

JOURNAL OF SCIENCE

PART A: ENGINEERING AND INNOVATION



Year | 2024

Volume | 11

Issue | 3

e-ISSN 2147-9542



| Owner |

on behalf of Gazi University
Rector
Prof. Dr.

Uğur ÜNAL

| Publishing Manager |

Prof. Dr.

Uğur GÖKMEN
Gazi University

| Chief Editor |

Prof. Dr.

Sema BİLGE OCAK
Gazi University

| Managing Editor |

Prof. Dr.

Mustafa Gürhan YALÇIN
Akdeniz University

| Assistant Editors |

Assoc. Prof. Dr.

Defne AKAY
Ankara University

Assoc. Prof. Dr.

Murat BİLEN
Gazi University



| Editorial Board |

Prof. Dr. Adem TATAROĞLU	Gazi University Physics
Prof. Dr. Adnan SÖZEN	Gazi University Energy Systems Engineering
Prof. Dr. Ali KESKİN	Çukurova University Automotive Engineering
Prof. Dr. Ali Osman SOLAK	Ankara University Chemistry
Prof. Dr. Alper BÜYÜKKARAGÖZ	Gazi University Civil Engineering
Prof. Dr. Atilla BIYIKOĞLU	Gazi University Mechanical Engineering
Prof. Dr. Aynur KAZAZ	Akdeniz University Civil Engineering
Prof. Dr. Cevriye GENCER	Gazi University Industrial Engineering
Prof. Dr. Çağlayan AÇIKGÖZ	Bilecik Şeyh Edebali University Chemical Engineering
Prof. Dr. Çetin ÇAKANYILDIRIM	Hitit University Chemical Engineering
Prof. Dr. Demet CANSARAN DUMAN	Ankara University The Institute of Biotechnology
Prof. Dr. Elif ORHAN	Gazi University Physics
Prof. Dr. Erdal IRMAK	Gazi University Electrical-Electronic Engineering
Prof. Dr. Fatih ÖZ	Atatürk University Food Engineering



| Editorial Board |

Prof. Dr. Feyza DİNÇER	Nevşehir Hacı Bektaş Veli University Geological Engineering
Prof. Dr. Gürhan İÇÖZ	Gazi University Mathematics
Prof. Dr. Hacer KARACAN	Gazi University Computer Engineering
Prof. Dr. Hüseyin Serdar YÜCESU	Gazi University Automotive Engineering
Prof. Dr. Meltem DOĞAN	Gazi University Chemical Engineering
Prof. Dr. Metin GÜRÜ	Gazi University Chemical Engineering
Prof. Dr. Mine TÜRKTAŞ	Gazi University Biology
Prof. Dr. Murat KAYA	Aksaray University Biotechnology and Nanotechnology
Prof. Dr. Nalan KABAY	Ege University Chemical Engineering
Prof. Dr. Nazife ASLAN	Ankara Hacı Bayram Veli University Chemistry
Prof. Dr. Niyazi Uğur KOÇKAL	Akdeniz University Civil Engineering
Prof. Dr. Nuran AY	Eskişehir Technical University Materials Science and Engineering
Prof. Dr. Nurdane İLBEYLİ	Akdeniz University Geological Engineering
Prof. Dr. Nursel AKÇAM	Gazi University Electrical-Electronic Engineering



| Editorial Board |

Prof. Dr. Ömer ŞAHİN	İstanbul Technical University Chemical Engineering
Prof. Dr. Selim ACAR	Gazi University Physics
Prof. Dr. Şükrü DURSUN	Konya Technical University Environmental Engineering
Prof. Dr. Veli ÇELİK	Ankara Yıldırım Beyazıt University Mechanical Engineering
Prof. Dr. Yılmaz ŞİMŞEK	Akdeniz University Mathematics
Prof. Dr. Yusuf URAS	Kahramanmaraş Sütçü İmam University Geological Engineering
Prof. Dr. Yücel ERCAN	TOBB University of Economics and Technology Mechanical Engineering
Prof. Dr. Zafer EVİS	Middle East Technical University Engineering Sciences
Assoc. Prof. Dr. Fusun YALÇIN	Akdeniz University Mathematics
Assoc. Prof. Dr. Yasemin LEVENTELİ	Akdeniz University Geological Engineering
Assist. Prof. Dr. Emine Şükran OKUDAN	Akdeniz University Basic Sciences Fisheries
Assist. Prof. Dr. Senai YALÇINKAYA	Marmara University Mechanical Engineering
Dr. Murat AKIN	Gazi University Computer Technologies
Dr. Silver GÜNEŞ	Gazi University Chemical Engineering



| Foreign Editorial Advisory Board |

Prof. Dr. Abdelmejid BAYAD	Université d'Évry Val d'Essonne FRANCE
Prof. Dr. Ali Behcet ALPAT	Istituto Nazionale di Fisica Nucleare (INFN) ITALY
Prof. Dr. Azra SPAGO	Dzermal Bijedic University of Mostar BOSNIA AND HERZEGOVINA
Prof. Dr. Bektay YERKIN	Satbayev University KAZAKHSTAN
Prof. Dr. Burçin BAYRAM	Miami University USA
Prof. Dr. Daeyeoul KIM	Jeonbuk National University SOUTH KOREA
Prof. Dr. Elvan AKIN	Missouri University of Science and Technology USA
Prof. Dr. Filiz DİK	Rockford University USA
Prof. Dr. Homer RAHNEJAT	Loughborough University UNITED KINGDOM
Prof. Dr. Loksha VEERABHADRIAH	Vijayanagara Sri Krishnadevaraya University INDIA
Prof. Dr. Mehmet DİK	Rockford University USA
Prof. Dr. Nedim SULJIC	University of Tuzla BOSNIA AND HERZEGOVINA
Prof. Dr. Rob DWYER-JOYCE	The University of Sheffield UNITED KINGDOM
Prof. Dr. Snezana KOMATINA	University Novi Sad SERBIA
Prof. Dr. Toni NIKOLIC	University Džermal Bijedić Mostar BOSNIA AND HERZEGOVINA
Prof. Dr. Turysbekova Gaukhar SEYTKHANOVNA	Satbayev University KAZAKHSTAN
Assoc. Prof. Dr. Burcin ŞİMŞEK	Associate Director Biostatistics (Oncology) at Bristol Myers Squibb USA
Assist. Prof. Dr. Alisa BABAJIC	University of Tuzla BOSNIA AND HERZEGOVINA
Dr. Daniel Ganyi NYAMSARI	Mining Company Researcher CAMEROON



| English Language Editors |

Lecturer
Gizem AÇELYA AYKAN Gazi University
School of Foreign Languages

Lecturer
Tuğçe BÜYÜKBAYRAM Gazi University
School of Foreign Languages

| Technical Editors |

Research Assistant Dr.
Fatih UÇAR Akdeniz University

Research Assistant
Özge ÖZER ATAĞOĞLU Akdeniz University



| Correspondence Address |

Gazi University Graduate School of Natural and Applied Sciences
Emniyet Neighborhood, Bandırma Avenue No:6/20B, 06560,
Yenimahalle - ANKARA
B Block, Auxiliary Building

| e-mail |

gujsa06@gmail.com

| web page |

<https://dergipark.org.tr/tr/pub/gujsa>

Gazi University Journal of Science Part A: Engineering and Innovation
is a peer-reviewed journal.



| INDEXING |



| ACCESSIBILITY |



This work are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



| CONTENTS |

Page | Articles

A Hybrid MADM Approach Based on Simple Additive Weighting and TOPSIS: An Application on Comparison of Innovation Performances of the EU Countries

419-430

F. GÖKTAŞ 

Statistics

Research Article

[10.54287/guj.1474940](https://doi.org/10.54287/guj.1474940)

A Study on the Evaluation of Image-Guided Treatment Use in Radiotherapy Centers in Türkiye

431-442

A. M. ŞENİŞİK 

Physics

Research Article

[10.54287/guj.1442069](https://doi.org/10.54287/guj.1442069)

Current Feedback Operation Amplifier Based on Floating Passive and Active Inductors in Filter Applications

443-450

H. DEMİREL 

Electrical & Electronics Engineering

Research Article

[10.54287/guj.1496604](https://doi.org/10.54287/guj.1496604)

Abrasion Behavior of Natural Aggregates and Slags in Turkey Aided by Micro-Deval Test

451-462

İ. GÖKALP  V. E. UZ 

Civil Engineering

Research Article

[10.54287/guj.1521679](https://doi.org/10.54287/guj.1521679)

| CONTENTS |

Page | Articles

Modeling Cardiovascular Flow with Artificial Viscosity: Analyzing Navier-Stokes Solutions and Simulating Cardiovascular Diseases

463-480

H. KARADAVUT  G. HAÇAT  A. ÇIBIK 

Mathematics

Research Article

[10.54287/guj.1485920](https://doi.org/10.54287/guj.1485920)

Comparing Machine Learning Algorithms for Rice Yield Prediction in Adamawa and Cross Rivers States of Nigeria

481-496

J. A. INGIO  A. S. NSANG  A. IORLIAM 

Information and Computing Sciences

Research Article

[10.54287/guj.1503494](https://doi.org/10.54287/guj.1503494)

TB-SMGAN: A GAN Based Hybrid Data Augmentation Framework on Chest X-ray Images and Reports

497-506

H. B. ÖZFİDAN  M. U. ŞİMŞEK 

Information and Computing Sciences

Research Article

[10.54287/guj.1501098](https://doi.org/10.54287/guj.1501098)

Development of a P(L-D,L)LA Foam as a Dura Substitute and Its In Vitro Evaluation

507-517

D. YÜCEL 

Biomedical Engineering

Research Article

[10.54287/guj.1514851](https://doi.org/10.54287/guj.1514851)



| CONTENTS |

Page | Articles

Neutral Benchmarking of Survival Models in Health Sciences: Comparative Study of Classical and Machine Learning Techniques

518-534 S. ABUBAKARI  F. KARAMAN 

Statistics

Research Article

[10.54287/guj.1505905](https://doi.org/10.54287/guj.1505905)

Further Results for Hermite-Based Milne-Thomson Type Fubini Polynomials with Trigonometric Functions

535-545 N. KILAR 

Mathematics

Research Article

[10.54287/guj.1546375](https://doi.org/10.54287/guj.1546375)

Drone Detection Performance Evaluation via Real Experiments with Additional Synthetic Darkness

546-562 F. ORUÇ  H. B. YILMAZ 

Information and Computing Sciences

Research Article

[10.54287/guj.1526979](https://doi.org/10.54287/guj.1526979)

Super Twisting Sliding Mode Control of Four-Phase Interleaved Boost Converter

563-576 V. YENİL  S. ÖZDEMİR  Z. ORTATEPE 

Electronics, Sensors and Digital Hardware

Research Article

[10.54287/guj.1529271](https://doi.org/10.54287/guj.1529271)

| CONTENTS |

Page | Articles

Investigation of the Effect of Hexagonal Boron Nitride Addition on the Mechanical Properties of Flax Fiber-Reinforced Composite Materials

577-588

A. ERYILMAZ  H. Y. ÜNAL  Y. PEKBEY 

Metallurgical and Materials Engineering

Research Article

[10.54287/guj.1533769](https://doi.org/10.54287/guj.1533769)

Kinetics of Colour, Clarity Changes and HMF Formation in Pear Juice Concentrate During Storage

589-597

F. KARADENİZ  B. IŞIK  S. KAYA  O. ASLANALI  F. MİDİLLİ 

Food Engineering

Research Article

[10.54287/guj.1529814](https://doi.org/10.54287/guj.1529814)

GIS-based AHP and MCDA Modeling for Cropland Suitability Analysis: A Bibliometric Analysis

598-621

D. C. WIJESINGHE 

Geoinformatics

Research Article

[10.54287/guj.1510527](https://doi.org/10.54287/guj.1510527)

Comparative Performance Analysis of Selected Machine Learning Algorithms and the Stacking Ensemble Method for Prediction of the Type II Diabetes Disease

622-646

N. A. ZOAKAH  A. S. NSANG  A. A. AJIBESIN  A. I. ZOAKAH 

Information and Computing Sciences

Research Article

[10.54287/guj.1531997](https://doi.org/10.54287/guj.1531997)



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1474940>

A Hybrid MADM Approach Based on Simple Additive Weighting and TOPSIS: An Application on Comparison of Innovation Performances of the EU Countries

Furkan GÖKTAŞ^{1*} ¹ Karabük University, Management Faculty, Department of Business, Karabük, Türkiye

Keywords	Abstract
Convex Optimization Innovation Multi-Attribute Decision-Making Simple Additive Weighting TOPSIS	This study aims to objectively compare the long-term innovation performances of the EU countries. In this context, we propose a hybrid multi-attribute decision-making (MADM) approach combining Simple Additive Weighting (SAW) and TOPSIS. The proposed approach (CST) uses the alternatives' quadratic utility functions considering the weighted sum value and distance to the positive ideal solution. It also uniquely determines the criteria weight vector using a strictly concave maximization problem. Using the Summary Innovation Index (SII) data for the 2016-2023 period, CST reveals that Sweden, Denmark, and the Netherlands are in the first three ranks. In contrast, Romania, Bulgaria, and Latvia are in the last three ranks.

Cite

Göktaş, F. (2024). A Hybrid MADM Approach Based on Simple Additive Weighting and TOPSIS: An Application on Comparison of Innovation Performances of the EU Countries. *GU J Sci, Part A, 11(3)*, 419-430. doi:10.54287/guj.1474940

Author ID (ORCID Number)

0000-0001-9291-3912 Furkan GÖKTAŞ

Article Process

Submission Date	28.04.2024
Revision Date	05.06.2024
Accepted Date	04.07.2024
Published Date	04.09.2024

1. INTRODUCTION

Innovation is the word of our age. An original or substantially upgraded product offered to the market or the initiation within a business of an original or substantially upgraded process is called innovation. That is, two main types of innovation are product innovation and process innovation. Product innovation is related to goods or services. Process innovation may have organizational or marketing aspects (Eurostat, 2023). European Innovation Scoreboard (EIS) can measure countries' innovation performances. EIS has four dimensions, twelve sub-dimensions, and thirty-two indicators. Based on EIS, selected countries have a total innovation score called the Summary Innovation Index (SII). SII is an important tool for monitoring and comparing countries' innovation performances and helps evaluate the effectiveness of the EU's innovation policies (European Commission, 2023).

Comparing the innovation performances of countries is of great importance in evaluating and improving the effectiveness of policies. These comparisons help countries identify their strengths and weaknesses and make strategic decisions accordingly (Anderson & Stejskal, 2019). Although it is a very important subject, few studies have compared the countries with respect to their innovation performances. This is because there are already some important and reliable innovation indices, such as the Global Innovation Index, Summary Innovation Index, Bloomberg Innovation Index, etc. Kaynak et al. (2017) compare the EU candidate countries using TOPSIS. Brodny et al. (2023) compare the EU countries using EDAS. Ozkaya et al. (2021) compare many countries using different methods. Namazi and Mohammadi (2018) compare many countries using DEA-based TOPSIS. Ecer and Aycin (2023) compare G7 countries using different methods. Aytekin et al. (2022) compare the EU member and candidate countries using DEA-EATWIOS. Satı (2024) compares the EU

*Corresponding Author, e-mail: furkangoktas@karabuk.edu.tr

member and candidate countries using TOPSIS. Do Carmo Silva et al. (2020) compare many countries using TOPSIS and PROMETHEE. Erdin and Caglar (2023) compare the OECD countries using a DEA-based approach. Murat (2020) compares the OECD countries using DEA. Jeon et al. (2022) and Selvaraj and Jeon (2021) compare the OECD countries using a fuzzy approach. Anderson and Stejskal (2019) and Jurickova et al. (2019) compare the EU member countries using DEA. Kabadurmuş and Karaman Kabadurmuş (2019) compare the Eastern Europe and Central Asia countries using TOPSIS. Chen et al. (2011) compare many countries using DEA.

SAW is one of the most used MADM methods (Taherdoost, 2023). Similarly, TOPSIS is widely used since it is an understandable method with a strong mathematical structure. On the other hand, the issue of determining criteria weights is a matter of criticism for TOPSIS. Because the results obtained with it depend significantly on the criteria weights (Bouslah et al., 2023). One of the motivations of this study is to combine SAW and TOPSIS with a hybrid approach, which also determines the criteria weights inherently. Thus, we propose a hybrid MADM approach combining SAW and TOPSIS. We call it CST. Another motivation for this study is to provide a different perspective for innovation-based analyses. Thus, this study aims to compare the long-term innovation performances of the EU countries using an objective approach. Due to its objectivity, we use CST to achieve this aim. This study differs from the above studies and existing innovation index reports since it directly makes a long-term analysis using SII's yearly data, whereas the studies in the literature consider many criteria using a specific year's data for each analysis. This study also differs from them by analyzing the annual progress of the countries. Thus, the originality of the paper comes from the long-term analysis and the annual progress analysis in addition to the proposed MADM approach.

The rest of the paper is organized as follows. Section 2 gives the steps of SAW and TOPSIS. It also presents the theory of CST. Section 3 illustrates CST by comparing the innovation performances of the EU countries for the 2016-2023 period based on SII (EIS, 2023). It also presents and discusses the results of the innovation performance comparisons. Section 4 concludes the paper.

2. MATERIAL AND METHOD

Many normalization techniques can be used in SAW. We prefer the vector normalization. Then, the steps of SAW can be given as follows (Vafaei et al., 2022; Taherdoost, 2023).

Step 1: The decision matrix $A_{n \times m}=(a_{ij})$ is formed, where a_{ij} is the positive value of the i^{th} alternative for the j^{th} criterion, n is the number of alternatives, and m is the number of criteria.

Step 2: The normalized decision matrix $B_{n \times m}=(b_{ij})$ is formed using (1a) for the benefit criterion and (1b) for the cost criterion (Acuña-Soto et al., 2021).

$$b_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^n a_{ij}^2}} \quad (1a)$$

$$b_{ij} = \frac{1/a_{ij}}{\sqrt{\sum_{i=1}^n 1/a_{ij}^2}} \quad (1b)$$

Step 3: The criteria weight vector $w=(w_j)$ is determined using an approach.

Step 4: The alternatives' weighted sum vector $p=(p_i)$ is determined using the equation $p=Bw$.

Step 5: The alternatives are ranked in descending order using the p_i values.

Different normalization techniques can be used in TOPSIS, which uses Euclidean distance (Acuña-Soto et al., 2021). We prefer (1a)-(1b) for normalization. Then, the first three steps of TOPSIS are identical to the first three steps of SAW. The other steps of TOPSIS are as follows (Acuña-Soto et al., 2021; Bouslah et al., 2023).

Step 4: Let the maximum element of the j^{th} row of B is β_j , whereas the minimum element of the j^{th} row of B is α_j . The positive ideal solution $x=(x_j)$ is determined using the equation $x_j=w_j\beta_j$ for all j . The negative ideal solution $y=(y_j)$ is determined using the equation $y_j=w_j\alpha_j$ for all j .

Step 5: The i^{th} alternative's relative proximity to the ideal solution is calculated as in (2). The numerator term equals the alternative's distance to the negative ideal solution. The denominator term equals the sum of the alternative's distance to the negative ideal solution and the alternative's distance to the positive ideal solution.

$$s_i = \frac{\sqrt{\sum_{j=1}^m w_j^2 (b_{ij} - \alpha_j)^2}}{\sqrt{\sum_{j=1}^m w_j^2 (b_{ij} - \alpha_j)^2} + \sqrt{\sum_{j=1}^m w_j^2 (\beta_j - b_{ij})^2}} \quad (2)$$

Step 6: The alternatives are ranked in descending order using the s_i values.

To combine SAW and TOPSIS, we assume that the quadratic utility function of the i^{th} alternative is as in (3), which considers two objectives: higher weighted sum value and lower distance to the positive ideal solution. We ignore the distance to the negative ideal solution since it is not a concave function. On the other hand, these distances generally give similar rankings. Thus, we believe that the loss of information is ignorable.

$$U_i(w) = \sum_{j=1}^m w_j b_{ij} - m \sum_{j=1}^m w_j^2 (\beta_j - b_{ij})^2 \quad (3)$$

Then, we form the following concave maximization problem, which determines the criteria weight vector (w) by maximizing the worst alternative's utility. Clearly, y is a variable associated with the worst alternative.

$$\begin{aligned} & \max y \\ & \text{s.t. } y - \left(\sum_{j=1}^m w_j b_{ij} - m \sum_{j=1}^m w_j^2 (\beta_j - b_{ij})^2 \right) \leq 0, \forall i \\ & \sum_{j=1}^m w_j = 1 \\ & w_j \geq 0, \forall j \end{aligned} \quad (4)$$

Since (4) is a concave maximization problem, its optimal solution set is convex. The optimal solution closest to the origin can be found approximately using Tikhonov's regularized problem, which gives a unique solution (Boyd & Vandenberghe, 2004; Beck & Sabach, 2014). Thus, we use Tikhonov's regularized problem given in (5) instead of (4), where ε is a positive constant close to 0. We take $\varepsilon=2^{-23}$ in this study. Since (5) is a concave maximization problem, it can be solved with MATLAB software CVX (Grant & Boyd, 2008).

$$\begin{aligned} & \max y - 0.5\varepsilon \sum_{j=1}^m w_j^2 \\ & \text{s.t. } y - \left(\sum_{j=1}^m w_j b_{ij} - m \sum_{j=1}^m w_j^2 (\beta_j - b_{ij})^2 \right) \leq 0, \forall i \\ & \sum_{j=1}^m w_j = 1 \\ & w_j \geq 0, \forall j \end{aligned} \quad (5)$$

The first two steps of CST are identical to the first two steps of SAW. The other steps of CST are as follows.

Step 3: The squared difference matrix $D=(d_{ij})$ is formed using the equation $d_{ij}=(\beta_j-b_{ij})^2$ for all i and j , where β_j is the maximum element of the j^{th} row of B.

Remark: When the criteria weights are equal, the i^{th} alternative's utility is as in (6).

$$U_i = \frac{\sum_{j=1}^m (b_{ij} - d_{ij})}{m} \quad (6)$$

Step 4: (5) is solved using suitable software to determine the criteria weight vector (w) uniquely. (It can also be used to find the criteria weights in SAW and TOPSIS.) The CVX code for (5) is given in (7).

```

cvx_solver mosek
cvx_begin
variables w(m) y;
maximize(y - 0.5 * eps("single") * transpose(w) * w);
subject to
    y * ones(n,1) - (B * w - m * D * (w.^2)) <= zeros(n,1);
    ones(1,m) * w == 1;
    w >= zeros(m,1);
cvx_end

```

(7)

Step 5: (3) is calculated for each alternative. Then, the alternatives are ranked in descending order using them.

3. RESULTS AND DISCUSSION

In this section, we use SAW, TOPSIS, and CST to compare the innovation performances of the EU countries for the 2016-2023 period based on SII. Any country's innovation performance increases with the increase in SII. We use SII for each year as a criterion, whereas the EU countries are the alternatives. We apply CST with the following steps.

Step 1: We form the decision matrix $A=(a_{ij})$ as in Table 1 using SII data of the EU countries for the 2016-2023 period.

Table 1. The decision matrix

	2023	2022	2021	2020	2019	2018	2017	2016
Austria	130.00	129.40	124.58	124.01	124.31	122.97	123.96	123.60
Belgium	136.44	136.92	136.00	127.60	129.26	125.94	123.69	122.35
Bulgaria	50.63	44.58	45.07	46.90	46.31	44.77	45.90	46.28
South Cyprus	114.29	113.70	108.41	88.70	84.54	82.23	79.09	78.67
Czechia	102.73	92.27	89.10	85.73	83.24	82.11	81.71	81.75
Germany	127.79	129.04	127.06	122.07	121.61	121.12	120.41	120.16
Denmark	149.24	146.47	144.53	140.06	137.79	134.16	134.52	133.25
Estonia	107.00	118.39	114.10	98.78	96.41	77.84	78.83	77.72
Greece	86.22	85.73	80.69	75.21	72.19	64.89	64.75	63.98
Spain	96.80	92.84	91.77	91.94	90.47	89.02	88.16	87.08
Finland	145.63	141.68	137.66	134.42	133.03	125.83	125.66	127.32
France	114.21	115.55	113.12	114.71	114.24	116.43	115.57	115.84

Table 2. continued

	2023	2022	2021	2020	2019	2018	2017	2016
Croatia	75.44	74.30	69.83	61.91	60.06	55.79	61.45	60.66
Hungary	76.31	73.87	70.42	67.96	66.92	68.52	68.48	68.57
Ireland	125.61	121.68	118.40	120.78	123.23	125.10	123.88	123.32
Italy	97.99	103.63	101.96	92.94	89.87	84.40	83.56	82.37
Lithuania	90.92	87.18	82.06	84.09	81.71	74.13	73.89	74.19
Luxembourg	127.15	126.17	125.87	128.45	129.71	128.47	128.76	128.68
Latvia	56.97	56.37	54.68	55.95	56.29	53.32	52.98	53.44
Malta	93.11	95.70	98.98	95.96	94.17	91.06	84.25	82.23
Netherlands	139.56	138.55	135.41	137.81	137.17	133.66	132.34	130.96
Poland	68.09	62.87	60.98	58.34	58.87	56.54	56.43	54.79
Portugal	92.88	89.90	87.53	96.97	93.75	83.91	85.02	85.15
Romania	35.85	38.27	37.36	33.79	31.91	34.51	35.42	34.41
Sweden	145.92	147.31	145.16	138.19	138.18	138.34	137.66	135.49
Slovenia	103.10	99.84	96.40	91.53	92.40	97.38	98.08	100.17
Slovakia	71.18	66.97	65.63	66.69	65.99	63.23	66.08	64.81

Step 2: We form the normalized decision matrix $B=(b_{ij})$ as in Table 2 using (1a).

Table 2. The normalized decision matrix

	2023	2022	2021	2020	2019	2018	2017	2016
Austria	0.2346	0.2356	0.2322	0.2382	0.2408	0.2448	0.2477	0.2482
Belgium	0.2462	0.2493	0.2535	0.2450	0.2504	0.2507	0.2471	0.2457
Bulgaria	0.0914	0.0812	0.0840	0.0901	0.0897	0.0891	0.0917	0.0929
South Cyprus	0.2063	0.2070	0.2021	0.1703	0.1637	0.1637	0.1580	0.1580
Czechia	0.1854	0.1680	0.1661	0.1646	0.1612	0.1634	0.1633	0.1641
Germany	0.2306	0.2350	0.2368	0.2344	0.2355	0.2411	0.2406	0.2413
Denmark	0.2693	0.2667	0.2694	0.2690	0.2669	0.2671	0.2688	0.2676
Estonia	0.1931	0.2155	0.2127	0.1897	0.1867	0.1550	0.1575	0.1560
Greece	0.1556	0.1561	0.1504	0.1444	0.1398	0.1292	0.1294	0.1285
Spain	0.1747	0.1690	0.1710	0.1766	0.1752	0.1772	0.1761	0.1749
Finland	0.2628	0.2580	0.2566	0.2582	0.2577	0.2505	0.2511	0.2556
France	0.2061	0.2104	0.2108	0.2203	0.2213	0.2318	0.2309	0.2326
Croatia	0.1361	0.1353	0.1301	0.1189	0.1163	0.1111	0.1228	0.1218
Hungary	0.1377	0.1345	0.1312	0.1305	0.1296	0.1364	0.1368	0.1377
Ireland	0.2267	0.2215	0.2207	0.2320	0.2387	0.2490	0.2475	0.2476
Italy	0.1768	0.1887	0.1900	0.1785	0.1741	0.1680	0.1670	0.1654
Lithuania	0.1641	0.1587	0.1529	0.1615	0.1583	0.1476	0.1476	0.1490
Luxembourg	0.2295	0.2297	0.2346	0.2467	0.2512	0.2557	0.2573	0.2584

Table 2. continued

	2023	2022	2021	2020	2019	2018	2017	2016
Latvia	0.1028	0.1026	0.1019	0.1075	0.1090	0.1061	0.1059	0.1073
Malta	0.1680	0.1742	0.1845	0.1843	0.1824	0.1813	0.1683	0.1651
Netherlands	0.2518	0.2523	0.2524	0.2647	0.2657	0.2661	0.2644	0.2630
Poland	0.1229	0.1145	0.1136	0.1120	0.1140	0.1125	0.1127	0.1100
Portugal	0.1676	0.1637	0.1631	0.1862	0.1816	0.1670	0.1699	0.1710
Romania	0.0647	0.0697	0.0696	0.0649	0.0618	0.0687	0.0708	0.0691
Sweden	0.2633	0.2682	0.2705	0.2654	0.2676	0.2754	0.2751	0.2720
Slovenia	0.1861	0.1818	0.1797	0.1758	0.1790	0.1938	0.1960	0.2011
Slovakia	0.1285	0.1219	0.1223	0.1281	0.1278	0.1259	0.1320	0.1301

Step 3: We form the $\beta=(\beta_j)$ vector as $(0.2693, 0.2682, 0.2705, 0.2690, 0.2676, 0.2754, 0.2751, 0.2720)^T$. Then, we determine the squared difference matrix $D=(d_{ij})$ matrix as in Table 3 using the equation $d_{ij}=(\beta_j-b_{ij})^2$.

Table 3. The squared difference matrix

	2023	2022	2021	2020	2019	2018	2017	2016
Austria	0.0012	0.0011	0.0015	0.0010	0.0007	0.0009	0.0007	0.0006
Belgium	0.0005	0.0004	0.0003	0.0006	0.0003	0.0006	0.0008	0.0007
Bulgaria	0.0317	0.0350	0.0348	0.0320	0.0317	0.0347	0.0336	0.0321
South Cyprus	0.0040	0.0037	0.0047	0.0097	0.0108	0.0125	0.0137	0.0130
Czechia	0.0070	0.0100	0.0109	0.0109	0.0113	0.0125	0.0125	0.0116
Germany	0.0015	0.0011	0.0011	0.0012	0.0010	0.0012	0.0012	0.0009
Denmark	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
Estonia	0.0058	0.0028	0.0033	0.0063	0.0065	0.0145	0.0138	0.0135
Greece	0.0129	0.0126	0.0144	0.0155	0.0163	0.0214	0.0212	0.0206
Spain	0.0090	0.0098	0.0099	0.0085	0.0085	0.0096	0.0098	0.0094
Finland	0.0000	0.0001	0.0002	0.0001	0.0001	0.0006	0.0006	0.0003
France	0.0040	0.0033	0.0036	0.0024	0.0022	0.0019	0.0019	0.0016
Croatia	0.0177	0.0177	0.0197	0.0225	0.0229	0.0270	0.0232	0.0226
Hungary	0.0173	0.0179	0.0194	0.0192	0.0190	0.0193	0.0191	0.0181
Ireland	0.0018	0.0022	0.0025	0.0014	0.0008	0.0007	0.0008	0.0006
Italy	0.0086	0.0063	0.0065	0.0082	0.0088	0.0115	0.0117	0.0114
Lithuania	0.0111	0.0120	0.0138	0.0116	0.0120	0.0163	0.0162	0.0151
Luxembourg	0.0016	0.0015	0.0013	0.0005	0.0003	0.0004	0.0003	0.0002
Latvia	0.0277	0.0274	0.0284	0.0261	0.0252	0.0286	0.0286	0.0271
Malta	0.0103	0.0088	0.0074	0.0072	0.0073	0.0089	0.0114	0.0114
Netherlands	0.0003	0.0003	0.0003	0.0000	0.0000	0.0001	0.0001	0.0001
Poland	0.0214	0.0236	0.0246	0.0246	0.0236	0.0265	0.0263	0.0263
Portugal	0.0103	0.0109	0.0115	0.0068	0.0074	0.0117	0.0111	0.0102

Table 3. continued

	2023	2022	2021	2020	2019	2018	2017	2016
Romania	0.0419	0.0394	0.0404	0.0416	0.0424	0.0427	0.0417	0.0412
Sweden	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Slovenia	0.0069	0.0075	0.0083	0.0087	0.0079	0.0066	0.0063	0.0050
Slovakia	0.0198	0.0214	0.0220	0.0199	0.0195	0.0224	0.0205	0.0201

Step 4: We uniquely find $w=(0.1195, 0.1348, 0.1315, 0.1204, 0.1138, 0.1229, 0.1289, 0.1281)^T$ using the CVX code given in (7). That is, CST objectively determines the criteria weights as follows. The 2023's weight is 0.1195, the 2022's weight is 0.1348, the 2021's weight is 0.1315, the 2020's weight is 0.1204, the 2019's weight is 0.1138, the 2018's weight is 0.1229, the 2017's weight is 0.1289, and the 2016's weight is 0.1281.

Step 5: We find the EU countries' utilities as in Table 4 using the quadratic utility function given in (3) and the criteria weight vector (w) given in Step 4. We also rank them in descending order. We find the first three ranks as Sweden, Denmark, and the Netherlands, respectively, whereas Romania is at the last rank.

Table 4. The innovation performances of the EU countries for the 2016-2023 period

	SAW		TOPSIS		CST	
	p_i	Rank	s_i	Rank	U_i	Rank
Austria	0.2402	7	0.8474	7	0.2393	7
Belgium	0.2485	5	0.8883	5	0.2480	5
Bulgaria	0.0886	26	0.1057	26	0.0553	26
South Cyprus	0.1790	13	0.5464	13	0.1701	13
Czechia	0.1670	18	0.4883	18	0.1561	18
Germany	0.2370	8	0.8328	8	0.2358	8
Denmark	0.2681	2	0.9800	2	0.2681	2
Estonia	0.1835	12	0.5662	12	0.1752	12
Greece	0.1417	20	0.3665	20	0.1248	20
Spain	0.1743	16	0.5239	16	0.1648	16
Finland	0.2562	4	0.9216	4	0.2560	4
France	0.2205	10	0.7494	10	0.2179	10
Croatia	0.1243	23	0.2810	23	0.1026	23
Hungary	0.1344	21	0.3284	21	0.1157	21
Ireland	0.2353	9	0.8191	9	0.2339	9
Italy	0.1762	14	0.5341	14	0.1671	14
Lithuania	0.1548	19	0.4286	19	0.1412	19
Luxembourg	0.2452	6	0.8648	6	0.2445	6
Latvia	0.1053	25	0.1858	25	0.0778	25
Malta	0.1759	15	0.5318	15	0.1668	15
Netherlands	0.2599	3	0.9390	3	0.2597	3
Poland	0.1140	24	0.2286	24	0.0892	24

Table 4. continued

	SAW		TOPSIS		CST	
	p_i	Rank	s_i	Rank	U_i	Rank
Portugal	0.1710	17	0.5069	17	0.1608	17
Romania	0.0675	27	0.0000	27	0.0261	27
Sweden	0.2698	1	0.9884	1	0.2698	1
Slovenia	0.1868	11	0.5860	11	0.1796	11
Slovakia	0.1270	22	0.2921	22	0.1062	22

Table 4 also presents the results obtained with SAW and TOPSIS when the criteria weight vector (w) is found with (5). We see that CST determines the same rank as SAW and TOPSIS. We also compare the annual progress of the EU countries using these methods. Criterion 1 equals SII 2023 minus SII 2022. Criterion 2, 3, 4, 5, 6, and 7 are determined similarly. Then, the criteria weights found with CST are 0.1699, 0.3058, 0.0849, 0.1009, 0.0761, 0.0930, and 0.1692, respectively. Furthermore, we find the annual progress rankings as in Table 5. South Cyprus is at the first rank based on the annual progress, whereas Luxembourg is at the last rank. The Spearman rank correlation between the CST and SAW results equals 0.9206, whereas the Spearman rank correlation between the CST and TOPSIS results equals 0.9640. CST incorporates SAW and TOPSIS using convex optimization and worst-case analysis. Thus, as in this study, the CST results may be similar to their results when the same criteria weights are used.

Table 5. The annual progress rankings of the EU countries for the 2017-2023 period

	SAW		TOPSIS		CST	
	p_i	Rank	s_i	Rank	U_i	Rank
Austria	0.1000	13	0.5872	9	-0.0407	12
Belgium	0.0859	17	0.5411	17	-0.0714	17
Bulgaria	0.0230	25	0.4882	23	-0.1869	24
South Cyprus	0.2426	1	0.6881	1	0.1609	1
Czechia	0.1815	3	0.6705	2	0.0892	3
Germany	0.0599	21	0.5145	19	-0.1054	19
Denmark	0.1371	5	0.6087	7	0.0267	5
Estonia	0.1293	8	0.5186	18	-0.1212	20
Greece	0.1877	2	0.6547	3	0.0926	2
Spain	0.1093	11	0.5840	10	-0.0175	8
Finland	0.1015	12	0.5697	14	-0.0582	14
France	0.0298	23	0.4927	22	-0.1583	22
Croatia	0.1182	9	0.5830	11	-0.0355	10
Hungary	0.0946	15	0.5799	12	-0.0384	11
Ireland	0.0841	18	0.5893	8	-0.0518	13
Italy	0.0988	14	0.5144	20	-0.0855	18
Lithuania	0.1595	4	0.6391	4	0.0505	4
Luxembourg	-0.0027	27	0.4537	27	-0.2162	27
Latvia	0.0274	24	0.4879	24	-0.1584	23

Table 5. continued

	SAW		TOPSIS		CST	
	p_i	Rank	s_i	Rank	U_i	Rank
Malta	0.0707	20	0.4787	26	-0.2162	25
Netherlands	0.1298	7	0.6132	6	0.0176	7
Poland	0.1337	6	0.6242	5	0.0249	6
Portugal	0.0911	16	0.5555	16	-0.0617	15
Romania	0.0414	22	0.4816	25	-0.1566	21
Sweden	0.1171	10	0.5773	13	-0.0260	9
Slovenia	0.0126	26	0.5005	21	-0.2162	26
Slovakia	0.0830	19	0.5600	15	-0.0681	16

Brodny et al. (2023) compare the EU countries' innovation performances for the 2013-2020 period by making a separate EDAS analysis for each year in this period. They find that Luxembourg, Sweden, Finland, Denmark, Germany, and the Netherlands have the highest performances, whereas Poland, Latvia, Romania, and Bulgaria have the lowest performances. We find similar results, except that Belgium and Austria have higher innovation performances than Germany in our results. The less similar results are also given in the literature. Based on the Global Innovation Index 2020 data, Aytakin et al. (2022) find that the Netherlands, Germany, and Sweden have the highest performances, whereas Lithuania and Greece have the lowest performances. Using the Global Innovation Index 2021 data and many other data, Satı (2024) finds that Austria, Denmark, and Germany have the highest performances, whereas Croatia has the lowest performance. Jurickova et al. (2019) compare the EU countries' innovation performances for the 2015-2016 period by making a separate DEA analysis for each year in this period. They find efficient countries such as Cyprus, Luxembourg, Malta, and Romania. Anderson and Stejskal (2019) compare the EU countries using the European Innovation Scoreboard data collated in 2018. They find that many countries (including Luxembourg, the Netherlands, Ireland, Bulgaria, Romania, etc.) are efficient decision-making units.

Table 4 and Table 5 give some important results about the EU countries' long-term innovation performances. We use a further analysis with the following procedure to integrate and discuss these results.

1. The average U_i value in Table 4 and the average U_i value in Table 5 are calculated. We show them with the x and y values, respectively.
2. If a country's U_i value in Table 4 is lower (higher) than the x value i.e. $-x+U_i$ is negative (positive), then we call that this country has innovation performance below (above) the average.
3. If a country's U_i value in Table 5 is lower (higher) than the y value i.e. $-y+U_i$ is negative (positive), then we call that this country has annual progress below (above) the average.
4. Countries are divided into four categories.
 - If a country is below (above) the average with respect to these two different analyses, then we call it as a problematic (star) country.
 - If a country is above the average with respect to innovation performance and below the average with respect to annual progress, then we call it as a question mark country.
 - If a country is below the average with respect to innovation performance and above the average with respect to annual progress, then we call it as a climbing country.

Based on the above procedure and information given in Table 4 and Table 5, we divide the EU countries into four categories as in Table 6.

Table 6 shows that the star countries are Austria, Denmark, Finland, Ireland, Netherlands, and Sweden. Austria (Ireland) has the seventh (ninth) rank based on innovation performance, whereas the other four countries are in the first four ranks. This information is compatible with the fact that continuous improvement is necessary to stay at the top. The problematic countries are Bulgaria, Italy, Latvia, Malta, Portugal, Romania, and

Slovakia. Since both their innovation performances and annual progress are below the average, these countries should take immediate and serious actions to increase their innovation performances. In addition, the EU should support these actions. These countries could determine the star countries as the guiding countries. Belgium, Germany, Estonia, France, Luxembourg, and Slovenia are the question mark countries. They should deeply analyze themselves and then take necessary actions. The climbing countries are South Cyprus, Czechia, Greece, Spain, Croatia, Hungary, Lithuania, and Poland. These countries could surpass some of the question marks countries in the future if they continue their progress.

Table 6. Categorization of the EU countries based on further analysis

	$-x+U_i$	$-y+U_i$	Category
Austria	0.0669	0.0196	Star
Belgium	0.0757	-0.0111	Question Mark
Bulgaria	-0.1170	-0.1266	Problematic
South Cyprus	-0.0022	0.2212	Climbing
Czechia	-0.0162	0.1494	Climbing
Germany	0.0635	-0.0451	Question Mark
Denmark	0.0958	0.0870	Star
Estonia	0.0029	-0.0609	Question Mark
Greece	-0.0475	0.1529	Climbing
Spain	-0.0075	0.0428	Climbing
Finland	0.0837	0.0021	Star
France	0.0455	-0.0980	Question Mark
Croatia	-0.0697	0.0247	Climbing
Hungary	-0.0567	0.0219	Climbing
Ireland	0.0616	0.0085	Star
Italy	-0.0052	-0.0252	Problematic
Lithuania	-0.0311	0.1108	Climbing
Luxembourg	0.0722	-0.1559	Question Mark
Latvia	-0.0946	-0.0981	Problematic
Malta	-0.0055	-0.1559	Problematic
Netherlands	0.0874	0.0779	Star
Poland	-0.0831	0.0852	Climbing
Portugal	-0.0115	-0.0014	Problematic
Romania	-0.1462	-0.0963	Problematic
Sweden	0.0975	0.0343	Star
Slovenia	0.0073	-0.1559	Question Mark
Slovakia	-0.0661	-0.0078	Problematic

4. CONCLUSION

This study proposes a hybrid MADM approach called CST. Since it combines SAW and TOPSIS using convex optimization and worst-case analysis, it has strong properties of these methods. CST also has two superiorities. It objectively determines the criteria weight vector and conveys more information than SAW or TOPSIS. Its main limitation is that the information is restricted to the total information derived with SAW and TOPSIS. In

addition, it does not consider the distance to the negative ideal solution, unlike TOPSIS. Furthermore, it finds the approximate result since it uses Tikhonov's regularized problem instead of the original problem. Moreover, special software is needed for its implementation, unlike SAW and TOPSIS.

This study objectively compares the EU countries' innovation performances for the 2016-2023 period using CST. Sweden has the maximum innovation performance for this period, whereas Romania has the minimum innovation performance. This study also compares the annual progress of the EU countries in the 2017-2023 period. The annual progress is maximum for South Cyprus and minimum for Luxembourg in this period. Based on these separate analyses, we divide the EU countries into four categories. Denmark, Finland, Ireland, Netherlands, and Sweden are the star countries corresponding to the best category. Belgium, Germany, Estonia, France, Luxembourg, and Slovenia are the question mark countries corresponding to the second-best category. South Cyprus, Czechia, Greece, Spain, Croatia, Hungary, Lithuania, and Poland are the climbing countries corresponding to the second-worst category. Bulgaria, Italy, Latvia, Malta, Portugal, Romania, and Slovakia are the problematic countries corresponding to the worst category. The main limitation of this study is that we only compare the EU countries using the Summary Innovation Index (SII). Future research could increase the number of countries and used innovation indices. Furthermore, CST could be used for other MADM problems.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- Acuña-Soto, C., Liern, V., & Pérez-Gladish, B. (2021). Normalization in TOPSIS-based approaches with data of different nature: application to the ranking of mathematical videos. *Annals of Operations Research*, 296(1), 541-569. <https://doi.org/10.1007/s10479-018-2945-5>
- Anderson, H. J., & Stejskal, J. (2019). Diffusion efficiency of innovation among EU member states: a data envelopment analysis. *Economies*, 7(2), 34. <https://doi.org/10.3390/economies7020034>
- Aytekin, A., Ecer, F., Korucuk, S., & Karamaşa, Ç. (2022). Global innovation efficiency assessment of EU member and candidate countries via DEA-EATWIOS multi-criteria methodology. *Technology in Society*, 68, 101896. <https://doi.org/10.1016/j.techsoc.2022.101896>
- Beck, A., & Sabach, S. (2014). A first order method for finding minimal norm-like solutions of convex optimization problems. *Mathematical Programming*, 147(1), 25-46. <https://doi.org/10.1007/s10107-013-0708-2>
- Bouslah, K., Liern, V., Ouenniche, J., & Pérez-Gladish, B. (2023). Ranking firms based on their financial and diversity performance using multiple-stage unweighted TOPSIS. *International Transactions in Operational Research*, 30(5), 2485-2505. <https://doi.org/10.1111/itor.13143>
- Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Brodny, J., Tutak, M., Grebski, W., & Bindzár, P. (2023). Assessing the level of innovativeness of EU-27 countries and its relationship to economic, environmental, energy and social parameters. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(2), 100073. <https://doi.org/10.1016/j.joitmc.2023.100073>
- Chen, C.-P., Hu, J.-L., & Yang, C.-H. (2011). An international comparison of R&D efficiency of multiple innovative outputs: The role of the national innovation system. *Innovation*, 13(3), 341-360. <https://doi.org/10.5172/impp.2011.13.3.341>
- Do Carmo Silva, M., Gavião, L. O., Gomes, C. F. S., & Lima, G. B. A. (2020). Global Innovation Indicators analysed by multicriteria decision. *Brazilian Journal of Operations & Production Management*, 17(4), 1-17. <https://doi.org/10.14488/BJOPM.2020.040>
- Ecer, F., & Aycin, E. (2023). Novel comprehensive MEREC weighting-based score aggregation model for measuring innovation performance: The case of G7 countries. *Informatica*, 34(1), 53-83. <https://doi.org/10.15388/22-INFOR494>

- EIS (2023). European Innovation Scoreboard 2023, the Summary Innovation Index. (Accessed:07/03/2024) <https://projects.research-and-innovation.ec.europa.eu/en/statistics/performance-indicators/european-innovation-scoreboard/eis>
- Erdin, C., & Çağlar, M. (2023). National innovation efficiency: A DEA-based measurement of OECD countries. *International Journal of Innovation Science*, 15(3), 427-456. <https://doi.org/10.1108/IJIS-07-2021-0118>
- European Commission (2023). European Innovation Scoreboard 2023 methodology report. (Accessed:07/03/2024) https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/european-innovation-scoreboard_en
- Eurostat (2023). Glossary: Innovation. (Accessed:07/03/2024) <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Innovation>
- Grant, M. C., & Boyd, S. P. (2008). Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, 95-110. Springer, London. https://doi.org/10.1007/978-1-84800-155-8_7
- Jeon, J., Geetha, S., Kang, D., & Narayanamoorthy, S. (2022). Development of the evaluation model for national innovation capability. *Technology Analysis & Strategic Management*, 34(3), 335-348. <https://doi.org/10.1080/09537325.2021.1900561>
- Jurickova, E., Pilik, M., & Kwarteng, M. A. (2019). Efficiency measurement of national innovation systems of the European Union countries: DEA model application. *Journal of International Studies*, 12(4), 286-299. <https://doi.org/10.14254/2071-8330.2019/12-4/19>
- Kabadurmuş, Ö., & Karaman Kabadurmuş, F. N. (2019). Innovation in Eastern Europe & Central Asia: A Multi-Criteria Decision-Making Approach. *Business & Management Studies: An International Journal*, 7(3), 98-121. <https://doi.org/10.15295/bmij.v7i3.1234>
- Kaynak, S., Altuntas, S., & Dereli, T. (2017). Comparing the innovation performance of EU candidate countries: an entropy-based TOPSIS approach. *Economic Research-Ekonomska Istraživanja*, 30(1), 31-54. <http://doi.org/10.1080/1331677X.2016.1265895>
- Murat, D. (2020). The measurement of innovation performance in OECD countries. *Journal of Management & Economics Research*, 18(4), 209-2026. <http://doi.org/10.11611/yead.822303>
- Namazi, M., & Mohammadi, E. (2018). Natural resource dependence and economic growth: A TOPSIS/DEA analysis of innovation efficiency. *Resources Policy*, 59, 544-552. <https://doi.org/10.1016/j.resourpol.2018.09.015>
- Ozkaya, G., Timor, M., & Erdin, C. (2021). Science, technology and innovation policy indicators and comparisons of countries through a hybrid model of data mining and MCDM methods. *Sustainability*, 13(2), 694. <https://doi.org/10.3390/su13020694>
- Satı, Z. E. (2024). Comparison of the criteria affecting the digital innovation performance of the European Union (EU) member and candidate countries with the entropy weight-TOPSIS method and investigation of its importance for SMEs. *Technological Forecasting and Social Change*, 200, 123094. <https://doi.org/10.1016/j.techfore.2023.123094>
- Selvaraj, G., & Jeon, J. (2021). Assessment of national innovation capabilities of OECD countries using trapezoidal interval type-2 fuzzy ELECTRE III method. *Data Technologies and Applications*, 55(3), 400-429. <https://doi.org/10.1108/DTA-07-2020-0154>
- Taherdoost, H. (2023). Analysis of simple additive weighting method (SAW) as a multi-attribute decision-making technique: A step-by-step. *Journal of Management Science & Engineering Research*, 6(1), 21-24. <https://doi.org/10.30564/jmser.v6i1.5400>
- Vafaei, N., Ribeiro, R. A., & Camarinha-Matos, L. M. (2022). Assessing normalization techniques for simple additive weighting method. *Procedia Computer Science*, 199, 1229-1236. <https://doi.org/10.1016/j.procs.2022.01.156>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1442069>

A Study on the Evaluation of Image-Guided Treatment Use in Radiotherapy Centers in Türkiye

Ahmet Murat ŞENİŞİK^{1*} ¹ Altınbaş University, SHMYO Radiotherapy Program, Istanbul, Türkiye

Keywords	Abstract
Radiotherapy Image-Guided Radiation Therapy (IGRT) Questionnaire Radiotherapy Centers	This study presents a preliminary investigation into the use of image-guided radiation therapy (IGRT) in radiotherapy centers in Türkiye, based on survey data. Radiation therapy technologists were administered a 67-item Internet-based questionnaire. The first five items focused on demographic data, while the remaining items pertained to IGRT applications in the clinics where the technicians worked. Statistical data of the answers obtained were used in SPSS 28.0. While evaluating the study data, descriptive statistical methods (frequency, percentage) were used. A total of 314 individuals participated in the survey, with 45.9% (n=144) being female and 54.1% (n=170) being male. The study focused on patients with stomach (20.6%), rectum (18.9%), prostate (17.6%), head and neck (14.5%), lung (14.3%), and breast (13.9%) cancers. It was observed that megavoltage (MV) ports (35.7%) were used more frequently during daily checks. It has been stated that 1-25 MV ports are drawn daily. This is followed by Cone Beam Computer Tomography (CBCT) and kilovoltage (kV) port checks. In our country, it is seen that important steps have been taken in IGRT and its use is quite common. It has been observed that the technical infrastructure and manpower in the clinics are sufficient. It is predicted that the development of systems that reduce the doses received by patients and employees with IGRT and the development of wage policies will increase the use of IGRT. We hope that this result will be supported by future large-scale studies. Although we think that the use of IGRT is at an adequate level, there are differences in practices between clinics. It is thought that the establishment of national protocols for the use of IGRT will be in favour of patients and technicians.

Cite

Şenışık, A. M. (2024). A Study on the Evaluation of Image-Guided Treatment Use in Radiotherapy Centers in Türkiye. *GU J Sci, Part A, 11(3)*, 431-442. doi:10.54287/guj.1442069

Author ID (ORCID Number)

0000-0002-7781-3355 Ahmet Murat ŞENİŞİK

Article Process

Submission Date	06.03.2024
Revision Date	23.05.2024
Accepted Date	05.07.2024
Published Date	05.09.2024

1. INTRODUCTION

Image Guided Radiation Therapy (IGRT) is an imaging procedure that allows increasing target accuracy and precision by correcting anatomical and biological deviations applied during conformal radiotherapy (RT) (Nabavizadeh et al., 2016). IGRT allows the identification of the target and making small adjustments just before delivering radiation to detect the target and irradiate the correct location. This helps deliver radiation more accurately. In this way, fewer side effects and a more accurate treatment are provided. With IGRT, while the radiation dose is kept at a minimum level in the normal tissues adjacent to the tumour, high doses can be applied to the tumour. In this way, the success of radiotherapy can be increased (De Los Santos et al., 2013; Şenışık et al., 2022).

IGRT provides precise localization and monitoring of the tumour before and during treatment. During treatment, however, bladder filling, peristalsis, or tumor movement may occur, requiring intrafractional monitoring. Respiratory or cardiac activity may also cause faster movements that affect treatment accuracy. Real-time monitoring of these movements requires a high temporal frequency (De Los Santos et al., 2013;

*Corresponding Author, e-mail: ahmet.senisik@altinbas.edu.tr

Bertholet et al., 2019). Imaging and tracking systems used to monitor the tumour and surrounding structures in detail can be called tumour tracking systems. Tumour tracking systems (TS) used in clinics are CT TS, US TS, Fiducial TS, MR TS, Respiratory monitoring system (RPM) TS, Surface TS, Cone Beam Computer Tomography (CBCT), Orthogonal TS, Xsight Spine TS.

Thanks to tumour tracking with real-time imaging, it helps to ensure that radiation beams are delivered correctly to the tumour in case of tumour movement caused by the patient's breathing or other internal organ movements. Meanwhile, it minimizes the risk of damaging healthy tissues (De Los Santos et al., 2013; Franzone et al., 2016; Lievens et al., 2020). Reducing the risk of unnecessary radiation exposure to healthy tissues and organs also minimizes possible side effects. By precisely targeting the tumour, higher radiation doses can be delivered while preserving critical structures, leading to improved treatment outcomes and reduced complications (Zelefsky et al., 2012; Kilburn et al., 2016; Fiorino et al., 2020). Thanks to these imaging methods, the treatment margin is minimized, and organ movements are considered, reducing the need for repositioning and readjustment during each treatment session (De Los Santos et al., 2013; Sun et al., 2017; Bertholet et al., 2019). Therefore, IGRT provides increased patient comfort and improved overall patient experience.

One of the important advantages of IGRT is that the treatment plan can be adapted with Adaptive radiotherapy according to the changes observed during treatment. Because tumours can change in size, shape, and location over time, IGRT adapts the treatment plan to take these changes into account and provides accurate and effective treatment. By comparing the actual treatment location with the planned location, clinicians can ensure the intended treatment is delivered correctly. This verification process increases quality assurance and helps maintain the highest standards of care in radiation therapy (Kilburn et al., 2016; Sun et al., 2017; Pinitpatcharalert et al., 2019).

The specific monitoring systems used for IGRT may vary depending on the treatment facility, the equipment available, and the type of cancer being treated. Radiation oncologists and medical physicists evaluate and select the most appropriate monitoring system based on each patient's needs and treatment requirements. IGRT has become increasingly common and widely adopted over the past decade due to its numerous benefits in improving treatment fidelity and outcomes. It has been incorporated into standard radiation therapy practice in many developed countries and is considered a standard of care for certain types of cancer. Due to its accuracy in target tracking, advances in IGRT demonstrate clinically superior benefits of stereotactic RT (SBRT or SRS) over conventional treatments in specific disease regions such as lung, brain, liver, and prostate (Zelefsky et al., 2012; Sun et al., 2017; Chen et al., 2020).

However, the use of IGRT depends on the capabilities of the clinics and the experience of the staff. In high-resource settings such as North America, Western Europe, and parts of Asia, IGRT is relatively common and routinely used in radiation therapy departments. In these regions, the frequency of IGRT use is generally higher. In environments where resources are limited or in areas where access to advanced technology is limited, the frequency of IGRT use is lower. Challenges such as cost, infrastructure, training, and availability of imaging equipment limit the widespread use of IGRT in some regions (Simpson et al., 2010; Nabavizadeh et al., 2016; Luh et al., 2020).

Although it is known that IGRT applications increase the accuracy and precision of radiotherapy treatments, the frequency of its clinical application has not been investigated before. This study aimed to evaluate the use of image-guided radiotherapy in Radiotherapy centers in Türkiye.

2. MATERIAL AND METHOD

The research was evaluated with a survey applied to associate degree students and graduates studying in the Radiotherapy program, with the permission of the Altınbaş University Ethics Committee numbered 30.03.2023-48240. The survey was conducted by the Principles of the Declaration of Helsinki. Since there is no previous study on this subject, the content validity of the questions prepared was determined by applying the Kendall agreement coefficient W correlation test in line with expert opinions ($p > 0.05$). In the analysis, the factor loadings of the questions vary between 0.455 and 0.765. The total variance explained is 35,673. High factor loadings indicate that the questions are related to each other and explain 45% to 76% of the total

variance. This means that the questions share a common variance with other questions. As a result of the pilot study, the Cronbach Alpha value was determined to be 0.842 in the reliability test, and then a full-scale application was carried out ($0.60 \leq \alpha \leq 0.80$ is quite reliable). The survey was conducted both face-to-face and online. Participants in the research study were informed about the research and volunteers were included. Consent was obtained from volunteer participants.

In the first part of the survey, demographic questions such as age and gender were asked. The second part of the study comprised 9-31 questions designed to assess the technical infrastructure and competency levels of participants in radiotherapy clinics. The 32-53 questions assess the frequency of utilization of the existing infrastructure by technicians engaged in radiotherapy clinics. A 5-point Likert scale was used when creating some questions.

The data obtained after the survey was collected and evaluated in a single center and SPSS 28.0 was used to analyze the data. While evaluating the study data, descriptive statistical methods (frequency, percentage) were used.

3. RESULTS AND DISCUSSION

IGRT has become increasingly common over the past decade due to its numerous benefits in improving the accuracy and outcomes of radiotherapy treatment. It has been incorporated into standard radiation therapy practice in many developed countries. In our study to determine the usage status and frequency in our country, a total of 314 people participated in the survey. 45.9% of the participants were female ($n = 144$) and 54.1% were male ($n = 170$). The average age was 42.7% and was between 25-31 ($n=134$). The number and brands of LINACs and the number of technicians and physicists in the participants' clinics are given in Tables 1 and 2. While only 17.2% of the centres participating in the survey have a single LINAC, a significant portion have two LINACs. While the number of clinics working with a single technician is 1.6%, 79.3% of the clinics have more than four technicians. While 23.2% of the centres employ a single medical physicist, two physicians work in 34.1% of the clinics. The majority of respondents (41.4%) were aged between 18 and 24 years, while 11.5% were aged between 32 and 38 years, 3.2% were aged between 39 and 45 years, and the remainder were aged 46 years or older. The majority of participants (65%) were employed in private clinics ($n = 204$), while the remaining 35% were employed in public hospitals ($n = 110$). The majority of participants (39.2%, $n=123$) have between one and three years of experience, while 30.9% ($n=97$) have between four and six years of experience. The remaining 19.4% ($n=61$) have more than six years of experience. The average working hours of employees are 7-8 hours. It was observed that the clinics participating in the study had appropriate equipment and technical personnel to perform IGRT. It seems that the number of personnel per device in clinics is sufficient.

Table 1. Number of Linear Accelerators, Technicians, and Physicists in the clinics where the participants are located

	Number of Linear Accelerators in the Clinics				
	1	2	3	4 and above	Total
n	54	156	72	32	314
%	17,2	49,7	22,9	10,2	100
	Number of Technicians in the Clinics				
	1	2	3	4 and above	Total
n	5	15	45	249	314
%	1,6	4,8	14,3	79,3	100
	Number of Physicists in the Clinics				
	1	2	3	4 and above	Total
n	73	107	35	99	314
%	23,2	34,1	11,1	31,5	100

It is important to note that the figures shown in Figure 1 do not indicate the number of devices. According to the 2022 report of the Nuclear Regulatory Board (NDK), there are a total of 381 treatment devices in our country, including 264 LINAC, 3 Co-60, 4 MR-LINAC, 12 CyberKnife, 14 GammaKnife, 24 Tomotherapy and 60 Brachytherapy devices (NDK, 2022). According to Figure 1, 29% of the participants work with Elekta, 23% with Varian, 16% with Siemens, 13% with Tomotherapy, 14% with CyberKnife and 5% with MRLinac.

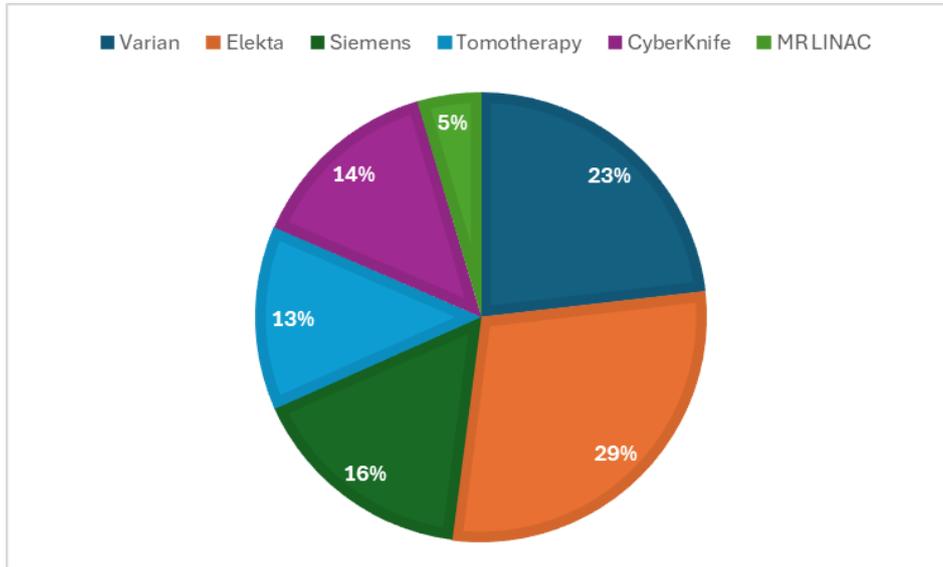


Figure 1. Distribution of linear accelerators used by participants according to brands

Simpson et al. (2010) demonstrated that IGRT users primarily treated genitourinary (91.1%), head and neck (74.2%), central nervous system (71.9%), and lung (66.9%) disease areas (Simpson et al., 2010). Pan et al. (2011) reported in their study with American oncologists that SBRT of the lung (51.8%), spine (39.0%) and liver (31.0%) was performed more commonly (Pan et al., 2011). In our study, daily patients treated mostly had stomach cancer (20.6%), rectal cancer (18.9%), prostate cancer (17.6%), head and neck cancer (14.5%), lung cancer (14.3%), and breast cancer (13.9%) patients. It is seen that the number of patients received daily in a center varies between 40-59. The number of patients treated by the participants daily is given in Table 2.

Tumour tracking systems used in the clinics where the participants worked are shown in Figure 2.

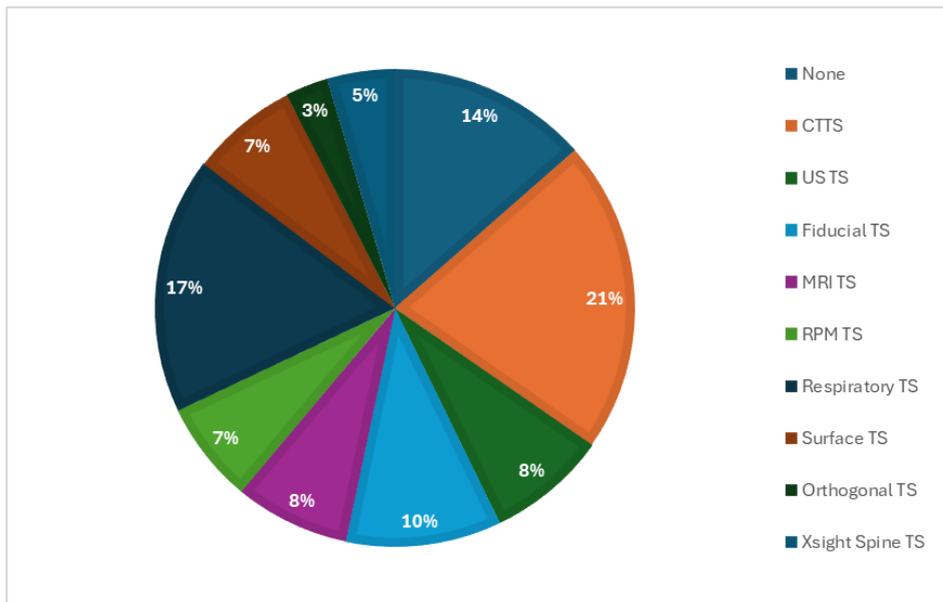


Figure 2. Percentage distribution of tumour tracking systems used by participants in the clinics where they work

Table 2. Daily number of patients treated in the clinics where the participants are located

Daily number of breast cancer patients							
	None	1-10	11-15	16-20	21-40	41-60	Total
n	13	102	93	60	25	21	314
%	4,1	32,5	29,6	19,1	8	6,7	100
Daily number of prostate cancer patients							
	None	1.Eki	Kas.15	16-20	21-40	41-60	Total
n	27	129	79	31	23	25	314
%	8,6	41,1	25,2	9,9	7,3	8	100
Daily number of lung cancer patients							
	None	1.Eki	Kas.15	16-20	21-40	41-60	Total
n	12	105	103	50	21	23	314
%	3,8	33,4	32,8	15,9	6,7	7,3	100
Daily number of stomach cancer patients							
	None	1.Eki	Kas.15	16-20	21-40	41-60	Total
n	19	151	63	35	10	36	314
%	6,1	48,1	20,1	11,1	3,2	11,5	100
Daily number of rectum cancer patients							
	None	1.Eki	Kas.15	16-20	21-40	41-60	Total
n	30	139	68	28	9	40	314
%	9,6	44,3	21,7	8,9	2,9	12,7	100
Number of patients per day head and neck							
	None	1.Eki	Kas.15	16-20	21-40	41-60	Total
n	28	106	83	36	36	25	314
%	8,9	33,8	26,4	11,5	11,5	8	100

According to the HERO survey by the European Society for Health Economics in Radiation Oncology, less than half of all linear accelerators in Europe (49%) are capable of IGRT. The percentages of linear accelerators capable of IGRT in different countries are as follows: 53%; Netherlands: 95%) (Vaandering et al., 2023). It is seen that 81.1% SRS/SBRT, 62.1% TBI, 54.5% Brachytherapy and 82.2% Adaptive radiotherapy applications are performed in the clinics participating in the survey in our country. The clinics where the participants work, the treatment modalities and the frequencies they use are shown in Table 3.

Table 3. Clinics where the participants worked, treatment modalities and frequencies used

Is SRS/SBRT applied?			
	Yes	No	Total
n	257	57	314
%	81,8	18,2	100
Is TBI performed?			
	Yes	No	Total
n	195	119	314
%	62,1	37,9	100

Table 3. continued

	Is Brachytherapy applied?				
	Yes		No		Total
n	171		143		314
%	54,5		45,5		100
	Is Adaptive RT applied?				
	Yes		No		Total
n	258		56		314
%	82,2		17,8		100
	Monthly frequency of SRS/SBRT application				
	0-5	6-10	11-15	16 and above	Total
n	147	79	47	41	314
%	46,8	25,2	15	13,1	100
	Monthly frequency of TBI application				
	0-5	6-10	11-15	16 and above	Total
n	251	42	18	3	314
%	79,9	13,4	5,7	1	100
	Monthly frequency of brachytherapy application				
	0-5	6-10	11-15	16 and above	Total
n	209	49	39	17	314
%	66,6	15,6	12,4	5,4	100
	Monthly frequency of Adaptive RT application				
	0-5	6-10	11-15	16 and above	Total
n	164	89	47	14	314
%	52,2	28,3	15	4,5	100

Nabavizadeh et al. (2016) administered a survey containing 5979 questions regarding IGRT to members of the American Society of Radiation Oncology (ASTRO) and found that the most preferred method was portal imaging (67.4%). This was reported to be followed by kV planar imaging (32%) and CBCT (10.4%). In their study, it was reported that ultrasound-based (most commonly in the intact prostate, 4.8%), fluoroscopy-based (most commonly in the lung, 1.7%), or CT-guided tracking systems were used (most commonly in the prostate fossa, 2.4%) (Nabavizadeh et al., 2016).

Beasley et al. (2019) administered a survey containing a maximum of 32 questions regarding lung stereotactic ablative radiotherapy (SABR) to 62 radiotherapy centers in the UK. Centers reported 6% use of kV-based monitoring and 88% use of CBCT. They reported in their study that two-thirds (66.7%) of the centers participating in the survey did not use any active tumour tracking system, and 11% used breath holding and surface tracking systems. It was also stated that 14% used kV tumour tracking, while 8% used an internal tracking system (Beasley et al., 2019).

Batumalai et al. (2017) investigated the frequency of IGRT use in Australia. CBCT (97%), kV electronic portal image (EPI) (89%) and MV EPI (75%) methods were most frequently used in their studies. They reported in their study that RPM was most commonly applied in patients with central nervous system (CNS) (12%), breast (12%) and lung (6%). While the spirometer (ABC) system is used in breast (6%) and gastrointestinal (GI) cancers (3%), optical (VisionRT & C-rad) systems are used in the CNS (3%) and breast (3%) patients. They

reported that the radiofrequency (Calypso) system was used in the CNS (3%), lung (3%) and GI (3%), while the ultrasound tracking system was used in the GI (6%) and breast (3%) (Batumalai et al., 2017).

Simpson et al. (2010) reported that the most commonly used IGRT methods are MV planar (62.7%), volumetric (58.8%) and kV planar imaging (57.7%). They reported that the percentage of use of ultrasound, video, megavoltage (MV) planar, kilovoltage (kV) planar and volumetric technologies is 22.3%, 3.2%, 62.7%, 57.7%, and 58.8% respectively (Simpson et al., 2010).

According to the results obtained from our study, it was observed that MV ports (35.7%) were used more frequently in daily checks. It has been stated that 1-25 MV ports are drawn daily. This is followed by CBCT and kV port controls. It was observed that the most common monitoring system was RPM. These are followed by fiducial, surface, X-spine, orthogonal kV and US tracking systems, respectively. It is thought that these numbers are shaped by both the availability of follow-up systems and the incoming patient profile. The daily and monthly numbers of patients treated with these follow-up systems are given in Table 4.

Table 4. Number of daily and monthly follow-ups received by follow-up systems in the clinics where the participants are located

	Number of kV Ports per Day					
	0	1-25	26-50	51-75	76-100	Total
n	75	81	60	25	73	314
%	23,9	25,8	19,1	8	23,2	100
	Daily Number of MV Ports					
	0	1-25	26-50	51-75	76-100	Total
n	90	112	61	31	20	314
%	28,7	35,7	19,4	9,9	6,4	100
	Number of CBCTs per day					
	0	1-25	26-50	51-75	76-100	Total
n	76	92	73	36	37	314
%	24,2	29,3	23,2	11,5	11,8	100
	Daily Number of CT Follow-Ups					
	0	1-25	26-50	51-75	76-100	Total
n	133	99	51	21	10	314
%	42,4	31,5	16,2	6,7	3,2	100
	Daily Number of MRI Follow-ups					
	0	1-25	26-50	51-75	76-100	Total
n	178	74	42	18	2	314
%	56,7	23,6	13,4	5,7	0,6	100
	Daily RPM Tracking Count					
	0	1-25	26-50	51-75	76-100	Total
n	163	87	43	16	5	314
%	51,9	27,7	13,7	5,1	1,6	100

Table 4. continued

Monthly Number of Surface Tracking						
	0	1-25	26-50	51-75	76-100	Total
n	141	92	59	15	7	314
%	44,9	29,3	18,8	4,8	2,2	100
Monthly Respiratory Monitoring Number						
	0	1-25	26-50	51-75	76-100	Total
n	113	126	45	19	11	314
%	36	40,1	14,3	6,1	3,5	100
Monthly Number of Fiducial Follow-Ups						
	0	1-25	26-50	51-75	76-100	Total
n	165	98	35	15	1	314
%	52,5	31,2	11,1	4,8	0,3	100
Monthly Orthogonal kV Monitoring Number						
	0	1-25	26-50	51-75	76-100	Total
n	215	60	28	8	3	314
%	68,5	19,1	8,9	2,5	1	100
Monthly Number of Xspine Followers						
	0	1-25	26-50	51-75	76-100	Total
n	208	73	20	13	0	314
%	66,2	23,2	6,4	4,1	0	100
Monthly US Tracking Number						
	0	1-25	26-50	51-75	76-100	Total
n	197	57	40	14	6	314
%	62,7	18,2	12,7	4,5	1,9	100

The answers given by the participants regarding the use of tumour tracking systems are shown in Table 5.

Table 5. Participants' answers regarding the use of tumour tracking systems

Is it necessary to use IGRT systems in the clinic?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	0	14	19	221	60	314
%	0	4,5	6,1	70,4	19,1	100
Do you think IGRT systems are used adequately in your clinic?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	8	21	202	51	32	314
%	2,5	6,7	64,3	16,2	10,2	100

Table 5. continued

Do you think the reason why IGRT systems are less used is the lack of equipment?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	12	29	219	47	7	314
%	3,8	9,2	69,7	15	2,2	100
Do you think the reason for the low use of IGRT systems is the large number of patients?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	9	36	210	42	17	314
%	2,9	11,5	66,9	13,4	5,4	100
Do you think the reason why IGRT systems are less used is the lack of technicians?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	15	37	207	43	12	314
%	4,8	11,8	65,9	13,7	3,8	100
Do you think the reason why IGRT systems are less used is the lack of physicists?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	15	216	40	33	10	314
%	4,8	68,8	12,7	10,5	3,2	100
Do you think the reason for the low use of IGRT systems is the lack of oncologists?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	17	205	59	29	4	314
%	5,4	65,3	18,8	9,2	1,3	100
Do you think the reason for the underuse of IGRT systems is lack of education?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	24	203	32	46	9	314
%	7,6	64,6	10,2	14,6	2,9	100
Do you think the reason why IGRT systems are less used is to reduce overdose intake?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	16	18	217	54	9	314
%	5,1	5,7	69,1	17,2	2,9	100
Do you think the reason why IGRT systems are less used is because they are not paid for?						
	Definitely not	Not at all	No idea	To some extent	Quite a bit	Total
n	23	20	216	43	12	314
%	7,3	6,4	68,8	13,7	3,8	100

When asked whether the use of IGRT is necessary, 70.4% of the participants answered Quite a bit. To the question of whether the use of IGRT is sufficient, 64.3% of the survey participants answered "I have no idea" and 16.2% answered "Quite a bit". When asked whether the use of IGRT was related to lack of equipment, large number of patients, and lack of technicians, the majority answered "I have no idea." The majority of participants think that the number of oncologists and physicists and the lack of training are not an obstacle to the use of IGRT. The majority of those who said they had no idea about the questions investigating the relationship between the low use of IGRT and the cost or low doses of patients were the ones who said they had no idea.

A certain number of centers and personnel could be reached with the survey applied. There is a need for larger studies that will include all centers in the country. Considering that the use of IGRT will increase in the future, dissemination of the necessary protocols and applications, control of the application and detection of deficiencies can be solved with such surveys.

4. CONCLUSION

There has been a significant increase in the number of radiotherapy machines over the past 10 years. However, the use and prevalence of tumor imaging techniques in treatment is questionable. The International Atomic Energy Agency (IAEA) provides a nuanced interpretation of the staffing recommendations for radiation oncology clinics outlined by the World Health Organization (WHO). While the WHO specifies the need for 1 radiotherapy technician for every 60 patients, 1 health physicist for every 400 patients, and 1 megavoltage device for every 300 new patients, the IAEA expands on these guidelines by emphasizing the complexity and context-specific nature of staffing needs. The IAEA acknowledges these baseline recommendations but argues that staffing levels should also account for the complexity of the treatments provided, the specific equipment used, and the percentage of time staff spend on non-clinical duties such as teaching, research, and management. This approach aims to ensure safe, efficient, and high-quality radiotherapy services tailored to each clinic's unique situation (Podgorsak, 2005; WHO, 2020). According to the survey results we conducted with a limited number of participants, it is thought that the number of patients, technicians and personnel per machine is sufficient in the surveyed clinics. However, a more comprehensive investigation is needed. However, considering that the general population and the incidence of the disease are increasing, it is thought that there is a need for technicians, physicists and physicians.

IGRT has become a mainstay of modern RT. By using imaging techniques such as portal imaging, CT, MRI or ultrasound, it is possible to cover the tumor with a sufficient radiation dose and protect normal tissues. However, the routine use of tumor tracking systems appears to vary from clinic to clinic. It is thought that national protocols regarding the use of IGRT have not yet been established and efforts to develop them should be initiated. The number of image-guided treatments is expected to increase as reimbursement policies change.

ACKNOWLEDGEMENT

We thank the participants who volunteered to participate in this study and those who helped deliver this study to participants.

This study was discussed at the meeting number 4 of Altınbaş University Clinical Research Ethics Committee on March 16, 2023 and was found ethically appropriate (Date and Number of Documents: 30.03.2023-48240).

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- Batumalai, V., Holloway, L. C., Kumar, S., Dundas, K., Jameson, M. G., Vinod, S. K., & Delaney, G. P. (2017). Survey of image-guided radiotherapy use in Australia. *Journal of Medical Imaging and Radiation Oncology*, 61(3), 394-401. <https://doi.org/10.1111/1754-9485.12556>
- Beasley, M., Brown, S., McNair, H., Faivre-Finn, C., Franks, K., Murray, L., Van Herk, M., & Henry, A. (2019). The advanced radiotherapy network (ART-NET) UK lung stereotactic ablative radiotherapy survey: national provision and a focus on image guidance. *British Journal of Radiology*, 92(1098), 20180988. <https://doi.org/10.1259/bjr.20180988>
- Bertholet, J., Knopf, A., Eiben, B., McClelland, J., Grimwood, A., Harris, E., Menten, M., Poulsen, P., Nguyen, D. T., Keall, P., & Oelfke, U. (2019). Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in Medicine & Biology*, 64(15), 15TR01. <https://doi.org/10.1088/1361-6560/ab2ba8>

- Chen, G.-P., Tai, A., Keiper, T. D., Lim, S., & Li, X. A. (2020). Technical Note: Comprehensive performance tests of the first clinical real-time motion tracking and compensation system using MLC and jaws. *Medical Physics*, 47(7), 2814-2825. <https://doi.org/10.1002/mp.14171>
- De Los Santos, J., Popple, R., Agazaryan, N., Bayouth, J. E., Bissonnette, J. P., Bucci, M. K., Dieterich, S., Dong, L., Forster, K. M., Indelicato, D., Langen, K., Lehmann, J., Mayr, N., Parsai, I., Salter, W., Tomblin, M., Yuh, W. T. C., & Chetty, I. J. (2013). Image Guided Radiation Therapy (IGRT) Technologies for Radiation Therapy Localization and Delivery. *International Journal of Radiation Oncology, Biology, Physics*, 87(1), 33-45. <https://doi.org/10.1016/J.IJROBP.2013.02.021>
- Fiorino, C., Guckenberger, M., Schwarz, M., van der Heide, U. A., & Heijmen, B. (2020). Technology-driven research for radiotherapy innovation. *Molecular Oncology*, 14(7), 1500-1513. <https://doi.org/10.1002/1878-0261.12659>
- Franzone, P., Fiorentino, A., Barra, S., Cante, D., Masini, L., Cazzulo, E., Todisco, L., Gabriele, P., Garibaldi, E., Merlotti, A., Redda, M. G. R., Alongi, F., & Corvò, R. (2016). Image-guided radiation therapy (IGRT): practical recommendations of Italian Association of Radiation Oncology (AIRO). *La Radiologia Medica*, 121(12), 958-965. <https://doi.org/10.1007/s11547-016-0674-x>
- Kilburn, J. M., Soike, M. H., Lucas, J. T., Ayala-Peacock, D., Blackstock, W., Isom, S., Kearns, W. T., Hinson, W. H., Miller, A. A., Petty, W. J., Munley, M. T., & Urbanic, J. J. (2016). Image guided radiation therapy may result in improved local control in locally advanced lung cancer patients. *Practical Radiation Oncology*, 6(3), e73-e80. <https://doi.org/10.1016/J.PRRO.2015.10.004>
- Lievens, Y., Borrás, J. M., & Grau, C. (2020). Provision and use of radiotherapy in Europe. *Molecular Oncology*, 14(7), 1461-1469. <https://doi.org/10.1002/1878-0261.12690>
- Luh, J. Y., Albuquerque, K. V., Cheng, C., Ermoian, R. P., Nabavizadeh, N., Parsai, H., Roeske, J. C., Weiss, S. E., Wynn, R. B. Yu, Y., Rosenthal, S. A., & Hartford, A. (2020). ACR–ASTRO Practice Parameter for Image-guided Radiation Therapy (IGRT). *American Journal of Clinical Oncology*, 43(7), 459-468. <https://doi.org/10.1097/COC.0000000000000697>
- Nabavizadeh, N., Elliott, D. A., Chen, Y., Kusano, A. S., Mitin, T., Thomas, C. R., & Holland, J. M. (2016). Image Guided Radiation Therapy (IGRT) practice patterns and IGRT's impact on workflow and treatment planning: Results from a national survey of american society for radiation oncology members. *International Journal of Radiation Oncology, Biology, Physics*, 94(4), 850-857. <https://doi.org/10.1016/j.ijrobp.2015.09.035>
- NDK, Nükleer Düzenleme Kurulu (2022). *Türkiye’de Radyasyon Kaynakları 2022*.
- Pan, H., Simpson, D. R., Mell, L. K., Mundt, A. J., & Lawson, J. D. (2011). A survey of stereotactic body radiotherapy use in the United States. *Cancer*, 117(19), 4566-4572. <https://doi.org/10.1002/cncr.26067>
- Pinitpatcharalert, A., Happersett, L., Kollmeier, M., McBride, S., Gorovets, D., Tyagi, N., Varghese, M., & Zelefsky, M. J. (2019). Early Tolerance Outcomes of Stereotactic Hypofractionated Accelerated Radiation Therapy Concomitant with Pelvic Node Irradiation in High-risk Prostate Cancer. *Advances in Radiation Oncology*, 4(2), 337-344. <https://doi.org/10.1016/J.ADRO.2018.12.001>
- Podgorsak, E. B. (2005). *Radiation Oncology Physics: A Handbook for Teachers and Students*. International Atomic Energy Agency, Austria.
- Şenışık, A. M., Tunçman, D., Mutlu, E., & Erge, E. (2022). *50 Soruda Kanser ve Radyoterapi*. Altınbaş Üniversitesi Yayınları.
- Simpson, D. R., Lawson, J. D., Nath, S. K., Rose, B. S., Mundt, A. J., & Mell, L. K. (2010). A survey of image-guided radiation therapy use in the United States. *Cancer*, 116(16), 3953-3960. <https://doi.org/10.1002/cncr.25129>
- Sun, B., Chang, J., & Rong, Y. (2017). The more IGRT systems, the merrier? *Journal of Applied Clinical Medical Physics*, 18(4), 7-11. <https://doi.org/10.1002/acm2.12126>
- Vaandering, A., Jansen, N., Weltens, C., Moretti, L., Stellamans, K., Vanhoutte, F., Scalliet, P., Remouchamps, V., & Lievens, Y. (2023). Radiotherapy-specific quality indicators at national level: How to make it happen. *Radiotherapy and Oncology*, 178, 109433. <https://doi.org/10.1016/j.radonc.2022.11.022>

WHO, World Health Organization (2020). WHO report on cancer: setting priorities, investing wisely and providing care for all. (2020). <https://www.who.int/publications/i/item/9789240001299>

Zelevsky, M. J., Kollmeier, M., Cox, B., Fidaleo, A., Sperling, D., Pei, X., Carver, B., Coleman, J., Lovelock, M., & Hunt, M. (2012). Improved Clinical Outcomes With High-Dose Image Guided Radiotherapy Compared With Non-IGRT for the Treatment of Clinically Localized Prostate Cancer. *International Journal of Radiation Oncology, Biology, Physics*, 84(1), 125-129. <https://doi.org/10.1016/j.ijrobp.2011.11.047>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1496604>

Current Feedback Operation Amplifier Based on Floating Passive and Active Inductors in Filter Applications

Hüseyin DEMİREL^{1*} ¹ Ankara Yıldırım Beyazıt University, Information Management Systems, Ankara, Türkiye

Keywords	Abstract
Passive and Active Inductance Simulator CFOA Band-Stop (BS) Filter	This paper proposes using current feedback operation amplifiers (CFOAs) as an active filter in creating floating passive and active inductance simulators is a versatile and efficient solution for circuit design. CFOAs are popular for their high slew rate and wide bandwidth which makes them ideal candidates for applications demanding rapid response time alongside high frequencies. It is possible to customize these simulated inductances for different design specifics by changing certain external resistor values. The circuit comprises a trio of CFOAs, an earthed capacitor and three resistors, thus making it quite simple to put into practice at a cheap price. Experiments and LTSPICE simulations have been conducted to examine the performance of the circuit, and closely confirmed theoretical expectations. In other words, this confirms the trustworthiness and preciseness of the introduced floating active inductance simulator. Voltage-mode band-stop filtering is one real-world scenario where this circuit is used. This indicates that our proposed technique can have more applications than one in circuit design. This is further proof that our idea can be used in other circuits as well. In conclusion, using CFOAs in active filters for building a grounded passive active inductance simulator is a solid and effective answer to circuit designers. This makes it an appealing choice for numerous applications because of its simplicity in designing the network as well as its high performance and customizability. Circuit design and application can be greatly improved if more study is done in this field.

Cite

Demirel, H. (2024). Current Feedback Operation Amplifier Based on Floating Passive and Active Inductors in Filter Applications. *GU J Sci, Part A, 11(3)*, 443-450. doi: [10.54287/guj.1496604](https://doi.org/10.54287/guj.1496604)

Author ID (ORCID Number)

[0000-0003-2983-1425](https://orcid.org/0000-0003-2983-1425) Hüseyin DEMİREL

Article Process

Submission Date	05.06.2024
Revision Date	14.06.2024
Accepted Date	04.07.2024
Published Date	06.09.2024

1. INTRODUCTION

Filters are circuits that allow signals within a certain frequency range to pass and do not allow signals at other frequencies to pass. Because of these features, filters are also called selective circuits. Active filters include active circuit elements such as transistors and operational amplifiers, as well as passive circuit elements such as resistors and capacitors. Passive circuit elements determine the cut-off frequency, while active circuit elements provide voltage gain. Filters are generally classified according to the state of the output voltage (signal) in response to the change in frequency of the input voltage (signal). Accordingly, there are generally four types of active filters. These are low-pass, high-pass, band-pass and non-band-pass filters (Demirel, 2017).

Operational amplifiers, one of the most important elements used in filters, are very high-gain differential amplifiers that use voltage feedback to provide a specified voltage gain. These amplifiers have a very high open-loop gain, high input impedance and low output impedance design. Operational amplifiers are used in many circuits such as addition, subtraction, multiplication, integration, and derivative operations in analog calculators, phase shifting, signal processing, instrumentation, communication, alarm, measurement, test circuits and filter applications (Demirel, 2017).

*Corresponding Author, e-mail: husdemirel06@gmail.com

The interest in developing signal processing circuits is increasing day by day. In particular, the focus is on the design and production of circuits such as active filters without the need for real coils. The advantages of this approach include the ability to design and produce circuits with smaller dimensions compared to physical coils by using spiral inductors in integrated circuits. However, this method also has some disadvantages. For example, there may be some limitations in terms of space usage and we may encounter some difficulties in terms of adjustability, cost and efficiency.

Inductance simulators play an important role in overcoming such disadvantages. These simulators allow the simulation of inductance behavior without physically using real coils. Thus, they can provide more flexible and efficient solutions in active and passive filter designs, analog phase shifters, parasitic element cancellation applications and oscillator designs. This allows the industry to develop more compact, more flexible and more efficient signal processing circuits. Subsequently, the emphasis switches to the inductance simulation, which makes use of many high-performing active building components, including operational trans-conductor amplifier (O.T. A), current conveyors, current feedback op-amps (CCCFO), four-terminal floating nullar (F.T.F.Ns), current differencing buffered amplifiers (CDBAs), etc. Most of the circuits that have been documented require the usage of active components with numerous outputs. Furthermore, some earlier works make use of active components that aren't yet ICs that are sold commercially. Therefore, utilizing off-the-shelf components to create the circuits is not a simple task.

An intriguing active component that is particularly well-suited for a certain kind of analog signal processing is the current feedback amplifier (C.F.A) or current feedback operational amplifier (CFOA) (Demirel & Ahmed, 2023a; Dikicioğlu & Polat, 2024; Mohammed et al., 2024). This device offers versatility and allows for a range of circuit configurations. It can function in both current and voltage modes. Furthermore, it can provide beneficial attributes such fast slew rate, absence of parasitic capacitances, broad bandwidth, and ease of implementation (Feng et al., 2011; Shkir & Abdulazeez, 2017; Ahmed & Demirel, 2023; Demirel & Ahmed, 2023b; Mohammed et al., 2023; Mohammed & Demirel, 2023). Currently, the CFA is available for purchase; one example is Analog Devices Inc.'s AD844.

LTspice (Linear Technology SPICE) is a powerful SPICE simulation software tool created by Analog Devices, originally by Linear Technology. It is used to simulate electronic circuits to predict their behavior. Key features and applications of LTspice include:

- **Circuit Simulation:** Enables users to simulate the behavior of analog and mixed-signal circuits.
- **Schematic Capture:** Allows for the creation of circuit diagrams using a graphical interface.
- **Waveform Viewer:** Visualizes simulation results as voltage, current, and other waveforms.
- **Model Library:** Offers a comprehensive library of components such as transistors, diodes, capacitors, resistors, and more, with the option to add custom models.
- **Transient Analysis:** Simulates circuit behavior over time for time-domain analysis.
- **AC Analysis:** Analyzes frequency response and other AC characteristics of circuits.
- **DC Analysis:** Evaluates the DC operating points of the circuit.
- **Noise Analysis:** Simulates the noise performance of circuits.
- **Parametric Sweeps:** Allows variation of component values to observe their effect on circuit performance.
- **Monte Carlo Analysis:** Assesses the statistical variation in circuit performance due to component tolerances.

LTspice is widely utilized by electrical engineers and designers for circuit design, testing, and optimization due to its accuracy, speed, and comprehensive features. Additionally, it is freely available, making it accessible for both professional and educational purposes.

This work presents a floating active and passive inductance simulator that uses identical CFOAs as its focus. Using off-the-shelf components makes the suggested circuit straightforward to execute realistically. It is also possible to regulate the inductances by using external resistors. Simulations using LT-Spice and experimentation demonstrate the performances of the suggested circuits, which correlate quite well with the

diagrams. To demonstrate the value of the circuit discussed, an application as a voltage-mode (VM) band stop filter is provided (Bhaskar et al., 1999; Jaikla & Lahiri, 2012; Senani et al., 2016).

2. MATERIAL AND METHOD

The method and study format of the study are given below under separate headings. These headings are operational principle of current feedback operational amplifiers, proposed and simulated floating active inductance and non-ideal illustration.

2.1. Operational Principle of Current Feedback Operational Amplifiers (CFOA)

Analog Devices' high-speed current feedback operational amplifiers, operating at speeds greater than 50 MHz, allow for enhanced performance at elevated speeds. These amplifiers typically exhibit broader bandwidths and higher slew rates compared to voltage feedback amplifiers and maintain consistent bandwidth regardless of gain. A current feedback op amp processes an error current at its low-impedance negative input terminal, rather than an error voltage, and generates a corresponding output voltage.

The following equation displays the characteristics of CFOA.

$$\begin{bmatrix} \mathbf{I}_y. \\ \mathbf{V}_x. \\ \mathbf{I}_z. \\ \mathbf{V}_w. \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_x. \\ \mathbf{V}_y. \\ \mathbf{V}_z. \\ \mathbf{V}_w. \end{bmatrix} \quad (1)$$

Figure 1a and 1b show the CFOA's symbol and analogous circuit, respectively.

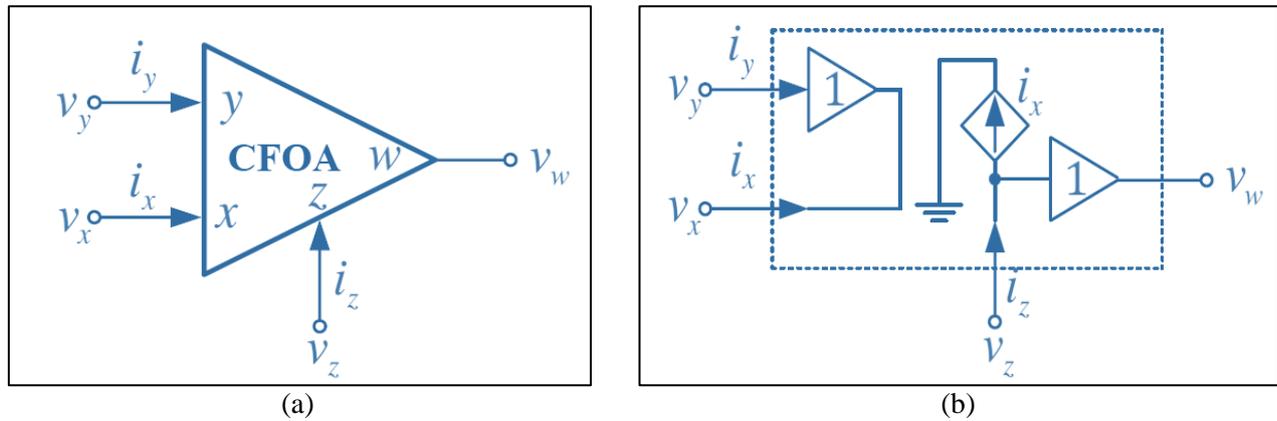


Figure 1. CFOA **a)** the corresponding symbol; **b)** the comparable circuit

2.2. Proposed and Simulated Floating Active Inductance

The suggested floating active inductance simulator is seen in Figure 2. Three CFOAs, three resistors, and one capacitor with grounded are all part of the circuit. The circuit in Figure 2 may be readily analyzed to obtain the input impedance using the CFOA characteristics from section 1. (Prommee & Dejhan, 2002; Maheshwari, 2009; Sa-Ngiamvibool & Jantakun, 2014).

$$Z_L(s) = \frac{V_1 - V_2}{I_{in}} = sC.R.R_1 \quad (2)$$

$R_2=R_3=R$, and $I_1=I_2=I_{in}$. According to Eq. (2), Figure 2's circuit emulates a floating active inductance with a specified value.

$$Leq = C.R.R_1 \quad (3)$$

Equation (3) demonstrates that altering the resistance and capacitor will change the inductance value (Leq).

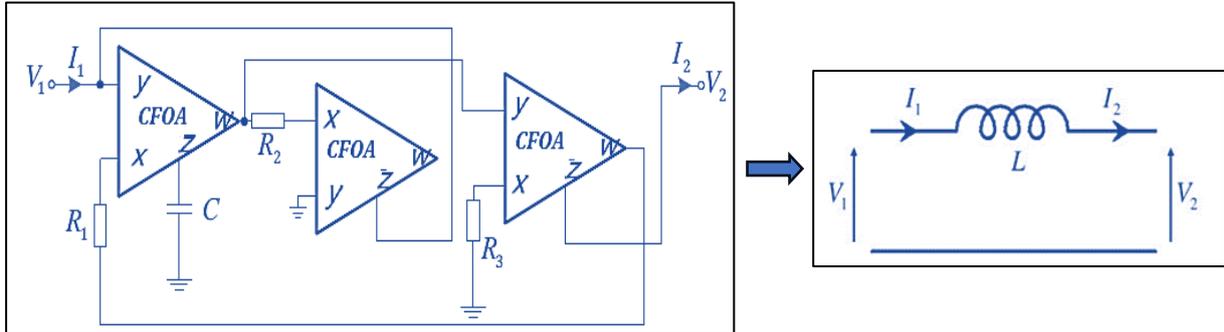


Figure 2. Proposed Floating Active and passive Inductance

2.3. Non-ideal Illustration

The following equation illustrates the CFOA qualities for the non-ideal instance (Horng et al., 2010; Horng, 2011).

$$\begin{bmatrix} Iy. \\ Vx. \\ Iz. \\ Vw. \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta. & 0 & 0 \\ \alpha. & 0 & 0 & 0 \\ 0 & 0 & \gamma. & 0 \end{bmatrix} \begin{bmatrix} Ix. \\ Vy. \\ Vz. \\ Vw. \end{bmatrix} \quad (4)$$

The transmitted error values, deviating from one, are β , α , and γ . Considering the effects of β , α , and γ on the circuit in Figure 2, I_1 and I_2 may be written as follows:

$$I_1 = (\beta_1 V_1 - \gamma_3 V_2) \frac{\gamma_1 \alpha_1 \alpha_2}{sCR_1 R_2} \quad (5)$$

and

$$I_2 = (\beta_1 V_1 - \gamma_3 V_2) \frac{\gamma_1 \alpha_1 \alpha_2 \beta_3}{sCR_1 R_3} \quad (6)$$

If the conditions $\beta_1 = \gamma_3 = \epsilon_1$, $\alpha_2 = \alpha_3 * \beta_3 = \epsilon_2$, the resistance $R_2 = R_3 = R$, and currents $I_1 = I_2 = I_{in}$ the input impedance is approximated, then Eq. (5) and (6) are as follows:

$$Z_L(s) = \frac{V_1 - V_2}{I_{in}} = \frac{sCRR_1}{\gamma_1 \alpha_1 \epsilon_1 \epsilon_2} \quad (7)$$

Based on Eq. (7), the electronic circuit as illustrated in Figure 2 represents a floating active and passive Inductance with a specified value for non-ideal consideration.

$$Leq = \frac{sCRR_1}{\gamma_1 \alpha_1 \epsilon_1 \epsilon_2} \quad (8)$$

The magnitude of the inductance value is affected by these errors as determined by Eq. (8). Considering that these errors are temperature dependent, the magnitude of the inductance value will also be slightly

temperature dependent. In order to reduce the effects of these results, a suitable CFOA (Current Feedback Operational Amplifier) design should be considered correctly. Correct approaches in CFOA design can minimize the effects of such errors and thus increase the stability of the inductance value. In particular, controlling the effects caused by temperature changes can provide more reliable and consistent operation of the circuits. Therefore, determining the inductance values correctly and reducing the effects of temperature changes can significantly increase the performance of electronic systems.

3. RESULTS AND DISCUSSION

The results of the simulations obtained as a result of the simulations performed in this study are given below. The LT-SPICE simulation program was utilized for the analysis to demonstrate the capabilities of the suggested active and passive inductance simulator. Using the AD.844 macro-model from Analog Devices, the CFOAs were implemented. The supply voltages used in the circuit were biased at $\pm 5V$. The circuit in Figure 2. is designed with $R1 = R2 = R3 = 2\text{ K}\Omega$ and $C = 1\text{ pF}$.

Figure 3. shows the phase and magnitude responses for the impedances in the suggested circuit.

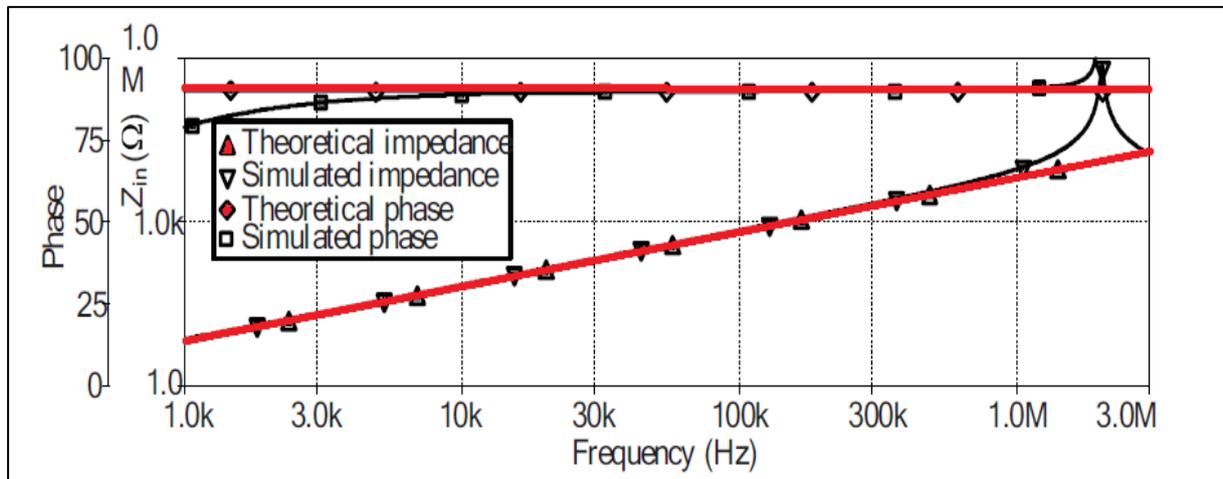


Figure 3. Circuit phases and impedances for the floating active and passive inductance

By changing resistance (R), as seen in Eq. (3), tuning ability is likewise simulated and displayed in Figure 4.

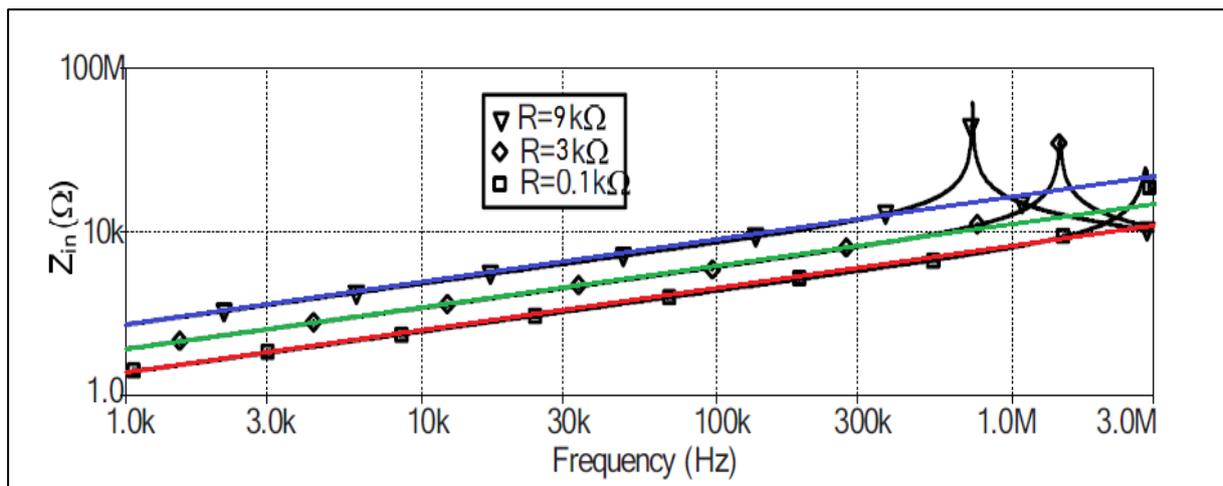


Figure 4. Waveforms of the proposed circuits' typical voltage & current

The voltage & current signal through the suggested floating active inductor when V_2 are grounded are seen in Figure 5.

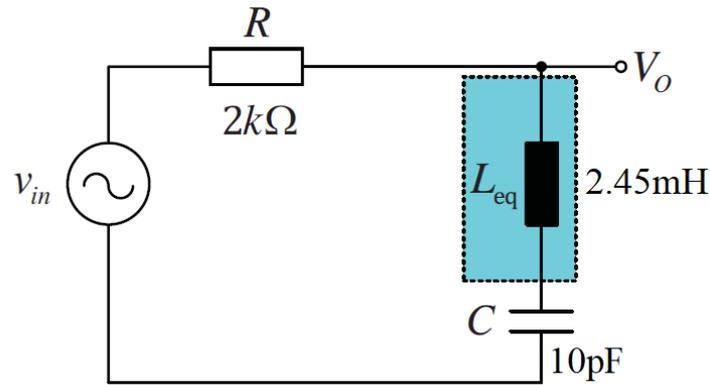


Figure 5. A simulator with floating passive inductance is used in a series resonant circuit

The circuit's simulated magnitude response is shown in Figure 6.

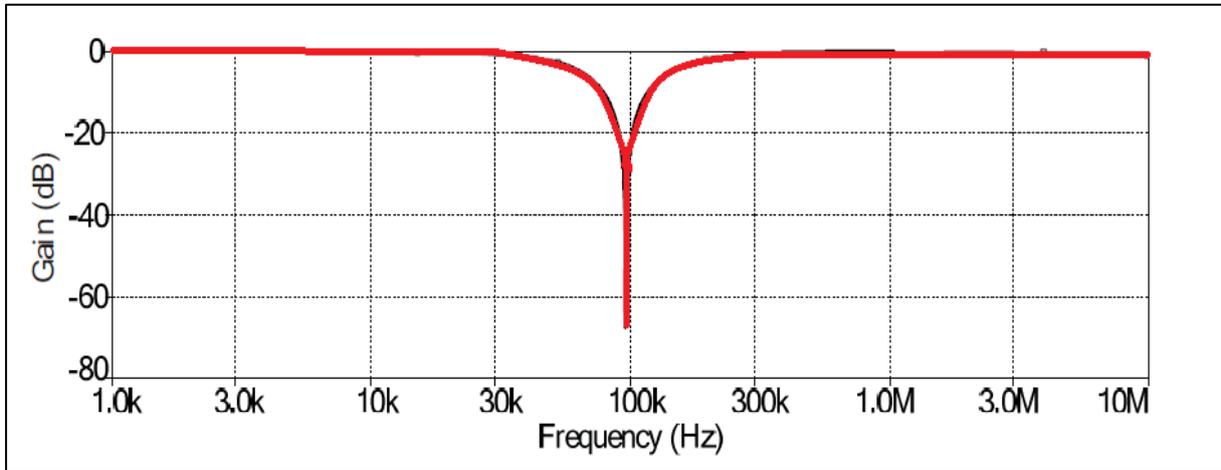


Figure 6. The circuit's simulated magnitude response

Figure 7 shows waveforms of the proposed active floating inductance simulator that are being tested for voltage & current.



Figure 7. Waveforms of the proposed active floating inductance simulator that are being tested for voltage & current.

The circuit's experimental magnitude response is shown in Figure 8.

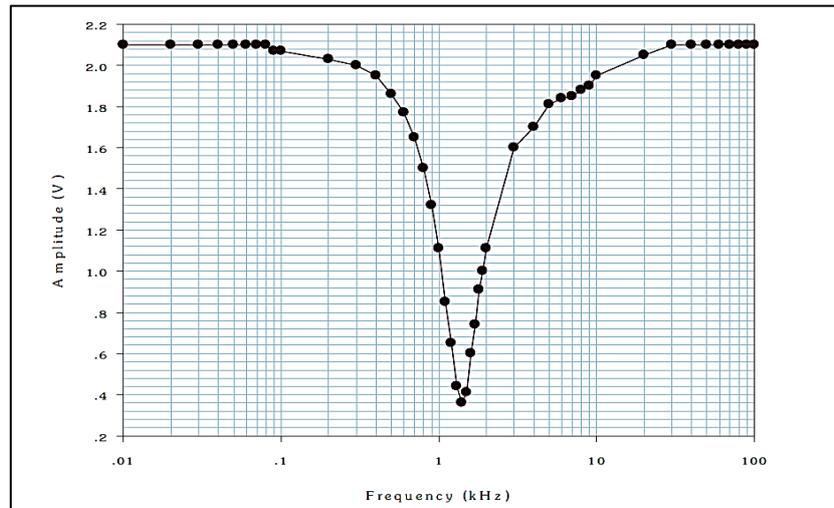


Figure 8. The circuit's experimental magnitude response

4. CONCLUSION

This study introduces both active and passive floating inductance simulators. The proposed circuit can be effectively built using components that are readily available in the market, making it highly accessible and practical. The active building blocks used in the study, especially Current Feedback Operational Amplifiers (CFOA), are shown to be suitable for both voltage and current signal processing modes. A significant advantage of this design is that the capacitors are connected to the Z terminal, which has a high output impedance (Z). This configuration prevents unnecessary additional poles from being created in the circuit.

Both proposed circuits offer a wide frequency response range, making them superior to circuits built with traditional methods. The wide range of frequency responses increases the usability of these circuits in different applications and provides flexibility. For example, these circuits can be used in various analog signal processing applications, such as filtering, signal conversion, and signal conditioning.

Another important contribution of the study is the support of simulation and experimental results. LTspice simulations and laboratory experiments confirm the advantages and effectiveness of the proposed design. Simulations show that the circuit performs as expected theoretically, while experimental results demonstrate its reliability and practicality for real-world applications.

In conclusion, this work provides significant contributions from both theoretical and practical perspectives and proposes an innovative and effective solution for floating inductance simulation. The development of such circuits opens up new opportunities in the fields of electronic circuit design and signal processing, pushing the boundaries of existing technologies.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- Ahmed, A., & Demirel, H. (2023). DESIGN Third order Sinusoidal Oscillator Employing Current Differencing Cascaded Trans conductance Amplifiers. *Gazi University Journal of Science Part C: Design and Technology*, 11(3), 735-743. <https://doi.org/10.29109/gujsc.1290137>
- Bhaskar, D. R., Sharma, V. K., Monis, M., & Rizvi, S. M. I. (1999). New current-mode universal biquad filter. *Microelectronics Journal*, 30(9), 837-839. [https://doi.org/10.1016/S0026-2692\(99\)00019-1](https://doi.org/10.1016/S0026-2692(99)00019-1)

- Demirel, H. (2017). *Elektronik II*. Birsen Publishing, İstanbul.
- Demirel, H., & Ahmed, A. (2023a). A Low-Power 30MHz, 6th Order Bandpass Differential Gm-C Filter on Chip Utilizing Floating Current Source. *Kastamonu University Journal of Engineering and Sciences*, 9(2), 96-103. <https://doi.org/10.55385/kastamonujes.1395608>
- Demirel, H., & Ahmed, A. (2023b). New FinFet Transistor Implementation of Floating and Grounded Inductance Simulator Based on Active Elements. *Gazi Journal of Engineering Sciences*, 9(3), 647-653. <https://doi.org/10.30855/gmbd.0705095>
- Dikicioğlu, E., & Polat, B. (2024). Analysis of Current-Voltage Properties of Al/p-si Schottky Diode with Aluminium Oxide Layer. *Gazi University Journal of Science Part A: Engineering and Innovation*, 11(1), 137-146. <https://doi.org/10.54287/gujisa.1413932>
- Feng, J., Wang, C., Zang, M., & Ren, Y. (2011, January 18-20). *Realization of current-mode general nth-order filter based on current mirrors*. In: Proceedings of the 3rd International Conference on Advanced Computer Control (pp. 367-370), Harbin, China. <https://doi.org/10.1109/ICACC.2011.6016433>
- Horng, J. W., Lee, H., & Wu, J. Y. (2010). Electronically tunable third-order quadrature oscillator using CDTAs. *Radioengineering*, 19(2), 326-330.
- Horng, J. W. (2011). Current/voltage-mode third order quadrature oscillator employing two multiple outputs CCIs and grounded capacitors. *Indian Journal of Pure and Applied Physics*, 49(7), 494-498.
- Jaikla, W., & Lahiri, A. (2012). Resistor-less current-mode four-phase quadrature oscillator using CCCDTAs and grounded capacitors. *AEU - International Journal of Electronics and Communications*, 66(3), 214-218. <https://doi.org/10.1016/j.aeue.2011.07.001>
- Maheshwari, S. (2009). Quadrature oscillator using grounded components with current and voltage outputs. *IET Circuits, Devices & Systems*, 3(4), 153-160. <https://doi.org/10.1049/iet-cds.2009.0072>
- Mohammed, A. A., Mahmood, Z. K., & Demirel, H. (2024). New Z copy-current differencing transconductance amplifier active filter using FinFET transistor based current Mode Universal Filter. *Global Journal of Engineering and Technology Advances*, 18(02), 001-005. <https://doi.org/10.30574/gjeta.2024.18.2.0019>
- Mohammed, A. A., Demirel, H., & Mahmood, Z. K. (2023). Analysis fin field-effect transistor design with high-k insulators. *Nexo Revista Científica*, 36(06), 892-905. <https://doi.org/10.5377/nexo.v36i06.17445>
- Mohammed, A. A. & Demirel, H. (2023). Integration of Quadrature Oscillator and Floating Inductor in FinFET Transistor Design: Innovations and Applications. *Iranian Journal of Electrical & Electronic Engineering*, 19(4), 117-123. <https://doi.org/10.22068/ijeee.19.4.2862>
- Prommee, P., & Dejhan, K. (2002). An integrable electronic controlled sinusoidal oscillator using CMOS operational transconductance amplifier. *International Journal of Electronics*, 89(5), 365-379. <https://doi.org/10.1080/713810385>
- Sa-Ngiamvibool, W., & Jantakun, A. (2014). Quadrature oscillator using CCCCTAs and grounded capacitors with amplitude controllability. *International Journal of Electronics*, 101, 1737-1758. <https://doi.org/10.1080/00207217.2014.883905>
- Senani, R., Bhaskar, D. R., Singh, V. K., & Sharma, R. K. (2016). *Sinusoidal Oscillators and Waveform Generators Using Modern Electronic Circuit Building Blocks*. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-319-23712-1>
- Shkir, A. A. M., & Abdulazeez, M. Z. (2017). Design of Voltage Mode 6th Order Elliptic Band-pass Filter Using Z-Copy Current Follower Transconductance Amplifier) ZC-CFTA. *Kirkuk University Journal-Scientific Studies*, 12(2), 271-285. <https://doi.org/10.32894/kujss.2017.124963>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1521679>

Abrasion Behavior of Natural Aggregates and Slags in Turkey Aided by Micro-Deval Test

İslam GÖKALP^{1*} Volkan Emre UZ² ¹ Batman University, Faculty of Engineering and Architecture, Department of Civil Engineering, Batman, Türkiye² Izmir Institute of Technology, Faculty of Engineering, Department of Civil Engineering, İzmir, Türkiye

Keywords	Abstract
Abrasion Resistance	The properties of aggregates over their lifetime vary depending on their petrographic origin, procurement and sampling methodologies, regional characteristics and testing procedures. This study was designed to investigate the short and long-term wear properties of aggregates, including slags, using a simple, effective and inexpensive Micro-Deval (MD) test. To accomplish this, ten types of products were obtained from different regions, including natural aggregates and slags of different origin. The aggregates were abraded with different MD drum speeds ranging from 5250 to 52500 in accordance with ASTM D6928 standard. Percentage mass losses (PML) were determined after the treatments to analyze the extent of abrasion caused by short and long term abrasive forces depending on the aggregate type. Scanning electron microscopy (SEM) images of representative aggregates were taken to monitor the effect of abrasion on aggregate microstructure. According to the results, the PML of natural aggregates was observed significantly higher than that of slags, and the PML of slags after certain abrasion treatment tends to be stable, but not for natural ones. Based on regression analyses, a strong relationship between PML for individual aggregates was calculated, but a weak relationship was found based on the origin of the samples and total samples. SEM images taken from the surface of the aggregates confirmed the compatibility of the PML results with the abrasion characteristics showing the current situation.
Micro Deval	
Slags	
Natural Aggregate	
Durability	

Cite

Gökalp, İ., & Uz, V. E. (2024). Abrasion Behavior of Natural Aggregates and Slags in Turkey Aided by Micro-Deval Test. *GU J Sci, Part A, 11(3)*, 451-462. doi:[10.54287/guj.1521679](https://doi.org/10.54287/guj.1521679)

Author ID (ORCID Number)	Article Process
0000-0003-3198-3508	İslam GÖKALP
0000-0002-9328-4756	Volkan Emre UZ
	Submission Date 24.07.2024
	Revision Date 15.08.2024
	Accepted Date 04.09.2024
	Published Date 12.09.2024

1. INTRODUCTION

The properties of aggregates such as size, shape, texture, cleanliness, toughness, abrasion, polishing resistance, durability, robustness and chemical compositions significantly affect both the functional and structural performance of road pavements (Gökalp et al., 2016). Aggregates are exposed to different abrasive forces and environmental influences that can cause crushing, degradation, fragmentation and abrasion (Saghafi et al., 2024). Considering these, the selection of suitable aggregates is an important consideration for pavement construction that can withstand traffic loads and environmental conditions throughout its service life (Yildirim & Prezzi, 2011). In order to decide on the use of any material as an aggregate for pavement, a series of tests must be carried out according to the material specification. These can be listed as fragmentation, polishing, abrasion, segregation resistance tests, etc. These tests are used to determine toughness, durability and strength properties for short and long periods (Omary et al., 2015). Existing literature (Tutumluuer & Pan, 2008; Kwon et al., 2017; Gökalp et al., 2018) has proved that the aforementioned properties of aggregate vary significantly depending on petrographic origin, procurement and sampling methodology, regional characteristics and testing procedures, etc.

The Micro Deval (MD) test can be implemented on samples with numerous standard test methods (Strzałkowski & Kaźmierczak, 2021). In this study, the most common test method for measuring abrasion resistance aggregates was used to evaluate short and long-term abrasion characterizations based on ASTM D6928 (2017), which provide different grade type for evaluation. This test method has been widely used by some researchers in combination with some other tests such as Polishing Stone Value, Wehner-Schulze, Auckland Veneer Polishing, Los Angeles Fragmentation, Soundness with $MgSO_4$, Aachen Polishing Machine and Road Testing Machine, Scanning Electron and Fluorescence Light Microscopy, and Image Processing. To summarize the general results of those studies (Cafiso & Taormina, 2007; Mahmoud & Masad, 2007; Lane et al., 2011; Shabani et al., 2013; Ortiz & Mahmoud, 2014; Wang et al., 2015), the following can be emphasized that the MD test were used for abrasion assessment by percentage mass losses (PML) before and after different abrasion levels is not taken into account. This situation, which is seen as a gap in scientific studies, is thought to be worth studying, and in this context, this study was established.

As the need for new roads and rehabilitation and maintenance of existing roads increases, so does the demand for aggregates of the required quality. Unfortunately, sources of raw materials available in nature are diminishing (Gao et al., 2017). Researchers' interest has in recent years turned towards exploring alternative materials for sustainable transportation. The substitution of natural aggregates with alternatives is considered as an urgent need to overcome the supply problem and the search for alternative construction materials has focused on industrial waste or by-products (Aslani et al., 2023). The steel industry is the second largest in the world after oil and gas, and steel is a versatile material used as a main component in industries such as machinery, automobiles, construction, infrastructure, shipbuilding and electronics. Therefore, by-products (slags) account for 10-15% of crude steel production, which means that more than 210 million tons of slag has the potential to be generated after steel production (Gökalg et al., 2018; Yonar & Dikbař, 2022).

In the light of prior studies, it can be concisely stated that they have some superior properties compared to natural aggregates, such as resistance to polishing; abrasion, etc. Pasetto et al. (2023) indicate that researchers have also investigated both their possible uses and their impact on the environment. These studies have indicated slags as inert and/or hazardous materials and highlighted their environmentally friendly properties in a wide range of areas, including concrete-based construction works and granular particle-based coating applications for different layers. Using these materials instead of natural aggregates can reduce the supply problem and help conserve natural resources, which are dwindling day by day. Environmental regulations also require minimizing the disposal of such industrial by-products and mandate the reuse of these waste materials in many countries around the world. This problem must be overcome for present and future generations to live in a healthier environment.

In this study, the abrasion behavior of aggregates of different origins including slags simulated for short and long periods of time by MD device, which is widely used in determining the abrasion resistance of materials, was investigated. The aggregate materials used in this study were selected from a wide range of mineralogical properties from various quarries, industrial plants and/or disposal sites in different geographical regions of Turkey, with a total of ten types including limestone, basalt, rock and steel slags. A series of tests were performed to demonstrate the specific properties of each aggregate. Short and long term abrasion was achieved with the MD apparatus according to the B rating of ASTM D6928 (2017) standard with the number of revolutions ranging from 5250 to 52500. After the treatments, the PML was determined for all aggregates to examine the extent of wear caused by short and long term abrasive forces depending on the aggregate type. Relative differences in percentages with respect to the result obtained from standard practice are presented. SEM visually monitored the microstructure of the aggregate for representative aggregates before and after abrasion to show the effect of each abrasion level. As a result, the results of all laboratory-based experiments were analyzed in detail, not only with PML, which shows the wear characteristics of each material at certain levels, but also with SEM images, which provide information on the physical, mechanical and chemical properties and show the surface micro-texture development at certain wear levels.

2. MATERIAL AND METHOD

There was a rationale for material selection based on the geological structure of Turkey. Limestone was chosen because it is the most common natural aggregate and is widely used in construction works. Basalt and boulders

were preferred based on their better physical and mechanical properties (Gökalp et al., 2018). Therefore, ten types of coarse aggregates, both natural aggregate and slag were evaluated in this study. These materials, consisting of boulders, limestone, basalt, ferrochrome slag and electric arc furnace steel slag, were obtained from privately operated quarries located in different regions of Turkey and from the facilities and/or disposal areas of companies operating in steel production. The identification number, material lithology, types of rocks source region and regional distribution of the aggregates are presented in Table 1. For showing the region of supplied aggregate and slags in a map, locations of the material sources on geological map of Turkey is given in Figure 1.

Table 1. Aggregate distribution and identification

Sample ID	Lithology	Types of Rocks	Source District	Region
LS-1	Limestone	Sedimentary	Adana- Ceyhan	Mediterranean
LS-2	Limestone	Sedimentary	Antakya - Kırıkhan	Mediterranean
LS-3	Limestone	Sedimentary	Mersin- Tarsus	Mediterranean
BS-1	Basalt	Igneous	Niğde - Bor	Central Anatolia
BS-2	Basalt	Igneous	Kayseri	Central Anatolia
BLD	Boulder	River-Basin Aggregate	Kahramanmaraş- Aksu	Mediterranean
EAF-1	EAF Slag	By-Product	Antakya- İskenderun	Mediterranean
EAF-2	EAF Slag	By-Product	Osmaniye	Mediterranean
EAF-3	EAF Slag	By-Product	Antakya- İskenderun	Mediterranean
FER	Ferrochrome Slag	By-Product	Elazığ	Eastern Anatolia

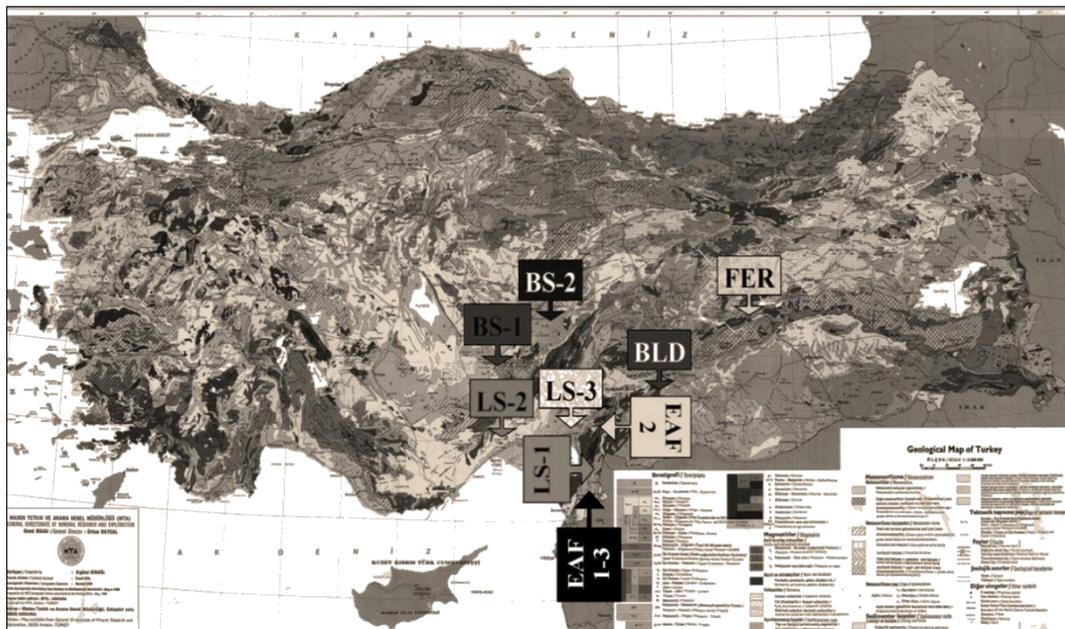


Figure 1. Locations of the material sources on geological map of Turkey

2.1. Material Characterization

The materials used in this study were characterized for their physical and mechanical properties by a series of standard tests. These tests included fragmentation, polishing resistance, soundness with $MgSO_4$ solution, dry unit weight and water absorption. Table 2 shows the average test results obtained for all specimens. Chemical composition determination of materials is an extremely important issue. In this context, the chemical

composition of each aggregate was determined by X-ray fluorescence analysis according to the standard test method EN 15309 (2007). Figure 2 shows the chemical composition of each aggregate.

The results in Table 2 showed that naturally occurring aggregates exhibited different properties depending on their source. Among them, polishing resistance of natural aggregates are week compared to slags except of EAF-3. Dry unit weight of slags are higher than that of natural ones due to their chemical composition. Figure 2 indicated that the chemical composition of aggregates varies depending on their geological formation in nature and production processes in industrial plants, and all samples have chemical contents reflecting their formation.

Table 2. Aggregate physical and mechanical properties

Sample ID	Fragmentation Resistance (%) (EN 1097-2)	Soundness (%) (EN 1367-2)	Polishing Resistance (PSV) (EN 1097-8)	Dry Unit Weight (g/cm ³) (EN 1097-6)	Water Absorption (%) (EN 1097-6)
LS-1	24.4	2.4	41.2	2.7	0.4
LS-2	16.2	3.0	43.2	2.7	0.2
LS-3	24.4	8.1	41.6	2.7	0.3
BS-1	12.0	6.9	61.0	2.6	2.0
BS-2	25.9	9.4	52.4	2.7	1.4
BLD	17.6	6.2	57.9	2.7	0.9
EAF-1	22.9	2.3	76.1	3.4	1.8
EAF-2	25.3	8.3	59.0	3.4	2.5
EAF-3	29.7	3.7	54.1	3.3	2.9
FER	16.5	6.1	61.7	2.9	1.1

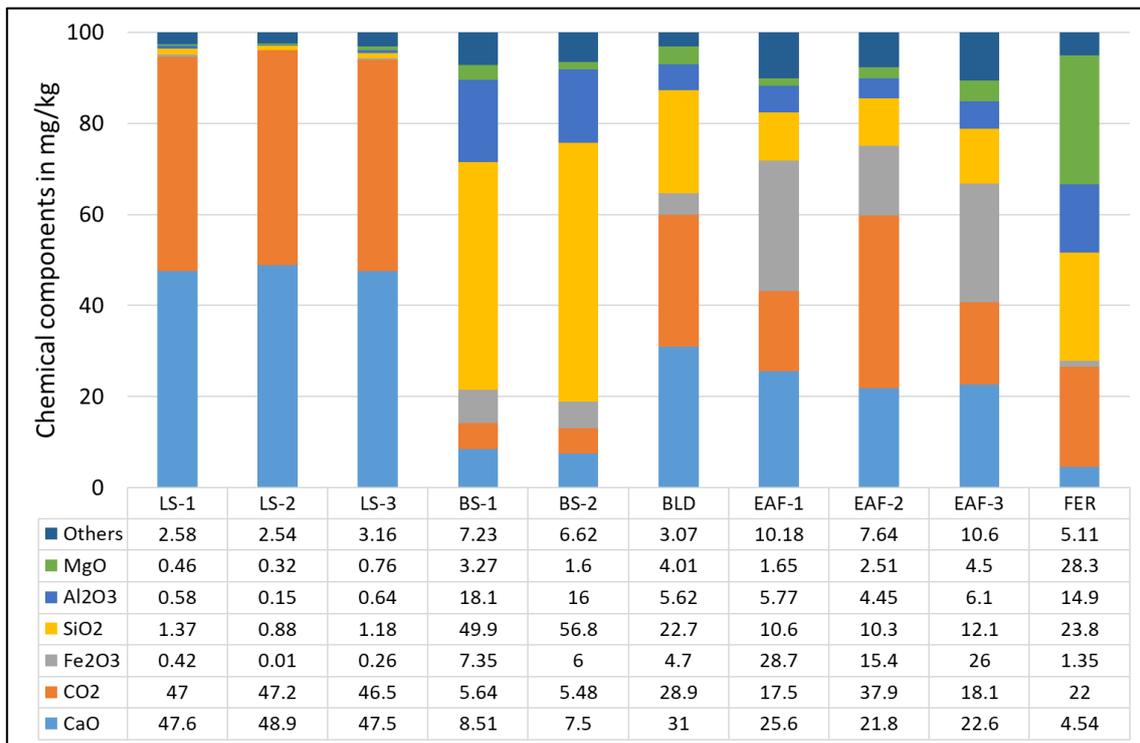


Figure 2. Chemical compositions of aggregates

2.2. Aggregate Abrasion Process

The standard size of the aggregate tested in accordance with ASTM D6928 (2017) was varied. The aggregate sizes ranged from 19.0 mm to 4.75 mm according to three different application types. Table 3 presents the aggregate sizes and mass configurations defined in the standard. It is worth noting here that all aggregates collected were extracted from such long and platy aggregates according to ASTM D4791 (2012) and the cubic particles remaining after separation were subjected to tests. Our concern was to avoid misleading test results and to obtain specimens that standardized in shape. The abrasion levels of the aggregates were evaluated according to the rotation numbers (RNs) of the MD test according to the ASTM D6928-B standard given in Table 4.

Table 3. Aggregate physical and mechanical properties

Sieve Size (mm)	Passing (mm)	19.0	16.0	12.5	9.5	6.3
Grade of ASTM D6928	A (g)	375	375	750	-	-
	B (g)	-	-	750	375	375
	C (g)	-	-	-	750	750

Table 4. Abrasion level identification

Abrasion Level	Abrasion-Free	1st	2nd	3rd	4th	5th
Revolution Numbers	0	5250	10500	21000	31500	52500
Case of Standard RN	None	Half	Standard	Double	Triple	Quintuple

The application steps for the test are as follows:

1. 1500 ± 5 g of test samples are taken from a given amount in the pile.
2. The samples are washed to remove dust and dried in an oven at 110 ± 5 °C to evaporate moisture until reaching constant mass.
3. 2.0 ± 0.05 liters of water is added to the MD drum at a given temperature and the aggregate is poured over the water-filled drum for conditioning for at least 1 hour.
4. A 5000 ± 5 g steel sphere with a diameter of 10 ± 0.5 mm is placed inside the drum and the drum is closed for being ready to rotate.
5. Rotation is conducted with 100 ± 5 rpm for three cases as follows;
 - 12000 ± 100 revolutions for 19/9.5 mm classification (Grade A),
 - 10500 ± 100 revolutions for 12.5/4.75 mm classification (Grade B), and
 - 9500 ± 100 revolutions for 9.5/4.75 mm classification (Grade C).

After the test, the aggregates are dried and sieved through a 1.18 mm sieve. The mass (m) of aggregate retained on the 1.18 mm sieve is weighted and operator determines PML in percent with the Equation 1.

$$\text{PML (\%)} = (1500 - m)/1500 \quad (1)$$

3. RESULTS AND DISCUSSION

This part of the present study consists of two parts. Firstly, the results of the wear properties for each aggregate at the applied wear levels are presented and discussed in detail. Secondly, SEM images for representative aggregates in the original condition and at the highest wear levels are given and concisely discussed.

3.1. Abrasion Characteristic of Aggregates

The abrasion levels achieved by the MD apparatus with reference to different drum speeds in the B classification standard according to ASTM D6928-B Grade is described here. PML results at different abrasion levels and statistical data is given in Table 5.

Table 5. PML at different abrasion levels with statistical data

Abrasion Level	Revolutions Numbers	Sample ID	LS-1	LS-2	LS-3	BLD	BS-1
1 st	5250	Mean	5.42	9.85	5.30	4.98	4.00
		Std Dev	0.164	0.147	0.385	0.065	0.092
		CoV	0.030	0.015	0.073	0.013	0.023
2 nd	10500	Mean	9.05	16.28	9.70	8.14	8.14
		Std Dev	0.115	0.200	0.040	0.189	0.195
		CoV	0.013	0.012	0.004	0.023	0.024
3 rd	21000	Mean	14.94	26.08	13.57	13.70	11.43
		Std Dev	0.084	0.081	0.085	0.193	0.465
		CoV	0.006	0.003	0.006	0.014	0.041
4 th	31500	Mean	19.83	32.93	18.26	17.57	15.35
		Std Dev	0.026	0.886	0.001	0.255	0.020
		CoV	0.001	0.027	0.000	0.015	0.001
5 th	52500	Mean	27.33	51.71	26.05	23.15	22.20
		Std Dev	0.030	0.037	0.250	0.030	0.300
		CoV	0.001	0.001	0.010	0.001	0.014
Abrasion Level	Revolutions Numbers	Sample ID	BS-2	EAF-1	EAF-2	EAF-3	FER
1 st	5250	Mean	4.80	5.64	5.33	8.97	5.47
		Std Dev	0.020	0.184	0.120	0.191	0.125
		CoV	0.004	0.033	0.023	0.021	0.023
2 nd	10500	Mean	6.81	8.01	8.01	12.24	8.31
		Std Dev	0.020	0.290	0.175	0.011	0.185
		CoV	0.003	0.036	0.022	0.001	0.022
3 rd	21000	Mean	10.33	11.61	10.69	16.63	11.68
		Std Dev	0.050	0.078	0.190	0.137	0.077
		CoV	0.005	0.007	0.018	0.008	0.007
4 th	31500	Mean	14.01	14.25	14.46	19.91	15.67
		Std Dev	0.060	0.146	0.265	0.360	0.010
		CoV	0.004	0.010	0.018	0.018	0.001
5 th	52500	Mean	18.65	16.91	17.70	20.84	19.95
		Std Dev	0.150	0.053	0.100	0.230	0.250
		CoV	0.008	0.003	0.006	0.011	0.013

Std Dev: Standard Deviation, Mean: Arithmetic Mean of results; CoV: Coefficient of Variations

From Table 5 that the abrasion characteristics of the aggregates in terms of PMLs differ from each other depending on their type and abrasion levels. Regarding the PML of natural aggregates, LS-2 showed the worst abrasion characteristics at all applied cycles, where the PML values were 9.85% at the lowest abrasion level. On the other hand, basalt and rock exhibited better abrasion properties compared to other natural materials. However, basalts were relatively better than boulders. It is important to note that the initial conditions were quite close for these. It is apparent that the level of abrasion increases and this trend of loss proceeds almost linearly. On the other hand, the slags are quite superior in terms of abrasion characteristics. This is because at the highest abrasion level the PML for slags ranges from 16.91 to 20.84, which can be a kind of close range, while for natural ones the PML ranges from 18.69 to 51.73, which means about three times far apart. It is worth noting here that the slags exhibit a stationary phase against abrasion as the abrasion level increases.

3.2. Regression Analysis and Mathematical Models

Regression analysis based on different mathematical models, different material types were used. In this context, different mathematical models were used in the study in linear, exponential, logarithmic and polynomial functional cases. Aggregates are known to exhibit different physical and mechanical properties due to their morphologic structure. For this reason, the materials were considered individually and regression analyses based on the mathematical models in question were formulated in this direction (Table 6). Based on the results of the regression analysis of the individual materials as given in Table 6, the author can state that the R^2 values for natural aggregates ranged roughly from 0.85 to 0.99. On the other hand, the values for slag aggregates ranged from 0.78 to 0.99. The higher R^2 values were calculated polynomial-based mathematical models, where the lower were for exponential one.

The other case considered for the regression analyses is related to the natural and by-product supplied methods. Thus, the use of both natural aggregates and slags in construction is based on the results obtained from the same test. Regression analysis was also carried out for each aggregate by origin separately and then all together in order to make an assessment (Table 7). According to the results presented in Table 7, lower R^2 values were observed for each combined case compared to the individual cases, ranging from 0.562 to 0.844. The R^2 values determined for slags are larger than those obtained for both natural aggregates and all aggregates. Also, lower R^2 values were observed for the sample population when all aggregates were considered. Such an analysis may be the result of a new approach.

3.3. SEM-based Visual Analyses

SEM has been used as a powerful magnification tool that focuses electron beams to obtain the desired detailed information about materials. The purpose of using SEM in this study is to monitor the aggregate surface at the micro level for the highest level of wear and no wear monitored through the instrument. Representative samples of all aggregate types selected for imaging at a fixed magnification (60 X) such as LS-3, BS-2, BLD, EAF-1 and FER are 8-10 mm in size and cubic in shape. A series of SEM images taken during the analysis of aggregate without abrasion and at the highest abrasion levels are given in Figure 3.

It is expected that the surface texture of each aggregate in its original condition is distinctly different and the surface of each becomes smoother at the highest levels of abrasion. When analyzed the figures individually, the following can be remarked;

Limestone have a smoother surface than the other aggregates before and after different abrasion level. Basalt, an igneous rock, has a naturally porous structure, which is demonstrated in the images. Boulder show both surface characteristics of limestone and basalt, where some part is smooth and some part is porous, which is expected case. Slags shows porous structure likewise in the state of abrasion free case after 5th abrasion level and this can be linked with the industrial production procedure, which show re-texturing the surface by itself.

Table 6. Different regression analysis and mathematical models for individual-based

Sample ID	Mathematical Models (where, y= PML (%) and x: MD Drum RN number)			
	Linear	Exponential	Logarithmic	Polynomial
LS-1	R ² = 0.984 y = 0.0005x + 4.2291	R ² = 0.880 y = 6.0758e ^{3E-05x}	R ² = 0.960 y = 9.3892ln(x) - 76.709	R ² = 0.999 y = -4E-09x ² + 0.0007x + 1.9716
LS-2	R ² = 0.996 y = 0.0009x + 6.5332	R ² = 0.914 y = 10.688e ^{3E-05x}	R ² = 0.920 y = 17.176ln(x) - 140.97	R ² = 0.996 y = -2E-09x ² + 0.001x + 5.6999
LS-3	R ² = 0.988 y = 0.0004x + 4.3484	R ² = 0.881 y = 6.0839e ^{3E-05x}	R ² = 0.947 y = 8.5829ln(x) - 69.545	R ² = 0.994 y = -2E-09x ² + 0.0006x + 3.0344
BLD	R ² = 0.969 y = 0.0004x + 4.3152	R ² = 0.861 y = 5.643e ^{3E-05x}	R ² = 0.975 y = 7.9073ln(x) - 63.991	R ² = 0.999 y = -5E-09x ² + 0.0007x + 1.6956
BS-1	R ² = 0.984 y = 0.0004x + 3.3364	R ² = 0.857 y = 4.8179e ^{3E-05x}	R ² = 0.949 y = 7.485ln(x) - 61.134	R ² = 0.992 y = -2E-09x ² + 0.0005x + 2.1318
BS-2	R ² = 0.986 y = 0.0003x + 3.8098	R ² = 0.916 y = 4.9896e ^{3E-05x}	R ² = 0.948 y = 5.9757ln(x) - 47.647	R ² = 0.999 y = -2E-09x ² + 0.0004x + 2.5317
EAF-1	R ² = 0.937 y = 0.0002x + 5.6105	R ² = 0.855 y = 6.1794e ^{2E-05x}	R ² = 0.989 y = 4.9956ln(x) - 37.678	R ² = 0.999 y = -4E-09x ² + 0.0005x + 3.2861
EAF-2	R ² = 0.962 y = 0.0003x + 5.0014	R ² = 0.883 y = 5.8023e ^{2E-05x}	R ² = 0.969 y = 5.3656ln(x) - 41.349	R ² = 0.994 y = -3E-09x ² + 0.0005x + 3.197
EAF-3	R ² = 0.839 y = 0.0002x + 9.7614	R ² = 0.780 y = 10.017e ^{2E-05x}	R ² = 0.979 y = 5.5172ln(x) - 38.354	R ² = 0.999 y = -8E-09x ² + 0.0007x + 5.6302
FER	R ² = 0.974 y = 0.0003x + 4.9074	R ² = 0.891 y = 5.9455e ^{3E-05x}	R ² = 0.965 y = 6.2416ln(x) - 48.955	R ² = 0.997 y = -3E-09x ² + 0.0005x + 3.0921

Table 7. Different mathematical models for origin and unified-based

Aggregate Type	Mathematical Models (where, y= PML (%) and x: MD Drum RN number)			
	Linear	Exponential	Logarithmic	Polynomial
Natural	R ² = 0.604 y = 0.0005x + 4.428	R ² = 0.696 y = 6.1403e ^{3E-05x}	R ² = 0.578 y = 9.4194ln(x) - 76.666	R ² = 0.608 y = -3E-09x ² + 0.0006x + 2.8442
Slag	R ² = 0.793 y = 0.0003x + 6.3202	R ² = 0.738 y = 6.7978e ^{2E-05x}	R ² = 0.833 y = 5.53ln(x) - 41.584	R ² = 0.844 y = -5E-09x ² + 0.0005x + 3.8013
Combined	R ² = 0.572 y = 0.0004x + 5.1853	R ² = 0.632 y = 6.3953e ^{3E-05x}	R ² = 0.562 y = 7.8636ln(x) - 62.633	R ² = 0.582 y = -4E-09x ² + 0.0006x + 3.227

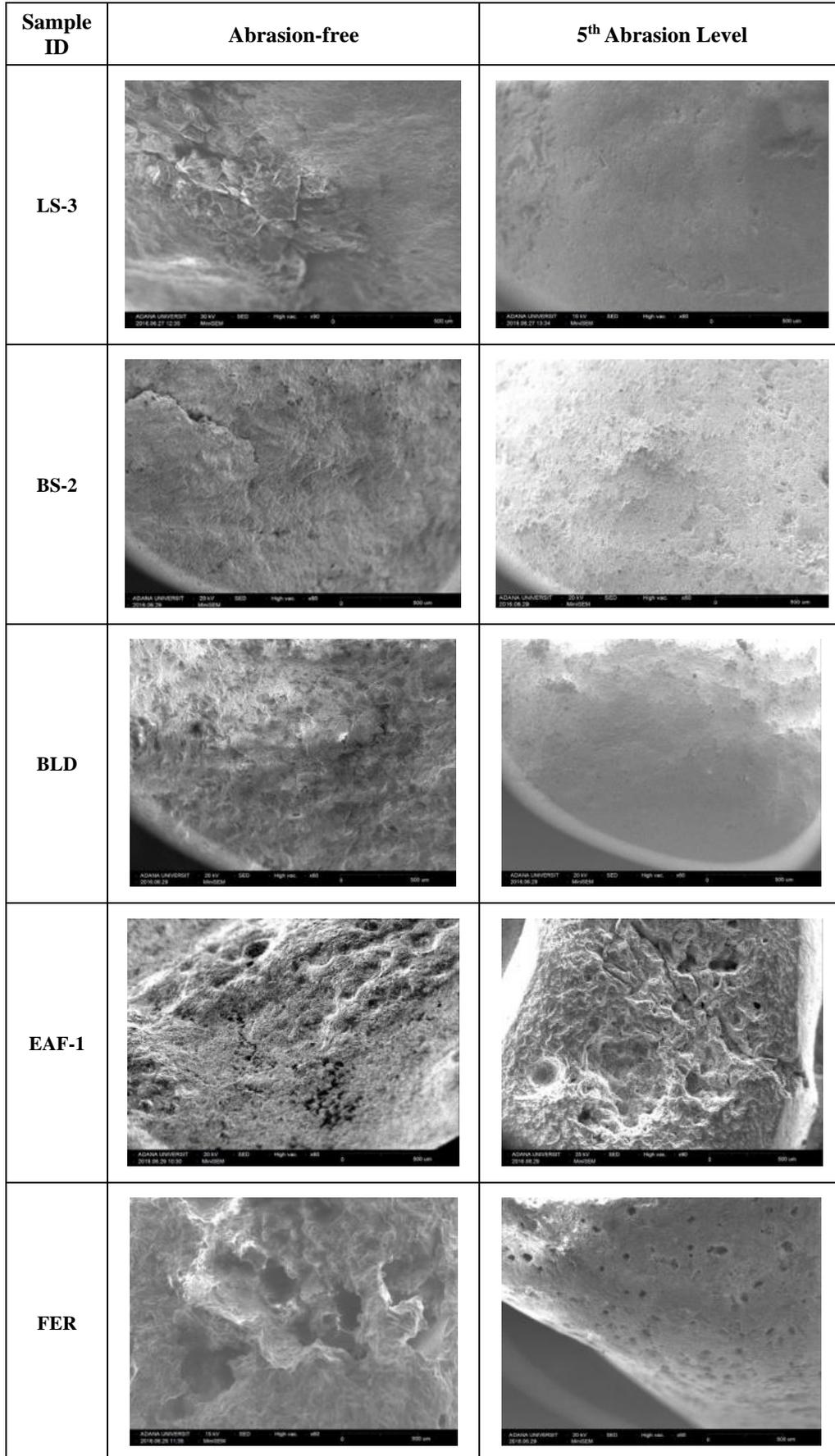


Figure 3. Visual analyses with SEM images of representative aggregate and slags

4. CONCLUSION

In the present study, the short and long term abrasion properties of aggregates obtained from different sources containing slag were investigated by simulating abrasion with different drum speeds between 5250 and 52500 according to the B grading of ASTM D6928 with MD apparatus. In this context, a total of ten types of aggregates including limestone, basalt, rock and steel slag were obtained from different regions of Turkey. A series of tests were performed on each aggregate to determine certain physical, mechanical and chemical properties. The PML after abrasion treatments for all aggregates was determined to assess the status caused by short and long term abrasion. Finally, SEM images were taken for representative aggregates to provide a visual comparison of the effect of wear on the surface structure of the aggregate. According to the laboratorial test and statically and visual analyses, followings can be highlighted.

1. Slags have significant superior properties compared to natural aggregates by means of abrasion resistance tested with MD.
2. The abrasion properties of aggregates differ from each other depending on their origin and abrasion levels.
3. The PMLs of natural aggregates were significantly higher than that of slags at each abrasion level.
4. The abrasion level increases and the PML progresses for natural aggregates, but this trend is not observed for slags, as they exhibit a more stable phase against abrasion.
5. Slags have a more porous surface than natural aggregates, as clearly seen from the SEM images.
6. Individual, origin and compound based regression analyses showed significantly different R^2 values and mathematical models.

As a result, aggregate is an important component in the production of bituminous mixtures, concrete, mortars to be used in buildings, filling materials, railway ballast, etc. in all areas of the construction industry. Aggregates must be robust and abrasion resistant under mechanical forces that cause fragmentation, polishing, crushing and deterioration during storage, mixing, paving, and compaction and under traffic loads of different intensities. This is because the properties of aggregate significantly affect both the functional and structural performance of any construction work. The selection of suitable aggregates is of utmost importance when considering their short and long-term performance for both economic and safety reasons. Slags can be considered the most abrasion resistant materials when their properties can become more stable after a certain abrasion process. Therefore, the use of slags in pavements when their surfaces are subjected to intense abrasion can bring not only technical, but also economic and ecological benefits.

AUTHOR CONTRIBUTIONS

Conceptualization, İ.G. and V.E.U.; methodology, I.G. and V.E.U.; title, I.G.; validation, I.G. and V.E.U. laboratory work, İ.G.; formal analysis, İ.G.; research, I.G. and V.E.U.; sources, V.E.U.; data curation, I.G. and V.E.U.; manuscript-original draft, İ.G.; manuscript-review and editing, I.G. and V.E.U.; visualization, İ.G.; supervision, V.E.U.; project management, V.E.U. funding, V.E.U. All authors have read and legally accepted the final version of the article published in the journal.

ACKNOWLEDGEMENT

The authors gratefully acknowledge financial support from the Scientific and Technological Research Council of Turkey-TÜBİTAK (Project No: 215M049). They would also like to thank the Department of Chief Engineering of Research and Development of 5th Regional Directorate of Highways within General Directorate of Highways for providing some of the test samples and for assistance.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Aslani, A., Hachem-Vermette, C., & Zahedi, R. (2023). Environmental impact assessment and potentials of material efficiency using by-products and waste materials. *Construction and Building Materials*, 378, 131197. <https://doi.org/10.1016/j.conbuildmat.2023.131197>
- ASTM (2012). Standard Test Method for Flat Particles, Elongated Particles, or Flat and Elongated Particles in Coarse Aggregate (ASTM D4791).
- ASTM (2017). Standard Test Method for Resistance of Coarse Aggregate to Degradation by Abrasion in the Micro-Deval Apparatus (ASTM D6928-17).
- Cafiso, S., & Taormina, S. (2007). Texture analysis of aggregates for wearing courses in asphalt pavements. *International Journal of Pavement Engineering*, 8(1), 45-54. <https://doi.org/10.1080/10298430600898307>
- EN (2007). Characterization of waste and soil - Determination of elemental composition by X-ray fluorescence (EN 15309).
- Gao, J., Sha, A., Wang, Z., Tong, Z., & Liu, Z. (2017). Utilization of steel slag as aggregate in asphalt mixtures for microwave deicing. *Journal of Cleaner Production*, 152, 429-442. <https://doi.org/10.1016/j.jclepro.2017.03.113>
- Gökalg, I., Uz, V. E., & Saltan, M. (2016). Testing the abrasion resistance of aggregates including by-products by using Micro Deval apparatus with different standard test methods. *Construction and Building Materials*, 123, 1-7. <https://doi.org/10.1016/j.conbuildmat.2016.06.141>
- Gökalg, İ., Uz, V. E., Saltan, M., & Tutumluer, E. (2018). Technical and environmental evaluation of metallurgical slags as aggregate for sustainable pavement layer applications. *Transportation Geotechnics*, 14, 61-69. <https://doi.org/10.1016/j.trgeo.2017.10.003>
- Kwon, J., Kim, S.-H., Tutumluer, E., & Wayne, M. H. (2017). Characterisation of unbound aggregate materials considering physical and morphological properties. *International Journal of Pavement Engineering*, 18(4), 303-308. <https://doi.org/10.1080/10298436.2015.1065997>
- Lane, D. S., Druta, C., Wang, L., & Xue, W. (2011). Modified micro-deval procedure for evaluating the polishing tendency of coarse aggregates. *Transportation Research Record: Journal of the Transportation Research Board*, 2232(1), 34-43. <https://doi.org/10.3141/2232-04>
- Mahmoud, E., & Masad, E. (2007). Experimental methods for the evaluation of aggregate resistance to polishing, abrasion, and breakage. *Journal of Materials in Civil Engineering*, 19(11), 977-985. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2007\)19:11\(977\)](https://doi.org/10.1061/(ASCE)0899-1561(2007)19:11(977))
- Omary, S., Ghorbel, E., & Wardeh, G. (2015, September 21-23). *Physical and Mechanical Properties of Recycled Demolition Gravels Submitted to Freeze/Thaw in Comparison to Natural Ones*. In: C.Hellmich, B. Pichler, & J. Kollegger (Eds.) Proceedings of the 10th International Conference on Mechanics and Physics of Creep, Shrinkage, and Durability of Concrete and Concrete Structures, CONCREEP 10 (pp. 833-842). American Society of Civil Engineers, Vienna, Austria. <https://doi.org/10.1061/9780784479346.099>
- Ortiz, E. M., & Mahmoud, E. (2014). Experimental procedure for evaluation of coarse aggregate polishing resistance. *Transportation Geotechnics*, 1(3), 106-118. <https://doi.org/10.1016/j.trgeo.2014.06.001>
- Pasetto, M., Baliello, A., Giacomello, G., & Pasquini, E. (2023). The use of steel slags in asphalt pavements: A state-of-the-art review. *Sustainability*, 15(11), 8817. <https://doi.org/10.3390/su15118817>
- Saghafi, M., Abdallah, I. N., Ahmad, M., Nazarian, S., & Herrera, C. (2024). Correlating Aggregate Friction Test Results Under Accelerated Laboratory Polishing and Aggregate Crushing. *International Journal of Pavement Research and Technology*, 17(2), 423-434. <https://doi.org/10.1007/s42947-022-00245-z>
- Shabani, S., Ahmadinejad, M., & Ameri, M. (2013). Developing a model for estimation of polished stone value (PSV) of road surface aggregates based on petrographic parameters. *International Journal of Pavement Engineering*, 14(3), 242-255. <https://doi.org/10.1080/10298436.2012.693179>
- Strzałkowski, P., & Kaźmierczak, U. (2021). Wear and fragmentation resistance of mineral aggregates—a review of micro-Deval and Los Angeles tests. *Materials*, 14(18), 5456. <https://doi.org/10.3390/ma14185456>

- Tutumluer, E., & Pan, T. (2008). Aggregate morphology affecting strength and permanent deformation behavior of unbound aggregate materials. *Journal of Materials in Civil Engineering*, 20(9), 617-627. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2008\)20:9\(617\)](https://doi.org/10.1061/(ASCE)0899-1561(2008)20:9(617))
- Wang, D., Wang, H., Bu, Y., Schulze, C., & Oeser, M. (2015). Evaluation of aggregate resistance to wear with Micro-Deval test in combination with aggregate imaging techniques. *Wear*, 338-339, 288-296. <https://doi.org/10.1016/j.wear.2015.07.002>
- Yildirim, I. Z., & Prezzi, M. (2011). Chemical, mineralogical, and morphological properties of steel slag. *Advances in Civil Engineering*, 2011, 463638. <https://doi.org/10.1155/2011/463638>
- Yonar, F., & Dikbaş, H. A. (2022). Çelikhane Cürufunun Türkiye’de Balastaltı Malzemesi Olarak Kullanılabilirliğinin Ulusal Ve Uluslararası Şartnameler Kapsamında Değerlendirilmesi. *Kırklareli Üniversitesi Mühendislik ve Fen Bilimleri Dergisi*, 8(2), 439-469. <https://doi.org/10.34186/klujes.1184990>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1485920>

Modeling Cardiovascular Flow with Artificial Viscosity: Analyzing Navier-Stokes Solutions and Simulating Cardiovascular Diseases

Hilal KARADAVUT¹ Gülnur HAÇAT^{1,2*} Aytekin ÇIBIK³ ¹ Gazi University, Graduate School of Natural and Applied Sciences, Ankara, Türkiye² Yalova University, Department of Business Administration, Yalova, Türkiye³ Gazi University, Department of Mathematics, Ankara, Türkiye

Keywords	Abstract
Artificial Viscosity	In this paper, the numerical solutions of the Navier-Stokes equations (NSE) used for modeling the flow in the cardiovascular system are investigated using the Finite Element Method (FEM). A fully discrete solution scheme of the NSE and its stability and error analysis are presented. Artificial viscosity stabilization is added to the fully discrete scheme to better model the real flow structure and to remove non-physical oscillations. Numerical tests are also presented to demonstrate the effectiveness of the resulting scheme. Simulations analyzing the flow structure in the case of cardiovascular diseases such as atherosclerosis and brain aneurysm are presented in detail along with wall shear stress values.
Cardiovascular Diseases	
Finite Element Method	
Navier-Stokes Equations	
Cite	
Karadavut, H., Haçat, G., & Çıbık, A. (2024). Modeling Cardiovascular Flow with Artificial Viscosity: Analyzing Navier-Stokes Solutions and Simulating Cardiovascular Diseases. <i>GU J Sci, Part A, 11(3)</i> , 463-480. doi:10.54287/guj.1485920	
Author ID (ORCID Number)	Article Process
0000-0003-1247-8004	Submission Date 17.05.2024
0000-0001-7343-8466	Revision Date 18.07.2024
0000-0003-3571-4137	Accepted Date 06.08.2024
	Published Date 24.09.2024

1. INTRODUCTION

The cardiovascular system consists of three main components: blood, the heart that pumps blood to the body, and the vessels that distribute the pumped blood to the body. The cardiovascular system maintains the body's functions by transporting oxygen, nutrients and hormones to the body's cells and by removing waste products and carbon dioxide produced by the cells (Nair, 2017; Alimov, 2023). Cardiovascular diseases occur when impairments in the functions of the cardiovascular system affect blood circulation. Cardiovascular diseases are defined as disorders of the heart, blood vessels and blood function. These diseases can include vascular narrowing, blockage, weakening or widening, heart failure and stroke. According to the World Health Organization, cardiovascular diseases are the leading cause of death worldwide (WHO, 2017; Gaidai et al., 2023). Although statistics on cardiovascular disease provide data, a better understanding of the cardiovascular system is important.

Mathematical modeling and simulation of blood flow aim to better understand the conditions that influence the functioning of the system when cardiovascular diseases occur. The emergence of more powerful computers with improved image processing and geometry extraction techniques and the development of better algorithms have increased the demand for blood flow modeling among researchers (Formaggia et al., 2009; Taylor et al., 2023).

Mathematical models of blood flow can be divided into Newtonian models and non-Newtonian models. Newtonian models are models that neglect shear analysis and viscoelastic effects and can be applied to large vessels. Non-Newtonian models are those that take viscoelastic effects into account and can be applied to veins with a diameter of less than 1 mm. The non-Newtonian behavior makes the viscosity dependent on the shear

*Corresponding Author, e-mail: gulnur.yilmazoglu@yalova.edu.tr

rate, which increases the computational cost, so Newtonian models are usually used for simplicity. Therefore, models that exhibit Newtonian flow, such as the NSE, are often preferred to model blood flow in the cardiovascular system. The NSE is considered an important tool in understanding the movement of blood between vessels and its effects on the overall health of the system. The most general forms of these incompressible equations governing blood flow are as follows:

$$\begin{aligned}
 u_t + (u \cdot \nabla)u - \nu \Delta u + \nabla p &= f, & (0, T] \times \Omega \\
 \nabla \cdot u &= 0 & (0, T] \times \Omega \\
 u(x, 0) &= u_0(x) & x \in \Omega \\
 u &= 0 & \Gamma_{Wall} \\
 n \cdot \sigma &= 0 & \Gamma_{out} \\
 u &= c & \Gamma_{in} .
 \end{aligned} \tag{1}$$

Here u represents the fluid velocity, p is the pressure, f is the force per unit mass of the fluid, ν the viscosity and σ the Cauchy stress tensor. Suppose that Ω represents a fluid domain and the boundaries of Ω consist of two distinct boundaries, Γ_{in} and Γ_{out} , with Γ_{Wall} representing the inner wall of the artery. The boundary of the section of the artery under consideration's input flow is denoted by Γ_{in} , and the border of its outflow flow by Γ_{out} . Neumann boundary condition on Γ_{out} , no-slip Dirichlet boundary condition on Γ_{Wall} , and Dirichlet boundary condition on Γ_{in} are all taken into consideration (Formaggia et al., 2009). Also, the function c , which stabilizes the input flow, is set as a parabolic function for the simulations.

Numerical solutions of these equations are used to understand the dynamics of blood flow, pressure changes in vessel walls and flow velocities (Quarteroni et al., 2002; Selmi et al., 2019; Ali et al., 2024). Besides the NSE, the most common method for the numerical solution of many partial differential equations is the finite element method, which involves dividing the domain into elements and approximating the solution as a linear combination of basis functions on these elements. In this formulation, large variations in some fluid properties can adversely affect the stability of numerical methods. Therefore, the artificial viscosity method is a technique used to provide stabilization in such cases. If the parameter of the method is chosen small, there is less artificial viscosity effect and more accurate results can be obtained with a better selection. On the other hand, if the parameter of the method is chosen larger, more artificial viscosity effects occur in the flow and the solution becomes more stable, but at the same time, more physical flow details may be lost. Therefore, a balanced selection of the parameter of the method is crucial to achieve accurate results. By stabilizing the fluid structure, artificial viscosity controls fluctuations and increases the stability of the numerical solution. Artificial viscosity techniques help CFD simulations produce more accurate results and predictions closer to real-world fluid dynamics (Manzari, 1999; Ma et al. 2022). In their work, Margolin and Lloyd-Ronning (2023) summarized the history of artificial viscosity, from its origins to current research and new directions for improvement. The work of Cook and Cabot (2005) has a similar structure.

Together with solutions of velocity, pressure fields and streamlines, we can calculate the wall shear stress ($WSS = \mu(du/dy)$) by analyzing the solution data. These data provide extremely valuable clues for understanding blood flow phenomena. Research shows that a continuous flow or oscillating shear stress can be signs of vessel anomalies, causing continuous damage to endothelial cells (Fisher et al., 2001; Chiu & Chien, 2011). However, measuring these stress values in a real patient can be quite challenging. This is where mathematical models and simulations can help us better understand real-world situations (Velten et al. 2024). These models can be integrated into the diagnosis and treatment of patients, thus contributing significantly to clinical practice and disease prognosis (Reneman & Hoeks, 2008; Formaggia et al., 2010; Arjmandi-Tash et al., 2011).

In this paper, a fully discretized scheme of NSE with artificial viscosity is introduced and numerical analysis and experiments are presented. In this context, after the necessary preliminaries in section 2, the stability and error analysis of the proposed scheme are presented in section 3. In section 4, numerical tests are performed to verify the accuracy of the theoretical results and simulations of cardiovascular conditions such as atherosclerosis and aneurysm are presented. In section 5, the benefits of the obtained results and future directions are discussed.

2. MATHEMATICAL PRELIMINARIES AND MATHEMATICAL MODELING

Since the FEM is used, the standard function spaces for the velocity field and pressure spaces are selected as follows (Adams, 1975):

$$X = H_0^1(\Omega), \text{ and } Q = L_0^2(\Omega).$$

Let V be a weak subspace of X , then

$$V := \{v \in X, (\nabla \cdot q, v) = 0, \forall q \in Q\}.$$

The dual norm of a function is given by

$$\|f\|_{-1} := \sup_{v \in H_0^1(\Omega)} \frac{(f, v)}{\|\nabla v\|}.$$

For $(v, q) \in (X, Q)$, the weak form of the NSE to be used is given below:

$$(u_t, v) + \nu(\nabla u, \nabla v) + b(u, u, v) - (p, \nabla \cdot v) + (\nabla \cdot u, q) = (f, v). \quad (2)$$

The nonlinear term in equation (2) is expressed in the following skew-symmetric trilinear forms:

$$b(u, v, w) := \frac{1}{2}((u \cdot \nabla)v, w) - \frac{1}{2}((u \cdot \nabla)w, v). \quad (3)$$

Below are some properties of b that will be used later in error analysis.

Lemma 2.1. For $u, v, w \in X$, $b(u, v, w)$ satisfies the following bounds:

$$|b(u, v, w)| \leq C_0(\Omega) \|\nabla u\| \|\nabla v\| \|\nabla w\|, \quad (4)$$

$$|b(u, v, w)| \leq C_0(\Omega) \|u\|^{\frac{1}{2}} \|\nabla u\|^{\frac{1}{2}} \|\nabla v\| \|\nabla w\|.$$

Proof: The first inequality is obtained by taking the absolute value of the skew-symmetric trilinear form and making use of Cauchy-Schwarz and Poincaré-Friedrich inequalities and similar arguments work for the second inequality too (Layton, 2008).

The application of the finite element method for spatial discretization requires the construction of a triangulation π_h of the domain Ω with maximum diameter h . Finite element spaces as given follows:

$$X^h \subset X \text{ and } Q^h \subset Q.$$

Lemma 2.2. It is assumed that the appropriate finite element spaces of velocity and pressure spaces satisfy the discrete inf-sup condition, i.e. the Ladyzenskaya-Brezzi-Babuska (LBB) condition. For example, let β be a constant, then

$$\inf_{v \in X^h} \sup_{q \in Q^h} \frac{(q^h, \nabla \cdot v^h)}{\|\nabla v^h\| \|q^h\|} \geq \beta > 0.$$

We introduce the discretely divergence-free subspace $V^h \subset X^h$ given by

$$V^h := \{v^h \in X^h, (\nabla \cdot q^h, v^h) = 0, \forall q^h \in Q^h\}.$$

Under the inf-sup condition, V^h is a nonempty, closed subspace of X^h and the formulation in X^h is equivalent to V^h (John, 2004).

It is also assumed that the velocity and pressure finite element spaces satisfy the following well-known approximation properties (Layton, 2008):

$$\inf_{v \in X_h} (\|u - v^h\| + h\|\nabla(u - v^h)\|) \leq Ch^{k+1}\|u\|_{k+1}, \quad u \in H^{k+1}(\Omega), \tag{4}$$

$$\inf_{q \in Q_h} \|p - q^h\| \leq Ch^{s+1}\|p\|_{s+1}, \quad p \in H^{s+1}(\Omega).$$

for $(v^h, q^h) \in (X^h, Q^h)$.

Discrete Gronwall’s inequality will be used in the error analysis (Girault & Raviart, 1979).

Lemma 2.3. (Discrete Gronwall’s inequality) For $n \geq 0$ and a_n, b_n, c_n, d_n, e_n be non-negative integers. If

$$a_{N+1} + \Delta t \sum_{n=0}^{N+1} b_n \leq \Delta t \sum_{n=0}^{N+1} a_n d_n + \Delta t \sum_{n=0}^{N+1} c_n + H, \quad N \geq 0$$

holds, then

$$a_{N+1} + \Delta t \sum_{n=0}^{N+1} b_n \leq \exp\left(\Delta t \sum_{n=0}^{N+1} \frac{e_n}{1 - ke_n}\right) \left(\Delta t \sum_{n=0}^{N+1} c_n + H\right), \quad N \geq 0$$

for $ke_n < 1$.

The fully discretized finite element approximation of the model is given below.

Algorithm 2.1: Given u_n^h , find $u_{n+1}^h \in X^h$ satisfying:

$$\left(\frac{u_{n+1}^h - u_n^h}{\Delta t}, v^h\right) + \nu(\nabla u_{n+1}^h, \nabla v^h) + b(u_{n+1}^h, u_{n+1}^h, v^h) - (p_{n+1}^h, \nabla \cdot v^h) + \alpha(\nabla u_{n+1}^h, \nabla v^h) = (f_{n+1}, v^h), \tag{5}$$

$$(\nabla \cdot u_{n+1}^h, q^h) = 0,$$

where $\alpha(\nabla u_{n+1}^h, \nabla v^h)$ is the stabilization term that will increase the stability of the solution, representing the artificial viscosity, and α is the artificial viscosity parameter.

3. ANALYSIS OF THE SCHEME

This section focuses on the numerical analysis of Algorithm 2.1. In this context, the stability of the fully discrete approximation scheme is first analyzed. Then, the error analysis of the developed algorithm is performed.

3.1. Stability Analysis

Theorem 3.1: Let $f \in L^2(0, T; H^{-1}(\Omega))$. (5) is unconditionally stable and the solution found satisfies the following equation for any $h > 0$ and $u^h, v^h \in X^h; p^h, q^h \in Q^h$:

$$\|u_N^h\|^2 + \sum_{n=0}^{N-1} \Delta t \left(\frac{\nu}{2} + \alpha\right) \|\nabla u_{n+1}^h\|^2 \leq \|u_0^h\|^2 + \sum_{n=0}^{N-1} \frac{\Delta t}{\nu} \|f_{n+1}\|_{-1}^2.$$

Proof: Given the schemes (5) on X^h , the equivalent formulation on V^h is

$$\left(\frac{u_{n+1}^h - u_n^h}{\Delta t}, v^h\right) + \nu(\nabla u_{n+1}^h, \nabla v^h) + b(u_{n+1}^h, u_{n+1}^h, v^h) + \alpha(\nabla u_{n+1}^h, \nabla v^h) = (f_{n+1}, v^h). \tag{7}$$

If $v^h = u_{n+1}^h$ is taken in (7), since $b(u_{n+1}^h, u_{n+1}^h, u_{n+1}^h) = 0$ from (4), the equation becomes as follows:

$$\left(\frac{u_{n+1}^h - u_n^h}{\Delta t}, u_{n+1}^h\right) + \nu(\nabla u_{n+1}^h, \nabla u_{n+1}^h) + \alpha(\nabla u_{n+1}^h, \nabla u_{n+1}^h) = (f_{n+1}, u_{n+1}^h).$$

Here, we first multiply both sides of the equation by Δt and then rearrange the terms,

$$\|u_{n+1}^h\|^2 + \Delta t(\nu + \alpha)\|\nabla u_{n+1}^h\|^2 = (u_n^h, u_{n+1}^h) + \Delta t(f_{n+1}, u_{n+1}^h) \tag{8}$$

The terms on the right-hand side of the equation are analyzed separately. Using the Cauchy-Schwarz and Young’s inequalities for the first term on the right, the following bound is obtained:

$$(u_n^h, u_{n+1}^h) \leq \|u_n^h\| \|u_{n+1}^h\| \leq \frac{1}{2}\|u_n^h\|^2 + \frac{1}{2}\|u_{n+1}^h\|^2.$$

For the second term, the definition of dual norm is applied first, followed by Cauchy-Schwarz and Young inequalities respectively:

$$\begin{aligned} \Delta t(f_{n+1}, u_{n+1}^h) &= \frac{\Delta t(f_{n+1}, u_{n+1}^h)}{\|\nabla u_{n+1}^h\|} \|\nabla u_{n+1}^h\| \leq \Delta t \|f_{n+1}\|_{-1} \|\nabla u_{n+1}^h\| \\ &\leq \frac{\Delta t \nu}{2} \|\nabla u_{n+1}^h\|^2 + \frac{1}{2\nu} \Delta t \|f_{n+1}\|_{-1}^2. \end{aligned}$$

Substituting the calculated terms in equation (8) yields the following inequality

$$\frac{1}{2}\|u_{n+1}^h\|^2 + \Delta t\left(\frac{\nu}{2} + \alpha\right)\|\nabla u_{n+1}^h\|^2 \leq \frac{1}{2}\|u_n^h\|^2 + \frac{\Delta t}{2\nu}\|f_{n+1}\|_{-1}^2.$$

If we take the sum for $n = 0, 1, \dots, N - 1$, we get

$$\|u_N^h\|^2 + \sum_{n=0}^{N-1} \Delta t\left(\frac{\nu}{2} + \alpha\right)\|\nabla u_{n+1}^h\|^2 \leq \|u_0^h\|^2 + \sum_{n=0}^{N-1} \frac{\Delta t}{\nu}\|f_{n+1}\|_{-1}^2$$

Then the inequality in Theorem 3.1 is obtained.

3.2. Error Analysis

In addition to the discrete inf-sup condition, the approximations in Lemma (2.2) are assumed to be valid for the selection of velocity-pressure spaces. First, we state the regularity assumptions required for NSE solutions.

$$u \in L^4(0, T; H^{k+1}(\Omega)) \cap L^4(0, T; L^2(\Omega)) \cap L^\infty(0, T; L^2(\Omega)),$$

$$p \in L^2(0, T; H^{s+1}(\Omega)),$$

$$u_t \in L^\infty(0, T; L^2(\Omega)).$$

Theorem 3.2: Let (u, p) be a solution of the NSE (1). Then the error of (5) defined as $e_{n+1} = u_{n+1} - u_{n+1}^h$ satisfies the following inequality:

$$\begin{aligned} & \|e_N^h\|^2 + \Delta t(v + \alpha) \sum_{n=0}^{N-1} \|\nabla e_{n+1}^h\|^2 \\ & \leq \exp(Cv^{-3}\Delta t\|\nabla u_{n+1}\|^4) \left[Cv^{-1}\Delta t \sum_{n=0}^{N-1} \left((\Delta t)^{-1}h^{2k+2} \int_{t_n}^{t_{n+1}} \|u_t\|_{k+1}^2 dt + h^{2k}\|\nabla u_{n+1}\|^2 \|u\|_{k+1}^2 + \right. \right. \\ & h^{2k}\|\nabla u_{n+1}^h\|^2 \|u\|_{k+1}^2 + h^{2k}\|p_{n+1}\|_k^2 + (\Delta t)^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt (\|\nabla u_{n+1}\|^2 + \\ & \|\nabla u\|^2) + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|p_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|f_{tt}\|^2 dt \left. \right) + Cv\Delta th^{2k}\|u\|_{k+1}^2 + Cv(\Delta t)^4\|\nabla u_{tt}\|^2 + \\ & \alpha\Delta t(h^{2k}\|u\|_{k+1}^2 + \|\nabla u_{n+1}\|^2) \left. \right]. \end{aligned}$$

Proof: The difference between the true solution and the approximate solution is estimated to obtain the minimum error. The equation

$$\begin{aligned} & \left(\frac{u_{n+1} - u_n}{\Delta t}, v^h \right) + v(\nabla u_{n+1}, \nabla v^h) + b(u_{n+1}, u_{n+1}, v^h) - (p_{n+1}, \nabla \cdot v^h) + (\nabla \cdot u_{n+1}, q^h) + \alpha(\nabla u_{n+1}, \nabla v^h) \\ & - \alpha(\nabla u_{n+1}, \nabla v^h) = (f_{n+1}, v^h) \end{aligned} \quad (9)$$

is obtained by writing $t \rightarrow t_{n+1}$ in the true solution of (1) and adding and subtracting the $\alpha(\nabla u_{n+1}, \nabla v^h)$ term. Subtract (5) from (9),

$$\begin{aligned} & \left(\frac{e_{n+1} - e_n}{\Delta t}, v^h \right) + v(\nabla e_{n+1}, \nabla v^h) + b(u_{n+1}, u_{n+1}, v^h) - b(u_{n+1}^h, u_{n+1}^h, v^h) - (p_{n+1} - q^h, \nabla \cdot v^h) \\ & + \alpha(\nabla e_{n+1}, \nabla v^h) = \tau(u_n, v^h) \end{aligned} \quad (10)$$

where $e_{n+1} = u_{n+1} - u_{n+1}^h$. The nonlinear terms on the right side of (10) can be arranged as follows:

$$\begin{aligned} & b(u_{n+1}, u_{n+1}, v^h) - b(u_{n+1}^h, u_{n+1}^h, v^h) \\ & = b(u_{n+1}, u_{n+1}, v^h) - b(u_{n+1}^h, u_{n+1}, v^h) + b(u_{n+1}^h, u_{n+1}, v^h) - b(u_{n+1}^h, u_{n+1}^h, v^h) \\ & = b(e_{n+1}, u_{n+1}, v^h) - b(u_{n+1}^h, e_{n+1}, v^h). \end{aligned}$$

The error can be decomposed as

$$e_{n+1} = u_{n+1} - \tilde{u}_{n+1} + \tilde{u}_{n+1} - u_{n+1}^h = \mu_{n+1} - \phi_{n+1}^h,$$

where \tilde{u}_{n+1} is the arbitrary interpolant. Choosing the test function $v^h = \phi_{n+1}^h$ yields

$$\begin{aligned} & \left(\frac{\phi_{n+1}^h - \phi_n^h}{\Delta t}, \phi_{n+1}^h \right) + v(\nabla \phi_{n+1}^h, \nabla \phi_{n+1}^h) + b(\phi_{n+1}^h, u_{n+1}, \phi_{n+1}^h) + (p_{n+1} - q^h, \nabla \cdot \phi_{n+1}^h) \\ & + \alpha(\nabla \phi_{n+1}^h, \nabla \phi_{n+1}^h) \\ & = \left(\frac{\mu_{n+1} - \mu_n}{\Delta t}, \phi_{n+1}^h \right) + v(\nabla \mu_{n+1}, \nabla \phi_{n+1}^h) + b(\mu_{n+1}, u_{n+1}, \phi_{n+1}^h) + b(u_{n+1}^h, \mu_{n+1}, \phi_{n+1}^h) \\ & + \alpha(\nabla \mu_{n+1}, \nabla \phi_{n+1}^h) + \tau(u_n, \nabla \phi_{n+1}^h). \end{aligned}$$

Here,

$$\begin{aligned} \tau(u_n, \nabla \phi_{n+1}^h) &= \left(\frac{u_{n+1} - u_n}{\Delta t} - u_t, \phi_{n+1}^h \right) - (p_{n+1} - p, \nabla \cdot \phi_{n+1}^h) + (f - f_{n+1}, \phi_{n+1}^h) \\ &\quad + \nu(\nabla(u_{n+1} - u_n), \nabla \phi_{n+1}^h) + b(u_{n+1} - u_n, u_{n+1}, \phi_{n+1}^h) + b(u, u_{n+1} - u, \phi_{n+1}^h) \\ &\quad + \alpha(\nabla u_{n+1}, \nabla \phi_{n+1}^h). \end{aligned}$$

Using the equation $a^2 - ab = \frac{a^2 + (a-b)^2 - b^2}{2}$, the first term is organized as follows:

$$(\phi_{n+1}^h - \phi_n^h, \phi_{n+1}^h) = \frac{1}{2} (\|\phi_{n+1}^h\|^2 + \|\phi_{n+1}^h - \phi_n^h\|^2 - \|\phi_n^h\|^2).$$

Accordingly,

$$\begin{aligned} &\frac{1}{2\Delta t} (\|\phi_{n+1}^h\|^2 + \|\phi_{n+1}^h - \phi_n^h\|^2 - \|\phi_n^h\|^2) + \nu \|\nabla \phi_{n+1}^h\|^2 + \alpha \|\nabla \phi_{n+1}^h\|^2 \\ &= \left(\frac{\mu_{n+1} - \mu_n}{\Delta t}, \phi_{n+1}^h \right) + \nu(\nabla \mu_{n+1}, \nabla \phi_{n+1}^h) + b(\mu_{n+1}, u_{n+1}, \phi_{n+1}^h) + b(u_{n+1}^h, \mu_{n+1}, \phi_{n+1}^h) \\ &\quad - b(\phi_{n+1}^h, u_{n+1}, \phi_{n+1}^h) - (p_{n+1} - q^h, \nabla \cdot \phi_{n+1}^h) + \alpha(\nabla \mu_{n+1}, \nabla \phi_{n+1}^h) + \tau(u_n, \nabla \phi_{n+1}^h) \\ &= T_1 + T_2 + \dots + T_8. \end{aligned} \tag{11}$$

Here, the terms obtained on the right side of the equation are named T_1, T_2, \dots, T_8 respectively. Each term will be analyzed separately, and necessary adjustments will be made.

T_1 is bounded by Cauchy-Schwarz, and Young's inequalities:

$$\left(\frac{\mu_{n+1} - \mu_n}{\Delta t}, \phi_{n+1}^h \right) \leq \frac{1}{\Delta t} \left\| \int_{t_n}^{t_{n+1}} \mu_t dt \right\| \|\phi_{n+1}^h\| \leq \frac{C\nu^{-1}}{\Delta t} \int_{t_n}^{t_{n+1}} \|\mu_t\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2.$$

T_2 is bounded by Young's inequalities:

$$\nu(\nabla \mu_{n+1}, \nabla \phi_{n+1}^h) \leq C\nu \|\nabla \mu_{n+1}\|^2 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2.$$

For T_3, T_4 , and T_5 first apply Lemma 2.2, then Young's inequality to get:

$$b(\mu_{n+1}, u_{n+1}, \phi_{n+1}^h) \leq C\nu^{-1} \|\nabla u_{n+1}\|^2 \|\nabla \mu_{n+1}\|^2 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$b(u_{n+1}^h, \mu_{n+1}, \phi_{n+1}^h) \leq C\nu^{-1} \|\nabla u_{n+1}^h\|^2 \|\nabla \mu_{n+1}\|^2 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$b(\phi_{n+1}^h, u_{n+1}, \phi_{n+1}^h) \leq C\nu^{-3} \|\phi_{n+1}^h\|^2 \|\nabla u_{n+1}\|^4 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2.$$

Young's inequality is applied for T_6, T_7 , and T_8

$$(p_{n+1} - q^h, \nabla \cdot \phi_{n+1}^h) \leq C\nu^{-1} \|p_{n+1} - q^h\|^2 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$\alpha(\nabla \mu_{n+1}, \nabla \phi_{n+1}^h) \leq C\alpha \|\nabla \mu_{n+1}\|^2 + \frac{\alpha}{4} \|\nabla \phi_{n+1}^h\|^2.$$

The terms given in equation $T_8 = \tau(u_n, \nabla \phi_{n+1}^h)$ are named S_1, S_2, \dots, S_7 and analyzed separately.

$$\begin{aligned} \tau(u_n, \nabla \phi_{n+1}^h) &= \left(\frac{u_{n+1} - u_n}{\Delta t} - u_t, \phi_{n+1}^h \right) + v(\nabla(u_{n+1} - u), \nabla \phi_{n+1}^h) + b(u_{n+1} - u_n, u_{n+1}, \phi_{n+1}^h) \\ &\quad + b(u, u_{n+1} - u, \phi_{n+1}^h) - (p_{n+1} - p, \nabla \cdot \phi_{n+1}^h) + (f - f_{n+1}, \phi_{n+1}^h) \\ &= S_1 + S_2 + \dots + S_7. \end{aligned}$$

Taylor Series expansion and Young's inequality are applied for these terms:

$$\left(\frac{u_{n+1} - u_n}{\Delta t} - u_t, \phi_{n+1}^h \right) \leq C\nu^{-1}(\Delta t)^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$v(\nabla(u_{n+1} - u), \nabla \phi_{n+1}^h) \leq C\nu \|\nabla(u_{n+1} - u)\|^2 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2 \leq C\nu(\Delta t)^3 \int_{t_n}^{t_{n+1}} \|\nabla u_{tt}\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$b(u_{n+1} - u_n, u_{n+1}, \phi_{n+1}^h) \leq C\nu^{-1}(\Delta t)^3 \|\nabla u_{n+1}\|^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$b(u, u_{n+1} - u, \phi_{n+1}^h) \leq C\nu^{-1}(\Delta t)^3 \|\nabla u\|^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$(p_{n+1} - p, \nabla \cdot \phi_{n+1}^h) \leq C\nu^{-1}(\Delta t)^3 \int_{t_n}^{t_{n+1}} \|p_{tt}\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$(f - f_{n+1}, \phi_{n+1}^h) \leq C\nu^{-1} \|f - f_{n+1}\|^2 + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2 \leq C\nu^{-1}(\Delta t)^3 \int_{t_n}^{t_{n+1}} \|f_{tt}\|^2 dt + \frac{\nu}{24} \|\nabla \phi_{n+1}^h\|^2,$$

$$\alpha(\nabla u_{n+1}, \nabla \phi_{n+1}^h) \leq C\alpha \|\nabla u_{n+1}\|^2 + \frac{\alpha}{4} \|\nabla \phi_{n+1}^h\|^2.$$

Substituting all the terms obtained in (11) and subtracting the non-negative term on the left side of the equation gives

$$\begin{aligned} &\frac{1}{2\Delta t} \|\phi_{n+1}^h\|^2 - \frac{1}{2\Delta t} \|\phi_n^h\|^2 + \frac{\nu}{2} \|\nabla \phi_{n+1}^h\|^2 + \frac{\alpha}{2} \|\nabla \phi_{n+1}^h\|^2 \\ &\leq C\nu^{-1} \left[(\Delta t)^{-1} \int_{t_n}^{t_{n+1}} \|\mu_t\|^2 dt + \|\nabla u_{n+1}\|^2 \|\nabla \mu_{n+1}\|^2 + \|\nabla u_{n+1}^h\|^2 \|\nabla \mu_{n+1}\|^2 \right. \\ &\quad + \|\mu_{n+1} - q^h\|^2 + (\Delta t)^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt (\|\nabla u_{n+1}\|^2 + \|\nabla u\|^2) \\ &\quad \left. + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|p_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|f_{tt}\|^2 dt \right] + C\nu \|\nabla \mu_{n+1}\|^2 + C\nu^{-3} \|\phi_{n+1}^h\|^2 \|\nabla u_{n+1}\|^4 \\ &\quad + C\alpha (\|\nabla \mu_{n+1}\|^2 + \|\nabla u_{n+1}\|^2) + C\nu(\Delta t)^3 \int_{t_n}^{t_{n+1}} \|\nabla u_{tt}\|^2 dt \end{aligned}$$

Multiplying inequality by $2\Delta t$ and applying Lemma 2.2 yields the following:

$$\begin{aligned} & \|\phi_{n+1}^h\|^2 - \|\phi_n^h\|^2 + \Delta t(\nu + \alpha) \|\nabla \phi_{n+1}^h\|^2 \\ & \leq C\nu^{-1}\Delta t \left[(\Delta t)^{-1}h^{2k+2} \int_{t_n}^{t_{n+1}} \|u_t\|_{k+1}^2 dt + h^{2k} \|\nabla u_{n+1}\|^2 \|u\|_{k+1}^2 + h^{2k} \|\nabla u_{n+1}^h\|^2 \|u\|_{k+1}^2 \right. \\ & \quad + h^{2k} \|p_{n+1}\|_k^2 + (\Delta t)^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt (\|\nabla u_{n+1}\|^2 + \|\nabla u\|^2) \\ & \quad \left. + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|p_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|f_{tt}\|^2 dt \right] + C\nu\Delta t h^{2k} \|u\|_{k+1}^2 \\ & \quad + C\nu^{-3}\Delta t \|\phi_{n+1}^h\|^2 \|\nabla u_{n+1}\|^4 + \alpha\Delta t (h^{2k} \|u\|_{k+1}^2 + \|\nabla u_{n+1}\|^2) \\ & \quad + C\nu(\Delta t^3)\Delta t \int_{t_n}^{t_{n+1}} \|\nabla u_{tt}\|^2 dt \end{aligned}$$

If the sum is taken for $n = 0, 1, \dots, N - 1$, we obtain

$$\begin{aligned} & \|\phi_N^h\|^2 + \Delta t(\nu + \alpha) \sum_{n=0}^{N-1} \|\nabla \phi_{n+1}^h\|^2 \\ & \leq \left[C\nu^{-1}\Delta t \sum_{n=0}^{N-1} \left((\Delta t)^{-1}h^{2k+2} \int_{t_n}^{t_{n+1}} \|u_t\|_{k+1}^2 dt + h^{2k} \|\nabla u_{n+1}\|^2 \|u\|_{k+1}^2 \right. \right. \\ & \quad + h^{2k} \|\nabla u_{n+1}^h\|^2 \|u\|_{k+1}^2 + h^{2k} \|p_{n+1}\|_k^2 + (\Delta t)^2 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt \\ & \quad \left. + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|u_{tt}\|^2 dt (\|\nabla u_{n+1}\|^2 + \|\nabla u\|^2) + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|p_{tt}\|^2 dt + (\Delta t)^3 \int_{t_n}^{t_{n+1}} \|f_{tt}\|^2 dt \right) \\ & \quad + C\nu\Delta t h^{2k} \|u\|_{k+1}^2 + C\nu(\Delta t)^4 \|\nabla u_{tt}\|^2 \\ & \quad \left. + \alpha\Delta t (h^{2k} \|u\|_{k+1}^2 + \|\nabla u_{n+1}\|^2) \right] \exp(C\nu^{-3}\Delta t \|\nabla u_{n+1}\|^4) \end{aligned}$$

Finally, Theorem 3.2 is obtained by applying the triangle inequality $\|e_{n+1}\| = \|\mu_{n+1} - \phi_{n+1}^h\| \leq \|\mu_{n+1}\| + \|\phi_{n+1}^h\|$ to the error terms using approximation properties.

Result 3.1: The error approximation obtained in Theorem 3.2 is an optimal error approximation. For example, if $k = 2$, the order of the error according to the polynomial choices (P^2, P^1) is 2. It is also clear from the statement of the theorem that $u \rightarrow u_h$ when $h, \Delta t, \alpha \rightarrow 0$.

4. NUMERICAL SIMULATION

In this section, some simulations are presented to verify the theoretical results. Firstly, a quantitative numerical test is carried out to reveal the convergence behavior of the scheme in (5) and to verify the result obtained in Theorem 1. In qualitative experiments, blood and body temperature are assumed to be constant. For the understanding of blood flow, blood is assumed to be a Newtonian fluid. Therefore, blood density $\rho=1060$ kg/m³ and blood viscosity $\nu=0.0035$ mPas were chosen (Kleinstreuer, 2016). A no-slip boundary condition is applied to the artery walls. Also, to determine the blood flow velocity, $c = \nu Re/\rho l$ is taken as the inflow

velocity, where l is the diameter of the lumen. The value of c is important to estimate the velocity at a specific point of blood flow or to analyze the overall performance of blood vessels. In this paper, all simulations were performed using FreeFem++ (Hecht., 2012), a publicly licensed finite element software package.

4.1. Convergence Study

First, the order of the spatial error is analyzed. Since the Taylor-Hood finite element pair mentioned in Result 1 is considered here, the expected order of the error is 2. The model problem computational domain is $\Omega = (0,1)^2$. The computational domain is triangulated with various coarse mesh resolutions and the mesh width is taken from $h = 2^{-1}$ to $h = 2^{-6}$. These choices were sufficient to verify the convergence rates. The converged solution was computed in the time interval $[0, 1]$. To minimize the effect of temporal error and to fully reveal the spatial effect on the error, $\Delta t = 0,000625$ was taken. The H_1 norms of the error for $\nu = 1$ and $\alpha = 0.01$ are evaluated. These values were optimally chosen to minimize the error rates in the newly derived model. For this test problem, the analytical velocity and pressure variables are considered as follows:

$$u = \begin{bmatrix} e^t \cos(y) \\ e^t \sin(x) \end{bmatrix}, p = (x - y)(1 + t).$$

These values are substituted in (1) and the function f is obtained. The results of the numerical solution are given in Table 1 for different values of h .

Table 1. Spatial velocity errors and rates of convergence

h	$\ \nabla(u - u^h)\ $	Rate
2^{-1}	6,46611e-4	-
2^{-2}	1,57286e-4	2,03
2^{-3}	3,85477e-5	2,02
2^{-4}	9,56907e-6	2,01
2^{-5}	2,38566e-6	2,00
2^{-6}	5,96731e-7	2,00

As can be seen from this table, the degree of convergence in the spatial sense is 2, which is the expected optimal degree of convergence. Similarly, in order to find the degree of temporal error, a fixed mesh size of $h = 2^{-6}$ was taken in the same region using the same true solution functions and the right-hand side. Δt values are halved iteratively in time interval $[0, 1]$ to obtain a rate. The results are presented in Table 2.

Table 2. Temporal velocity errors and rates of convergence

Δt	$\ \nabla(u - u^h)\ $	Rate
2^{-1}	1,98947e-1	-
2^{-2}	1,70164e-2	3,55
2^{-3}	2,91767e-3	2,54
2^{-4}	2,17324e-4	1,98

The effect of the α stabilization term on the error rates calculated at different Re values for $h = 2^{-3}$ and $\Delta t = 0,125$ will be compared in Table 3.

Table 3. Velocity errors for different α values

Re	$\ \nabla(u - u^h)\ $ for $\alpha = 0$	$\ \nabla(u - u^h)\ $ for $\alpha = 0,01$
100	0,0903558	0,0797131
200	0,140067	0,11151
300	0,186865	0,134197
400	0,233123	0,151754
500	0,27908	0,165875

4.2. Study of Two-Necked Vessels

As a result of plaque formation in the vessel, knuckles, i.e. hardness, may occur on the upper and lower surfaces of the vessel in a reciprocal or non-reciprocal manner. This situation negatively affects the blood flow. In the numerical experiments, we tried to obtain the closest and most accurate solutions to the real blood flow with artificial viscosity stabilization. For this purpose, different Reynolds numbers were examined and the α 's that would give the most accurate result were tuned.

As can be seen from Figure 1, artificial viscosity stabilization has a significant impact on the solution when analyzing the velocity of blood flow.

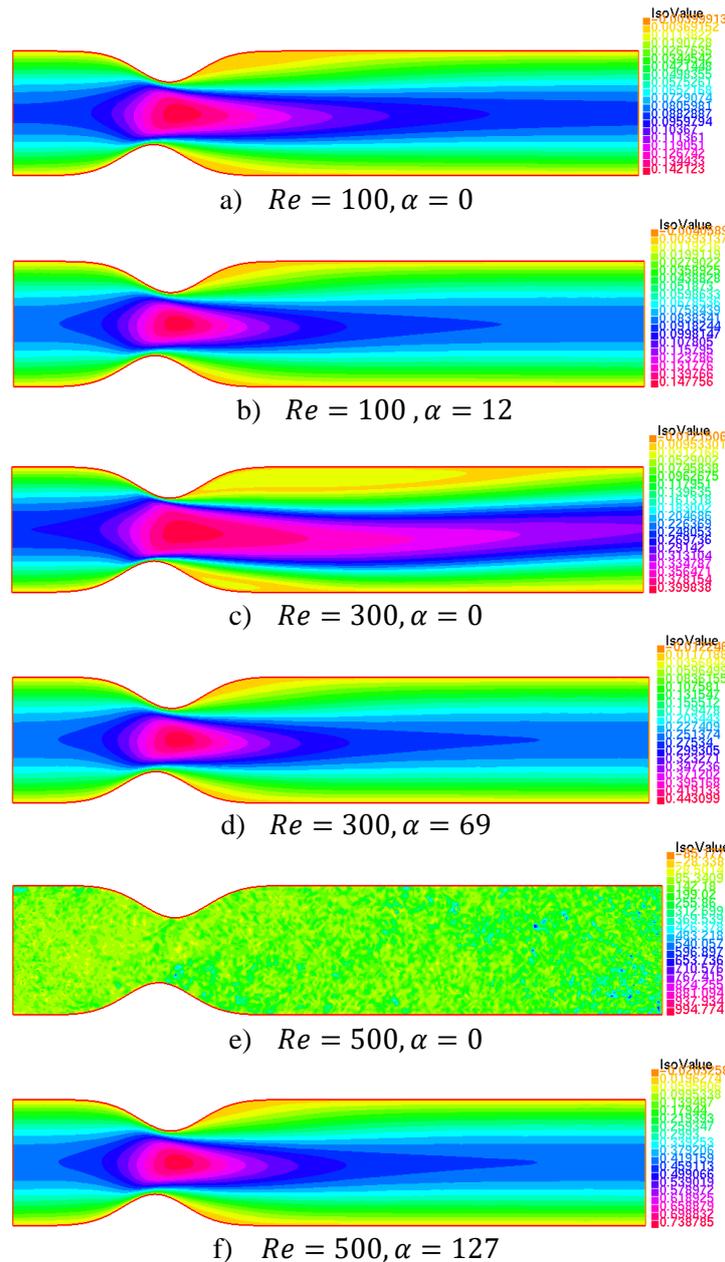


Figure 1. Comparison of the effect of artificial viscosity stabilization for different values of the Reynolds number on the blood flow velocity in a two-necked vessel

Figure 2 is important for understanding the effects of artificial viscosity stabilization on blood pressure. Determining how the pressure profile changes when blood flow has different Re and how artificial viscosity affects this profile can help in the development of better designed and effective medical devices and treatment

methods. Furthermore, this analysis can provide the necessary information to better model the behavior of blood flow in different conditions.

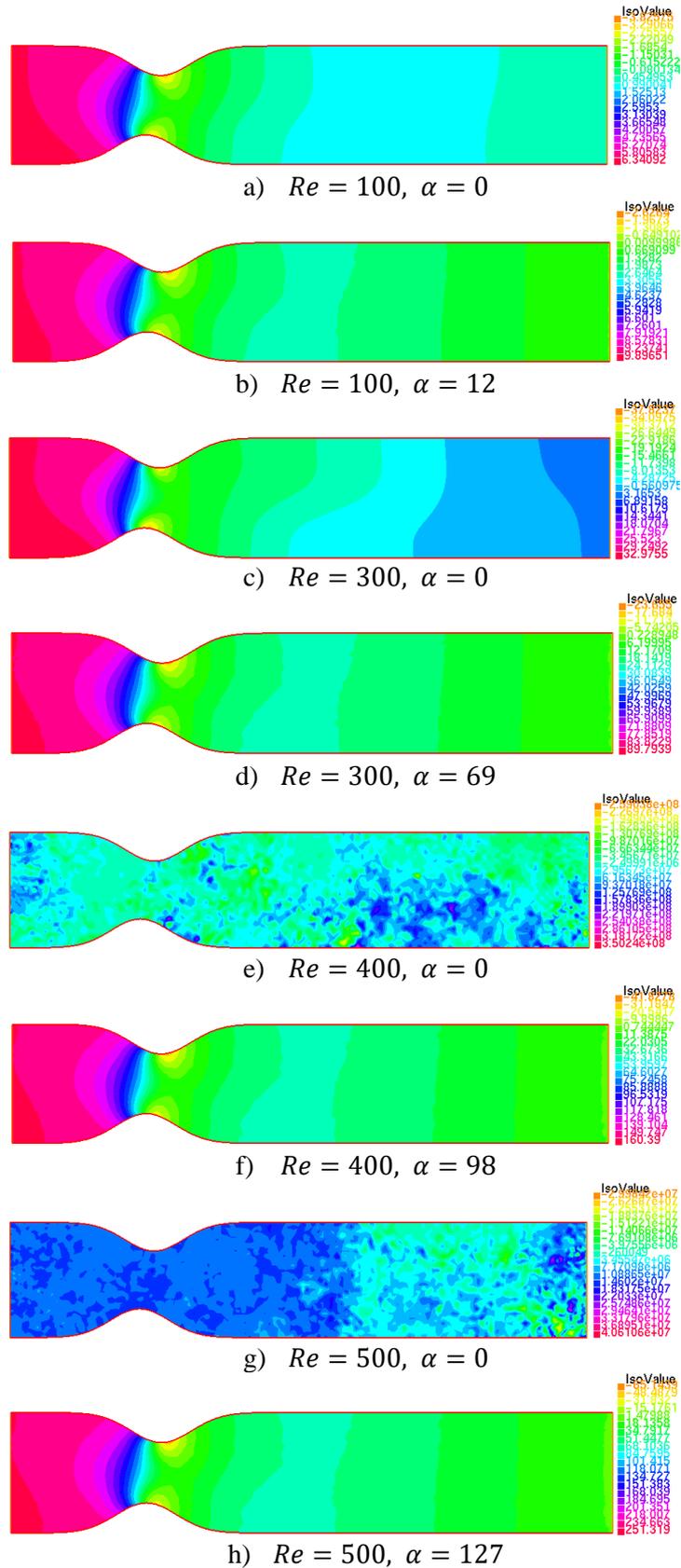


Figure 2. Comparison of the effect of artificial viscosity stabilization for different values of Re on blood pressure in a two-necked vessel

Figure 3 shows the errors at $Re = 500$ for different values of α to illustrate the effect of the stabilization.

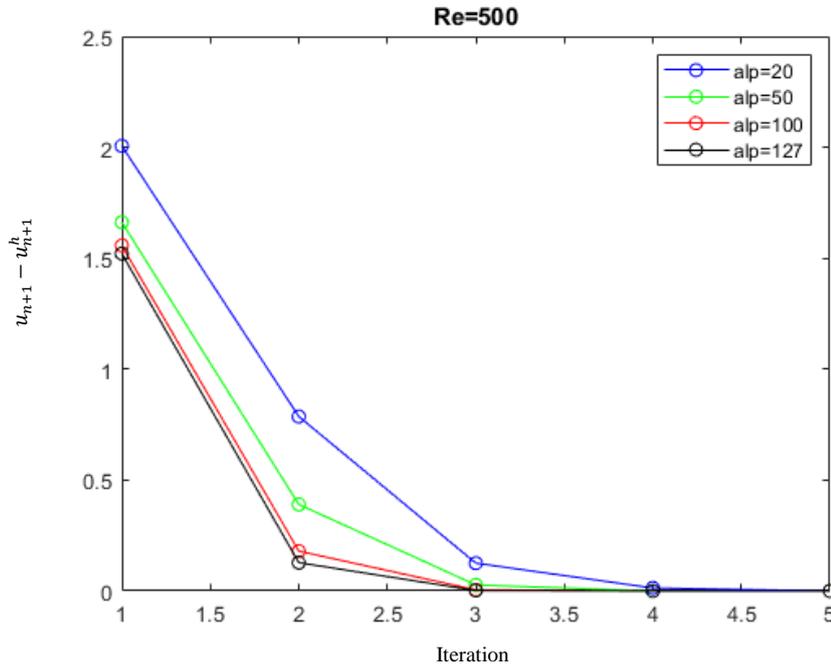


Figure 3. Errors for different α values for $Re=500$

For a given number Re , iterations are obtained with different α values. Considering the error rates resulting from these iterations, the α value that will give the minimum error is selected. As can be seen from the figure, for $Re = 500$, error rates were calculated considering different α values and it was seen that $\alpha = 127$ gave the minimum error.

4.3. Wall Shear Stress in Brain Aneurysm

Figure 4 shows the section focusing on WSS in a brain aneurysm. This figure shows simulation results using FreeFem++ software to model blood flow and its effects on the aneurysm.

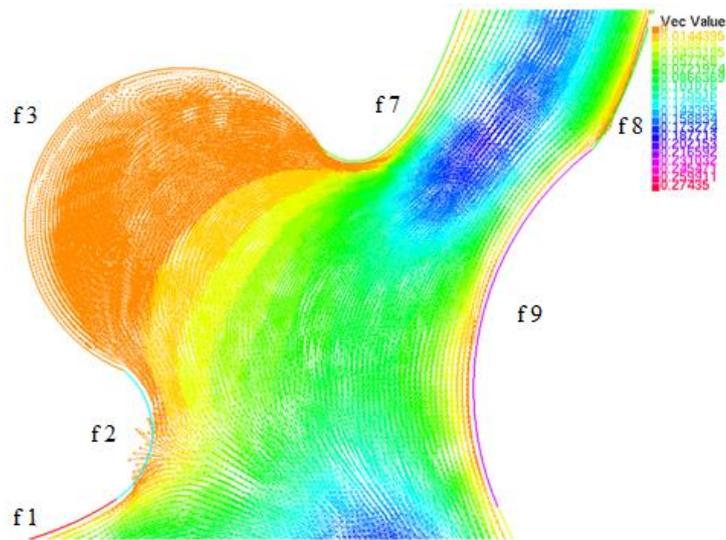


Figure 4. FreeFem++ simulation of a brain aneurysm model

Figures (5-10) of the regions with and without stabilization for $Re = 200$ and $\alpha = 205$ values are given below.

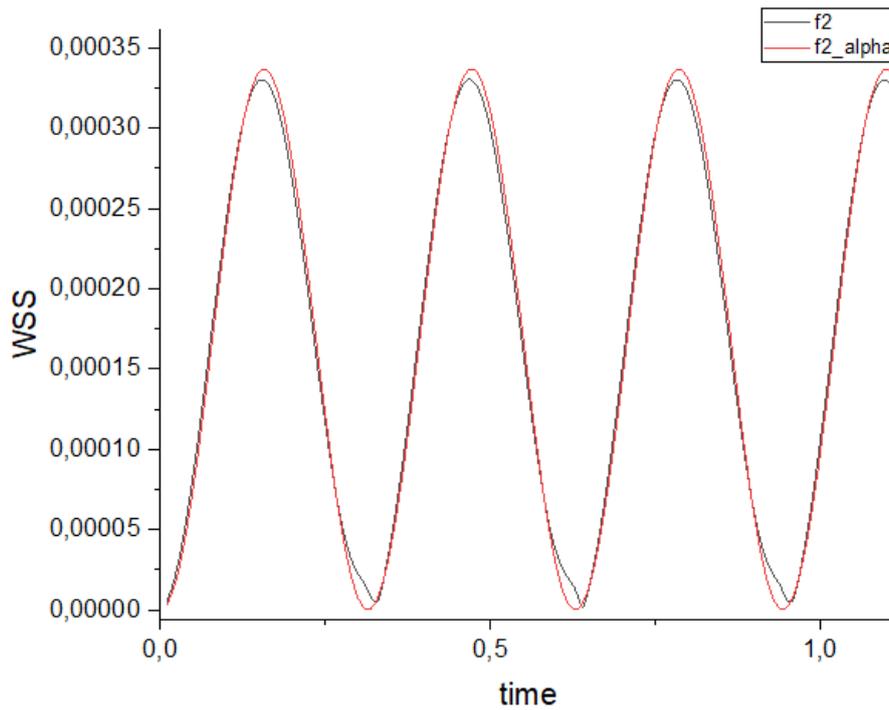


Figure 5. WSS values for the f_2 region

The maximum WSS value in this region was approximately 0,000355 Pa, close to $\alpha = 0$. The flow in the vessels of the brain is usually pulsatile. This is caused by the expansion and contraction of the arteries with the beating of the heart. This is due to changes in the direction of blood flow in the inner walls of the arteries.

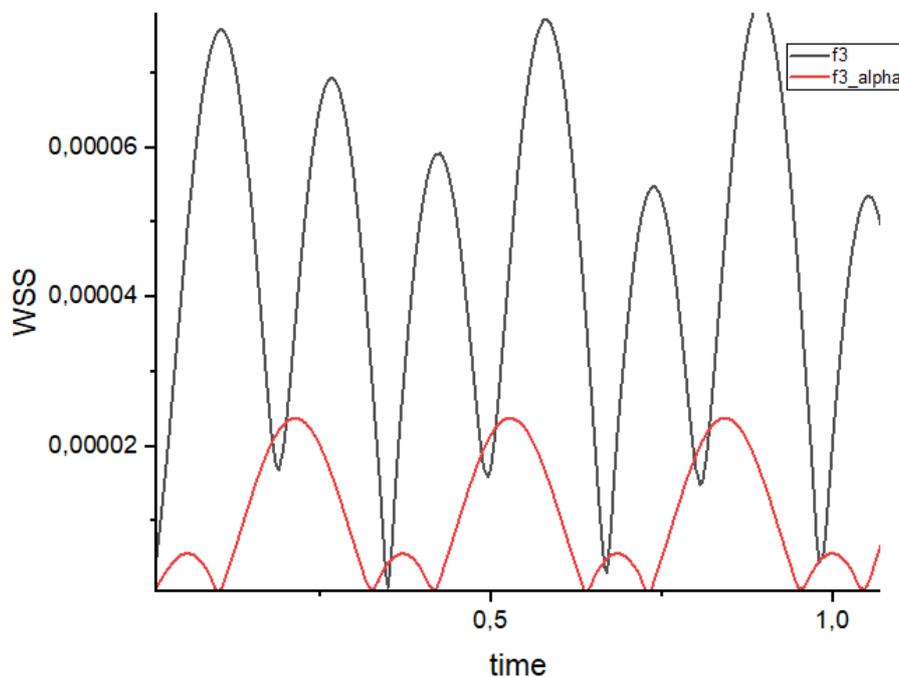


Figure 6. WSS values for the f_3 region

As seen in the graph for the f_3 region, it is clear that when there is no stabilization, results far from the general structure of the flow is obtained. The reason why the WSS value in this region is lower than f_2 and f_7 is due to the slowdown in the flow in these regions, as can be seen from the model (Figure 4).

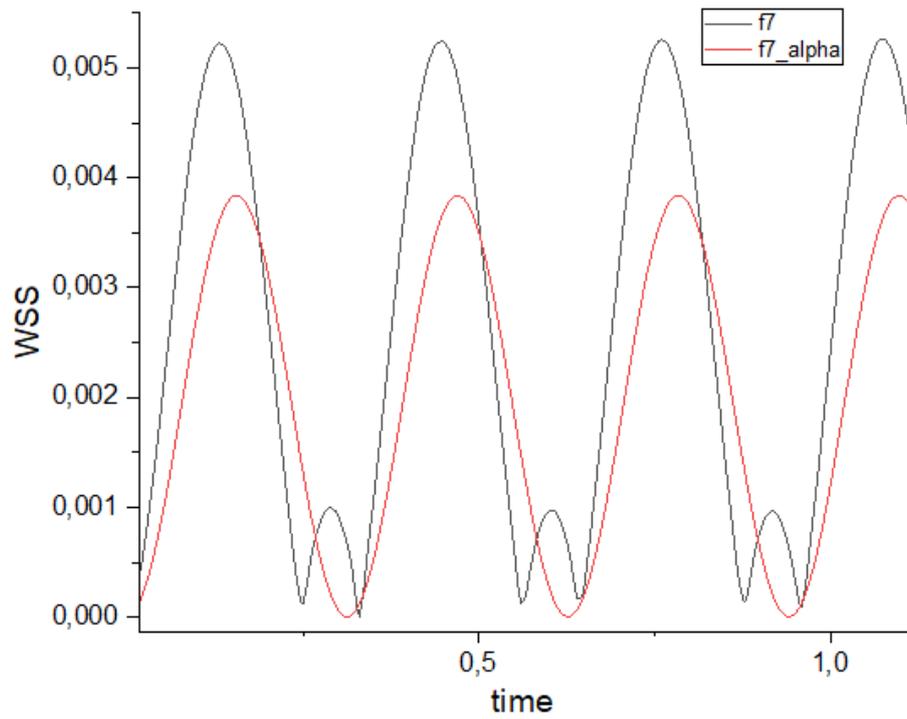


Figure 7. WSS values for the f7 region

The region where the WSS value is the highest, that is, the region where the stress is the highest, is the f7 region. This is because the flow is concentrated in this region and it contains the choke point.

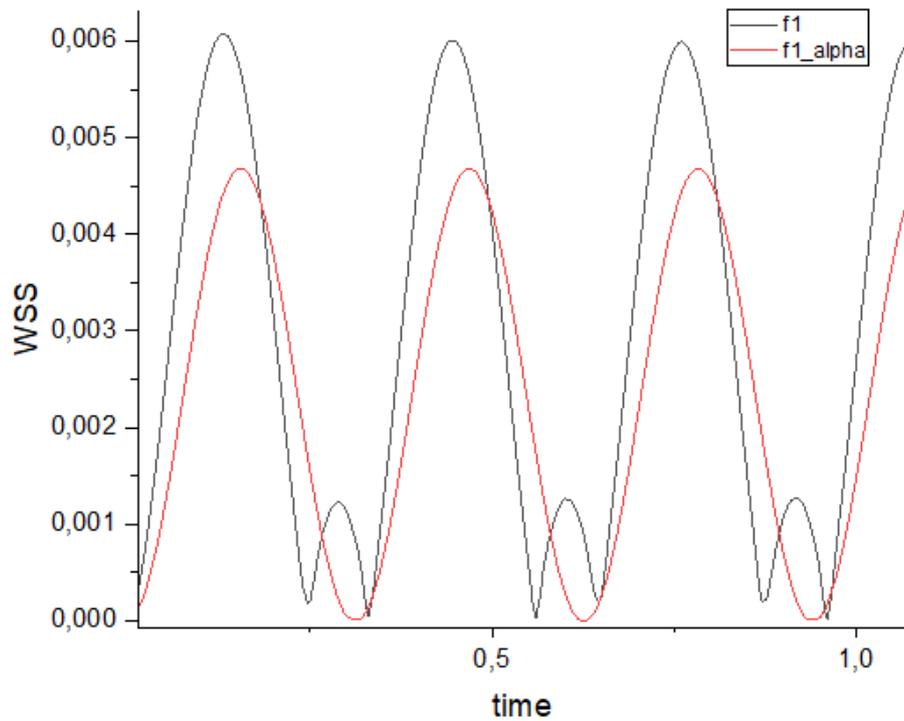


Figure 8. WSS values for the f1 region

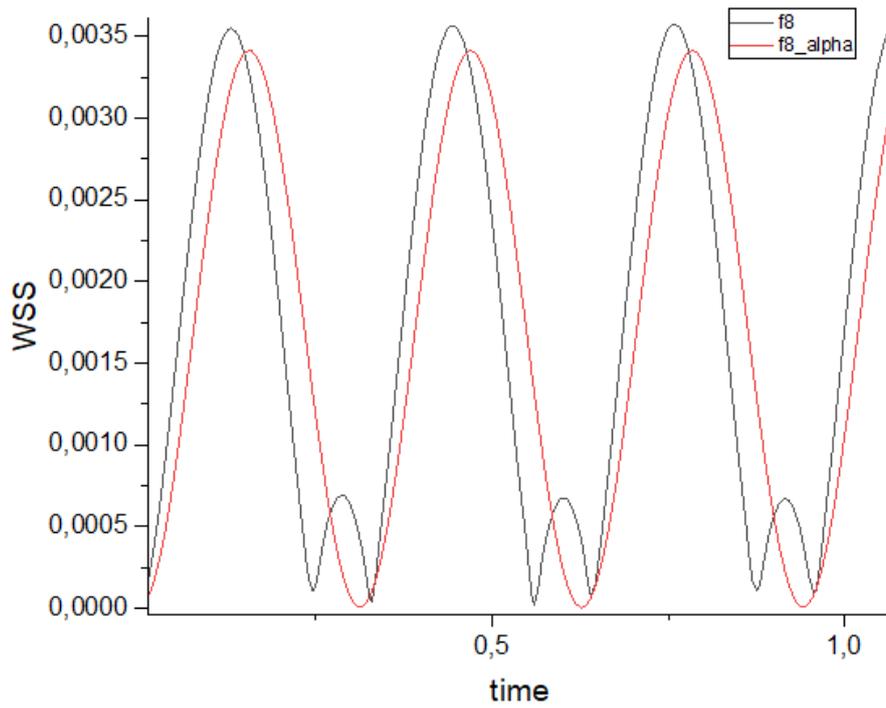


Figure 9. WSS values for the $f8$ region

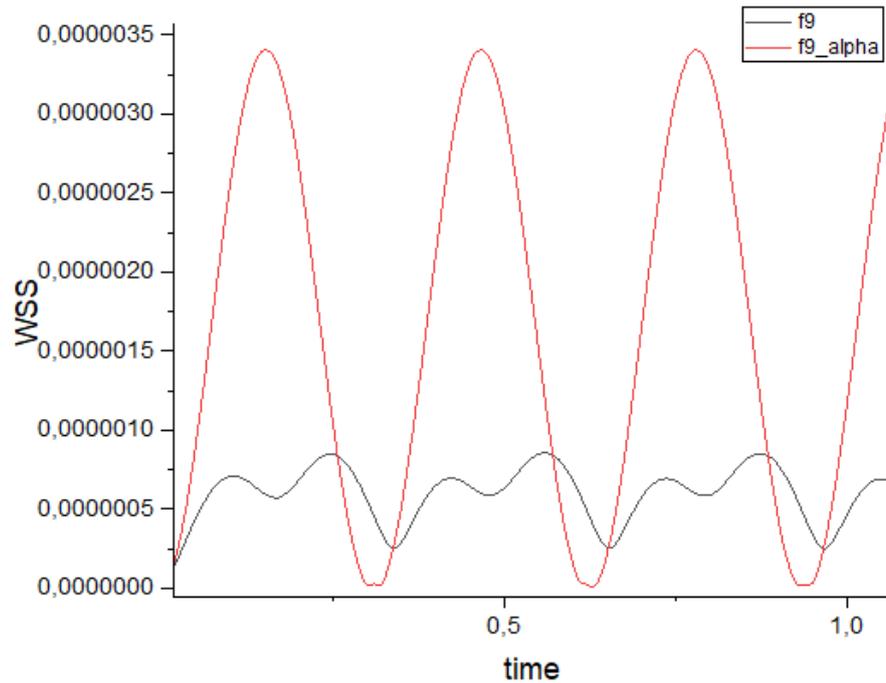


Figure 10. WSS values for the $f9$ region

When we examine the graphs of $f1$, $f8$ and $f9$ regions, it is in the $f9$ region that we will see the effect of stabilization the most. This is because the flow is in a more complex situation. As can be seen from $f9$ graph, when α is higher, the WSS values are more regular in accordance with the blood flow.

5. CONCLUSION

In this paper, changes in vascular flow due to cardiovascular diseases are analyzed with the NSE model. Stability and error analysis of the fully discrete scheme are performed. A stabilization term $\alpha(\nabla u, \nabla v)$

representing the artificial viscosity is used to stabilize the fully discrete finite element scheme. Numerical convergence tests are presented to verify the theoretical convergence rates, demonstrating the effectiveness of the proposed scheme. In addition, the effects of cardiovascular diseases on blood flow velocity, blood pressure and wall shear stress in vessels are modeled. Throughout the obtained models, it is shown how effective the scheme is and thanks to the stabilization, accurate, understandable and realistic results are obtained.

The results of this study provide a better understanding of the changes in blood flow in vessels caused by cardiovascular diseases. The analysis using the NSE model reveals the complexity of the dynamics in vessel flow and shows that these dynamics can be reliably modeled through stability and error analysis of the fully discrete scheme. The stabilization term representing the artificial viscosity improves the accuracy and stability of the model, making the results realistic and reliable. The verification of the theoretical convergence rates by numerical convergence tests reinforces the effectiveness of the proposed scheme and demonstrates the usability of this approach in practical applications. With this model, the effects of cardiovascular diseases on blood flow velocity, blood pressure and wall shear stress can be studied in detail and the effects of these parameters on disease progression can be better understood.

The results provide important information for planning medical interventions and disease management. The accuracy of the model and the reliable results provided by stabilization can contribute to the development of decision support systems in clinical applications. This could enable more efficient management of patients' treatment processes and early detection of diseases. Therefore, this study can be considered as an important step towards a better understanding and management of cardiovascular diseases.

AUTHOR CONTRIBUTIONS

Conceptualization, A.Ç.; methodology, A.Ç.; fieldwork, H. K.; software, A. Ç. and G.H.; title, G. H.; validation, A. Ç.; formal analysis, A. Ç. and G.H.; research, H. K.; sources, H. K.; manuscript-original draft, G.H.; manuscript-review and editing, A.Ç.; visualization, G.H.; supervision, A.Ç.

All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Adams, R. A. (1975). *Sobolev Spaces*. Academic Press, New York.
- Ali, S., Najjar, I. M. R., Sadoun, A. M., & Fathy, A. (2024). Navigating cardiovascular dynamics through mathematical modeling of arterial blood flow. *Ain Shams Engineering Journal*, 15(4), 102594. <https://doi.org/10.1016/j.asej.2023.102594>
- Alimov, N. (2023). Blood Supply to the Human Body, Vascular Anatomy and Blood Components. *Western European Journal of Medicine and Medical Science*, 1(4), 4-14.
- Arjmandi-Tash, O., Razavi, S. E., & Zambouri, R. (2011). Possibility of atherosclerosis in an arterial bifurcation model. *BioImpacts*, 1(4), 225-228. <https://doi.org/10.5681/bi.2011.032>
- Chiu, J.-J., & Chien, S. (2011). Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives. *Physiological Reviews*, 91(1), 327-387. <https://doi.org/10.1152/physrev.00047.2009>
- Cook, A. W., & Cabot, W. H. (2005). Hyperviscosity for shock-turbulence interactions. *Journal of Computational Physics*, 203(2), 379-385. <https://doi.org/10.1016/j.jcp.2004.09.011>
- Fisher, A. B., Chien, S., Barakat, A. I., & Nerem, R. M. (2001). Endothelial cellular response to altered shear stress. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 281(3), L529-L533. <https://doi.org/10.1152/ajplung.2001.281.3.L529>
- Formaggia, L., Quarteroni, A., & Veneziani, A. (Eds.). (2010). *Cardiovascular Mathematics: Modeling and simulation of the circulatory system* (Vol. 1). Springer Science & Business Media.

- Gaidai, O., Cao, Y., & Loginov, S. (2023). Global cardiovascular diseases death rate prediction. *Current Problems in Cardiology*, 48(5), 101622. <https://doi.org/10.1016/j.cpcardiol.2023.101622>
- Girault, V., & Raviart, P. A. (1979). *Finite element approximation of the Navier-Stokes equations* (Vol. 749). Berlin: Springer.
- Hecht, F. (2012). New development in FreeFem++. *Journal of Numerical Mathematics*, 20(3-4), 251-266. <https://doi.org/10.1515/jnum-2012-0013>
- John, V. (2004). Reference values for drag and lift of a two-dimensional time-dependent flow around a cylinder. *International Journal for Numerical Methods in Fluids*, 44(7), 777-788. <https://doi.org/10.1002/flid.679>
- Kleinstreuer, C. (2016). *Biofluid dynamics: Principles and selected applications*. CRC Press.
- Layton, W. (2008). *Introduction to the numerical analysis of incompressible viscous flows*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718904>
- Ma, L., Yan, C., & Yu, J. (2022). Suitability of an Artificial Viscosity Model for Compressible Under-Resolved Turbulence Using a Flux Reconstruction Method. *Applied Sciences*, 12(23), 12272. <https://doi.org/10.3390/app122312272>
- Manzari, M. T. (1999). An explicit finite element algorithm for convection heat transfer problems. *International Journal of Numerical Methods for Heat & Fluid Flow*, 9(8), 860-877. <https://doi.org/10.1108/09615539910297932>
- Margolin, L. G., & Lloyd-Ronning, N. M. (2023). Artificial viscosity—then and now. *Meccanica*, 58(6), 1039-1052. <https://doi.org/10.1007/s11012-022-01541-5>
- Nair, M. (2017). *Circulatory system. Fundamentals of anatomy and physiology for nursing and healthcare students* (2nd Ed.). Chichester: Wiley Blackwell.
- Quarteroni, A., Veneziani, A., & Zunino, P. (2002). Mathematical and numerical modeling of solute dynamics in blood flow and arterial walls. *SIAM Journal on Numerical Analysis*, 39(5), 1488-1511. <https://doi.org/10.1137/S0036142900369714>
- Reneman, R. S., & Hoeks, A. P. (2008). Wall shear stress as measured in vivo: consequences for the design of the arterial system. *Medical & Biological Engineering & Computing*, 46(5), 499-507. <https://doi.org/10.1007/s11517-008-0330-2>
- Selmi, M., Belmabrouk, H., & Bajahzar, A. (2019). Numerical study of the blood flow in a deformable human aorta. *Applied Sciences*, 9(6), 1216. <https://doi.org/10.3390/app9061216>
- Sjösten, W., & Vadling, V. (2018). *Finite Element Approximations of 2D Incompressible Navier-Stokes Equations Using Residual Viscosity*. Uppsala Universitet.
- Taylor, C. A., Petersen, K., Xiao, N., Sinclair, M., Bai, Y., Lynch, S. R., & Schaap, M. (2023). Patient-specific modeling of blood flow in the coronary arteries. *Computer Methods in Applied Mechanics and Engineering*, 417(Part B), 116414. <https://doi.org/10.1016/j.cma.2023.116414>
- Velten, K., Schmidt, D. M., & Kahlen, K. (2024). *Mathematical modeling and simulation: introduction for scientists and engineers*. John Wiley & Sons.
- WHO (World Health Organization) (2019). Cardiovascular diseases (CVDs). (Accessed 11/06/2024) https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1503494>

Comparing Machine Learning Algorithms for Rice Yield Prediction in Adamawa and Cross Rivers States of Nigeria

Joseph Abunimye INGIO¹ Augustine Shey NSANG¹ Aamo IORLIAM^{2*} ¹ Department of Computer Science, SITC, AUN, Yola, Nigeria² Data Science Department, SITC, AUN, Yola, Nigeria

Keywords	Abstract
Machine Learning Algorithms Cross River Adamawa Prediction	Rice production is critical for global food security, and accurate yield prediction empowers informed decision-making. This paper investigates machine learning (ML) techniques for rice yield prediction in Adamawa and Cross River states, with distinct agroclimatic conditions. Traditional yield prediction methods that are commonly used often have limitations such as less insights into the available data and reduced accuracy. Hence, this research explores the potential of machine learning for improved prediction accuracy. We leverage climatic data and historical rice yields to train and evaluate Decision Trees, Random Forest, Support Vector Regressor, Polynomial Regressor, Multiple Linear Regression and Long Short-Term Memory (LSTM) models. Performance is compared using Mean Squared Error, Root Mean Squared Error, Coefficient of Determination, Mean Absolute Error, and Mean Absolute Percentage Error. Feature selection identifies All-sky Photosynthetically Active Radiation (PAR) as the most influential factor. Linear Regression emerges as the superior model, achieving an R ² of 0.90 (Adamawa) and 0.91 (Cross River), demonstrating robust generalizability across regions. This research contributes to the development of ML-powered Agro-information systems for two Nigerian regions, enhancing agricultural practices and food security.

Cite

Ingio, J. A., Nsang, A. S., & Iorliam, A. (2024). Comparing Machine Learning Algorithms for Rice Yield Prediction in Adamawa and Cross Rivers States of Nigeria. *GU J Sci, Part A, 11(3)*, 481-496. doi:[10.54287/guj.1503494](https://doi.org/10.54287/guj.1503494)

Author ID (ORCID Number)	Article Process
0009-0007-3574-8560	Joseph Abunimye INGIO
0000-0002-6466-9032	Augustine Shey NSANG
0000-0001-8238-9686	Aamo IORLIAM
	Submission Date 23.06.2024
	Revision Date 11.07.2024
	Accepted Date 06.08.2024
	Published Date 28.09.2024

1. INTRODUCTION

The role of agriculture and food production in achieving one of the major sustainable development goals of the United Nations (UN) has made it a major topic of discussion on a global scale with a focus on improving food security and decreasing hunger to a considerable extent by 2030 (Rosa, 2017). The startling surge in the number of people facing food crises and hunger is the basis for this goal. There were 691–783 million hungry people in the world in 2022 alone, which is approximately 122 million more than the figures in 2019 (FAO et al., 2023). This shows an obvious need for the production of more food, particularly the most important and widely eaten ones, to meet the global demand which is increasing at a rapid pace.

Rice is a food crop that is consumed by a great number of people constituting over half of the world's population (Gnanamanickam, 2009; Das et al., 2018) and it has been termed "the world's most important food crop" (Zeigler & Barclay, 2008). To raise awareness of the role of rice in reducing poverty and malnutrition, the United Nations declared 2004 to be the "International Year of Rice". This further registered its importance as a food source and its widespread consumption globally (Gnanamanickam, 2009). In addition, rice is seen as a commodity that can boost a nation's economic growth as it is a major export commodity for countries like China, India, The Philippines, etc.

*Corresponding Author, e-mail: aamo.iorliam@aun.edu.ng

In Nigeria, rice has emerged as a staple food over the past few decades, enjoyed in every part of the country (Gyimah-Brempong et al., 2016; Kamai et al., 2020). Rice production amounts to about 8.3 million metric tonnes of unmilled rice per year and about 5.4 million metric tonnes of milled rice per year which constitutes 46% of the total rice produced in Africa (Sasu, 2023). Nevertheless, this rate of production is insufficient to meet the nation's rising rice demand, which has increased reliance on rice importation to satisfy the teeming population of rice eaters in the country. In 2014 half the quantity of rice consumed in Nigeria was imported (Gyimah-Brempong et al., 2016) and in 2018 over 7 trillion Naira was spent importing rice into Nigeria (Okonkwo et al., 2021). The surge in demand can be attributed to various factors, including shifts in consumer preferences, population growth, growing incomes and a swift urbanization process (Kamai et al., 2020). The rice being produced in Nigeria is cultivated in about 21 states in Nigeria with 8 states producing over 50% of the total amount of rice produced in Nigeria. Most of the states with the highest quantity of rice produced are in the Northern region of the country and they include Kebbi, Kaduna, Kano, and Borno while a few are in the Southern region, and they include Cross River, and Ebonyi. Cultivation of rice is usually done in rainfed lowland fields as well as rainfed highland fields during the raining season which typically spans between May and August. This, however, spells some challenges for rice farmers in northern Nigeria as about 1200mm to 1600mm of rainfall is needed for optimum growth of rice and this volume of rainfall does not occur in the North. In addition, there is the challenge of pest infestation and poor soil fertility as a result of increased pressure on land resources due to population expansion (Kamai et al., 2020). Farmers in the southern region of Nigeria who cultivate rice also face some challenges despite the increased volume of rainfall in the region when compared to the north. This is because of the general inconsistency of rainfall as in times past hence there is a risk of planting and not having enough rain for proper growth of seeds. These challenges and trends in agriculture have been captured in comprehensive datasets that are available for the 36 states in Nigeria.

The Agricultural data available can provide valuable insights on trends and patterns that can be used in analysis and prediction. Using data mining techniques is one way to accomplish this. Data mining is a process in which large datasets are searched through in an attempt to uncover new patterns and relationships (Geetha, 2015; Vanitha et al., 2019) with a goal to extract knowledge from the data and convert it to a human understandable format. This constitutes a major preliminary step towards the application of machine learning methods to forecast, or take action based on the knowledge found in the data.

The predominant technique for predicting crop yield among farmers in Nigeria mostly employs a crude method of estimating the yield of a particular crop based on previous yield with very little consideration given to possible climatic and environmental factors that may have changed after the previous yield. Data mining and Machine learning techniques can help increase the prediction accuracy as those factors are taken into account when building machine learning models for crop prediction thereby increasing the predictability and accuracy of the predicted yield.

The integration of machine learning into agriculture holds promise as it can bring advantages. One major benefit is the ability to make predictions, which helps reduce errors that occur when relying on manual forecasting, enabling informed decision-making processes, and promoting further growth in the agricultural sector. This technological advancement has the potential to address challenges previously mentioned, such as bridging the gap, between rice demand and production in Nigeria.

This research aimed at comparing and evaluating the performance of different machine learning algorithms for rice yield prediction in two distinct geo-climatic regions of Nigeria (Adamawa and Cross River States). Our specific objectives were to:

- (i) identify key features and potential discrepancies in the rice yield and climate datasets from the two regions through exploratory data analysis. This helped in understanding the data's characteristics and any potential biases.
- (ii) build and compare the performance of six machine learning models namely: Decision Trees, Random Forest, Support Vector Regressor, Polynomial Regressor, Multiple Linear Regression and Long Short-Term Memory (LSTM) which were implemented in Python using a Jupyter notebook.

- (iii) evaluate the performance of models implemented in (ii) using standard metrics like the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). This allowed us to compare the effectiveness of these different algorithms in predicting rice yields.

By achieving these objectives, this research contributed valuable insights into the feasibility and effectiveness of machine learning techniques for rice yield prediction in Adamawa and Cross River States of Nigeria for the first time. Furthermore, this study's findings lays the groundwork for a crop yield prediction system utilizing the best-performing machine learning model, in our case Linear Regression. This system can empower farmers and decision-makers to optimize resource allocation and improve agricultural planning. As such, increasing rice production and enhancing food security across the nation. The rest of this paper is organized into these sections; Section two presents a Literature Review, Section three explains the methodology employed in the research, section four deals with the results and discussions, and section five presents the conclusion, implication and future direction.

2. REVIEW OF RELATED STUDIES

Agricultural processes have long been carried out manually and much of it is still done that way in most developing countries including Nigeria. In sub-Saharan Africa, up to 65% of farming is done manually, about 25% is done using animal traction (donkeys, bulls' carts etc.) and about 10% is mechanized (Onwude et al., 2018). As a result, farming is seen to be a laborious task. This notion continued until the introduction of mechanized farming and a widespread introduction of tractors into land processing, facilitated by the shortage of food, workers, and draft animals caused by the World War (Karasev, 2023). With this new development came the advantages of large-scale farming and an increased efficiency in food production. However, the introduction of modern technologies for agricultural mechanization encountered some hindrances in many developing countries due to factors such as compatibility with the environment, availability of resources, cost, government policies, adequacy, and appropriateness. Consequently, farmers in these countries have inadequately used available resources, resulting in low productivity and high production costs (Onwude et al., 2018). These hindrances are not the only factors responsible for the low agricultural productivity. Challenges such as climate changes, urban encroachment, and a lack of qualified farmers, have brought about new practices for sustainable agriculture and food supply (Elbeheiry & Balog, 2022). Precision agriculture, also referred to as smart farming, has arisen as a cutting-edge approach to tackle these existing challenges threatening the sustainability of agricultural practices (Sharma et al., 2021). Sometimes shortened to digital agriculture, it utilizes modern information technologies, software, and smart devices to enable data-driven, sustainable farm management. Essentially, it employs technology-enabled tools to assist decision-making in agricultural operations (Pierce & Nowak, 1999; Sharma et al., 2021). This is ultimately aimed at reducing the cost of food production and the environmental impact of agricultural practices while maintaining an optimum yield and profitability.

Precision agriculture technologies (Figure 1) can be categorized into five groups according to Pierce and Nowak (1999) – Geographic Information Systems (GIS), Global Positioning Systems (GPS), sensors, computers, and application control tools.

Yield Prediction appears to be one of the most challenging tasks in Precision Agriculture (van Klompenburg et al., 2020) because there are several parameters that contribute to the optimum yield of a particular crop specie and these parameters vary from one species to another. As a result, many models have been proposed so far. Conventional approaches to predicting rice yields prior to harvest have predominantly consisted of statistical regression models (Mariappan & Ben Das, 2017), process-based crop simulation models grounded in agronomic principles like the CERES model (Ritchie et al., 1998), and traditional farmer knowledge and observations.

While valuable, these traditional statistical and simulation modelling techniques face several limitations in accurately capturing the multitude of complex, often non-linear interactions between the diverse factors that influence rice yield in the real world (Khaki & Wang, 2020). Crop models are data-hungry, requiring extensive inputs that may not be available, and make assumptions that restrict their generalizability (Wart et al., 2013).

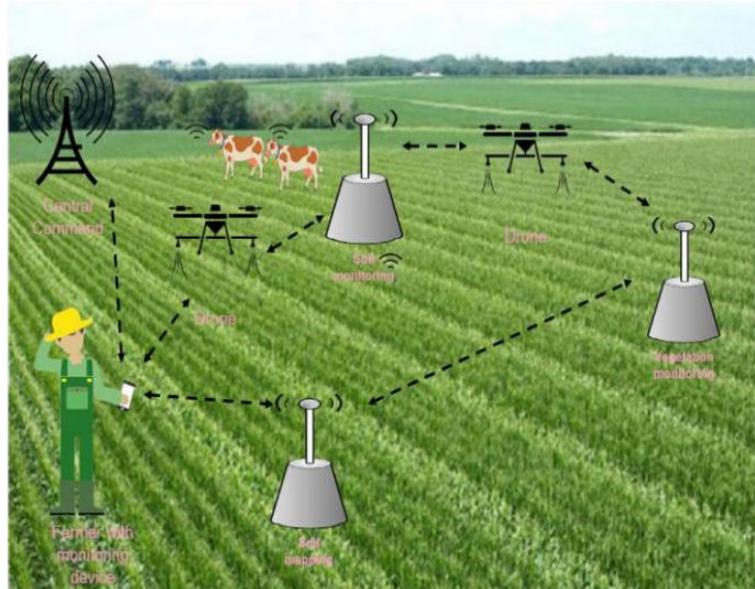


Figure 1. Precision Agriculture source: (Sharma et al., 2021)

Traditional farmer knowledge is grounded in local experience but can lack quantitative rigor and predictive precision (Van Asten et al., 2009). It may also fail to holistically integrate the array of biotic and abiotic stresses across the crop cycle that cumulatively shape final yields.

These limitations have motivated increasing research into leveraging machine learning techniques as an alternative, data-driven approach for developing more accurate and robust yield prediction models.

van Klompenburg et al. (2020) performed a detailed review of literature based on crop yield prediction using Machine learning and deep learning over the span of more than a decade and their findings revealed the most used machine learning algorithms, the most preferred features for crop yield prediction and which evaluation parameters occur in literature relating crop yield prediction. The research concluded that Deep learning algorithms such as Convolutional Neural Networks (CNN) was widely used followed by Linear regression which is commonly used as a benchmark but not necessarily the best performing algorithm. They identified the following as the most preferred features for crop yield prediction, Temperature, soil type, rainfall, and crop information. And the most used evaluation parameters include R^2 (Coefficient of Determination), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

Another major contributor to this research domain is Paudel et al. (2021). In their research, they developed a machine learning workflow that can be used for large scale prediction of crop yield. Having identified that the methods and data used in predicting the yield of a particular crop may not be transferable to another crop or location, their workflow focuses on a modular application of machine learning that ensures correctness and reusability and can be applied in different countries with minimal configuration changes.

Also notable is the work of Patrio et al. (2024) who compared the performance of Random Forest Regression, Gradient Boosting, SVR, K-Nearest Neighbours, Regression and Decision Tree Regression in predicting rice yield using climatic and yield data from the Sumatra island. Their study identified Linear regression as the best performing model with an R^2 -score of 85.53%.

In Nigeria, Iorliam et al. (2021) utilised machine learning techniques like Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Naïve Bayes, and Decision Tree, for Okra shelf life prediction and observed that Support Vector Machine, Naïve Bayes, and Decision Tree excellently predicted the shelf life of Okra better as compared to the other machine learning techniques they used.

Jiya et al. (2023) performed a study in Nigeria using rice yield and climatic data from Katsina state between 1970 and 2017 in which they employed various models such as Random Forest, Artificial Neural Network, Random Trees, Logistic Regression, and Naïve Bayes in predicting rice yield in Katsina State and compared

the performance of each machine learning technique. Their result showed that, Random Forest and Random Trees demonstrated better performance in predicting rice yields, offering a tool for proactive measures to ensure food security in the region. Even though this research is closely related to ours, it focused on a different location (Katsina State) and most of the machine learning algorithms we utilised were different from Jiya et al. (2023). This research is therefore motivated by Iorliam et al. (2021) and Jiya et al. (2023) with a focus on predicting rice yield in Adamawa State and Cross River State of Nigeria using machine learning techniques (Decision Trees, Random Forest, Support Vector Regressor, Polynomial Regressor, Multiple Linear Regression and LSTM).

3. METHODOLOGY

Our methodology consists of six phases and is described below and summarized in Figure 2:

- i. Data Collection – Secondary data from NASA POWER and the National Bureau of Statistics database (NBS) was utilised in this study.
- ii. Data Exploration phase – This stage involved visualizing the data using charts to comprehend the data.
- iii. Data Preprocessing phase – This phase involved data cleaning (removing missing values/outliers) and getting the right features for our proposed models.
- iv. Model Development phase - The Decision Trees, Random Forest, Support Vector Regressor, Polynomial Regressor, Multiple Linear Regression and LSTM were implemented using appropriate libraries and tools in python.
- v. Model Evaluation phase – Systematically evaluating model performance using metrics like RMSE, MAE, R-squared based on train/test splits.
- vi. Model Optimization phase – In this phase hyperparameters were tuned to achieve the best performing model for rice prediction.

This multi-phase methodology provides a rigorous framework for testing the machine learning algorithm models based on their ability to accurately predict rice yield from the available dataset features.

3.1. Study Area and Data Collection

Adamawa State is located in northeastern Nigeria within the savannah vegetation zone. It has an area of about 36,917 km² and an estimated population of 4.9 million (NBS, 2020; ADSPC, 2022). The tropical climate in the state experiences wet and dry seasons with an average annual rainfall ranging from 75 -103 mm, concentrated in the wet season months of May to September. Mean annual temperatures vary from 22°C to 31°C (Adebayo, 1999; ADSPC, 2022). The vegetative landscape consists primarily of short grasses, scattered trees, and shrubs. Major cash crops grown in the state include maize, rice, cotton, sorghum, and sugarcane.

Cross River State is located in the southern coastal region of Nigeria within the tropical rainforest vegetation zone. It covers an area of 20,156 km² and has a population of approximately 4.2 million (NBS, 2020). The state has abundant rainfall exceeding 3036 mm annually, along with high relative humidity. Temperatures remain relatively constant throughout the year, averaging between 15°C to 30°C. The natural vegetation is dense rainforest rich in timber resources. Major crops grown include rice, cassava, oil palm, cocoa, rubber, and plantains. The Cross River basin provides favourable conditions for wetland rice cultivation.

3.1.1. Justification for Study Area Selection

Adamawa and Cross River states were strategically selected for this paper due to their importance for rice production in Nigeria combined with their distinct geo-climatic characteristics. Both states contribute a substantial amount to the total quantity of rice produced in Nigeria. Comparing model performance between these two Agro-ecological zones with different climates, soil conditions, and farming practices provides

insights into the transferability of the machine learning algorithms. Any model that consistently performs well in both locations is likely to generalize effectively to other rice-growing regions of Nigeria. The multi-year timeseries data from the two states also enables training sophisticated machine learning models for yield forecasting particularly the deep-learning model. This paper provides a template for expanding prediction efforts to more rice-producing states in the future.

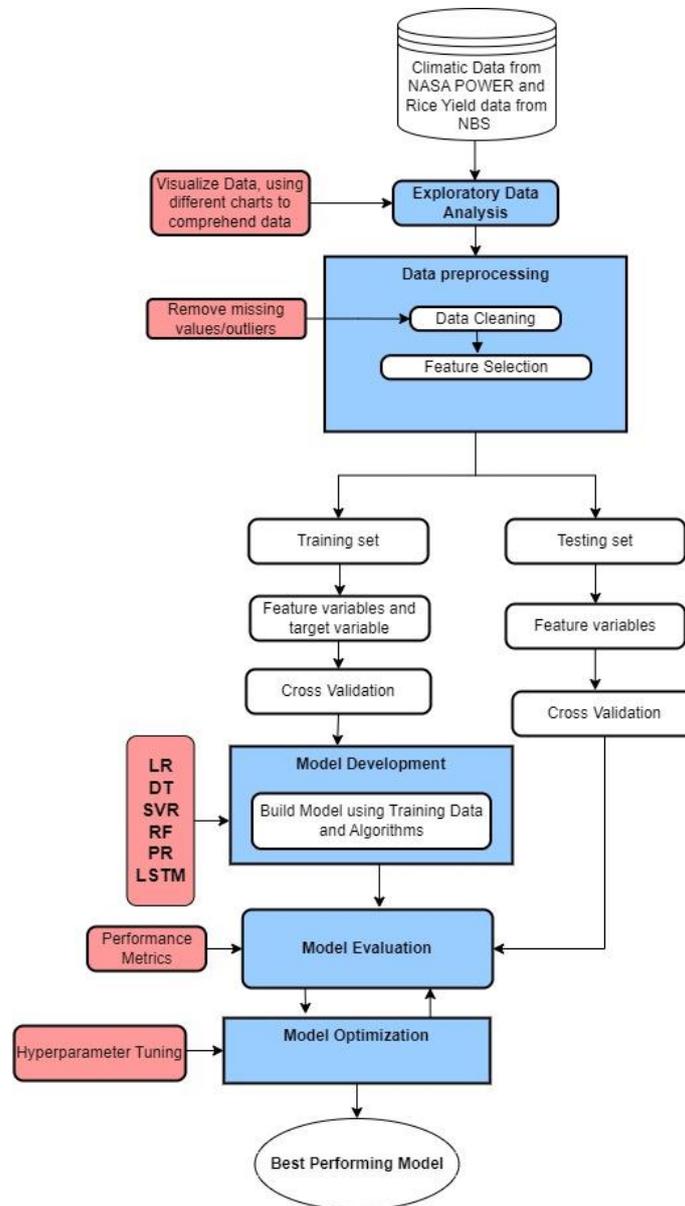


Figure 2. Proposed Methodology workflow.

3.1.2. Data Collection

Annual rice yield data (tons/hectare) for the period 1997 to 2020 was collected for each state from the National Bureau of Statistics database. The NBS data is compiled from state-level agricultural production surveys and provides authoritative aggregated statistics on crop yields.

Corresponding climatic data was obtained from the National Aeronautics and Space Agency(NASA)'s POWER (Prediction of Worldwide Energy Resource) project, which provides global meteorological data derived from satellite observations and numerical weather prediction models. Specific location coordinates within each state were used to retrieve POWER API data: Long. 11.41° (+2.02°), Lat. 8.02° (+2.72°) for Adamawa and Long. 8.39° (+0.51°), Lat. 4.99° (+1.77°) for Cross River.

The NASA POWER data parameters include:

- Precipitation - Total monthly rainfall (mm)
- Minimum temperature - Monthly minimum temps (°C)
- Maximum temperature - Monthly maximum temps (°C)
- Specific humidity - Monthly average specific humidity (kg/kg)
- Photosynthetically active radiation - Monthly average downward surface shortwave flux (W/m^2)
- Wind speed – the average wind at 2 metres above the ground (m/s)
- Average Temperature – Monthly average temperature
- Relative Humidity – Monthly average relative Humidity

The NBS yield data was combined with the 18-year POWER climatic data for each state to compile the input dataset that was used in training machine learning models and also used in the testing too. The dataset was screened for any missing values and outliers. Rows with missing values were removed.

3.2. Machine Learning Algorithms Utilised

3.2.1. Decision Trees (Regression trees)

A decision tree is a binary tree that separates data into pure leaf nodes, or data that belong to a single class (homogeneous class), repeatedly. The decision node (parent) and the leaf node(child) are the main components in a decision tree. Leaf nodes determine the class of a new data point, while decision nodes carry a condition to split data into them. For regression analysis, this process continues until each class has just one leaf. The information gain of a node is measured by the Variance Reduction, and the trees use this information to determine which decision node to select (Chauhan, 2022). We adapt the Regression decision trees approach presented by Veenadhari et al. (2014) for rice yield dataset as follows:

1. For all rice datapoints, examine potential base cases.
 - Case 1: If all rice data points have the same value for the target variable, Return a leaf node with the predicted value as the average of the target attribute in the data.
 - Case 2: If no attributes remain, return a leaf node with the predicted value as the average of the target attribute in the data
2. For each attribute a ,
 - Calculate the normalized information gain for splitting the data based on a .
3. Choose the attribute $best_a$ with the highest normalized information gain.
4. Create a decision node that splits the data based on the value of $best_a$.
5. For each possible value v of a_best :
 - Create a child node by recursively calling `BuildRegressionTree` on the subset of data where $best_a$ has the value v .
 - Set the child node as a branch of the current decision node.
6. Return the root node

3.2.2. Support Vector Regressor (SVR)

SVR is a machine learning algorithm applied in regression analysis in a similar manner to how Support Vector Machines are used in Classification tasks. SVR has the capacity to carry out nonlinear multivariate regression with remarkable robustness and efficiency. (Cortes & Vapnik, 1995). The fundamental idea is that a linear relationship in a higher dimensional space can describe a complex nonlinear relationship between some variables. This is possible through the application of linear optimization techniques to the projection of

variables of interest into a high-dimensional space. The regression function evaluated, is then applied back to the low-dimensional phase space of the variables that were first observed (Oguntunde et al., 2018). The support vector regressor (Drucker et al., 1996) is given by.

$$y = f(x) + \epsilon. \quad (1)$$

The function $f(x)$ is estimated by:

$$f(x) = \sum_{i=1}^n y_i \alpha_i K(x_i, x) + b, \quad (2)$$

where $K(x_i, x_j)$ = kernel function,

α_i = vector of weight of i^{th} point (Lagrange Multiplier)

b = constant scalar, and ϵ = error term.

3.2.3. Polynomial Regression

This is performed by regressing the dependent variable on the powers of the independent variable (Ostertagová, 2012).

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \dots + \beta_k x_i^k + e_i, \text{ for } i = 1, 2, \dots, n \quad (3)$$

the polynomial's degree is denoted by k and it is the order of the model. This is basically the same as having multiple linear regression models with $X_1 = X$, $X_2 = X^2$, $X_3 = X^3$, etc

3.2.4. Random Forest

Random Forest Regressor is an algorithm used for regression machine learning tasks. It belongs to the ensemble learning family, specifically based on the Random Forest algorithm. It is constructed using multiple decision trees on different subsets of a given dataset, averaging their outcomes to raise the dataset's estimated accuracy (Mwiti, 2022). Mwiti (2022) explained that:

- A distinct sample of rows is used to create each tree, and another sample of features is chosen for splitting at each node.
- Every single tree makes a unique prediction.

To arrive at a single result, these predictions are then averaged.

3.2.5. Multiple Linear Regression

Multiple Linear Regression, also known as Multilinear Regression, is a machine learning algorithm that utilizes statistical regression analysis in predicting the value of an output variable based on a set of input variables. It is an extension of Linear Regression, which is a multivariate technique. Regression analysis aims to construct mathematical models that describe or explain the relationships that may exist between variables (Seber & Lee, 2003). The simplest case is Simple Linear Regression, where there is only one dependent variable and one independent variable. In contrast, Multiple Linear Regression involves more than one independent variable to predict one or more dependent variables. Machine Learning algorithms based on regression analysis are commonly applied in forecasting and, in some cases, to determine the causal relationship between the dependent and independent variables (Maulud & Abdulazeez, 2020). Forecasting in regression analysis occurs when an equation of the form:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon = \beta_0 + \sum_{i=1}^p X_i \cdot \beta_i + \epsilon \quad (4)$$

Where $X_1, X_2, X_3, \dots, X_p$ are the independent variables or features used to predict the dependent variables or target variable: y is evaluated, where ε is an random variable that cannot be observed, and is also referred to as the error component, with mean 0 and variance σ^2 .

The relationship described by (4) is known as a multiple linear regression model, β_0 is the intercept, $\beta_1 \dots \beta_p$ are the slope coefficients for each independent variable and $\sigma^2 > 0$ is an unknown error variance (Pečkov, 2012).

3.2.6. Long short-term Memory (LSTM)

LSTM is a deep learning algorithm that is ideally used in regression analysis and works well with time series data. It is a variant of recurrent neural network (RNN) that has the ability to learn long-term dependencies in time series or sequence data and is mostly used for forecasting (Arras et al., 2019; Özdoğan-Sarıkoç et al., 2023). This is because standard RNNs have limited memory capacity and struggle to learn dependencies between sequence elements that are separated by long gaps. To address this limitation, the long short-term memory (LSTM) network was developed as an extension of RNNs, constructed from specialized memory blocks or cells. These LSTM cells act as memory units, with the specific purpose of retaining information over extended periods. By maintaining an internal cell state, LSTMs can preserve knowledge of past context to better link widely spaced events and make predictions informed by long-range sequential correlations. The explicit memory of LSTM cells provides the ability to learn temporal dependencies that conventional RNNs lack (Imani, 2019).

4. RESULTS AND DISCUSSION

The results derived from applying the phased methodology described in the previous section on the datasets gotten from Adamawa and Cross river states respectively are presented in this section.

4.1. EDA and Pre-processing

During the Exploratory Data Analysis phase, the datasets from both states were visualized to view its properties and distribution. Several missing values were observed in the Cross-river dataset. These missing values were removed, by removing the rows containing them.

4.1.1. Correlation Matrix between Variables

To better understand the relationship between the features and the variable in terms of correlation and which features are most important for prediction, a correlation matrix for each of the states was generated as shown in Figure 3 and 4.

Figure 3 shows a positive correlation between specific humidity (sp_humidity) and precipitation (precipitation). The coefficient of correlation between these variables is 0.72 suggesting an increase in sp_humidity whenever precipitation increases.

A strong positive correlation also exists between re_humidity and sp_humidity (0.90) and with precipitation (0.84) suggesting multicollinearity among these variables. This means that these variables contain redundant information. The model might struggle to distinguish the independent effect of each on yield, leading to inaccurate coefficient estimates and increased variance. Hence such variables are not ideal choices as predictors.

The negative correlation between temperature and yield can be helpful as it clarifies the relationship between temperature and yield (indirectly through precipitation).

From Figure 4, some variables are observed to have relatively high correlations with each other, indicating potential multicollinearity issues. For instance, "t_max" (maximum temperature) and "av_temp" with a correlation of 0.72. The variables "cl_sky_par" (clear sky radiation) and all_sky_par (all sky radiation) have a correlation of 0.31. Other variables like "t_min" (minimum temperature), sp_humidity (specific humidity),

all_sky_par (all sky radiation) have low or near-zero correlations with the "yield" variable, suggesting they may have little predictive power for the yield. Features with high multicollinearity were omitted from the training and testing set during feature selection.

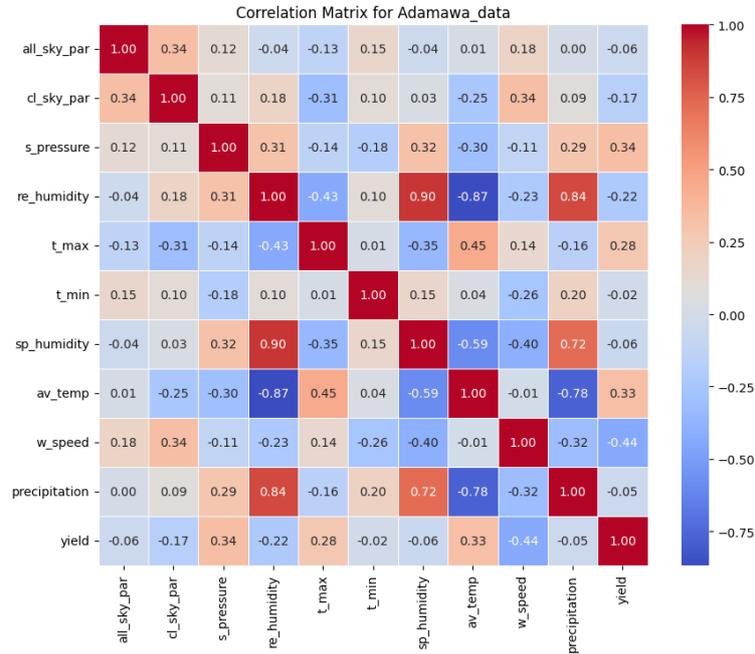


Figure 3. Correlation Matrix for Adamawa State Data

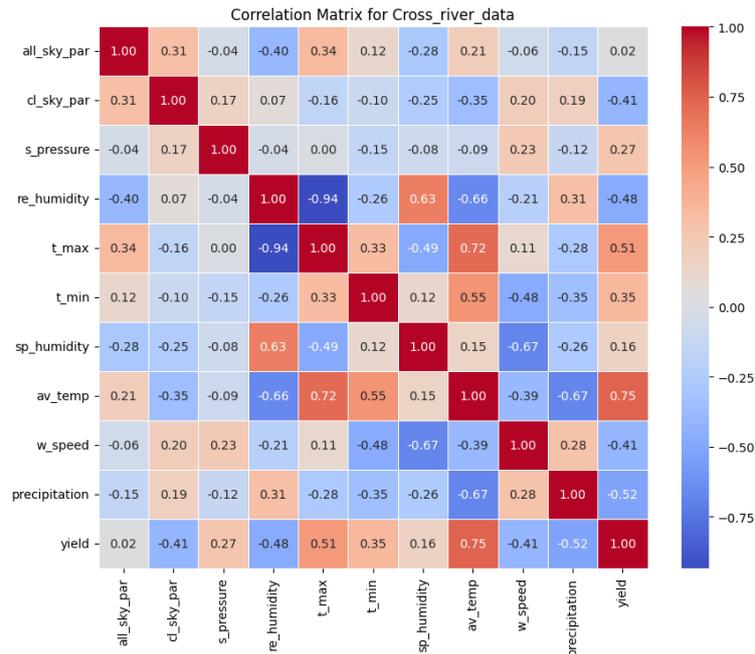


Figure 4. Correlation Matrix for Cross River State Data

4.1.2. Feature Selection

A Recursive feature elimination process was carried out to select the most relevant features that will best predict the target variable and the features selected were; s_pressure, t_max, av_temp, w_speed, all_sky_par, t_min. A similar set of features were selected when a different feature selection (F-regression) technique was used. The F-regression selection technique reduces the dimensionality of data by selecting a subset containing the most relevant features for our regression models based on their F-value scores or scores from analysis of

variance. The following features were selected: s_pressure, re_humidity, sp_humidity, av_temp, w_speed, precipitation, cl_sky_par, t_min.

However, several iterations of the training and testing of the selected algorithms did not yield a good performance using the features listed above. Hence, the number of features to select was reduced to 5 (for RFE: n_features_to_select= 5, and k = 5 F-regression). Both feature selection techniques presented the following features as the most relevant: all_sky_par, re_humidity, t_max, w_speed, s_pressure. Therefore, these features were used in the building the models for rice yield prediction.

4.2. Model Evaluation

The following performance metrics were applied in evaluating the performance of the Models built using the machine learning algorithms mentioned in previous sections. Below is a brief description of each metric and its significance.

- **MSE:** This stands for Mean Squared Error. It is a measure of the average squared difference between the predicted rice yields and the actual yields. Lower MSE indicates a better fit between predictions and actual values.
- **R2_Score:** This is R-squared also referred to as the coefficient of determination. It is the variance (squared correlation) in the dependent variable (rice yield) that can be explained by the independent variables (features used in the model). Values closer to 1 generally indicate a better fit.
- **MAE:** This stands for Mean Absolute Error. It represents the average absolute difference between predicted and actual rice yields. Lower MAE indicates better model performance.
- **RMSE:** This stands for Root Mean Squared Error. It's the square root of the MSE and represents the standard deviation of the prediction errors. Lower RMSE indicates better performance.
- **MAPE:** This stands for Mean Absolute Percentage Error. It represents the average absolute percentage difference between predicted and actual rice yields. Lower MAPE indicates better performance.

The metrics mentioned above were used to measure the accuracy of each model's performance in predicting the yield for both states and the following section contains an analysis of the results.

4.3. Analysis of Evaluation Results

Table 1 and 2 contains a concise summary of the values for each model's performance capture using the metrics mention in the previous section. Providing insights on the accuracy of each model built during this study.

Table 1. Results from Cross River Dataset

	Model_Name	MSE	R2_Score	MAE	RMSE	MAPE
0	Linear Regression	4.719138e+07	0.917511	5693.183176	6869.597977	2.495869
1	Polynomial Regression	3.188555e+09	-4.573497	44272.198092	6869.597977	18.972183
2	Decision Tree	3.887149e+08	0.320538	9747.800000	19715.854640	4.092578
3	Random Forests	1.274798e+08	0.777169	9087.848000	11290.693672	3.858125
4	Support Vector Regression	6.053853e+08	-0.058195	24461.999912	24604.578759	10.907300
5	LSTM	4.681418e+10	-91.935435	215198.636118	216365.849919	99.999360

Based on the results displayed in Table 1, Multiple Linear Regression outperforms the other models in predicting the yield using the Cross River state dataset, with an MSE value of 4.72e+07 which appears to be the lowest among the listed models. It also has a relatively high R2_Score (0.91), indicating a good fit between predictions and actual yields. Additionally, Multiple Linear Regression has a lower MAE (5,693.18) and RMSE (6,869.59) compared to other models, suggesting a good balance between underestimation and overestimation of rice yields. The MAPE value for Multiple Linear Regression is also relatively low (2.49%) showing a low average percentage deviation between the predicted and actual values. Its performance is closely

followed by that that of Random Forest which demonstrated a moderately high predictive accuracy with and MSE value of 1.274E+8, an R2-score of 0.77 and relatively low values for MAE, RMSE and MAPE.

Table 2. Results from Adamawa Dataset

	Model_Name	MSE	R2_Score	MAE	RMSE	MAPE
0	Linear Regression	5.781650e+07	0.903143	6211.029384	7603.715936	2.804602
1	Polynomial Regression	5.781650e+07	-5.586385	43435.593867	62702.307730	19.325156
2	Decision Tree Regression	1.820654e+08	0.694995	8221.547143	13493.160431	3.395051
3	Random Forests Regression	5.889005e+07	0.901344	5461.879000	7673.985520	2.356664
4	Support Vector Regression	1.045834e+09	-0.752036	24379.475714	32339.362564	10.239843
5	LSTM	4.391004e+10	-118.700195	208670.084874	7603.715936	99.999351

From the results in Table 2 Multiple Linear Regression also demonstrated the best performance on the Adamawa dataset when compared with the other algorithms. It has the lowest MSE (5.78e+07), and a relatively high R2_Score (0.903), indicating a very good fit between predictions and actual yields. Additionally, Multiple Linear Regression has a lower RMSE (7,673) compared to other models, suggesting a good balance between underestimation and overestimation of rice yields.

However, the MAPE and MAE value for Multiple Linear Regression seems to be slightly higher than that of Random Forest which has the lowest values of 5461.87 for MAE and 2.35% for MAPE, showing a low average percentage deviation between the predicted and actual values. Multiple Linear regression achieved the best performance in 3 out of the five metrics hence it is considered the best performing algorithm on the Adamawa dataset closely followed by Random Forest algorithm.

5. CONCLUSION, IMPLICATION AND FUTURE DIRECTION

This paper aimed at comparing the performance of five different machine learning algorithms in predicting rice yield in two distinct geo-climatic regions in Nigeria. The machine learning algorithms include Decision Trees, Random Forests, Multiple Linear Regression, Polynomial Regression, Support Vector Regression and Long Short-Term Memory (LSTM) neural networks.

Extensive data on rice yields, and weather pattern, were obtained. These datasets underwent preprocessing, cleaning, and separation into training as well as testing sets. In each of the five machine learning models, data for each region was trained and tested. To evaluate their predictive accuracy a complete range of model evaluation metrics like mean squared error, R-squared and mean absolute errors were computed. Multiple Linear Regression turned out to be superior to all other algorithms across both geographic regions when it came to yield prediction precision. It has an advantage over the other algorithms since it can capture long-range dependencies in time-series data. The results also showed that Random forests displayed a relatively high predictive capacity while the remaining Models exhibited poor performances due to their inability to handle non-linear relationships between features and the target variable.

As climate change continues to impact agricultural systems globally, the application of machine learning algorithms offers valuable insights and tools to address challenges in food production. Rice, a staple crop worldwide and in Nigeria, is a crucial component in tackling food insecurity and hunger crises. Timely and accurate prediction of rice yields across different regions of Nigeria can provide invaluable information to improve overall rice production and ensure food security in the country.

This study contributes significantly to the field of agricultural machine learning by providing a comprehensive comparison of multiple algorithms for rice yield prediction across diverse geo-climatic regions in Nigeria. Our findings, particularly the superior performance of Multiple Linear Regression, offer valuable insights for researchers developing crop yield prediction models in similar contexts as there is still limited work in this

domain within Nigeria. This work also underscores the importance of considering regional variations in climate and agricultural practices when developing predictive models.

In terms of practical implications for rice farming in Nigeria, this work presents the following: 1. Improved yield forecasting: Farmers and agricultural planners can use our model to make more accurate predictions of rice yields, enabling better resource allocation and planning. 2. Climate adaptation: The insights into the relationship between climatic variables and rice yields can help farmers adapt their practices to changing climate conditions in their respective regions. 3. Policy support: Government agencies can use these models to inform agricultural policies and support programs tailored to different regions.

In the future, we propose the development of a Rice Yield Prediction Support System (RYPSS) based on our best-performing Multiple Linear Regression model as its backend. This system could include: 1. A user-friendly mobile application that allows farmers to input local weather data and receive yield predictions. 2. Integration with weather APIs to automatically fetch relevant climatic data for the farmer's location. 3. Customizable features that allow farmers to adjust inputs based on their specific farming practices. 4. Regular updates to the underlying model as more data becomes available, ensuring continued accuracy. 5. An interface that allows farmers to enter current yield data that will be stored in a database that is accessible by government parastatals like the Ministry of Agriculture which will in turn facilitate informed decision-making at the government level.

Furthermore, we propose the following directions for future work: 1. Expand the study to include more regions in Nigeria, capturing a wider range of geo-climatic conditions. 2. Incorporate additional variables such as soil quality, fertilizer use, and pest incidence to enhance model accuracy. 3. Explore the integration of remote sensing data to improve yield predictions over larger areas. 4. Investigate other advanced machine learning techniques, such as ensemble methods or deep learning models, to further improve the predictive accuracy. 5. Establish robust and comprehensive agricultural databases to enable more advanced analyses and facilitating the development of even more sophisticated predictive models.

Sustained efforts in data gathering, coupled with ongoing research in machine learning techniques tailored for agricultural applications, will not only enhance our understanding of the complex interplay between various factors influencing crop yields but also empower stakeholders with actionable insights to make informed decisions and implement effective strategies for sustainable and resilient food production systems.

AUTHOR CONTRIBUTIONS

Conceptualization, J.A.I. A.S.N. and A.I.; methodology, A.I. and J.A.I.; fieldwork, J.A.I. and A.I.; software, J.A.I. and A.I.; title, J.A.I. A.S.N. and A.I.; validation, A.I. and A.S.N.; laboratory work, J.A.I. and A.I.; formal analysis, A.I. and A.S.N.; research, J.A.I. A.S.N. and A.I.; sources, J.A.I. A.S.N. and A.I.; data curation, J.A.I. and A.I.; manuscript-original draft, J.A.I. and A.I.; manuscript-review and editing, J.A.I. A.S.N. and A.I.; visualization, A.I. and J.A.I.; supervision, A.S.N. and A.I.; project management, J.A.I. A.S.N. and A.I.; funding, J.A.I. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Adebayo, A. A. (1999). *Adamawa State in Maps*. Paraclete Publishers.
- ADSPC (Adamawa State Planning Commission) (2022). Adamawa State At A Glance. <https://adspc.ad.gov.ng/adamawa-state/>
- Arras, L., Arjona-Medina, J., Widrich, M., Montavon, G., Gillhofer, M., Müller, K.-R., Hochreiter, S., & Samek, W. (2019). Explaining and Interpreting LSTMs. In: W. Samek, G. Montavon, A. Vedaldi, L. K.

- Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (pp. 211-238). Springer-VerlagBerlin, Heidelberg. https://doi.org/10.1007/978-3-030-28954-6_11
- Chauhan, N. S. (2022, February 9). Decision Tree Algorithm, Explained. KDnuggets. <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Das, B., Nair, B., Reddy, V. K., & Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology*, 62(10), 1809-1822. <https://doi.org/10.1007/s00484-018-1583-6>
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. (1996, December 3-5). *Support Vector Regression Machines*. In: M. C. Mozer, M. Jordan, & T. Petsche (Eds.) Proceedings of the Advances in Neural Information Processing Systems 9 (NIPS 1996) (pp. 155-161), Denver Colorado.
- Elbeheiry, N., & Balog, R. S. (2022). Technologies Driving the Shift to Smart Farming: A Review. *IEEE Sensors Journal*, 23(3), 1752-1769. <https://doi.org/10.1109/JSEN.2022.3225183>
- FAO, IFAD, UNICEF, WFP, & WHO. (2023). The State of Food Security and Nutrition in the World 2023. Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum. Rome, FAO. <https://doi.org/10.4060/cc3017en>
- Geetha, M. C. S. (2015). A Survey on Data Mining Techniques in Agriculture. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(2), 887-892.
- Gnanamanickam, S. S. (2009). Rice and Its Importance to Human Life. In: S. S. Gnanamanickam (Eds.), *Biological Control of Rice Diseases* (pp. 1-11). Springer Netherlands. https://doi.org/10.1007/978-90-481-2465-7_1
- Gyimah-Brempong, K., Johnson, M., & Takeshima, H. (2016). Chapter 1. Rice in the Nigerian Economy and Agricultural Policies. In: K. Gyimah-Brempong, M. Johnson, & H. Takeshima (Eds.), *The Nigerian Rice Economy* (pp. 1-20). University of Pennsylvania Press. <https://doi.org/10.9783/9780812293753-005>
- Imani, M. (2019, August 26-27). *Long Short-Term Memory Network and Support Vector Regression for Electrical Load Forecasting*. In: Proceedings of the 2019 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), Istanbul, Türkiye. <https://doi.org/10.1109/PGSRET.2019.8882730>
- Iorliam, I. B., Ikyo, B. A., Iorliam, A., Okube, E. O., Kwaghtyo, K. D., & Shehu, Y. I. (2021). Application of Machine Learning Techniques for Okra Shelf Life Prediction. *Journal of Data Analysis and Information Processing*, 9(3), 136-150. <https://doi.org/10.4236/jdaip.2021.93009>
- Jiya, E. A., Illiyasu, U., & Akinyemi, M. (2023). Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms. *Journal of Information Systems and Informatics*, 5(2), 785-799. <https://doi.org/10.51519/journalisi.v5i2.506>
- Kamai, N., Omoigui, L. O., Kamara, A. Y., & Ekeleme, F. (2020). *Guide to rice production in Northern Nigeria*. International Institute of Tropical Agriculture (IITA).
- Karasev, A. (2023). Excursion to the History of Tractor Building and the Introduction of Tractors in Agriculture. *Tekhnicheskij Servis Mashin*, 61(1), 155-163. <https://doi.org/10.22314/2618-8287-2023-61-1-155-163>
- Khaki, S., & Wang, L. (2020). *Crop Yield Prediction Using Deep Neural Networks*. In: H. Yang, R. Qiu, & W. Chen (Eds.), Proceedings of the 2019 INFORMS International Conference on Service Science (pp. 139-147). https://doi.org/10.1007/978-3-030-30967-1_13
- Mariappan, A. K., & Ben Das, J. A. (2017, April 07-08). *A paradigm for rice yield prediction in Tamilnadu*. In: Proceedings of the 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), (pp. 18-21), Chennai, India. <https://doi.org/10.1109/TIAR.2017.8273679>

- Maulud, D. H., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(2), 140-147. <https://doi.org/10.38094/jastt1457>
- Mwiti, D. (2022, July 21). Random Forest Regression: When Does It Fail and Why? Neptune.Ai. <https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why>
- NBS (National Bureau of Statistics) (2020). Demographic Statistics Bulletin. <https://nigerianstat.gov.ng/download/1241121>
- Oguntunde, P. G., Lischeid, G., & Dietrich, O. (2018). Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. *International Journal of Biometeorology*, 62(3), 459-469. <https://doi.org/10.1007/s00484-017-1454-6>
- Okonkwo, U. U., Ukaogo, V., Kenekukwu, D., Nwanshindu, V., & Okeagu, G. (2021). The politics of rice production in Nigeria: The Abakaliki example, 1942-2020. *Cogent Arts & Humanities*, 8(1), 1880680. <https://doi.org/10.1080/23311983.2021.1880680>
- Onwude, D. I., Chen, G., Hashim, N., Esdaile, J. R., Gomes, C., Khaled, A. Y., Alonge, A. F., & Ikrang, E. (2018). Mechanization of Agricultural Production in Developing Countries. In: G. Chen (Eds.), *Advances in Agricultural Machinery and Technologies* (pp. 3-26). CRC Press. <https://doi.org/10.1201/9781351132398-1>
- Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering*, 48, 500-506. <https://doi.org/10.1016/j.proeng.2012.09.545>
- Özdoğan-Sarıkoç, G., Sarıkoç, M., Celik, M., & Dadaser-Celik, F. (2023). Reservoir volume forecasting using artificial intelligence-based models: Artificial Neural Networks, Support Vector Regression, and Long Short-Term Memory. *Journal of Hydrology*, 616, 128766. <https://doi.org/10.1016/j.jhydrol.2022.128766>
- Patrio, U., Yuliska, Y., & Widayari, Y. D. L. (2024). Predicting Rice Production In Sumatra Island Using Linear Regression. In: B. Santoso, B. Bustami & A. Satria (Eds.), *Proceedings of the 11th International Applied Business and Engineering Conference (ABEC 2023)*, (2023, September 21). Bengkalis, Riau, Indonesia. <http://doi.org/10.4108/eai.21-9-2023.2342997>
- Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, 103016. <https://doi.org/10.1016/j.agsy.2020.103016>
- Pečkov, A. (2012). *A Machine Learning Approach to Polynomial Regression*. PhD Thesis. Jožef Stefan International Postgraduate School.
- Pierce, F. J., & Nowak, P. (1999). Aspects of Precision Agriculture. *Advances in Agronomy*, 67, 1-85. [https://doi.org/10.1016/S0065-2113\(08\)60513-1](https://doi.org/10.1016/S0065-2113(08)60513-1)
- Ritchie, J. T., Singh, U., Godwin, D. C., & Bowen, W. T. (1998). Cereal growth, development and yield. In: G. Y. Tsuji, G. Hoogenboom, & P. K. Thornton (Eds.), *Understanding Options for Agricultural Production* (pp. 79-98). Springer Netherlands. https://doi.org/10.1007/978-94-017-3624-4_5
- Rosa, W. (Eds.) (2017). Transforming Our World: The 2030 Agenda for Sustainable Development. In: *A New Era in Global Health* (pp. 529-567). Springer Publishing Company. <https://doi.org/10.1891/9780826190123.ap02>
- Sasu, D. D. (2023, November 9). Nigeria: Production of milled rice 2010-2023. Statista. <https://www.statista.com/statistics/1134510/production-of-milled-rice-in-nigeria/>
- Seber, G. A. F., & Lee, A. J. (2003). *Linear Regression Analysis* (2nd Ed.). John Wiley & Sons. <https://doi.org/10.1002/9780471722199>
- Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*, 9, 4843-4873. <https://doi.org/10.1109/ACCESS.2020.3048415>

- Van Asten, P. J. A., Kaaria, S., Fermont, A. M., & Delve, R. J. (2009). Challenges and lessons when using farmer knowledge in agricultural research and development projects in Africa. *Experimental Agriculture*, 45(1), 1-14. <https://doi.org/10.1017/S0014479708006984>
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>
- Vanitha, C. N., Archana, N., & Sowmiya, R. (2019, March 15-16). *Agriculture Analysis Using Data Mining And Machine Learning Techniques*. In: Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), (pp. 984-990), Coimbatore, India. <https://doi.org/10.1109/ICACCS.2019.8728382>
- Veenadhari, S., Misra, B., & Singh, C. (2014, January 03-05). *Machine learning approach for forecasting crop yield based on climatic parameters*. In: Proceedings of the 2014 International Conference on Computer Communication and Informatics, (pp. 1-5), Coimbatore, India. <https://doi.org/10.1109/ICCCI.2014.6921718>
- Wart, J. V., Kersebaum, K. C., Peng, S., Milner, M., & Cassman, K. G. (2013). Estimating crop yield potential at regional to national scales. *Field Crops Research*, 143, 34-43. <https://doi.org/10.1016/j.fcr.2012.11.018>
- Zeigler, R. S., & Barclay, A. (2008). The Relevance of Rice. *Rice*, 1(1), 3-10. <https://doi.org/10.1007/s12284-008-9001-z>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1501098>

TB-SMGAN: A GAN Based Hybrid Data Augmentation Framework on Chest X-ray Images and Reports

Hasan Berat ÖZFİDAN¹ Mehmet Ulvi ŞİMŞEK^{1*} ¹ ASELSAN, Ankara, Türkiye

Keywords	Abstract
Generative Augmentation Latent Space Manipulation Generative Adversarial Networks Chest x-ray Data Augmentation	Data augmentation is a common practice in image classification, employing methods such as reflection, random cropping, re-scaling, and transformations to enhance training data. These techniques are prevalent when working with extended real-world datasets, focusing on improving classification accuracy through increased diversity. The use of Generative Adversarial Networks (GANs), known for their high representational power, enables learning the distribution of real data and generating samples with previously unseen discriminative features. However, intra-class imbalances in augmentations are problematic for conventional GAN augmentations. Hence, we propose a framework named Text-Based Style-Manipulated GAN augmentation framework (TB-SMGAN) aims to leverage the generative capabilities of StyleGAN2-ADA. In this framework, we utilize StyleCLIP to control disentangled feature manipulations and intra-class imbalances. We enhance the efficiency of StyleCLIP by fine-tuning CLIP with x-ray images and information extractions from corresponding medical reports. Our proposed framework demonstrates an improvement in terms of mean PR-AUC score when employing the text-based manipulated GAN augmentation technique compared to conventional GAN augmentation.

Cite

Özfidan, H. B., & Şimşek, M. U. (2024). TB-SMGAN: A GAN Based Hybrid Data Augmentation Framework on Chest X-ray Images and Reports. *GU J Sci, Part A, 11(3)*, 497-506. doi:10.54287/guj.1501098

Author ID (ORCID Number)	Article Process
0009-0006-0950-2017	Submission Date 13.06.2024
0000-0001-8017-4753	Revision Date 06.08.2024
	Accepted Date 12.09.2024
	Published Date 28.09.2024

1. INTRODUCTION

Accurately classifying medical images is crucial for early diagnosis and effective treatment, but limited training data remains a critical obstacle (Liu et al., 2023). Traditional data augmentation techniques like geometric transformations and color jittering are used to expand datasets in literature. Their ability to introduce clinically relevant features is essential for applying medical tasks properly. Nevertheless, these methods are constrained to precise classification within medical images (Shorten & Khoshgoftaar, 2019).

Several approaches are explored to address data scarcity in medical image classification, each with its own set of limitations. One such approach is Conventional Data Augmentations, such as geometric transformations and color jittering, which are employed to expand datasets. However, these methods may not introduce clinically relevant features necessary for accurate diagnosis. Furthermore, they have the drawback of creating unrealistic augmentations that can potentially lead to biased models (Shorten & Khoshgoftaar, 2019; Jablonski et al., 2022; Wang & Qi, 2022). Another strategy is Self-Supervised Learning, which holds promise by learning representations from unlabeled data. Despite its potential, aligning learned features with specific diagnosis tasks proves to be challenging, and the performance in medical domains often falls short of optimal levels (Benčević et al., 2022; Ke et al., 2022; Hochberg et al., 2022). Additionally, Transfer Learning, where leveraging pre-trained models on large-scale datasets can enhance performance on smaller medical datasets. However, the presence of domain mismatch, characterized by differences between the training and target

*Corresponding Author, e-mail: mehmetulvi@gmail.com

datasets, poses a substantial challenge and can significantly impact the model's generalizability (Dao et al., 2022; 2023; Kariuki et al., 2023). Another notable approach is the use of the Deep Synthetic Minority Oversampling Technique (SMOTE), designed to address class imbalance by oversampling minority classes. Despite its effectiveness in tackling imbalanced datasets, this technique has drawbacks, including the risk of overfitting and the potential generation of unrealistic samples (Tarawneh et al., 2022; Altwaijry, 2023). Lastly, Physics-Informed Neural Networks offer an avenue to improve model generalizability by integrating physical constraints into the learning process. However, a significant challenge lies in the incorporation of accurate and relevant physical models specifically tailored for medical images (Wang et al., 2021; Islam & Mondal, 2019).

Considering these limitations, a robust data augmentation technique is required to solve this problem. This paper introduces a novel Text-Based Style-Manipulated GAN augmentation framework (TB-SMGAN) that overcomes these limitations by leveraging the combined capabilities of StyleGAN2-ADA (Karras et al., 2020) and StyleCLIP (Patashnik et al., 2021). Utilizing StyleCLIP's ability to manipulate disentangled features based on textual descriptions, TB-SMGAN extracts key clinical findings from medical reports to fine-tune the generation of disease-specific variations within each class. This allows us to control precisely over intra-class imbalances and targeted generation of discriminative features relevant to specific medical diagnoses. To further refine TB-SMGAN's ability to generate clinically relevant augmentations, we fine-tune Contrastive Language-Image Pre-Training (CLIP) (Radford et al., 2021) with x-ray images and extracted information from corresponding medical reports. This domain-specific adaptation improves the framework's effectiveness in medical tasks. Our framework demonstrates significantly improved classification performance, as measured by PR-AUC score, compared to standard GAN augmentation. This improvement showcases the potential of TB-SMGAN for overcoming data scarcity and enhancing the accuracy and generalizability of medical image classification models, ultimately contributing to improved patient care.

Our proposed framework presents several contributions to the field of medical Generative Adversarial Network (GAN) augmentation:

- 1) Rule-Based Information Extraction from X-ray Reports: We develop a novel rule-based algorithm for accurately extracting relevant information from X-ray reports. This extracted information provides additional context for medical image analysis and enhances the effectiveness of downstream tasks.
- 2) Fine-Tuning CLIP for Medical Domain Adaptation: We fine-tune CLIP using various text extraction methods specifically tailored for the medical domain. This domain adaptation ensures the learned representations are relevant and informative for medical applications, leading to improved performance in subsequent tasks.
- 3) Text-Based Latent Space Manipulations for Medical Data Augmentation: This work introduces a novel approach that utilizes text-based information to manipulate the latent space of GANs. This approach enables the generation of synthetic medical data that is not only realistic but also semantically aligned with the extracted textual information, further enriching the training dataset and improving the generalizability of deep learning models trained on augmented data.

These contributions collectively address essential challenges within the domain of medical image analysis, offering substantial advancements that contribute to the refinement of deep learning models. The results show that the proposed framework enhances accuracy and efficacy, establishing a more robust foundation for diverse clinical applications.

The rest of the paper is organized as follows: Section II overviews and discusses the related works in the literature. Then the details of TB-SMGAN framework are given in section III. We demonstrate the comparative results in performance evaluations in section IV. Finally, Section V summarizes key findings and the outputs of the paper.

2. RELATED WORKS

GAN is a type of deep learning model capable of generating synthetic data that is statistically similar to the real data (Alqahtani et al., 2021). This is achieved by training two neural networks in an adversarial manner: a generator network that attempts to create realistic synthetic data, and a discriminator network that attempts

to distinguish between real and synthetic data. This adversarial process allows the generator to progressively learn to generate increasingly realistic and diverse data (Li et al., 2022).

In recent years, GAN-based data augmentation has emerged as a powerful alternative to classical data augmentation techniques (Lacan et al., 2023). One such approach is proposed to populate training datasets with synthetic data (Bowles et al., 2018). This approach utilizes the Progressive Growing of GANs (PGGAN) architecture for effective modeling of the input data distribution. The authors compare classical and GAN augmentation across various datasets to assess the generalizability of their synthetic data augmentation strategy. Additionally, they investigated the effects of varying augmentation strengths on model performance. Experiments demonstrate that GAN augmentation improves classification accuracy compared to classical data augmentation techniques.

Medical GAN augmentation emerges as a beneficial tool for enhancing the performance of medical image classification tasks. One study collected data from Sheba Medical Center for three specific diseases and experimented with DCGAN, ACGAN, and ACGAN discriminator for data augmentation (Frid-Adar et al., 2018). The best performance is observed when the classifier is trained with additional generated images alongside real data. Similarly, another study compares both classical and GAN data augmentation for pneumonia recognition and finds that the DCGAN-augmented classifier achieved the highest accuracy, recall, and F1 score (Kora Venu & Ravula, 2020). Low classification accuracies in medical datasets are attributed to class imbalance (Deepshikha & Naman, 2020). Researchers address this by balancing the dataset with samples generated by a DCGAN, leading to improved classification performance. Moreover, another study reported superior classification performance for underrepresented classes augmented with GANs compared to classical augmentation techniques (Sundaram & Hulkund, 2021). However, this work only focused on ROC-AUC score, which is insensitive to data imbalance. Our study addresses this limitation by reporting PR-AUC scores, providing a more comprehensive and robust evaluation of the proposed approach. We leverage the conditional StyleGAN2-ADA (Karras et al., 2020) model for synthetic data generation, further exploring its potential for enhancing medical image classification.

The manipulation of latent spaces within GANs has emerged as a rapidly evolving field with significant potential for medical applications. However, the current research landscape exploring this technique in medical contexts remains relatively sparse, with limited investigations into its full capabilities.

One notable contribution employed StyleGAN architecture for manipulating latent spaces, achieving the generation of synthetic CT images corresponding to T2-weighted MR images and vice versa (Fetty et al., 2020). This work demonstrated the feasibility of cross-modality image synthesis through latent space walks between modalities. However, deeper investigations into the disentanglement properties of the style transfer process were not undertaken, leaving this area open for future research.

3. TB-SMGAN: A GAN BASED HYBRID DATA AUGMENTATION FRAMEWORK

In this section, we present the TB-SMGAN framework, which is shown Figure 1, the components of the framework are explained in detail.

3.1. Proposed TB-SMGAN System Architecture

We use a GAN framework, StyleGAN2-ADA, to generate realistic samples with a limited representational power dataset. The proposed adaptive discriminator augmentation technique stabilizes GAN training with tiny datasets and allows for diverse and realistic output generation without augmentation leakage or distortion (Karras et al., 2020). Our framework utilizes GAN augmentation to generate synthetic data samples and augment the dataset. Compared to traditional data augmentation techniques, GAN augmentation provides creative and realistic sample generation, preserving distinctive features with enough variation and fidelity. Therefore, we utilize StyleGAN2-ADA for both GAN augmentation and text-based style-manipulated augmentation experimentations.

Recent developments from OpenAI enable us to utilize the representational power of the CLIP pre-trained model for latent space manipulation (Radford et al., 2021). StyleCLIP enables semantic meaningful latent

space manipulation with text guidance. Although several approaches enable text-based style manipulations in StyleCLIP, we focus solely on the input-agnostic global direction method, which does not require further optimization. For a given pair of text that defines neutral and target attributes, the global direction method exploits the colinearity between CLIP's image embedding space and StyleGAN2-ADA's style space for the difference of the given text pair. Since the global direction method offers style manipulation on real images, we train the encoder4editing framework for manipulating real x-ray images (Tov et al., 2021). After training, we combine the pre-trained StyleGAN2-ADA generator, encoder4encoder and StyleCLIP in order to create text-based style-manipulated augmentations.

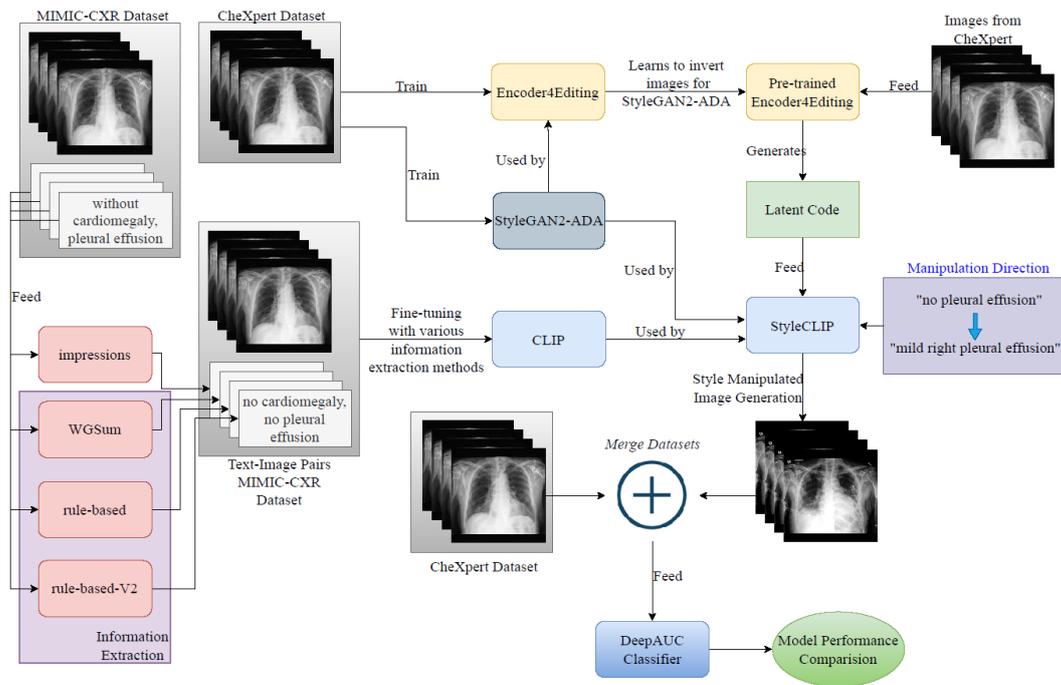


Figure 1. Style manipulated data augmentation design overview

During our initial experiments, we observed that the global optimization method does not yield the desired image manipulations in the context of chest X-rays since StyleCLIP using default CLIP weights. We infer this is due to the lack of medical representation in CLIP's training set. To address this issue, we fine-tune CLIP with medical image-text pairs originating from the MIMIC-CXR dataset (Johnson et al., 2019). We improve the fine-tuning process by incorporating NLP techniques for medical reports.

To create more informative reports, we extract entities using scispaCy, which contains spaCy models for processing biomedical-specific textual data (Neumann et al., 2019; Honnibal et al., 2020). In our experiments, we use the entities extracted by the `encore_sci_scibert` model, which is a complete spaCy pipeline built on top of a transformer and pre-trained on a large biomedical corpus consisting of nearly 785,000 words. We use only the entities extracted for each report, instead of the full report, and also use dependency relations between words. We modify the entities to create less complex semantics for each report as shown in Algorithm 1. For instance, considering an X-ray report stating "There is no focal consolidation, pleural effusion, or pneumothorax." Our rule-based approach combines entities linked to the word "No" directly or indirectly, resulting in "no focal consolidation", "no pleural effusion" and "no pneumothorax" as modified entities.

However, the results of our experiments show that the rule-based approach performs poorly for some classes. To address this issue, we analyze our rule-based approach to refine the disentanglement of disease-specific features, particularly for Edema and Consolidation classes. These classes present challenges due to their overlapping textual semantics. To overcome this, we introduce new indicator words and name the fine-tuning strategy "rule-based-V2".

Algorithm 1 Rule-Based Information Extraction

Require: Medical report text T , scispaCy model M , list of words affecting semantic complexity L

Ensure: Modified list of entities E'

```

1: Extract entities  $E$  from  $T$  using  $M$ 
2: Construct dependency relation graph  $G$  for  $T$ 
3:  $E' \leftarrow \emptyset$ 
4: for all entity  $e \in E$  do
5:   Find words  $W$  in  $L$  directly or indirectly related to  $e$  in  $G$ 
6:   for all word  $w \in W$  do
7:     if relation between  $w$  and  $e$  then
8:       Add  $w$  prefix to  $e$ 
9:        $E' \leftarrow E' \cup \{e\}$ 
10:    else
11:      Add  $e$  to  $E'$ 
12:    end if
13:  end for
14: end for
15: return  $E'$ 

```

To quantify the effectiveness of style manipulations, we utilize DeepAUC, the top-performing solution in the CheXpert competition (Yuan et al., 2021) for evaluating the manipulated images.

To connect the dots formally, we incorporated all of the aforementioned approaches into a framework named TB-SMGAN. Our proposed framework consists of several steps and uses various models and methods for information extraction, image generation, and classification as in Figure 1. TB-SMGAN proceeds with the following steps:

- 1) Train StyleGAN2-ADA: The algorithm begins by training a StyleGAN2-ADA model, denoted as (G) , on the set of real chest X-ray images $X_{CheXpert}$
- 2) Train Encoder4Encoder for GAN Inversion: Next, an Encoder4Encoder model, denoted as (E) , is trained using the previously trained StyleGAN2-ADA model ($G_{CheXpert}$) and the set of real chest X-ray images ($X_{CheXpert}$).
- 3) Extract Information and Fine-tune CLIP: For each information extraction method (i) in the set (I) consists of WGSUM, RuleBasedV1, RuleBasedV2 and Impressions, and for each pair of X-ray image and report (x, r) in $(X_{MIMIC-CXR}, R_{MIMIC-CXR})$ dataset, the algorithm fine-tunes a CLIP model, denoted as (T_i) , using the information extracted from the report $(F_i(r))$ and the X-ray image (x) .
- 4) Determine Style Manipulation Direction: For each information extraction method (i) and each disease (d) in the set of diseases (\mathcal{D}), the algorithm computes a manipulation direction (Δ_{z+d}) using the StyleCLIP model (S_i) which includes fine-tuned CLIP (T_i) , the positive and negative text prompts for the disease (P_{+d}) and (P_{-d}) respectively.
- 5) Generate Style-Manipulated Images: The algorithm generates manipulated images for each information extraction method (i), each disease (d), and each X-ray image (x) in the set of X-ray images $(X_{CheXpert}^d)$ with only disease (d) by applying the manipulation direction (Δ_{z+d}) to $(S_i(x, X_{CheXpert}))$ which generates a synthetic image (\hat{x}_{+d})
- 6) Data Selection with Classifier: Select the top performing style manipulated synthetic images for each disease (d) in the set of diseases (\mathcal{D}) by comparing PR-AUC scores using DeepAUC classifier trained on X-ray images and labels $(X_{CheXpert}, L_{CheXpert})$.
- 7) Merge Datasets: The algorithm then merges the original dataset $(X_{CheXpert}, L_{CheXpert})$ with the set of manipulated images and corresponding labels $(\hat{x}_{+d}, \hat{l}_{+d})$ to create a new dataset $(\hat{X}_{\square}, \hat{L}_{\square})$.
- 8) Evaluate Representation Power of Augmented Dataset: Finally, the algorithm trains a DeepAUC model on the augmented dataset $(\hat{X}_{\square}, \hat{L}_{\square})$ and reports the results.

Our proposed method represents a sophisticated approach to data augmentation, leveraging advanced models and techniques to generate new, manipulated images that are capable of improving the representational power of the final augmented dataset.

4. EXPERIMENTAL RESULTS

4.1. Datasets

In this paper, we employ the CheXpert dataset (Irvin et al., 2019), comprising about 225,000 chest X-ray images from Stanford University Medical Center. The dataset includes 14 classes, with uncertain labels categorized as positive, negative, or uncertain.

On our framework, we focus on five specific classes—Atelectasis, Cardiomegaly, Consolidation, Edema, and Pleural Effusion—central to the challenges in the CheXpert competition (Irvin et al., 2019). By excluding multi-class labeled samples, we refined the dataset to approximately 85,000 samples, optimizing our model's training for more focused and efficient learning in these specific classes.

4.2. TB-SMGAN Framework

This section presents the experimental results of our proposed method for augmenting chest X-ray datasets using text-based style manipulations. We evaluate the representation capacity of the augmentations by DeepAUC classifier. The key focus for classification performance evaluation is on PR-AUC scores instead of ROC-AUC scores. This metric provides a robust evaluation of our methodology, including its effectiveness in addressing class imbalance issues often encountered in medical image analysis tasks.

Table 1 shows the PR-AUC scores of the DeepAUC classifier for different fine-tuning strategies on only synthetic data generated by TB-SMGAN. The DeepAUC classifier used for evaluation is trained on the CheXpert dataset without any augmentations. Each row represents a distinct fine-tuning strategy along with plus or minus signs that indicate positive or negative manipulation direction. Columns denote the PR-AUC scores for each disease. The "Mean AUC" column presents the average PR-AUC across all diseases. Results demonstrate significant variations in performance across strategies. Notably, rule-based (+) strategy achieves the highest PR-AUC for Cardiomegaly emphasizing the effectiveness of incorporating rule-based information during fine-tuning. The impression (+) strategy leads in PR-AUC for Consolidation and Atelectasis, underlining the importance of specific fine-tuning strategies for different diseases.

Table 1. Classification results on only synthetic data generated by TB-SMGAN

Fine-Tuning Strategy	Cardiomegaly (PR-AUC)	Edema (PR-AUC)	Consolidation (PR-AUC)	Atelectasis (PR-AUC)	Pleural-Effusion (PR-AUC)	Mean AUC
rule-based (-)	0,147	0,198	0,196	0,194	0,193	0,186
impression (-)	0,191	0,203	0,159	0,169	0,175	0,179
rule-based-V2 (-)	0,234	0,228	0,126	0,169	0,227	0,197
original-CLIP (-)	0,151	0,41	0,225	0,197	0,411	0,279
WGSum-generated (-)	0,254	0,241	0,207	0,247	0,24	0,238
randomized (-)	0,19	0,269	0,182	0,261	0,312	0,243
inverted	0,241	0,294	0,219	0,221	0,277	0,251
rule-based (+)	0,579	0,235	0,17	0,364	0,432	0,356
impression (+)	0,27	0,232	0,353	0,516	0,455	0,365
rule-based-V2 (+)	0,208	0,382	0,222	0,281	0,619	0,342
original-CLIP (+)	0,339	0,38	0,336	0,26	0,205	0,304
WGSum-generated (+)	0,159	0,579	0,281	0,221	0,315	0,311
randomized (+)	0,318	0,181	0,288	0,217	0,187	0,238

Table 2 presents the results of style-manipulated GAN augmentations. The augmented dataset includes the full CheXpert dataset and the corresponding generated data, depending on the fine-tuning strategy. Experiments are conducted for various fine-tuning strategies. "Pure-dataset" and "randomized" strategies are included for benchmarking and sanity-check purposes respectively. On the one hand, "Pure-dataset" demonstrates the results of the classifier on pure CheXpert dataset without any augmentations. On the other hand, "randomized" indicates the fine-tuning strategy which uses reports with random word order. This strategy demonstrates the significance of the word order in the context of fine-tuning. Additionally, "Inverted" fine-tuning strategy demonstrates the results for the inverted images originating from Encoder4Encoder. Moreover, "StyleGAN2-Augmentation" is included to compare GAN augmentation with our proposed framework. This strategy includes class conditional image generation of StyleGAN2-ADA. In addition to the singular fine-tuning strategies, we experiment with the combinations of fine-tuning strategies. The ensemble method unifies the generated samples from various training strategies with maximum PR-AUC scores based on Table 1. For example, we include synthetic Cardiomegaly samples generated from the rule-based (+) method and synthetic Edema samples generated from the WGSUM-Generated (+) method. To account for the unavailability of an impressions section for other datasets, we also employ an ensemble-without-impression method, in which we ignore the data created by fine-tuning strategies using impressions and select the data with the maximum PR-AUC scores. For instance, we include data created by the original-CLIP (+) method for Consolidation, and we select data generated by the rule-based-V2(+) method instead of the impression (+) method. Based on the results in Table 2, the TB-SMGAN(ensemble) framework performs slightly better than GAN augmentation and performs the best among all manipulation strategies in terms of PR-AUC score.

Table 2. Classification performance results of fine-tuning strategies on text-based style manipulated GAN augmented dataset

Fine-Tuning Strategy	Cardiomegaly (PR-AUC)	Edema (PR-AUC)	Consolidation (PR-AUC)	Atelectasis (PR-AUC)	Pleural-Effusion (PR-AUC)	Mean AUC
pure-dataset	0,719	0,745	0,615	0,633	0,842	0,711
rule-based (+), inverted and rule-based-V2 (+)	0,707	0,756	0,592	0,644	0,847	0,709
impression (+)	0,688	0,766	0,562	0,68	0,855	0,71
original-clip (+)	0,738	0,779	0,566	0,667	0,843	0,719
rule-based-V2 (+)	0,658	0,8	0,643	0,66	0,841	0,72
rule-based (-)	0,641	0,782	0,728	0,624	0,837	0,722
inverted	0,7	0,788	0,66	0,651	0,841	0,728
rule-based (+)	0,684	0,796	0,659	0,693	0,821	0,731
WGSUM-generated (-)	0,659	0,791	0,671	0,698	0,834	0,731
randomized (+)	0,678	0,818	0,614	0,712	0,835	0,731
rule-based-V2 (-)	0,649	0,787	0,703	0,707	0,838	0,737
randomized (-)	0,682	0,791	0,688	0,701	0,827	0,738
WGSUM-generated (+)	0,687	0,79	0,669	0,691	0,852	0,738
ensemble-without-impression	0,694	0,796	0,69	0,701	0,818	0,74
impression (-)	0,664	0,796	0,758	0,682	0,823	0,744
StyleGAN2-Augmentation	0,71	0,772	0,676	0,702	0,863	0,745
original-clip (-)	0,67	0,81	0,706	0,727	0,827	0,748
rule-based (+) and inverted	0,723	0,796	0,658	0,743	0,847	0,753
TB-SMGAN (ensemble)	0,672	0,802	0,783	0,667	0,845	0,754

5. CONCLUSION

GANs have demonstrated significant capabilities in representing complex data distributions. By effectively capturing the underlying variance of real-world data, GANs enable the generation of synthetic images with highly discriminative features. This work leverages the generative power of StyleGAN2-ADA to perform data augmentation for medical image datasets. Moreover, we introduce a text-based style manipulated GAN augmentation technique named TB-SMGAN for the medical domain. We utilize DeepAUC, the top solution of the CheXpert competition, to demonstrate the effectiveness of our GAN augmentation technique. Our methodology reveal that the classification performance of TB-SMGAN outperforms the classical GAN augmentation technique due to the increased representation of augmentations created by TB-SMGAN.

While our findings are promising, there are several directions for future research. First, it would be beneficial to explore the application of our proposed framework to other domains beyond medical imaging, to assess its generalizability. Second, developing a consistent measurement methodology for the quality of the dataset could yield valuable insights. Lastly, as text-based manipulated GAN augmentation has shown promise, future work could focus on refining this technique and exploring its potential in other contexts. This could involve the incorporation of more complex textual information or the development of more sophisticated text-to-image translation methods.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, manuscript-review, editing supervision, Mehmet Ulvi ŞİMŞEK; field and laboratory works, sources, software, data curation and visualization Hasan Berat Özfıdan; research, validation, formal analysis, manuscript-original draft and funding, both authors. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering*, 28, 525-552. <https://doi.org/10.1007/s11831-019-09388-y>
- Altwaijry, N. (2023). Probability-based synthetic minority oversampling technique. *IEEE Access*, 11, 28831-28839. <https://doi.org/10.1109/ACCESS.2023.3260723>
- Benčević, M., Habijan, M., Galić, I., & Pizurica, A. (2022, August 29 - September 02). *Self-supervised Learning as a Means to Reduce the Need for Labeled Data in Medical Image Analysis*. In: Proceedings of the 30th European Signal Processing Conference (EUSIPCO 2022) (pp. 1328-1332). Belgrade, Serbia. <https://doi.org/10.23919/EUSIPCO55093.2022.9909542>
- Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., & Rueckert, D. (2018). Gan augmentation: Augmenting training data using generative adversarial networks. <https://doi.org/10.48550/arXiv.1810.10863>
- Dao, H. N., Quang, T. N., & Paik, I. (2022, October 26-28). *Transfer Learning for Medical Image Classification on Multiple Datasets using PubMedCLIP*. In: Proceedings of the 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) (pp. 1-4). Yeosu, Korea, Republic of. <https://doi.org/10.1109/ICCE-Asia57006.2022.9954669>
- Dao, H. N., Nguyen, T., Mugisha, C., & Paik, I. (2023). A Multimodal Transfer Learning Approach Using PubMedCLIP for Medical Image Classification. *IEEE Access*, 12, 75496-75507. <https://doi.org/10.1109/ACCESS.2024.3401777>
- Deepshikha, K., & Naman, A. (2020). Removing Class Imbalance using Polarity-GAN: An Uncertainty Sampling Approach. <https://doi.org/10.48550/arXiv.2012.04937>

- Fetty, L., Bylund, M., Kuess, P., Heilemann, G., Nyholm, T., Georg, D., & Löfstedt, T. (2020). Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Zeitschrift für Medizinische Physik*, 30(4), 305-314. <https://doi.org/10.1016/j.zemedi.2020.05.001>
- Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321, 321-331. <https://doi.org/10.1016/j.neucom.2018.09.013>
- Hochberg, D. C., Greenspan, H., & Giryas, R. (2022). A self supervised StyleGAN for image annotation and classification with extremely limited labels. *IEEE Transactions on Medical Imaging*, 41(12), 3509-3519. <https://doi.org/10.1109/TMI.2022.3187170>
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in python.
- Islam, S. M., & Mondal, H. S. (2019, July 06-08). *Image Enhancement Based Medical Image Analysis*. In: Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). Kanpur, India. <https://doi.org/10.1109/ICCCNT45670.2019.8944910>
- Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghighi, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019, January 27 - February 1). *Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison*. In: Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 590-597). Honolulu, Hawaii, USA. <https://doi.org/10.1609/aaai.v33i01.3301590>
- Jablonski, J. A., Angadi, S. S., Sharma, S., & Brown, D. E. (2022, March 10-11). *Enabling Clinically Relevant and Interpretable Deep Learning Models for Cardiopulmonary Exercise Testing*. In: Proceedings of the 2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT) (pp. 50-53). Houston, TX, USA. <https://doi.org/10.1109/HI-POCT54491.2022.9744068>
- Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., Mark, R. G., & Horng, S. (2019). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*, 6(1), 317. <https://doi.org/10.1038/s41597-019-0322-0>
- Kariuki, P. W., Gikunda, P. K., & Wandeto, J. M. (2023, April 14-16). *Deep Transfer Learning Optimization Techniques for Medical Image Classification: A Review*. In: Proceedings of the 2022 International Conference on Intelligent Computing and Machine Learning (2ICML) (pp. 7-15). Qingdao, China. <https://doi.org/10.1109/2ICML58251.2022.00013>
- Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020, December 6-12). Training generative adversarial networks with limited data. In: H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.) Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20) (pp. 12104-12114). Vancouver BC Canada.
- Ke, B., Lu, H., Huo, W., & Wang, Y. (2022, July 22-24). *Semi-supervised Medical Image Classification Combining Metric Pseudo-Label and Classification Pseudo-Label*. In: P. Lin, & Y. Yang (Eds.) Proceedings of the 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI) (pp. 794-799). Shijiazhuang, China. <https://doi.org/10.1109/ICCEAI55464.2022.00166>
- Kora Venu, S., & Ravula, S. (2020). Evaluation of Deep Convolutional Generative Adversarial Networks for Data Augmentation of Chest X-ray Images. *Future Internet*, 13(1), 8. <https://doi.org/10.3390/fi13010008>
- Lacan, A., Sebag, M., & Hanczar, B. (2023). GAN-based data augmentation for transcriptomics: survey and comparative assessment. *Bioinformatics*, 39(S1), i111-i120. <https://doi.org/10.1093/bioinformatics/btad239>
- Li, Z., Xia, P., Tao, R., Niu, H., & Li, B. (2022). A New Perspective on Stabilizing GANs Training: Direct Adversarial Training. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(1), 178-189. <https://doi.org/10.1109/TETCI.2022.3193373>

- Liu, L., Zhang, Y., & Sun, L. (2023). Medimatrix: innovative pre-training of grayscale images for rheumatoid arthritis diagnosis revolutionises medical image classification. *Health Information Science and Systems*, 11(1), 44. <https://doi.org/10.1007/s13755-023-00246-7>
- Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019, August 1). *ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing*. In: Dina D.-F., Kevin B. C., Sophia A., & J. Tsujii (Eds.) Proceedings of the 18th BioNLP Workshop and Shared Task (pp. 319-327). Florence, Italy. <https://doi.org/10.18653/v1/W19-5034>
- Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., & Lischinski, D. (2021, October 10-17). *StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery*. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 2065-2074). Montreal, QC, Canada. <https://doi.org/10.1109/ICCV48922.2021.00209>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, July 18-24). *Learning transferable visual models from natural language supervision*. In: M. Meila, & T. Zhang (Eds.) Proceedings of the 38th International Conference on Machine Learning, PMLR (pp. 8748-8763). <https://doi.org/10.48550/arXiv.2103.00020>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Sundaram, S., & Hulkund, N. (2021). GAN-based Data Augmentation for Chest X-ray Classification. <https://doi.org/10.48550/arXiv.2107.02970>
- Tarawneh, A. S., Hassanat, A. B., Altarawneh, G. A., & Almuhaimeed, A. (2022). Stop oversampling for class imbalance learning: A review. *IEEE Access*, 10, 47643-47660. <https://doi.org/10.1109/ACCESS.2022.3169512>
- Tov, O., Alaluf, Y., Nitzan, Y., Patashnik, O., & Cohen-Or, D. (2021). Designing an encoder for StyleGAN image manipulation. *ACM Transactions on Graphics (TOG)*, 40(4), 133. <https://doi.org/10.1145/3450626.3459838>
- Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., & Yao, Y. (2021). Deep Learning in Medical Ultrasound Image Analysis: A Review. *IEEE Access*, 9, 54310-54324. <https://doi.org/10.1109/ACCESS.2021.3071301>
- Wang, X., & Qi, G.-J. (2022). Contrastive Learning With Stronger Augmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5), 5549-5560. <https://doi.org/10.1109/TPAMI.2022.3203630>
- Yuan, Z., Yan, Y., Sonka, M., & Yang, T. (2021, October 10-17). *Large-scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification*. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (pp. 3040-3049). Montreal, QC, Canada. <https://doi.org/10.1109/ICCV48922.2021.00303>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1514851>

Development of a P(L-D,L)LA Foam as a Dura Substitute and Its *In Vitro* Evaluation

Deniz YÜCEL^{1,2*} ¹ Acıbadem Mehmet Ali Aydınlar University (ACU), School of Medicine, Department of Histology and Embryology, İstanbul, Türkiye² Acıbadem Mehmet Ali Aydınlar University (ACU), ACU Biomaterials Center, İstanbul, Türkiye

Keywords	Abstract
Dura Substitute Synthetic Polymer Porous Foam Meningeal Cells Cell Ingrowth	Dura substitutes are used to reduce the risk of postoperative complications following neurosurgical interventions, and to facilitate the healing of dura damages or defects caused by injuries. Traditional tissue transplants have limitations like limited tissue availability, potential risk of immune rejection and disease transmission. The use of biomaterials composed of synthetic polymers as dura substitutes offers a promising approach to overcome these limitations to replace and treat damaged dura mater. Potential biocompatible porous scaffolds still need to be developed to minimize the risks of immune response and disease transmission, while also ensuring effective cell migration and cell ingrowth in three dimension. The aim of the present study was to develop a poly(L-lactide-co-D,L-lactide) (P(L-D,L)LA) foam with an optimal pore size for dura mater substitution, investigate its morphological characteristics, and evaluate its potential for dura mater regeneration by assessing the spreading and growth of meningeal cells within it through <i>in vitro</i> studies. Foams were produced by lyophilization using different concentrations of P(L-D,L)LA solution. A GMP-grade P(L-D,L)LA, suitable for medical device applications, was used in this study. Morphological analysis was performed using scanning electron microscopy, and porosity of the foams was studied with mercury porosimetry. In <i>in vitro</i> studies, meningeal cells were seeded onto the polymeric foams, and their behavior and proliferation in these scaffolds were investigated with cytoskeleton and nucleus staining, and colorimetric cell proliferation assay, respectively. Scanning electron microscopy results showed that the foams prepared with 2.5% and 3% polymer solutions displayed good structural integrity and convenient interconnectivity, with pore sizes ranging from 80 to 150 μm . However, the foams prepared with 2% and 4% polymer solution demonstrated poor structural integrity and low interconnectivity, respectively. <i>In vitro</i> studies showed that the foams prepared with 2.5% and 3% polymer solutions served effectively as scaffolds for meningeal cells, and the cells attached, spread and homogeneously distributed. In addition, the cells proliferated and increased in number over time within these polymeric scaffolds. These findings suggest that the foams produced with 2.5% and especially 3% P(L-D,L)LA polymer solutions could effectively serve as a suitable substitute for the dura mater, providing an appropriate environment for cell ingrowth and tissue integration. This indicates that the developed foam could be a promising treatment for dura mater damage or defects, with the potential approach to promote regeneration in future <i>in vivo</i> and clinical studies.

Cite

Yucel, D. (2024). Development of a P(L-D,L)LA Foam as a Dura Substitute and Its In Vitro Evaluation. *GU J Sci, Part A, 11(3)*, 507-517. doi:10.54287/guj.1514851

Author ID (ORCID Number)	Article Process
0000-0002-1373-5183 Deniz YÜCEL	Submission Date 11.07.2024 Revision Date 28.08.2024 Accepted Date 16.09.2024 Published Date 28.09.2024

1. INTRODUCTION

The meninges are connective tissue layers covering brain and spinal cord, and consist of pia, arachnoid and dura mater from the innermost to the outermost layer (Patel & Kirmi, 2009). The meninges are important in protecting the central nervous system (CNS) by serving as a structural support and facilitating the circulation of cerebrospinal fluid (CSF). The pia mater, thin and highly vascular connective tissue layer, surrounds the

*Corresponding Author, e-mail: deniz.yucel@acibadem.edu.tr

nervous tissue of brain and spinal cord, providing nourishment to these tissues. The arachnoid mater, comprised of a delicate sheet of connective tissue and extending trabeculae to the pia mater, forms an intricate network structure. Serving as a barrier between pia and dura maters, it establishes the subarachnoid space, housing cerebral arteries and cerebrospinal fluid (CSF). Trabeculae consist of loose connective tissue with elongated fibroblasts. The outermost layer, dura mater is a relatively thick layer of dense connective tissue continuing at the outer surface with the periosteum of the skull and provides a protective barrier against physical impacts. The dura mater contains a periosteal dura layer that attaches to the inner side of the skull and serves as the periosteum. This outer layer of dura mater consists of organized bundles of collagen fibers with fibroblasts and osteoblasts. The innermost layer, meningeal dura mater, is composed of sheet-like layers containing fine collagen fibers with fibroblasts.

After surgeries involving the brain or spinal cord, such as tumor resection, discectomy/artificial disc replacement, laminectomy, or minimally invasive surgeries, the dura mater needs to be repaired or replaced (Dong et al., 2023). In addition, there may be defects or tears in the dura mater due to injury by trauma like skull fracture and spinal burst fracture, infection, or congenital conditions. When a dural defect is too large for direct repair with suturing, the use of dura grafts to patch or close the defect through a duraplasty procedure is a potential treatment strategy to repair or reinforce the damaged dura mater and prevents CSF leakage. Dura grafts and dura substitutes serve to restore the integrity of the dura mater, to prevent complications such as infections or CSF leaks, and provide protection to the underlying neural tissues. There are traditional tissue transplants like autografts, allografts, and xenografts for the treatment of the dural defects (Khurana et al., 2024). Despite the successful results of autologous grafts with a low risk of immune rejection, they have inherent disadvantages such as limited tissue supply and donor site morbidity. Even though allografts and xenografts are available, these grafts may carry a high risk of immune rejection and a potential risk of disease transmission (Shijo et al., 2017). An alternative treatment approach is the use of biomaterials as a dura substitute to repair dural defects.

The biomaterial used as a dura substitute should be biocompatible, provide a scaffold for cell adhesion and tissue ingrowth, and promote the regeneration of dural tissue. Even though non-biodegradable materials like silicone have been used for dural repair, they are more prone to foreign-body reactions for a long time and may provoke significant complications (Ohbayashi et al., 1994). However, biodegradable dura substitutes are advantageous because they provide essential support during the healing of the dural defect, gradually dissolve and are replaced by newly formed tissue, and their complete degradation eliminates the possible risk of long-term foreign-body reaction (Wang & Ao, 2019, Cho et al., 2024). Biodegradable biomaterials can be natural polymers like collagen, or synthetic polymers like polyesters, or a combination of both (Shi et al., 2016; Deng K. et al., 2017; Deng W. et al, 2021; Liu et al., 2021). Natural polymer-based dural substitutes consist of mostly collagen (Mai et al., 2024) as well as bacterial cellulose (Deng W. et al., 2021). However, natural polymers have some drawbacks like potential risk of disease transmission associated with some animal-derived polymers, limited control over *in vivo* degradation and poor mechanical properties (Khurana et al., 2024). Synthetic polymers with their tunable properties and lack of disease transmission risk may address these issues associated with the natural polymers (Shi et al., 2016). In addition, synthetic biodegradable materials offer advantages such as being available in larger quantities, eliminating the need for a secondary surgery, allowing for the manufacturing of desired dimensions and properties, and ensuring consistent quality. Biodegradable synthetic polymers like the most commonly used polyesters poly(lactic acid) (PLA), poly(glycolic acid) (PGA), and poly(ϵ -caprolactone) (PCL) are biocompatible and tunable according to demand, which makes them preferable in medical applications (Deng K. et al., 2017). Poly-L-lactide (PLLA) has been commonly used to develop dura substitutes (Shi et al., 2016; Klopp et al., 2004; Wang Y.-f. et al., 2013; Deng K. et al., 2017). Studies have demonstrated its full compatibility with dural tissue, its integration with the surrounding tissue, its ability to reduce tissue adhesion and prevent CSF leakage (Klopp et al., 2004; Wang Y.-f. et al., 2013). Poly(L-lactide-co-D,L-lactide) (P(L-D,L)LA), a biocompatible polymer composed of poly-L-lactide and racemic DL-lactide, has been used in clinical applications in ratios of 70:30, 80:20, 85:15, and 96:4 (Schachtner et al., 2019). The commonly utilized P(L-D,L)LA (70:30) displays degradation within 1-3 years, decreasing the potential for adverse tissue reactions.

Dural substitutes can be in various forms, including film, mesh, membrane and sponge/foam, or combination of these. The film consisting of P(L-D,L)LA (70:30) polymer, with its hydrophobic nature, was used only as

a barrier to reduce the peridural tissue adhesion (Welch et al., 2002; Klopp et al., 2004). There are also electrospun fibrous meshes consisting of PLLA (Shi et al., 2016; Liao et al., 2021; Cho et al., 2024), PLLA-gelatin (Deng K. et al., 2017), and PLA-PCL-collagen (Wang Y.-f. et al., 2013) to be used for dural repair. Electrospun fibrous meshes have porosity that depends on fiber diameter and process parameters. However, due to the sheet-like assembly process of electrospinning, adequate porosity is typically achieved on the surface, while porosity in the bulk region is often limited (Guimarães et al., 2010). The tightly packed fibrous layers may restrict cell infiltration into the scaffold and inhibit the growth of 3D tissues. Porous scaffolds provide voids that facilitate the penetration of the patient's autologous cells and promote tissue renewal (Deng K. et al., 2017). The porosity within the entire scaffold is crucial, as it enables cell migration and tissue growth to ensure the degradation of the scaffold and the reformation of the dura mater. It was reported that by the fourth day after surgery, fibroblasts and histocytes started moving towards the implanted dural substitute and entered the implant through pores of the scaffold (Laun et al., 1990).

In this study, the aim was to develop a biodegradable foam with optimum pore size as a dura substitute and to assess this biomaterial's potential effectiveness in promoting dura regeneration by investigating its suitability for meningeal cell growth through *in vitro* studies. A GMP grade polymer of P(L-D,L)LA (70:30) was selected for its biocompatibility, biodegradability, and its ability to be tailored to achieve the desired scaffold morphology. In this study, foam/sponge of P(L-D,L)LA (70:30) was developed as a dura substitute to provide the necessary porosity for cell infiltration and proliferation considering the ultimate aim of enhancing the regeneration and healing process of the dura mater in the future clinical use. Dura scaffolds were produced in sponge form using different concentrations of P(L-D,L)LA. Their porous structures were shown morphologically, and pore size distribution was analyzed with mercury porosimetry. The meningeal cell integration, growth and morphology of cells in the 3D foams were demonstrated by colorimetric cell proliferation assay and Phalloidin-DAPI (2-(4-Amidinophenyl)-6-indolecarbamide) staining. This 3D porous polymeric scaffold could be a promising dura substitute to be used in the treatment of dural defects.

2. MATERIAL AND METHOD

2.1. Preparation of Polymeric Foams

2.1.1. Production of Polymeric Foams

3D foams were produced using a biodegradable, synthetic polymer, P(L-D,L)LA (70:30) (GMP grade, Mw. ca. 1.500.000 g/mol, PURAC, Corbion, Netherland). The stock polymer solution (4%, w/v%) was prepared by dissolving P(L-D,L)LA in 1,4-dioxane (ACS reagent, $\geq 99.8\%$) by occasionally mixing for 24 h (Kenar et al., 2010; Choi et al., 2010). Afterwards, this stock polymer solution was diluted with 1,4-dioxane to prepare 2%, 2.5% and 3% working polymer solutions. Different concentrations of P(L-D,L)LA solutions were poured into petri plates, and kept at -20°C overnight. Frozen samples were completely dried with lyophilization, and 3D porous foams were obtained.

2.1.2. Characterization of Polymeric Foams

The morphology and the porosity of the foams' surface and bulk were investigated with scanning electron microscope (SEM). The polymeric foams were coated with Au under vacuum, and then examined using SEM. In addition to surface analysis, a set of samples was cut before coating, and cross-section images were taken to assess the porosity inside the sponge. The porosity of the foams was quantitatively assessed by measuring pore area with Image J software using SEM images of three different regions. In addition, the distribution of pore sizes of the foams was analyzed using mercury porosimetry under low vacuum.

2.2. Culture of Meningeal Cells

Human meningeal cells (cell line, HMC, ScienCell Research Laboratories) were seeded on fibronectin ($5\mu\text{g}/\text{cm}^2$) coated tissue culture plates, and cultured with a growth medium composed of α -MEM: F12 (1:1) including 10% FBS and Pen/Strep (100 U/mL, 100 $\mu\text{g}/\text{mL}$) in an incubator (5% CO_2 , 37°C) by refreshing medium every three days. Once the cell confluency was approximately 90%, they were detached from the tissue culture plates by treatment of trypsin/EDTA. The cells were centrifuged at 1000 rpm, and then

subcultured by transferring into new fibronectin coated tissue culture plates, and kept in the growth medium containing 10% DMSO in the vapor of liquid nitrogen until use.

2.3. Investigation of Meningeal Cell Behavior in the Polymeric Foams

2.3.1. Seeding and Culture of Cells in the Foams

Before cell seeding, the foams prepared with two different polymer concentrations (2.5% and 3%) were kept in 70% ethanol at 4°C for 2 h for sterilization. After removing the ethanol and air drying of the foams, the foams were coated with fibronectin (50 µg/mL, 100 µL/foam) for 2h at 37°C to promote the attachment of meningeal cells. The foams were ready for cell seeding after drying slightly to a wet state. Meningeal cells were collected by trypsinization from tissue culture plates, centrifuged at 1000 rpm, and then resuspended in the growth medium. After counting with a hemocytometer, the cells at a density of 5×10^4 cells/foam were seeded onto the foams and cultured with the growth medium in a humidified 5% CO₂ incubator at 37°C by refreshing medium every three days.

2.3.2. Distribution and Morphology of the Meningeal Cells on/in the Foams

In order to visualize the meningeal cells in the foams, fixation of the cells on day 14 of the culture was done with 4% paraformaldehyde for 1h at room temperature (RT). The morphology of the cells was examined by staining with fluorescein isothiocyanate (FITC) conjugated Phalloidin and 2-(4-Amidinophenyl)-6-indolecarbamidine (DAPI) for cytoskeletal actin filaments and nucleus, respectively (Yucel et al., 2010). The cells were kept in 0.1% Triton X-100 solution for 5 min at RT to permeabilize cell membrane and allow dye penetration. After PBS wash, the samples were incubated in 1% bovine serum albumin (BSA) for 30 min at 37°C to prevent non-specific binding. Subsequently, the samples were incubated in FITC-Phalloidin solution (1:100 diluted with 0.1% BSA) for 1 h at 37°C. Following another wash with PBS, the nuclei of cells were stained with DAPI (1:5000, diluted with PBS) for 10 min at RT. After a final wash with PBS, the cells in the foams were examined and viewed with a Laser Scanning Confocal Microscope (LSCM, Zeiss LSM700) using 488 nm and 405 nm lasers. Optical sections were taken in z-direction, and the images were obtained using the ZenPlus program to investigate the distribution and morphology of the meningeal cells in the 3D foams.

2.3.3. Proliferation of the Cells in the Foams

In order to determine the meningeal cell survival and growth in the foams, the cell-seeded foams were transferred into new 24 well plates at the 1st, 7th, 14th and 21st days of the culture, and a colorimetric cell proliferation assay (MTS) was applied to the samples which were studied in triplicates (Yucel et al., 2010). A 10% solution of MTS-PMS prepared in DMEM low glucose including 10% FBS and Pen/Strep (100 U/mL, 100 µg/mL) was added to the cell-seeded foam samples, and the cells were incubated in an incubator (5% CO₂, 37°C) for 2 h. At the end of incubation, 200 µL of the reaction solution from each sample was placed into 96 well-plate, and the absorbance value was measured at 490 nm using Elisa-Plate Reader. The cell numbers were determined using the slope of the calibration graph ($y=0.0201x$). The calibration graph was constructed by measuring the absorbance of known, specific cell seeding numbers with an ELISA plate reader. The slope of the graph was used to calculate the cell numbers based on the absorbance values measured from the samples.

2.4. Statistical Analyses

GraphPad Prism 10.0 software was used for statistical analyses. The sample data was confirmed by Shapiro-Wilk test to check if the data was normally distributed. A one-way analysis of variance (ANOVA) was used to compare mean cell numbers by day in both groups, while a two-way ANOVA was used for the comparison between groups. A minimum confidence level of 95% was set, and p-values smaller than or equal to 0.05 were considered statistically significant. All values were reported as the mean ± standard deviation.

3. RESULTS AND DISCUSSION

3.1. Morphology and Porosity Analysis of Polymeric Foams

Biomaterials should be designed to address the target tissue architecture, appropriate porosity with interconnected pores, and a tissue-specific 3D shape. Therefore, foam structures in a circular shape with a

diameter of 9 mm were constructed as a dura substitute in this study for characterization and *in vitro* studies. However, the shape and the dimensions of the foams could be altered according to the specific application. The surface and bulk morphology of the foams were examined with SEM (Figure 1). It was observed that the foams with a thickness of 0.93 ± 0.32 mm, prepared using a 2% P(L-D,L)LA polymer concentration, exhibited compromised structural integrity compared to those prepared with higher polymer concentrations (Figure 1a and 1e). In addition, the pore sizes were around 300 μm , and there was a significant variation in pore size distribution. This less robust foam network can limit its handling and performance, particularly in clinical applications that require consistent and durable materials. SEM results showed that the foams with thicknesses of 1.74 ± 0.11 mm and 1.92 ± 0.15 mm, produced using polymer concentrations of 2.5% and 3% P(L-D,L)LA, respectively, exhibited high structural integrity, and their pore sizes ranged between 80-150 μm (Figure 1b and 1c, Figure 2a and 2c). In addition, it was seen in the cross-section views that this porosity continued throughout the interior of the foams, and there were interconnected pores (Figure 1f and 1g). In the foams prepared with a polymer concentration of 4% P(L-D,L)LA and a thickness of 2.48 ± 0.22 mm, the structural integrity was good, and the pore sizes were in the range of 80-100 μm (Figure 1d and 1h). However, upon examining cross-sectional images, it was noticed that some pores were closed and the interconnectivity of pores in the foam prepared with a polymer concentration of 4% was poor. Therefore, it was thought that this structure may limit the cell ingrowth and provide a less favorable environment for dura regeneration and tissue integration compared to other foams. Consequently, as expected, increasing the polymer concentration to levels such as 2.5% and 3% provided better structural integrity and led to a reduction in pore size. However, further increase in polymer concentration to 4% decreased the interconnectivity of pores. Therefore, it was decided to continue the further analysis with foams prepared with 2.5% and 3% P(L-D,L)LA solution. Using the surface and cross-section SEM images, the porosities of the foams' surface and bulk were determined by calculating the percentage of pore area (Table 1). The foams produced with polymer concentrations of 2.5% and 3% P(L-D,L)LA exhibited similar porosity levels both at the surface (ca. 58%) and in the bulk (ca. 48%). It was revealed that scaffolds with a porosity of 46.9% promoted the growth of osteocytes (Wang M. O. et al., 2015). It was also reported in the literature that fibroblast could even infiltrate scaffolds with an average porosity of 34.4% and mean pore size of 11 μm (Rnjak-Kovacina et al., 2011). Therefore, it was thought that the porosity of the dura substitute developed in this study would provide a suitable environment for osteocyte ingrowth and the formation of the periosteal dura layer, as well as for the ingrowth of fibroblast-like meningeal cells that would regenerate the meningeal dura mater.

In Figure 2b and 2d, mercury porosimetry results show the distribution of the pore size throughout the foams produced with the polymer solutions of 2.5% and 3% P(L-D,L)LA. It was observed that the pore sizes varied in the range of 20-130 μm and 20-200 μm in the foams of 2.5% and 3% P(L-D,L)LA, respectively. The pore sizes ranging from 80 to 150 μm seen in the SEM images were in the range of pore sizes determined with mercury porosimetry, thus demonstrating consistency in these results. A porous scaffold made of chitin and bacterial cellulose was developed as a dura graft with pore sizes ranging from 90 to 200 μm and high pore connectivity, similar to the pore size range obtained in this study (Deng W. et al., 2021). In fact, the range determined in the present study is narrower, indicating that more homogeneous pore sizes were obtained. In addition, it was reported that a collagen dura substitute with interconnected pores ranging from a few micrometers to 150 μm was suitable for cell infiltration, proliferation, and the exchange of nutrients and waste (Liu et al., 2021). It was also revealed that endogenous cells would infiltrate and proliferate within the porous scaffold and gradually reconstruct the extracellular matrix to regenerate tissue. In another study, it was observed that as the porous scaffold degraded, the newly formed tissue replaced it, ultimately leading to the repair and regeneration of the dura mater (Ramot et al., 2024).

Consequently, considering their high structural integrity, appropriate pore size, and interconnectivity, the foams produced with 2.5% and 3% P(L-D,L)LA were utilized in the further studies to investigate the behaviour of meningeal cell on these constructs.

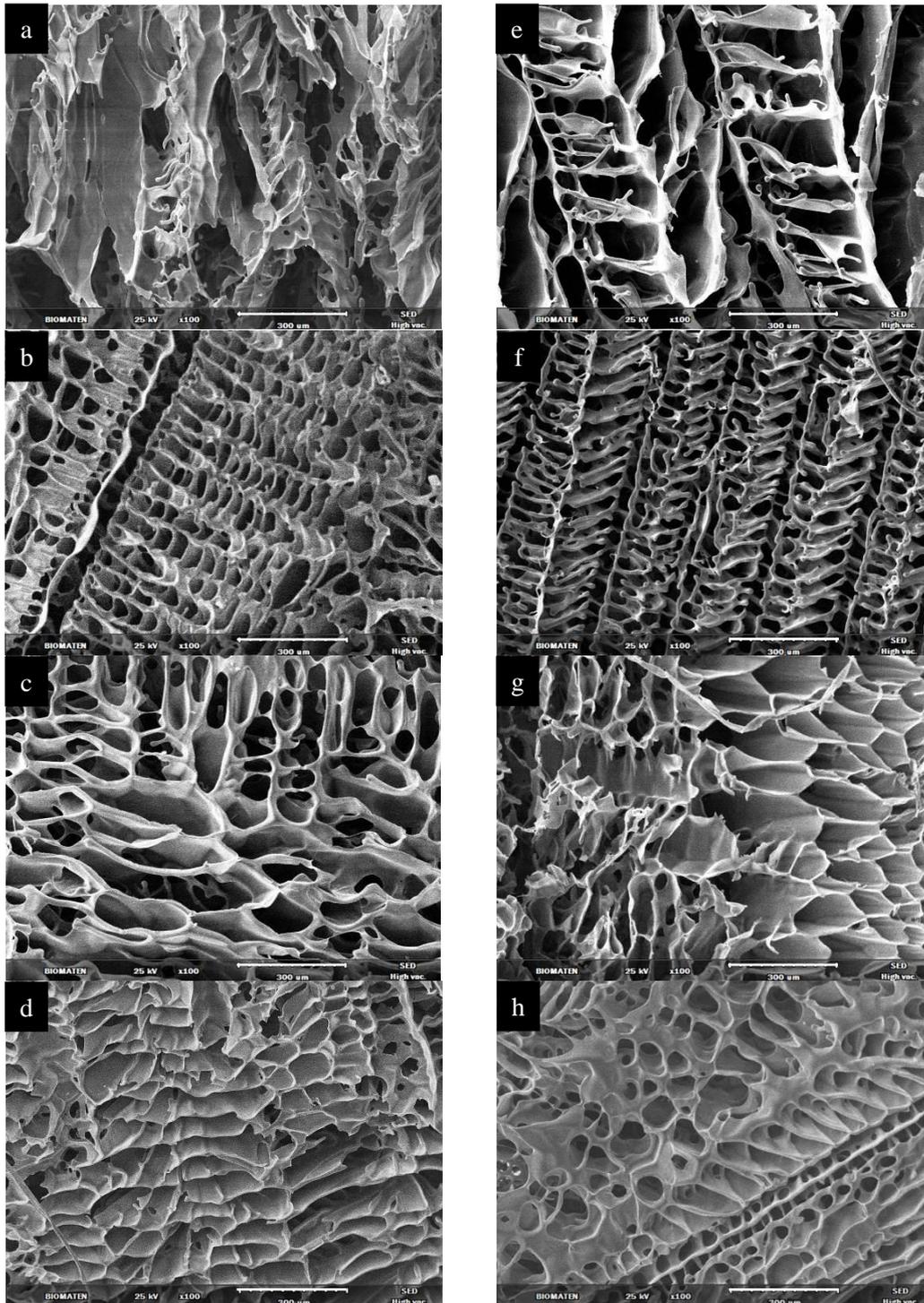


Figure 1. Scanning electron micrographs of the polymeric foams. The foams prepared using the polymer solutions of **a, e**) 2%, **b, f**) 2.5%, **c, g**) 3% and **d, h**) 4% P(L-D,L)LA. The images of **a-d**) the surface and **e-h**) cross-section of the foams. Magnification: 100X, scale bar: 300µm.

Table 1. Porosity analysis of the foams from the top and cross-section views.

	Foam with 2.5% P(L-D,L)LA	Foam with 3% P(L-D,L)LA
Surface	57.04 ± 0.58	58.41 ± 0.13
Cross-Section	48.16 ± 0.28	48.26 ± 1.63

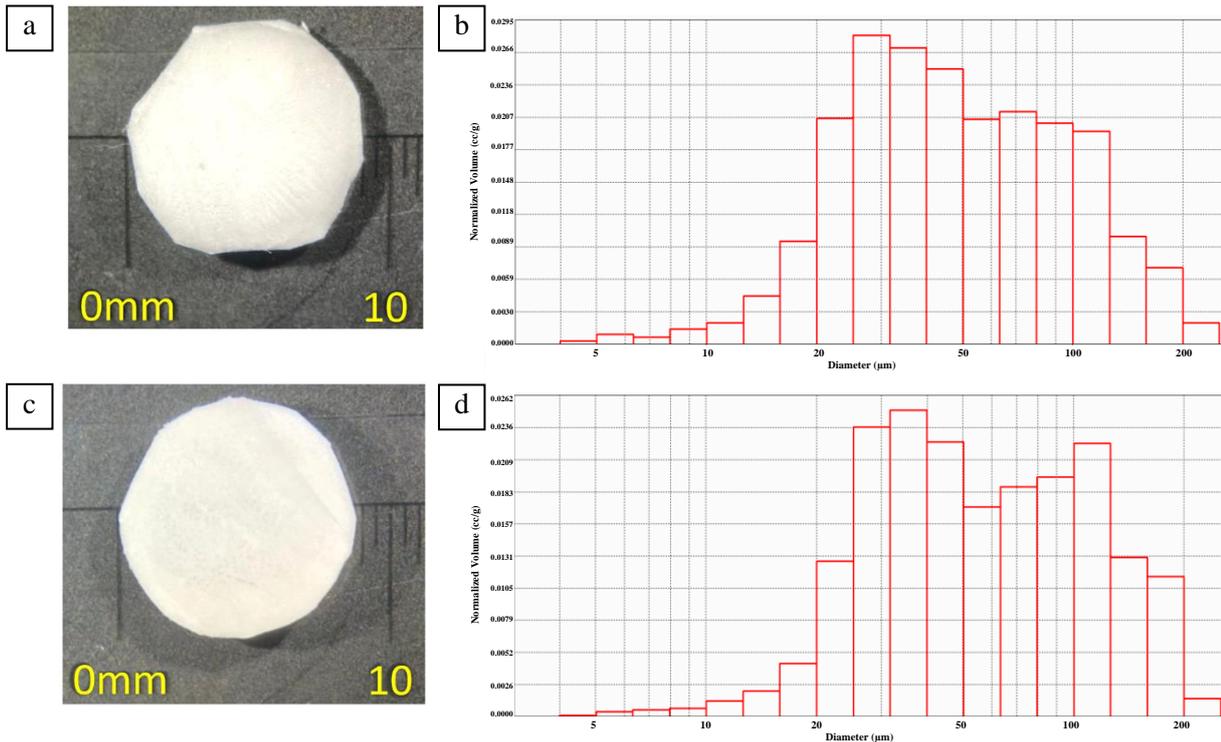


Figure 2. Stereomicrographs of foams prepared with **a)** 2.5% and **c)** 3% P(L-D,L)LA polymer solutions. Distribution of pore size of the polymeric foams prepared using the polymer solutions of **b)** 2.5% and **d)** 3% P(L-D,L)LA with normalized volume vs pore size histogram of the foams.

3.2. Culture of Meningeal Cells

Human meningeal cells cultured with the growth medium grew and increased in number on the fibronectin coated tissue culture plate (Figure 3a). Meningeal cells without any treatment could not properly adhere and spread on untreated P(L-D,L)LA due to its hydrophobic nature. Therefore, the foams were coated with fibronectin which takes role in cell adhesion, migration and growth (Mekhail et al., 2012). It was seen that the meningeal cells spread properly on the culture plate and their actin filaments were well-organized (Figure 3b). The scanning electron micrograph in Figure 3c showed the human meningeal cell interaction with the foam surface. The cells adhered well to the surface, formed filopodia, and expanded on the polymeric foam.

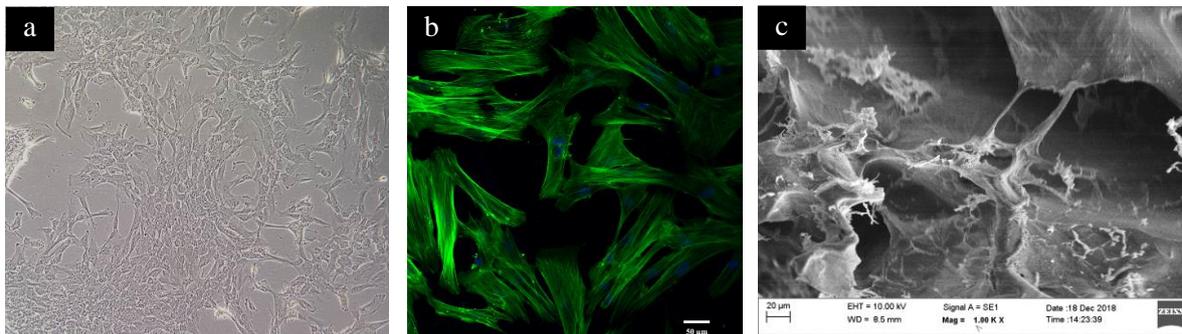


Figure 3. Morphology of the human meningeal cells. Images of the cells on the tissue culture plate **a)** with bright field microscope and **b)** with fluorescence microscope after Phalloidin-DAPI staining for actin filaments (green) and nucleus (blue). **c)** SEM image of the meningeal cells on the polymeric foam.

3.3. Meningeal Cell Behavior in the Polymeric Foams

3.3.1. Distribution and Morphology of the Meningeal Cells on/in the Foams

The distribution and morphology of human meningeal cells on/in the foams were shown by their actin organization after Phalloidin-DAPI staining (Figure 4). It was seen that meningeal cells adhered and spread on

both foams produced with polymer solutions of 2.5% and 3% P(L-D,L)LA. Even the polymer preparation concentration did not change the cell behavior in terms of cell adherence and spreading, it seems that more cells were found on/in the foams of 3% P(L-D,L)LA compared to the foams of 2.5% P(L-D,L)LA. The high magnified images (Figure 4b and 4e) showed that cells spread along the pore surfaces of the foams. The analysis of cells with LSCM displayed that the cells were homogeneously distributed throughout the foam along z-direction (Figure 4c and 4f). The cells were not limited to the surface of the foams, they were also able to penetrate through the pores of the foam. *In vitro* studies revealed that the porosity and pore size of the developed dura substitute ensure uniform distribution of cells in 3D.

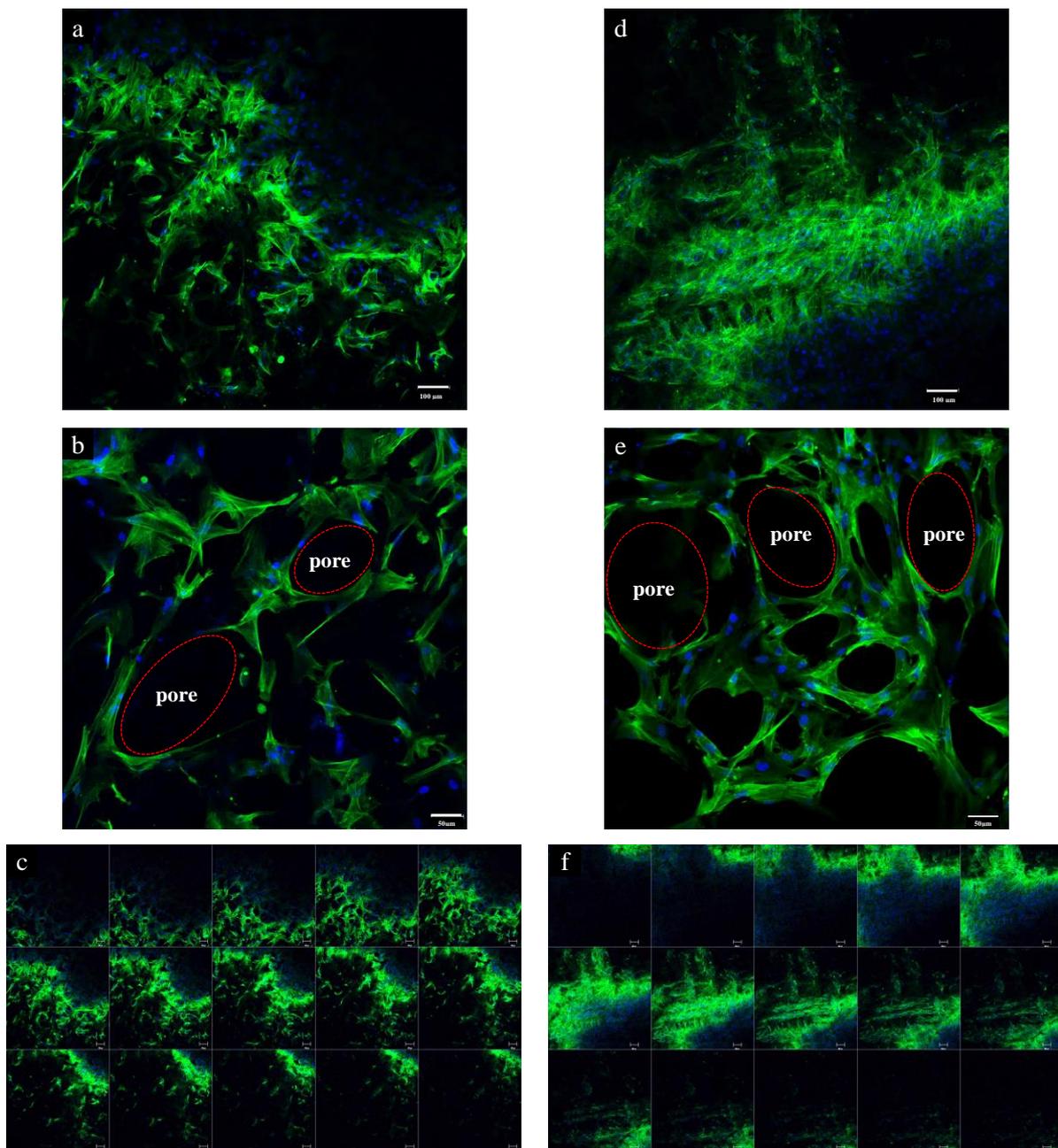


Figure 4. LSCM images of the human meningeal cells in the polymeric foams prepared with **a-c**) 2.5% and **d-f**) 3% P(L-D,L)LA solution. **c, f**) Images of optical sections through the samples in z-direction. The cells were stained with FITC-Phalloidin (green) for actin filaments and counterstained with DAPI (blue) for nucleus. Magnification, scale bar: **a, d, c, f**) 10X, 100 μ m, **b, e**) 20X, 50 μ m.

3.3.2. Proliferation of the Cells in the Foams

MTS assay results showed that meningeal cells exhibited similar growth profiles on both foams produced with polymer solutions of 2.5% and 3% P(L-D,L)LA (Figure 5). The results of Day 1 showed that the cells attached to both foams, and in fact, more cells were able to attach to the 3% P(L-D,L)LA foams. The number of cells on both foams increased significantly by at least 2-fold within 7 days, and this significant increase continued linearly, reaching 3- to 4-fold by Day 14. The rate of cell proliferation decreased after Day 14 in both foams, which could be due to cell-to-cell contact inhibition as the cell number increased. On Day 14, the cell number in the foam of 3% P(L-D,L)LA was significantly higher compared to the foam of 2.5% P(L-D,L)LA, confirming the Phalloidin-DAPI staining results.

In tissue engineering, it is suggested that a scaffold should have pores of sufficient size and quantity, typically ranging from 50 to 200 μm , to facilitate cell infiltration and proliferation (Li et al., 2017). In this study, it was observed that the meningeal cells, fibroblast-like cells, penetrated into the foams, survived, grew, and spread in the foams with pore sizes ranging from 80 to 150 μm . It was reported that the regeneration of dural defects is primarily driven by the development of connective tissue fibers and fibroblastic proliferation (Khurana et al., 2024). Consequently, these results suggest that the foams produced with 2.5% and especially 3% P(L-D,L)LA polymer solutions would serve as an appropriate dura substitute to provide cell ingrowth and tissue integration, and thus it would be a potential dura substitute to promote regeneration of the dura mater.

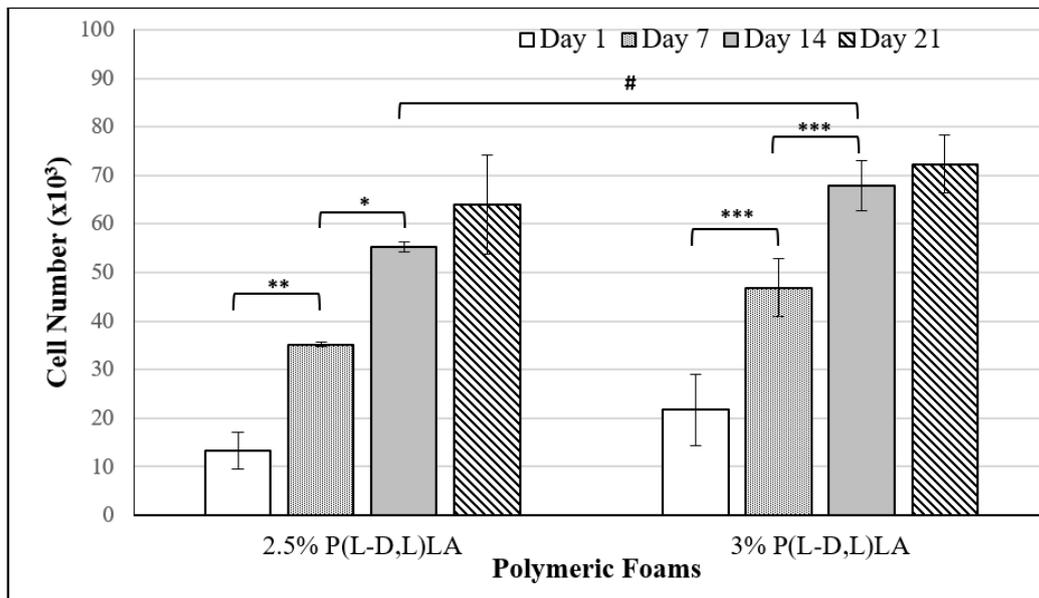


Figure 5. Proliferation of the human meningeal cells in the polymeric foams prepared with 2.5% and 3% P(L-D,L)LA solutions. The cell number in the foams at the 1st, 7th, 14th and 21st days of the culture. Statistical differences (# $p \leq 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$) are indicated.

4. CONCLUSION

Significant progress has been made in developing dural substitutes for repairing the dura mater. The studies in biomaterials and regenerative medicine aim to improve the performance and safety of these grafts, suggesting them as potential alternative treatments for future clinical applications. There are still some shortcomings with current dura substitutes, such as achieving ideal tissue compatibility and serving as a suitable scaffold to promote cell migration and tissue regeneration for structural and functional repair. The synthetic polymer P(L-D,L)LA (70:30) used in this study, as being GMP grade is appropriate to be used in medical device applications and could be suggested for the clinical studies after further *in vivo* performance analysis. Porosity is a critical factor for successful dural repair and regeneration, as it provides a favorable environment for cell integration and proliferation. In this study as well, it was found that the obtained foams with pore sizes in the range of 80 to 150 μm allowed the spreading and growth of meningeal cells. This suggests that the dura substitutes

produced with 2.5% and especially 3% P(L-D,L)LA polymer solutions could provide a required, suitable environment for cell ingrowth and promote dura mater regeneration.

ACKNOWLEDGEMENT

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) with the research project number SBAG 118S587. The author acknowledges the contributions of Acıbadem Mehmet Ali Aydınlar University (ACU) and Middle East Technical University (METU) Center of Excellence in Biomaterials and Tissue Engineering (BIOMATEN) for the use of the facilities and equipment, and METU Central Laboratory for mercury porosimetry analysis.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- Cho, M., Shim, K. M., Park, S. S., Kang, S. S., Jang, K., & Kim, S. E. (2024). Evaluation of Biocompatibility and Healing Properties of Dural Substitutes Produced by Electrospinning Technology. *In Vivo*, 38(3), 1119-1126. <https://doi.org/10.21873/invivo.13546>
- Choi, S.-W., Zhang, Y., & Xia, Y. (2010) Three-dimensional scaffolds for tissue engineering: the importance of uniformity in pore size and structure. *Langmuir*, 26(24), 19001-19006. <https://doi.org/10.1021/la104206h>
- Deng, K., Yang, Y., Ke, Y., Luo, C., Liu, M., Deng, Y., Tian, Q., Yuan, Y., Yuan, T., & Xu, T. (2017). A novel biomimetic composite substitute of PLLA/gelatin nanofiber membrane for dura repairing. *Neurological Research*, 39(9), 819-829. <https://doi.org/10.1080/01616412.2017.1348680>
- Deng, W., Tan, Y., Riaz Rajoka, M. S., Xue, Q., Zhao, L., & Wu, Y. (2021). A new type of bilayer dural substitute candidate made up of modified chitin and bacterial cellulose. *Carbohydrate Polymers*, 256, 117577. <http://www.doi.org/10.1016/j.carbpol.2020.117577>
- Dong, R.-P., Zhang, Q., Yang, L.-L., Cheng, X.-L., & Zhao, J.-W. (2023). Clinical management of dural defects: A review. *World Journal of Clinical Cases*, 11(13), 2903-2915. <http://www.doi.org/10.12998/wjcc.v11.i13.2903>
- Guimarães, A., Martins, A., Pinho, E. D., Faria, S., Reis, R. L., & Neves, N. M. (2010). Solving cell infiltration limitations of electrospun nanofiber meshes for tissue engineering applications. *Nanomedicine (Lond)*, 5(4), 539-554. <https://doi.org/10.2217/nmm.10.31>
- Kenar, H., Kose, G. T., & Hasirci, V. (2010). Design of a 3D aligned myocardial tissue construct from biodegradable polyesters. *Journal of Materials Science: Materials in Medicine*, 21(3), 989-997. <https://doi.org/10.1007/s10856-009-3917-8>
- Khurana, D., Suresh, A., Nayak, R., Shetty, M., Sarda, R. K., Knowles, J. C., Kim, H.-W., Singh, R. K., & Singh, B. N. (2024). Biosubstitutes for dural closure: Unveiling research, application, and future prospects of dura mater alternatives. *Journal of Tissue Engineering*, 15, 1-26. <http://www.doi.org/10.1177/20417314241228118>
- Klopp, L. S., Welch, W. C., Tai, J. W., Toth, J. M., Cornwall, G. B., & Turner, A. S. (2004). Use of polylactide resorbable film as a barrier to postoperative peridural adhesion in an ovine dorsal laminectomy model. *Neurosurgical Focus*, 16(3), 1-9. <http://www.doi.org/10.3171/foc.2004.16.3.3>
- Laun, A., Tonn, J. C., & Jerusalem, C. (1990). Comparative study of lyophilized human Dura mater and lyophilized bovine pericardium as dural substitutes in neurosurgery. *Acta Neurochirurgica (Wien)*, 107(1-2), 16-21. <http://www.doi.org/10.1007/BF01402607>
- Li, Q., Mu, L., Zhang, F., Sun, Y., Chen, Q., Xie, C., & Wang, H. (2017). A novel fish collagen scaffold as dural substitute. *Materials Science and Engineering C Materials for Biological Applications*, 80, 346-351. <http://www.doi.org/10.1016/j.msec.2017.05.102>

- Liao, J., Li, X., He, W., Guo, Q., & Fan, Y. (2021). A biomimetic triple-layered biocomposite with effective multifunction for dura repair. *Acta Biomaterialia*, 130, 248-267. <https://doi.org/10.1016/j.actbio.2021.06.003>
- Liu, W., Wang, X., Su, J., Jiang, Q., Wang, J., Xu, Y., Zheng, Y., Zhong, Z., & Lin, H. (2021). In vivo Evaluation of Fibrous Collagen Dura Substitutes. *Frontiers in Bioengineering and Biotechnology*, 9, 628129. <http://www.doi.org/10.3389/fbioe.2021.628129>
- Mai, R., Osidak, E., Mishina, E., Domogatsky, S., Andreev, A., Dergam, Y., & Popov, V. (2024). Collagen Membrane as Artificial Dura Substitute: A Comprehensive In Vivo Study of Efficiency and Substitution Compared to Durepair. *World Neurosurgery*, In Press. <https://doi.org/10.1016/j.wneu.2024.08.061>
- Mekhail, M., Almazan, G., & Tabrizian, M. (2012). Oligodendrocyte-protection and remyelination post-spinal cord injuries: a review. *Progress in Neurobiology*, 96(3), 322-239. <https://doi.org/10.1016/j.pneurobio.2012.01.008>
- Ohbayashi, N., Inagawa, T., Katoh, Y., Kumano, K., Nagasako, R., & Hada, H. (1994). Complication of silastic dural substitute 20 years after dural plasty. *Surgical Neurology*, 41(4), 338-341. [http://www.doi.org/10.1016/0090-3019\(94\)90187-2](http://www.doi.org/10.1016/0090-3019(94)90187-2)
- Patel, N., & Kirmi, O. (2009). Anatomy and imaging of the normal meninges. *Seminars in Ultrasound, CT and MR*, 30(6), 559-564. <http://www.doi.org/10.1053/j.sult.2009.08.006>
- Ramot, Y., Kronfeld, N., Steiner, M., Manassa, N. N., Bahar, A., & Nyska, A. (2024) Neural tissue tolerance to synthetic dural mater graft implantation in a rabbit durotomy model. *Journal of Toxicologic Pathology*, 37(2), 83-91. <https://doi.org/10.1293/tox.2023-0121>
- Rnjak-Kovacina, J., Wise, S. G., Li, Z., Maitz, P. K. M., Young, C. J., Wang, Y., & Weiss, A. S. (2011). Tailoring the porosity and pore size of electrospun synthetic human elastin scaffolds for dermal tissue engineering. *Biomaterials*, 32(28), 6729-6736. <http://www.doi.org/10.1016/j.biomaterials.2011.05.065>
- Schachtner, J., Frohbergh, M., Hickok, N., & Kurtz, S. (2019). Are Medical Grade Bioabsorbable Polymers a Viable Material for Fused Filament Fabrication? *Journal of Medical Devices*, 13(3), 0310081-310085. <http://www.doi.org/10.1115/1.4043841>
- Shi, Z., Xu, T., Yuan, Y., Deng, K., Liu, M., Ke, Y., Luo, C., Yuan, T., & Ayyad, A. (2016). A New Absorbable Synthetic Substitute With Biomimetic Design for Dural Tissue Repair. *Artificial Organs*, 40(4), 403-413. <http://www.doi.org/10.1111/aor.12568>
- Shijo, M., Honda, H., Koyama, S., Ishitsuka, K., Maeda, K., Kuroda, J., Tanii, M., Kitazono, T., & Iwaki, T. (2017). Dura mater graft-associated Creutzfeldt-Jakob disease with 30-year incubation period. *Neuropathology*, 37(3), 275-281. <http://www.doi.org/10.1111/neup.12359>
- Wang, M. O., Vorwald, C. E., Dreher, M. L., Mott, E. J., Cheng, M. H., Cinar, A., Mehdizadeh, H., Somo, S., Dean, D., Brey, E. M., & Fisher, J. P. (2015). Evaluating 3D-printed biomaterials as scaffolds for vascularized bone tissue engineering. *Advanced Materials*, 27(1), 138-144. <http://www.doi.org/10.1002/adma.201403943>
- Wang, W., & Ao, Q. (2019). Research and application progress on dural substitutes. *Journal of Neurorestoratology*, 7(4-5), 161-170. <http://www.doi.org/10.26599/JNR.2019.9040020>
- Wang, Y.-f., Guo, H.-f., & Ying, D.-j. (2013). Multilayer scaffold of electrospun PLA-PCL-collagen nanofibers as a dural substitute. *Journal of Biomedical Materials Research Part B Applied Biomaterials*, 101(8), 1359-1366. <http://www.doi.org/10.1002/jbm.b.32953>
- Welch, W. C., Cornwall, G. B., Toth, J. M., Turner, A. S., Thomas, K. A., Gerszten, P. C., & Nemoto, E. M. (2002). Use of polylactide resorbable film as an adhesion barrier. *Orthopedics*, 25(10), 1121-1130. <https://doi.org/10.3928/0147-7447-20021002-02>
- Yucel, D., Kose, G. T., & Hasirci, V. (2010). Tissue engineered, guided nerve tube consisting of aligned neural stem cells and astrocytes. *Biomacromolecules*, 11(12), 3584-3591. <https://doi.org/10.1021/bm1010323>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1505905>

Neutral Benchmarking of Survival Models in Health Sciences: Comparative Study of Classical and Machine Learning Techniques

Sumaila ABUBAKARI* Filiz KARAMAN¹ ¹ Yildiz Technical University, Institute of Applied Sciences, Department of Statistics, İstanbul, Türkiye

Keywords	Abstract
Survival Modeling	Survival analysis plays a central role in diverse research fields, especially in health sciences. As an analytical tool, it can be used to help improve patients' survival time, or at least, reduce the prospects of recurrence in cancer studies. However, approaches to the predictive performance of the current survival models mainly center on clinical data along with the classical survival methods. For censored "omics" data, the performance of survival models has not been thoroughly studied, either often due to their high dimensionality issues or reliance on binarizing the survival time for classification analysis. We aim to present a neural benchmark approach that analyzes and compares a broad range of classical and state-of-the-art machine learning survival models for "omics" and clinical datasets. All the methods considered in our study are evaluated using predictability as a performance measure. The study is systematically designed to make 36 comparisons (9 methods over 4 datasets, i.e., 2 clinical and 2 omics), and shows that, in practice, predictability of survival models does vary across real-world datasets, model choice, as well as the evaluation metric. From our results, we emphasize that performance criteria can play a key role in a balanced assessment of diverse survival models. Moreover, the Multitask Logistic Regression (MTLR) showed remarkable predictability for almost all the datasets. We believe this outstanding performance presents a unique opportunity for a wider use of MTLR for survival risk factors. For translational clinicians and scientists, we hope our findings provide practical guidance for benchmark studies of survival models, as well as highlight potential areas of research interest.
Machine Learning	
Benchmarking	
Biomarkers	
Oncology	

Cite
Abubakari, S., & Karaman, F. (2024). Neutral Benchmarking of Survival Models in Health Sciences: Comparative Study of Classical and Machine Learning Techniques. <i>GU J Sci, Part A, 11(3)</i> , 518-534. doi: 10.54287/guj.1505905

Author ID (ORCID Number)	Article Process
0000-0003-4375-6273 Sumaila ABUBAKARI	Submission Date 27.06.2024
0000-0002-8491-674X Filiz KARAMAN	Revision Date 05.07.2024
	Accepted Date 11.07.2024
	Published Date 30.09.2024

1. INTRODUCTION

Survival models are some of the most popular analytical techniques in the field of Statistics that are designed to handle censored observations. In the sense of application, they span a vast majority of fields; Medicine (Salerno & Li, 2023), Education (Arib, 2023), Gadget reliability (Karim & Islam, 2019), and Loan default (Thackham, 2022). Survival analysis is usually seen as a unique technique for its ability to deal with issues of censoring—a scenario where the exact survival time of a patient is not exactly known, either due to the event not being observed within the study time, or partial information of their survival time is known. For those subjects who are censored at the end of the study, we know that their survival time is, at least more than the stated time of the study. Censoring is often grouped into left-, right-, and interval censorship, with the most common being right censoring. The differences in these types lie in the range of the exact survival time we observe. For instance, in right-censoring, we observe the lower limit, the upper limit in left-censoring, and both in the interval-censoring. Comprehensive discussions on various forms of censoring are illustrated in (Klein & Moeschberger, 2003; Gijbels, 2010). In this article, we adopt the terminology of survival analysis in which the "status" variable is binarized, and the time to observe the event is referred to as the survival time.

Over the years, various survival models have been maintained, improved, or extended to achieve different research goals. These include both traditional and machine learning techniques—estimating survivor functions, comparing two or more survival curves, and/or the joint cumulative effect of complex risk factors on the survival time. Often, the inference in survival analysis is obtained from one or a hybrid of different modeling schemes. For instance, the Kaplan-Meier (KM) estimator (Kaplan & Meier, 1958) uses the nonparametric approach to estimate the survivor function while the log-rank test (Peto & Peto, 1972) is used to compare two or more survivor functions. Though the KM estimator is simple and easier to interpret, the parametric approach is preferred for instances in which the distribution of the survival times is pre-determined or assumed. When the target is to estimate the effect of risk factors on survival time, the most popular go-to technique is the CoxPH (Cox, 1972) since the baseline hazard is unspecified, while the effect of predictor(s) is specified parametrically.

In the last few decades, high-throughput techniques have enormously generated data at a faster rate and on large scale (“omics”) from cellular processes. For example, following the rapid progression of technology in DNA microarrays, survival prospects of cancer patients and other forms of diseases have efficiently been improved due to such technologies presenting better paths to evaluate gene expression levels (microarray data extraction). Thus, one can run a survival genomic analysis, focusing on specific genotypes for clinical insights. For instance, non-small cell lung cancer (NSCLC) is considered a chief cause of lung cancer mortalities today. It is believed that the survival rate is influenced by differentially expressed genes (DEGS) between normal lung tissue and NSCLC. It was found that the overall survival rate was highly correlated with DEGs, and enriched in factors such as angiogenesis, DNA replication, and cell cycle (Liu et al., 2019). A challenging task from such microarray data, however, is the enormity of gene expression data used to discriminate between defective cells and normal cells, even for a unit gene. Simply put, we have to deal with the problem of multiple simultaneous hypothesis testing. In fairly low-dimensional data, the Bonferroni correction (Dunn, 1961) sufficiently deals with the problem of multiple testing. However, due to its conservative nature, when tonnes of genes are tested, a small proportion is detected. To overcome this problem, the proposed false discovery rate (FDR) (Benjamini & Hochberg, 1995) uses a method that adjusts for the conservativeness in the Bonferroni correction approach. Even so, genomic data suffers from the problem of the “curse of dimensionality” ($p \gg n$). The statistical techniques to deal with this sort of data go beyond the traditional methods, due to the high dimensional space of the risk factors, coupled with the high collinearity of some genes in the gene expression levels. Again, to overcome this challenge, many researchers have proposed efficient approaches; for example, penalized Cox regression, which trains, tests, and validates the high dimensional data (Dai & Breheny, 2019; Shih & Emura, 2021).

Several survival models have been proposed in the last few decades, from the traditional approaches to the contemporary machine learning models. Numerous investigations in the literature provide a great overview of survival models using right-censored datasets, with little or no focus on time-dependent covariates (see (Wang et al., 2019)). Nonetheless, a limited number of these studies provide comprehensive real-world dataset comparisons, and very few also approach the analysis from a practical point of view. Our motivation for this study stems from providing a fairly broad benchmark study that uses clinical and omics datasets in the health sciences. Our study seeks to improve knowledge and understanding of survival models, as well as to guide clinical decisions. The rest of this paper is organized as follows: in Section 2, we give a general overview of the classical CoxPH model, its proposed regularized extensions for dealing with high-dimensionality, and the modern ML survival models in health science. We also give a comprehensive literature review of some carefully selected articles in cancer research, discussing them in light of commonly used ML survival models for clinical and omics data, and thereafter revisit some benchmark studies in Section 3. In Section 4, we introduce the 4 datasets considered in this study, as well as the 9 models and procedures used for the comparative study. In Section 5, we give the results, and the discussion is provided in Section 6. Finally, we conclude our comparative study in Section 7 with some concluding remarks and implications for clinical researchers.

2. OVERVIEW OF CLASSICAL SURVIVAL AND MACHINE LEARNING MODELS

Generally, there are two arms of survival models. Classical models comprise parametric, semiparametric, and nonparametric models. On the other hand, contemporary ML models comprise state-of-the-art deep neural

learning-based methods and ensemble-based methods. We briefly review both arms of modeling in the next sections.

2.1. Traditional Survival Models

The CoxPH model (Cox, 1972) is the most common traditional survival approach used to evaluate the dependency of survival time on risk factors (predictor variables). It is built on the validity of the PH assumption, mathematically stated by;

$$h(t|X) = h_0(t) \exp\left(\sum_{j=1}^p x_{ij} \beta_j\right) \quad (1)$$

where $h_0(t) \geq 0$ is the baseline hazard function, $X = (x_1, x_2, \dots, x_p)$ is a vector of covariates in the model, and $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ is a vector of coefficients. The unspecified baseline hazard function in equation (1) implies that it assumes no functional form, i.e., a ratio of two hazard functions is free of the baseline hazard (cancels out).

Estimating the regression coefficients in CoxPH requires maximizing the likelihood function. However, due to the presence of censored observations, a *full* maximum likelihood estimation (MLE) is impracticable. To overcome this problem, the partial likelihood method is proposed, which takes into consideration censored and uncensored observations in the dataset (Cox, 1975). For right-censored data, this likelihood function is stated by;

$$L(\beta) = \prod_{j=1}^n \left\{ \frac{\exp(\beta^T x_j)}{\sum_{k \in R(t_j)} \exp(\beta^T x_k)} \right\}^{\delta_j} \quad (2)$$

where $R(t_j) = \{k: t_k \geq t_j\}$ represents the risk set at time t_j . Note that the risk set comprises both the censored and uncensored subjects before time t_j . The goal is to estimate a vector of regression coefficients, $\hat{\beta}$, by maximizing Equation (2).

2.1.1. Penalized Cox Survival Models

Generally, microarray data are known to have overwhelmingly few observations for too many variables ($p \gg n$). For example, gene microarrays have been used to identify significant disease-related genes in a single-wise gene analysis. However, this approach is suboptimal as it fails to identify and explain the complex associations between diseases, genes, and environments. One solution to this problem is the so-called *regularization* approach. Regularization is a technique used to improve and prevent the overfitting of a probabilistic model. In the context of gene microarrays (omics), one of the aims is to identify the most significant risk factors/features among hundreds of thousands of features linked to the outcome variable. Here, the features are selected by imploring different penalty functions on the assumption of sparsity—thus, of the tens of thousands of genes, few of the genes may have significance on the patient's survival time (Hastie et al., 2015). Penalized Cox models are extensions of the CoxPH model, proposed with varying penalized functions—notably among them; lasso-, ridge-, elastic-net, and OSCAR-Cox models (Ye & Liu, 2012; Shih & Emura, 2021). Subject to their respective penalty functions, the regression coefficients are obtained by a minimization (negative) of the partial log-likelihood, shown below;

$$\begin{aligned}
\hat{\beta}_{lasso} &= \operatorname{argmin} \left\{ - \sum_{i=1}^n \delta_i \left(X_i \beta - \log \left(\sum_{i=1}^n \exp (X_i \beta) \right) \right) + \lambda \sum_{k=1}^p |\beta_k| \right\} \\
\hat{\beta}_{ridge} &= \operatorname{argmin} \left\{ - \sum_{i=1}^n \delta_i \left(X_i \beta - \log \left(\sum_{i=1}^n \exp (X_i \beta) \right) \right) + \lambda \sum_{k=1}^p |\beta_k^2| \right\} \\
\hat{\beta}_{EN} &= \operatorname{argmin} \left\{ - \sum_{i=1}^n \delta_i \left(X_i \beta - \log \left(\sum_{i=1}^n \exp (X_i \beta) \right) \right) + \lambda \sum_{k=1}^p |\beta_k^2| + (1 - \alpha) \sum_{k=1}^p \beta_k^2 \right\} \\
\hat{\beta}_{OSCAR} &= \operatorname{argmin} \left\{ - \sum_{i=1}^n \delta_i \left(X_i \beta - \log \left(\sum_{i=1}^n \exp (X_i \beta) \right) \right) + \lambda_1 \|\beta\|_1 + \lambda_2 \|T\beta\|_1 \right\}
\end{aligned} \tag{3}$$

where λ is the ‘*tuning parameter*’ used to regulate the degree of regularization, δ_i is an indicator representing uncensored observation. In the case of the Octagonal Shrinkage and Clustering Algorithm for Regress (OSCAR), T is the sparse (symmetric) edge set matrix obtained by setting a graph structure where each considered feature is a node. For regularization, $\lambda = 0$ means no regularization is performed, while for $\lambda \rightarrow \infty$, the regression coefficients tend to be contained (i.e., regularized). Lasso-Cox uses L_1 -norm regularizer while ridge-Cox uses L_2 -norm regularizer, with the elastic-net using a combination of both penalties. It is important to point out, though, that the various forms of penalization in Equation (3) could also be incorporated into the cost functions of recent machine learning techniques, which we next introduce in the section below.

2.2. Modern Machine Learning (ML) Methods

The traditional survival models are rarely sufficient for capturing complex nonlinear dependencies between the survival time and the predictor variables. To this end, there has been a rising increase in the use of ML models in healthcare, especially for their remarkable performance. Also for their domain adaptability and ability to improve predictive accuracy, applications of ML models span diverse areas of research, for example, security (Liang et al., 2019).

The basic underlying idea of ML is to make a computer run powerful algorithms on complex input data so as to recognize hard-to-discover patterns. Machine learning systems are generally classified based on the task or a set of tasks to accomplish. This classification depends on whether the ML system learns through human supervision (supervised learning) or via other means such as; unsupervised, semi-supervised, reinforcement ML. Of these systems, supervised ML is the most interesting for survival data analysis, especially for classifying and predicting the target variable. We briefly give an overview of common ML methods.

2.2.1. Support Vector Machines (SVMs)

As a supervised ML approach, SVMs have been successful in dealing with regression and classification problems, in addition to the success of their adaptability to fitting survival data (Smola & Schölkopf, 2004). The general aim of SVMs lies in maximizing the distance between two classes while at the same time finding a separate hyperplane that minimizes wrong classification. This hyperplane also attempts to stay so far from close observations so that the individuals found on the edge of the separating hyperplane constitute the supporting vectors, which on the whole determine the classification.

Although using linear classifiers in SVMs is often efficient and enhances performance, in the case of high-dimensional datasets, linear SVM classifiers tend to be poor discriminants. Interestingly, the SVM classifier overcomes this problem by using a high-dimensional kernel function that handles both nonlinearity and high dimensionality. SVMs have also been extended to handle regression and survival data. For example, an SVM-based method, SurvivalSVM, was proposed by Van Belle et al. (2008) for survival modeling. With a modified

penalty term, this model is a variation of the penalized log-likelihood function. SurvivalSVM differs from other models by treating the prognosis problem as a ranking problem, rather than directly incorporating hazard estimation. A further extension known as support vector regression for censored data, SVRc (Khan & Zubek, 2008) was developed to factor in an asymmetric cost function for uncensored and censored data.

2.2.2. Random Survival Forests (RSF)

Random forests constitute another ensemble method purposefully developed for making predictions via tree-structured models (Breiman, 2001). The modeling framework is similar to bagging—thus, to grow the trees, it involves randomly bootstrapping from the training set. The central difference between the two is that, instead of using all the covariates or attributes when splitting a node, an RF uses a random subset of attributes to search for the best variables. Random survival forests have been shown to improve predictive performance due to randomization which reduces the degree of correlation among the trees. As an ensemble learner, RSFs are formed by averaging several base learners, similar to how regression problems are modeled. In the framework of survival ML, the base learner is a survival tree while the ensemble is the cumulative hazard function obtained by averaging the Nelson-Aalen's cumulative hazard function of individual trees (Ishwaran et al., 2011).

Ishwaran et al. (2008) first proposed Random Survival Forests (RSF) as a random forest variation for modeling survival data. Multiple models are generated from a large number of resamples. The result of the ensemble prediction is then averaged across the base learners or the outcome of a majority vote. In our application, the core features of RSF are that we assess the performance of the survival tree rather than using mean square error (mse) as in traditional regression analysis, or the confusion matrix as in classification problems. Additionally, we employ each node's log-rank estimation as the stopping rule.

2.2.3. Boosting-Based Methods (CoxBoosting)

Boosting is another popular ensemble method based on the combination of base learners into a strong learner which represents the final output (Freund, 1990). The principal concept of boosting is to iteratively update a set of predictors by repeatedly learning weak classifiers and adding them to a final strong classifier. Updating is done by minimizing a pre-assigned loss function. Note that after a weak learner is added, weights in the data are readjusted, so-called “re-weighting”.

The Cox boosting model (De Bin, 2016) was proposed and designed based on the classical Cox model, where the boosting is applied in estimating the risk factors, i.e., regression coefficients, as in Equation (1). The β 's are updated iteratively either by using the *mboost* method or by the partial log-likelihood function. CoxBoost is therefore a gradient-boosting algorithm in which the L_2 -norm partial log-likelihood is used (our choice in this study). Note that there are 2 main factors to consider in a boosting procedure: the first one serves as a benchmark to control the weakness of the estimators, while the second parameter specifies the number of boosting iterations to be performed to meet the stopping criterion. The second parameter is necessary to avoid overfitting.

2.2.4. Neural Networks

In recent years, the ML methods discussed here, so far, have been classified as classical ML models, in contrast to the more emerging complex ML neural networks such as deep learning. Inspired by the complicated functionality of the human brain, artificial neural networks (ANN) are a collection of algorithms that are interconnected to process pieces of information in response to input data. Like boosting methods, many neural network methods such as *Cox-nnet* (Ching et al., 2018) and *DeepSurv* have been proposed as extensions to the popular classical Cox regression. The general idea is to have a collection of cost functions to estimate the survival probability or the hazard of patients, assisted by neurons in hidden layers of deep learning architecture. For example, gene microarrays can be represented with these hidden layers, with no stringent regard to the proportional hazard assumption.

While the Cox-nnet and DeepSurv perform favorably well on high-dimensional data, they are PH-based neural networks, hence their predictive power depends on the validity of the PH assumption. To overcome the PH

dependency, DNNSurv (Zhao & Feng, 2019) proposed a new deep neural network model that uses the pseudo-value approach to estimate the survival probability.

2.2.5. Multi-Task Learning (MTL)

Up till now, all the models discussed stress on optimizing one objective. Multi-task learning as an ML model focuses on training the model to perform multiple tasks (objectives) concurrently. Especially in survival models where the dataset is time-dependent, it is more informative to perform tasks concurrently and have the parameters estimated from a joint optimization of numerous likelihood functions—thus, each task corresponds to an objective. In MTL, the aim is to improve the generalized performance of the model by relying on the information shared across multiple tasks. This concept can be implemented in deep learning (DL). For example, Yu et al. (2011) proposed the multi-task logistic regression model (MTLR) as a survival model for multiple time points to use a logistic regression model to predict survival for each. In this case, parameters are jointly estimated by maximizing the joint likelihood function.

2.3 Performance Evaluation

When the time-to-event data are laced with censored observations, the predictive performance of survival models may not be adequately evaluated by the traditional ranking or classification metrics. The most common base metric for evaluating survival models is the *concordance index* (C-Index) (Harrell Jr et al., 1996). As a metric, the index is defined as the ratio of correctly ordered pairs (concordant) to the overall number of possible evaluation pairs. The index values of this metric lie in the range [0,1]. An index of 1 is interpreted as a *perfect* concordance between the event times and the risk. In the same vein, a value of 0 is interpreted as perfect discordance and a value of 0.5 means the model does no better prediction than the toss of a fair coin. Unfortunately, this metric is not unbiased when the amount of censoring in the data is high, thus far, Uno et al. (2011) have proposed an alternative estimator to deal with this situation.

The Brier score (Brier, 1950) is another popular metric for evaluating the predictive performance of survival models. It can be thought of as a cost function that measures the mean squared difference between the predicted probabilities and the true classes. Like the C-Index, it ranges within [0,1]. A metric score of 0 is interpreted as *perfect accuracy* and a metric score of 1 is interpreted as *perfect inaccuracy*. For a given survival time interval, a mathematical integration of multiple Brier scores can be computed as the overall average measure of the performance. This is referred to as the integrated Brier score (ibrier). Incorporating the censoring information into the Brier score method has also been extended (Graf et al., 1999). Other performance metrics for survival models include Royston's D index, Mean Absolute Error (MAE), and the time-dependent AUC.

3. REVIEW OF BENCHMARK STUDIES IN HEALTH RESEARCH

With a primary focus on omics and cancer datasets, we selected articles from popular scientific repositories (e.g., Scopus) by stressing ML techniques and the ubiquitous classical CoxPH and its extensions. We should point out that in doing this, the *exclusion criterion* was to drop articles not focused on survival analysis in health science. At the same, we searched for articles in this field with high citations in the last 5 years and with articles whose main foci are on “comparison”, “benchmarking”, and similar derivatives. After a careful review of the search results, we selected 8 of these papers to study the survival methods explored in them. Based on the techniques, we selected the most commonly used ML survival models in health science (Table 1).

3.1 Benchmark Studies in Health Science

In Moncada-Torres et al. (2021), the classical CoxPH is compared with SVMs, Random Forests, and XGBoost with decision trees as the base learner. To evaluate the performance of these methods, the authors used non-metastatic breast cancer data using the Concordance Index. The performance was identical for all the methods except for the XGBoost which outperformed the rest considered. In another comparative study of the Cox model against random survival forests (RSF) and support vector machine (SVM), Kim et al. (2022) found that the Cox performed slightly better than the RSF and the SVM, in terms of assessing the prognostic prospect of resected non-metastatic pancreatic ductal adenocarcinoma of patients.

In a study to evaluate recurrence patterns and the survivability of gastric cancer patients who underwent chemotherapy and radiation therapy, Akcay et al. (2020) explored ML techniques such as Random Forests, XGBoost, support vector classification, and the Naive Gaussian Bayes techniques. The study concluded the XGBoost and the Random Forest were the best predictors of overall survivability and peritoneal metastases.

Using the SEER database of Lung cancer patients to explore more predictive information, Lynch et al. (2017) explored, relative to the CoxPH model, the predictive performance of ML methods such as Decision Trees, SVMs, Gradient Boosting Machines (GBM), and a custom ensemble. Though the performance of these ML techniques was comparatively similar to the classical Cox proportional method, the Gradient Boosting approach proved to be the most efficient. In a similar retrospective study that focused on prognostic predictive modeling of Breast cancer patients, Xiao et al. (2022) compared the performance of three competing models, namely; Random survival Forests, penalized CoxPH, and SVMs. The study found the Cox model and the SVM to have marginally outperformed the RSF. In a cohort study of breast cancer, Aivaliotis et al. (2021) found the RSF to capture complex non-proportional hazard patterns, they also found out that the RSF overfits the data when compared to the classical CoxPH models—besides the fact that it is less interpretable. To explore the survival outcomes of bladder cancer patients, Bhambhvani et al. (2021) assessed the predictive performance of a multivariable CPM with Artificial Neural Networks (ANN). Intending to predict overall survival (OS) and 5-year specific survival (DSS), this study concluded that ANNs improve predictability in bladder cancer patients than in a multivariable CoxPH model, except for the complexities in the interpretation of ANNs. In a more comprehensive and large-benchmark study, Herrmann et al. (2021), under a multi-omic data setting, benchmarked some eleven methods built around random forests, boosting, and penalized regression. Using the KM estimation and the Cox model as reference methods, the block forest method—a variant of the Random Forest method was found to outperform the classical Cox method, however marginally. Further, the study pointed out that the performance of these methods, to a degree, varied on account of the multi-omic structure of the data.

Richter and Khoshgoftaar (2018) investigated the advances in statistical and ML techniques, the gaps in the literature, and several approaches for developing cancer risk models utilizing structured clinical patient data. The authors concluded that the most popular statistical technique in survival analysis is the CoxPH model while the most common ML approaches are neural networks, SVMs, and decision trees. In the same vein, other researchers have focused on recent advancements in cancer research, concluding that ensemble approaches, decision trees, and artificial neural networks are some of the most used ML methods for setting up survival models. More precisely, a recent comprehensive and methodological literature review by Deepa and Gunavathi (2022) summed up, to date, the majority of cancer research on machine- and deep-learning applications in survival analysis. The comprehensive review cited papers used in ML models—Support Vector Machines, Random Forests, and Support Vector Regression to predict the factors affecting survivability, as well as XGBoost to forecast recurrence and disease progression.

In light of this brief literature review, we deduced that the most widely used modeling techniques in cancer studies include CoxPH, boosting-based methods, Support Vector Machines (SVMs), deep learning-based models, and Random Survival Forests (RSF). Note that it is exceedingly difficult to include every technique in benchmark research. For this reason, our review of survival models considered commonly employed methods, both in the classical sense and the modern ML techniques in health sciences. In Table 1, we present a summary of the articles considered for the review of survival models—methods, performance metrics, and the cancer type.

4. MATERIALS AND METHODS

Clinical data sets—we considered 2 clinical datasets as summarized in Table 2.

- Lung dataset: This data contains 7 features of 228 patients diagnosed with advanced lung cancer from the North Central Cancer Treatment Group (NCCTG). This data can be obtained from the survival package in R and contains the survival information of the patients.

- Veteran dataset: Contains lung cancer data from a randomized trial of 2 treatment regimens with 6 features and 137 patients. This data is also readily available in the survival package in R.

Omics data sets—we considered 2 omics datasets as summarized in Table 2.

- Ovarian1 dataset: Ovarian cancer gene expression data from curatedOvarianData package with “GSE49997_eset” as the data ID. The study was on OVCAD Consortium conducting a study to validate the impact of a molecular subtype on ovarian cancer outcomes. The curation of this is given by Ganzfried et al. (2013).
- Ovarian2 dataset: Ovarian cancer gene expression data from curatedOvarianData package with “GSE30161_eset” as the data ID. The study was on multi-gene expression predictors of single drug responses to adjuvant treatment in ovarian cancer: predicting platinum resistance. The curation of this is given by Ganzfried et al. (2013).

Table 1. Summary of selected benchmarking studies in oncology (cancer) focusing on survival models of classical and modern machine learning model techniques.

Study	Technique	Type of cancer	Evaluation metric
Herrmann et al. (2021)	Boosting-based Penalized regression-based Random-forest-based	Liver, Blood, Lung, Skin Brain, Kidney, Stomach, Colon Ovarian, Pancreatic, Bladder, Breast Head-neck (SEER, TCGA)	Brier score Concordance Index
Moncada-Torres et al. (2021)	CoxPH Random forest SVMs XGBoosting	Breast cancer	Concordance Index
Xiao et al. (2022)	Random forest Penalized-based regression Support Vector Machines	Breast cancer	Brier score Concordance Index AUC D-index
Kim et al. (2022)	CoxPH Random survival forests SVMs	Pancreatic (SEER/KOTUS-BP)	AUC sensitivity specificity
Akçay et al. (2020)	XGBoosting Random forests SVMs Logistic regression Gaussian Naive Bayes (GNB) algo. Multi-layer perceptron	Gastric cancer	AUC sensitivity specificity
Bhambhani et al. (2021)	Multivariable CoxPH Artificial Neural Networks	Bladder cancer (SEER)	AUC
Lynch et al. (2017)	Decision Trees Multivariable CoxPH Support Vector Machines Gradient Boosting Linear regression	Lung cancer (SEER)	Root Mean Square Error
Aivaliotis et al. (2021)	CoxPH Random forests	Breast cancer (UK)	Brier score Concordance Index

Table 2. Summary of datasets considered in the comparative study in this paper. We round the rate of censorship to three decimal places

Dataset source	No. of observations	No. of features	Data class	Censoring rate
Veteran (Kalbfleisch & Prentice, 2011)	137	8	Clinical	0.066
NCCTG Lung (Loprinzi et al., 1994)	228	9	Clinical	0.276
Ovarian 1 (Ganzfried et al., 2013)	194	16050	Omics	0.706
Ovarian 2 (Ganzfried et al., 2013)	58	19818	Omics	0.379

4.1. Benchmarking Procedure and Methods

Our study is aimed at a neutral comparison of classical survival models to the state-of-the-art ML models. Generally, the performance of probabilistic models is often influenced by the design and choice of datasets, and for this reason, we chose two data types—clinical and omics datasets. For each of the 9 models considered in this study, we evaluated the models' performances in terms of their predictability. Here, predictability is in reference to 4 popular evaluation metrics. Note that when the amount of censoring in the test data is high, Harrell's c-index is known to be a biased estimator, and for this reason, we have included an alternative estimate, Uno's c-index which uses the inverse probability of censoring weighting (ipcw). We also calculated the standard deviation (SD). All the analysis in our comparative study is conducted using **R 4.3.1**.

The models considered in the study are outlined in Table 3. For the comparison of the methods to be on neutral grounds, we use the default settings for the hyperparameter tuning (except for CoxPH). The evaluation of the methods are carried out on real-world data sets (2 clinical, 2 omics) as shown in Table 2. Using RStudio, we run the analysis repeatedly with 5-fold cross validation. In each run, we split the entire data into a *training* data set (80%) and the remaining 20% used as the *testing* set. Each model is trained using the training set while the evaluation metrics are computed using the testing set. In a case where there is a feature selection step, the 5-fold cross validation is still applied to the nested feature selection. We provide a supplementary data (*Supplementary-Table-1*) on the details of packages and parameters for the methods.

All the 4 datasets in the comparative study are benchmark datasets—the ovarian datasets have already being manually curated (clinical) and the expression data in them have also being uniformly processed.

5. RESULTS

To exhaustively assess the strengths and weaknesses of the various survival models, we settled on 9 typical methods from our extensive review of the literature. We then study their statistical performance to 4 diverse datasets. The basis of our assessment is on 4 metrics on predictability; Harell's C-Index, Uno's C-Index, Brier's score, and the time-dependent AUC. The primary focus is on extensive comparison of models from the traditional approaches to the state-of-the-art machine learning—*Classical models*: Cox, Cox_lasso, Cox_ridge, Cox_elasticNet; *Advanced ML*: RSF, SurvivalSVM, CoxBoosting; *Neural Network model*: DeepSurv.

5.1. Practical Consideration in Performance Assessment

Owing to the varying characteristics of data collection in domains like medical field (e.g., omics data), not all survival models meet the feasibility criteria for application to all data types. For instance, the classical CoxPH methods cannot handle high-dimensional data (i.e., $p \gg n$) where there are more features (p) than samples (n). For this reason, the 2 omics datasets, Ovarian_1, Ovarian_2, are not feasible for the classical Cox method, as seen in Table 4. Further, the results highlight, for example, that SurvivalSVM is readily applicable to all the 2 clinical datasets, whereas it is not readily applicable to Ovarian_2 dataset.

The multitask linear regression (MTLR) approach, conspicuously, outperformed the rest of the learners on all the evaluation metrics as presented in Table 4 (bolden mean values), across all the data sets, except for

Ovarian_2 where the method is not applicable. More so, looking at the average degree of performance between the clinical and omic data sets suggests that the learners (methods) depend on the type of data set.

For instance, Figure 1 shows the comparative performance of all the methods using Uno's C-index, which uses the inverse probability of censoring weighted, IPWC, (See supplementary material for results of other metrics). To appreciate the gain in model predictability relative to the Cox-based ML method, we compare the CoxBoosting method with the conventional Cox-based methods (i.e., Cox_EN, Cox_Lasso), which are typically employed as the standard of comparison in many medical investigations. Figure 2 presents our results which demonstrate similar performance across the datasets. For this comparison, we observe similar predictability for both the clinical and omic datasets when measured by both Harrell's C-Index and Uno's C-Index. This observation suggests that the performance of clinical and complex health data using state-of-the-art ML may not be as straightforward as is the case in some fields.

Table 3. Summary table of traditional and state-of-the-art ML models considered in the benchmarking, R functions with their listed parameters.

Model name	Function	R package	Default parameters
Cox	coxph	survival	none
Lasso Cox (cox_lasso)	penalized glmnet	penalized glmnet (omics)	lambda1 = 1, lambda2 = 0 alpha = 1.0, nfold = 5, type.measure ='C'
Ridge Cox (cox_ridge)	penalized glmnet	penalized glmnet (omics)	lambda1 = 0, lambda2 = 1 alpha = 0.0, nfold = 5L, type.measure = 'C'
Elastic net (cox_en)	penalized glmnet	penalized glmnet (omics)	lambda1 = 1, lambda2 = 1 alpha = 0.50, nfold = 5L, type.measure = 'C'
Random survival forest	rfsrc	randomForestSRC	ntree = 1000, mtry = 10
SurvivalSVM	survivalsvm	survivalsvm	margin = 0.050, bound = 10, eig.tol = 1e-05 sgf.sv = 5, sigf = 7, maxiter = 20 conv.tol = 1e-07, posd.tol = 1e-08
Cox Boosting model	coxboost	Coxboost	stepnumber = 10, penalty number = 100
Deep survival (Deepsurv)	deepsurv	survivalmodels	frac = 0.3, activation = 'relu', dropout = 0.10, early_stopping = T, num_nodes = c(4L, 8L, 4L, 2L), epochs = 100, batch_size = 32
Multitask Logistic Regression	mtlr	MTLR	C1 = 1, normalize = T, train_biases = T

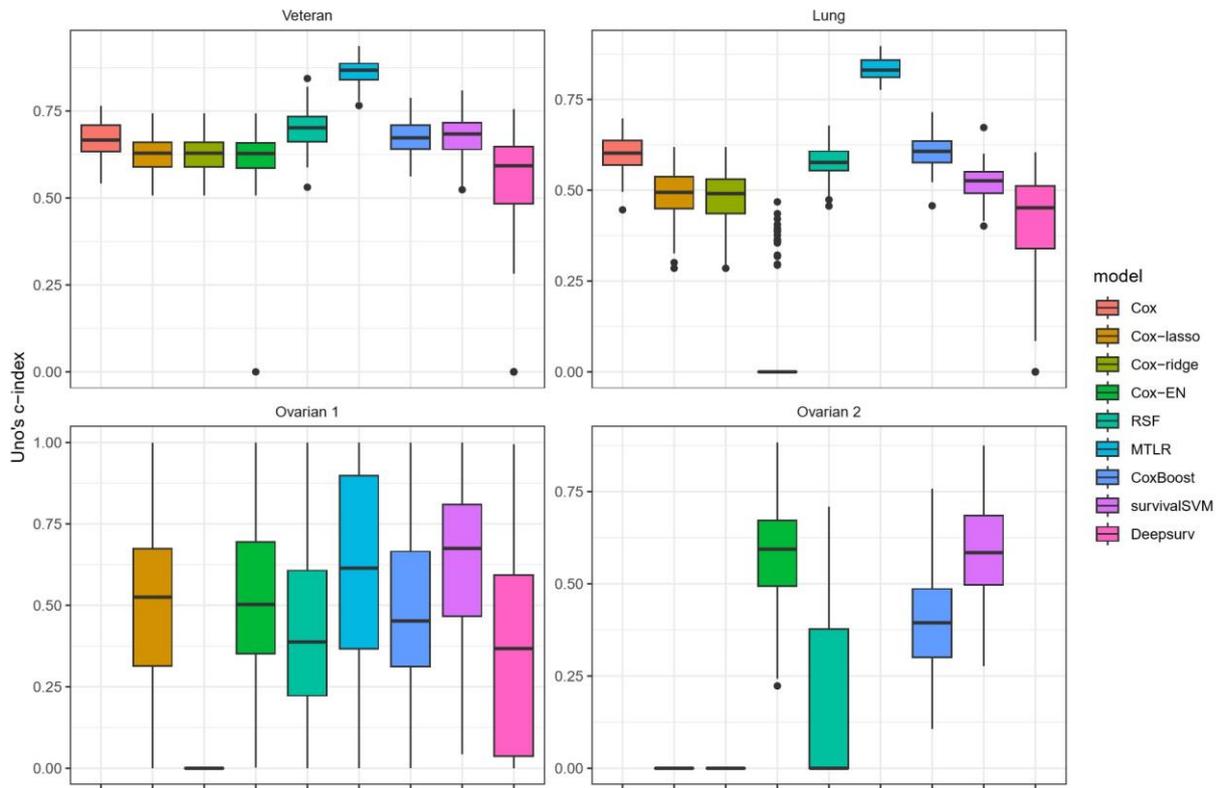


Figure 1. Visualization of Uno's concordance index for each of the 9 learners considered on all the 4

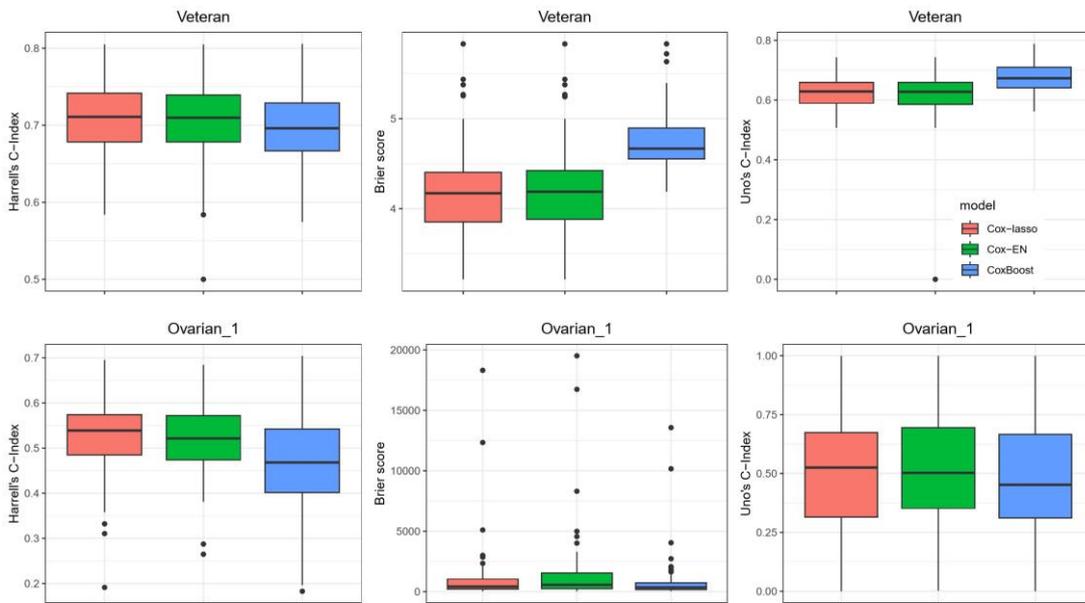


Figure 2. Comparison of model predictability between ML Cox-based method and the traditional penalized Cox-related learners. Top-left: Harrell's metric on Veteran dataset. Top-middle: Brier's metric on Veteran dataset. Top right: Uno's metric on Veteran dataset. Bottom-left: Harrell's metric on Ovarian 1 dataset. Bottom-middle: Brier's metric on Ovarian 1 dataset. Bottom-right: Uno's metric on Ovarian dataset.

6. DISCUSSION

More principally, caution should be exercised when interpreting the results of our comparative study. This is so because of the assessment methods used to measure the performance of the competing models. Stated differently, the relative standing of the models tends to change significantly from one scale of assessment measure to the other, in addition to the default settings of the packages considered in the benchmarking. For example, when “omics” data is considered, the *CoxBoosting* method under C-Index performs poorly as opposed to Uno’s index, integrated Brier score (ibrier), and the time-dependent measures. This observation is identical to the study by Zhang et al. (2022). But if we considered the low-dimensional datasets (i.e., NCCTG Lung cancer and Veteran Lung cancer datasets), the C-Index outperforms the Brier score. What this observation shows is that the predictive performance can significantly vary with the assessment scale employed in the modeling. Of the three measures (i.e., C-index, ibrier score, and time-dependent AUC), the AUC-scale measure in nearly all cases outperformed the other two measures. It is important however, to mention at this point that the C-Index as a discriminant measure should not be given preference over the ibrier score since according to Herrmann et al. (2021), the C-Index cannot be considered an appropriate scoring rule. Thus, for prognostic calibration, choose the Brier score over the Harrell’s index while for ease of interpretability, the C-Index might be preferred over the Brier score.

For predictive power, the MTLR model outperformed all the other models in nearly all the datasets we considered. A couple of reasons could be cited for this observation of performance and this is explained in detail in Yu et al. (2011). One reason worth mentioning is the fact that the MTLR model is robustly designed to allow concurrent fitting of multiple logistic regression models while directly taking into consideration the survivor function. Despite the great performance of the MTLR, not many researchers have employed its application in the modeling of omics survival data Zhang et al. (2022). This presents an opportunity for more translational medical researchers to consider this approach in their analytical toolkit.

Ideally, the best-performing model in a comparative study of competing methods is unlikely to exist in all likely scenarios because a high-performing model under one consideration could be a low-performing model under a different criterion. This should necessarily imply that the methods in our study that did not perform well might be great competitors under different criteria—interpretability, predictability, or problems with the “curse of dimensionality”.

7. CONCLUSION

In our benchmark comparative study, we extensively evaluated the importance of survival methods in practice, where the selected models are based on our comprehensive review of ML benchmark studies, and applying these models to diverse datasets in biomedical studies. Using model predictability as the basis of performance, we explored a wide range of survival models from the traditional CoxPH methods to the state-of-the-art ML models, where the evaluation metric was assessed on 4 scales—Harrell-, Uno- c-index, Brier score, and time-dependent AUC. In our study, we did not attempt a comprehensive survey of tuning procedures for the 9 models considered. The core reason, in practice, is that the default set of hyperparameters is often used, hence the decision to rely on the default parameters in our study. Nevertheless, we point out that different targeted penalized techniques for a given data might cause varying performances for the learners covered in our study. Our research’s conclusions will provide some degree of guidance for clinicians and translation researchers, while at the same time pointing to areas of potential study in the scope of benchmarking methodology and survival approaches.

There has been, in recent years, a distinct change in course in how time-to-event data are modeled, from the classical approach of directly modeling the hazard estimator to building several models using the survivor function. In theory, though, modeling the hazard function paves a great way to identify key biomarkers/risk factors related to the survival prospects of patients. However, if the primary goal is to accurately predict the patient’s survival, then modeling directly on the survival probability greatly improves predictability. To this end, modeling methods such as the survivalSVM, and MTLR, which employed direct modeling of the survivor, demonstrated better performance based on predictability. This finding is consistent with the analysis by Yu et al. (2011) about the effective performance of their MTLR method.

In our benchmark study, it is interesting that MTLR comparatively showed exceptionally high model predictability. This high performance may be explained on several fronts. As Yu et al. (2011) extensively discussed, the performance of the MTLR is attributable to 3 central reasons; dynamic modeling, direct modeling of the survivor function, and concurrent construction of multiple logistic regression learners. To account for nonlinearity in datasets, many researchers built on the proposed MTLR by incorporating neural networks (see Fotso (2018)). Surprisingly, very few studies have used MTLR, either by using clinical data or high-dimensional omics data. Thus far, we think there is potential to employ MTLR more extensively for predicting risk factors in survival modeling, given its remarkably high model predictability.

One of the most important measures for evaluating survival studies is model predictability, with evaluation indices such as Harrell's c-index being the most widely used. Though Harrell's c-index is a ranked-based metric capable of evaluating predicted outcomes with censored data, a couple of versions of it are also available, for example, the Uno's c-index uses the IPCW technique to evaluate predicted outcomes. Other than the concordance indices, other metrics for evaluation include the Brier score and the time-dependent AUC, where the time interval is divided into several time points for evaluation, similar to the general idea of AUC in binary classification. Since model predictability can be evaluated on several metrics, we suggest that a combination of different metrics should be employed to help in comprehensively assessing the fitted model.

Though many survival models have algorithms capable of fitting both omics and clinical data, several recently developed methods are uniquely designed for high-dimensionality in omics data (e.g., genomics, transcriptomics). For example, CoxBoost as an ML learner is tailored to handle the curse of dimensionality ($p \gg n$) in omics data. The goal of developing these data-specific techniques and/or learners is to efficiently capture the distinctive features of the omics or clinical data. Apart from the fact that the performance of a model's predictability (real-world datasets) is influenced by the data type (clinical or omics), another aspect of the data that can affect predictability is *data modality*, which we did not cover in this paper.

AUTHOR CONTRIBUTIONS

Conceptualization, S.A; methodology, S.A and F.K, Software; S.A and F.K, Validation, F.K, formal analysis; S.A and F.K, data curation, S. A, manuscript-original draft, S.A; visualization, S.A and F.K, supervision, F.K. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Aivaliotis, G., Palczewski, J., Atkinson, R., Cade, J. E., & Morris, M. A. (2021). A comparison of time to event analysis methods, using weight status and breast cancer as a case study. *Scientific Reports*, 11(1), 14058. <https://doi.org/10.1038/s41598-021-92944-z>
- Akcay, M., Etiz, D., & Celik, O. (2020). Prediction of survival and recurrence patterns by machine learning in gastric cancer cases undergoing radiation therapy and chemotherapy. *Advances in Radiation Oncology*, 5(6), 1179-1187. <https://doi.org/10.1016/j.adro.2020.07.007>
- Arib, M. A. A. (2023). Survival analysis of students not graduated on time using cox proportional hazard regression method and random survival forest method. *Journal of Statistics and Data Science*, 13-21. <https://doi.org/10.33369/jdsds.v2i1.24312>
- Bengio, Y., & Grandvalet, Y. (2003). No unbiased estimator of the variance of k-fold cross-validation. *Advances in Neural Information Processing Systems*, 16.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bhambhani, H. P., Zamora, A., Shkolyar, E., Prado, K., Greenberg, D. R., Kasman, A. M., Liao, J., Shah, S., Srinivas, S., Skinner, E. C., & Shah, J. B. (2021). Development of robust artificial neural networks for

- prediction of 5-year survival in bladder cancer. *Urologic Oncology: Seminars and Original Investigations*, 39(3), 193.e7-193.e12. <https://doi.org/10.1016/j.urolonc.2020.05.009>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1-3. [https://doi.org/10.1175/1520-0493\(1950\)078%3C0001:VOFEIT%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078%3C0001:VOFEIT%3E2.0.CO;2)
- Ching, T., Zhu, X., & Garmire, L. X. (2018). Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Computational Biology*, 14(4), e1006076. <https://doi.org/10.1371/journal.pcbi.1006076>
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2), 269-276. <https://doi.org/10.1093/biomet/62.2.269>
- Dai, B., & Breheny, P. (2019). Cross-validation approaches for penalized Cox regression. *Statistical Methods in Medical Research*, 33(4), 702-715. <https://doi.org/10.1177/09622802241233770>
- De Bin, R. (2016). Boosting in cox regression: A comparison between the likelihood-based and the model-based approaches with focus on the r-packages CoxBoost and mboost. *Computational Statistics*, 31, 513-531. <https://doi.org/10.1007/s00180-015-0642-2>
- Deepa, P., & Gunavathi, C. (2022). A systematic review on machine learning and deep learning techniques in cancer survival prediction. *Progress in Biophysics and Molecular Biology*. <https://doi.org/10.1016/j.pbiomolbio.2022.07.004>
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293), 52-64. <https://doi.org/10.1080/01621459.1961.10482090>
- Fotso, S. (2018). Deep neural networks for survival analysis based on a multi-task framework. *arXiv Preprint arXiv:1801.05512*. <https://doi.org/10.48550/arXiv.1801.05512>
- Freund, Y. (1990). Boosting a weak learning algorithm by majority. In M. FULK & J. CASE (Eds.), *Colt proceedings 1990* (pp. 202-216). Morgan Kaufmann. <https://doi.org/10.1016/B978-1-55860-146-8.50019-9>
- Ganzfried, B. F., Riester, M., Haibe-Kains, B., Risch, T., Tyekucheva, S., Jazic, I., Wang, X. V., Ahmadifar, M., Birrer, M. J., Parmigiani, G., Huttenhower, C., & Waldron, L. (2013). curatedOvarianData: Clinically annotated data for the ovarian cancer transcriptome. *Database*, 2013. <https://doi.org/10.1093/database/bat013>
- Gijbels, I. (2010). Censored data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), 178-188. <https://doi.org/10.1002/wics.80>
- Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18(17-18), 2529-2545. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990915/30\)18:17/18%3C2529::AID-SIM274%3E3.0.CO;2-5](https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18%3C2529::AID-SIM274%3E3.0.CO;2-5)
- Harrell Jr, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4), 361-387. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4%3C361::AID-SIM168%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4)
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. CRC press. <https://doi.org/10.1201/b18401>
- Herrmann, M., Probst, P., Hornung, R., Jurinovic, V., & Boulesteix, A. (2021). Large-scale benchmark study of survival prediction methods using multi-omics data. *Briefings in Bioinformatics*, 22(3), bbaa167. <https://doi.org/10.1093/bib/bbaa167>
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). *Random survival forests*. <https://doi.org/10.1214/08-AOAS169>

- Ishwaran, H., Kogalur, U. B., Chen, X., & Minn, A. J. (2011). Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(1), 115-132. <https://doi.org/10.1002/sam.10103>
- Kalbfleisch, J. D., & Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. (2nd Ed.). John Wiley & Sons.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), 457-481. <https://doi.org/10.1080/01621459.1958.10501452>
- Karim, Md. R., & Islam, M. A. (2019). *Reliability and Survival Analysis*. Springer Singapore. <https://doi.org/10.1007/978-981-13-9776-9>
- Khan, F. M., & Zubek, V. B. (2008). Support vector regression for censored data (SVRc): A novel tool for survival analysis. *2008 Eighth IEEE International Conference on Data Mining*, 863-868. <https://doi.org/10.1109/ICDM.2008.50>
- Kim, H., Park, T., Jang, J., & Lee, S. (2022). Comparison of survival prediction models for pancreatic cancer: Cox model versus machine learning models. *Genomics & Informatics*, 20(2). <https://doi.org/10.5808/gi.22036>
- Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed). Springer.
- Liang, F., Hatcher, W. G., Liao, W., Gao, W., & Yu, W. (2019). Machine learning for security and the internet of things: The good, the bad, and the ugly. *IEEE Access*, 7, 158126-158147. <https://doi.org/10.1109/ACCESS.2019.2948912>
- Liu, C., Chen, Y., Deng, Y., Dong, Y., Jiang, J., Chen, S., Kang, W., Deng, J., & Sun, H. (2019). Survival-based bioinformatics analysis to identify hub genes and key pathways in non-small cell lung cancer. *Translational Cancer Research*, 8(4). <https://tcr.amegroups.org/article/view/30209>
- Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M., & Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North central cancer treatment group. *Journal of Clinical Oncology*, 12(3), 601-607. <https://doi.org/10.1200/JCO.1994.12.3.601>
- Lynch, C. M., Abdollahi, B., Fuqua, J. D., de Carlo, A. R., Bartholomai, J. A., Balgeman, R. N., van Berkel, V. H., & Frieboes, H. B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International Journal of Medical Informatics*, 108, 1-8. <https://doi.org/10.1016/j.ijmedinf.2017.09.013>
- Moncada-Torres, A., Maaren, M. C. van, Hendriks, M. P., Siesling, S., & Geleijnse, G. (2021). Explainable machine learning can outperform cox regression predictions and provide insights in breast cancer survival. *Scientific Reports*, 11(1), 6968. <https://doi.org/10.1038/s41598-021-86327-7>
- Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, 135(2), 185-198. <https://doi.org/10.2307/2344317>
- Richter, A. N., & Khoshgofaar, T. M. (2018). A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. *Artificial Intelligence in Medicine*, 90, 1-14. <https://doi.org/10.1016/j.artmed.2018.06.002>
- Salerno, S., & Li, Y. (2023). High-dimensional survival analysis: Methods and applications. *Annual Review of Statistics and Its Application*, 10(1), 25-49. <https://doi.org/10.1146/annurev-statistics-032921-022127>
- Shih, J., & Emura, T. (2021). Penalized cox regression with a five-parameter spline model. *Communications in Statistics-Theory and Methods*, 50(16), 3749-3768. <https://doi.org/10.1080/03610926.2020.1772305>
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Thackham, M. (2022). *Survival analysis: Applications to credit risk default modelling* [PhD thesis, Macquarie University]. <https://doi.org/10.25949/19436723.v1>

- Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L.-J. (2011). On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, 30(10), 1105-1117. <https://doi.org/10.1002/sim.4154>
- Van Belle, V., Pelckmans, K., Suykens, J. A., & Van Huffel, S. (2008). *Survival SVM: A practical scalable algorithm*. In: Proceedings of the 16th European Symposium on Artificial Neural Networks (pp. 89-94).
- Wang, P., Li, Y., & Reddy, C. K. (2019). *Machine learning for survival analysis: A survey*. 51(6). <https://doi.org/10.1145/3214306>
- Xiao, J., Mo, M., Wang, Z., Zhou, C., Shen, J., Yuan, J., He, Y., & Zheng, Y. (2022). The application and comparison of machine learning models for the prediction of breast cancer prognosis: Retrospective cohort study. *JMIR Medical Informatics*, 10(2), e33440. <https://doi.org/10.2196/33440>
- Ye, J., & Liu, J. (2012). Sparse methods for biomedical data. *ACM Sigkdd Explorations Newsletter*, 14(1), 4-15. <https://doi.org/10.1145/2408736.2408739>
- Yu, C., Greiner, R., Lin, H., & Baracos, V. (2011). Learning patient-specific cancer survival distributions as a sequence of dependent regressors. *Advances in Neural Information Processing Systems*, 24.
- Zhang, Y., Wong, G., Mann, G., Muller, S., & Yang, J. Y. (2022). SurvBenchmark: Comprehensive benchmarking study of survival analysis methods using both omics data and clinical data. *GigaScience*, 11, giac071. <https://doi.org/10.1093/gigascience/giac071>
- Zhao, L., & Feng, D. (2019). Dnnsurv: Deep neural networks for survival analysis using pseudo values. *IEEE Journal of Biomedical and Health Informatics*, 24(11), 3308-3314. <https://doi.org/10.1109/JBHI.2020.2980204>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1546375>

Further Results for Hermite-Based Milne-Thomson Type Fubini Polynomials with Trigonometric Functions

Neslihan KILAR^{1*} ¹ Niğde Ömer Halisdemir University, Bor Vocational School, Department of Computer Technologies, Türkiye

Keywords	Abstract
Hermite-Based Milne-Thomson Type Polynomials Stirling Type Polynomials and Numbers Trigonometric Functions Generating Functions	This paper examines generating functions of r -parametric Hermite-based Milne-Thomson polynomials. Using generating function methods, the relationships among these polynomials, Fubini type polynomials, and trigonometric functions are given. Moreover, new formulas are derived by utilizing not only the generating functions of these polynomials but also associated functional equations. These formulas pertain to r -parametric Hermite-based sine-and cosine-Milne-Thomson Fubini polynomials, as well as Stirling type polynomials and numbers. Additionally, by analyzing special cases of newly obtained results, some known formulas are also derived. Furthermore, some identities involving secant and cosecant numbers are derived through the properties of trigonometric functions. Special polynomials and their generating functions are an important tool for solving some problems in many areas such as combinatorics and number theory. By introducing new formulas, this paper significantly enhances these problems-solving abilities in these areas. Consequently, these results have potential to shed light on important applications in mathematics, engineering, and mathematical physics.

Cite

Kılar, N. (2024). Further Results for Hermite-Based Milne-Thomson Type Fubini Polynomials with Trigonometric Functions. *GU J Sci, Part A, 11(3)*, 535-545. doi:10.54287/guj.1546375

Author ID (ORCID Number)	Article Process
0000-0001-5797-6301 Neslihan KILAR	Submission Date 09.09.2024 Revision Date 16.09.2024 Accepted Date 23.09.2024 Published Date 30.09.2024

1. INTRODUCTION

Trigonometric functions have many significant applications. These functions are also related to special polynomials with their generating functions. These functions enable to define various families of special polynomials. Further, these functions play a crucial role in examining their fundamental properties of related polynomials. They used to construct mathematical modeling and solve differential equations in applied science. These frameworks of trigonometric functions and special polynomials in mathematical analysis, as well as in addressing physical problems and engineering applications. For example, many applications and formulas for these have been given by authors (Agyuz, 2024;-; Zayed et al., 2024). Inspired the above explanations, the motivation of this paper is to deal with the r -parametric Hermite-based Milne-Thomson polynomials with trigonometric functions and the Fubini type numbers in detail. The first motivation begins with the notations and definitions related to certain special polynomials and numbers. The second motivation use generating functions in order to obtain some novel results for certain families of polynomials.

Let $e^v = \exp(v)$ and

$$\frac{(\xi)_d}{d!} = \binom{\xi}{d} = \frac{\xi(\xi-1)(\xi-2)\dots(\xi-d+1)}{d!},$$

where ξ is complex numbers: \mathbb{C} , $d \in \mathbb{N} = \{1, 2, 3, \dots\}$, and $(\xi)_0 = 1$ (Bayad & Simsek, 2014;-; Srivastava, 1976).

The generating function that pertains to secant numbers is expressed by

$$\sec(v) = \sum_{k=0}^{\infty} s_k \frac{v^{2k}}{(2k)!} \quad (1)$$

(Kim & Kim, 2018 (Equation (1.12))).

The first few terms of the numbers s_k are presented below:

$$s_0 = 1, \quad s_1 = 1, \quad s_2 = 5, \quad s_3 = 61, \quad s_4 = 1385, \quad s_5 = 50521, \quad s_6 = 2702765, \dots$$

The secant numbers are also called “zig numbers” (Kim & Kim, 2018).

The generating function that pertains to hyperbolic cosecant numbers is expressed by

$$v \operatorname{csch}(v) = \sum_{k=0}^{\infty} d_k \frac{v^k}{k!} \quad (2)$$

(Kim & Kim, 2018 (Equation (2.22))).

By using (2), one has

$$v \operatorname{csc}(v) = (vi) \operatorname{csch}(vi) = \sum_{k=0}^{\infty} (-1)^k d_{2k} \frac{v^{2k}}{(2k)!}, \quad (3)$$

where $i^2 = -1$ (Kim & Kim, 2018 (Equation (2.24))).

The first few terms of the numbers d_k are presented below:

$$d_0 = 1, \quad d_1 = 0, \quad d_2 = -\frac{1}{3}, \quad d_3 = 0, \quad d_4 = \frac{7}{15}, \quad d_5 = 0, \quad d_6 = -\frac{31}{21}, \dots$$

The numbers d_k are pertains to Bernoulli numbers and polynomials (Kim & Kim, 2018).

The generating function that pertains to generalized Hermite-Kampè de Fèriet polynomials is expressed by

$$\exp\left(\sum_{j=1}^r u_j v^j\right) = \sum_{k=0}^{\infty} H_k(\vec{u}, r) \frac{v^k}{k!}, \quad (4)$$

where $\vec{u} = (u_1, u_2, \dots, u_r)$ (Dattoli et al., 1994; 1996; Kilar, 2021; Kilar & Simsek, 2021). Many researchers have studied these types of polynomials using various techniques, such as degenerate versions and q -calculus (Cesarano et al., 2022; Fadel et al., 2024; Zayed et al., 2024).

From (4), we get

$$H_k(\vec{u}, r) = k! \sum_{j=0}^{\lfloor \frac{k}{r} \rfloor} \frac{u_r^j H_{k-rj}(\vec{u}, r-1)}{j! (k-jr)!},$$

where $[m]$ is the largest integer m (Dattoli et al., 1994; 1996; Kilar, 2021; Kilar & Simsek, 2021).

The generating function that pertains to λ -array polynomials is expressed by

$$(\lambda \exp(v) - 1)^p \exp(vx) = \sum_{k=0}^{\infty} \frac{p! S_p^k(x; \lambda)}{k!} v^k, \quad (5)$$

where $p \in \mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ and $\lambda \in \mathbb{C}$ (Simsek, 2013).

For $x = 0$ in (5) yields λ -Stirling numbers of the second kind

$$(\lambda \exp(v) - 1)^p = \sum_{k=0}^{\infty} \frac{p! S_2(k, p; \lambda)}{k!} v^k \quad (6)$$

(Simsek, 2013).

For $\lambda = 1$ in (6) yields Stirling numbers of the second kind

$$S_2(k, p; 1) := S_2(k, p).$$

The generating function that pertains to numbers $S_{22}(k, p)$ is expressed by

$$(\exp(v) - v - 1)^p = \sum_{k=0}^{\infty} \frac{p! S_{22}(k, p)}{k!} v^k \quad (7)$$

(Charalambides, 2005 (pp. 123-127); Simsek, 2024).

The generating functions that pertain to polynomials are expressed by

$$\exp\left(\sum_{j=1}^r u_j v^j\right) \cos(yv) = \sum_{k=0}^{\infty} C_k(\vec{u}, y; r) \frac{v^k}{k!} \quad (8)$$

and

$$\exp\left(\sum_{j=1}^r u_j v^j\right) \sin(yv) = \sum_{k=0}^{\infty} S_k(\vec{u}, y; r) \frac{v^k}{k!} \quad (9)$$

(Kilar & Simsek, 2021; and also Kilar, 2021; 2023a).

When $r = 1$ in (8) and (9), one has

$$C_k(u_1, y; 1) := C_k(u_1, y) \quad \text{and} \quad S_k(u_1, y; 1) := S_k(u_1, y).$$

That is, the polynomials $C_k(x, y)$ and $S_k(x, y)$ are expressed by

$$\exp(vx) \cos(yv) = \sum_{k=0}^{\infty} C_k(x, y) \frac{v^k}{k!} \quad (10)$$

and

$$\exp(vx) \sin(yv) = \sum_{k=0}^{\infty} S_k(x, y) \frac{v^k}{k!}. \quad (11)$$

Kilar & Simsek (2021) defined the r -parametric Hermite-based Milne-Thomson type polynomials as follows:

$$2(b + f(v, a))^p \exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \cos(yv) = \sum_{k=0}^{\infty} \mathfrak{h}_1(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!} \quad (12)$$

and

$$2(b + f(v, a))^p \exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \sin(yv) = \sum_{k=0}^{\infty} \mathfrak{h}_2(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!}, \quad (13)$$

where $f(v, a)$ refers to a meromorphic function or an analytic function, $p \in \mathbb{N}_0$ and a, b are real numbers: \mathbb{R} (see also Kilar, 2021).

When $b = 0$ and $f(v, 1) = \frac{2}{(2 - \exp(v))^2}$ in (12) and (13), we have the r -parametric Hermite-based cosine- and sine-Milne-Thomson type Fubini polynomials, respectively:

$$\frac{2^{p+1}}{(2 - \exp(v))^{2p}} \exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \cos(yv) = \sum_{k=0}^{\infty} \mathfrak{F}\mathfrak{h}_1(k, x, y, p; \vec{u}, r) \frac{v^k}{k!} \quad (14)$$

and

$$\frac{2^{p+1}}{(2 - \exp(v))^{2p}} \exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \sin(yv) = \sum_{k=0}^{\infty} \mathfrak{F}\mathfrak{h}_2(k, x, y, p; \vec{u}, r) \frac{v^k}{k!} \quad (15)$$

(Kilar, 2021; Kilar & Simsek, 2021).

Using Eqs. (14) and (15), some special cases of these are given as follows.

When $y = 0$ and $\vec{u} = (0, 0, \dots, 0) = \vec{0}$ in (14), one has

$$\mathfrak{F}\mathfrak{h}_1(k, x, 0, p; \vec{0}, r) = 2a_k^{(p)}(x)$$

and putting $x = 0$ into the above equation,

$${}_{\mathbb{F}}\mathfrak{h}_1(k, 0, 0, p; \vec{0}, r) := 2a_k^{(p)},$$

where $a_k^{(p)}(x)$ and $a_k^{(p)}$ refer to the p th ordered Fubini type polynomials and numbers (Kilar & Simsek, 2017). We note that the Fubini type polynomials and numbers play a crucial role in combinatorics, probability, statistics, and number theory. Furthermore, these numbers and polynomials have comprehensive applications in fields such as mathematical physics, engineering, and optimization problems. Many researchers have extensively studied these numbers and polynomials using techniques including approximations, operators, special functions, and generating functions, and also they found many interesting results (see Agyuz, 2024; Ali & Paris, 2022; Diagana & Maïga, 2017; Kereskényi-Balogh & Nyul, 2021; Kilar, 2023b; Kilar & Simsek, 2017; Srivastava et al., 2021).

2. MAIN RESULTS

Here, using generating functions of the r -parametric Hermite-based Milne-Thomson type polynomials, novel formulas and identities for these type polynomials are obtained. These relations include certain special polynomials and numbers mentioned in the section above, such as the polynomials $a_k^{(p)}(x)$, the numbers $a_k^{(p)}$, the secant and cosecant numbers, the polynomials $S_p^k(x; \lambda)$, the numbers $S_2(k, p; \lambda)$, and the numbers $S_{22}(k, p)$.

Theorem 2.1. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$$C_k(\vec{u}, y; r) = 2^{p-1}(2p)! \sum_{j=0}^k \binom{k}{j} S_{2p}^j\left(-x; \frac{1}{2}\right) {}_{\mathbb{F}}\mathfrak{h}_1(k-j, x, y, p; \vec{u}, r). \quad (16)$$

Proof. From (14), we have

$$\exp\left(\sum_{j=1}^r u_j v^j\right) \cos(yv) = \frac{2^{2p} \left(\frac{1}{2} \exp(v) - 1\right)^{2p}}{2^{p+1}} \exp(-xv) \sum_{k=0}^{\infty} {}_{\mathbb{F}}\mathfrak{h}_1(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Merging the above equation with (5) and (8), we derive

$$\sum_{k=0}^{\infty} C_k(\vec{u}, y; r) \frac{v^k}{k!} = 2^{p-1}(2p)! \sum_{k=0}^{\infty} S_{2p}^k\left(-x; \frac{1}{2}\right) \frac{v^k}{k!} \sum_{k=0}^{\infty} {}_{\mathbb{F}}\mathfrak{h}_1(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Thus

$$\sum_{k=0}^{\infty} C_k(\vec{u}, y; r) \frac{v^k}{k!} = 2^{p-1}(2p)! \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} S_{2p}^j\left(-x; \frac{1}{2}\right) {}_{\mathbb{F}}\mathfrak{h}_1(k-j, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Equating the terms of $\frac{v^k}{k!}$ on each side of the equation mentioned above yields Eq. (16).

Substituting $x = 0$ into (16) yields the below formula:

Corollary 2.2. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$$C_k(\vec{u}, y; r) = 2^{p-1}(2p)! \sum_{j=0}^k \binom{k}{j} S_2\left(j, 2p; \frac{1}{2}\right) {}_{\mathbb{F}}\hbar_1(k-j, 0, y, p; \vec{u}, r).$$

Remark 2.3. Substituting $y = 0$ and $\vec{u} = (0, 0, \dots, 0)$ into (16), after some calculations, for $k \in \mathbb{N}$, we get

$$0 = \sum_{j=0}^k \binom{k}{j} S_{2p}^j\left(-x; \frac{1}{2}\right) a_{k-j}^{(p)}(x)$$

(Kilar, 2023b).

Theorem 2.4. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$$S_k(\vec{u}, y; r) = 2^{p-1}(2p)! \sum_{j=0}^k \binom{k}{j} S_{2p}^j\left(-x; \frac{1}{2}\right) {}_{\mathbb{F}}\hbar_2(k-j, x, y, p; \vec{u}, r). \quad (17)$$

Proof. By aid of (15), we can write

$$\exp\left(\sum_{j=1}^r u_j v^j\right) \sin(yv) = 2^{p-1} \left(\frac{1}{2} \exp(v) - 1\right)^{2p} \exp(-xv) \sum_{k=0}^{\infty} {}_{\mathbb{F}}\hbar_2(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Merging the above equation with (5) and (9), we derive

$$\sum_{k=0}^{\infty} S_k(\vec{u}, y; r) \frac{v^k}{k!} = 2^{p-1}(2p)! \sum_{k=0}^{\infty} S_{2p}^k\left(-x; \frac{1}{2}\right) \frac{v^k}{k!} \sum_{k=0}^{\infty} {}_{\mathbb{F}}\hbar_2(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Thus

$$\sum_{k=0}^{\infty} S_k(\vec{u}, y; r) \frac{v^k}{k!} = 2^{p-1}(2p)! \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} S_{2p}^j\left(-x; \frac{1}{2}\right) {}_{\mathbb{F}}\hbar_2(k-j, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Equating the terms of $\frac{v^k}{k!}$ on each side of the equation mentioned above yields Eq. (17).

When $x = 0$ in (17) yields the below formula:

Corollary 2.5. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$$S_k(\vec{u}, y; r) = 2^{p-1}(2p)! \sum_{j=0}^k \binom{k}{j} {}_{\mathbb{F}}\hbar_2(k-j, 0, y, p; \vec{u}, r) S_2\left(j, 2p; \frac{1}{2}\right).$$

Theorem 2.6. For $\vec{t} = (x + u_1, u_2, u_3, \dots, u_r)$, $\vec{u} = (u_1, u_2, u_3, \dots, u_r)$ and $k \in \mathbb{N}$, we have

$$C_k(\vec{t}, y; r) = 2^{-p-1} \sum_{j=0}^{2p} (2p)_j \sum_{q=0}^k \sum_{m=0}^q \binom{q}{m} \binom{k}{q} (-1)^{q-m+j} (2p-j)_{q-m} S_{22}(m, j) {}_{\mathbb{F}}\mathfrak{h}_1(k-q, x, y, p; \vec{u}, r).$$

Proof. Using (14), we get

$$\exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \cos(yv) = \frac{(\exp(v) - 2)^{2p}}{2^{p+1}} \sum_{k=0}^{\infty} {}_{\mathbb{F}}\mathfrak{h}_1(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Merging the above equation with (8), we have

$$\sum_{k=0}^{\infty} C_k(\vec{t}, y; r) \frac{v^k}{k!} = \frac{1}{2^{p+1}} \sum_{j=0}^{2p} \binom{2p}{j} (\exp(v) - v - 1)^j (v - 1)^{2p-j} \sum_{k=0}^{\infty} {}_{\mathbb{F}}\mathfrak{h}_1(k, x, y, p; \vec{u}, r) \frac{v^k}{k!},$$

where $\vec{t} = (x + u_1, u_2, u_3, \dots, u_r)$. Combining the above equation with (7), we obtain

$$\sum_{k=0}^{\infty} C_k(\vec{t}, y; r) \frac{v^k}{k!} = \frac{1}{2^{p+1}} \sum_{j=0}^{2p} \binom{2p}{j} j! \sum_{k=0}^{\infty} S_{22}(k, j) \frac{v^k}{k!} \sum_{k=0}^{\infty} (-1)^{k+j} (2p-j)_k \frac{v^k}{k!} \sum_{k=0}^{\infty} {}_{\mathbb{F}}\mathfrak{h}_1(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Thus

$$\begin{aligned} & \sum_{k=0}^{\infty} C_k(\vec{t}, y; r) \frac{v^k}{k!} \\ &= \frac{1}{2^{p+1}} \sum_{j=0}^{2p} (2p)_j \sum_{k=0}^{\infty} \sum_{q=0}^k \sum_{m=0}^q \binom{q}{m} \binom{k}{q} (-1)^{q-m+j} (2p-j)_{q-m} S_{22}(m, j) {}_{\mathbb{F}}\mathfrak{h}_1(k-q, x, y, p; \vec{u}, r) \frac{v^k}{k!}. \end{aligned}$$

Equating the terms of $\frac{v^k}{k!}$ on each side of the equation mentioned above yields the result.

Theorem 2.7. For $k \in \mathbb{N}$, $\vec{t} = (x + u_1, u_2, u_3, \dots, u_r)$ and $\vec{u} = (u_1, u_2, u_3, \dots, u_r)$, we have

$$S_k(\vec{t}, y; r) = 2^{-p-1} \sum_{j=0}^{2p} (2p)_j \sum_{q=0}^k \sum_{m=0}^q \binom{q}{m} \binom{k}{q} (-1)^{q-m+j} (2p-j)_{q-m} {}_{\mathbb{F}}\mathfrak{h}_2(k-q, x, y, p; \vec{u}, r) S_{22}(m, j).$$

Proof. From (15), we get

$$\exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \sin(yv) = \frac{(\exp(v) - 2)^{2p}}{2^{p+1}} \sum_{k=0}^{\infty} {}_{\mathbb{F}}\mathfrak{h}_2(k, x, y, p; \vec{u}, r) \frac{v^k}{k!}.$$

Merging the previous equation with (9) and (7) yields

$$\sum_{k=0}^{\infty} S_k(\vec{t}, y; r) \frac{v^k}{k!} = \frac{1}{2^{p+1}} \sum_{j=0}^{2p} \binom{2p}{j} j! \sum_{k=0}^{\infty} S_{22}(k, j) \frac{v^k}{k!} \sum_{k=0}^{\infty} (-1)^{k+j} (2p-j)_k \frac{v^k}{k!} \sum_{k=0}^{\infty} {}_F \hat{h}_2(k, x, y, p; \vec{u}, r) \frac{v^k}{k!},$$

where $\vec{t} = (x + u_1, u_2, u_3, \dots, u_r)$. Hence

$$\begin{aligned} & \sum_{k=0}^{\infty} S_k(\vec{t}, y; r) \frac{v^k}{k!} \\ &= \frac{1}{2^{p+1}} \sum_{j=0}^{2p} (2p)_j \sum_{k=0}^{\infty} \sum_{q=0}^k \sum_{m=0}^q \binom{q}{m} \binom{k}{q} (-1)^{q-m+j} (2p-j)_{q-m} {}_F \hat{h}_2(k-q, x, y, p; \vec{u}, r) S_{22}(m, j) \frac{v^k}{k!}. \end{aligned}$$

Comparing the terms of $\frac{v^k}{k!}$ on each side of the equation mentioned above yields the result.

Theorem 2.8. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$$\hat{h}_1(k, x, y, p; \vec{u}, r, a, b) = \sum_{j=0}^{\lfloor \frac{k+1}{2} \rfloor} \binom{k+1}{2j} \frac{y^{2j-1}}{k+1} \hat{h}_2(1-2j+k, x, y, p; \vec{u}, r, a, b) \frac{(-1)^j}{2} d_{2j}. \tag{18}$$

Proof. From (13), we derive

$$2(b + f(v, a))^p \exp(xv) \exp\left(\sum_{j=1}^r u_j v^j\right) \sin(2yv) = \sum_{k=0}^{\infty} \hat{h}_2(k, x, 2y, p; \vec{u}, r, a, b) \frac{v^k}{k!}.$$

Using the above equation and (12), we have

$$\sum_{k=0}^{\infty} \hat{h}_1(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!} = \frac{1}{2 \sin(yv)} \sum_{k=0}^{\infty} \hat{h}_2(k, x, 2y, p; \vec{u}, r, a, b) \frac{v^k}{k!}.$$

Joining the above equation with (3) yields

$$\sum_{k=0}^{\infty} \hat{h}_1(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!} = \frac{1}{2yv} \sum_{k=0}^{\infty} (-1)^k y^{2k} d_{2k} \frac{v^{2k}}{(2k)!} \sum_{k=0}^{\infty} \hat{h}_2(k, x, 2y, p; \vec{u}, r, a, b) \frac{v^k}{k!}.$$

After some performing calculations, we have

$$\sum_{k=0}^{\infty} \hat{h}_1(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!} = \sum_{k=0}^{\infty} \sum_{j=0}^{\lfloor \frac{k+1}{2} \rfloor} \binom{k+1}{2j} \frac{y^{2j-1} (-1)^j}{2(k+1)} \hat{h}_2(1-2j+k, x, 2y, p; \vec{u}, r, a, b) d_{2j} \frac{v^k}{k!}.$$

Comparing the terms of $\frac{v^k}{k!}$ on each side of the equation mentioned above yields Eq. (18).

Theorem 2.9. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$$\hbar_2(k, x, y, p; \vec{u}, r, a, b) = \frac{1}{2} \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{2j} \hbar_2(k-2j, x, 2y, p; \vec{u}, r, a, b) y^{2j} s_j. \quad (19)$$

Proof. From (13), we derive

$$\sum_{k=0}^{\infty} \hbar_2(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!} = \frac{1}{2 \cos(yv)} \sum_{k=0}^{\infty} \hbar_2(k, x, 2y, p; \vec{u}, r, a, b) \frac{v^k}{k!}. \quad (20)$$

Joining (20) with (1), some performing calculations, we get

$$2 \sum_{k=0}^{\infty} \hbar_2(k, x, y, p; \vec{u}, r, a, b) \frac{v^k}{k!} = \sum_{k=0}^{\infty} \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{2j} \hbar_2(k-2j, x, 2y, p; \vec{u}, r, a, b) y^{2j} \frac{s_j}{k!} v^k.$$

Comparing the terms of $\frac{v^k}{k!}$ on each side of the equation mentioned above yields Eq. (19).

When $b = 0$ and $f(v, 1) = \frac{2}{(2 - \exp(v))^2}$ in (18) and (19) yields the below formula:

Corollary 2.10. For $k \in \mathbb{N}_0$ and $\vec{u} = (u_1, u_2, \dots, u_r)$, we have

$${}_{\mathbb{F}}\hbar_1(k, x, y, p; \vec{u}, r) = \sum_{j=0}^{\lfloor \frac{k+1}{2} \rfloor} \binom{k+1}{2j} \frac{y^{2j-1}}{k+1} {}_{\mathbb{F}}\hbar_2(1-2j+k, x, 2y, p; \vec{u}, r) \frac{(-1)^j}{2} d_{2j}$$

and

$${}_{\mathbb{F}}\hbar_2(k, x, y, p; \vec{u}, r) = \frac{1}{2} \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{2j} {}_{\mathbb{F}}\hbar_2(k-2j, x, 2y, p; \vec{u}, r) y^{2j} s_j.$$

3. CONCLUSION

In this paper, by aid of generating function methods, the r -parametric Hermite-based Milne-Thomson polynomials were investigated. By these functions and trigonometric functions, many formulas and identities pertaining to the Stirling type numbers and polynomials, the r -parametric Hermite-based sine-and cosine-Milne-Thomson Fubini polynomials, and the secant and cosecant numbers were obtained. These types of polynomials, Fubini and Hermite polynomials, have significant potential for applications in various fields such as combinatorics, probability theory, statistics, and number theory. Furthermore, the solutions offered by these polynomials are not limited to pure mathematics but can also be applied to solve real-world problems encountered in engineering, physics, and other areas. Therefore, the results presented in this paper provide valuable contributions to both theoretical mathematics and applied sciences, offering a broad range of benefits across many areas.

In the future, we plan to investigate the relations among r -parametric Hermite-based sine-and cosine-Milne-Thomson Fubini polynomials, determinantal expressions, Faa di Bruno formula, and other special numbers and polynomials, as well as their applications in differential equations and mathematical models.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- Agyuz, E. (2024). On convergence properties of Fubini-type polynomials. *AIP Conference Proceedings*, 3094(090012). <https://doi.org/10.1063/5.0210608>
- Ali, M., & Paris, R. B. (2022). Multi-index Fubini-type polynomials. *Montes Taurus Journal of Pure and Applied Mathematics*, 4(1), 97-106.
- Bayad, A., & Simsek, Y. (2014). Convolution identities on the Apostol-Hermite base of two variables polynomials. *Differential Equations and Dynamical Systems*, 22(3), 309-318. <https://doi.org/10.1007/s12591-013-0181-7>
- Cesarano, C., Ramírez, W., & Khan, S. (2022). A new class of degenerate Apostol-type Hermite polynomials and applications. *Dolomites Research Notes on Approximation*, 15(1), 1-10. <https://doi.org/10.14658/PUPJ-DRNA-2022-1-1>
- Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*. Hoboken: John Wiley & Sons Inc., Publication.
- Dattoli, G., Chiccoli, C., Lorenzutta, S., Maino, G., & Torre, A. (1994). Theory of generalized Hermite polynomials. *Computers & Mathematics with Applications*, 28(4), 71-83. [https://doi.org/10.1016/0898-1221\(94\)00128-6](https://doi.org/10.1016/0898-1221(94)00128-6)
- Dattoli, G., Lorenzutta, S., Maino, G., Torre, A., & Cesarano, C. (1996). Generalized Hermite polynomials and supergaussian forms. *Journal of Mathematical Analysis and Applications*, 203, 597-609. <https://doi.org/10.1006/jmaa.1996.0399>
- Diagana, T., & Maïga, H. (2017). Some new identities and congruences for Fubini numbers. *Journal of Number Theory*, 173, 547-569. <https://doi.org/10.1016/j.jnt.2016.09.032>
- Fadel, M., Raza, N., & Du, W.-S. (2024). On q -Hermite polynomials with three variables: Recurrence relations, q -differential equations, summation and operational formulas. *Symmetry*, 16(4), 1-19. <https://doi.org/10.3390/sym16040385>
- Kereskényi-Balogh, Z., & Nyul, G. (2021). Fubini numbers and polynomials of graphs. *Mediterranean Journal of Mathematics*, 18(230), 1-10. <https://doi.org/10.1007/s00009-021-01838-x>
- Kilar, N. (2021). *Generating Functions of Hermite Type Milne-Thomson Polynomials and Their Applications in Computational Sciences*. Antalya: PhD Thesis, University of Akdeniz.
- Kilar, N. (2023a). On computational formulas for parametric type polynomials and its applications. *Journal of Balikesir University Institute of Science and Technology*, 25(1), 13-30. <https://doi.org/10.25092/baunfbed.1083754>
- Kilar, N. (2023b). Combinatorial sums and identities associated with functional equations of generating functions for Fubini type polynomials. *GUJ Sci*, 36(2), 807-817. <https://doi.org/10.35378/gujs.989270>
- Kilar, N., & Simsek, Y. (2017). A new family of Fubini numbers and polynomials associated with Apostol-Bernoulli numbers and polynomials. *Journal of the Korean Mathematical Society*, 54(5), 1605-1621. <https://doi.org/10.4134/JKMS.j160597>
- Kilar, N., & Simsek, Y. (2021). Computational formulas and identities for new classes of Hermite-Based Milne-Thomson type polynomials: Analysis of generating functions with Euler's formula. *Mathematical Methods in the Applied Sciences*, 44(8), 6731-6762. <https://doi.org/10.1002/mma.7220>
- Kim, D. S., & Kim, T. (2018). Some p -adic integrals on \mathbb{Z}_p associated with trigonometric functions. *Russian Journal of Mathematical Physics*, 25(3), 300-308. <https://doi.org/10.1134/S1061920818030032>

- Simsek, Y. (2013). Generating functions for generalized Stirling type numbers, Array type polynomials, Eulerian type polynomials and their applications. *Fixed Point Theory Applications*, 2013(87), 1-28. <https://doi.org/10.1186/1687-1812-2013-87>
- Simsek, Y. (2024). Formulas for p -adic q -integrals including falling-rising factorials, combinatorial sums and special numbers. *RACSAM*, 118(92), 1-52. <https://doi.org/10.1007/s13398-024-01592-1>
- Srivastava, H. M. (1976). A note on a generating function for the generalized Hermite polynomials. *Indagationes Mathematicae (Proceedings)*, 79(5), 457-461. [https://doi.org/10.1016/S1385-7258\(76\)80009-1](https://doi.org/10.1016/S1385-7258(76)80009-1)
- Srivastava, H. M., Srivastava, R., Muhyi, A., Yasmin G., Islahi H., & Araci, S. (2021). Construction of a new family of Fubini-type polynomials and its applications. *Advances in Difference Equations*, 2021(36), 1-25. <https://doi.org/10.1186/s13662-020-03202-x>
- Zayed, M., Wani, S. A., Oros, G. I., & Ramírez, W. (2024). A study on extended form of multivariable Hermite-Apostol type Frobenius-Euler polynomials via fractional operators. *AIMS Mathematics*, 9(6), 16297-16312. <https://doi.org/10.3934/math.2024789>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1526979>

Drone Detection Performance Evaluation via Real Experiments with Additional Synthetic Darkness

Furkan ORUÇ¹ H. Birkan YILMAZ^{1*} ¹ Boğaziçi University, Department of Computer Engineering, İstanbul, Türkiye

Keywords	Abstract
Drone Detection Yolo LSTM Vision Transformers Darkness	Detecting drones is increasingly challenging, particularly when developing passive and low-cost defense systems capable of countering malicious attacks in environments with high levels of darkness and severe weather conditions. This research addresses the problem of drone detection under varying darkness levels by conducting an extensive study using deep learning models. Specifically, the study evaluates the performance of three advanced models: Yolov8, Vision Transformers (ViT), and Long Short-Term Memory (LSTM) networks. The primary focus is on how these models perform under synthetic darkness conditions, ranging from 20% to 80%, using a composite dataset (CONNECT-M) that simulates nighttime scenarios. The methodology involves applying transfer learning to enhance the base models, creating Yolov8-T, ViT-T, and LSTM-T variants. These models are then tested across multiple datasets with varying darkness levels. The results reveal that all models experience a decline in performance as darkness increases, as measured by Precision-Recall and ROC Curves. However, the transfer learning-enhanced models consistently outperform their original counterparts. Notably, Yolov8-T demonstrates the most robust performance, maintaining higher accuracy across all darkness levels. Despite the general decline in performance with increasing darkness, each model achieves an accuracy above 0.6 for data subjected to 60% or greater darkness. The findings highlight the challenges of drone detection under low-light conditions and emphasize the effectiveness of transfer learning in improving model resilience. The research suggests further exploration into multi-modal systems that combine audio and optical methods to enhance detection capabilities in diverse environmental settings.

Cite

Oruç, F., & Yılmaz, H. B. (2024). Drone Detection Performance Evaluation via Real Experiments with Additional Synthetic Darkness. *GU J Sci, Part A, 11(3)*, 546-562. doi:10.54287/guj.1526979

Author ID (ORCID Number)	Article Process
0000-0001-6866-5285	Furkan ORUÇ
0000-0002-4773-2028	H. Birkan YILMAZ
	Submission Date 02.08.2024
	Revision Date 13.08.2024
	Accepted Date 19.09.2024
	Published Date 30.09.2024

1. INTRODUCTION

Drones are widely used unmanned aerial vehicles for defense systems and object detection technologies. Adam (2020) claim that the use of drone network applications has significantly increased. Drones serve various roles, including acting as sensors, providing network services, and facilitating delivery services, among other applications. We can observe and utilize drones in various ways, such as size and shape. Thus, there is always a requirement for such a technological system for drone detection (Moustafa et al., 2017). Traditional detection methods, such as radar and RF-based systems, often struggle with small, low-flying drones, particularly in cluttered or dark environments. This study's focus on enhancing drone detection models with transfer learning under varying levels of synthetic darkness is thus both timely and relevant.

Detecting drones in an accurate manner is crucial for preventing threats and ensuring public safety. Traditional radar and radio frequency (RF) detection methods, while effective in some scenarios, often struggle with the small size, low altitude, and low radar cross-section of modern drones (Moustafa et al., 2017). These limitations

*Corresponding Author, e-mail: birkan.yilmaz@bogazici.edu.tr

have contributed to the development of alternative detection techniques, particularly those based on computer vision and machine learning, which offer the potential for more precise and adaptable detection systems.

On the other hand, defense systems may observe malicious drone attacks that might cause vulnerabilities that might even lead to a war. To be able to handle this problem, the detection and surveillance of small drones under severe weather conditions are quite necessary. Drone detection systems are highly popular nowadays. Thus, analyzing and understanding when and how an unmanned aerial vehicle is approaching any target is crucial. As Moustafa and Jolfaei (2020) state, there might be developed a drone intrusion system with information sources, communication links, and autonomous control basis components. For this work, we would like to focus on information sources on the vision dataset among these components.

In recent years, models such as YOLO (You Only Look Once), Vision Transformers (ViT), and Long Short-Term Memory (LSTM) networks have shown significant impact in object detection tasks, including drone detection. These models leverage large datasets and complex architectures to learn and recognize patterns in visual data, making them well-suited for detecting drones in diverse conditions. However, the performance of these models can still be significantly impacted by factors such as darkness, which can obscure visual features and reduce detection accuracy.

This work will focus on the results and methods of drone detection using different models on a certain level of darkness. We also conducted our methodology by gathering real data as an experimental setup. After the collection part, we applied darkness procedures and compared the results in a precise manner.

The structure contains an introduction, literature review, and related work, our setup for experiments, results, and discussion. The literature review is about previous work done on drone detection and model performance comparison on dark images. In addition, the experimental setup talks about what equipment we used during our experiments, such as cameras, drones, and other electronic devices.

The Methodology section briefly is about model comparison and how we added darkness. After that, YOLOv8, LSTM (long-short-term memory), and vision transformers are discussed. The results section explains the detection performance comparison and decrease-increase among different levels of darkness using Precision and Recall Curves and ROC (Receiver operating characteristic) Curves.

Experimental Results section presents the results of the experiments, comparing the performance of the original models against their transfer learning-enhanced versions across different levels of darkness. It includes detailed discussions of the models' Precision-Recall and ROC curves and the impact of darkness on detection accuracy.

Discussion section interprets the results, discussing the implications of the findings for drone detection in real-world scenarios. It also explores the strengths and limitations of the models tested, with particular attention to how transfer learning improves performance under challenging conditions.

Finally, our paper concludes by summarizing the key findings, emphasizing the advantages of transfer learning for drone detection under varying levels of darkness, and suggesting directions for future research, including the potential for multi-modal detection systems.

One of the novelties of the approach is the impact varying levels of synthetic darkness (20%, 40%, 60%, 80%) on drone detection performance, using models such as YOLOv8-T, ViT-T (Vision Transformer with Transfer Learning), and LSTM-T (Long-Short Term Memory with Transfer Learning). Those methods will be mentioned in the methodology section. This approach will address a significant gap in the literature where the gradual impact of darkness on detection models is less explored. Additionally, we have a comprehensive dataset, which includes a mix of darkness levels providing more robust and challenging test environments as compared to other datasets used in previous studies. Transfer learning enhancements also provide an optimized pre-trained network for drone detection under dark conditions, which improves performance in scenarios where training data is limited or highly variable. Lastly, an insight to our model behavior such as CONNECT-M (which will be defined in methodology section later) challenges models more comprehensively than single-condition datasets, leading to a better understanding of model generalizability and robustness.

Lastly, This study fills this research gap by systematically evaluating the performance of state-of-the-art drone detection models—YOLOv8, Vision Transformers (ViT), and Long Short-Term Memory (LSTM) networks—under varying levels of synthetic darkness. By using experimental data and applying transfer learning to these models (resulting in YOLOv8-T, ViT-T, and LSTM-T), the study investigates whether pre-trained models can be fine-tuned to improve performance in low-light conditions, which will be addressing the current limitations of drone detection systems.

2. LITERATURE REVIEW

Khan et al. (2023) proposes a drone detection system, which is GAANet (Ghost auto anchor network) Using Yolov5. The images are under a high level of darkness at night and low visibility. They also implemented an auto anchor calculation for their model architecture for adding and removing convolutions. As compared to Yolov5, GAANet experiences overall precision and recall around 2.3 and 1.4 percent more respectively.

As Misbah et al. (2023) mentioned, we can observe that TF-net which is defined as TensorFlow network by authors is another approach that focuses on drone detection under nighttime. They propose a tiny feature network, which contains detection of unmanned aerial vehicles under night vision based on infrared images. TF-Net also focuses on detection with complex background images. Misbah et al. (2023) work with four different Yolo algorithms with different hyperparameters. Their conclusion shows that 95.7 precision results with 84 percent of mean absolute precision. Lastly, their detection threshold for intersection over union is 44.8 percent to consider the frame as true positive.

Not only darkness addition, but also severe weather conditions such as rainy, foggy, and stormy weather are important to make analysis for drone detection, which will provide challenges to the dataset. Methods of the authors of Munir et al. (2024) mention about the drone dataset with a complex and severe condition. Methods for benchmarks are Yolov8, Yolov5, and Faster R-CNN (Region based convolutional neural network) respectively. It is shown that Faster-RCNN and Yolov5 perform better when it comes to detecting unmanned aerial vehicle image frames with rainy conditions. In conclusion of their research, they claim average precision of their models for Yolov5, Yolov8, and Faster R-CNN are 69.3, 67.2, and 59.2 respectively

Yi et al. (2019) propose a drone detection and classification approach which will separate drones from birds or different backgrounds. Yolov2 is used as an object detection model. However, they also tried transfer learning from ImageNet under low level of light with additional data augmentation. They called this approach Yolov2 DarkNet. Lastly, average intersection and over union and recall are analyzed and compared as evaluation metrics. Andraši et al. (2017) defined their detection system as Unmanned Aerial Vehicles (UAVs) detection for safety and security. They challenged their image dataset with low visibility, urban environments, and the night. Their approach proposes a thermal infrared camera object detection. At the end, they claim that they tested the applicability of low cost long-wave infrared cameras for various examples of swarm Unmanned Aerial Vehicles in flight.

Svanström et al. (2022) developed an automatic detection of flying drones. However, they combine thermal infrared cameras and microphone sensors for multi-sensor drone detection. The solution of Svanström et al. (2022) also integrates an implementation of ADS-B receiver and a GPS receiver. However, the detection range is limited for certain distances. Their claim shows that the detection system can be more efficient in low-level of light and more robust to mitigate false detection with the fusion of audio and vision modalities.

Another tiny object detection method is proposed by Zhai et al. (2023). They developed the detection algorithm by Yolov8 similar to ours. First, a high-resolution detection head is added for small targets. Then, redundant network layers are cut to mitigate network parameters and improve detection performance. Thirdly, preprocessing steps such as multi-scale extraction were applied with SPD-Convolution. As compared to the baseline model of Yolov8, their method improved by 11.9 percent, 15.2 percent, and 9 percent in terms of precision, recall, and mean average precision. In addition, the deployment of their system indicates that the number of parameters decreased by 59.9 percent.

Ramadan et al. (2021) propose a novel approach of drone detection technology called Ad Hoc Network (FANET). It involves frameworks such as Recurrent Neural Networks (RNN) as a base. It includes gathering

network data and using big data analytics to detect drone anomalies. The data collection part of this procedure is working inside of each Ad Hoc Network for intrusion drone detection. They claim that they set extensive ways of experiments based on various datasets to examine the efficiency of the proposed framework. Their results show that FANET outperforms RNN as a baseline detection model.

Finally, Jamil et al. (2022) built a drone detection system for malicious drone attacks using vision transformers, which is one of our models for this work. They proposed a vision transformer based on a framework with drone image dataset into fixed-size patches. In addition, linear embeddings and position embeddings are applied. Their proposed work is compared to deep convolutional neural networks (D-CNN) which revealed that their proposed model has experienced an accuracy of 98.3 percent overall.

3. MATERIAL AND METHOD

3.1. Setup Architecture

Figure 1 shows the rectangular area we used for our experiments. In addition, locations are calculated. The distance between first and last lines is 2800 cm, and each line is 700 cm. In addition, a line (drain) is detected through the right corner of the area as an inner side. The setup takes place as approximately 410cm with the drone flight experiments.

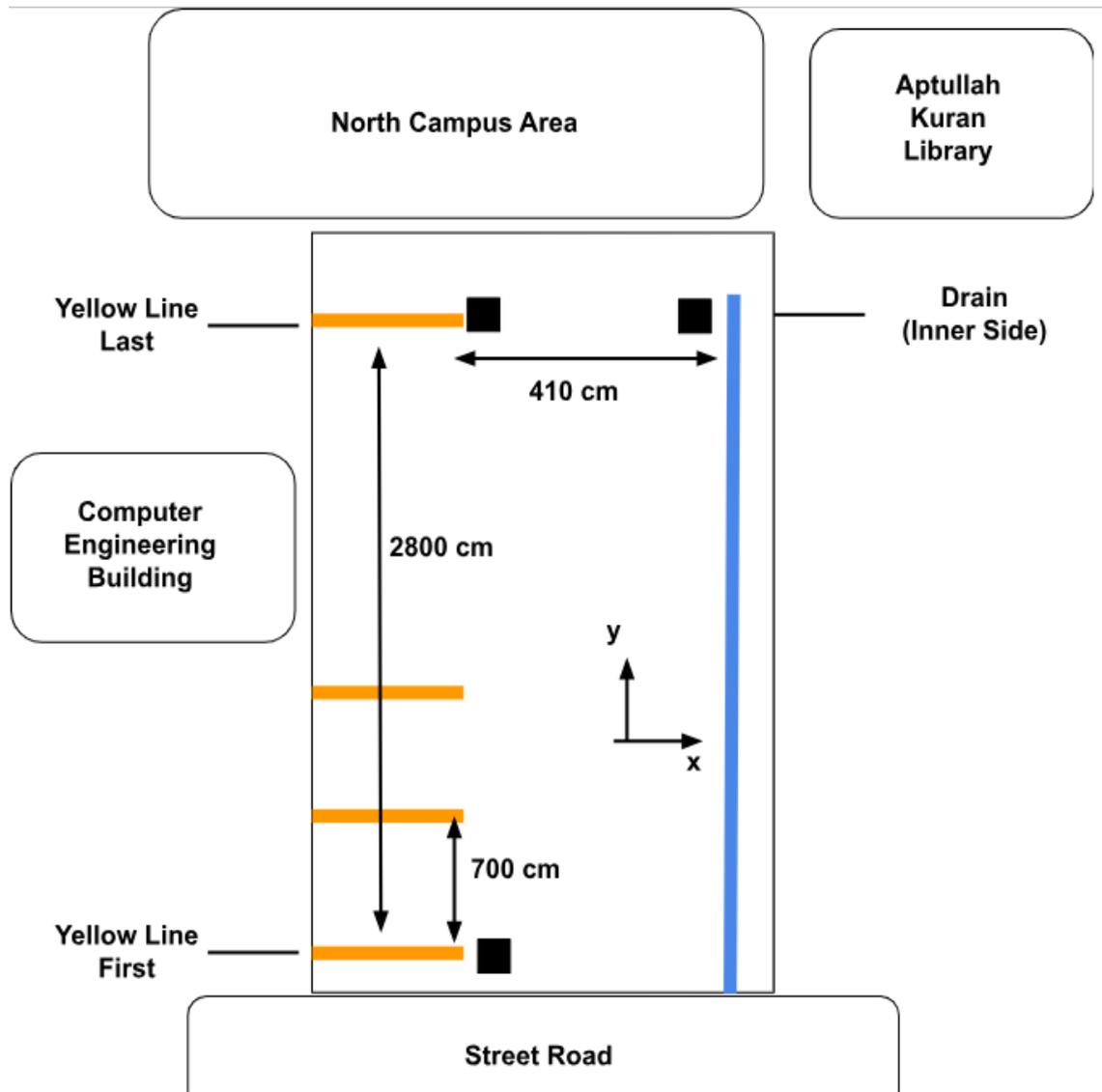


Figure 1. Schematic map of the experimental area

The first and last yellow line coordinates as latitudes and longitudes are demonstrated in Table 1. Also, latitude and longitude coordinates of the Inner side are given.

Figure 2 indicates our setup with different modalities. However, we will only focus on the vision modality part in this approach. There are parallel-located cameras for vision modality. One computer is designed as a controller. Two others are showing the footage for left and right cameras simultaneously. The closest point between the left and right tripods is 87 cm. In addition to setup and data collection, Dell Precision t3600, 64 GiB, NVIDIA GA102GL RTX A5000 is used for data analysis, learning, visualization and testing. The following features of the setup are as follows:

- Processor: 12th Gen Intel Core i9-12900 with 24 cores, providing the necessary computational power for real-time data processing and model training.
- Disk Capacity: 4.5TB of disk capacity, ensuring sufficient space for storing high-resolution images, video data, and trained model weights.
- OS Name and Type: Ubuntu 22.04.3 LTS, 64-bit,
- GNOME Version: 42.9,
- GPU: NVIDIA GA102GL RTX A5000, a high-performance graphics card essential for deep learning tasks, enabling the acceleration of model training and inference processes.

Table 1. Inner Side and Yellow Line Coordinates

Label	Latitude	Longitude
Yellow Line - Last	41.0857753	29.0437393
Yellow Line - First	41.08570841	29.04412709
Drain (Inner Side)	41.08581119	29.04373023

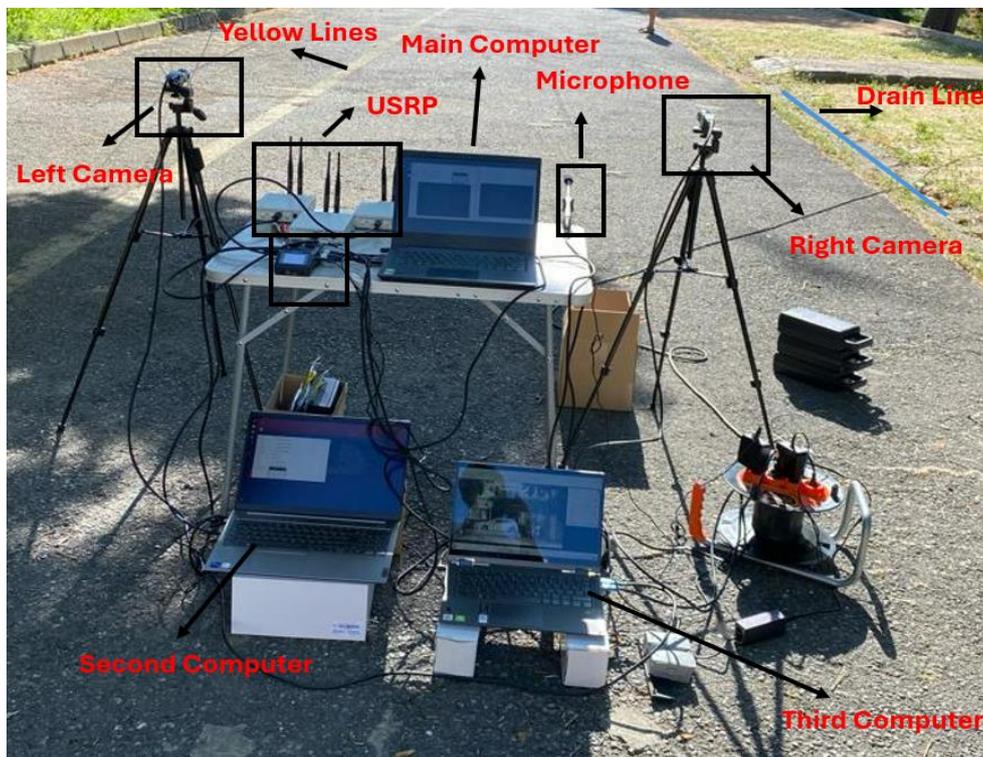


Figure 2. Designed setup outside

Figure 3 shows the industrial cameras used to obtain drone footage for image dataset. These cameras are placed on the right and the left sides of the table of the setup. Further features of our cameras are as follows: USB 3.0 Interface, 2/3-inch Sony CMOS Pregius IMX250 Sensor, 2,448×2,048 (5 MP) Resolution, Global Shutter, Trigger and I/O Inputs, 29x29x43 mm width, height, and length.



Figure 3. *a) Image source camera, b) Image source camera with fixing device*

The model and several of the drone flew during the experiment is shown in Figure 4. DJI Mavic Air 2 is used with its control device and four propellers. Device controller has three options: Cine, Normal and Sport 5 modes. We mostly completed our experiments in normal mode. In addition, the control device is used for up, down, left, right, back, and forward moves. Weight of the drone is 570g and the diagonal distance is 302mm. There is also a gimbal with 3-axis (tilt, roll, and pan) which we did not use and record any frame from the drone's point of view. Further features of our drone for experiments are as follows:

- Battery Capacity: 2500 mAh,
- Max Flight Time: 34 minutes,
- Control Modes: The drone was primarily operated in Normal Mode for the experiments, although Cine and Sport modes were also available on the controller.
- Angular Velocity: 250 °/s (N Mode),
- Propellers: Quick release, low noise, folding,
- Gimbal: A 3-axis (tilt, roll, and pan) gimbal was included, although it was not utilized in this study as the focus was on the drone's external detection rather than its onboard camera.



Figure 4. *a) DJI Mavic Air 2 drone, b) Controller of DJI Mavic Air 2 drone*

3.2. Dataset and Structure

The dataset consists of 28 .avi videos having the duration ranging between 5-20 minutes with 30 fps and in total 518 700 frames. For each experiment, two videos were created at the same time, both left and right cameras were recording during the process from different angles. While half of them belong to the left, the others belong to the right camera.

The images usually contain one drone in each frame with a complex background with external objects such as buildings, a gate, and trees. Figure 5 shows a randomly selected frame among the other frames of the dataset. In Figure 5, there is a drone on the left part of the gate. The dataset consists of 28 videos of which 14 of them are recorded from left, and the remaining are recorded from right camera. The video format has been recorded as AVI format then, we obtained 30 frames per second since the cameras have 30 fps settings.



Figure 5. Sample image from the dataset

Our methodology uses this data architecture with three different models: Yolov8, Vision Transformer (ViT), and Long-short term memory (LSTM). However, the detection procedure is not only evaluated in one dataset. There are other datasets with combined or single darkness addition. Entire datasets used for this approach are defined as follows:

- CONNECT: Original data without darkness,
- CONNECT-D20 Dataset: 20 percent darkness addition of entire data,
- CONNECT-D40 Dataset: 40 percent darkness addition of entire data,
- CONNECT-D60 Dataset: 60 percent darkness addition of entire data,
- CONNECT-D80 Dataset: 80 percent darkness addition of entire data.

Lastly, we defined a mixture of darkness levels as the CONNECT-M dataset. CONNECT-M dataset contains the followings:

- 30 percent of the Data: 20 percent darkness addition,
- 30 percent of the Data: 40 percent darkness addition,
- 20 percent of the Data: 60 percent darkness addition,
- 20 percent of the Data: 80 percent darkness addition.

3.3. Yolov8-T

Authors of Reis et al. (2023) propose a generalized method for real-time object detection including flying objects such as drones. This state-of-art method is called Yolov8, which we also conducted for our research. The authors provide an in-depth explanation of the new architecture and functionality that YOLOv8 has adapted. According to evaluation metrics mAP50-95 with the 50 fps 1080p videos. The algorithm gets an average of 0.835 mAP50-95 as a result.

Furthermore, Yolov8 uses Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). It is important to note that they make our annotation part easier. The first parameter of Yolo architecture is input image (Reis et al., 2023). However, PAN architecture utilized different layers from different levels as part of the network architecture (Li et al., 2023).

Li et al. (2023) also states the backbone of YOLOv8 is a deep Convolutional Neural Network (CNN) designed to extract hierarchical features from the input image. YOLOv8 typically uses CSPDarknet (Cross Stage Partial Darknet) as its backbone, which is known for its efficient feature extraction and reduction of computational complexity. YOLOv8 also utilizes a Feature Pyramid Network (FPN) combined with a Path Aggregation Network (PAN). This neck architecture is crucial for fusing feature maps from different layers, allowing the model to detect objects at multiple scales (Minderer et al., 2022).

Those approaches make the Yolov8 approach capture tiny and small objects. In addition, since we work with additional layers with parameter arrangement, our model structures for Yolov8 are defined as Yolov8-T. The term shows the yolov8 model with transfer learning applied. Following parameters have been used for model testing and parameter tuning: Input Image Size, Batch Size, Number of Epochs, Learning Rate, Anchor Boxes, Confidence Threshold. Various image sizes were tested, ranging from 320x320 to 640x640 pixels. The final choice of 416x416 was selected as it provided a balance between detection accuracy and computational efficiency, ensuring that small drones were detectable without overwhelming the GPU. Batch sizes of 8, 16, and 32 were tested. The model was trained with a batch size of 16, which was found to offer a good trade-off between stability and the ability to utilize the GPU effectively without memory overflow issues. As a tuning process, the model was initially trained for 50 epochs, with performance monitored on the validation set. Then, Learning rates were experimented with in the range of 0.001 to 0.0001. A learning rate of 0.0005 was chosen as it allowed the model to converge properly.

3.4. Vision Transformer-T (ViT-T)

Minderer et al. (2022) introduced a novel technique for detecting objects on multiple scales by integrating both textual and visual elements using Vision Transformers. This technique is officially termed "Open-Vocabulary Object Detection." It employs a conventional vision transformer that encompasses both an image encoder and a text encoder. A token pooling and projection layer is typically used in this model to derive image classification embeddings. The core of ViT consists of multiple transformer encoder layers, each containing multi-head self-attention mechanisms and feed-forward neural networks. However, in our specific application, where we focus solely on single-object scenarios, we have omitted this aspect of the model. The process begins with the utilization of a pre-trained COCO model in tandem with a text encoder, resulting in a combined image-text output. This is subsequently adapted for open-vocabulary object detection by bypassing the token pooling step in the image-encoding phase. For object detection, we have employed specific textual inputs. The final classification head is a simple Multilayer Perceptron with a softmax output, responsible for determining the class probabilities and the presence of a drone in each patch.

The output for object detection can be achieved using any one of the four options provided, as this enhances the likelihood of successfully detecting an object in a single attempt. The specific hyper parameters implemented in this method are as follows:

- focal loss with $\alpha = 0.3$ and $\gamma = 2.0$,
- Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, which improved the model's ability to adapt to different levels of darkness,

- Cosine learning rate, initial learning rates were set between 0.0001 and 0.001. A learning rate of 0.0002 was chosen,
- Text encoder input length to 16 characters,
- Dropout rates of 0.1, 0.3, and 0.5 were tested. A dropout rate of 0.3 was selected, providing a good level of regularization while maintaining model performance, especially under darker conditions where overfitting could lead to significant performance drops.

And this specific model with hyperparameters tuned and transfer learning applied is called ViT-T.

3.5. Long-short Term Memory-T (LSTM-T)

The concept of Long Short-Term Memory (LSTM) was first introduced by Hochreiter and Schmidhuber (1997). They described LSTM as a method to truncate gradients, enabling the model to minimally bridge time intervals using specialized units. Over time, contemporary LSTM models have evolved, incorporating additional features such as dropout layers, dense layers, and various parameter enhancements. In our drone detection model, we have integrated both dropout and dense layers.

The dropout layer, as explained by Goodfellow et al. (2016), serves as a regularization strategy in artificial neural networks to address the problem of overfitting. It works by randomly omitting certain output features during the training phase. Additionally, they also describe a dense layer as a fully connected layer, where every input is linked to each output through a learnable weight (Goodfellow et al., 2016). In addition, Hochreiter and Schmidhuber (1997) also propose that an input vector described as x , and the output described as y of a dense layer can be calculated as:

$$y = \phi(Wx + b) \quad (1)$$

where W is the weight matrix, b is the bias vector, ϕ is the activation function. The equation shows that the matrix multiplication ' Wx ' calculates the weighted sum of the inputs, and then biases ' b '. In addition, since we work with additional layers with parameter arrangement, our model structure for LSTM is defined as LSTM-T. The term shows that long-short term memory model with transfer learning applied.

Our LSTM-T model uses the following parameters for tuning and performance development:

- Sequence length: 9, sequence lengths of 5, 9, and 15 were tested. A sequence length of 9 was chosen, as it provided enough temporal context to detect drones moving across frames without overwhelming the model with too much data.
- Number of layers: 5, LSTM - Dropout - Dense - Dropout - Dense, models with 2, 3, and 5 layers were tested. A 3-layer LSTM network was selected,
- Activation function: ReLu for LSTM and sigmoid for Dense layer,
- Optimizer: Adam, Loss: binary cross-entropy,
- Hidden Size: 256, hidden sizes of 128, 256, and 512 units were tested. A hidden size of 256 was optimal, allowing the model to capture sufficient information,
- Number of Classes: 1,
- Learning Rate 0.0001,
- Batch Size: 10,
- Number of Epochs: 150.

3.6. Key Evaluation Metrics

We used several key metrics to evaluate the performance of drone detection models under varying conditions of darkness.

Firstly, precision is the ratio of true positive detections to the sum of true positives and false positives, which measures the accuracy of positive predictions made by the model and Hochreiter and Schmidhuber (1997) uses the precision as

$$\text{Precision} = \text{True Positives (TP)} / (\text{True Positives (TP)} + \text{False Positives (FP)}) \quad (2)$$

Also, they used the metric Recall as

$$\text{Recall} = \text{True Positives (TP)} / (\text{True Positives (TP)} + \text{False Negatives (FN)}) \quad (3)$$

In terms of ROC (Receiver Operator Characteristic) curve calculations, we used True Positive Rate (TPR) or sensitivity and False Positive Rate (FPR). TPR is defined as

$$\text{TPR} = \text{True Positives (TP)} / (\text{True Positives (TP)} + \text{False Negatives (FN)}) \quad (4)$$

Similarly, False Positive Rate (FPR) is defined as

$$\text{FPR} = \text{False Positives (FP)} / (\text{False Positives (FP)} + \text{True Negatives (TN)}) \quad (5)$$

which provides a single value representing the overall performance of the model across all classification thresholds. AUC-ROC value indicates better model performance, with 1.0 being perfect and 0.5 being equivalent to random guessing. In terms of justification, precision is crucial in applications where false positives can lead to significant consequences, such as unnecessary defensive actions or false alarms in security systems. Precision also measures true positive detections among all positive detections made by the model, indicating how many of the detected drones are actual drones rather than background noise or other objects. Recall is also chosen because a metric to be balanced to ensure results with accuracy and AUC tables. Lastly, AUC-ROC is useful when comparing different models or configurations and that is quite important in understanding the trade-offs between precision and recall.

4. EXPERIMENTAL RESULTS

The collected experimental data is expanded with synthetic darkness, and we tested our trained network on different datasets. For evaluation, we consider precision-recall curves and ROC curves. First, we need to see the differences between standard Yolo, Vit, and LSTM algorithms as compared to Yolov8-T, ViT-T, and LSTM-T in order to measure the development.

Table 2 shows the difference between three original algorithms and three transfer learning applied algorithms on CONNECT dataset.

Table 2. AUC scores of original and transfer learning applied models on CONNECT dataset

Dataset	Method	AUC (Precision-Recall)	AUC (ROC Curve)
CONNECT	Yolov8-T	0.80	0.78
CONNECT	ViT-T	0.73	0.74
CONNECT	LSTM-T	0.72	0.71
CONNECT	Yolov8	0.56	0.52
CONNECT	ViT	0.55	0.53
CONNECT	LSTM	0.59	0.56

Table 2 shows that The AUC score for Yolov8-T is 0.80, significantly higher than Yolov8's 0.56. Similarly, the AUC (ROC Curve) for Yolov8-T is 0.78 compared to Yolov8's 0.52. This clearly indicates that the application of transfer learning dramatically improves the model's ability to accurately detect drones, enhancing both precision and recall, as well as overall classification performance. In addition, ViT-T also shows a substantial improvement over ViT, with a Precision-Recall AUC of 0.73 versus 0.55 for the original

model. The ROC AUC improves from 0.53 to 0.74. These results suggest that transfer learning has also a strong positive effect on the performance of Vision Transformers in drone detection tasks. Lastly, The LSTM-T model outperforms the original LSTM model, with an AUC (Precision-Recall) of 0.72 compared to LSTM's 0.59. The AUC (ROC Curve) shows a similar improvement, with LSTM-T scoring 0.71 versus LSTM's 0.56. This improvement underscores the benefit of transfer learning in recurrent neural networks like LSTM for this application. There is an improvement on the CONNECT Dataset with all three different methods. Therefore, we are going to apply the same logic to different levels of darkness with added datasets.

4.1. Results of CONNECT Dataset

Figure 6 shows the results of Yolov8-T, ViT-T, and LSTM-T according to Precision-Recall Curves and ROC Curves. The results are 0.80, 0.73, and 0.72 for the Precision Recall area under the curve.

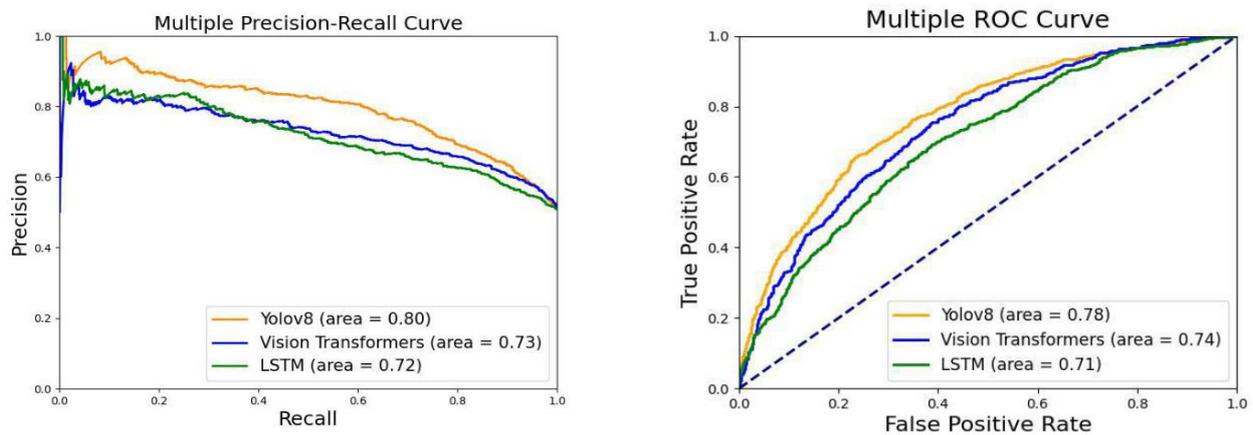


Figure 6. Precision-Recall and ROC curves under no additional darkness.

4.2. Results of CONNECT-D20 Dataset

Figure 7 indicates the ROC Curves and Precision-Recall Curve of the following models with 20 percent darkness added. Yolov8-T, ViT-T, and LSTM-T's performances are shown with AUC scores (Area under the curve) which are 0.69, 0.68, and 0.63, respectively. We observe there is a performance reduction as compared to the CONNECT Dataset. Similarly, while precision-recall curves of Yolov8-T, ViT-T, and LSTM-T are similarly approached in terms of AUC, the AUC score of ViT-T is closer to Yolov8-T.

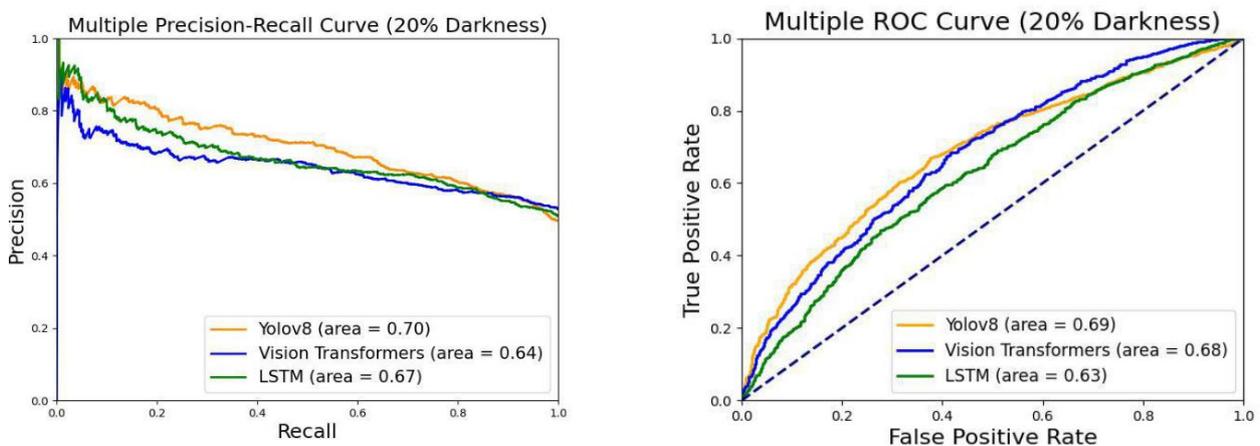


Figure 7. Precision-Recall and ROC curves on the CONNECT-D20 dataset

4.3. Results of CONNECT-D40 Dataset

Figure 8 indicates the ROC Curve and Precision-Recall Curves of the following models with 40 percent darkness added. The areas under the curve are 0.68, 0.66, and 0.64, respectively. We can observe that ViT-T models experienced a decrease in terms of AUC score, unlike LSTM-T and Yolov8-T. In addition, the AUC (Area under curve) of the ROC curve for LSTM-T has increased by 0.1 compared to 20 percent darkness.

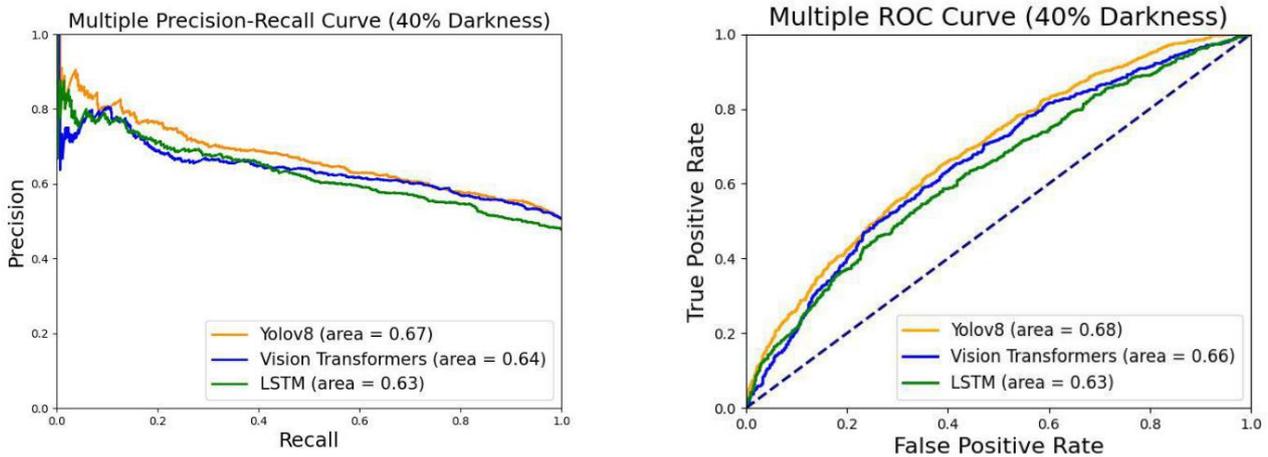


Figure 8. Precision-Recall and ROC curves on the CONNECT-40 dataset

4.4. Results of CONNECT-D60 Dataset

Figure 9 displays the calculated areas under curves, using True Positive Rate and False Positive Rates. These rates are 0.64 for Yolov8-T, 0.63 for ViT-T, and 0.57 for LSTM-T. It's noted that the LSTM-T exhibited a more consistent decrease than both Yolov8-T and Transformers when subjected to an increase in darkness by 20 percent. Furthermore, Figure 9 reveals that in terms of precision and recall values, ViT-T performed less effectively compared to their accurate positive and false favorable rates.

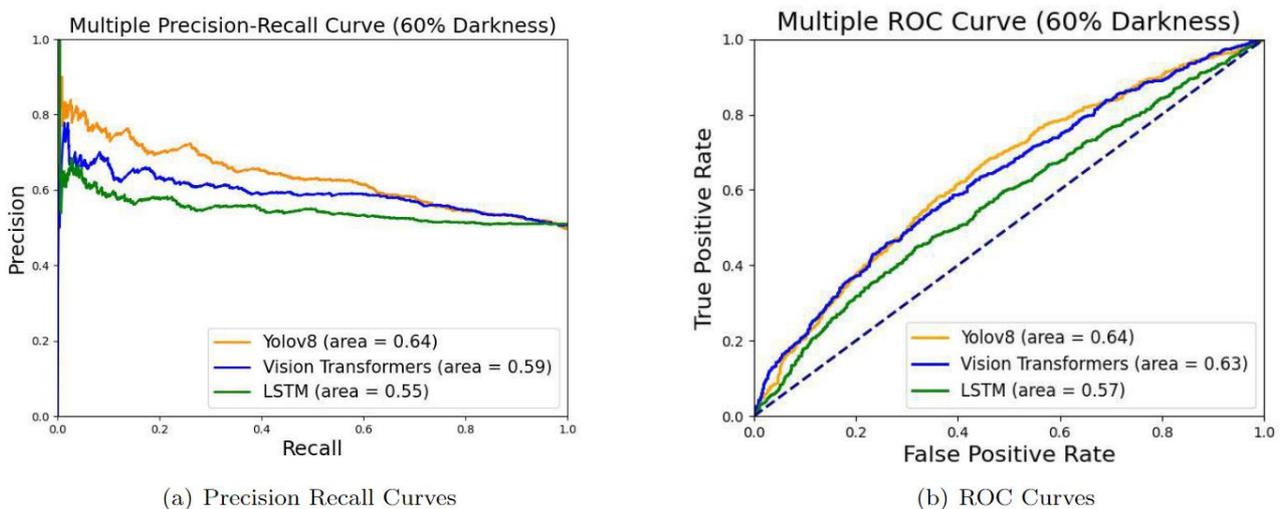


Figure 9. Precision-Recall and ROC curves on the CONNECT-60 dataset

4.5. Results of CONNECT-D80 Dataset

At this increased level of darkness, The ROC Curve outcomes depicted in Figure 10 indicate a minor decline in performance across all three deep learning models. Specifically, the areas under the curve for Yolov8-T, ViT-T, and LSTM-T are 0.56, 0.53, and 0.51, respectively. Moreover, in the Precision-Recall Curve, the areas under the curves for these models are 0.58, 0.54, and 0.51, respectively. These findings suggest that as darkness

intensifies, the effectiveness of the models' accurate favorable rates deteriorates. It is observed that the detection results are nearing those of a randomized model with 80 percent darkness addition.

In summary, the addition of darkness leads to a decline in the AUC performance for all three models, although LSTM-T shows some exceptions at specific values of the constant darkness level. The darkest samples show the most significant drops in performance metrics.

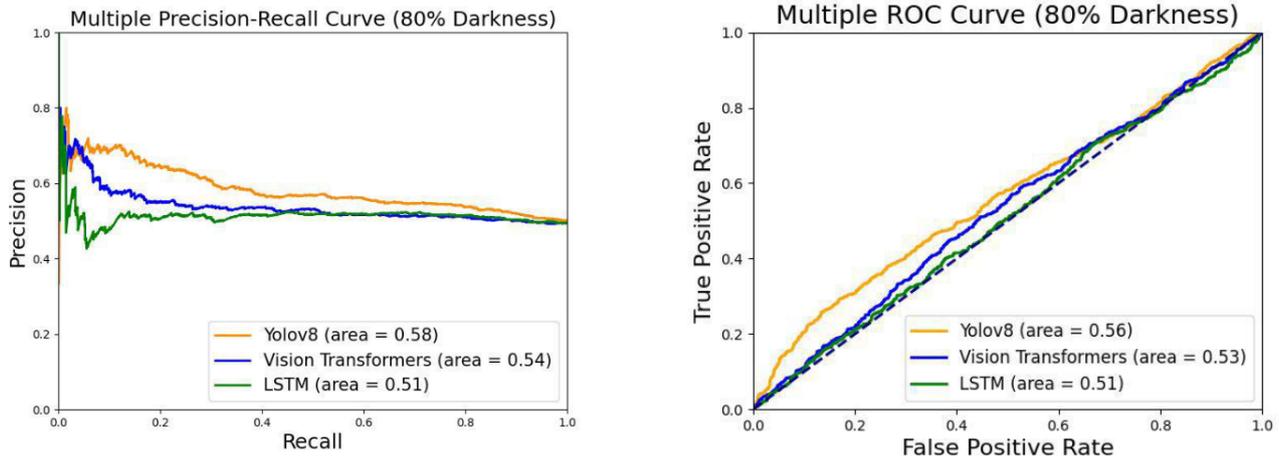


Figure 10. Precision-Recall and ROC curves on the CONNECT-80 dataset

4.6. Results of CONNECT-M Dataset

The CONNECT-M dataset presents more significant challenges than the CONNECT dataset, as it incorporates images with varying degrees of darkening to mimic nighttime conditions. Specifically, 30 percent of its data is darkened by 20 percent, another 30 percent by 40 percent, 20 percent by 60 percent, and the remaining by 80 percent. Figure 11 illustrates that different darkness levels impact the performance of the models. Furthermore, there might be a decline in model performance when encountering new image frames with unfamiliar darkness levels that are not seen during training. To elaborate, when the dataset includes a darkened frame from one video, another frame with the same level of darkness is selected from a different video subset to address potential issues like over-fitting.

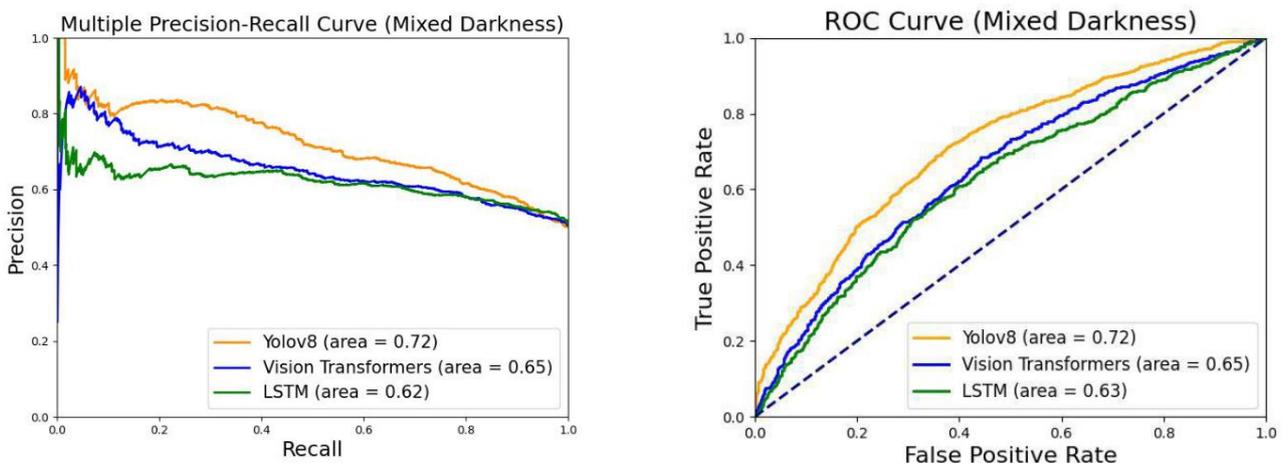


Figure 11. Precision-Recall and ROC curves on the CONNECT-M dataset

As seen in Table 3, all-AUC scores of the Precision-Recall Curve and ROC Curve of Yolov8-T, ViT-T, and LSTM-T vision models. They are shown for each darkness-added variation of the CONNECT dataset and the CONNECT dataset itself.

Table 3. AUC scores of vision models on all datasets

Dataset	Method	AUC (Precision-Recall)	AUC (ROC Curve)
CONNECT	Yolov8-T	0.80	0.78
CONNECT	ViT-T	0.73	0.74
CONNECT	LSTM-T	0.72	0.71
CONNECT-D20	Yolov8-T	0.70	0.69
CONNECT-D20	ViT-T	0.64	0.68
CONNECT-D20	LSTM-T	0.67	0.63
CONNECT-D40	Yolov8-T	0.67	0.68
CONNECT-D40	ViT-T	0.64	0.66
CONNECT-D40	LSTM-T	0.63	0.63
CONNECT-D60	Yolov8-T	0.64	0.64
CONNECT-D60	ViT-T	0.59	0.63
CONNECT-D60	LSTM-T	0.55	0.57
CONNECT-D80	Yolov8-T	0.58	0.56
CONNECT-D80	ViT-T	0.54	0.53
CONNECT-D80	LSTM-T	0.51	0.51
CONNECT-M	Yolov8-T	0.72	0.72
CONNECT-M	ViT-T	0.65	0.65
CONNECT-M	LSTM-T	0.62	0.63

We can say that across all models (Yolov8-T, ViT-T, LSTM-T), there is a consistent decrease in the AUC scores as the level of darkness increases. For instance, the AUC (Precision-Recall) for Yolov8-T drops from 0.80 in the CONNECT dataset (no additional darkness) to 0.58 in the CONNECT-D80 dataset (80% darkness). This trend is similarly observed in both the Precision-Recall and ROC Curve AUC scores for all models. However, Yolov8-T generally outperforms ViT-T and LSTM-T across all datasets. For example, in the CONNECT dataset, Yolov8-T achieves an AUC (Precision-Recall) of 0.80, while ViT-T and LSTM-T achieve 0.73 and 0.72, respectively. This suggests that Yolov8-T is more robust in maintaining performance under varying levels of darkness compared to the other models. Lastly, The LSTM-T model shows more variability and less robustness in its performance compared to Yolov8-T and ViT-T. For instance, the AUC (Precision-Recall) for LSTM-T in the CONNECT-D60 dataset is 0.55, which is significantly lower than the 0.64 achieved by Yolov8-T in the same dataset. This suggests that LSTM-T may require more fine-tuning or additional modifications to handle varying darkness levels effectively.

5. DISCUSSION

One of the most striking findings is the consistent decline in model performance as the level of darkness increases. This trend is evident across all three models—Yolov8, Vision Transformer (ViT), and Long Short-Term Memory (LSTM)—and is particularly pronounced at the highest levels of darkness (80%). The decrease in AUC (Precision-Recall and ROC) scores suggests that these models, while effective in well-lit conditions, struggle to maintain accuracy in low-light environments. This performance degradation likely stems from the reduced visibility of drones in darker images, which challenges the models' ability to discern relevant features and make accurate predictions.

Another notable observation in our analysis is the varying performance of specific deep learning models at different levels of darkness. For instance, the LSTM-T model exhibited lower performance under 20 percent darkness compared to 40 percent darkness. Specifically, the areas under the curves were 0.67 for CONNECT-D20 and 0.63 for CONNECT-D40 datasets. This suggests that changes may influence the tuning of model parameters in the pixel composition of the frames. However, it is interesting to note that LSTM-T's performance on the ROC curve remained consistent at 0.63 for both 20 percent and 40 percent darkness levels. Since recall and actual positive rate are considered identical in this context, the calculation of the false positive rate could be impacting the area under the curve to a certain extent.

Among the models tested, Yolov8-T consistently outperforms the others, demonstrating superior accuracy and robustness across all darkness levels. This can be attributed to several factors, including the advanced architecture of Yolov8, which is optimized for real-time object detection, and the benefits of transfer learning, which allows the model to build on existing knowledge rather than starting from scratch. Yolov8-T's ability to maintain high AUC scores even as darkness increases indicates its strong potential for practical deployment in scenarios where lighting conditions are variable and unpredictable.

5.1. Limitations

The primary focus on synthetic darkness, while useful for controlled experiments, may not fully capture the complexity of real-world low-light environments, where factors such as shadows, reflections, and varying light sources can further complicate detection tasks. Future studies should consider incorporating more diverse datasets that include naturally dark conditions to better simulate real-world scenarios.

5.2. Future Research

While the transfer learning has been shown to significantly improve model performance, it is also important to explore other advanced techniques, such as self-supervised learning or generative adversarial networks (GANs), which could provide additional benefits by enabling models to learn more effectively from limited data. These techniques, combined with ongoing improvements in model architectures, could lead to even more robust and reliable drone detection systems. We can also say that the variability observed in the LSTM-T model suggests that further research is needed to optimize recurrent neural networks for static image detection tasks. Investigating hybrid models that combine the strengths of different architectures (e.g., CNN-LSTM or Transformer-LSTM hybrids) could lead to new insights for effective solutions for drone detection in challenging conditions.

6. CONCLUSION

In this study, we have undertaken a thorough comparison of model performance, focusing on object detection through optical methods, particularly under conditions of added darkness. The results demonstrate that models with transfer learning, specifically Yolov8-T, ViT-T, and LSTM-T, consistently outperform their non-enhanced counterparts (Yolov8, ViT, LSTM) across all conditions. For instance, Yolov8-T achieves an AUC (Precision-Recall) of 0.80 on the CONNECT dataset, compared to 0.56 for Yolov8. This highlights the substantial benefit of transfer learning in improving model robustness and accuracy. For drone detection in images, we utilized Yolov8-T, vision transformers (ViT-T), and LSTM-T. The findings offer valuable insights into the advantages and constraints of vision-based modalities when darkness is introduced. In addition, the research highlighted the effects of environmental factors, such as darkness, on detection performance in multi-modal systems under real-world conditions. By means of observed trends, across all models, there was a

consistent decline in performance metrics as the level of synthetic darkness increased. This trend was most pronounced in the CONNECT-D80 dataset because of high level of darkness. In addition to that, YOLOv8-T consistently outperformed ViT-T and LSTM-T across all levels of darkness. Because transfer-learning method was more robust with pre-trained processes and fine-tuning parameters Also, we can conclude that YOLOv8-T is the most robust model. Among the models tested, YOLOv8-T consistently shows the highest performance across all datasets, including under varying levels of darkness. It not only maintains higher AUC scores in both Precision-Recall and ROC curves but also exhibits better generalization capabilities, making it the most reliable model for drone detection in challenging conditions. On the other hand, The LSTM-T model, while improved by transfer learning, shows more variability in performance compared to YOLOv8-T and ViT-T. This indicates that LSTM architectures might be more sensitive to changes in darkness levels, requiring further fine-tuning or potentially alternative approaches to handle extreme conditions effectively. In our final approach, even though models with 60 percent and 80 percent added darkness showed accuracy below 0.6 in precision-recall and ROC curve results, an accuracy above 0.6 was observed for each model when considering the total 40 percent of data subjected to 60 percent or greater darkness.

AUTHOR CONTRIBUTIONS

F. Oruç and H. B. Yılmaz wrote the manuscript, F. Oruç and H. B. Yılmaz developed the approach, F. Oruç conducted the experiments, and F. Oruç and H. B. Yılmaz interpreted the results.

ACKNOWLEDGEMENT

This study has been supported by Boğaziçi University Scientific Research Projects Under Grant 17862 (BAP-SUP-17862). We also want to thank our research team, who helped a lot during the experiments in NetLab of the Computer Engineering Department of Bogazici University. Special thanks will go to Yusuf Özben, Kerem Avcı, and Ömer Şahin Albayram at this point.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Adam, E. Y. (2020). Connectivity considerations for mission planning of a search and rescue drone team. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(4), 2228-2243. <https://doi.org/10.3906/elk-1912-46>
- Andraši, P., Radišić, T., Muštra, M., & Ivošević, J. (2017). Night-time detection of UAVs using thermal infrared camera. *Transportation Research Procedia*, 28, 183-190. <https://doi.org/10.1016/j.trpro.2017.12.184>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jamil, S., Abbas, M. S., & Roy, A. M. (2022). Distinguishing malicious drones using vision transformer. *AI*, 3(2), 260-273. <https://doi.org/10.3390/ai3020016>
- Khan, M. U., Misbah, M., Kaleem, Z., Deng, Y., & Jamalipour, A. (2023, June 20-23). *GAANet: Ghost auto anchor network for detecting varying size drones in dark*. In: Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring) (pp. 1-5). Florence, Italy. <https://doi.org/10.1109/VTC2023-Spring57618.2023.10200720>
- Li, Y., Fan, Q., Huang, H., Han, Z., & Gu, Q. (2023). A modified YOLOv8 detection network for UAV aerial image recognition. *Drones*, 7(5), 304. <https://doi.org/10.3390/drones7050304>
- Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., Wang, X., Zhai, X., Kipf, T., & Houlby, N. (2022, October 23-27). *Simple Open-Vocabulary Object Detection*. In: S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.),

- Proceedings of the 17th European Conference on Computer Vision (ECCV 2022) (pp. 728-755). Tel Aviv, Israel. https://doi.org/10.1007/978-3-031-20080-9_42
- Misbah, M., Khan, M. U., Yang, Z., & Kaleem, Z. (2023, March 12-13). *Tf-net: Deep learning empowered tiny feature networks for night-time UAV detection*. In: J. Zhao (Eds.), Proceedings of the 13th EAI International Conference on Wireless and Satellite Systems (pp. 3-18). Virtual Event, Singapore. https://doi.org/10.1007/978-3-031-34851-8_1
- Moustafa, N., & Jolfaei, A. (2020). Autonomous detection of malicious events using machine learning models in drone networks. In: Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond (pp. 61-66). <https://doi.org/10.1145/3414045.3415951>
- Moustafa, N., Slay, J., & Creech, G. (2017). Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks. *IEEE Transactions on Big Data*, 5(4), 481-494. <https://doi.org/10.1109/TBDDATA.2017.2715166>
- Munir, A., Siddiqui, A. J., & Anwar, S. (2024, January 01-06). *Investigation of UAV Detection in Images with Complex Backgrounds and Rainy Artifacts*. In: Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW) (pp. 221-230). Waikoloa, HI, USA. <http://doi.org/10.1109/WACVW60836.2024.00031>
- Ramadan, R. A., Emara, A. H., Al-Sarem, M., & Elhamahmy, M. (2021). Internet of drones intrusion detection using deep learning. *Electronics*, 10(21), 2633. <https://doi.org/10.3390/electronics10212633>
- Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. <https://doi.org/10.48550/arXiv.2305.09972>
- Svanström, F., Alonso-Fernandez, F., & Englund, C. (2022). Drone detection and tracking in real-time by fusion of different sensing modalities. *Drones*, 6(11), 317. <https://doi.org/10.3390/drones6110317>
- Yi, K. Y., Kyeong, D., & Seo, K. (2019). Deep learning-based drone detection and classification. *The Transactions of the Korean Institute of Electrical Engineers*, 68(2), 359-363. <http://doi.org/10.5370/KIEE.2019.68.2.359>
- Zhai, X., Huang, Z., Li, T., Liu, H., & Wang, S. (2023). YOLO-Drone: an optimized YOLOv8 network for tiny UAV object detection. *Electronics*, 12(17), 3664. <https://doi.org/10.3390/electronics12173664>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1529271>

Super Twisting Sliding Mode Control of Four-Phase Interleaved Boost Converter

Veli YENİL¹ Sadık ÖZDEMİR² Zafer ORTATEPE^{3*} ¹ Department of Çardak OSB Vocational School, Pamukkale University, Denizli, Türkiye² Department of Mechatronics Engineering, Pamukkale University, Denizli, Türkiye³ Department of Automotive Engineering, Pamukkale University, Denizli, Türkiye

Keywords	Abstract
Fuel Cell Interleaved Boost Converter PI Control Super Twisting Sliding Mode Control	This paper presents a novel control method that integrates super-twisting sliding mode (STSM) voltage control with proportional-integral (PI) current control for a four-phase interleaved boost converter (IBC) in fuel cell applications. The STSM control, employed in the outer voltage loop, provides robust voltage regulation by generating precise reference currents for each phase. The conventional PI control in the inner current loop utilizes these reference currents to generate pulse width modulation (PWM) signals for each phase. The effectiveness of the proposed control strategy is evaluated through comprehensive simulation studies in MATLAB/Simulink, demonstrating an improvement in dynamic performance and enhanced robustness compared to conventional methods. Quantitative analysis shows that the output voltage quickly rises to the reference voltage within approximately 0.25 seconds in the proposed STSM-PI control method and improves transient response by 16 times compared to the conventional PI-PI method. This integrated STSM-PI control strategy offers significant advancements in reliability and efficiency making it a promising solution for high-performance fuel cell power systems.

Cite

Yenil, V., Özdemir, S., & Ortatepe, Z. (2024). Super Twisting Sliding Mode Control of Four-Phase Interleaved Boost Converter. *GU J Sci, Part A, 11(3)*, 563-576. doi:10.54287/guj.1529271

Author ID (ORCID Number)	Article Process
0000-0002-0257-5305	Submission Date 07.08.2024
0000-0001-7676-7484	Revision Date 28.08.2024
0000-0001-7771-1677	Accepted Date 11.09.2024
	Published Date 30.09.2024

1. INTRODUCTION

Fossil fuels, including coal, oil, and natural gas, have been extensively used across a wide range of applications for many decades. While fossil fuels have provided immense benefits in terms of energy access and economic development, their widespread use has also led to environmental and climate change concerns due to greenhouse gas emissions and other pollutants (Abas et al.; 2015; Eriksson & Gray, 2017). This has spurred efforts to transition towards more sustainable and cleaner energy sources and technologies that provide high-power efficiency and compactness.

Among these technologies, fuel cells have gained significant attention as an encouraging alternative. A fuel cell is a type of electrochemical device that uses a chemical reaction to transform chemical energy held in fuels into electrical energy. It operates similarly to a battery but does not require recharging; instead, it continues to produce electricity if it is supplied with fuel and an oxidizing agent (usually oxygen from the air) (Mekhilef et al., 2012). Fuel cells are known for their high efficiency, low emissions (since water is the primary by-product), and quiet operation. They have various applications including powering vehicles (fuel cell vehicles), and among all cleaner and sustainable energy sources, fuel cells are gaining more and more attention and becoming popular for electric transportation systems (Manoharan et al., 2019; Sazali et al., 2020).

However, the most important challenge in fuel cell technology is low output voltage, typically ranging between 0.5 to 0.8 volts per cell. This low output voltage is often insufficient for most applications, necessitating voltage

conversion to higher voltage levels. Multiple fuel cells can be connected in series to achieve desired voltage levels, but this approach can reduce system reliability and add complexity (Hao et al., 2021). Thus, DC/DC boost converters have become essential in fuel cell systems to efficiently step up the voltage to a suitable level for end-use applications.

Conventional DC/DC boost converters are frequently used because to their simplicity and cost-effectiveness. However, they may suffer from low efficiency at low power levels, which is a common problem in fuel cell systems due to varying load demands. In addition, the inductors and capacitors used in conventional boost converters can be bulky and heavy, which is a significant drawback in applications where space and weight are critical, such as in automotive or portable fuel cell systems (Sagar Bhaskar et al., 2020).

Interleaved boost converter (IBC) offers a promising solution to these challenges by using multiple phases (or channels) of boost converters operating in parallel with interleaved switching signals. This design significantly reduces input and output ripple currents, enhances efficiency, and improves thermal management. By interleaving the switching phases, the converter minimizes electromagnetic interference (EMI) and reduces the size and cost of input and output filter components. Additionally, the IBC is scalable, allowing for higher power levels without significantly increasing the size of individual components. The reduction in ripple current also decreases conduction losses in the power switches and inductors, further enhancing overall efficiency (Seyezhai & Mathur, 2012; Khosroshahi et al., 2015; Mallikarjuna Reddy & Samuel, 2020).

To achieve optimal performance, various control methods have been introduced for IBC topology, including proportional-integral (PI) control, deadbeat control, sliding mode control (SMC), model predictive control (MPC) and super-twisting sliding mode (STSM) control (Kabalo et al., 2013; Nikhar et al., 2016). These control methods have different advantages related to fast dynamic response, energy efficiency and robustness.

In the SMC, the system is driven along a predefined sliding surface toward a desired state (Komurcugil et al., 2021). However, in practice, achieving and maintaining this sliding condition can be challenging in nonlinear systems (Sankar et al., 2021). STSM control is an extension of conventional SMC and a robust control technique used in nonlinear systems to achieve precise tracking performance (Gonzalez et al., 2012; Napole et al., 2021). Different from the SMC, the STSM control generates a second-order sliding motion along the sliding surface. This additional term helps in achieving faster convergence and smoother control action, even in highly nonlinear systems (Hao et al., 2023). Besides, the MPC is an effective nonlinear control method providing a fast dynamic response. Nevertheless, this approach has a chattering problem, and the MPC's performance is dependent on the system model (Schwenzer et al., 2021). Chattering can degrade system performance and lead to instability or reduced control accuracy. To decrease chattering effect, STSM control is applied to the different types of DC-DC converters (Hao et al., 2022; Guler et al., 2023).

In this paper, an integrated control method is proposed that combines STSM voltage control with conventional PI current control to enhance the robustness of four-phase IBC used in fuel cell applications. The STSM control ensures precise voltage regulation by generating accurate reference currents for each phase, while the PI control minimizes the error between the measured and reference inductor currents, ensuring equal current distribution across all phases. Proposed control strategy is thoroughly evaluated through simulation studies conducted in MATLAB/Simulink.

This study presents several contributions to show the superiority of the suggested topology over previous literature investigations and its application. The following can be used to summarize their contributions:

1. A combined STSM-PI control method is proposed and implemented to the IBC topology instead of conventional PI-PI control.
2. Thanks to the proposed control method, the chattering problem is reduced.

This article is organized as follows: Section 2 provides a detailed analysis of the four-phase IBC. Section 3 describes the proposed control scheme, including both STSM and conventional PI controls. Section 4 gives

the simulation results that verify the robustness of the control scheme. Finally, conclusion is given in Section 5.

2. ANALYSIS OF FOUR PHASE IBC

Four-phase IBC, as depicted in Figure 1, can be regarded as four conventional boost converters connected in parallel.

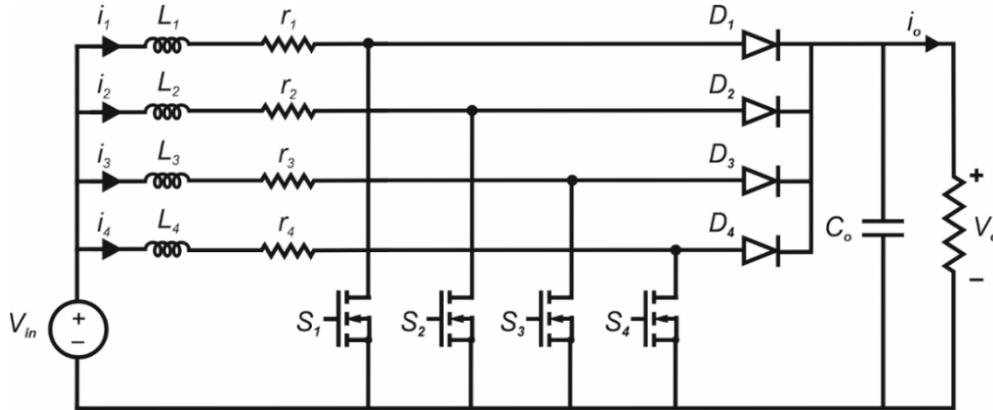


Figure 1. Four-phase IBC topology

In Figure 1, each boost converter has an inductor L_i ($i = 1, 2, 3, 4$), a diode D_i ($i = 1, 2, 3, 4$), a semiconductor switch S_i ($i = 1, 2, 3, 4$), and an output capacitor (C_o). r_i ($i = 1, 2, 3, 4$) denotes the parasitic resistance of each inductor. Besides, V_{in} denotes input voltage source, I_{in} denotes total input current, V_o denotes output voltage, I_o denotes output current, and R_o denotes load resistance. Four-phase IBC operates with continuous conduction mode (CCM), based on the drive signal of each converter.

A conventional boost converter has two operation modes. When the S_i is on, the input current of the inductor can be calculated with Equation (1). When the S_i is off, the capacitor and load are charged with the voltage source and inductor. Therefore, input current can be calculated with Equation (2).

$$\frac{di_{Li}}{dt} = \left(\frac{1}{L_i}\right)(V_{in} - R_o I_o) \quad (1)$$

$$\frac{di_{Li}}{dt} = \left(\frac{1}{L_i}\right)(V_{in} - R_o I_o - V_o) \quad (2)$$

By using Kirchoff's current law, the input current of the IBC can be obtained as follows:

$$i_{in} = i_{L1} + i_{L2} + i_{L3} + i_{L4} \quad (3)$$

By using Equation (1) and (2), the dynamic model of IBC can be obtained as follows:

$$\frac{di_{Li}}{dt} = \left(\frac{1}{L_i}\right)(V_{in} - (1 - d_i)V_o) \quad (4)$$

Where d_i ($i=1,2,3,4$) is duty cycle for the power switches.

$$d_i = \begin{cases} 0 & S_i \text{ is OFF state} \\ 1 & S_i \text{ is ON state} \end{cases} \quad (5)$$

The four-phase IBC model can be obtained using Kirchoff's current/voltage law and the previously mentioned assumptions.

$$\frac{di_i}{dt} = -(1 - d_i) \frac{v_o}{L} + \frac{r_L}{L} (v_{in} - i_i) \quad (6)$$

$$\frac{dv_o}{dt} = \frac{1}{C} \sum_1^4 i_i - \frac{1}{f_L C} V_o - \frac{1}{C} \sum_1^4 d_i i_i \quad (7)$$

Voltage gain of a four-phase IBC in CCM can be calculated as (8).

$$\frac{V_{out}}{V_{in}} = \frac{1}{1 - D} \quad (8)$$

V_{out} , V_{in} and D are average DC values of V_o , V_{in} and d_i . Besides, the capacitor and inductor equations are given in (9).

$$\begin{cases} V_{in} - L_1 \frac{di_{L_1}(t)}{dt} = 0 \\ V_{in} - L_{2,3,4} \frac{di_{L_{2,3,4}}(t)}{dt} = V_o(t) \\ \sum_{i=2}^4 i_{L_i}(t) - C \frac{dV_o(t)}{dt} = \frac{V_o(t)}{R} \end{cases} \quad (9)$$

2.1. Component Design

Inductors play a crucial role in energy storage and transfer process in a boost converter. In an IBC operating in continuous conduction mode (CCM), inductor current never falls to zero during the switching cycle, providing a continuous energy transfer to the output. The inductance value is calculated to provide that the converter operates in CCM, which minimizes current ripple and enhances efficiency. The inductance value (L) can be calculated using the following equation:

$$L = \frac{V_{in} D}{f_{sw} \Delta I_L} \quad (10)$$

where V_{in} denotes input voltage, D denotes duty cycle, f_{sw} denotes switching frequency and ΔI_L denotes peak-to-peak inductor current ripple. The effective inductance is divided among the phases for an interleaved converter. Therefore, the ripple current per phase is decreased, which allows for a smaller inductor size while maintaining the desired performance.

The output capacitor in an IBC is responsible for filtering the output voltage ripple and providing a stable DC output voltage. In CCM operation, the capacitor must be designed to handle the ripple current and maintain the desired output voltage. The capacitor value (C) is calculated based on the allowable output voltage ripple (ΔV_o), switching frequency, duty cycle and load current as follows:

$$C = \frac{I_o D}{f_{sw} \Delta V_o} \quad (11)$$

where I_o denotes output current, and ΔV_o denotes peak-to-peak output voltage ripple. Although a higher capacitance value reduces output voltage ripple, the circuit may become larger and more expensive. The ESR of the capacitor affects the output voltage ripple. Low ESR capacitors, such as tantalum or ceramic capacitors, are preferred to minimize the voltage ripple. The voltage ripple due to the ESR can be estimated as follows:

$$\Delta V_o = \Delta I_L \times ESR \quad (12)$$

The choice of output diode is determined by reverse recovery time, voltage and current ratings. In addition to the V_o voltage, there needs to be a margin for the voltage value because of the ringing effect. Within the range

of the maximum output current, V_f should be as low as possible to reduce conduction losses. In order to reduce switching losses, the reverse recovery time (t_{rr}) must be as little as feasible.

3. CONTROL SCHEME

In this paper, the control scheme consists of two control methods. Voltage control loop is constructed with STSM, and a current control loop is constructed with conventional PI control. The objective of the STSM control is to provide good voltage tracking performance, and the objective of the conventional PI control is to minimize the error between the measured inductor current and reference values of inductor current, generated from the STSM voltage control loop. As a result, output voltage regulation and equal current for each phase are provided. The proposed control scheme is given in Figure 2.

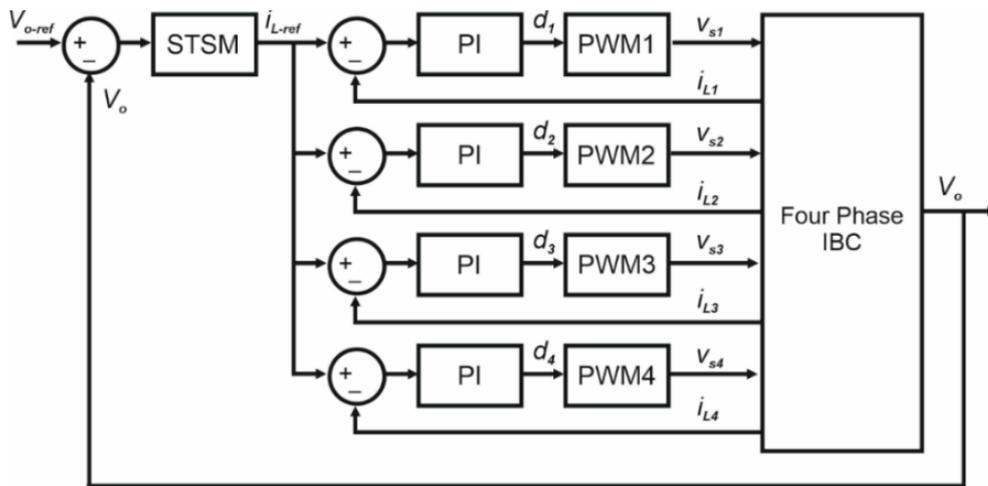


Figure 2. Control scheme of four-phase IBC using STSM and PI control

In four-phase IBC, each adjacent drive signals have a $T_s/4$ delay as given in Figure 3.

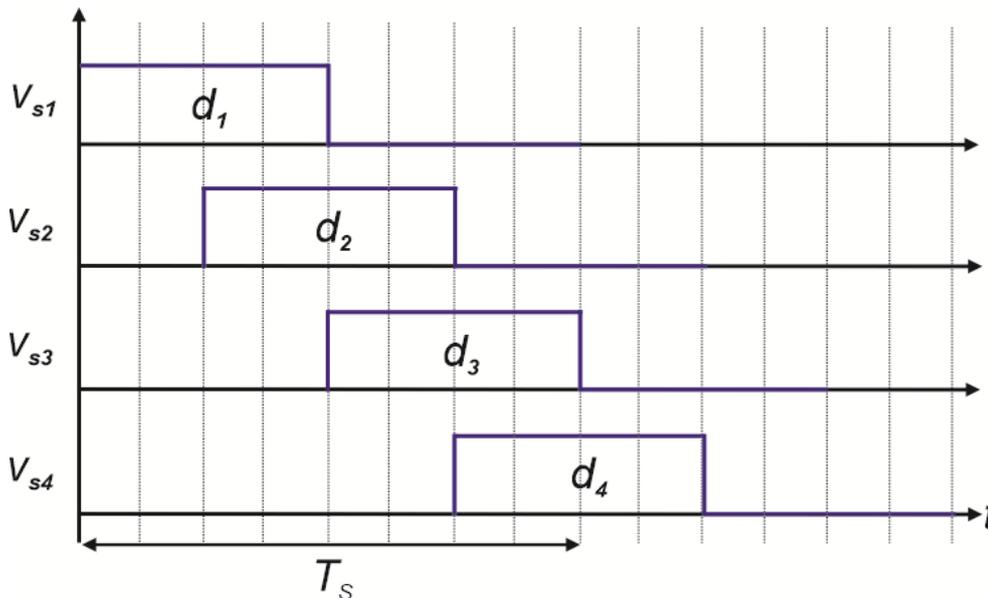


Figure 3. Control signals of four-phase IBC

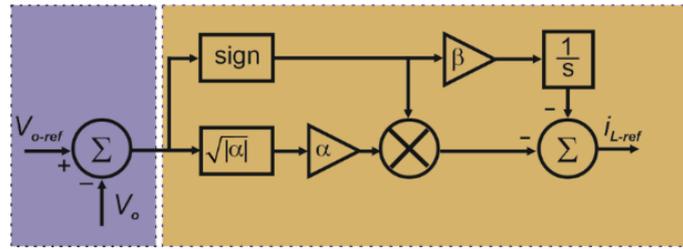
In this paper, only CCM is considered in the simulation analysis. As shown in Figure 1, the proposed four-phase IBC topology contains four switches, diodes, and one capacitor. The transient states of the switches are ignored so, the switches have two switching states. “1” denotes on-state and “0” gives off-state. Table 1 gives the switching states of four-phase IBC topology, and S1000 says the S_1 switch is on-state and the rest of the switches are off-state.

Table 1. Four-phase IBC operation modes

Conditions		Switching States
Mode 1	$0 < d_1(t) + d_2(t) + d_3(t) + d_4(t) \leq 1$	S1000→S0000→S0100→S0000→S0010→S0000→S0001→S0000
Mode 2	$1 < d_1(t) + d_2(t) + d_3(t) + d_4(t) \leq 2$	S1000→S1100→S0100→S0110→S0010→S0011→S0001→S1001
Mode 3	$2 < d_1(t) + d_2(t) + d_3(t) + d_4(t) \leq 3$	S1011→S1001→S1101→S1100→S1110→S0110→S0111→S0011
Mode 4	$3 < d_1(t) + d_2(t) + d_3(t) + d_4(t) \leq 4$	S0111→S1111→S1011→S1111→S1101→S1111→S1110→S1111

3.1. Implementation of Super Twisting Sliding Mode (STSM) Control

The SMC is an effective nonlinear control method that is insensitive to system parameters. It has been widely used in power converter control. However, SMC has a chattering problem resulting in uncontrolled switching frequency. To reduce the chattering problem of conventional SMC, the STSM control can be implemented with different types of converters. To generate the inductor current reference, i_{Lref} for the IBC, the STSM control is selected and implemented in this paper. The STSM control block diagram is given in Figure 4.

**Figure 4.** STSM control method

STSM control consists of two terms as equivalent control term (u_{eq}) and discontinuous control term (u_{st}):

$$u(t) = u_{eq}(t) + u_{st}(t) \quad (13)$$

$$u_{st}(t) = u_1(t) + u_2(t) \quad (14)$$

Also, the u_1 and u_2 can be calculated as follows:

$$\begin{cases} u_1(t) = -\alpha \cdot \text{sign}(S) \\ u_2(t) = -\beta \cdot |S|^{0.5} \text{sign}(S) \end{cases} \quad (15)$$

3.2 Implementation of Conventional PI Control

PI control is formed by combining proportional (P) and integral (I) controllers and is widely used in various power electronics applications due to its effectiveness, and simplicity and it is preferred to regulate current and voltage loops for the IBC. The proportional term provides an immediate response to the error signal. It is responsible for the speed of the transient response and helps to reduce the rise time. The integral term accumulates the error over time and eliminates steady-state error. It ensures that output voltage precisely matches the reference voltage in the steady state. Proper tuning of the proportional and integral gains (K_p and K_i) is crucial for optimal performance. PI controller effectively regulates the output voltage. Variations in input voltage or load are quickly corrected by adjusting the duty cycle. Interleaving of phases reduces the input and output current ripple, leading to smoother operation and lower electromagnetic interference (EMI). By distributing the load across multiple phases, the interleaved boost converter improves efficiency and reduces thermal stress on individual components. Theoretical equation of the PI controller is given in (16).

The block diagram of a conventional PI control is given in Figure 5. The controller compares the reference $R(s)$ with the output control signal $C(s)$ for having the error signal $E(s)$. The error signal is minimized in the

PI controller then control signal $U(s)$ is gained. Distortion signal $D(s)$ and the control signal are added and employed to transfer function (plant, $G_p(s)$) to have the response signal.

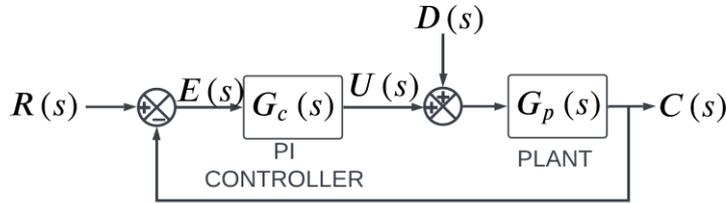


Figure 5. Conventional PI controller block diagram

$$G_c(s) = K_p + \frac{K_i}{s} \tag{16}$$

4. SIMULATION RESULTS

Simulation studies have been conducted using the MATLAB/Simulink platform in order to evaluate the control performance and dynamic response of the proposed control method. Figure 6 shows the block design of the simulation configuration and Table 2 contains a list of the simulation parameters.

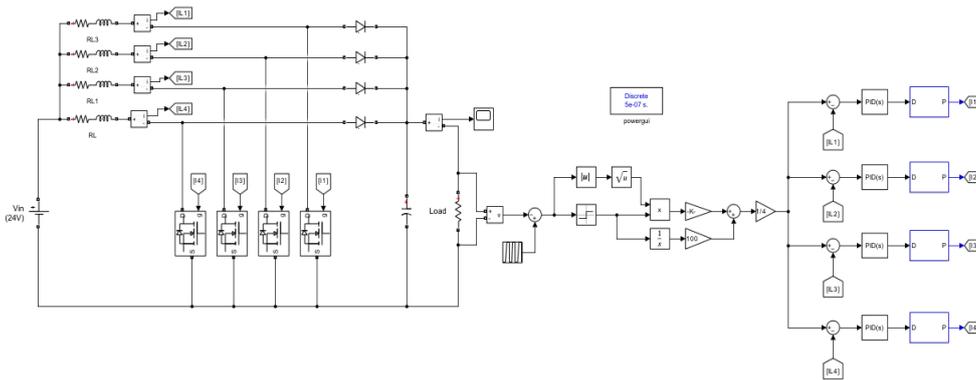


Figure 6. Simulation layout of the system

Table 2. Simulation parameters

Parameters	Symbol	Values
Input voltage	V_{in}	24 V
Load resistor	R_o	12 Ω
Inductor	$L_i (i=1,..4)$	500 μ H
Inductor resistance	$r_i (i=1,..4)$	0.3 Ω
Output voltage	V_o	48 – 60 V
Output capacitor	C_o	1000 μ F
Controller gains	α, β	100, 0.1
Controller gains	K_p, K_i	0.1, 10
Number of phases		4
Switching frequency	f_s	10 kHz

Figure 7 shows the output voltage performance of the proposed STSM-PI control method and conventional PI-PI control method, respectively. The controller gain (α and β) parameters of the proposed control method and the controller gain (K_p, K_i) parameters of PI controller are given in Table 2. In Figure 7, $V_{in} = 24V$ is applied to the input as a reference DC voltage. Also, $V_{oref} = 48V$ is applied as the reference output voltage until the 2. seconds. After 2 seconds, a reference ramp function ($V_{oref} = 48V$ to $60V$) is applied as a reference voltage, and after 3 seconds, $V_{oref} = 60V$ is applied constantly. In a transient state, maximum voltage ripple of the output voltage is measured as 2 % as shown in Figure 7a. It is also observed that even if the reference output voltage changes, the actual output voltage shows a good tracking performance in transient state and settles at the new reference after a few hundred milliseconds. Figure 7b shows the output voltage performance of the conventional PI-PI control method. The figure shows that the output voltage can never follow the reference voltage. As shown in the Figure 7 the proposed controller has better tracking performance under steady-state and transient responses compared to the conventional PI-PI control method.

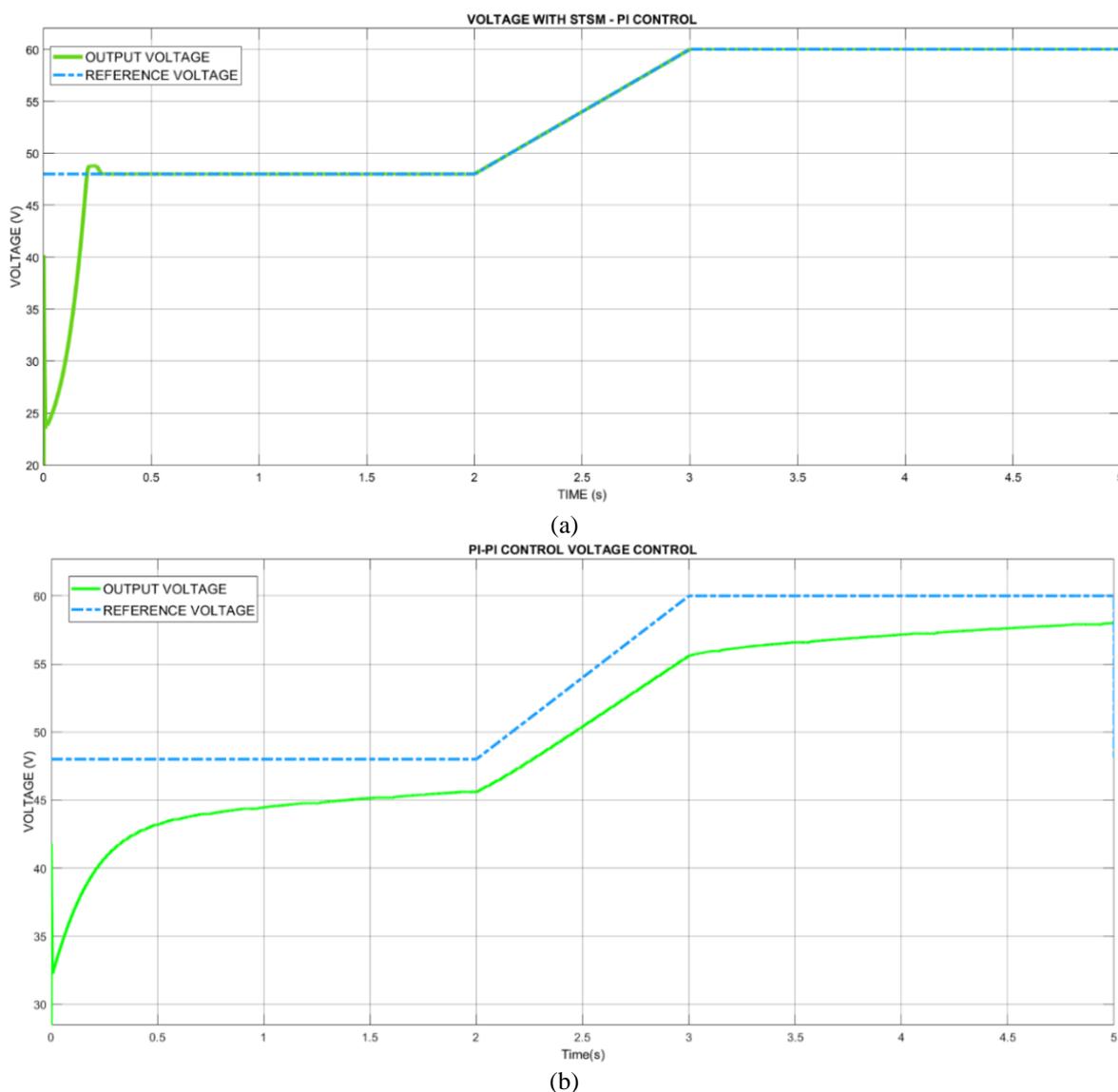
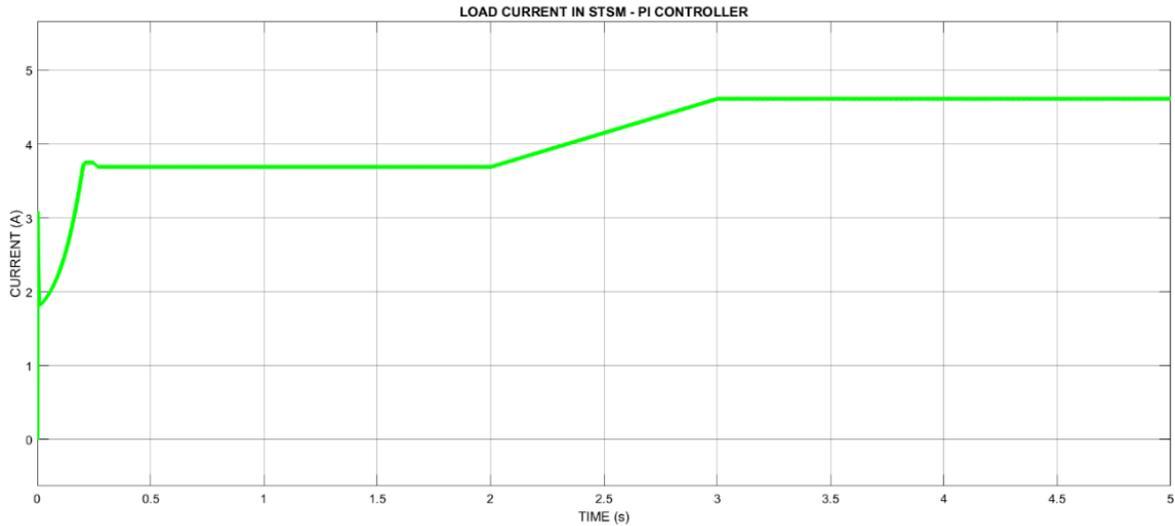


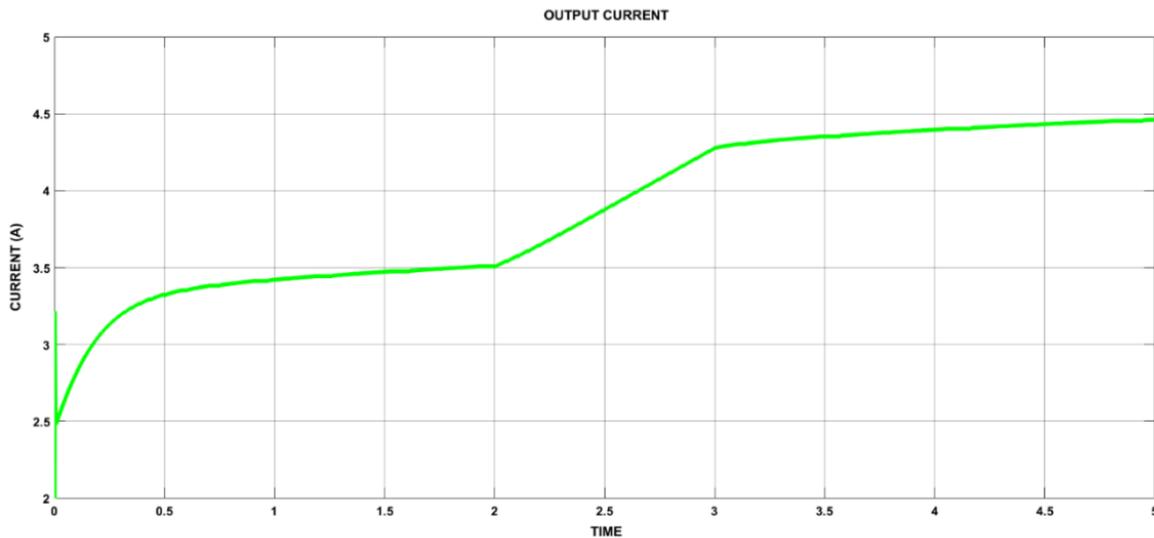
Figure 7. The comparison of the output voltage performance **a)** Proposed STSM-PI controller, **b)** conventional PI-PI controller

Figure 8 shows the output current performance of the proposed STSM-PI control method and conventional PI-PI control method, respectively. In Figure 8, $V_{in} = 24V$ is applied to the input as a reference DC voltage. Also, $V_{oref} = 48V$ is applied as the reference output voltage until the 2. seconds. After 2 seconds, a reference ramp

function ($V_{o_{ref}} = 48V$ to $60V$) is applied as a reference voltage, and after 3 seconds, $V_{o_{ref}} = 60V$ is applied constantly. As shown in the Figure 8a and 8b, the proposed controller has better output current tracking performance under steady-state and transient responses compared to the conventional PI-PI control method. The superior performance of the proposed method is mainly due to its high robustness to uncertainties and external disturbances in the system. This controller provided smoother control by minimizing the chattering effect while responding more efficiently and quickly to rapid changes in load current. Moreover, using the conventional PI-PI controller, the system is found to be less effective against nonlinear dynamics and large parameter variations, which resulted in slower and unstable load current reaching the desired reference values.



(a)



(b)

Figure 8. The comparison of the output current performance **a)** Proposed STSM-PI controller, **b)** conventional PI-PI controller

Besides, inductor currents of each phase and output current waveforms of the converter is given in Figure 9. As seen in the figure, the circuit operates in CCM mode since the phase currents never drop to zero. Also, from the figure it is seen that the total current is shared as phase currents.

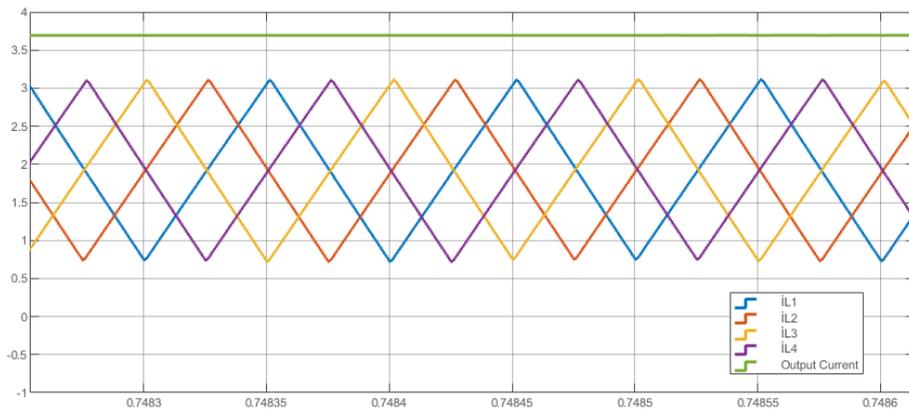


Figure 9. Inductor currents and output current waveforms of the system

Table 3 presents a comparative analysis of two control methods STSM-PI control and conventional PI-PI control based on several key performance characteristics: robustness, response time, stability, chattering, and complexity. The comparison in Table 3 shows that while the proposed STSM-PI controller offers superior performance in terms of robustness, response time, and stability, it does so at the cost of increased complexity and potential implementation challenges. In contrast, the conventional PI-PI controller is simpler and easier to implement but may not provide the same level of performance in dynamic or uncertain environments. The choice between these controllers depends on the specific requirements of the application, such as the need for fast response and high robustness versus the desire for simplicity and ease of implementation.

Table 3. Comparison of proposed STSM-PI and conventional PI-PI controllers

Characteristic	STSM-PI Controller	PI-PI Controller
Robustness	High	Low
Response Time	Fast	Moderate
Stability	High	Moderate
Chattering	Reduced	None
Complexity	Moderate	Low

Table 4 compares the proposed method with the existing studies. The proposed STSM-PI control method, while not the fastest in terms of settling time, offers a balanced performance with respect to voltage overshoot, and phase count. The four-phase topology ensures better current distribution and reduced ripple, while the discrete inductance design simplifies the implementation. 2% voltage overshoot is a reasonable trade-off, offering a balance between dynamic response and system stability. Compared to existing studies, the proposed method is competitive in terms of settling time, robustness, and reliability, especially when high efficiency and reduced ripple are important considerations.

Table 4. Comparison of proposed method with the existing studies

	Proposed	(Zhuo et al., 2021)	(Alajmi et al., 2022)	(Banerjee et al., 2017)	(Nahar & Uddin, 2018)
Number of phase	4	2	3	2	2
Inductance type	discrete	discrete	discrete	coupled	discrete
Voltage overshoot	% 2	none	none	% 1.63	%30
Inductor current	CCM	CCM	DCM	CCM	CCM
Settling time	0.2 sec	0.1 sec	0.25 sec	0.02 sec	0.1 sec

4.1. Chattering Phenomenon

Chattering is a common issue in control systems, especially in conventional SMC. It causes to the rapid oscillations that occur around the desired trajectory or sliding surface. These oscillations are caused by the high-frequency switching actions inherent in SMC that are designed to keep the system state on the sliding surface. However, in practical applications, chattering can lead to several undesired effects, including increased wear and tear on mechanical components, reduced system efficiency and higher energy consumption. In some cases, chattering effect can also destabilize the system, making it challenging to achieve precise control.

MPC is a popular control strategy that predicts future behavior of the system based on a model and optimizes control inputs to achieve desired performance. While MPC is effective in providing fast dynamic responses, it can still suffer from chattering, particularly in systems with non-linearities or high-frequency dynamics. The root cause of chattering in MPC often lies in the discrete nature of the control systems and the approximation errors in the predictive model.

The STSM control, advance variant of SMC, is specifically designed to address the chattering issue. Unlike conventional SMC, STSM employs a second-order sliding mode approach. This approach includes two control terms: an equivalent control term and a discontinuous control term. The equivalent control term provides the system trajectory remains close to the sliding surface, while the discontinuous term is responsible for maintaining the system on the surface.

Consequently, STSM effectively addresses the chattering problem by introducing a second-order sliding mode that smooths the control action and reduces the frequency of switching. This contrasts with MPC, where chattering can still occur due to model mismatches and discrete control actions. The main advantage of STSM over conventional SMC and MPC lies in its ability to reduce or even eliminate chattering. This is achieved through the following mechanisms:

Second-Order Sliding Surface: STSM proposes a second-order sliding motion, which smooths out the control action, remarkably reducing the high-frequency switching that causes chattering. This second-order approach allows for a more gradual convergence to the sliding surface, minimizing the abrupt variations in control input that typically lead to chattering.

Robustness to nonlinearities and uncertainties: STSM is highly robust to external disturbances and system nonlinearities, which are often the source of chattering in other control methods. By accounting for these uncertainties within the control system, STSM can maintain stable performance without the need for excessive switching, further reducing chattering. This makes STSM particularly suitable for real-time applications where both computational efficiency and performance are critical.

5. CONCLUSION

In this paper, the performance of a four-phase IBC controlled by an integrated STSM and PI control strategy is thoroughly investigated. The primary objectives are to enhance the efficiency of the converter, improve transient response, and ensure robust performance under varying load and reference voltage conditions. Simulation results show that the proposed control method provides superior robustness against parameter variations, maintaining stability across a wide range of operating conditions. Quantitative analysis shows that the output voltage quickly rises to the reference voltage within approximately 0.25 seconds in the proposed STSM-PI control method and improves transient response by 16 times compared to the conventional PI-PI method. The STSM control also effectively minimizes the chattering effects, a common issue in conventional SMC, thereby enhancing overall system stability. The PI control, while simpler to implement, provides satisfactory steady-state performance with minimum steady-state error. Besides, the integrated STSM-PI approach outperforms the PI-PI control strategy in both dynamic and steady-state scenarios, making it a more reliable choice for high-performance fuel cell applications. These results suggest that the integrated control strategy is a viable solution for improving the efficiency and reliability of fuel cell power systems. Future work may focus on experimental validation to further confirm the simulation results and explore the application of this control strategy in real-world power electronic systems.

AUTHOR CONTRIBUTIONS

Conceptualization, V.Y.; methodology, V.Y. and S.Ö.; software, V.Y.; title, Z.O.; validation, V.Y., S.Ö. and Z.O.; research, S.Ö., and Z.O.; sources, S.Ö.; data curation, S.Ö.; manuscript-original draft, S.Ö.; manuscript-review and editing, Z.O.; visualization, Z.O.; supervision, Z.O. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Abas, N., Kalair, A., & Khan, N. (2015). Review of Fossil Fuels and Future Energy Technologies. *Futures*, 69, 31-49. <https://doi.org/10.1016/J.FUTURES.2015.03.003>
- Alajmi, B. N., Marei, M. I., Abdelsalam, I., & Ahmed, N. A. (2022). Multiphase Interleaved Converter Based on Cascaded Non-Inverting Buck-Boost Converter. *IEEE Access*, 10, 42497-42506. <https://doi.org/10.1109/ACCESS.2022.3168389>
- Banerjee, S., Ghosh, A., & Rana, N. (2017). An Improved Interleaved Boost Converter with PSO-Based Optimal Type-III Controller. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 5(1), 323-337. <https://doi.org/10.1109/JESTPE.2016.2608504>
- Sagar Bhaskar, M., Ramachandaramurthy, V. K., Padmanaban, S., Blaabjerg, F., Ionel, D. M., Mitolo, M., & Almakhlis, D. (2020). Survey of DC-DC Non-Isolated Topologies for Unidirectional Power Flow in Fuel Cell Vehicles. *IEEE Access*, 8, 178130-178166. <https://doi.org/10.1109/ACCESS.2020.3027041>
- Eriksson, E. L. V., & Gray, E. MacA. (2017). Optimization and Integration of Hybrid Renewable Energy Hydrogen Fuel Cell Energy Systems – A Critical Review. *Applied Energy*, 202, 348-364. <https://doi.org/10.1016/J.APENERGY.2017.03.132>
- Gonzalez, T., Moreno, J. A., & Fridman, L. (2012). Variable Gain Super-Twisting Sliding Mode Control. *IEEE Transactions on Automatic Control*, 57(8), 2100-2105. <https://doi.org/10.1109/TAC.2011.2179878>
- Guler, N., Bayhan, S., Fesli, U., Blinov, A., & Vinnikov, D. (2023). Super Twisting Sliding Mode Control Strategy for Input Series Output Parallel Converters. *IEEE Access*, 11, 107394-107403. <https://doi.org/10.1109/ACCESS.2023.3320178>
- Hao, X., Salhi, I., Laghrouche, S., Ait-Amirat, Y., & Djerdir, A. (2021). Robust Control of Four-Phase Interleaved Boost Converter by Considering the Performance of PEM Fuel Cell Current. *International Journal of Hydrogen Energy*, 46(78), 38827-38840. <https://doi.org/10.1016/J.IJHYDENE.2021.09.132>
- Hao, X., Salhi, I., Laghrouche, S., Ait-Amirat, Y., & Djerdir, A. (2022). Backstepping Supertwisting Control of Four-Phase Interleaved Boost Converter for PEM Fuel Cell. *IEEE Transactions on Power Electronics*, 37(7), 7858-7870. <https://doi.org/10.1109/TPEL.2022.3149099>
- Hao, X., Salhi, I., Laghrouche, S., Ait Amirat, Y., & Djerdir, A. (2023). Multiple Inputs Multi-Phase Interleaved Boost Converter for Fuel Cell Systems Applications. *Renewable Energy*, 204, 521-531. <https://doi.org/10.1016/J.RENENE.2023.01.021>
- Kabalo, M., Paire, D., Blunier, B., Bouquain, D., Godoy Simões, M., & Miraoui, A. (2013). Experimental Evaluation of Four-Phase Floating Interleaved Boost Converter Design and Control for Fuel Cell Applications. *IET Power Electronics*, 6(2), 215-226. <https://doi.org/10.1049/IET-PEL.2012.0221>
- Khosroshahi, A., Abapour, M., & Sabahi, M. (2015). Reliability Evaluation of Conventional and Interleaved DC-DC Boost Converters. *IEEE Transactions on Power Electronics*, 30(10), 5821-5828. <https://doi.org/10.1109/TPEL.2014.2380829>
- Komurcugil, H., Biricik, S., Bayhan, S., & Zhang, Z. (2021). Sliding Mode Control: Overview of Its Applications in Power Converters. *IEEE Industrial Electronics Magazine*, 15(1), 40-49. <https://doi.org/10.1109/MIE.2020.2986165>

- Mallikarjuna Reddy, B., & Samuel, P. (2020). Analysis, Modelling and Implementation of Multi-Phase Single-Leg DC/DC Converter for Fuel Cell Hybrid Electric Vehicles. *International Journal of Modelling and Simulation*, 40(4), 279-290. <https://doi.org/10.1080/02286203.2019.1610689>
- Manoharan, Y., Hosseini, S. E., Butler, B., Alzahrani, H., Senior, B. T. F., Ashuri, T., & Krohn, J. (2019). Hydrogen Fuel Cell Vehicles; Current Status and Future Prospect. *Applied Sciences*, 9(11), 2296. <https://doi.org/10.3390/APP9112296>
- Mekhilef, S., Saidur, R., & Safari, A. (2012). Comparative Study of Different Fuel Cell Technologies. *Renewable and Sustainable Energy Reviews*, 16(1), 981-989. <https://doi.org/10.1016/J.RSER.2011.09.020>
- Nahar, S., & Uddin, M. B. (2018, September 13-15). *Analysis the Performance of Interleaved Boost Converter*. In: Proceedings of the 4th International Conference on Electrical Engineering and Information and Communication Technology (ICEEICT 2018) (pp. 547-551). Dhaka, Bangladesh. <https://doi.org/10.1109/CEEICT.2018.8628104>
- Napole, C., Derbeli, M., & Barambones, O. (2021). A Global Integral Terminal Sliding Mode Control Based on a Novel Reaching Law for a Proton Exchange Membrane Fuel Cell System. *Applied Energy*, 301, 117473. <https://doi.org/10.1016/J.APENERGY.2021.117473>
- Nikhar, A. R., Apte, S. M., & Somalwar, R. (2016, December 22-24). *Review of Various Control Techniques for DC-DC Interleaved Boost Converters*. In: Proceedings of the International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC 2016) (pp. 432-437). Jalgaon, India. <https://doi.org/10.1109/ICGTSPICC.2016.7955340>
- Sankar, K., Saravanakumar, G., & Jana, A. K. (2021). Nonlinear Multivariable Control of an Integrated PEM Fuel Cell System with a DC-DC Boost Converter. *Chemical Engineering Research and Design*, 167, 141-156. <https://doi.org/10.1016/J.CHERD.2021.01.011>
- Sazali, N., Wan Salleh, W. N., Jamaludin, A. S., & Mhd Razali, M. N. (2020). New Perspectives on Fuel Cell Technology: A Brief Review. *Membranes*, 10(5), 99. <https://doi.org/10.3390/MEMBRANES10050099>
- Schwenzer, M., Ay, M., Bergs, T., & Abel, D. (2021). Review on Model Predictive Control: An Engineering Perspective. *The International Journal of Advanced Manufacturing Technology*, 117(5-6), 1327-1349. <https://doi.org/10.1007/S00170-021-07682-3>
- Seyezhai, R., & Mathur, B. L. (2012). Design and Implementation of Interleaved Boost Converter for Fuel Cell Systems. *International Journal of Hydrogen Energy*, 37(4), 3897-3903. <https://doi.org/10.1016/J.IJHYDENE.2011.09.082>
- Zhuo, S., Xu, L., Huangfu, Y., Gaillard, A., Paire, D., & Gao, F. (2021). Robust Adaptive Control of Interleaved Boost Converter for Fuel Cell Application. *IEEE Transactions on Industry Applications*, 57(6), 6603-6610. <https://doi.org/10.1109/TIA.2021.3113262>

NOMENCLATURE

$\alpha - \beta$	Controller gains
C_o	Output capacitor
<i>CCM</i>	Continuous conduction mode
d_i	Duty cycles for the power switches
D_i	Diodes of the circuit
<i>EMI</i>	Electromagnetic interface
<i>IBC</i>	Interleaved boost converter
I_o	Output current
$i_{L_{ref}}$	Reference inductor current
K_p, K_i	Proportional and integral coefficients
L_i	Inductors of the circuit
<i>MPC</i>	Model predictive control
<i>PI</i>	Proportional-Integral
<i>PWM</i>	Pulse width modulation
r_i	Parasitic resistance
R_o	Resistive load
S_i	Power switches
<i>SMC</i>	Sliding mode control
<i>STSM</i>	Super-twisting sliding mode
V_i	Input voltage
V_o	Output voltage



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1533769>

Investigation of the Effect of Hexagonal Boron Nitride Addition on the Mechanical Properties of Flax Fiber-Reinforced Composite Materials

Ahmet ERYILMAZ¹ Hasan Yavuz ÜNAL¹ Yeliz PEKBEY^{1*} ¹ Ege University, Faculty of Engineering, Department of Mechanical Engineering, İzmir, Türkiye

Keywords	Abstract
Natural Fiber-Reinforced Composites Flax Fabric Flax-Epoxy Composites Vacuum Bagging Hexagonal Boron Nitride	Hexagonal boron nitride (h-BN) has recently been utilized as a reinforcement in composite materials due to its properties such as hardness, thermal conductivity, electrical insulation, and strong chemical stability. The aim of this study is to investigate the effect of nano-sized hexagonal boron nitride (h-BN) on the mechanical properties of flax fiber-reinforced composite material. For this purpose, initially, hexagonal boron nitride was added to epoxy resin in different weight ratios and homogenized without agglomeration using ultrasonic treatment. Then, by employing the hand lay-up method, the mixture was applied to flax fiber fabrics and the flax fiber-epoxy composites were produced using the vacuum bagging method. Mechanical performance of the composites, produced with 0.5%, 1%, and 1.5% by weight of hexagonal boron nitride, was determined through tensile, flexural, shear, and compression tests. Experimental results indicated that the addition of hexagonal boron nitride to flax fiber epoxy composite material increased the flexural strength and modulus compared to the unreinforced flax fiber epoxy composite material. The highest flexural strength and modulus were observed in the samples with 1.5% by weight of hexagonal boron nitride (h-BN). Consequently, it can be considered that flax fiber-epoxy composite material with hexagonal boron nitride (h-BN) addition holds potential, especially for applications subjected to bending moments.

Cite

Eryilmazi A., Ünal, H. Y., & Pekbey, Y. (2024). Investigation of the Effect of Hexagonal Boron Nitride Addition on the Mechanical Properties of Flax Fiber-Reinforced Composite Materials. *GU J Sci, Part A, 11(3)*, 577-588. doi:10.54287/guj.1533769

Author ID (ORCID Number)	Article Process
0009-0008-3917-8363	Submission Date 15.08.2024
0000-0003-1934-7445	Revision Date 12.09.2024
0000-0002-1024-8806	Accepted Date 19.09.2024
	Published Date 30.09.2024

1. INTRODUCTION

In today's technology, studies on nanotechnology continue rapidly due to its use in all kinds of applications. Boron nitride, which is an advanced technology product and is not constitute in nature, is yielded by chemical reaction of boron oxide, carbon and nitrogen at high temperatures. Since about 1950, it has been produced and then synthesized. Corrosion resistant crucibles have frequently used in engineering practices such as refractory in casting walls in applications requiring high temperature resistance, dielectric parts in electronic fields and armor technologies (Haubner et al., 2002).

Due to their different physical properties and applications, they are available in forms such as hexagonal (h-BN), cubic BN (c-BN), turbostratic BN(t-BN), wurtzite BN (w-BN), rhombohedral BN (r-BN), and amorphous BN (a-BN) (Chen et al., 2022). The hexagonal form of boron nitride is also known as white graphene in addition to its resistance to thermal shocks at high temperatures such as about 2000°C and its electrical insulation properties. Boron nitride in cubic form is the hardest material after diamond due to its short bond and is used in the processing and cutting of hard materials, especially at high temperatures and high speeds, and to reduce wear due to its lubricity. The fatigue life of parts made of boron nitride, especially in cubic form, is high due to their resistance to high temperature and hardness. Hexagonal form boron nitride is mostly used in applications.

*Corresponding Author, e-mail: yeliz.pekbey@ege.edu.tr

Because of its excellent thermal transmittance and electrical insulation properties, hexagonal boron nitride (h-BN) is regarded a promising filler in high-voltage insulation engineering applications for polymer-based composites (Meng et al., 2014). Kusunose and Sekino (2016) found significant improvement in the thermal conductivity of materials sintered with hexagonal boron nitride in the direction of hot pressing and perpendicular to this direction. Additionally, literature supports that thermal conductivity increases in composites produced by various methods with added hexagonal boron nitride. These studies suggest that incorporating h-BN into nanocomposites can form materials capable of withstanding extreme conditions and efficiently dissipating heat, offering a promising path for thermal management applications (Wang et al., 2023; Xu et al., 2023; Liu et al., 2024; Jeong et al., 2024; Jiang et al., 2024).

Some investigations have been carried out to reduce the agglomeration of h-BN at high ratios in the resin. Li et al. (2002) investigated the machinability of $\text{Al}_2\text{O}_3/\text{BN}$ composite ceramics with nanosized boron nitride dispersions ranging from 0 to 30% by volume by hot pressing of turbostratic BN (t-BN) coated $\alpha\text{-Al}_2\text{O}_3$ powders prepared by chemical processes using boric acid. They found that nanocomposite ceramics with more than 20% boron nitride content by volume exhibited excellent machinability compared to conventional hard materials that can be drilled using metal alloy drills.

Eichler and Lesniak (2008) reviewed boron nitride (BN) and BN composites for high-temperature applications, and Karim et al. (2023) provided a literature review on h-BN-based ceramics and their composites with oxides, nitrides, carbides, and metals based on different production methods.

Wen et al. (2000) produced a BN- SiO_2 composites for high-temperature dielectric material applications via a mixture of BN and glass powders. Similarly, BN-reinforced SiO_2 (BNW/ SiO_2) ceramics, varying from 5% to 20% by mass, were produced by hot pressing at temperatures of 1250°C, 1300°C, 1350°C, and 1450°C for 10 minutes and 30 MPa pressure, and their microstructure, mechanical, and dielectric properties were investigated (Duan et al., 2023).

Yang et al. (2024) investigated the thermal conductivity of epoxy-based nanocomposites added boron nitride-based nanostructures using both experimental and atomistic numerical simulations (smoothed-particle hydrodynamics (SPH)). The numerical study examined effects of nano filler volume fraction, aspect ratio, and orientation on thermal conductivity.

Boron nitride is also used in the literature as an additive in composite materials. In recent years, it has been observed that nanomaterials enhance performance of composite materials when added to composite materials (Jia et al., 2011). Zou et al. (2024) investigated mechanical and dielectric properties of Si_3N_4 -BN composite to evaluate effect of hexagonal boron nitride (h-BN) content. The study found that as the BN content increased, the relative density of Si_3N_4 -BN-MAS composites decreased from 92.3% to 79.9%, flexural strength decreased from 1000 MPa to 225 MPa, and dielectric constant decreased from 7.0 to 4.6. However, the highest fracture toughness was observed with 20% by weight BN content.

Srikhar and Omprakash (2024) examined the mechanical properties of aluminum hybrid metal matrix composites reinforced with varying weight percentages (2%, 3%, 4%, 5%, and 6%) of boron nitride (BN) produced by sintering. Composites including 5% by weight BN displayed the highest tensile strength and hardness, while composites with 4% by weight BN exhibited the highest flexural strength. This improvement in mechanical properties achieved since the BN particles within the aluminum matrix are homogeneously distributed acting as an effective load-bearing and reinforcing agent.

Demircan and Kalaycı (2024) investigated the shear strength of glass and carbon fiber reinforced thermoset composites by using epoxy containing different (0, 1, 2 and 3 wt%) nano hexagonal boron nitride (nano-h-BN) by weight. The shear strengths of the joints bonded with epoxy adhesives containing 3 wt% nano hexagonal boron nitride (nano-h-BN) were the highest.

To the authors' knowledge, there is no publication in the literature in which boron nitride is added as a secondary reinforcement to natural fiber reinforced composites and the mechanical properties are characterized. In this study, three different weight ratios of boron nitride nanoparticles were mixed with epoxy

resin to produce nano-reinforced composite materials. The main purpose of this research is to ascertain the impact of boron nitride, which exhibits favorable thermal and electrical insulation characteristics, on the mechanical properties of fiber-reinforced polymers. In order to gain this objective, boron nitride was incorporated into the resin at concentrations of 0.5, 1, and 1.5% by weight. The mechanical properties of the nano-reinforced samples as well as the composite sample without reinforcement were determined by tensile, three-point bending, compression and shear tests.

2. MATERIAL AND METHOD

Flax Fiber

The dmc linen (flax) woven fabrics used as fiber reinforcement material were purchased from Bursa İpek Tekstil San. Ve Tic. Ltd. Şti. The fabric is of plain weave type (1x1) and has a mass of 400 g/m². It has a weaving density of 13 ct, representing the number of stitches per inch (2.54 cm), which corresponds to 5.12 stitches. This equates to a stitch count of 5.12 pts/cm, indicating the number of stitches per centimeter. Figure 1a shows an image of the linen fabric used as fiber in the production of composite plates. Since the obtained fabrics may contain residues, dust, and other unwanted contaminants, they were pre-washed with tap water (Figure 1b). The purpose of this pre-treatment is to make the fabric surfaces cleaner, thereby improving the adhesion between the matrix and the linen fabric. The washed linen fabrics were left to dry at room temperature (Figure 1c).

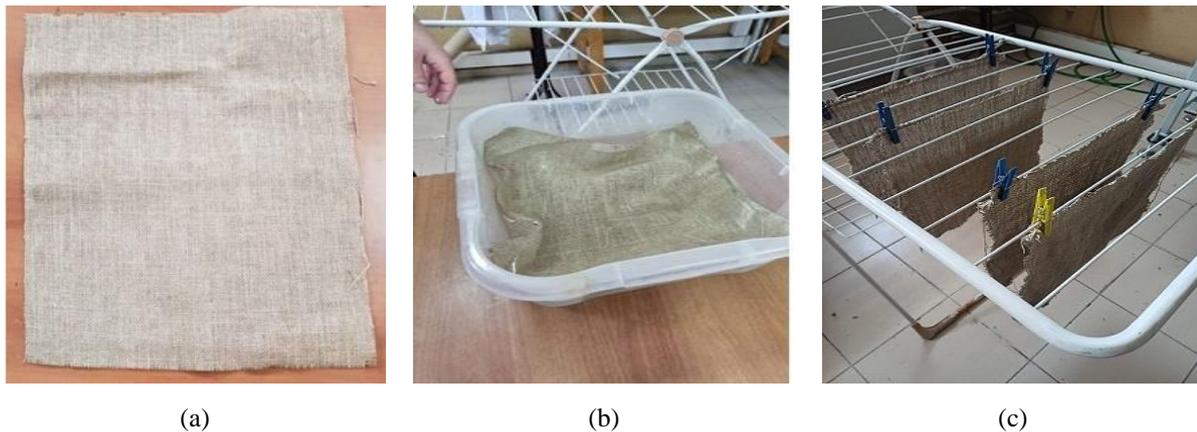


Figure 1. Flax Woven Fabric and Its Cleaning Process
a) Purchased Linen Fabric; b) Washing and c) Dried Linen Fabrics

Epoxy and Hardener

Epoxy resin with a viscosity of 700-900 mPas and a density of 1.13-1.17 g/cm³, MGS Laminating Epoxy Resin L160, and a hardener with a viscosity of 10-50 mPas, MGS Laminating Epoxy Hardener H160, were used. The epoxy resin and hardener were bought from Dost Kimya.

Hexagonal Boron Nitride Powder

Hexagonal boron nitride powder with a purity greater than 99% and a particle size of 65-75 microns was obtained from Nanografi. The properties of the hexagonal boron nitride used, as detailed in Table 1, were provided by the manufacturer and supplemented by literature (Srikhar & Omprakash, 2024).

Figure 2 schematically illustrates the preparation of composite materials with epoxy resin reinforced with flax fibers and hexagonal boron nitride (h-BN) for determining their mechanical performance (Eryılmaz, 2024).

Table 1. Mechanical Properties of Hexagonal Boron Nitride

Element Analysis (%)	Fe ₂ O ₃	CaO	MgO	B ₂ O ₃
		0,03	0,002	0,04
Purity (%)	99,85			
Average Particle Size (nm)	65-75			
Melting Point (°C)	3000			
Boiling Point (°C)	3927			
Hardness (Knoop 100 g) (kg.mm ²)	400–3000			
Density (g/cc)	2,3			
Fracture Toughness (MPa.m ^{-1/2})	2,5–6,5			
Young Modulus (GPa)	20–103			
Electrical Conductivity (at 25°C) (S)	140			
Thermal Conductivity (at 25°C) (W/m-K)	20–27			

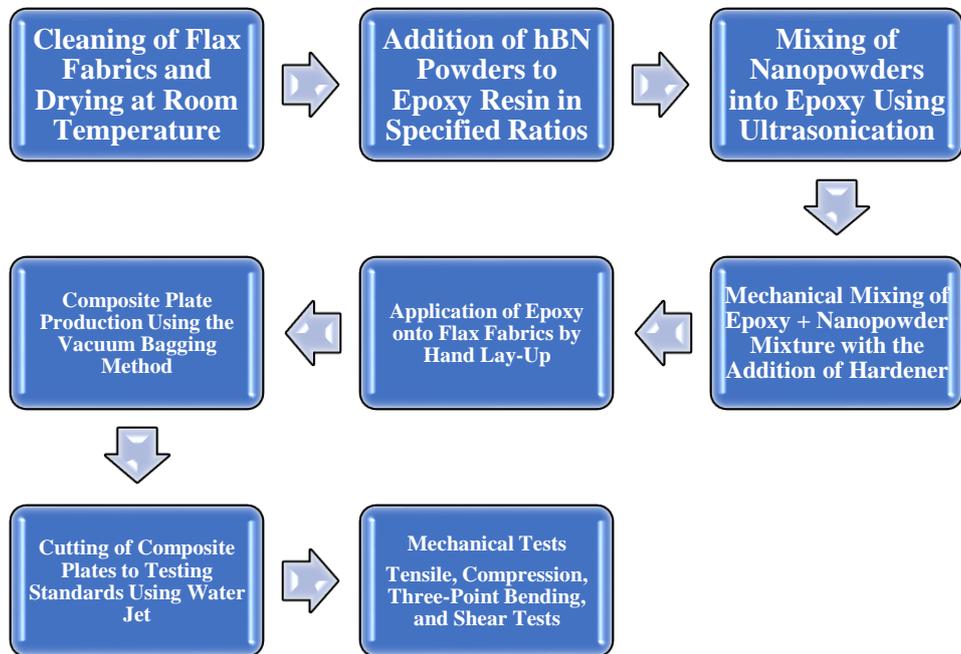
**Figure 2. Production Stages of Flax Fiber- (h-BN) Reinforced Epoxy Resin Composites**

Figure 3 shows supplementation of hexagonal boron nitride (h-BN) nanopowders in various weight ratios into epoxy resin.

**Figure 3. Addition of Hexagonal Boron Nitride (h-BN) Nanopowders in Different Weight Ratios into Epoxy Resin**

Ultrasonication

The process of homogeneously mixing hexagonal boron nitride was carried out using a Hielscher UP400St ultrasonic mixer. The probe of the mixer was positioned 2 cm below the surface of the resin, and mixing was performed. Ultrasonic sound waves facilitated the dispersion of hexagonal boron nitride within the resin, resulting in a homogeneous mixture. To prevent the resin from curing due to the temperature increase during the mixing process, the container was cooled with another vessel filled with water and ice (Figure 4).



Figure 4. Mixing of Hexagonal Boron Nitride into Epoxy Using an Ultrasonic Mixer

Production of Composite Plates

Flax fabric measuring 38 x 38 cm² was placed on a release film, and the prepared resin mixture was applied with a brush until all fibers were saturated. A second layer of fabric was then placed, and the process was continued until a total of four layers were achieved. The application of the epoxy matrix prepared with a brush on the flax fabrics is shown in Figure 5.



Figure 5. Application of (h-BN) Nanopowder Reinforced Epoxy Resin onto Flax Fabrics by Hand Lay-Up

To facilitate the production process and ensure higher quality parts, a 45 x 55 cm peel ply was used (Figure 6). The peel ply placed on the top surface of the composite piece was used to remove air bubbles, enhance the even distribution of the matrix in a vacuum environment, and facilitate the easier separation of the produced plate from the system. The use of vacuum bags improved the production process by aiding in the absorption of excess resin. The vacuum blanket absorbed excess resin within the vacuum bag and reduced the amount of resin that could escape into the vacuum hose, allowing the vacuum pump to operate more safely.

After cutting the vacuum cover (nylon) to the required dimensions, spiral hoses were attached to the T-connector fittings on the vacuum hose. Sealing tapes were placed around these hoses to prevent air leaks from the corners. Finally, the vacuum cover was sealed with sealing tape to ensure that air ingress and egress were blocked. Once the system was prepared for vacuuming, the vacuum process was initiated (Figure 6). During this process, the composite plate underwent a thermal treatment from the surface where the vacuum bagging was performed. On the recommendation of the supplier, the initial temperature was set at 60°C, maintained for eight hours. Immediately afterwards, it was allowed to cool to room temperature. The cooled composite

plate was carefully removed from the vacuum bag. As a result of the production, four composite plates measuring 38 x 38 cm were obtained, with one being unreinforced and the other three containing hexagonal boron nitride at weight percentages of 0.5%, 1.0%, and 1.5%, respectively.

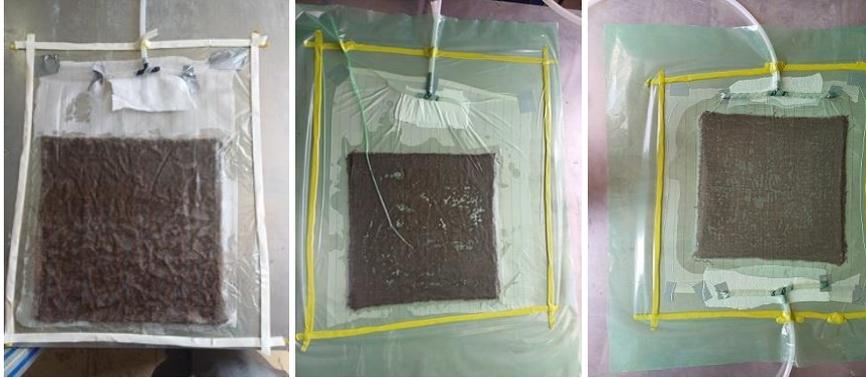


Figure 6. Production of Composite Plates Using the Vacuum Bagging Method

Standard Cutting of Plates

The plates were cut using a water jet (Figure 7). The plates were cut with high precision using a mixture of diamond and sand in addition to pure water at 3000 bar pressure, and each test specimen was cut to standard dimensions.



Figure 7. Cutting of Composite Plates to Testing Standards Using Water Jet

3. EXPERIMENTAL METHODS

To determine the mechanical performance of the produced composite plates, tensile, compressive, three-point bending, and shear tests were conducted (Figures 8-9). For each mechanical test, at least three samples were tested.

Mechanical tests were conducted according to standards using the Shimadzu Autograph AG-IS uniaxial 100 kN testing machine located at the Biomechanics Laboratory of the Department of Mechanical Engineering, Ege University. The tensile test was performed according to ASTM D3039-17. Tensile tests were performed at a rate of 2 mm/min on three samples per plate, with each sample measuring 250 mm in length and 25 mm in width. The elongation of two parallel lines drawn on the sample was monitored with a dual-camera system to obtain the unit strain value. The tensile test results determined the maximum normal stress, the unit strain at this stress, the unit strain at fracture, and the modulus of elasticity.

ASTM D7264-21 at a speed of 1 mm/min was carried out for the three-point bending test. The sample used in the tests had a width of 13 mm and a span distance between supports that was 16 times the thickness. The bending strength and unit strain values were computed using Equation 1 and Equation 2. The bending modulus was figured out slope of the stress-strain curve.

$$\sigma_{bending} = \frac{3.F.L}{2.w.t^2} \quad (1)$$

$$\varepsilon = \frac{6.\delta.t}{L^2} \quad (2)$$

In Equations (1) and (2), $\sigma_{bending}$: represents bending stress [MPa], F : the applied load [N], L : the span distance [mm], w : width [mm], t : thickness [mm] and δ : the vertical displacement [mm].

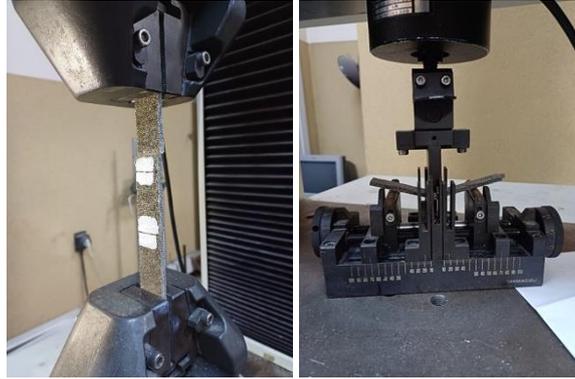


Figure 8. Tensile and Three-Point Bending Test Apparatus

The shear test was controlled according to ASTM D7078-20. At a constant speed of 2 mm/min, shear stress was applied by moving one half of the sample upward. Shear stress was calculated using the formulas in Equation 3 and Equation 4. Since strain gauges were not used in the tests, unit strain was not calculated.

$$\tau_i = \frac{F_i}{A} \quad (3)$$

$$A = w.t \quad (4)$$

In Equation 3 and Equation 4, τ_i represents shear stress at point i [MPa], F_i represents shear force at point i [N], A represents cross-sectional area [mm²], w represents the width [mm], and t represents the thickness [mm].

The compressive test was conducted according to ASTM D6641-16. Using the apparatus shown in Figure 9, only compressive stress was obtained without causing buckling. The sample dimensions were set to 140 mm in length and 13 mm in width, with a measurement distance of 13 mm. The test speed was applied at 1.3 mm/min. Compressive stress was calculated using the formulas in Equation 5 and Equation 6. Since strain gauges were not used in the tests, unit strain was not calculated.

$$\sigma_{compressive} = \frac{F_i}{A} \quad (5)$$

$$A = w.t \quad (6)$$

In Equation 5 and Equation 6, $\sigma_{compression}$ represents compressive stress at point i [MPa], F_i represents compressive force at point i [N], A represents cross-sectional area [mm²], w represents the width [mm], and t represents the thickness [mm].



Figure 9. Shear and Compression Test Apparatus

4. RESULTS AND DISCUSSION

Tensile Test Results

Typical results from the tensile tests of epoxy resin composites with flax fibers and hexagonal boron nitride (h-BN) nanopowder additions under constant load are depicted in Figure 10 and Table 2. For composites without hexagonal boron nitride (h-BN), the modulus of elasticity and tensile strength were 3.71 GPa and 53.02 MPa, respectively, while the addition of hexagonal boron nitride reduced the tensile properties. The modulus of elasticity showed the greatest decrease in the sample with 1% h-BN by weight. The tensile strength decreased by 36% in the sample with 0.5% h-BN by weight, and by 28% and 24% in the samples with 1% and 1.5% h-BN by weight, respectively.

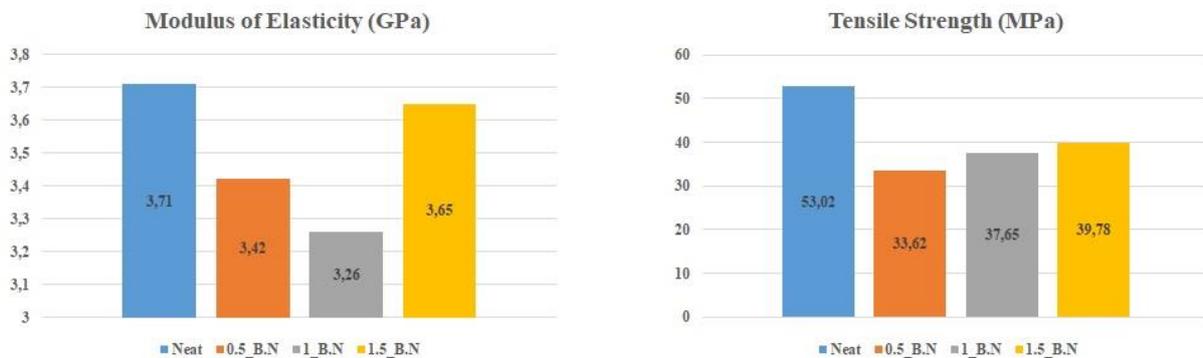


Figure 10. Comparing to Modulus of Elasticity and Tensile Strengths of Epoxy Composites with Varying Weight Percentages of (h-BN)

When examining area under the typical stress-strain curve for hexagonal boron nitride (h-BN) changes in flax fiber-epoxy composites, i.e., toughness, it is observed that the highest toughness values were found in the unreinforced samples, with the 1.5% h-BN reinforced composites showing the best performance. Table 2 compares the mechanical properties of flax fiber-epoxy composites with different weight percentages of hexagonal boron nitride (h-BN).

Table 2. Comparison of Mechanical Properties with Variation in Weight Percentages of (h-BN) in Flax Fiber-Epoxy Composites

Samples	Modulus of Elasticity (GPa)	Tensile Strength (MPa)	Max. Load (kN)	Strain (%)	Toughness (kJ/m ³)
Neat	3.71 ± 0.16	53.02 ± 1.48	6.22 ± 0.21	1.66 ± 0.04	549.89 ± 37.70
0.5% h-BN	3.42 ± 0.08	33.62 ± 3.00	4.64 ± 0.18	1.06 ± 0.14	208.40 ± 56.50
1% h-BN	3.26 ± 0.15	37.65 ± 0.91	5.33 ± 0.06	1.13 ± 0.07	249.78 ± 23.26
1.5% h-BN	3.65 ± 0.11	39.78 ± 3.61	5.43 ± 0.29	1.27 ± 0.13	303.87 ± 64.09

Three-Point Bending Test Results

The results of the three-point bending tests are presented in Figure 11 and Table 3. The addition of hexagonal boron nitride (h-BN) caused an increase the bending modulus of the flax fiber-epoxy composites, with the greatest increase observed in samples with 1.5% h-BN by weight. A similar trend was seen in bending strength. The bending modulus and bending strength increased by 24.4% and 14.7%, respectively, compared to unreinforced flax fiber-epoxy composites. The addition of hexagonal boron nitride (h-BN) reduced unit strain.

Table 3. Comparison of Bending Properties with Variation in Weight Percentages of (h-BN) in Flax Fiber-Epoxy Composites

Samples	Flexural Modulus (GPa)	Flexural Strength (MPa)	Strain (%)
Neat	3.28 ± 0.05	73.04 ± 3.19	3.79 ± 0.42
0.5% h-BN	3.65 ± 0.20	80.36 ± 4.12	2.95 ± 0.16
1% h-BN	3.49 ± 0.14	61.37 ± 4.84	2.35 ± 0.36
1.5% h-BN	4.08 ± 0.26	83.77 ± 3.65	3.24 ± 0.22

The increase in bending strength with the addition of hexagonal boron nitride (h-BN) improved the rigidity of the epoxy while also enhancing its ductility. However, compared to unreinforced samples, tensile strengths were reduced. This reduction may be attributed to poor adhesion of hexagonal boron nitride (h-BN) within the epoxy resin and the occurrence of agglomeration during ultrasonication.

Similar results have shown in literature. In 2019, Kartal and Boztoprak produced boron nitride particle reinforced vinylester composite materials using vinyl ester resin as the matrix material and hexagonal boron nitride particles at different ratios as reinforcing material and investigated tensile strength, flexural strength, impact strength, abrasion resistance. The addition of boron nitride into vinyl ester resin started at 0.5 wt% and continued up to 2 wt%. A decrease in tensile, % elongation and bending properties of the composite material obtained was observed (Kartal & Boztoprak, 2019).

The % elongation decreases along with the tensile strength, which indicates that ductility decreases and brittleness increases with the increase in the amount of boron nitride, so the 1.5 wt% boron nitride reinforced composite material breaks more quickly due to being more brittle. In addition, the formation of agglomeration also contributed to the decrease in tensile strength. Agglomeration formation also has an effect on the decrease in tensile strength.

Although mixing was done manually and in an ultrasonic homogenizer to prevent agglomeration, agglomeration still occurred in some areas.



Figure 11. Comparison of Bending Modulus and Bending Strength with Variation in Weight Percentages of (h-BN)

Shear and Compressive Test Results

Figure 12 shows typical stress-displacement curves from shear tests. The shear strength for unreinforced flax fiber-epoxy composites was 35.66 MPa. For specimen with 0.5%, 1%, and 1.5% h-BN by weight, the shear strengths were 33.91 MPa, 32.97 MPa, and 33.44 MPa, respectively. Shear strengths decreased by approximately 4.91%, 8.05%, and 6.23% compared to unreinforced samples.

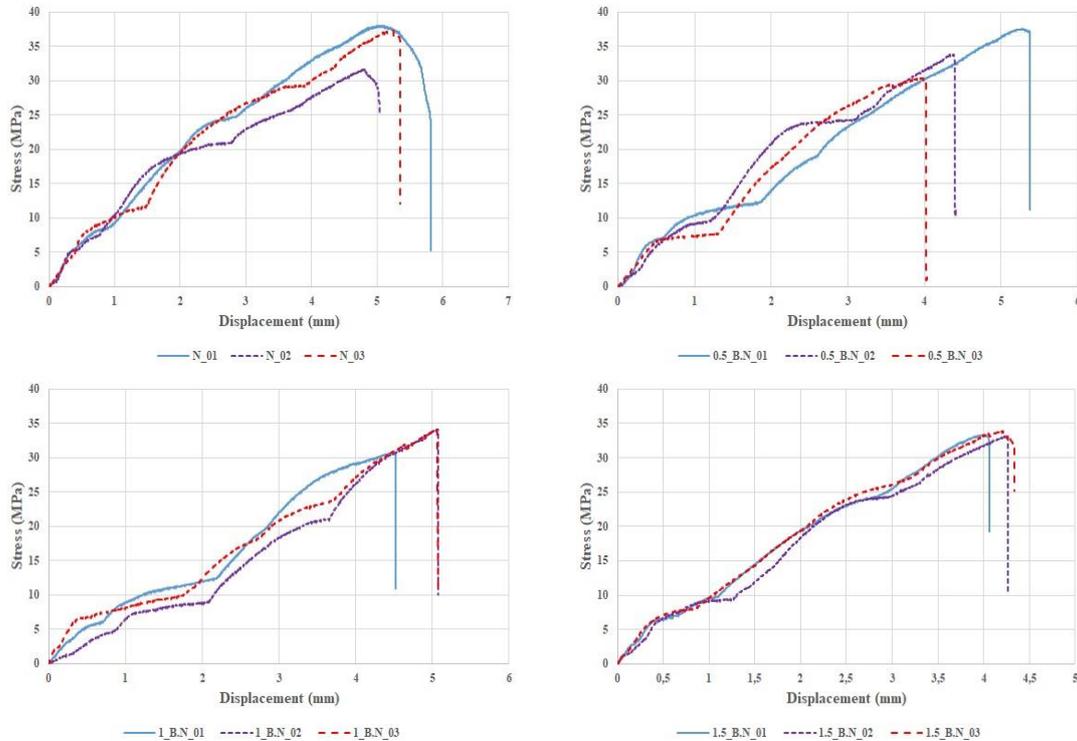


Figure 12. Shear Test Stress-Strain Curves for Flax Fiber- (h-BN) Reinforced Epoxy Resin Composites

Table 4 shows the results from the compressive tests. The compressive strength for unreinforced flax fiber-epoxy composites was 75.73 MPa. The highest increase in compressive strength was 15.2% in specimen with 1% h-BN by weight. Additionally, specimen with 0.5% and 1.5% h-BN showed increases in compressive strength of 12.9% and 8.4%, respectively.

In terms of the atomic structure/array characteristics of hexagonal boron nitride, it is in the form of layers (laminar) like the structure of graphite and made up of hexagonal rings formed by boron and nitrogen atoms. There are strong covalent bonds between the atoms forming the rings (Watanabe & Taiguchi, 2011). There are Van der Waals bonds between the layers consisting of rings formed by strong covalent bonds. As shown in Table 4, in compressive strength tests, in specimens with more h-BN reinforcement by weight, more compressive strength is necessitated to break the strong covalent bonds between these layers. Therefore, compressive strength is considered to increase in compressive strength tests.

Table 4. Comparison of Shear and Compression Test Properties with Variation in Weight Percentages of (h-BN) in Flax Fiber-Epoxy Composites

Samples	Shear Strength (MPa)	Compressive Strength (MPa)
Neat	35.66 ± 3.53	75.73 ± 4.07
0.5% h-BN	33.91 ± 3.58	85.53 ± 5.14
1% h-BN	32.97 ± 1.92	87.24 ± 3.94
1.5% h-BN	33.44 ± 0.35	82.11 ± 6.40

5. CONCLUSION

This study investigated the effects of hexagonal boron nitride (h-BN) addition on the mechanical performance of flax fiber-epoxy composite materials. For this purpose, composites containing 0.5%, 1%, and 1.5% by weight of hexagonal boron nitride (h-BN) were produced and changes in tensile, bending, shear, and compressive tests were examined. Adding hexagonal boron nitride (h-BN) at specific levels caused enhancements in bending strength and bending modulus of the composites. Composites containing 1.5% h-BN by weight exhibited the highest bending strength and bending modulus. This improvement in mechanical properties can be attributed to the homogeneous dispensation of h-BN particles within the epoxy matrix and good adhesion with the flax fibers.

However, when the h-BN content exceeded a certain weight, no further improvement in mechanical properties was observed. This is likely due to the agglomeration of h-BN nanopowders within the epoxy and their non-uniform distribution. The experimental results suggest that flax fiber-hexagonal boron nitride (h-BN) epoxy composites could meet needs for high-strength applications.

AUTHOR CONTRIBUTIONS

Conceptualization, A.E., H.Y.Ü. and Y.P.; methodology, A.E., H.Y.Ü. and Y.P.; fieldwork, A.E., H.Y.Ü. and Y.P.; software, A.E., H.Y.Ü. and Y.P.; title, A.E. and Y.P.; validation, A.E., H.Y.Ü. and Y.P.; laboratory work, A.E. and H.Y.Ü.; formal analysis, H.Y.Ü. and Y.P.; research, A.E. and Y.P.; sources, A.E., H.Y.Ü. and Y.P.; data curation, A.E. and H.Y.Ü.; manuscript-original draft, A.E. and Y.P.; manuscript-review and editing, A.E., H.Y.Ü. and Y.P.; visualization, H.Y.Ü. and Y.P.; supervision, Y.P.; project management, Y.P.; funding, A.E., H.Y.Ü. and Y.P.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Chen, M., Pan, L., Xia, X., Zhou, W., & Li, Y. (2022). Boron nitride (BN) and BN based multiple-layer interphase for SiC/SiC composites: A review. *Ceramics International*, 48(23), Part A, 34107-34127. <https://doi.org/10.1016/j.ceramint.2022.07.021>
- Demircan, O., & Kalaycı, A. (2024). Investigation of mechanical properties of nano hexagonal boron nitride integrated epoxy adhesives in bonded GFRP and CFRP composite joints. *International Journal of Adhesion and Adhesives*, 133, 103759. <https://doi.org/10.1016/j.ijadhadh.2024.103759>
- Duan, W., Yang, H., Yang, Z., Li, B., Jia, D., & Zhou, Y. (2023). The microstructural evolution, mechanical and dielectric properties of BN_w/SiO₂ ceramics. *Ceramics International*, 49(8), 11968-11977. <https://doi.org/10.1016/j.ceramint.2022.12.046>
- Eichler, J. & Lesniak, C. (2008). Boron nitride (BN) and BN composites for high-temperature applications. *Journal of the European Ceramic Society*, 28, 1105-1109. <https://doi.org/10.1016/j.jeurceramsoc.2007.09.005>
- Eryılmaz, A. (2024). Keten Fiberli-Hekzagonal Bor Nitridür Katkılı Epoksi Reçineli Kompozit Malzemelerin Mekanik Performanslarının Belirlenmesi [BSc Thesis, University of Ege]. Ege University Repository.
- Haubner, R., Wilhelm, M., Weissenbacher, R., & Lux, B. (2002). Boron nitrides — properties, synthesis and applications. In: M. Jansen (Eds.), *High Performance Non-Oxide Ceramics II. Structure and Bonding* (pp. 1-45). Springer. https://doi.org/10.1007/3-540-45623-6_1
- Jeong, H., Lyu, J., Choi, H., Kim, M. W., Kim, J., Yoo, H., Lee, Y., Youk, J. H., & Chae, H. G. (2024). Enhanced thermal conductivity and mechanical property via improvement of hydrogen bonding between hexagonal boron nitride and aramid copolymer. *Composites Science and Technology*, 253, 110652. <https://doi.org/10.1016/j.compscitech.2024.110652>
- Jia, D., Zhou, L., Yang, Z., Duan, X., & Zhou, Y. (2011). Effect of Preforming Process and Starting Fused SiO₂ Particle Size on Microstructure and Mechanical Properties of Pressurelessly Sintered BN_p/SiO₂ Ceramic Composites. *Journal of American Ceramic Society*, 94(10), 3552-3560. <https://doi.org/10.1111/j.1551-2916.2011.04540.x>

- Jiang, H., Cai, Q., Mateti, S., Bhattacharjee, A., Yu, Y., Zeng, X., Sun, R., Huang, S., & Chen, Y. I. (2024). Recent research advances in hexagonal boron nitride/polymer nanocomposites with isotropic thermal conductivity. *Advanced Nanocomposites*, 1(1), 144-156. <https://doi.org/10.1016/j.adna.2024.03.004>
- Karim, M. R. A., Khan, M. A., Zaman, A. U., & Hussain, A. (2023). Hexagonal boron nitride-based composites: an overview of processing approaches and mechanical properties, *Journal of the Korean Ceramic Society*, 60, 1-23. <https://doi.org/10.1007/s43207-022-00251-8>
- Kartal, İ., & Boztoprak, Y. (2019). Investigation of Mechanical Properties of Vinyl Ester Composites Reinforced With Boron Nitride Particles, *El-Cezeri Journal of Science and Engineering*, 6(1), 43-50. <https://doi.org/10.31202/ecjse.450790>
- Kusunose, T., & Sekino, T. (2016). Thermal conductivity of hot-pressed hexagonal boron nitride. *Scripta Materialia*, 124, 138-141. <https://doi.org/10.1016/j.scriptamat.2016.07.011>
- Li, Y., Qiao, G., & Jin, Z. (2002). Machinable Al₂O₃/BN composite ceramics with strong mechanical properties. *Materials Research Bulletin*, 37(8), 1401-1409. [https://doi.org/10.1016/S0025-5408\(02\)00786-9](https://doi.org/10.1016/S0025-5408(02)00786-9)
- Liu, M., Zhang, H., Wu, Y., Wang, D., & Pan, L. (2024). Effect of functionalization on thermal conductivity of hexagonal boron nitride/epoxy composites. *International Journal of Heat and Mass Transfer*, 219, 124844. <https://doi.org/10.1016/j.ijheatmasstransfer.2023.124844>
- Meng, W., Huang, Y., Fu, Y., Wang, Z., & Zhi, C. (2014). Polymer composites of boron nitride nanotubes and nanosheets. *Journal of Materials Chemistry C*, 47, 10049-10061. <https://doi.org/10.1039/C4TC01998A>
- Srikhar, K., & Omprakash, B. (in press). Evaluation of mechanical behaviour of aluminium 2024 composites with boron nitride. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2024.05.057>
- Yang, X., Zhang, X., Yu, T., Li, Y., & Kirca, M. (2024). Coarse-grained simulation of thermal conductivity of boron nitride/epoxy composites based on DPD and SPH method. *Computational Materials Science*, 241, 113036. <https://doi.org/10.1016/j.commatsci.2024.113036>
- Wang, S. G., Feng, H. F., & Guo, Z.-X. (2023). Stacking and thickness effects on cross-plane thermal conductivity of hexagonal boron nitride. *Computational Materials Science*, 228, 112345. <https://doi.org/10.1016/j.commatsci.2023.112345>
- Wen, G., Wu, G. L., Lei, T. Q., Zhou, Y., & Guo, Z. X. (2000). Co-enhanced SiO₂-BN ceramics for high-temperature dielectric applications. *Journal of the European Ceramic Society*, 20(12), 1923-1928. [https://doi.org/10.1016/S0955-2219\(00\)00107-2](https://doi.org/10.1016/S0955-2219(00)00107-2)
- Watanabe, K., Taiguchi, T. (2011). Hexagonal Boron Nitride as a New Ultraviolet Luminescent Material and Its Application. *International Journal of Applied Ceramic Technology*, 5, 977-989. <https://doi.org/10.1111/j.1744-7402.2011.02626.x>
- Xu, Z., Bai, L., Zhang, Y., Cao, J., & Zheng, J. (2023). Mechanically strong, reprocessable and thermally stable PDMS/Hexagonal boron nitride composites with desirable thermal conductivities based on dynamic silyl ether linkages. *Composites Science and Technology*, 240, 110076. <https://doi.org/10.1016/j.compscitech.2023.110076>
- Zou, W., Cai, D., Wang, B., Zhu, Q., Yang, Z., Duan, X., Duan, W., Jia, D., & Zhou, Y. (2024). Effect of boron nitride content on mechanical, dielectric and thermal shock resistance properties of Si₃N₄-BN-MAS composites. *Ceramics International*, 50(17), Part A, 29866-29876. <https://doi.org/10.1016/j.ceramint.2024.05.282>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1529814>

Kinetics of Colour, Clarity Changes and HMF Formation in Pear Juice Concentrate During Storage

Feryal KARADENİZ^{1*} Betül IŞIK¹ Soner KAYA¹ Osman ASLANALI¹ Fatma MİDİLLİ¹ ¹ Göknur Gıda R&D Center, Merkez-Niğde, Türkiye

Keywords	Abstract
Pear Juice Concentrate	In terms of providing delicious taste and excellent source of nutritional content, pears are highly consumed fruit all around the world. Pear juice concentrate (PJC) is one of the most frequently used products in the manufacturing processes of beverages, syrup base for canning fruits, vinegar, and wine. Colour, clarity, and the formation of hydroxy methyl furfural (HMF), a mark of Maillard browning reactions, were appraised as quality indices changes of PJC throughout storage. The samples were kept at 47°, 37°, and 27°C for 32 weeks and kinetic criteria were calculated for changes in color, clarity, and the generation of HMF. According to the model of zero-order reaction, the results demonstrated that the quantity of HMF rose linearly with temperature and storage period. The values of clarity and colour were both linearly reduced to align with the zero-order reaction kinetic. The dependence on temperature of the rate constant of the reactions examined and the activation energy values were calculated as 153.14, 93.81 and 61.48 kJ mol ⁻¹ for the generation of HMF, clarity, and colour changes with the help of Arrhenius equation.
Kinetics	
Colour	
HMF	
Storage	
Clarity	

Cite

Karadeniz, F., Işık, B., Kaya, S., Aslanali, O., & Midilli, F. (2024). Kinetics of Colour, Clarity Changes and HMF Formation in Pear Juice Concentrate During Storage. *GU J Sci, Part A, 11(3)*, 589-597. doi:10.54287/guj.1529814

Author ID (ORCID Number)	Article Process
0000-0001-6367-5660	Submission Date 07.08.2024
0000-0002-2722-5862	Revision Date 14.08.2024
0000-0001-7363-394X	Accepted Date 28.08.2024
0000-0002-6432-1087	Published Date 30.09.2024
0009-0004-8942-4172	

1. INTRODUCTION

Pear, one of the eldest cultivated species, is among the widely consumed fruits in the world thanks to its high nutritious content and satisfying taste (Nikicevic, 2005). Between the years of 1994 and 2022 top 10 pear producers were ranked as China, USA, Italy, Argentina, Spain, Türkiye, South Africa, Japan, Republic of Korea, and Belgium. In the year of 2022 Türkiye's pear harvested area was 23.706 ha while pear production was 551.086 tons (FAOSTAT, 2024).

Pear not only provides a great source of carbohydrates, phenolic compounds, vitamins, and minerals but also high amount of dietary fiber, which popularizes them in the production of food purees for babies and mix nectars (Saeduddin et al., 2015). Additionally, vinegar, wine, beverages, and syrup base for fruit canning are also among the uses of pear juice concentrate (Cornwell & Wrolstad, 1981). As a chemo-preventive agent, chlorogenic acid, one of the main phenolic compounds in pears, helps patients avoid chronic illnesses like cardiovascular disease and cancer. It also strengthens the immune system and lessens the side effects of chemotherapy medications. It is known that pears have been utilized as herbal derived medicine to mitigate cough, constipation, and alcoholism for more than 2000 years in China. While pears have low levels of fat and protein, they are abundant in vitamins C, E, and B complex as well as potassium, copper, and fiber (Öztürk et al., 2015).

*Corresponding Author, e-mail: fkardeniz@goknur.com.tr

Pears should be conserved or processed in time following the harvest period so as not to encounter quality changes through handling. For instance, pear juice, is prone to browning easily during processing and storage. Once brown color has occurred, the color of pear juice becomes an unacceptable dark color from attractive golden yellow. Thus, the commodity characteristics of pear juice, such as its nutritional value, organoleptic features, and marketability decreases, and manufacturers are mostly subjected to large losses and customer rejections (Cornwell & Wrolstad, 1981). Browning observed in fruits and vegetables stems from enzymatic and non-enzymatic browning reactions. For concentrated fruit juices browning is mainly occurring due to the reactions related to non-enzymatic browning produced by the Maillard reaction (MR) leading to black and brown pigments formation as well as differences in odor and taste (Ibarz et al., 2008). The nonenzymatic browning reactions that result from MR, degradations of pigments, caramelization, and ascorbic acid are the primary reasons of browning in processed foods (Cornwell & Wrolstad, 1981). It is widely known that the Maillard reaction causes to HMF formation, which serves as measure for the degree of heat treatment performed (Lee & Nagy, 1988; Zhu et al., 2023). MR is also observed through the storage and the processing of fruit juice concentrates. Chen et al. (2024) reported that the Maillard reaction is the main reason for browning in orange juice concentrate during storage. Maillard reaction is what causes the browning and unpleasant attributes in apple and pear juice (Kathuria et al., 2023).

Since foods' organoleptic characteristics are of high importance for consumers, the investigation of the process and storage conditions of the products is considered critical. The kinetic process of chemical browning reactions has been examined by numerous researchers (Beveridge & Harrison, 1984; Karadeniz et al., 2024a). The modeling of kinetic mechanism enables to assess the impact of process conditions on crucial quality indices. Degradation kinetics, naming reaction order, activation energy, and rate constant, is vital to have knowledge about the food quality loss through storage besides thermal process treatments.

The purpose of this study is to present the association between the data obtained from the afore mentioned variables evaluated and to characterize the kinetic mechanism of nonenzymatic browning in PJC by examining the HMF quantity, color, and clarity variations along the storage at 47°, 37°, and 27°C.

2. PRODUCTION OF PJC SAMPLES

Göknur (Göknur Gıda A.Ş., Niğde) fruit and vegetable juice concentrate factory produced pear juice concentrate according to flow chart given in Figure 1. This process is repeated to prepare duplicate samples. The prepared samples were put as duplicate sets in storage rooms adjusted at 47°, 37°, and 27°C. Samples that were randomly chosen were taken on weekly basis from a 47°C storage room on two weeks intervals from a 37°C storage room, and every three weeks from a 27°C storage room. The samples were then analyzed with respect to the parameters of HMF amount, colour, and clarity. The samples' pH, total soluble solids (TSS), and titratable acidity (% citric acid) were also assessed to define the characteristics of PJC.

3. CHEMICAL SUBSTANCES

For this research, Merck is the firm that all chemicals were provided from.

4. DETERMINATION OF PHYSICOCHEMICAL AND CHEMICAL ATTRIBUTES OF PJC

4.1. pH, TSS (°Bx), and Titratable Acidity

TSS was measured with an ATAGO refractometer with the model of RX-5000α (Atago Co., Ltd., Japan). Following the dilution process to reach the single strength °Bx level (11.9 °Bx), an automatic pH meter (Mettler-Toledo GmbH, Switzerland) was used to examine the pH value of the PJC. Using the same pH meter, the potentiometric method was used to analyze the titratable acidities of diluted samples, which were then reported as a percentage of anhydrous citric acid.

4.2. Measurement of Colour and Clarity

A Spectrophotometer was used to measure transmittance in diluted samples at wavelengths of 440 nm and 625 nm, respectively, to identify color and clarity changes in PJs. The colour and clarity results were expressed as follows: colour (%T@440nm @11,9°Bx) and clarity (%T@625nm, @11,9°Bx).

4.3. Spectrophotometric Analysis of Hydroxy -Methyl- Furfural

HMF concentration in stored PJC was obtained with employing a spectrophotometric method (Anonymous 1984). This practice is related to the reaction taking place among p-toluidine, barbituric acid, and HMF. Since the strengthens of red color occurred during this reaction, is correlated with the amount of HMF in the sample, spectrophotometer at 550 nm was used for the quantification of HMF.

5. KINETIC PARAMETERS COMPUTATION

Standard equation for a zero-order reaction model ($A = kt + A_0$) was used to analyze the differentiation in the HMF, colour, and clarity parameters of PJC that means HMF linearly rises, while colour and clarity decreases during storage.

Where; A_0 , amount of criteria at zero-time; A , amount of criteria at moment t ; k , rate constant of the reaction; t , storage period (week).

Arrhenius equation ($k = k_0 \times e^{-Ea/RT}$) was used for determination of temperature dependence of the HMF occurrence, and colour and clarity changes in PJC.

Where; k , speed constant of the reaction (week^{-1}); k_0 , frequency factor (week^{-1}); Ea , activation energy (kJ mol^{-1}); R ; universal gas constant ($8.314 \times 10^{-3} \text{ kJ mol}^{-1} \text{ K}^{-1}$); T , absolute temperature ($^{\circ}\text{K}$). Besides, Q_{10} values were stated by the below formula:

$$Q_{10} = (k_2/k_1)^{(10/T_2-T_1)} \quad (1)$$

where k_2 , speed constant of the reactions at T_2 temperature (mg HMF week^{-1} and colour T% @440 nm, clarity T% @625 nm); k_1 speed constant of the reactions at T_1 temperature (mg HMF week^{-1} and colour T% at 440 nm, clarity T% at 625 nm) (Karadeniz et al., 2024b).

6. STATISTICAL ANALYSIS

This analysis has been performed using the IBM SPSS Statistics program. The correlations between HMF, colour, and clarity parameters have been determined, and a correlation test has been conducted at a significance degree of 0.01.

7. RESULTS & DISCUSSION

No significant changes were determined in physicochemical parameters of pear juice concentrates during storage at different temperatures. The initial °Bx value (70.36) was found as 70.45°Bx, 70.57°Bx and 70.55°Bx when the storage period ended at storage temperatures of 27°C, 37°C and 47°C, respectively. The value of pH was initially 3.9. At the end of storage, it was determined as 3.8, 3.9 and 3.8 at storage temperatures of 27°C, 37°C and 47°C, respectively. The titratable acidity (% anhydrous citric acid) was found as 1.57 at the beginning of the storage while it was found as 1.60% at 27°C, 1.62% at 37°C and 1.59% at 47°C at the end of storage.

Changes in clarity, colour, and the formation of HMF in comparison with storage time had been described as a function of the temperatures (47°, 37°, 27°C) for pear juice concentrate's nonenzymatic browning reaction. The most appropriate model was chosen by defining the determination coefficients (R^2) through regression analysis. Best-fit regression equations were used to calculate the rate constants. Variation in HMF quantity, colour values and clarity through the time of storage is presented in Figure 2, 3 and 4, respectively starting from the lowest storage temperature to the highest storage temperature in this research article.

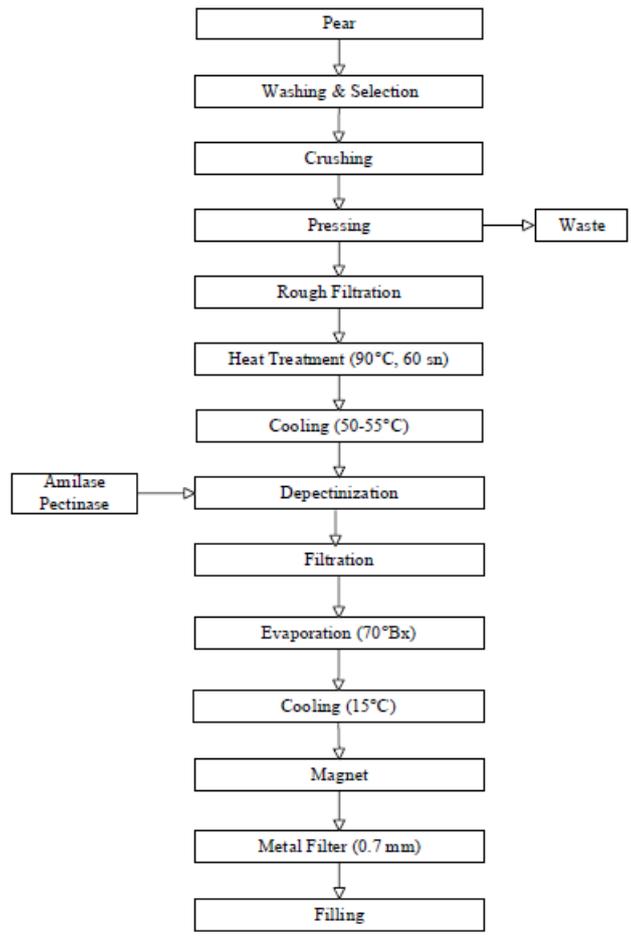


Figure 1. Pear Juice Concentrate Processing Flow Chart

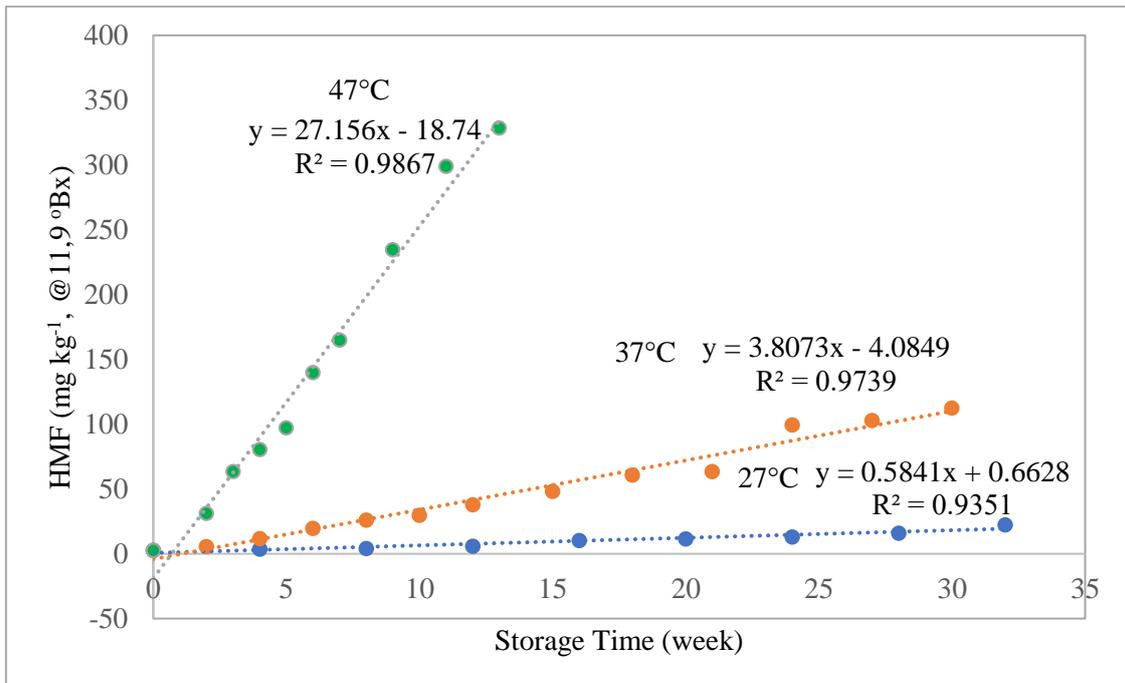


Figure 2. HMF occurrence in PJC during storage

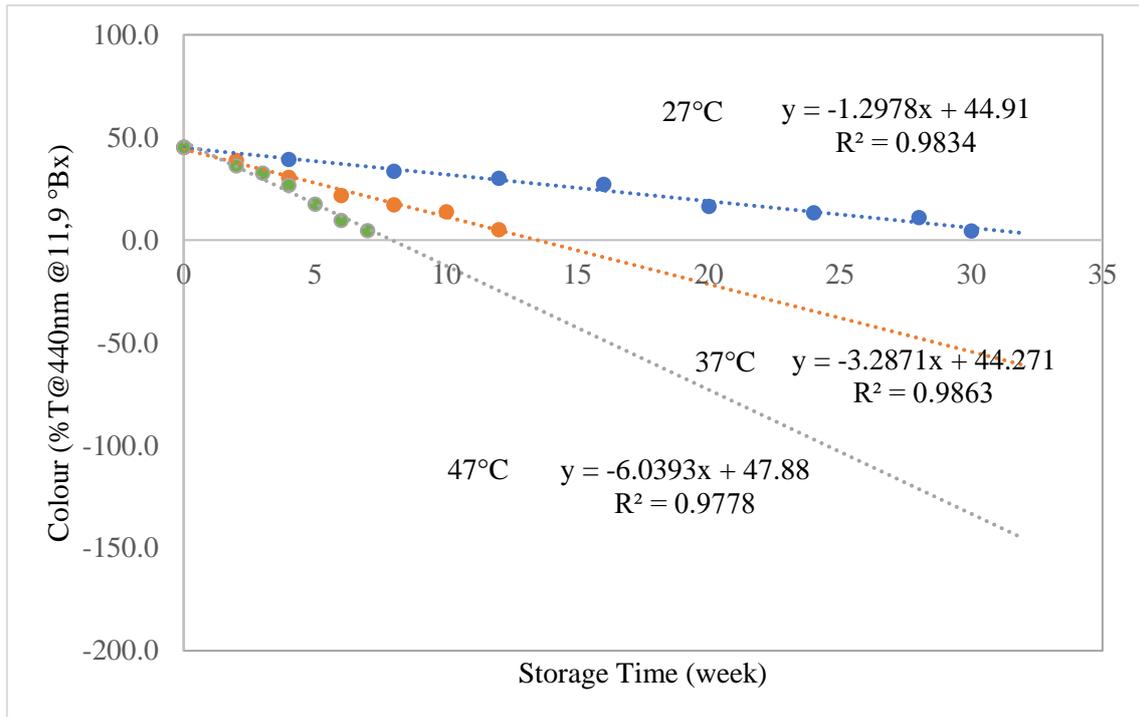


Figure 3. Colour changes in PJC during storage

The HMF content in pear juice concentrate increased over time, starting at 2.7 mg kg^{-1} and reaching 22.5, 112.5, and 328.5 mg kg^{-1} at 27° , 37° , and 47°C , respectively, at the end of storage time. The amount of HMF climbed by time and heat treatment applied in canned grapefruit juices and apple puree as indicated in the outputs of Lee and Nagy (1988) and Zhang et al. (2019) studies, respectively. The maximum increase for HMF content was detected at the rate of 12066%, this was followed by the value of 4066% at the storage temperatures of 47°C and 37°C , respectively. HMF amount in PJC, which showed a linear increase, stated that the reaction kinetics was in parallel with zero order model as presented in other research articles (Burdurlu et al., 2006; Duru et al., 2012). The model of sucrose–glutamic acid showed zero-order kinetics for HMF formation (Zhang et al., 2019). Beveridge and Harrison (1984) studied the effect of temperature and TSS on nonenzymatic browning reactions in PJC and reported that browning reaction occurred following zero order reaction kinetics.

The colour changes were evaluated for browning observed in PJC. It was detected that the colour value (%T@440nm) reduced by time at each storage temperatures examined (Figure 3) indicating that the brown colour intensity was increased. At the beginning of the storage the colour was found as 45.2 and it is less than 5 after 30 weeks of storage at 27°C , 12 weeks of storage at 37°C and 7 weeks of storage at 47°C . This is indicating that higher storage temperature increases the reaction rate. Reaction for colour intensity increase in PJC followed zero-order reaction kinetic which coincides with the findings of other studies (Beveridge & Harrison, 1984; Burdurlu et al., 2006). It has been demonstrated that the reactions of amino acids with reducing sugars have zero-order kinetics, leading to non-enzymatic changes (Ibarz et al., 2008).

Clarity change in PJC was found as 77% at 47°C stored samples. The reduction in clarity was found as 57% and 21% at 37°C and 27°C stored samples, respectively. The clarity changes in PJC also fit well with zero-order reaction kinetics which indicates there is a linear reduction of clarity during storage period.

Temperature dependence of the HMF formation, browning and clarity decrease was determined by Arrhenius equation and Arrhenius plots were given in Figure 5. Activation energy for HMF formation was found higher ($153.14 \text{ kJ mol}^{-1}$) than that of both clarity ($93.81 \text{ kJ mol}^{-1}$) and color ($61.48 \text{ kJ mol}^{-1}$) changes which shows that the occurrence of HMF slow in comparison with color and clarity. Activation energy values for HMF increase in pear puree were reported as $115.14 \text{ kJ mol}^{-1}$ by Tosi et al. (2002) and in modelling solution of apple juice found as $117.2\text{--}165.8 \text{ kJ mol}^{-1}$ by the research of Resnik and Chirife (1979) which coincides with the result of this research.

Temperature quotient (Q_{10}) values were computed (Table 1) for three temperature ranges: 27–37°C, 37–47°C, and 27–47°C. Table 1 illustrates how temperature increases had a greater impact on HMF occurrence in PJC than on clarity and browning reactions. It was discovered that the rise in HMF at 47°C was 7.13 times greater than that at 37°C.

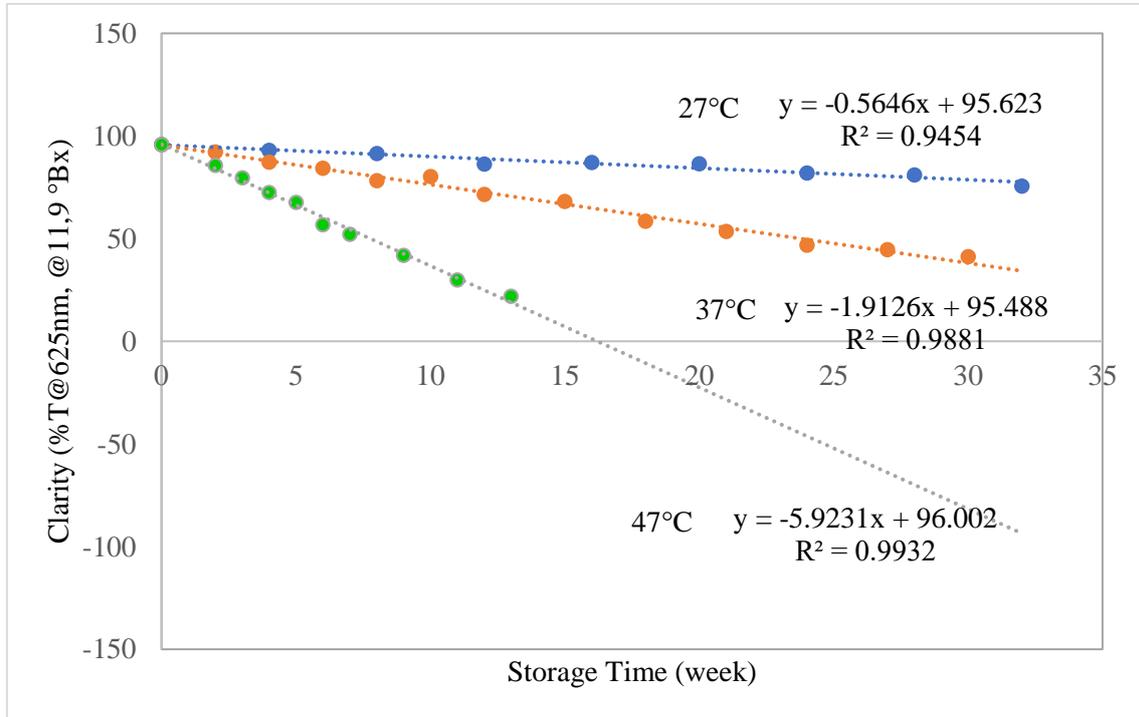


Figure 4. Clarity changes in PJC during storage

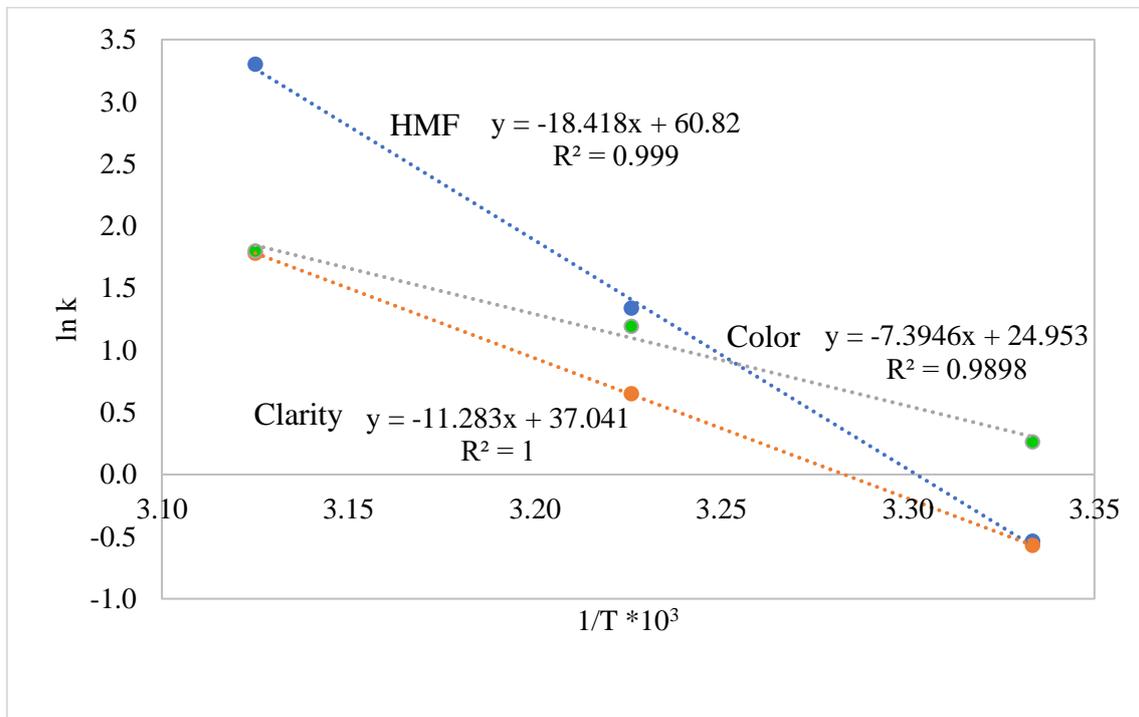


Figure 5. Arrhenius plots of HMF, Color and Clarity parameters in PJC

Table 1. Kinetic criteria for Hydroxy methyl furfural, Clarity and Colour of PJC*

Criteria	Storage Temperature (°C)	Reaction Equation	Activation Energy (kJ mol ⁻¹)	Q ₁₀		
				27-37°C	37-47°C	27-47°C
Hydroxy methyl furfural	27	y = 0,5841x + 0,6628 (0,9351)	153,14 (0,999)	6.52	7.13	6.82
	37	y = 3,8073x - 4,0849 (0,9739)				
	47	y = 27,156x - 18,74 (0,9867)				
Clarity (% T@625nm, @11,9 °Bx)	27	y = -0,5646x + 95,623 (0,9454)	93,81 (1)	3.39	3.10	3.24
	37	y = -1,9126x + 95,488 (0,9881)				
	47	y = -5,9231x + 96,002 (0,9932)				
Colour (% T@440nm @11,9 °Bx)	27	y = -1,2978x + 44,91 (0,9834)	61,48 (0,9898)	2.53	1.84	2.16
	37	y = -3,2871x + 44,271 (0,9863)				
	47	y = -6,0393x + 47,88 (0,9778)				

* The determination coefficients (R²) of the reactions are given in the brackets

Significant (P < 0.01) negative correlations were noticed between HMF formation and colour changes in PJC during storage (-0.959-(-0.992)). Clarity decrease, and HMF occurrence also shows significant (P < 0.01) negative correlations in PJC (-0.953-(-0.992)) (Table 2). Similar to the findings of this investigation, Capuano and Fogliano (2011) detected a positive correlation between the development of browning, color and HMF content.

Table 2. Correlation coefficients among measured criteria in PJC^a

Criteria	HMF		
	27°C	37°C	47°C
Color	-0.959	-0.992	-0.989
Clarity	-0.953	-0.982	-0.992

^a P < 0.01

8. CONCLUSION

In order to evaluate the nonenzymatic browning reactions seen in kept PJC at 27°, 37°, and 47°C, hydroxy methyl furfural occurrence as well as colour and clarity changes were measured. The values of titratable acidity, pH, and TSS (°Bx) did not differ when the storage period ended. The kinetics of zero order reactions were followed by HMF formation, colour changes, and clarity changes. In all storage temperatures examined, the HMF content and colour changes displayed correlations in a negative way ($r=-0.959$ - (-0.992)). Likewise, there is a significant negative correlation ($r=-0.953$ - (-0.992)) between the decrease in clarity and the increase in HMF for all the reviewed temperatures. For HMF formation, colour, and clarity, the corresponding activation energy (kJ mol^{-1}) values were 153.14, 61.48, and 93.81, respectively.

AUTHOR CONTRIBUTIONS

Conceptualization, strategy, documentation and reviewing, F.K.; research & project management, B.I., laboratory work and formal analysis, S.K.; software, sources and funding, O.A., manuscript-review and editing, F.M. Publication of the final version of the article in this Journal has been read and legally accepted by all the authors.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Anonymous. (1984). IFFJP Methods. Analysen-Analyses. Zug, 12, (pp. 1-2).
- Beveridge, T., & Harrison, J. E. (1984). Nonenzymatic Browning in Pear Juice Concentrate at Elevated Temperatures. *Journal of Food Science*, 49(5), 1335-1336. <https://doi.org/10.1111/j.1365-2621.1984.tb14984.x>
- Burdurlu, H. S., Koca, N., & Karadeniz, F. (2006). Degradation of vitamin C in citrus juice concentrates during storage. *Journal of Food Engineering*, 74(2), 211-216. <https://doi.org/10.1016/j.jfoodeng.2005.03.026>
- Capuano, E., & Fogliano, V. (2011). Acrylamide and 5-hydroxymethylfurfural (HMF): A review on metabolism, toxicity, occurrence in food and mitigation strategies. *LWT- Food Science and Technology*, 44(4), 793-810. <https://doi.org/10.1016/j.lwt.2010.11.002>
- Chen, Y., Zhang, M., Mujumdar, A. S., & Liu, Y. (2024). A simulation system to study the types and causes of browning in concentrated orange juice during storage. *Food and Bioprocess Technology*, 146, 160-169. <https://doi.org/10.1016/j.fbp.2024.05.018>
- Cornwell, C. J., & Wrolstad, R. E. (1981). Causes of Browning in Pear Juice Concentrate During Storage. *Journal of Food Science*, 46(2), 515-518. <https://doi.org/10.1111/j.1365-2621.1981.tb04899.x>
- Duru, N., Karadeniz, F., & Erge, H. S. (2012). Changes in Bioactive compounds, antioxidant activity and HMF formation in rosehip nectars during storage. *Food and Bioprocess Technology*, 5, 2899-2907. <https://doi.org/10.1007/s11947-011-0657-9>
- FAOSTAT (2024). Production and Yield quantities of Pears in Türkiye (Accessed: 13/03/2024). <https://www.fao.org/faostat/en/#data/QCL/visualize>
- Ibarz, A., Garza, S., & Pagán, J. (2008). Nonenzymatic browning of selected fruit juices affected by D-galacturonic acid. *International Journal of Food Science and Technology*, 43(5), 908-914. <https://doi.org/10.1111/j.1365-2621.2007.01541.x>
- Karadeniz, F., Atalay, D., Erge, H. S., Kaya, S., Işık, B., & Aslanali, O. (2024a). Kinetics of 5-hydroxymethylfurfural (5-HMF) formation and colour change in date fruit fillings stored at different temperatures. *Journal of Food Composition and Analysis*, 127, 105986. <https://doi.org/10.1016/j.jfca.2024.105986>

- Karadeniz, F., Işık, B., Kaya, S., Aslanali, O., & Midilli, F. (2024b). Kinetics of Nonenzymatic Browning Reactions in Pumpkin Puree During Storage. *Gazi University Journal of Science Part A: Engineering and Innovation*, 11(1), 101-111. <https://doi.org/10.54287/gujasa.1400745>
- Kathuria, D., Hamid, Gautam, S., & Thakur, A. (2023). Maillard reaction in different food products: Effect on product quality, human health and mitigation strategies. *Food Control*, 153, 109911. <https://doi.org/10.1016/j.foodcont.2023.109911>
- Lee, H. S., & Nagy, S. (1988). Quality changes and nonenzymic browning intermediates in grapefruit juice during storage. *Journal of Food Science*, 53(1), 168-172. <https://doi.org/10.1111/j.1365-2621.1988.tb10201.x>
- Nikicevic, N. (2005). Effect of Some Production Factors on Chemical Composition and Sensory Qualities of Williams Pear Brandy. *Journal of Agricultural Sciences*, 50(2), 193-206. <http://doi.org/10.2298/JAS0502193N>
- Öztürk, A., Demirsoy, L., Demirsoy, H., Asan, A., & Gül, O. (2015). Phenolic Compounds and Chemical Characteristics of Pears (*Pyrus Communis* L.). *International Journal of Food Properties*, 18(3), 536-546. <https://doi.org/10.1080/10942912.2013.835821>
- Resnik, S., & Chirife, J. (1979). Effect of moisture content and temperature on some aspects of nonenzymatic browning in dehydrated apple. *Journal of Food Science*, 44(2), 601-605. <https://doi.org/10.1111/j.1365-2621.1979.tb03845.x>
- Saeeduddin, M., Abid, M., Jabbar, S., Wu, T., Hashim, M. M., Awad, F. N., Hu, B., Lei, S., & Zeng, X. (2015). Quality assessment of pear juice under ultrasound and commercial pasteurization processing conditions. *LWT - Food Science and Technology*, 64(1), 452-458. <https://doi.org/10.1016/j.lwt.2015.05.005>
- Tosi, E., Ciappini, M., Ré, E., & Lucero, H. (2002). Honey thermal treatment effects on hydroxymethylfurfural content. *Food Chemistry*, 77(1), 71-74. [https://doi.org/10.1016/S0308-8146\(01\)00325-9](https://doi.org/10.1016/S0308-8146(01)00325-9)
- Zhang, L-L., Kong, Y., Yang, X., Zhang, Y-Y., Sun, B-G., Chen, H-T., & Sun, Y. (2019). Kinetics of 5-hydroxymethylfurfural formation in the sugar–amino acid model of Maillard reaction. *Journal of the Science of Food and Agriculture*, 99(5), 2340-2347. <https://doi.org/10.1002/jsfa.9432>
- Zhu, Y., Zhang, M., Mujumdar, A. S., & Liu, Y. (2023). Application advantages of new non-thermal technology in juice browning control: A comprehensive review. *Food Reviews International*, 39(7), 4102-4123. <https://doi.org/10.1080/87559129.2021.2021419>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1510527>

GIS-based AHP and MCDA Modeling for Cropland Suitability Analysis: A Bibliometric Analysis

Dilnu Chanuwan WIJESINGHE^{1*} ¹ School of Geographical Sciences, Southwest University, Beibei District 400715, Chongqing Province, China

Keywords	Abstract
Analytical Hierarchy Process Bibliometric Analysis Cropland Geographic Information System Suitability	The 'Land Suitability Analysis' is a useful management method for ensuring that agricultural lands are utilized sustainably and planned based on their potential. Geographic Information Systems (GIS) and the Analytic Hierarchy Process (AHP) for cropland suitability analysis have seen substantial contributions from researchers worldwide. This combination assesses and maps the suitability of land for different crops by utilizing the multi-criteria decision analysis (MCDA) strengths of AHP and the spatial analytic capabilities of GIS. This Bibliometric analysis involves examining publications to identify patterns and trends, such as the most prolific authors & Countries, influential journals, and highly cited papers. It helps in understanding the development and current state of a research field. Using Biblioshiny software, the researchers obtained 183 publications of 687 authors and 319 different institutions using the bibliographic information from the Scopus database. The bibliometric analysis uses the following subcategories: Country, Authors, Publication Sources, Annual Scientific Production, and keywords. By examining the outcomes of bibliometric analysis, methodology, and applications, it was discovered that AHP and MCDA are the most often utilized techniques in this respect. Also, the findings indicated a rising number of publications and a growing interest in the subject, especially in recent years. Over the previous 23 years, the overall trend of publications in this field grew gradually at an annual growth rate of 21.81%. Asian nations, especially China, India, and Iran, have had the biggest influence on the nation's scientific output in the discipline. During this period, India and Iran had the most research papers published. In addition, "GIS," "Land Suitability," and "AHP" are the top three most often used terms. Future trends in this subject are predicted by the current keywords: "GIS," "Land Suitability," "AHP," and "Remote Sensing." Moreover, this exhaustive investigation provides a basis for comprehending the present status and future direction of GIS-based cropland suitability research. These discoveries offer valuable insights for future modeling and research endeavors on the subject and aid in identifying research gaps in the existing literature.

Cite

Wijesinghe, D C. (2024). GIS-based AHP and MCDA Modeling for Cropland Suitability Analysis: A Bibliometric Analysis. *GU J Sci, Part A, 11(3)*, 598-621. doi:[10.54287/guj.1510527](https://doi.org/10.54287/guj.1510527)

Author ID (ORCID Number)	Article Process
0000-0002-3268-7869 Dilnu Chanuwan WIJESINGHE	Submission Date 04.07.2024 Revision Date 08.07.2024 Accepted Date 28.08.2024 Published Date 30.09.2024

1. INTRODUCTION

Of all the natural resources available, agricultural resources are considered to be the most important renewable and dynamic resource (Sathiyamurthi et al., 2022). In this manner, the term "suitability" refers to a function that is dependent on land elements and crop requirements; it is similar to a metric that describes the characteristics of the land unit to compare the needs of a particular land use (FAO, 1976). Analyzing a land area's suitability for a certain type of land use is a methodical process known as Land Suitability Analysis (LSA). Assessment of land suitability by FAO guidelines has been implemented in many parts of the world, especially in developing nations (FAO, 1976). A piece of land's appropriateness for agricultural production is determined by how well it can withstand crop production over time (Singha & Swain, 2016). To determine if a piece of land is suitable for a particular agricultural crop, many factors must be taken into consideration.

*Corresponding Author, e-mail: dilnuchanuwan@gmail.com

These include biophysical components involving relief (elevation), terrain slope, field drainage, soil qualities, nutrient content, atmospheric conditions, vegetative aspects, etc. When making significant judgements, this also takes into account social, ecological, and cultural factors (Abushnaf et al., 2013). To combine geographically referenced data, GIS is typically recognized as a decision support system (Özkan et al., 2019). Expert views are used in the multi-criteria decision-making (MCDM) process, which involves assessing a set of alternatives about several criteria (Liao et al., 2023). Many complex decision-making issues have been resolved with its help (Chiao, 2021; Torkayesh et al., 2021; Baydas et al., 2022). Some of the MCDM techniques used in site selection difficulties are the Elimination and Choice Translating Reality (ELECTRE), Weighted Linear Combination (WLC), AHP, and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). (Wigati et al., 2019). Geographical problems involving geographic data, such as site selection problems, are sometimes known as spatial or geographical decision problems. The advancement of spatial query and analysis systems in GIS has been greatly aided by the rapid advances in computer technologies during the early 2000s. Several studies have revealed that the use of ArcGIS software for geographic decision support systems has increased in this era, which is recognized for its long-lasting effects (Miller & Goodchild, 2014). A few advantages of MCDA-GIS integration include its ability to assist stakeholders and policymakers in the crucial process of defining evaluation criteria (Palmisano et al., 2016). Alfred Weber first put out the idea of site selection theory in 1909, concentrating on the need to locate the depot closest to every customer. The selection of location is a constantly fascinating novel issue. Finding the best location with the required circumstances while taking into account a variety of factors is known as site selection. A thorough assessment of the site selection procedure is offered by the application of GIS-based MCDA techniques, which combine various datasets and criteria using geospatial technology to provide a suitability map that shows regions appropriate for cropland.

Over the past few decades, GIS has become widely used in the field of agricultural science. Additionally, research on agricultural productivity has been done with fuzzy logic, fuzzy TOPSIS, and AHP-MCDA (Houshyar et al., 2014; Maleki et al., 2017; Kahsay et al., 2018; Amin et al., 2020). Continuous factors can be modeled for a suitability assessment within a GIS study thanks to the fuzzy set theory (Purnamasari et al., 2019). Land adaptability must be evaluated to facilitate productive planning and long-term, sustainable land usage. Tashayo et al. (2020) used AHP-Fuzzy and GIS to examine whether land in the saline, calcareous, and sodic soils of the Marvdasht Plain, which is in southern Iran, was suitable for growing wheat. Habibie et al. (2021) assessed the land suitability study with the use of several criterion considerations and GIS in order to ensure the sustainability of maize production in Tuban Regency, Indonesia. Further, according to fifteen socio-environmental, physical, and climatic characteristics, Musakwa (2018) investigated the appropriateness of cropland in South Africa. In Sri Lanka, Jayasinghe and Withanage (2021) have utilized GIS-MCDA techniques to determine potato land suitability. Their analysis included a diverse array of criteria such as soil pH, salinity, elevation, slope, aspect, market proximity, and land use. Also, the assessment of land is crucial in this regard since it offers details on the possibilities and limitations of the land for a specific type of land use about crop productivity (Habibie et al., 2021). Many models have been developed to estimate crop output through remote sensing applications, thanks to the quick growth of spectral reflectance and remote sensing technologies (Sakamoto et al., 2013; Lobell et al., 2015; Zabihi et al., 2015; Azzari et al., 2017). A significant amount of literature has been produced over the years by the many researchers who have created models and methods for creating maps of land suitability around the globe. Numerous fields have had a literature review on the application of GIS-MCDA integration for site selection problems that have been conducted, including cropland suitability (Purnamasari et al., 2019; Kılıç et al., 2022; Sathiyamurthi et al., 2022), landslide susceptibility (Huang et al., 2022; Özkan et al., 2019; Chinthaka et al., 2023), and solar photovoltaic power plant site selection (Al Garni & Awasthi, 2018; Rediske et al., 2018), for solid waste site selection (Wijesinghe & Fernando, 2023). For scholars, the literature—which consists of multiple papers has offered a useful summary of land suitability.

The appropriate results are important for academics to identify the main topics, consider potential avenues for future investigation, and seek collaboration with other organizations or nations in a certain field of study (Huang et al., 2022). It is necessary to do more thorough literature studies that examine the contributions of research experts, the field's evolution, topics, and upcoming academic possibilities related to cropland suitability. A popular strategy for examining published literature is bibliometric analysis, which is a contemporary approach in computer engineering, database management, and statistics (Qin et al., 2022).

Bibliometrics is a statistical analysis used in various scientific fields to assess books, journal articles, and other written communication formats on a specific topic. This paper employs bibliometrics analysis to demonstrate the current state of research and the direction of development for the usage of AHP-MCDA models in cropland suitability applications. Bibliometrics allows researchers to analyze the growth and development of research over time. By examining publication trends, one can identify peak periods of interest and the evolution of the field. By analyzing keywords and their co-occurrence, bibliometrics helps in identifying major themes, sub-fields, and emerging topics within the domain of GIS-based AHP and MCDA modeling for cropland suitability. The bibliometrics analysis to quantify and understand the landscape of research involving AHP-MCDA models in cropland suitability. It looks at the volume and trends of scientific production, identifies key contributors, most prolific countries, influential journals, and highly cited papers, and provides insights into where future research might be headed. Identifies which countries or regions are conducting the most research on this topic, indicating where cropland suitability is a significant concern. This helps researchers and stakeholders understand the current state of the field and make informed decisions about future research directions. Not only that, A thorough summary of the body of research is given by Bibliometrics, summarizing key findings, methodologies, and theoretical advancements in the field. The method facilitates the creation of visual representations (such as co-citation networks, thematic maps, and collaboration graphs), making complex data more accessible and understandable. As far as the researcher is aware, no research has ever undertaken a bibliometric review on cropland suitability anywhere in the world. Insights from bibliometric analysis can inform policymakers and planners about effective strategies for land use and agricultural development. Therefore, to fill the research gap in this area, this work investigates the gaps in the literature and conducts a thorough bibliometric study on GIS-based cropland suitability for the years 2010 to 2023. Additionally, this manuscript attempts to address the following questions:

RQ1: How are the Annual Scientific Productions related to the GIS-based cropland Suitability?

RQ2: Who are the most impactful (Cited) authors on the subject of GIS-based cropland Suitability?

RQ3: Which are the most influential journals (Sources) in this area of research?

RQ4: Which countries have contributed to international collaboration and what is the total number of publications related to this area of research?

RQ5: What are the relevant author keywords and trending topics for years for future research direction related to GIS-based cropland Suitability?

The problems above are essential to understanding the general context of GIS-based AHP-MCDA techniques used in the agriculture sectors. This work aims to elucidate the current state of research and development trends regarding the use of AHP-MCDA models in the agricultural environment by applying content analysis and bibliometrics.

2. MATERIALS AND METHODS

This work employed a bibliometric analysis to assess the quality of prior research and present an extensive overview of the current status of scientific output, offering a plethora of data on the respective topic. The database was visually inspected and filtered before outliers were discovered. After that, it was examined using the R environment's "Bibliometrics" package and the Biblioshiny app (R CoreTeam, 2016; Aria & Cuccurullo, 2017). Bibliometric analysis can highlight under-researched areas and gaps in the existing literature, guiding future research directions and strategies. It enables comparative analysis of different research entities, helping researchers and institutions benchmark their performance. By employing the bibliometrics method, researchers can systematically and comprehensively analyze the vast literature available in this field. This provides a structured approach to understanding the field's dynamics, identifying influential research, and guiding future studies. A standard bibliometric analysis consists of the following five steps: research design, data collection, analysis, visualization, and interpretation (Zupic & Čater, 2014). With the aid of bibliometric technology, information may be retrieved from the repository by classifying and analyzing vast volumes of historical data that were the result of a study done during a specific period (Ejaz et al., 2022). Doing a comprehensive

bibliometric analysis of the literature in GIS-based cropland suitability helps to achieve the goals of this study. Figure 1 illustrates this research methodology for gathering and analyzing data.

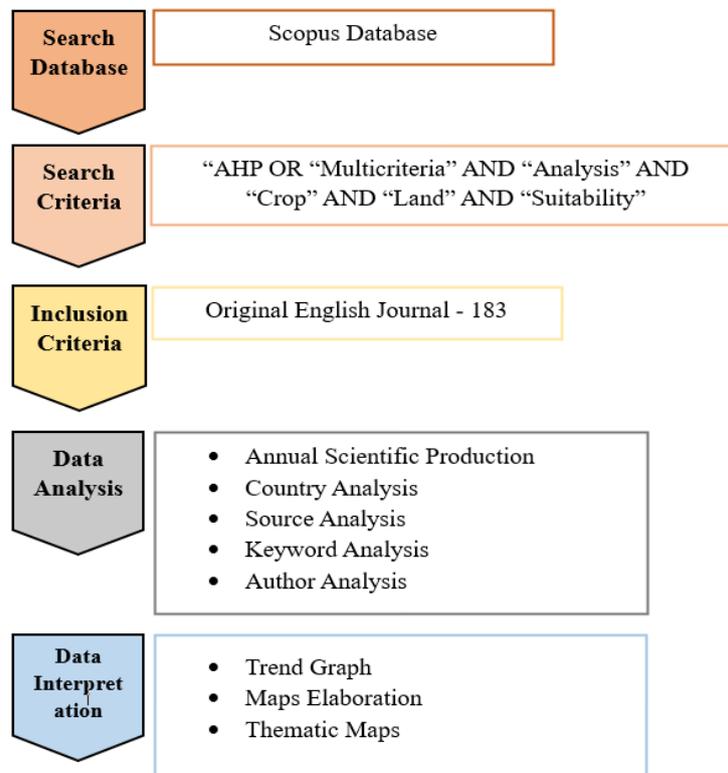


Figure 1. Methodology Process

2.1. Materials

2.1.1 Data Collection

The current study searches for pertinent documents using the Scopus database. Launched in 2004, the Scopus database is one of Elsevier's well-known databases. Compared to other academic databases, Scopus has a greater number of articles (Veloutsou & Mafe, 2020). Also, for a more in-depth analysis, this database's comprehension and ability to support numerous inclusion and exclusion criteria are essential (Rautela et al., 2024). Choosing and utilizing precise keywords that can yield precise database findings is one of the most crucial phases of bibliometric research. The search parameters included "Article title, Abstracts, and Keywords." Identified keywords were combined using the Boolean operators "AND/OR" to refine the search results. In the Scopus database, the search terms included "AHP OR "Multicriteria" AND "Analysis" AND "Crop" AND "Land" AND "Suitability." In the next step, the document type was 'Article', the language was selected as 'English Language' and limited to published documents from 2010 to 2023. Ensured to include relevant fields such as author(s), title, source (journal/conference), abstract, keywords, DOI, publication year, citations, and affiliations. To be used in the bibliometric study that followed, the final records were exported from the Scopus database in Comma-Separated Value (CSV) format. As a data cleaning part, then removed duplicates, irrelevant articles, and non-English papers if necessary. After, Extracted bibliographic information such as authors, publication year, journal, keywords, abstracts, and citations. Depending on the scope of the study, the researcher might choose to focus only on English-language papers to maintain consistency and comprehensibility. also, manually inspected titles and abstracts to confirm relevance, especially for borderline cases where keyword matching may not be conclusive. By thoroughly cleaning the data, researcher determine that the subsequent bibliometric analysis accurately reflects the research landscape of GIS-based AHP and MCDA modeling for cropland suitability analysis. Also, the data format is required to meet the input requirements of the Bibliometrix package in R. Ultimately, 183 publications were chosen for further investigation in this study.

2.1.2 Data Analysis and Interpretation

The researchers examined and visualized the present and future orientations of Cropland Suitability using GIS, using the Bibliometric program from R 4.4.0 Software and the Biblioshiny web interface. With bibliographic data taken from the Scopus database, Biblioshiny analyses a wide range of categories in analytics and graphs, from individual contributions to the social network (Moral-Muñoz et al., 2020). Using the R programming language and the Shiny package, Massimo Aria developed the Biblioshiny software (Aria & Cuccurullo, 2017). Also, a readily interpretable bibliometric map is displayed by the bibliometric visualization tool VOSViewer (Rautela et al., 2024). This study examined how the volume of research has evolved. Identify periods of increased interest and potential reasons for these trends. Identify dominant themes and emerging topics in the field of cropland suitability analysis using GIS, AHP, and MCDA. Also, highlight key papers and authors that have significantly contributed to the field. Understand the impact of foundational studies. Further, Quantify publications over time, journal distributions, country-wise contributions, and author productivity. In this context, the researcher Understands how AHP and MCDA are applied to weight criteria and makes decisions in land suitability analysis. Additionally, it might provide a visual representation of how the field of study has grown and changed. Moreover, create maps, charts, and graphs to visualize bibliometric patterns and GIS-based suitability results. These include trend graphs, the production of the top authors over the graph, corresponding author country figures, country maps, appropriate sources, and thematic maps. Finally, provide insights for policymakers on land-use planning and sustainable agricultural practices based on suitability analysis results.

3. RESULTS

3.1 Scientific Productivity in the Field of GIS-based Cropland Suitability

To understand a study field's development state and trend, it can be useful to analyze the number of publications over time. Using the Scopus Database, 183 research publications on GIS-based cropland suitability published between 2010 and 2023 were examined for this study. There has been a notable advancement in this subject recently, and numerous studies are being presented in this regard. In many countries, one can witness an attempt to concentrate on various investigations by utilizing the technical knowledge associated with this. As a result, from 2010 to 2023, the publication of research articles of the chosen scope directly increased, as seen in Figure 2. With an R-square of 83%, the equation shown in Figure 2 can forecast the documents published annually. Research paper publication has increased between 2018 and 2023, and this trend is distinct in that it will increase by 21.31% by that year. There were less than ten published research publications between 2010 and 2017. Researcher publications from 2018 to 2023 comprised a substantial amount of 145 publications, and 38 papers were published overall in those 07 years. It's evident from looking at the research articles' citations that there hasn't been any improvement.

Figure 3 displays the average number of citations per article and year for the article. Concerning the evolution of the number of citations annually, it should be noted that the decrease in citations recorded for the 2019–2023 period can be explained by the fact that, as would be expected, the temporal frame within which publications published in the years close to the conclusion of the analysis year can be cited is smaller. After receiving 8.38 Mean Total Citations annually for three years starting in 2010, there were a few minor increases noted. The paper "GIS-based land suitability assessment for tobacco production using AHP and fuzzy set in Shandong province of China," written by Zhang, Su, Wu, and Liang and published in 2015, has the most citations which received 197 (Zhang et al., 2015). Furthermore, the highest point in data for 2018 and 2019 can be seen in Figure 3, which shows the evolution of the average annual article citations. The paper "Development of a Model Using Matter Element, AHP, and GIS Techniques to Assess the Suitability of Land for Agriculture," written by Seyedmohammadi et al. and published in 2019, had 114 citations during this period (Seyedmohammadi et al., 2019). 93 papers were published between 2021 and 2023, with an average annual total citation of 7.09. Of those, 30 papers were published in 2018 and 2019, accounting for a total of 12.63 citations.

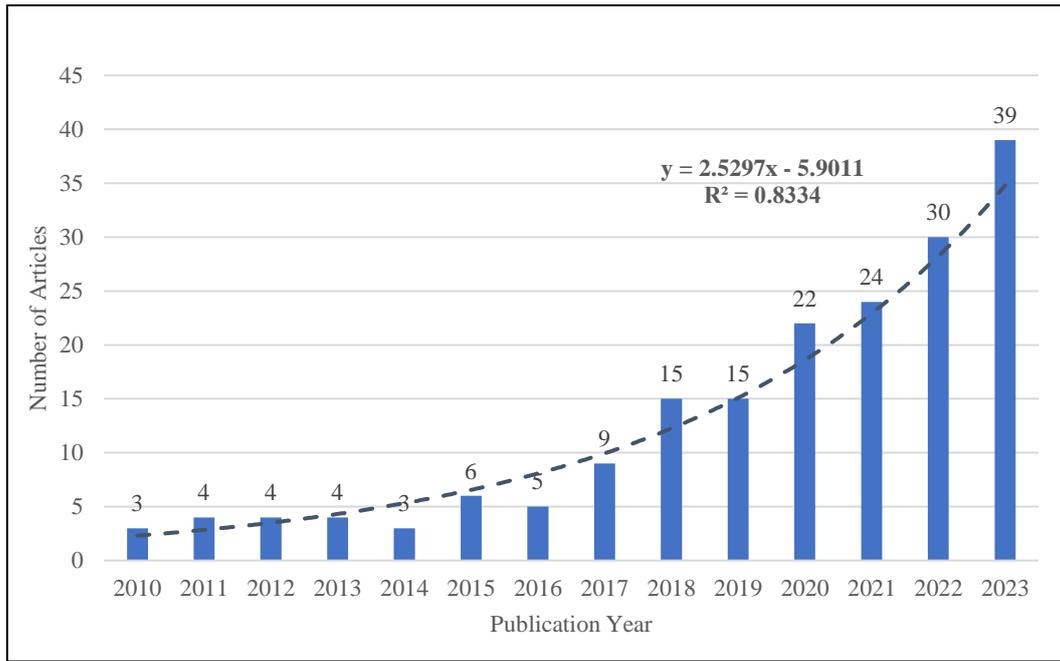


Figure 2. Annual Scientific Production

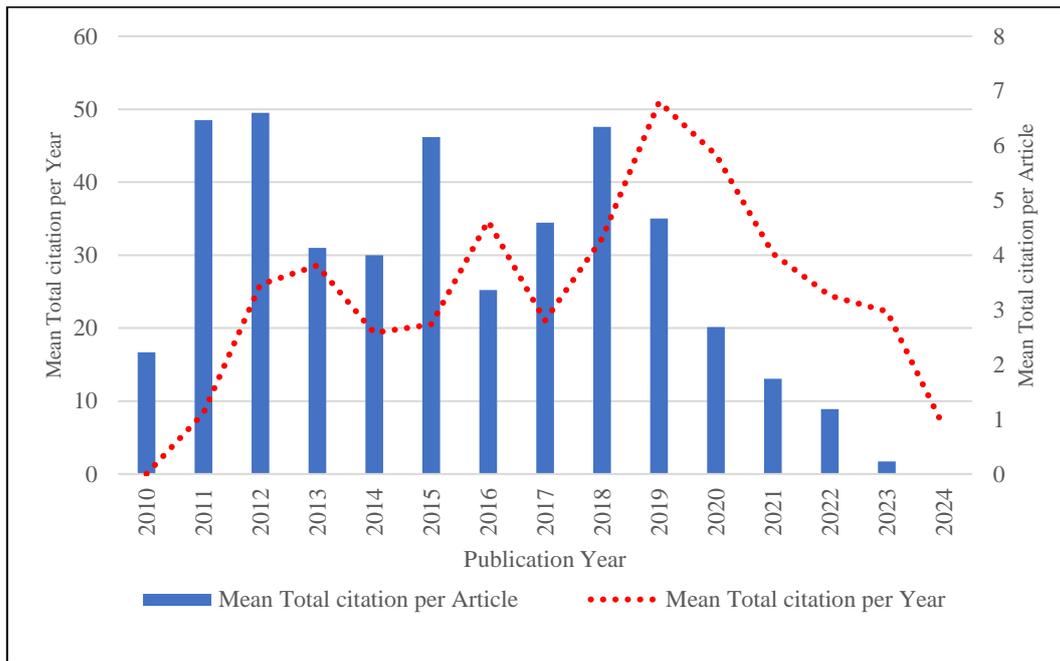


Figure 3. Distribution of Mean Citations of Documents Yearly

According to Figure 4, a three-field plot (Sankey diagram) of the Author's nation (AU_CO), Authors (AU), and Keyword (DE) of publication of the cited references was constructed to indicate the percentage of study topics for each nation and the recentness of the papers they cited. The names of the contributing researchers are displayed in the middle column, nations are represented in the left column, and the most popular keywords used by these writers are displayed in the right column. A precise and well-illustrated depiction of well-known academics, their nations, and the knowledge management-related fields of interest are highlighted with keywords. It demonstrates that the most often used terms, which are mostly associated with the most referenced references, are "Land Suitability," "GIS," and "AHP." The Analytical Hierarchy method, GIS, land suitability, and remote sensing are the main areas of interest for researchers utilizing GIS to investigate land suitability in Peru. The majority of studies on consent have been published in Croatia, Peru, Japan, China, Iran, and India publish a limited amount of research on GIS and land suitability.

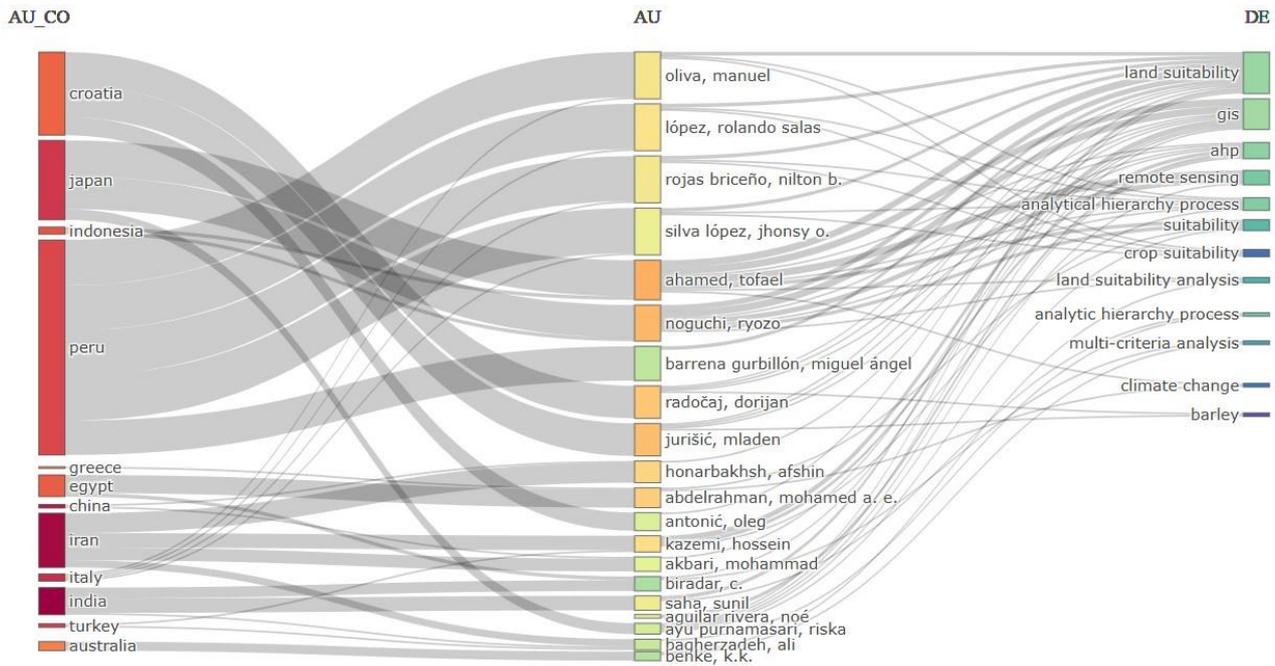


Figure 4. Three-field Plot (Sankey diagram) of Author's Country, Authors, and Keyword

3.2 Author Analysis

In the study, there were 687 authors in entirety. Table 1 identifies the highest-cited authors, with more than 100 citations, as Ahamed, T., Noguchi, R., Honarbaksh, A., and Kazemi, H. Since Ahamed and Noguchi have 'h' index values of 5 and 'g' index values of 7 and 6, respectively, they are the most prominent authors. Furthermore, Jurisic, M. (63 citations) and Radočaj, D. (63 citations) have both made significant contributions to the literature in terms of the total number of citations. Due to the excellent quality of the numerous publications, they have written and published on the subject, Ahamed T. and Noguchi R. are well-regarded in the field of GIS-based cropland suitability analysis. The publication began in 2018, as shown in Figure 5, where the size of the circle indicates the number of documents and the colour shade indicates the number of citations. The years 2019 to 2022 noticed the highest number of published papers as well as the highest frequency of average citations per item. 42 citations were found in an article by Ahmed that was presented in the Computers and Electronics in Agriculture journal and was titled "Land suitability assessments for yield prediction of cassava using geospatial fuzzy expert systems and remote sensing" (Purnamasari et al., 2019). Using LSA to evaluate suitable sites, the research goal was to create a yield prediction model. In addition, ecological categories were used in a multi-criteria decision-making process together with a fuzzy expert system to identify priority indicators. Noguchi was the second most published author with a 'h' index of 5 and a 'g' index of 6, respectively. Titled "Land Suitability Analysis for Maize Production in Indonesia Using Satellite Remote Sensing and GIS-based Multicriteria Decision Support System," the article from 2021 that has received the most citations belongs to him. The author Honarbaksh, A., produced the third-highest citation count with 'a' h index and a 'g' index of 3 respectively. He wrote articles about AHP–GIS for cropland Suitability Modelling as recently as 2020, and his initial publication dates back to 2019.

In the present investigation, an attempt has been made to calculate the relative concentration of GIS-based cropland suitability contributors by using Lotka's law. According to Figure 6, which depicts those 622 authors who published one article and one author who published more than six, individual productivity was determined to accord with Lotka's Law (1926) about the distribution frequency. It was determined that the correlation coefficient was $r = 0.4147$. According to Lotka's law, the productivity of authors in terms of publications is negatively correlated with their number of authors.

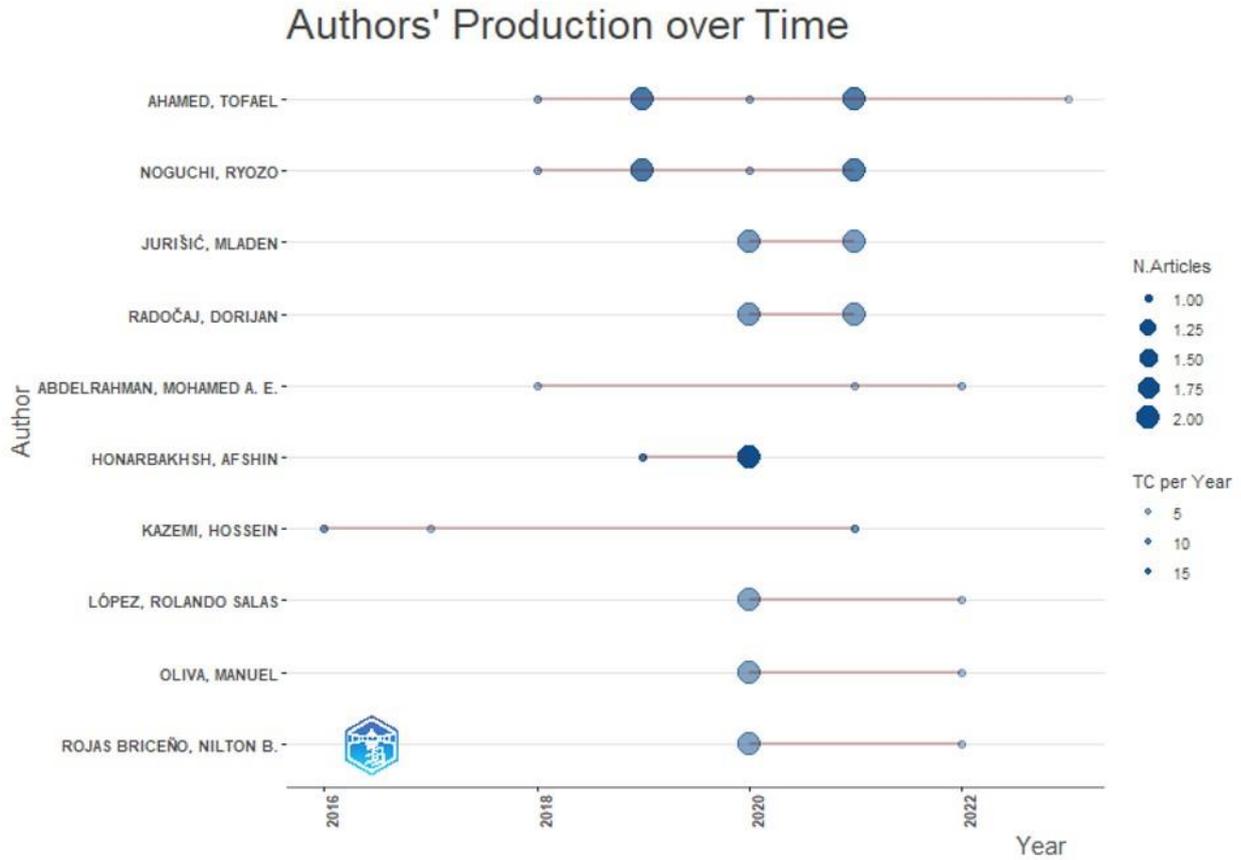


Figure 5. The Production of the top Authors Over Time

Table 1. Most Impactful Authors

Element	h_index	g_index	m_index	TC	NP	PY_start
Ahamed Tofael	5	7	0.71	166	7	2018
Noguchi Ryozo	5	6	0.71	165	6	2018
Jurišić Mladen	4	4	0.8	63	4	2020
Radočaj Dorijan	4	4	0.8	63	4	2020
Abdelrahman Mohamed A. E.	3	3	0.42	37	3	2018
Honarbakhsh Afshin	3	3	0.5	156	3	2019
Kazemi Hossein	3	3	0.33	114	3	2016
López Rolando Salas	3	3	0.6	34	3	2020
Oliva Manuel	3	3	0.6	34	3	2020
Rojas Briceño Nilton B.	3	3	0.6	34	3	2020

Source: Based on Biblioshiny Software

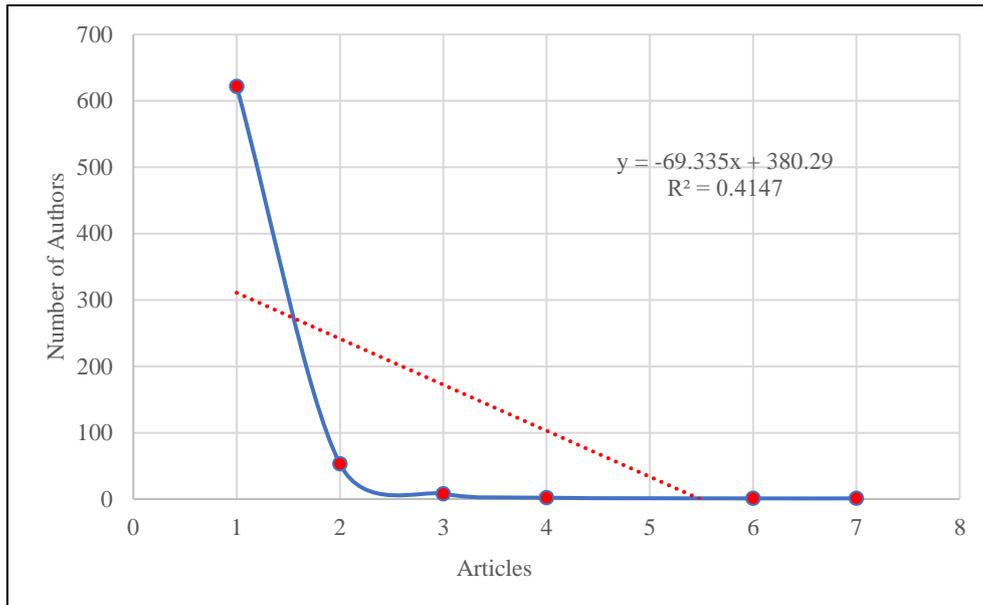


Figure 6. Lotka's Law in GIS-based Cropland Suitability Research

3.3 Source Analysis

Figure 7 displays the results of the top 10 most pertinent sites that have published research publications on GIS-based cropland Suitability. The outcome was produced based on Scopus data that was retrieved between 2010 and 2023. Sustainability is identified as the most important and pertinent source. The greatest number of publications and the most significant source of references for research on themes related to GIS-based cropland suitability analysis were published by this journal. For example, the studies included under these categories have been published in the Sustainability. "Crop-Suitability Analysis Using the Analytic Hierarchy Process and Geospatial Techniques for Cereal Production in North India" and "Land Suitability Planning for Sustainable Mango Production in Vulnerable Region Using Geospatial Multi-Criteria Decision Model". The top 10 most relevant journals, Sustainability Switzerland, Computer and Electronics in Agriculture, Ecological Indicators, and Modelling Earth System and Environment, are shown in Figure 8 based on citation, impact, and number of articles based on the H-index. Also, in the upcoming years, it is anticipated that the issue will continue to receive a significant number of citations in pertinent publications.

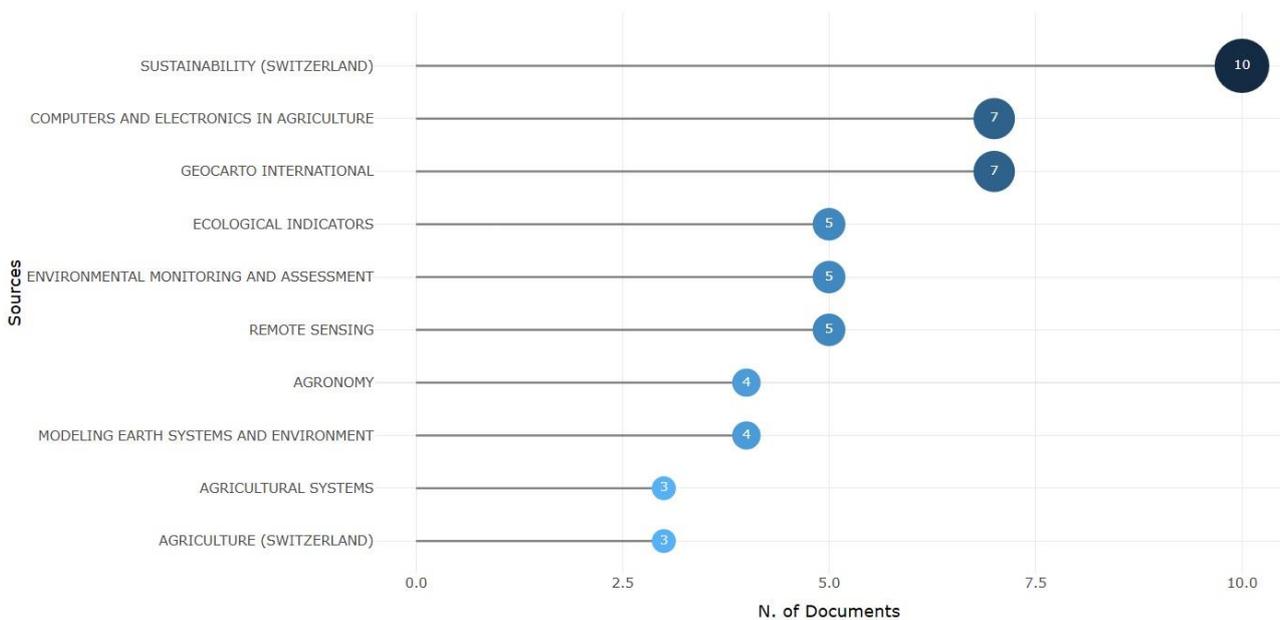


Figure 7. Distribution of Publications based on Different Sources

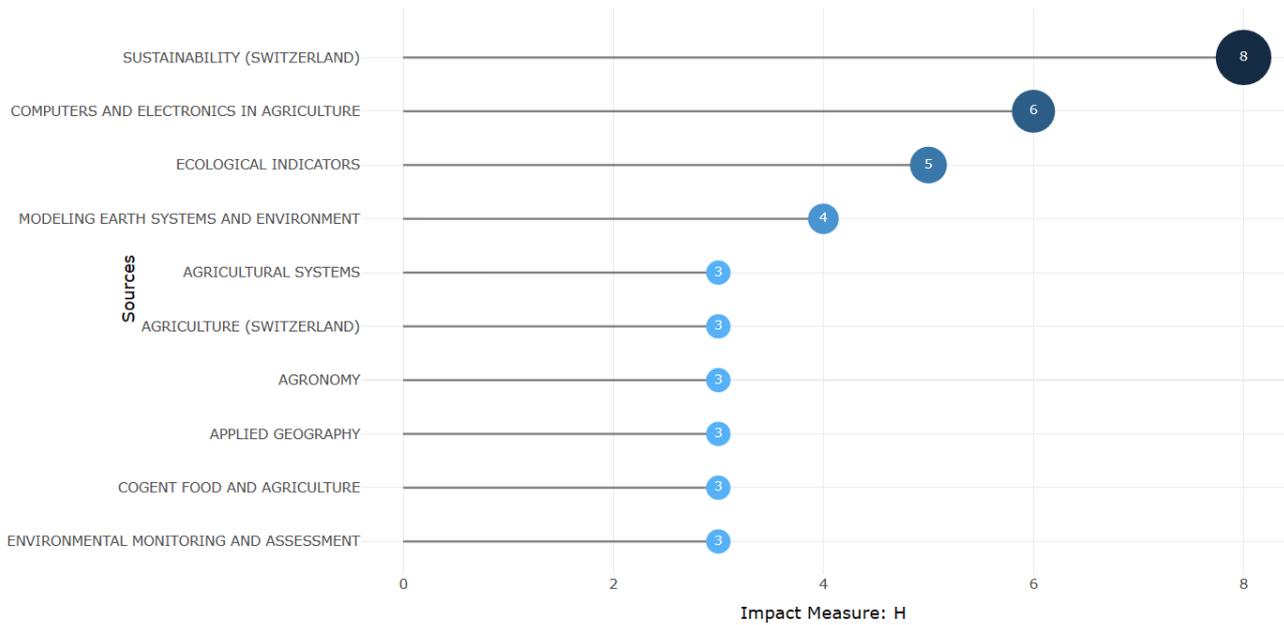


Figure 8. Sources Local Impact by h-index

According to Table 2, the term "The Most Local Cited Sources (from Reference Lists)" refers to a bibliometric analysis method that finds the sources that are most commonly mentioned in a certain field of study or research area. The process normally entails looking through the references of scientific publications to identify the sources that authors in a certain field, in this case, cropland suitability cite most frequently. According to the Zhang et al. (2015) study "GIS-based land suitability assessment for tobacco production using AHP and fuzzy set in Shandong province of China," which is presented in the table below, this approach offers an overview of the most important and pertinent document in the GIS-based cropland suitability area of research which, according to the Journal of Computers and Electronics in Agriculture, represented 40 local citations and 197 global citations, respectively. On the other hand, the other two most highly cited documents, both locally and globally, are Kazemi et al. (2016) from Ecological Indicators, which has 19, 64 local citations and global citations respectively, and Ostovari et al. (2019), also from Ecological Indicators, who has 68 worldwide citations. In particular, for the first three study publications, LC/GC ratios of 20.30%, 29.69%, and 25% can be found, respectively. In this field of study, this can be used to identify the principal theories, methodologies, approaches, and conclusions. Additionally, it can help observe how a field evolves and spot new patterns and potential study areas.

A pattern that calculates the exponentially declining rewards of looking up references in science publications is known as the Bradford Law. Samuel C. Bradford originally described this pattern in 1934 (Bradford, 1934; Black, 2004; Yatsko, 2012). The results of Bradford's law source clustering are shown in Figure 9 & Table 3, which attempts to visually depict the distribution of documents in a field among multiple journals so that we may determine which journals are most responsible for the area's relevant literature. Only 13 journals were determined to be in the core zone out of the 116 total sources using Bradford's formula. Additionally, the middle and minor zones contained, respectively, 43 and 60 of the journals. This indicates that a significant number of publications on the subject of GIS-based cropland suitability and AHP, MCDA are probably published in these 13 journals, whereas the content in the other sources may be less relevant or less frequently published on that particular issue. The most significant and fundamental sources are several journals, including "Sustainability (Switzerland)", "Computers and Electronics in Agriculture", "Geocarto International", and "Ecological Indicators". As a result of this, among the primary sources, Sustainability is the most frequently cited and highly relevant publication. This indicates that Bradford's law can be applied to the present research that has been filed and is relevant in this instance. This demonstrates the high calibre of the database and the articles that it has found and added about GIS-based cropland suitability research conducted country-wise. Nevertheless, citation counts alone do not indicate a journal's value and impact; other metrics, such as the quantity of publications, are also positively connected with it.

Table 2. Most Local Cited Document

Document	Year	Local Citations	Global Citations	LC/GC Ratio (%)	Normalized Local Citations	Normalized Global Citations
Zhang et al., 2015, Comput Electron Agric	2015	40	197	20.30	5.58	4.27
Kazemi et al., 2016, Ecol Indic	2016	19	64	29.69	4.52	2.54
Ostovari et al., 2019, Ecol Indic	2019	17	68	25.00	7.73	1.94
Seyedmohammadi et al., 2019, Geoderma	2019	10	114	8.77	4.55	3.26
Kihoro et al., 2013, Springerplus	2013	10	92	10.87	4.00	2.97
Tashayo et al., 2020a, Environ Manage	2020	8	40	20.00	6.07	1.99
Tashayo et al., 2020b, J Saudi Soc Agric Sci	2020	8	48	16.67	6.07	2.38
Sarkar et al., 2014, Geo-Spatial Inf Sci	2014	8	37	21.62	1.85	1.23
Worqlul et al., 2017, Appl Geogr	2017	7	98	7.14	4.20	2.85
Jamil et al., 2018, Agric Res	2018	6	31	19.35	2.81	0.65
Vasu et al., 2018, Land Use Policy	2018	5	68	7.35	2.34	1.43
Seyedmohammadi et al., 2018, Geoderma	2022	5	33	15.15	21.43	3.71

Source: Based on Biblioshiny Software

3.4 Country Analysis

Table 3 highlights the top 10 nations with the largest number of citations, even though authors from 38 different countries or regions contributed to the study. Italy and Portugal are the two European nations in the top 10, while the top five nations are overall—Iran, India, China, Korea, and Japan. The specialization is that the top five countries are all located in Asia, and the contribution of conducting research in those nations is noteworthy. The majority of research publications on GIS-based cropland Suitability are published in developing regions, such as Asia, as Table 4 demonstrates. These results imply that these areas are leading the field. With a significant disparity between India and China (with 418 and 364 citations, respectively), Iran was expected to hold the top spot with a sizable 708 citations with 32.2 Average Article Citations. Studying GIS-based analysis may have a greater theoretical and practical impact on countries that are developing. Further, the number of papers published in a given country on a certain topic may serve as a proxy for the country's stature and influence in the field of GIS-based analysis research.

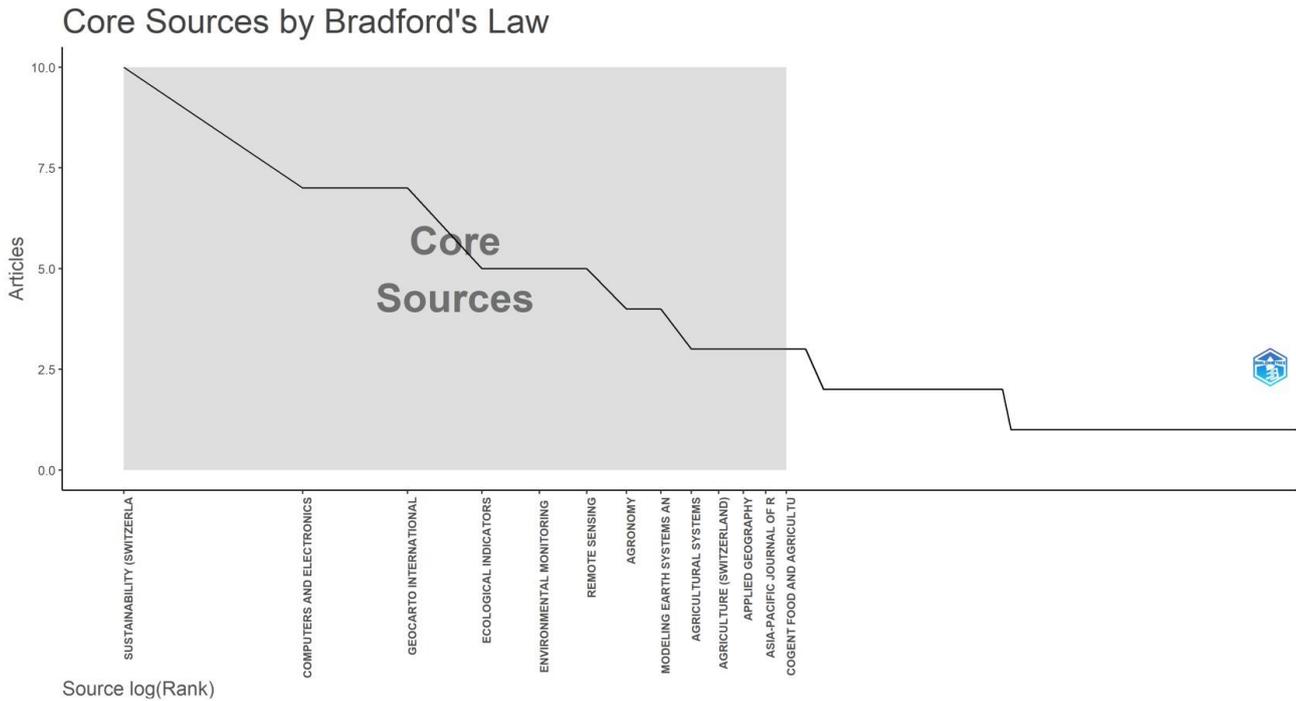


Figure 9. Core Sources by the Bradford's Law

Table 3. Core Sources based on Bradford's Law

Source	Freq	cumFreq	Zone
Sustainability (Switzerland)	10	10	Zone 1
Computers And Electronics In Agriculture	7	17	Zone 1
Geocarto International	7	24	Zone 1
Ecological Indicators	5	29	Zone 1
Environmental Monitoring And Assessment	5	34	Zone 1
Remote Sensing	5	39	Zone 1
Agronomy	4	43	Zone 1
Modeling Earth Systems And Environment	4	47	Zone 1

Figure 10 displays the articles' geographic distribution. The frequency observed to be higher than the number of articles analyzed is explained by the figure, which is based on the co-occurrence of the nation's according to the affiliations of the authors. Upon examining the top ten most productive nations, it is evident that most of them are found in the continents of North and South America, Asia, Africa, and Europe. India is an incredibly productive partner for other nations. With 19.78% of the 48 countries, India has a considerably higher document number (frequency) than any other country. However, the field of cropland suitability utilizing GIS in India has many documents, but the average article citation per item is only 12.7, suggesting that the article level still has to be further enhanced. Seven different countries have collaborated on at least one study. At a rate of 148, 73, 42, and 38 times a year, respectively, China, India, Iran, and the United States are all involved in frequent collaborations. Portugal is the nation with the most average citations per item, coming in at number 58.3, while its document count only makes it to the 7th. It demonstrates that the article level is high despite the low document count. Additionally, the graphic shows that there are more local authors when the shades of blue go from lighter to darker. In contrast, grey denotes the absence of local authors. In theory,

emerging nations and agricultural countries stand to gain more from the study of GIS-based agricultural land suitability. Among them, some agriculturally oriented regions were unable to support greater academic research due to a lack of funding and a relatively low level of science and technology. Moreover, it is challenging to hold a significant number of positions in the world due to the lack of technology, money, and talent pools in many emerging nations.

Table 4. Most Cited Countries

Country	Total Citation	Average Article Citations
Iran	708	32.2
India	418	12.7
China	364	40.4
Korea	210	10.5
Japan	203	18.5
USA	194	27.7
Portugal	175	58.3
Turkey	153	15.3
Italy	144	20.6
Ethiopia	114	12.7

Source: Based on Biblioshiny Software

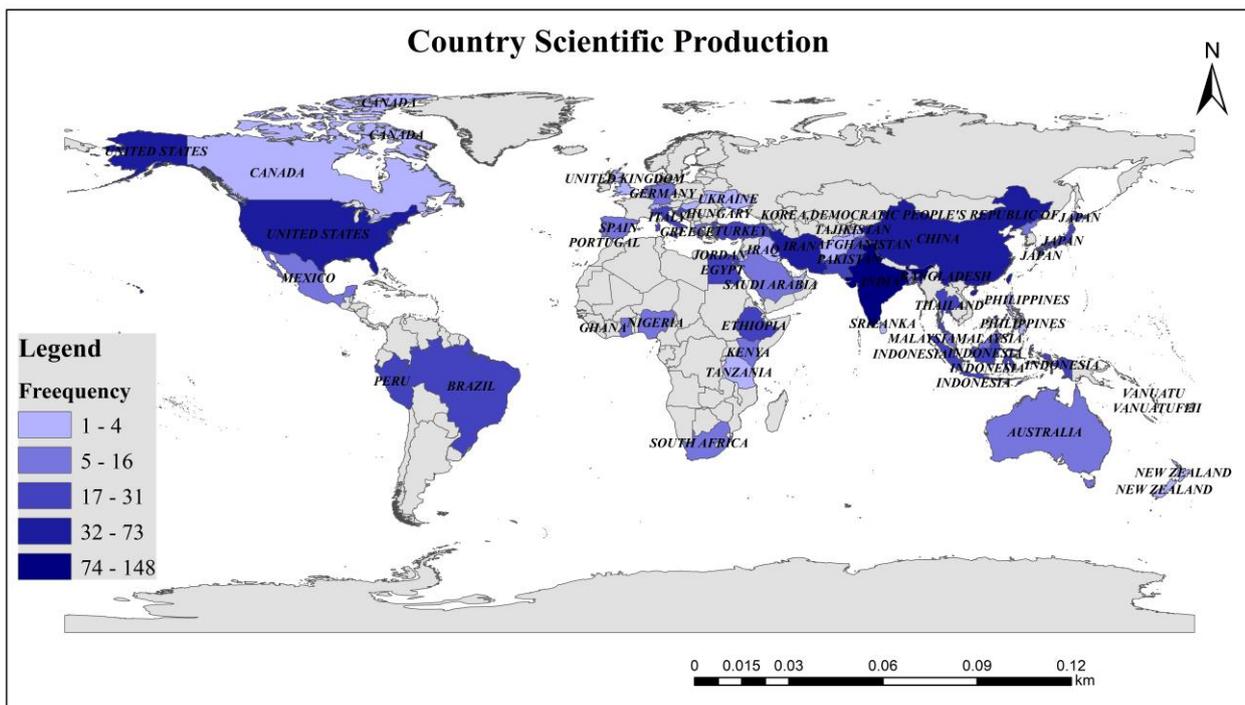


Figure 10. Country Scientific Production

Figure 11 displays the number of publications in which each article is categorized by the Corresponding author's association with a particular nation. This research is intended to ascertain the proportion of publications in which at least one author is linked to a nation other than the one of the corresponding author. Only domestic authors are featured in more publications than authors from other countries. More research on GIS-based cropland suitability has been published in India than in any other country, albeit the majority of these studies were carried out on their own. The findings showed that there were generally a considerable number of papers written by Indian authors. The GIS-based cropland Suitability collaboration data of sample countries reveal that 21.53 percent of Indian publications are prepared independently. While there is more than a 35% multiple-country collaboration ratio, nations like Egypt, Brazil, and Pakistan are involved in international collaboration. In particular, specifically, no articles spanning multiple countries are visible in Indonesia and Croatia. However, nine distinct nations were found to have a higher level of international collaboration (MCP ratio = 1.0) but have just one multiple-country article: Armenia, Bangladesh, Ghana, Netherlands, New Zealand, South Africa, Tanzania, Ukraine, and the United Arab Emirates.

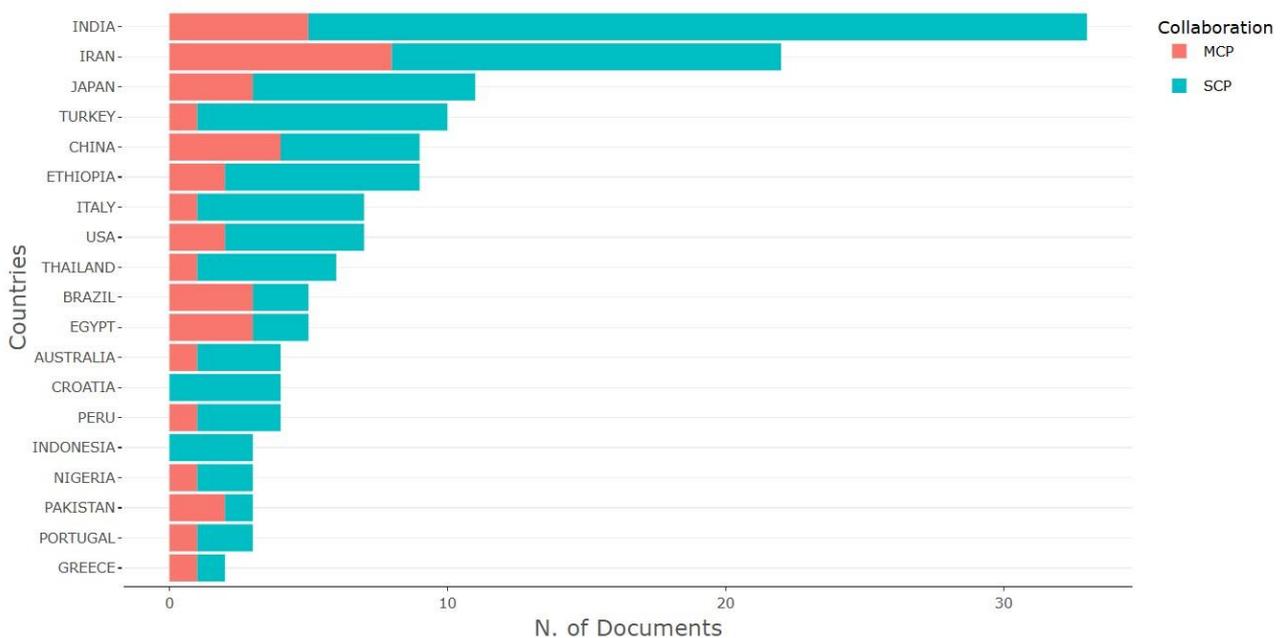


Figure 11. Corresponding Author's Country

3.5 Keyword Analysis

Using keywords, the article's main points are clarified and high-level summarized. The software program biblioshiny does statistical analysis and data mining on the research articles' high-frequency keywords. The most commonly occurring keywords in the GIS-based land suitability analysis are chosen and shown in Table 5 and Figure 12 as those with word frequencies greater than or equal to 7. The research papers' high-frequency keywords are subjected to statistical analysis and data mining using the software package Biblioshiny. The concept and writing approach to the paper in the GIS-based cropland suitability is succinct and reflected in the analysis of keywords like cluster analysis and multiple correspondence analysis. The keywords "GIS," "Land suitability," and "AHP" are most commonly used in the context of GIS-based cropland suitability; they account for 24.5%, 22.5%, and 16.8% of explorations, respectively. The reader can also comprehend the potential boundaries of GIS-based research through the keyword analysis.

The theme map relative to the appropriateness of cropland by GIS is shown in Figure 13. The themes that are currently popular in this field of study are presented visually. Based on the topics' density and centrality, thematic maps are separated into four quadrants, each of which symbolizes a distinct topic typology. Density is a measurement of the theme's evolution, while centrality indicates the theme's significance throughout the full investigation of this image. A high level of centralization and density is implied by the prevalence of motifs in the upper-right quadrant. Author keyword clusters and their relationships are the source of topics for the thematic analysis. These themes have characteristics that define them. A node is considered more central and

important if it has a higher number of relations with other nodes in the theme network and has an essential location within the network. Cohesion inside a node, which symbolizes the density of a field of study, similarly defines the field's capacity to grow and endure. The top-right portion of the map displays the themes that have recently gained significance. The driving themes are shown in the upper right quadrant (Q1), underlying themes are shown in the lower right quadrant (Q4), much-specialized themes are shown in the upper left quadrant (Q2), and emerging themes are shown in the lower left quadrant (Q3) (Pai et al., 2022). Author Keywords have a deeper and more varied ability to convey the substance of an article.

Table 5. Frequency Analysis of Authors' Keywords

Words	Occurrences	% Occurrences
GIS	51	24.5
Land Suitability	47	22.5
AHP	35	16.8
Analytical Hierarchy Process	18	8.6
Remote Sensing	13	6.25
Analytic Hierarchy Process	10	4.8
Analytical Hierarchy Process (AHP)	10	4.8
Suitability	9	4.3
Land Suitability Analysis	8	3.8
Multi-Criteria Analysis	7	3.3

Source: Based on Biblioshiny Software

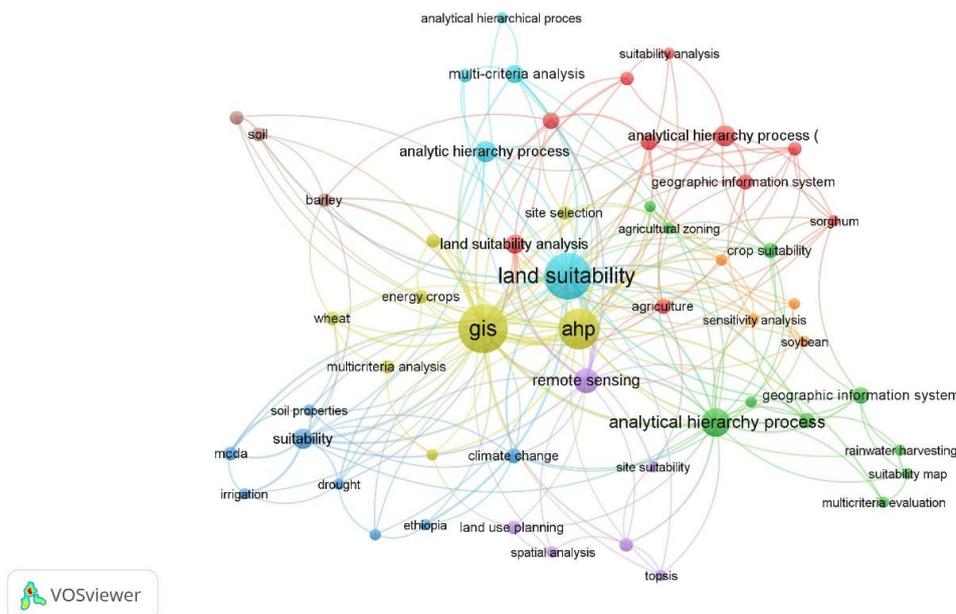


Figure 12. Keywords Trend of the Most Utilized Words

The motor themes are displayed in Q1. They are distinguished by high density as well as centrality. Two of the "motor themes" that have received the greatest attention in the literature are "Land Suitability" and "GIS". In this review, sustainable development is the main focus. Performance, progress, technical advancement, and sustainability are just a few of the numerous ideas that are related to this theme. Since our initial subject categories were "suitability," "modeling," "assessment," and "evaluation," this word has emerged as one of the most frequently used ones in the cluster sustainable development of this sector. The results of this study indicate that subjects related to dataset repositories, multi-criteria evaluation, rainwater collecting, crops, and irrigation, and results have further enumerated that India has the greatest potential in Q2. As a result, these themes are mainly central but sparsely distributed, and thus have a negligible impact on the field's growth of cropland suitability utilizing GIS. The fundamental and core research fields that serve as the field's cornerstones are included in the basic themes' quadrant. Keywords like "multi-criteria analysis," "topography," "AHP," "modeling," and "land suitability" are included in this instance's bubble. These keywords draw emphasis on the notion and prompt development of GIS-based cropland suitability. Additionally, some of the subjects within these themes are required to advance the area of GIS-based land suitability, as evidenced by the way the themes appear to be evolving while still intersecting with the core themes. Finally, study areas that are either gaining traction or losing significance over time are captured by the emerging or declining themes quadrant. Words like "AHP" and "agricultural land suitability" are associated with two bubbles in this quadrant.

Three main groups were formed by arranging study areas related to scientific production according to time intervals estimated in years (Figure 14). Strong evidence of the differentiation, integration, transfer, and regeneration of themes may be seen in the intricate relationships between the many thematic evolutions and the diverse spectrum of study themes across time. It is possible to emphasize how topics combine or divide into multiple themes using these longitudinal analyses (Aria & Cuccurullo, 2020). The qualities of the lines indicate the caliber of the connections between the keywords. Research subjects can be arranged into motor and basic themes using the map of the thematic progression of scientific production (Melega, 2022).

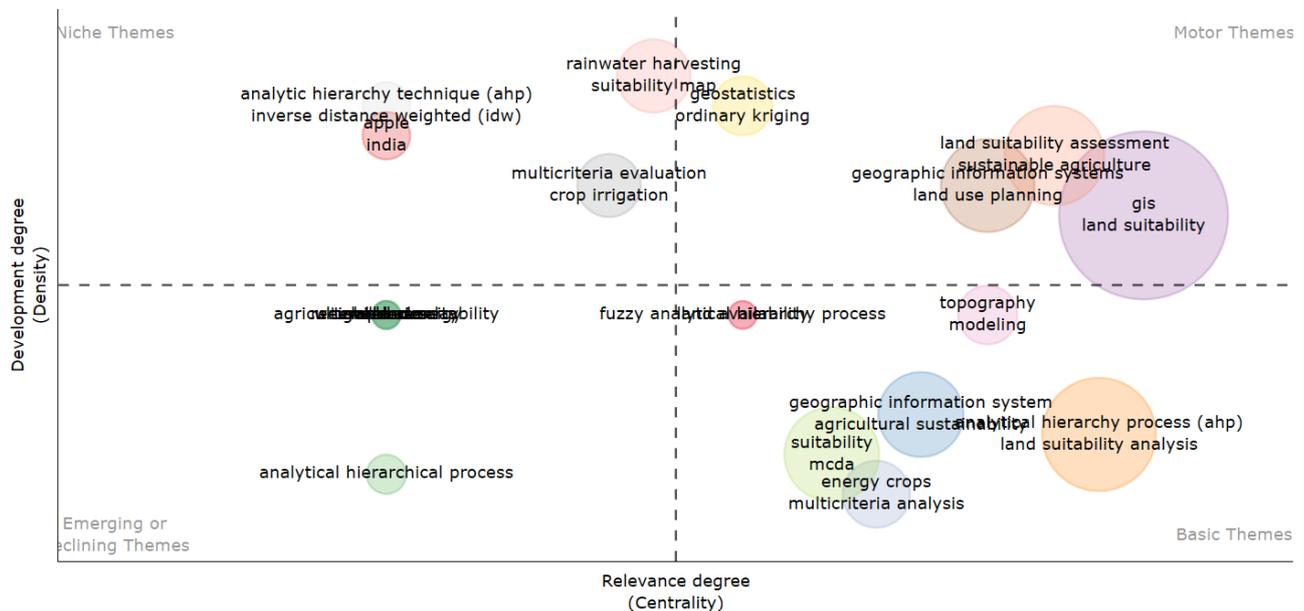


Figure 13. Thematic Map

Suitability evaluation, suitability analysis, the hierarchy process, and remote sensing are the most common topics in the first phase (2010–2018), which makes sense given that research was still in the early stages of trying to define and clarify the idea at this time. The study on the subject is divided into multiple themes throughout the second phase (2018–2020). The early growth stage concentrated on decision analysis in this context and the expansion of land suitability and cropland. The study "Development of a Model Using Matter Element, AHP, and GIS Techniques to Assess the Suitability of Land for Agriculture" by Seyedmohammadi et al. (2019) was one of the most cited articles in the second phase, with 114 global citations. The study looked

at the suitability of land for barley production in the Dasht-e-Moghan region using a new framework that integrates the matter element model, AHP, RS, and GIS. Parts of the themes from 2021 and 2023 continue, including the AHP, GIS, and land suitability assessment. Furthermore, MCDA for cropland suitability analysis emerged as a prominent research topic in this phase of the GIS-based study, which focused on applying AHP. A great deal of research has also been conducted on cropland to assist in the development of sustainable land development and food production. For instance, in the Tozanlı subbasin, which is situated in the upper region of the Yesilirmak Basin in Turkey, Kılıc et al. (2022) investigated land suitability for wheat cultivation and created a land suitability map for wheat by combining the AHP-Fuzzy algorithm with GIS. Also, the research field's evolution and history were explained through the utilization of keywords and topic evolution. It is noteworthy that compared to earlier times, a significantly greater number of publications were published in the years 2019–2023.

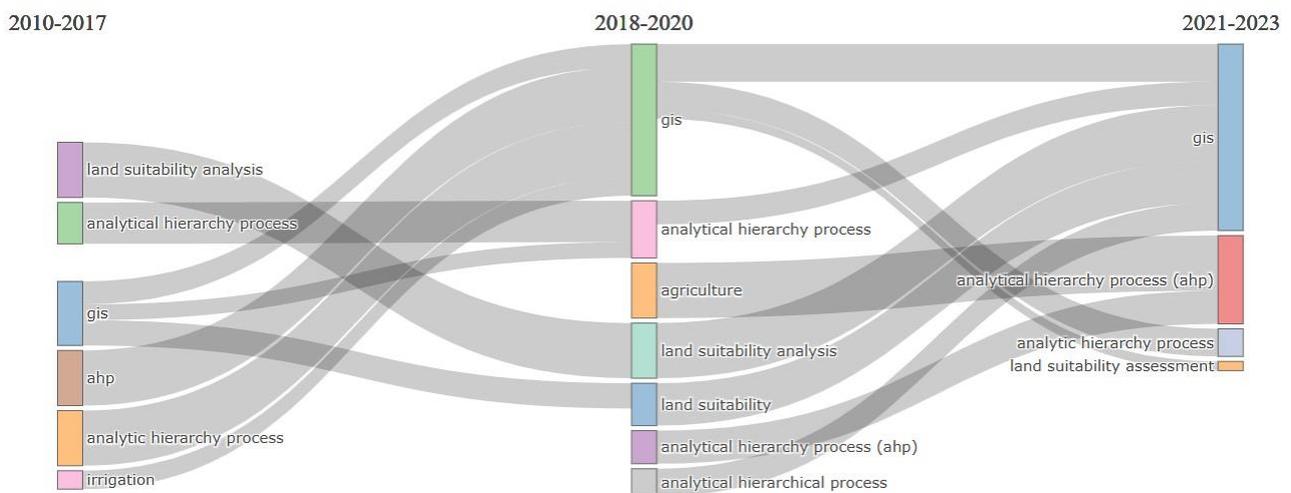


Figure 14. Thematic Evolution of GIS-based CropLand Suitability from 2010-2023

4. DISCUSSION

The major goal of this study was to comprehend the state of the literature and the patterns of recent publications in the field of GIS-based cropland suitability analyses. The study uses bibliometric analysis as a method of data analysis to explore the literature from the last ten years to achieve this goal. The topic has gained significance since the sector has seen an increase in global concern over the past ten years. The growing number of articles in this field of study amply demonstrates the significance and pertinence of this subject in the field. 183 publications that were published in 116 journals between October 2010 and October 2023 were included in the study. There has been an increasing tendency in the number of documents, particularly after 2019. In recent years, there has been an upsurge in research on site selection (suitability) issues utilizing GIS-MCDA in conjunction with the AHP techniques. AHP is frequently used in multi-criteria decision-making to establish the relative weights of each criterion. An instrument for pairwise comparison of the several criteria and sub-criteria taken into account for site selection issues is offered by AHP (Wigati et al., 2019). The relative importance of two criteria and the degree to which one criterion supersedes the other are determined via comparison. To determine each criterion's weight, a diagonal comparison matrix is created (Wigati et al., 2019). Research on AHP and GIS-integrated agricultural land suitability had not been done much until 2018, but since then, there has been a trend of presenting this type of data. Also, the contribution of researchers in countries like Croatia, Japan, Iran, and Peru is high for this kind of analysis. Conducting research in this field through high-level universities like the University of Tsukuba and Islamic Azad University has also been a factor in determining the productivity of the researchers as well as the field. Local citation and global citation factors have also had a positive impact on several types of research during the selected study period. Through that, the appropriateness and effectiveness of those researches are directly integrated into other research and it also contributes to the development status of this context. It is also unique that most of the articles selected for the study have been published in Journals with a 3.3 Impact Factor such as Sustainability. The journal

Sustainability (Switzerland) focuses on interdisciplinary studies related to sustainability, which aligns well with the themes of GIS-based AHP and cropland suitability. This alignment encourages researchers to publish their work in this journal. Publishing in a reputable journal like Sustainability (Switzerland) ensures high visibility and impact for researchers. The journal's emphasis on sustainability-related research attracts contributions that address critical global challenges.

Furthermore, the majority of the research has been conducted in developing nations, mostly in Asia, demonstrating the necessity of food production in these areas. AHP and GIS were the most often utilized site selection techniques, indicating their efficacy in identifying appropriate areas for cropland suitability. Decision rules aid in the evaluation of potential options for selection by helping decision-makers determine which ones are most appropriate. Although the applications of AHP and GIS have become popular tools for site selection, it is crucial to recognize that their efficacy may depend on several contextual factors (Subagyo et al., 2023). The growing body of studies on the best places to locate farms represents a step in the right direction toward more crop production and sustainable land development. Land sustainability in productivity can be addressed on a regional scale by researchers, politicians, and decision-makers by concentrating on these topics. The aforementioned study emphasizes the significance of implementing a comprehensive strategy that takes into account both the technical viability and wider socio-economic and environmental consequences. There needs to be more discussion on GIS spatial analytic modeling for land suitability, according to the vast number of academic papers. The results show that a large number of public universities have been instrumental in the advancement of GIS-based research. China, Japan, India, and Iran are home to the bulk of these colleges and universities. This suggests that GIS aids in the growth of nations with sizable economies and advanced technological sectors. The findings demonstrate that a co-occurrence network is useful in revealing the relationships between the frequently occurring keywords in a word cloud, in addition to helping to identify them. The more frequently occurring terms are larger and more apparent in the word cloud presented in Figure 15, whereas the less frequently occurring terms are smaller. Word clouds can be used to find the midpoint of written letters (Atenstaedt, 2012). Examining the keywords that writers use in their articles is a crucial method for learning about trendy topics and areas of interest for academics. With the aid of the most often used terms, word cloud analysis makes it possible to swiftly determine the publication's topic and concentration. Also, through the use of size and color, a word cloud converts texts into tags—words whose relative importance may be seen in the resultant cloud (Mulay et al., 2020). When these keywords are further examined, it becomes clear from their color code that the larger keyword (represented by its width) is related to the smaller keywords in a coordinated way. Choosing author keywords has the primary benefit of offering insight into popular subjects and emerging research trends. The GIS-based cropland suitability analysis, which is the main focus of this study, has a strong association with the analytical hierarchy process, multi-criteria analysis, GIS, and cultivation.

4.1. Trend Analysis

Numerous studies have focused on a new objective for text processing and data exchange once it was examined. Every year there has been a rise in the quantity of research articles as well; initially, they were centered upon AHP and MCDA, but now they include a larger range of cropland suitability topics. About problem structuring, element representation, quantification, linking parts to overarching objectives, and alternative solution evaluation, the AHP offers a reasonable and thorough framework (Saaty, 1980). The study's research trends show that there is an increase in cropland suitability based on GIS research data sources, as well as an expansion of application areas and a diversification of research methodologies. The intersection and integration of several disciplines will probably become a major trend in cropland suitability research in the future, and this field's application to the environmental background will likely grow wider. More research and analysis of GIS and AHP-based suitability analysis were suggested by the study. Policymakers, academics, researchers, and decision-makers should concentrate on developing early intervention techniques and raising essential awareness. On the other hand, numerous methods can be used to conduct a bibliometric analysis for the current study. Other than the ones employed in this work, future researchers can apply bibliometric techniques to obtain new and improved insights into this topic. Developing more sophisticated bibliometric analysis technologies in the future could enable the coverage and visualization of more parameters.



Figure 15. Cloud Illustration of the Keywords

4.2. Future Research Direction

Figure 16 displays the themes that are currently in trend as well as possible future advances in the discipline. Author-defined article keywords typically relate to this type of publication material and are adequate to extract subject elements of a field (Aria & Cuccurullo, 2017). Also, larger circles indicate more instances of the topic, while the circles on the lines show the number of occurrences of the topic (Ma et al., 2023). This research provides further context for the themes that have emerged over time about keyword occurrences in the literature on cropland suitability, AHP, and GIS. These topics may have various connections to the appropriateness of agricultural land. For example, within the issue of GIS-based cropland suitability in 2020, the most talked-about topics were "land suitability" and "GIS." Throughout 2018, themes like fuzzy logic and GIS were linked to the appropriateness of cropland based on GIS. Another important factor that has been emphasized since 2021 is the AHP and suitability. AHP, MCDA, and GIS scopes are likely to be given more attention in cropland suitability analyses. It is possible to confirm the efficiency and appropriateness of the analysis through these keywords in the period selected for the study. This can be described as one of the most prominent and productive areas of GIS-related research that is steadily growing. In the year 2023, the keyword "Climate change" has been discussed in several researches under a frequency of 10%. With the development of several keywords under trends topics in such research, the trend of such effective research is increasing. Effective goals can be reached with the use of such scientific methods in increasing the trend of future research directions.

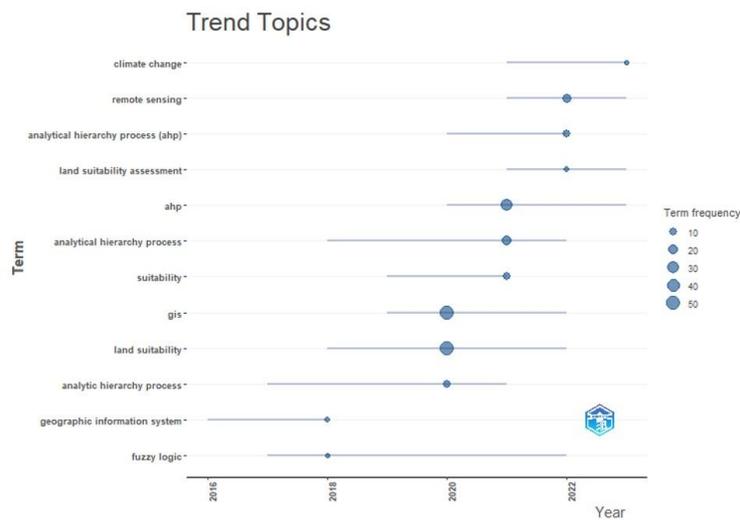


Figure 16. Trending Topics

5. CONCLUSION

Using Biblioshiny and VOSviewer, a comprehensive bibliometric analysis of the growth and evolution of GIS-based cropland suitability was conducted in this study using 183 papers from Scopus that were published over the last 23 years. Significant trends were found in the bibliometric examination of appropriate places for cropland utilizing AHP and GIS techniques. Preliminary findings, future directions for researchers, and gaps

in the literature concerning site selection and evaluation issues taken into consideration for cropland are all examined. The number of papers has increased, indicating a notable increase in research interest in recent years. The research papers' integration of diverse methods and approaches within planning disciplines has been accelerated by advancements in GIS and AHP. GIS is now a key component of this field's objective decision-making process due to the explosion of big data in data-driven studies. Researchers are conducting various research in the developing world of technology. At a time when an expanded research environment is being implemented through GIS, a growing impact has been added to the agricultural fields through selected studies. Through the use of multi-criteria methods, the yield and success of crops depend on such selected criteria. The productive environment of a cropland can in any way affect the variety of crops and its suitability to establish a successful plantation can be determined. The suitability of several criteria can be checked through AHP-MCDA in determining a suitable land for growing an agricultural crop. In this research as well, researchers using such methods have identified it as an effective tool. The selection of suitable and unsuitable areas for an agricultural crop depends on various environmental and soil criteria. Distribution of related crops can be conducted according to geographical regions. The researchers have presented whether it is possible to decide a way that the relevant criterion is suitable for the land through multiple decision analysis and the yield of the crop can also be determined according to the appropriateness of those criteria. Primarily, researchers have focused on the extent to which the hierarchical process in GIS and AHP can be used to select a suitable site. When it comes to the scientific development of systematic literature evaluations on cropland suitability using GIS, the Journal of Sustainability is a pioneer. India was the country that produced the most publications. Iran has the highest level of multi-country publication with other countries regarding the analysis of the corresponding Author's country, with the bulk of studies being single-country publications. Subsequent studies ought planners to embrace GIS as a crucial methodological technique in diverse land use procedures for multi-criteria decision analysis and decision-makers. For decision-makers, researchers, and practitioners involved in developing cropland, it gives a methodical approach to site selection and offers insightful information.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- Abushnaf, F. F., Spence, K. J., & Rotherham, I. D. (2013). Developing a land evaluation model for the Benghazi region in Northeast Libya using a geographic information system and multi-criteria analysis. *APCBEE Procedia*, 5, 69-75. <https://doi.org/10.1016/j.apcbee.2013.05.013>
- Amin, S., Rohani, A., Aghkhani, M. H., & Abbaspour-Fard, M. H., Asgharipour, M. R. (2020). Assessment of land suitability and agricultural production sustainability using a combined approach (Fuzzy-AHP-GIS): A case study of Mazandaran province, Iran. *Information Processing in Agriculture*, 7(3), 384-402. <https://doi.org/10.1016/j.inpa.2019.10.001>
- Al Garni, H. Z., & Awasthi, A. (2018). Solar PV Power Plants Site Selection: A Review. In: I. Yahyaoui (Eds.), *Advances in Renewable Energies and Power Technologies Volume 1: Solar and Wind Energies* (pp. 57-75). <https://doi.org/10.1016/B978-0-12-812959-3.00002-2>
- Aria, M., & Cuccurullo, C. (2020). Biblioshiny: Bibliometrix for No Coders. 2020. (Accessed:25/04/2024) <https://www.bibliometrix.org/home/index.php/layout/bibliometrix>
- Aria, M., & Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Atenstaedt, R. (2012). Word cloud analysis of the BJGP. *British Journal of General Practice*, 62(596), 148. <https://doi.org/10.3399/bjgp12X630142>
- Azzari, G., Jain, M., & Lobell, D. B. (2017). Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment*, 202, 129-141. <https://doi.org/10.1016/j.rse.2017.04.014>

- Baydas, M., Elma, O. E., & Pamučar, D. (2022). Exploring the specific capacity of different multi-criteria decision-making approaches under uncertainty using data from financial markets. *Expert Systems with Applications*, 197, 116755. <https://doi.org/10.1016/j.eswa.2022.116755>
- Black, P. E. (2004). "Bradford's law" in Dictionary of Algorithms and Data Structures, U.S. National Institute of Standards and Technology. <https://www.nist.gov/dads/HTML/bradfordsLaw.html>
- Bradford, S. C. (1934). Sources of Information on Specific Subjects. *Engineering: An Illustrated Weekly Journal*, 137, 85-86.
- Chiao, K. P. (2021). Multi-criteria decision making with interval type 2 fuzzy Bonferroni mean. *Expert Systems with Applications*, 176, 114789. <https://doi.org/10.1016/j.eswa.2021.114789>
- Chinthaka, M. L. C., Amaraweera, P. H., & Wijesinghe, W. M. D. C. (2023). The Identification of Landslide Risk-Prone Areas in the Imbulpe Divisional Secretariat in Sri Lanka: A GIS-Based Multi-Criteria Decision Analysis. *Journal of Social Sciences and Humanities Review (JSSHR)*, 8(1), 8-29. <https://doi.org/10.4038/jsshr.v8i1.115>
- Ejaz, H., Zeeshan, H. M., Ahmad, F., Bukhari, S. N. A., Anwar, N., Alanazi, A., Sadiq, A., Junaid, K., Atif, M., Abosalif, K. O. A., Iqbal, A., Hamza, M. A., & Younas, S. (2022). Bibliometric Analysis of Publications on the Omicron Variant from 2020 to 2022 in the Scopus Database Using R and VOSviewer. *International Journal of Environmental Research and Public Health*, 19(19), 12407. <https://doi.org/10.3390/ijerph191912407>
- FAO (1976). A framework for land evaluation. Food and Agriculture Organization of the United Nations, Soils Bulletin No. 32. FAO, Rome.
- Habibie, M. I., Noguchi, R., Shusuke, M., & Ahamed, T. (2021). Land suitability analysis for maize production in Indonesia using satellite remote sensing and GIS-based multicriteria decision support system. *GeoJournal*, 86(2), 777-807. <https://doi.org/10.1007/s10708-019-10091-5>
- Houshyar, E., SheikhDavoodi, M. J., Almassi, M., Bahrami, H., Azadi, H., Omidi, M., Sayyad, G., & Witlox, F. (2014). Silage corn production in conventional and conservation tillage systems. Part I: Sustainability analysis using combination of GIS/AHP and multi-fuzzy modeling. *Ecological Indicators*, 39, 102-114. <https://doi.org/10.1016/J.Ecolind.2013.12.002>
- Huang, J., Wu, X., Ling, S., Li, X., Wu, Y., Peng, L., & He, Z. (2022). A bibliometric and content analysis of research trends on GIS-based landslide susceptibility from 2001 to 2020. *Environmental Science and Pollution Research*, 29, 86954-86993. <https://doi.org/10.1007/s11356-022-23732-z>
- Jamil, M., Sahana, M., & Sajjad, H. (2018). Crop Suitability Analysis in the Bijnor District, UP, Using Geospatial Tools and Fuzzy Analytical Hierarchy Process. *Agricultural Research*, 7(4), 506-522 <https://doi.org/10.1007/s40003-018-0335-5>
- Jayasinghe, A. D. S., & Withanage, W. K. N. C. (2021). A geographical information system-based multi-criteria decision analysis of potato cultivation land suitability in Welimada divisional secretariat, Sri Lanka. *Potato Journal*, 47(2), 126-134.
- Kazemi, H., Sadeghi, S., & Akinci, H. (2016). Developing a land evaluation model for faba bean cultivation using geographic information system and multi-criteria analysis (A case study: Gonbad-Kavous region, Iran). *Ecological Indicators*, 63, 37-47. <http://doi.org/10.1016/j.ecolind.2015.11.021>
- Kahsay, A., Haile, M., Gebresamuel, G., Mohammed, M., & Tejada Moral, M. (2018). Land suitability analysis for sorghum crop production in northern semi-arid Ethiopia: Application of GIS-based fuzzy AHP approach. *Cogent Food & Agriculture*, 4(1), 1507184. <https://doi.org/10.1080/23311932.2018.1507184>
- Kilic, O. M., Ersayın, K., Gunal, H., Khalofahc, A., & Alsubeie, M. S. (2022). Combination of fuzzy-AHP and GIS techniques in land suitability assessment for wheat (*Triticum aestivum*) cultivation. *Saudi Journal of Biological Sciences*, 29(4), 2634-2644. <https://doi.org/10.1016/j.sjbs.2021.12.050>
- Kihoro, J., Bosco, N. J., & Murage, H. (2013). Suitability analysis for rice growing sites using a multicriteria evaluation and GIS approach in great Mwea region, Kenya. *SpringerPlus*, 2, 265. <https://doi.org/10.1186/2193-1801-2-265>

- Liao, H., Yang, S., Kazimieras, Zavadskas, E. K., & Škare, M. (2023). An overview of fuzzy multi-criteria decision-making methods in hospitality and tourism industries: bibliometrics, methodologies, applications and future directions. *Economic Research-Ekonomska Istraživanja*, 36(3), 2150871. <https://doi.org/10.1080/1331677X.2022.2150871>
- Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, 164, 324-333. <https://doi.org/10.1016/j.rse.2015.04.021>
- Ma, H., Ismail, L., Noordin, N., & Razali, A. B. (2023). Bibliometric analysis of willingness to communicate in the English as a second language (ESL) context. *Humanities and Social Sciences Communications*, 10, 702. <https://doi.org/10.1057/s41599-023-02109-8>
- Maleki, F., Kazemi, H., Siahmarguee, A., & Kamkar, B. (2017). Development of a land use suitability model for saffron (*Crocus sativus* L.) cultivation by multi-criteria evaluation and spatial analysis. *Ecological Engineering*, 106(Part A), 140-153. <https://doi.org/10.1016/J.Ecoleng.2017.05.050>
- Melega, A. (2022). Bibliometric Analysis of Scientific Production Regarding the Harmonization of Accounting In Brics Emerging Economies. *European Journal of Accounting, Finance & Business*, 10(1), 11-20. <https://doi.org/10.4316/EJAFB.2022.1012>
- Miller, H. J., & Goodchild, M. F. (2014). Data-driven geography. *GeoJournal*, 80(4), 449-461. <https://doi.org/10.1007/s10708-014-9602-6>
- Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional De La Información*, 29(1). <https://doi.org/10.3145/epi.2020.ene.03>
- Mulay, P., Joshi, R., & Chaudhari, A. (2020). Distributed incremental clustering algorithms: a bibliometric and word-cloud review analysis. *Science & Technology Libraries*, 39(3), 289-306. <https://doi.org/10.1080/0194262X.2020.1775163>
- Musakwa, W. (2018). Identifying land suitable for agricultural land reform using GIS-MCDA in South Africa. *Environment, Development and Sustainability*, 20, 2281-2299. <https://doi.org/10.1007/s10668-017-9989-6>
- Ostovari, Y., Honarbakhsh, A., Sangoony, H., Zolfaghari, F., Maleki, K., & Ingram, B. (2019). GIS and multi-criteria decision-making analysis assessment of land suitability for rapeseed farming in calcareous soils of semi-arid regions. *Ecological Indicators*, 103, 479-487. <https://doi.org/10.1016/j.ecolind.2019.04.051>
- Özkan., B., Özceylan., E., & Sariçiçek, I. (2019). GIS-based MCDM modeling for landfill site suitability analysis: A comprehensive review of the literature. *Environmental Science and Pollution Research*, 26, 30711-30730. <https://doi.org/10.1007/s11356-019-06298-1>
- Pai, R. Y., Shetty, A., Shetty, A. D., Bhandary, R., Shetty, J., Nayak, S., Dinesh, T. K., & D'souza, K. J. (2022). Integrating artificial intelligence for knowledge management systems – synergy among people and technology: a systematic review of the evidence, *Economic Research-Ekonomska Istraživanja*, 35(1), 7043-7065. <https://doi.org/10.1080/1331677X.2022.2058976>
- Palmisano, G. O., Loisi, R. V., Ruggiero, G., Rocchi, L., Boggia, A., Roma, R., & Dal Sasso, P. (2016). Using Analytic Network Process and Dominance-based Rough Set Approach for sustainable requalification of traditional farm buildings in Southern Italy. *Land Use Policy*, 59, 95-110. <https://doi.org/10.1016/j.landusepol.2016.08.016>
- Purnamasari, R. A., Ahamed, T., & Noguchi, R. (2019). Land suitability assessment for cassava production in Indonesia using GIS, remote sensing, and multi-criteria analysis. *Asia-Pacific Journal of Regional Science*, 3, 1-32. <https://doi.org/10.1007/s41685-018-0079-z>
- Qin, Z., Zhao, Z., Xia, L., & Ohore, O. E. (2022). Research trends and hot spots of aquatic biofilms in the freshwater environment during the last three decades: a critical review and bibliometric analysis. *Environmental Science and Pollution Research*, 29, 47915-47930. <https://doi.org/10.1007/s11356-022-20238-6>

- Rautela, S., Sharma, A., & Panackal, N. (2024). Exploring the mental well-being of higher educational institutions students: a bibliometric analysis. *Cogent Education*, *11*(1), 2343522. <https://doi.org/10.1080/2331186X.2024.2343522>
- R CoreTeam. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rediske, G., Siluk, J. C. M., Gastaldo, M. G., Rigo, P. D., & Rosa, C. B. (2018). Determinant factors in site selection for photovoltaic projects: A systematic review. *International Journal of Energy Research*, *43*(5), 1689-1701. <https://doi.org/10.1002/er.4321>
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill: New York, NY, USA.
- Sakamoto, T., Gitelson, A. A., & Arkebauer, T. J. (2013). MODIS-based corn in yield estimation model in incorporating crop phenology information. *Remote Sensing of Environment*, *131*, 215-231. <https://doi.org/10.1016/j.rse.2012.12.017>
- Sarkar, A., Ghosh, A., & Banik, P. (2014). Multi-criteria land evaluation for suitability analysis of wheat: a case study of a watershed in eastern plateau region, India. *Geo-spatial Information Science*, *17*(2), 119-128. <https://doi.org/10.1080/10095020.2013.774106>
- Sathiyamurthi, S., Saravanan, S., Sankriti, R., Aluru, M., Sivaranjani, S., & Srivel, R. (2022). Integrated GIS and AHP techniques for land suitability assessment of cotton crop in Perambalur District, South India. *International Journal of System Assurance Engineering and Management*, *15*(1), 267-278. <https://doi.org/10.1007/s13198-022-01705-2>
- Seyedmohammadi, J., Sarmadian, F., Jafarzadeh, A. A., Ghorbani, M. A., & Shahbazi, F. (2018). Application of SAW, TOPSIS and fuzzy TOPSIS models in cultivation priority planning for maize, rapeseed and soybean crops. *Geoderma*, *310*, 178-190. <https://doi.org/10.1016/j.geoderma.2017.09.012>
- Seyedmohammadi, J., Sarmadian, F., Jafarzadeh, A. A., & McDowell, R. W. (2019). Development of a model using matter element, AHP, and GIS techniques to assess the suitability of land for agriculture. *Geoderma*, *352*, 80-95. <https://doi.org/10.1016/j.geoderma.2019.05.046>
- Singha, C., & Swain, K. C. (2016). Land suitability evaluation criteria for agricultural crop selection: A review. *Agricultural Reviews*, *37*(2), 125-132. <https://doi.org/10.18805/ar.v37i2.10737>
- Subagyo, Moh. Yanuar, J. P., Bambang, P. N., Saleh, A., & Akhmad, F. W. (2023). The Suitable Location for a Hybrid Renewable Energy Wind Solar Power Plant: A Review by Bibliometric. *IOP Conference Series: Earth and Environmental Science*, *1266*, 012090. <https://doi.org/10.1088/1755-1315/1266/1/012090>
- Tashayo, B., Honarbakhsh, A., Azma, A., & Akbari, M. (2020a). Combined fuzzy AHP-GIS for agricultural land suitability modeling for a watershed in Southern Iran. *Environmental Management*, *66*(3), 364-376. <https://doi.org/10.1007/s00267-020-01310-8>
- Tashayo, B., Honarbakhsh, A., Akbari, M., & Eftekhari, M. (2020b). Land suitability assessment for maize farming using a GIS-AHP method for a semi- arid region, Iran. *Journal of the Saudi Society of Agricultural Sciences*, *19*(5), 332-338. <https://doi.org/10.1016/j.jssas.2020.03.003>
- Torkayesh, A. E., Malmir, B., & Asadabadi, M. R. (2021). Sustainable waste disposal technology selection: The stratified best-worst multi-criteria decision-making method. *Waste Management*, *122*, 100-112. <https://doi.org/10.1016/j.wasman.2020.12.040>
- Vasu, D., Srivastava, R., Patil, N. G., Tiwary, P., Chandran, P., & Singh, S. K. (2018). A comparative assessment of land suitability evaluation methods for agricultural land use planning at village level. *Land Use Policy*, *79*, 146-163. <https://doi.org/10.1016/j.landusepol.2018.08.007>
- Veloutsou, C., & Mafe, C. R. (2020). Brands as relationship builders in the virtual world: A bibliometric analysis. *Electronic Commerce Research and Applications*, *39*, 100901. <https://doi.org/10.1016/j.elerap.2019.100901>

- Wigati, S. S., Sopha, B. M., Sri Asih, A. M., & Sutanta, H. (2019). Bibliometric Analysis for Site Selection Problems Using Geographic Information Systems, Multi-Criteria Decision Analysis and Fuzzy Method. *Journal of Physics: Conference Series*, 1351, 012051. <https://doi.org/10.1088/1742-6596/1351/1/012051>
- Wijesinghe, W. M. D. C., & Fernando, P. W. S. (2023). Optimal Urban Waste Dumping Site Selection in Kalutara DS Division of Sri Lanka using GIS-based Multi-Criteria Decision Analysis. *Journal of Asian Geography*, 2(2), 45-58. <https://doi.org/10.36777/jag2023.2.2.7>
- Worqlul, A. W., Jeong, J., Dile, Y. T., Osorio, J., Schmitter, P., Gerik, T., Srinivasan, R., & Clark, N. (2017). Assessing potential land suitable for surface irrigation using groundwater in Ethiopia. *Applied Geography*, 85, 1-13. <https://doi.org/10.1016/j.apgeog.2017.05.010>
- Yatsko, V. A. (2012). The interpretation of Bradford's law in terms of geometric progression, *Automatic Documentation and Mathematical Linguistics*, 46(2), 112-117. <https://doi.org/10.3103/S0005105512020094>
- Zabihi, H., Ahmad, A., Vogeler, I., Said, M. N., Golmohammadi, M., Golein, B., & Nilashi, M. (2015). Land suitability procedure for sustainable citrus planning using the application of the analytical network process approach and GIS. *Computers and Electronics in Agriculture*, 117, 114-126. <https://doi.org/10.1016/j.compag.2015.07.014>
- Zhang, J., Su, Y., Wu, J., & Liang, H. (2015). GIS-based land suitability assessment for tobacco production using AHP and fuzzy set in Shandong province of China. *Computers and Electronics in Agriculture*, 114, 202-211. <http://doi.org/10.1016/j.compag.2015.04.004>
- Zupic, I., & Čater, T. (2014). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18(3), 429-472. <https://doi.org/10.1177/1094428114562629>



Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

<http://dergipark.org.tr/guj.1531997>

Comparative Performance Analysis of Selected Machine Learning Algorithms and the Stacking Ensemble Method for Prediction of the Type II Diabetes Disease

Nathan Ayuba ZOAKAH^{1*}  Augustine Shey NSANG¹  Abel Adeyemi AJIBESIN¹  Ayuba Ibrahim ZOAKAH² 

¹ School of Information Technology and Computing, American University of Nigeria, Yola

² Department of Community Medicine, University of Jos, Nigeria

Keywords	Abstract
Machine Learning Diabetes Support Vector Machine Stacked Ensemble Method	<i>Diabetes</i> is a prevalent non-communicable disease affecting many people globally. The common risk factors are obesity, age, lack of exercise, lifestyle, genetic factors, high blood pressure, and poor diet. Early identification of this condition can help prevent subsequent complications, including heart attacks, lower limb amputations, nerve damage, and blindness. Data mining and machine learning have become popular and successful methods of identifying numerous diseases, including Diabetes, using clinical data over the years. This study focuses on the principles and processes of Naïve Bayes, Support Vector Machines, Logistic Regression, Decision Tree, and Random Forest algorithms for diabetes prediction, using the Scikit-learn inbuilt libraries for the experiments. Furthermore, we ensemble all five machine learning models to produce a single stacked ensemble model. Data preprocessing techniques such as scaling, missing data removal, dimensionality reduction, and balancing of target class were performed on the Jos Urban Diabetes dataset used for this study. The comparison of the algorithms' performances across various evaluation metrics, demonstrates that the Support Vector Machines algorithm outperform all others in terms of Accuracy , Precision , Sensitivity , and Matthew's Correlation Coefficient with scores of 96.11% , 91.61% , 85.67% , and 82.59% respectively with 10-fold cross-validation. Furthermore, the Stacked Ensemble Method model had the best Area Under the Receiver Operating Characteristic Curve scores of 98.47% with 10-fold cross-validation.

Cite

Zoakah, N. A., Nsang, A. S., Ajibesin, A. A., & Zoakah, A. I. (2024). Comparative Performance Analysis of Selected Machine Learning Algorithms and the Stacking Ensemble Method for Prediction of the Type II Diabetes Disease. *GU J Sci, Part A, 11(3)*, 622-646. doi:10.54287/guj.1531997

Author ID (ORCID Number)	Article Process
0009-0000-3873-8471	Nathan Ayuba ZOAKAH
0000-0002-6466-9032	Augustine Shey NSANG
0000-0001-6518-0231	Abel Adeyemi AJIBESIN
0000-0003-1856-7753	Ayuba Ibrahim ZOAKAH
	Submission Date 12.08.2024
	Revision Date 09.09.2024
	Accepted Date 16.09.2024
	Published Date 30.09.2024

1. INTRODUCTION

Diabetes mellitus is a chronic disorder that arises when the body does not produce enough insulin or utilize the insulin produced adequately (IDF, 2021). It is a disorder that develops when the blood sugar or glucose in the blood reaches abnormal high levels (hyperglycemia) due to its inability to reach other cells in the body (Birjais et al., 2019). Diabetes is of two main types: Type I and Type II. Other kinds of Diabetes include monogenic Diabetes (formerly known as secondary Diabetes), gestational Diabetes, and other unclassified Diabetes.

Type I Diabetes occurs when the endocrine gland in the pancreas produces little or no insulin required to regulate the blood sugar level. It is caused by an autoimmune reaction in which the body's immune system affects the pancreatic insulin-producing beta-cells. This type of Diabetes is most common in children and adolescents (Choudhury & Gupta, 2019). Although the initial causes of this type of Diabetes are unknown, research shows that the cause of this autoimmune reaction may be the combination of environmental triggers like viral infections and genetic susceptibility (IDF, 2021).

*Corresponding Author, e-mail: nathanzoakah@gmail.com

Persons with type I Diabetes will require a daily dosage of insulin injections, regular glucose level monitoring, and support to live healthy lives and avert further complications. Some symptoms of type I Diabetes are constant hunger, feelings of fatigue, blurred vision, and diabetic ketoacidosis (DKA).

Type II diabetes, also known as "insulin resistance," occurs when the endocrine gland in the pancreas produces insulin that the body cannot adequately use to regulate the blood sugar level. Ninety percent (90%) of Diabetes worldwide is said to be type II diabetes making it the most common type of Diabetes (IDF, 2021). The reasons for type II diabetes are unknown; however, there is a clear association between being overweight or obese, becoming older, ethnicity, and having a family history. Although they have similar symptoms with type I diabetes, they tend to be less dramatic and sometimes show no symptoms. However, some benchmarks would enable someone with type II diabetes to watch for leading long, healthy lives and avert long-term complications. These include a healthy diet, physical activities, maintaining body weight, smoking cessation, regular checkups, and taking prescribed medication by medical personnel.

Blood glucose levels higher than usual but not yet high enough to be identified as type II diabetes are **prediabetes** (Punthakee et al., 2018).

Diabetes is a major public health problem. According to World Health Organization (WHO, 2021), around 1.5 million people worldwide died directly from Diabetes in 2019, making it the ninth-highest cause of death. Furthermore, according to the International Diabetes Federation (IDF), Diabetes affected 415 million people in 2015, which is anticipated to climb to 642 million by 2040 (Zimmet et al., 2016). In a recent report by the IDF (2021), Diabetes is shown to be one of the fastest-growing emergencies of the 21st century, as revealed in the following statistics. About 537 million people had Diabetes in 2021; the projected numbers for 2030 and 2045 are 643 million and 783 million, respectively. In 2021 alone, more than 1.2 million children and adolescents had type I diabetes.

Diabetes can cause a variety of dangerous long-term complications, including cardiovascular diseases, stroke, renal failure or eye damage (retinopathy), heart attack, lower limb amputation, kidney damage (nephropathy), peripheral artery disease, blood vessels, and nerve damage (neuropathy) (Tigga & Garg, 2020). However, Diabetes and all its associated problems can be significantly reduced or prevented if it is detected, treated early, and appropriately managed.

Machine learning and Data Mining techniques have been used in the medical domain as very reliable tools for predicting Diabetes from clinical data. Data mining is extracting information from data and uncovering numerous patterns inherent in the data that are accurate, novel, and beneficial (Bhatia, 2019). This process helps to uncover hidden trends in a vast amount of data to support decision-making. On the other hand, Machine Learning is a branch of Artificial Intelligence that enables computers to learn from data without being specifically programmed to do so (Bhatia, 2019). It uses various algorithms to make predictions from the data prepared through data mining processes with little or no human intervention. Machine learning aims to create a computer program that can access data and utilize it to learn (Ibrahim & Abdulazeez, 2021).

The usage of data mining has accelerated in the Big Data era. With their power and automation, data mining technologies can handle massive volumes of data and extract value (Shmueli et al., 2019). Health practitioners are progressively moving health and healthcare data from traditional to digital formats, and as a result, healthcare institutions are creating vast quantities of data (Armstrong, 2022). However, as with many real-world data, they are prone to inconsistencies and errors such as missing or noisy data. Data preprocessing is a preliminary step in the data mining and machine learning process, which helps to eliminate or reduce such inconsistencies in data. In addition, data quality is a crucial consideration in the data mining process for disease prediction and diagnosis since poor data quality might lead to erroneous or low prediction results during machine learning (Zhu et al., 2019).

With the advent of E-health records and databases and the massive quantity of data collected from hospitals and medical records, early identification of Diabetes is achievable through predictive analysis. This can be done through a physician's knowledge and expertise with the illness; nevertheless, such work is prone to mistakes and inaccuracies if performed manually, depriving the patient of adequate therapy (Khanam & Foo,

2021). Automating this process using machine learning and data mining can reveal hidden patterns in the data, allowing for better decision-making.

Therefore, the purpose of this research is to understand the basic principles of the selected ML algorithms, implement the algorithms on the diabetes dataset and, evaluating the algorithms performances while following best practices and important data mining and machine learning steps to achieve the expected results.

2. MATERIAL AND METHOD

In this research, we employed an ensemble approach and supervised machine learning techniques to evaluate the performance of five algorithms. In supervised learning, algorithms are trained on labeled data to generate predictions based on input data. This method learns a mapping between the input and output data, using pairs from a labeled dataset to understand this relationship. To accurately predict new, unknown data, the algorithm attempts to comprehend the link between the input and output data. Techniques such as regression and classification are incorporated.

For this study, binary classification was utilized to categorize the diabetes dataset into diabetic and non-diabetic (control) groups. The methodology comprised six major steps:

1. Data Collection: Gathering relevant data for analysis.
2. Exploratory Data Analysis: Examining the data to uncover patterns and insights.
3. Data Processing: Preprocessing the data, including handling missing values, scaling, and dimensionality reduction.
4. Machine Learning: Training the algorithms on the processed data.
5. Ensemble Learning: Combining multiple models to improve performance.
6. Performance Evaluation: Assessing the accuracy and effectiveness of the models.

Each of these steps is crucial for ensuring the robustness and reliability of the machine learning models.

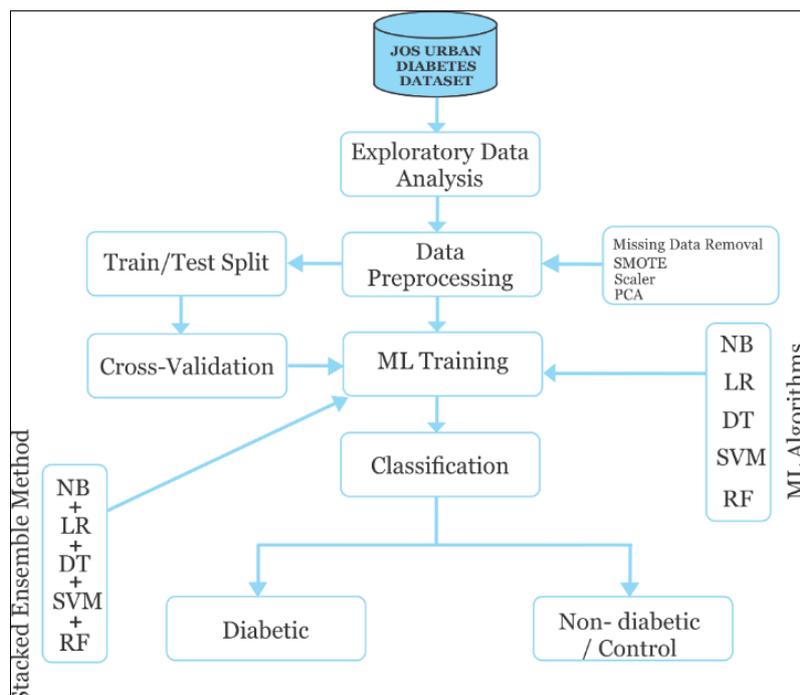


Figure 1. Overview of the General Methodological Workflow

Data Collection Phase

The Apex Clinic and Diabetes Screening Centre in Jos, Plateau State, Nigeria, owns the dataset utilized in this study. This data was gathered and recorded during routine screenings of individuals visiting the diabetes clinic mentioned above. Age, Fasting Plasma Glucose, Diastolic blood pressure, Systolic blood pressure, Weight, Height, Body Mass Index, Waist Circumference, Hip Circumference, and Gender were all recorded. The features of the diagnostic indicators or variables in the Jos Urban Diabetes dataset are shown in Table 1. This dataset comprises 753 records from various people gathered from the aforementioned diabetic clinic. For each attribute: (numeric-valued and strings) containing 753 observations.

Table 1. Description of the Jos Urban Diabetes Dataset

S/N	Attribute Name / Measurement	Attribute Data type	Range
1	Age (years)	Integer	10 – 85
2	Fasting Plasma Glucose (mg/ dl)	Integer	59 – 495
3	Diastolic blood pressure (mm Hg)	Integer	89 – 217
4	Systolic blood pressure (mm Hg)	Integer	26 – 148
5	Weight (Kilogram)	Integer	28 – 124
6	Height (Meters)	Float	1.30 – 1.96
7	Height Squared	Float	1.69 – 3.84
8	Body mass index (weight in kg / (height in m) ²)	Float	14.90 – 49.90
9	Waist Circumference (Centimeters)	Float	21.00 – 156.00
10	Hip Circumference (Centimeters)	Float	61.00 – 143.00
11	Waist Hip Ratio	Object or String	0.24 – 1.24
12	Gender	Object or String	Male and Female
13	Diagnosis Class Distribution: (class value 1 is interpreted as "tested positive for diabetes" and class value 0 is interpreted as "tested negative")	Integer	Class variable (0 or 1)

Exploratory Data Analysis Phase

The primary goal of this phase is to comprehend the dataset utilized for this study. Various graphical and non-graphical visualization techniques were used to identify missing values, features, the correlation between features, target variables, the data type of the variables, the shape of the dataset, and other statistical attributes such as mean and standard deviation.

Data Preprocessing Phase

In this phase, various techniques were used to handle inconsistencies in the dataset identified in the previous phase. The data was scaled and transformed; missing data values were removed and the imbalanced target variable was balanced using a resampling technique. The dataset was split into two (training and testing sets). After that, the data was ready to be sent to the next phase.

Machine Learning Phase

During this phase, five machine learning algorithms were implemented. The training dataset used Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest algorithms to produce a

prediction model or classifier. All of the algorithms stated above used the inbuilt sci-kit learn ML algorithms in this study.

Ensemble Learning Phase

All five algorithms used in this article were combined in this phase to generate a single stacked model. To learn from each of the models built by the different algorithms employed, a final estimator or classifier is used. The final estimator utilized for the ensemble approach to build the stacked model was the Logistic Regression algorithm from the sci-kit learn module.

Performance Evaluation Phase

The testing dataset was used to evaluate the performance of the different machine learning algorithms or classifiers in this phase of the proposed methodology. The accuracy, precision, sensitivity, MCC, and ROC area under the curve were then used to compare the findings.

Figure 2 provides a detailed description of the methodological workflow, beginning with data collection and exploratory data analysis. The first step in data preprocessing was the removal of missing data, followed by splitting the dataset into training and testing sets. Additional preprocessing steps included class balancing using SMOTE, scaling, and dimensionality reduction via PCA for the feature variables and target variable of the training dataset. For the testing dataset, only scaling and dimensionality reduction were performed.

Furthermore, 10-fold cross-validation was conducted on both splits of the dataset. The training dataset was utilized for the learning process with all five machine learning algorithms, including hyperparameter tuning and the stacked ensemble method. The trained classifiers or models were then applied to the testing dataset to make classifications and measure the performance of each algorithm. Finally, the best classifiers or models were saved.

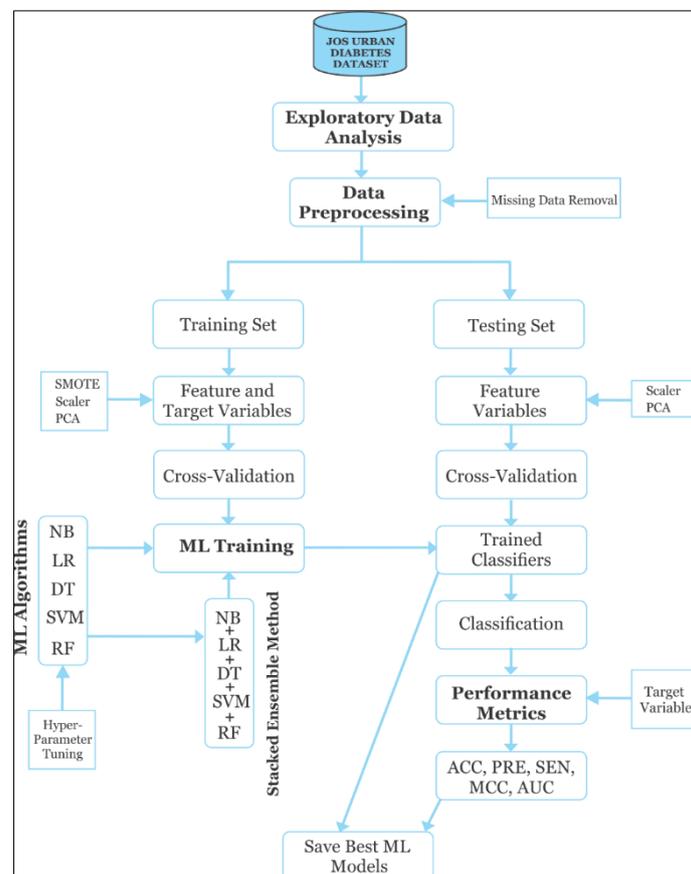


Figure 2. Detailed Methodological Workflow Diagram

SELECTED MACHINE LEARNING ALGORITHMS

The five selected supervised learning algorithms are discussed in this section. Some of the mathematical and statistical foundations as well as how these algorithms learn from data are also briefly explained.

Naïve Bayes Algorithm

The Naïve Bayes algorithm is a classification method that employs Bayes' theorem under the strong assumption that all variables or predictors are independent of one another (Prasanna, 2019). It indicates that features inside a class are unrelated to other features within the same class. Another strong assumption with this algorithm is that all the variables or predictors have an equal effect on the target variable (Gandhi, 2018). It can be used for either binary or multiclass classification problems. When using a Bayesian classification, our main aim is to find the probability of a label (L) given some observed features (F), also known as the posterior probability $P(L|F)$. It can be expressed using Bayes' Theorem as shown in the formula below:

$$P(L|F) = \frac{P(L) P(F|L)}{P(F)} \quad (1)$$

Where;

$P(L|F)$ is the posterior probability

$P(L)$ is the class prior probability

$P(F|L)$ is the class conditional probability or likelihood of a predictor given a class

$P(F)$ is the predictor of prior probability or evidence

The feature vector F can be written as

$$F = (f_1, f_2, f_3, \dots, f_n)$$

where $f_1, f_2, f_3, \dots, f_n$ correspond to the feature variables in a dataset.

Based on the assumption that all the features are mutually independent, we substitute F in equation (1) and expand using Chain's rule. The resulting equation is shown in (2):

$$P(L|f_1, \dots, f_n) = \frac{P(f_1|L) * P(f_2|L) * \dots * P(f_n|L) * P(L)}{P(f_1) * P(f_2) * \dots * P(f_n)} \quad (2)$$

Equation (3) is used to find the class with the highest probability and demonstrates how the label (L) value is calculated given the features.

$$L = \operatorname{argmax}_L P(f_1|L) * P(f_2|L) * \dots * P(f_n|L) * P(L) \quad (3)$$

Because the class probability for each feature is between the range 0 to 1, the values, when multiplied together, give minimal values, which can lead to overflow problems. The solution is to apply a Log function to each class probability ($P(f_i|L)$), and according to the logarithm law, the multiplication signs change to addition, as seen in equation (4).

$$L = \operatorname{argmax}_L \log P(f_1|L) + \log P(f_2|L) + \dots + \log P(f_n|L) + \log P(L) \quad (4)$$

The prior probability $P(L)$ is the frequency of the class in the training sample; that is,

$$P(L) = \frac{\text{Number of samples with class label } L}{\text{Total number of samples}}$$

To calculate the class conditional probability, it is modeled with the Gaussian distribution formula, as seen in equation (5).

$$P(f_i | L) = \frac{1}{\sqrt{2\pi\sigma_i^2}} * e^{-\left(\frac{(f_i - \mu_i)^2}{2\sigma_i^2}\right)} \quad (5)$$

where:

f_i are the feature variables

μ is the mean value of all feature variables in a given class

σ is the variance of all feature variables in a given class

π is a constant value of 3.142

e is a constant value of 2.7183

Finally, in equation (4), the class conditional probability and prior probability results are substituted to obtain the posterior probability for each class. The argmax method finds and returns the class with the highest probability (Loeber, 2019a).

Decision Tree Algorithm

A decision tree is generally a binary tree that recursively splits a dataset until we are left with only pure leaf nodes, that is, data with only one type of class (homogeneous class) (Normalized Nerd, 2021). The decision tree comprises two entities, the decision node (parent) and the leaf node (child). Decision nodes contain a condition to split data into leaf nodes, while leaf nodes are used to decide the class of a new data point. The trees choose the decision node based on a statistical calculation called information gain, where the information gain of a node is measured by the Entropy or Gini Index (Pranto et al., 2020).

To calculate the entropy (E), also known as the measure of uncertainty, we use the formula below.

$$E = - \sum_{i=1}^n p(X_i) * \text{Log}_2 p((X_i)) \quad (6)$$

where $p(X) = \frac{\text{Frequency of the class labels}}{\text{Total number of samples}}$

After computing the entropy of the parent node, the data is split into child nodes, and the entropy of the child nodes is computed to determine how much information was acquired after splitting. This measure is known as the information gain (IG) and is calculated using the formula below.

$$IG = E(\text{parent}) - [\text{weighted average}] * E(\text{children}) \quad (7)$$

Maniruzzaman et al. (2020) gives a brief description of the steps taken to build a decision tree:

1. Build a tree with its nodes as input features;
2. Choose the feature that delivers the most significant information gain when predicting the output from the input features to forecast the output;

- Repeat the previous stages to create subtrees based on characteristics not utilized in the previous nodes.

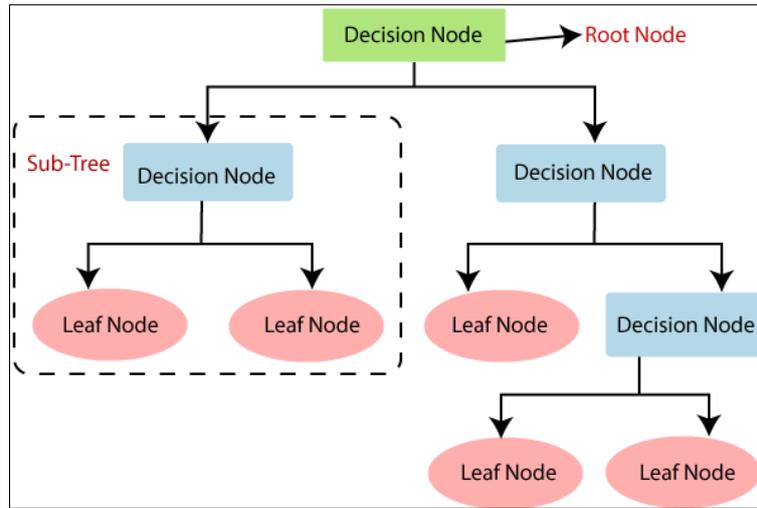


Figure 1. Graphical Representation of Decision Tree.

Loeber (2019b) gives a more extended explanation of the procedures necessary for creating the decision tree algorithm. He starts with the steps to construct the tree when training the algorithm and continues with the actions required to predict a new data point when traversing the constructed tree.

Training Phase

- Begin at the top node and pick the optimal split based on the best information gained at each node.
- Loop through all features and thresholds (all possible feature values)
- At each node, save the best-split feature and split the threshold.
- Recursively generate the tree.
- To stop the tree from growing, apply certain stopping conditions (max depth, min sample at node, or no more distribution in node (homogeneous class))
- When we have a leaf node, save its most common label.

Testing Phase (Prediction)

- Recursively navigate the tree.
- Look at the best-split feature of the test feature vector F at each node and move left or right based on whether F [feature index] equals the threshold.

When we reach the leaf node, we return the most common class label that was previously saved.

Random Forest Algorithm

Random forest is one of the most well-known and powerful supervised machine learning techniques that is based on ensemble learning, where several classifiers are combined to tackle complicated problems and enhance model accuracy. It is a classifier that uses several decision trees on distinct subsets of a given dataset and averages their results to improve the projected accuracy of that dataset. The new dataset or testing data is disseminated to all newly generated subtrees in the random forest. Each decision subtree in the forest is free to choose the dataset's class (Ibrahim & Abdulazeez, 2021). The model will then identify the best-suited class based on majority voting. Random Forest can be applied in various biomedical studies, particularly in diagnosing diabetes (Maniruzzaman et al., 2020). The following are simplified stages demonstrating how the

random forest algorithm works (Sruthi, 2021). Figure 4 shows the graphical illustration of the random forest algorithm.

1. In a Random Forest, n records are chosen randomly from a data collection of k records.
2. An independent decision tree is constructed for each sample (The same procedure is used in the decision tree algorithm above).
3. Each decision tree will provide a result.
4. Majority Voting determines the resolution of classification problems.

Bootstrapping is a random sampling procedure or approach used in the random forest algorithm in which the algorithm is trained on only a portion of the observations rather than all. The subtrees' outputs are aggregated, referred to as bagging when combined with the bootstrapping procedure (Choudhary, 2021).

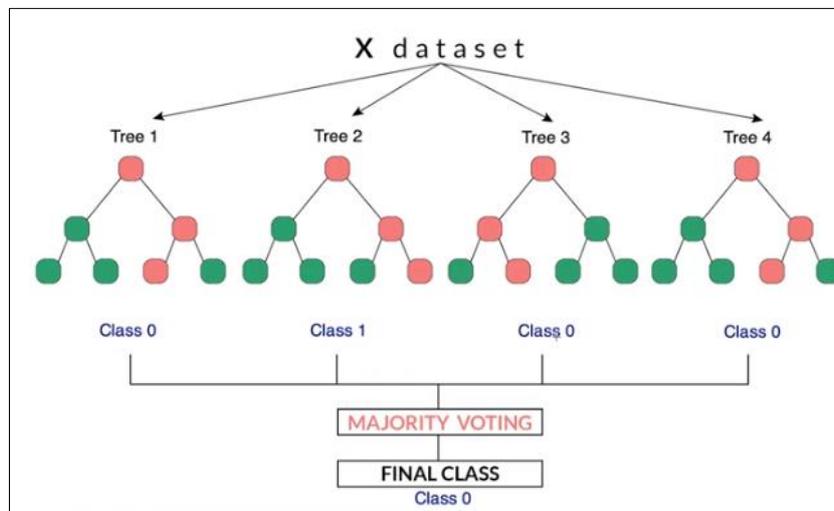


Figure 2. Graphical Representation of Random Forest Algorithm (Loeber, 2019c)

Support Vector Machine Algorithm

SVM is a supervised machine modeling technique that is widely used in classification (Sisodia & Sisodia, 2018). It is an algorithm that employs the hypothesis space of linear functions in a high-dimensional feature space and is taught with a learning algorithm from optimization theory that applies a learning bias derived from statistical learning theory (Jakkula, 2006). SVM aims to use an appropriate hyperplane (linear decision boundary) to classify data points in a multidimensional space. A *hyperplane* is a decision boundary used to categorize data points. A hyperplane far away from the data points in each category should be chosen; the farther our data points are from the hyperplane, the more confident we are that they have been correctly recognized. The support vectors are the spots closest to the classifier's margin (Tigga & Garg, 2020). The following explanation is limited to linear separable SVM models with linear kernels. Nonlinear SVM models include kernels such as polynomial and radial basis function kernels.

The linear model is represented in Equation (8)

$$W * X - b = 0 \quad (8)$$

The function should satisfy the following criteria to determine which class a data point belongs to, as shown in Figure 5 (positive red class or negative blue class) and as illustrated in the following equations (9) and (10) below.

$$W * X_i - b \geq 1 \quad \text{if } y_i = 1 \quad (9)$$

$$W * X_i - b \leq -1 \quad \text{if } y_i = -1 \quad (10)$$

where:

\mathbf{W} = weight vector

\mathbf{X}_i = feature label

\mathbf{b} = bias

y_i = class labels

So, multiplying our class label with our linear model gives us a single equation representing the condition to be satisfied, as shown in equation (11).

$$y_i (\mathbf{W} * \mathbf{X}_i - \mathbf{b}) \geq 1 \quad (11)$$

Furthermore, to maximize the margins between the two classes represented in Figure 5, we use the formula $\frac{2}{\|\mathbf{W}\|}$. It is equivalent to minimizing the distance using the formula $\frac{\|\mathbf{W}\|}{2}$ (Sisodia & Sisodia, 2018). Where $\|\mathbf{W}\|$ is the magnitude of the weight vector.

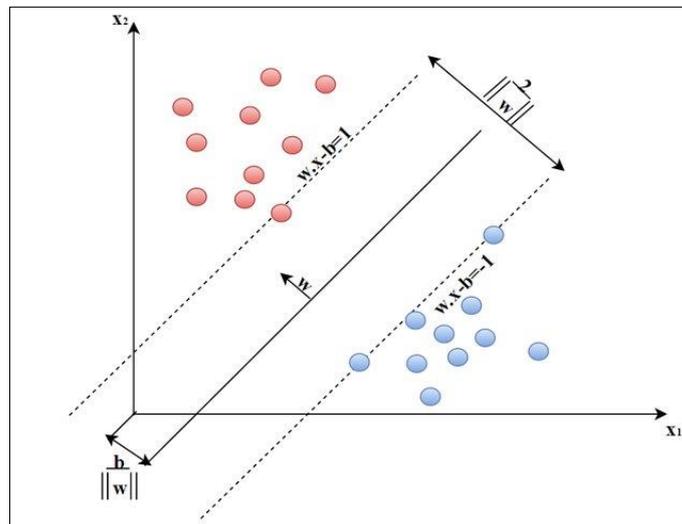


Figure 3. Support Vector Machine Classifying a Binary Task

Next, we describe the main objective function, which is made up of two parts: the one responsible for maximizing margins with an extra regularization parameter and the part responsible for establishing the separating hyperplane with a loss function termed Hinge loss (Lanhenke, 2022). Equation (12) shows the primary objective function.

$$\mathcal{J} = \lambda \|\mathbf{W}\|^2 + \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i (\mathbf{W} * \mathbf{X}_i - \mathbf{b})) \quad (12)$$

where:

$l = \max(0, 1 - y_i (\mathbf{W} * \mathbf{X}_i - \mathbf{b}))$ is the lost function called hinge loss.

and

$\lambda \|\mathbf{W}\|^2$ is the magnitude of the **squared Weight vector** multiplied by the regularization parameter **lambda**.

Both parts of the objective function work together to acquire the correct classification of our labels or classes by placing them on the correct side of the hyperplane and having a hyperplane with maximum margins.

The loss function is then optimized using gradient descent by identifying the derivatives for the weights (W) and biases (b). Based on the condition in equation (11), we divide the objective function in equation (12) into two classes.

If $y_i (W * X_i - b) \geq 1$

Then

$$J_i = \lambda ||W||^2 \quad (13)$$

Else

$$J_i = \lambda ||W||^2 + 1 - y_i (W * X_i - b) \quad (14)$$

To obtain the gradients, we apply the condition in equation (11) to the derivatives for both classes in equations (13) and (14), which is as follows:

If $y_i (W * X_i - b) \geq 1$

Then

$$\frac{dJ_i}{dW_k} = 2\lambda W_k \quad (15)$$

$$\frac{dJ_i}{db} = 0 \quad (16)$$

Else

$$\frac{dJ_i}{dW_k} = 2\lambda W_k - y_i * X_i \quad (17)$$

$$\frac{dJ_i}{db} = y_i \quad (18)$$

Finally, the weights and biases are updated with the formulas shown in equations (19) and (20):

$$W = W - \alpha . dw \quad (19)$$

$$b = b - \alpha . db \quad (20)$$

where:

W = Weight

b = bias

α = learning rate

dw = derivative of the weight and

db = derivate of the bias

Logistic Regression Algorithm

Logistic Regression is another supervised learning algorithm used for solving classification problems. It is a statistical model also known as a logit model which was borrowed into the machine learning field (Birjais et al., 2019). It is a regression model that predicts whether a given data item or entry (feature variables) falls within a certain class (target variable) (Ibrahim & Abdulazeez, 2021). This algorithm uses a Logistic function also known as a Sigmoid function to model a binary classification problem as in our case, predicting if a person is diabetic or not.

In Logistic Regression, we find the probabilistic values that lie between 0 and 1 by applying the sigmoid function in equation (22) to our linear model in equation (21) which models the probability of our data. Figure 6 shows the sigmoid function graph and Figure 7 shows the probabilistic values, threshold, and maximum values of 0 and 1.

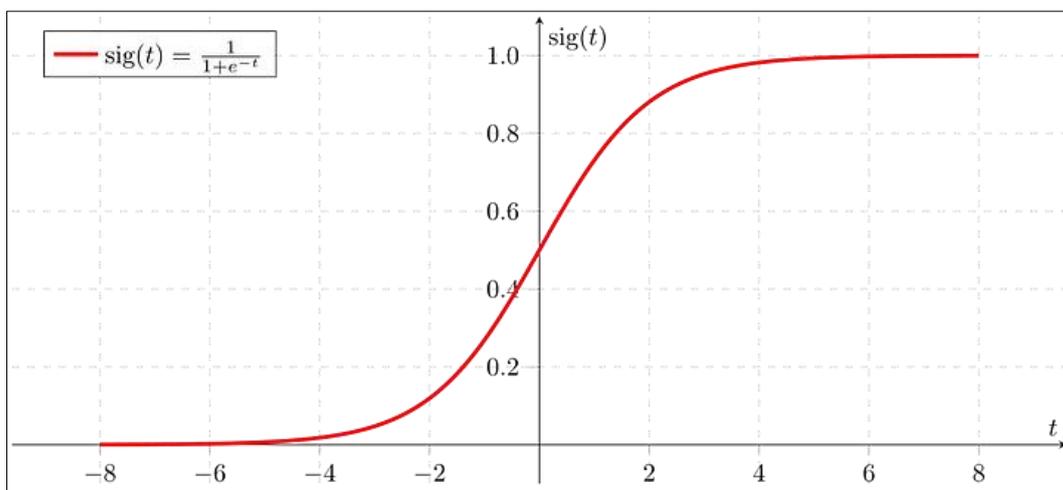


Figure 4. Sigmoid Function Graph

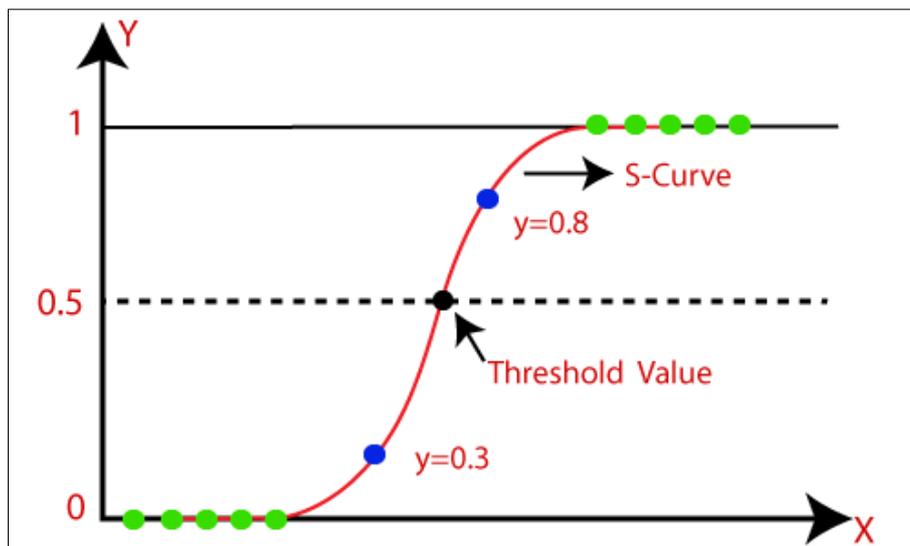


Figure 5. Probabilistic Values, Threshold, and Maximum Values on the Sigmoid Graph

The equations for the linear model and sigmoid function are shown below.

$$Z = W * X + b \quad (21)$$

$$S(x) = \frac{1}{1 + e^{-(x)}} \quad (22)$$

The sigmoid function is applied to the linear model to determine the approximation of our target variable y , producing equation (23).

$$\hat{y} = S(Z) = \frac{1}{1 + e^{-(W*X+b)}} \quad (23)$$

Gradient descent, similar to the Support Vector Machine Algorithm described above, is used to optimize the cost function in terms of weight (W) and bias (b). Cross Entropy is the cost function used in logistic regression, as indicated in equation (24) (Swaminathan, 2019).

$$J(\theta) = \frac{1}{N} \sum_{i=1}^n y_i \log(h_{\theta}(x_i)) + (1 - y_i) \log(1 - h_{\theta}(x_i)) \quad (24)$$

If $y_i = 1$, then the $(1 - y_i)$ term will become zero hence, $\log(h_{\theta}(x_i))$ alone will be left.

If $y_i = 0$, then the (y_i) term will become zero hence, $\log(1 - h_{\theta}(x_i))$ alone will be left.

Gradient descent begins at a certain position and iteratively updates the parameters (that is, weight and bias) by computing derivatives and advancing in the direction of the gradient (negative direction) depending on a specific learning rate (α).

The derivatives for the weight and bias are as follows;

$$\frac{dJ}{dw} = \frac{1}{N} \sum x_i (\hat{y} - y_i) \quad (25)$$

$$\frac{dJ}{db} = \frac{1}{N} \sum (\hat{y} - y_i) \quad (26)$$

Finally, the weights and biases are updated using the formulae shown in equations (27) and (28) :

$$W = W - \alpha \cdot dw \quad (27)$$

$$b = b - \alpha \cdot db \quad (28)$$

where:

W = Weight

b = bias

α = learning rate

dw = derivative of the weight and

db = derivative of the bias

Stacking Ensemble Method

This is a Stacked Generalization ensemble learning approach (Zhang & Ma, 2012). It is a strategy whose primary goal is to get the best generalization accuracy possible (Rokach, 2009). This is accomplished by combining predictions from many models to form a new model used to generate predictions on the test dataset (Singh, 2021). The diagram below depicts the general graphical illustration of stacking.

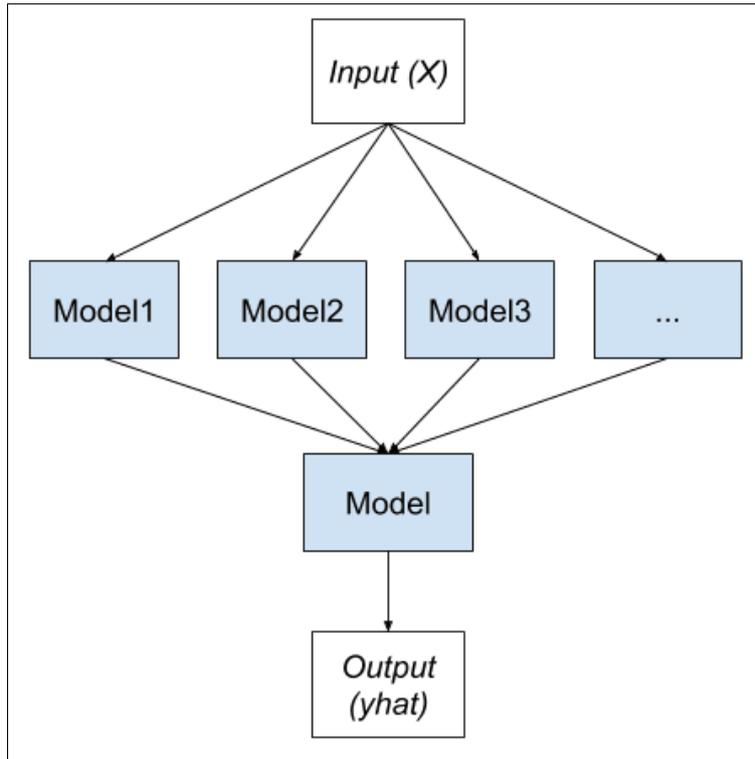


Figure 6. General Stacking Illustration

Harrison (2022) provided a more detailed understanding of the stacking ensemble learning approach in the stages below.

1. The dataset is divided into two parts: training and testing.
2. The stacking technique starts with individual model training, sometimes known as the level 0 model stage. Each model predicts on the training dataset, and the outcomes of each model prediction (target class) are added to the training dataset. This modifies the training dataset by adding new feature variables.

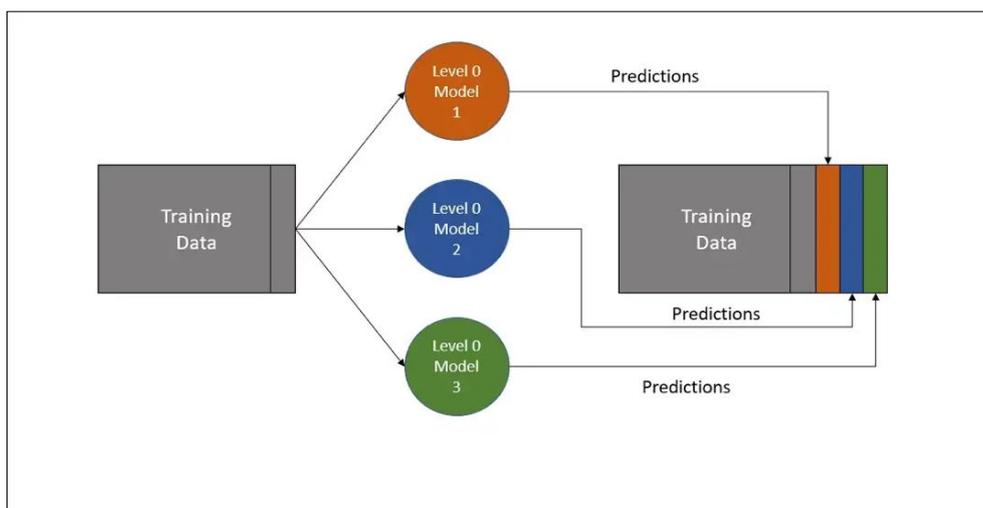


Figure 7. Level 0 Stacking Stage (Harrison, 2022).

3. The newly altered training dataset is then trained on another model, referred to as the level 1 model stage. This stage generates the trained stacking model, often known as the level 1 model.

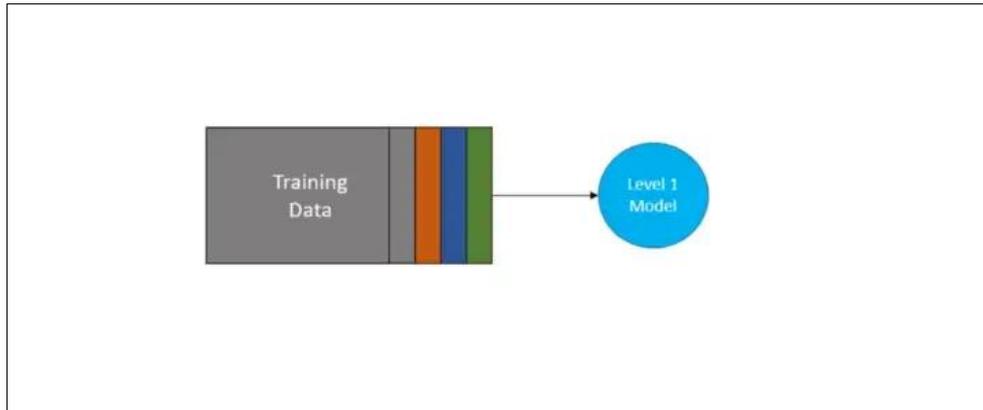


Figure 8. Level 1 Stacking Stage (Harrison, 2022).

- The last step makes final predictions using the level 1 model on the testing dataset. But first, the testing dataset must be changed to have the same shape as the modified training dataset. As a result, each level 0 trained model is applied to the testing dataset to generate new predictions, which are then added to the testing dataset, resulting in the modified testing dataset. Finally, the stacked model is utilized to produce the final prediction using the trained level 1 model on the modified testing dataset.

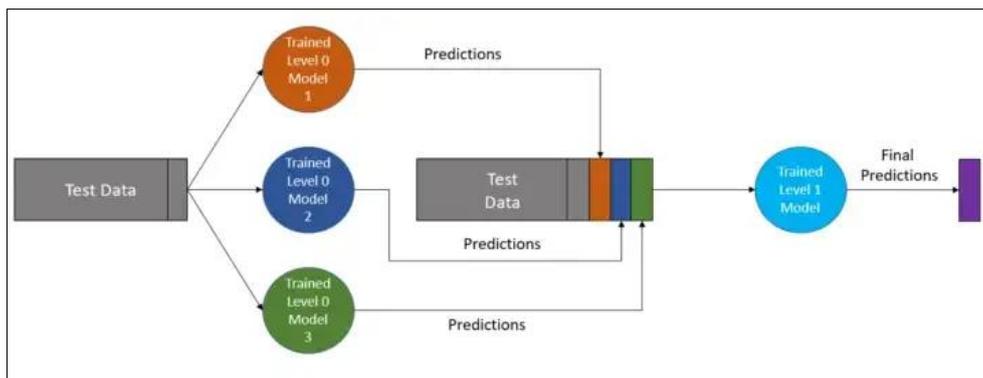


Figure 9. Final Stacking Stage (Harrison, 2022).

3. RESULTS AND DISCUSSION

The goal of this research was to implement machine learning (ML) algorithms to categorize the diabetes disease from a clinical dataset using the built-in scikit's learn algorithms. This presentation encompasses all the stages and procedures involved in understanding the data, preprocessing, training the algorithm to develop a model, and testing the model to evaluate classification outputs.

Exploratory Data Analysis Results

According to the EDA results, the dataset has fourteen attributes or columns and 753 records. Most of the data types are numerical—both floats and integers— but with two object or string data types that need to be converted to numerical data types to make them ideal for our algorithm to learn. Figures 12 and 13 give us further details of the percentage of missing values in the entire data. A total of 1.2% of data is missing from the dataset and the visualization shows that the majority of the data is not missing. However, the machine learning methods and Scikit Learn library used do not perform well when there are omissions in the data. Therefore, missing data may be replaced or removed out of the dataset.

The class or target variable distribution is shown in Figure 14 as well as the correlation between the feature variables and the target variable in Figure 15. It can be seen that there is a strong positive correlation between weight and body mass index, hip circumference, and waist circumference, as well as between the target variable and blood glucose level. However, there is also a strong negative correlation between height and the

target variable. These observations show us that we have an idea of the feature variables that are important for predicting the target variable. With 656 negative outcomes and 97 positive results, the target variable is likewise imbalanced. This simply indicates that because of the enormous disparity in the frequency of the target class, if we utilize this dataset in this way, our models will learn more from the majority class (negative outcomes) and less from the minority class (positive outcomes). Hence, the dataset must be balanced.

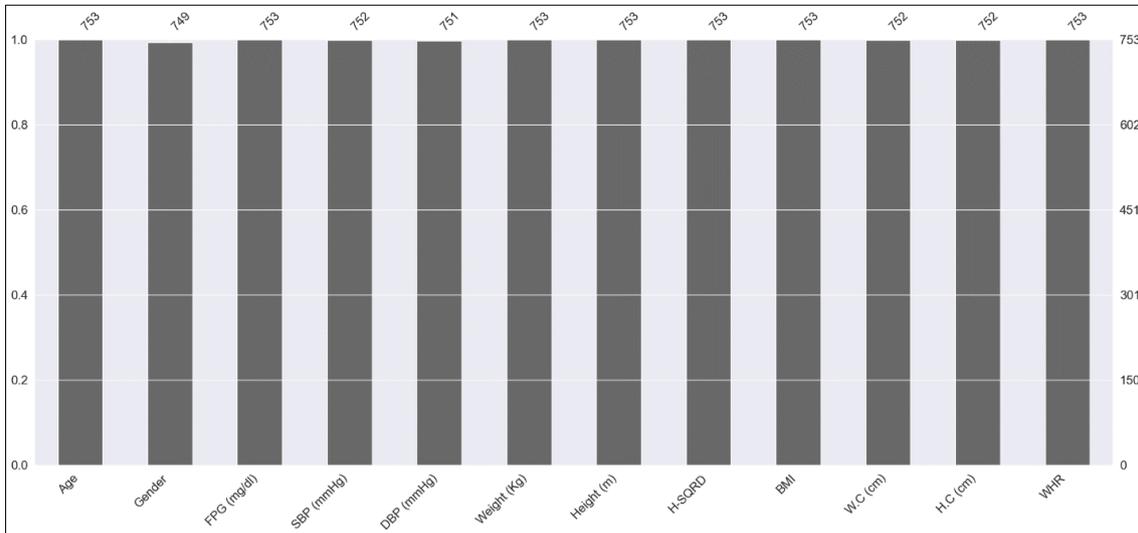


Figure 12. Visualization of Missing Data in the Dataset

```

Exploratory Data Analysis

Percentage of Missing Data

In [10]: percent_missing = diabetes_df.isnull().sum() * 100 / len(diabetes_df)

In [11]: percent_missing
Out[11]: Serial Number    0.000000
Age                    0.000000
Gender                 0.531288
FPG (mg/dl)           0.000000
SBP (mmHg)            0.132802
DBP (mmHg)            0.265604
Weight (Kg)           0.000000
Height (m)             0.000000
H-SQRD                 0.000000
BMI                    0.000000
W.C (cm)               0.132802
H.C (cm)               0.132802
WHR                    0.000000
diagnosis              0.000000
dtype: float64

In [12]: percent_missing.sum()
Out[12]: 1.195219123505976
    
```

Figure 13. Percentage of Missing Data in the Dataset

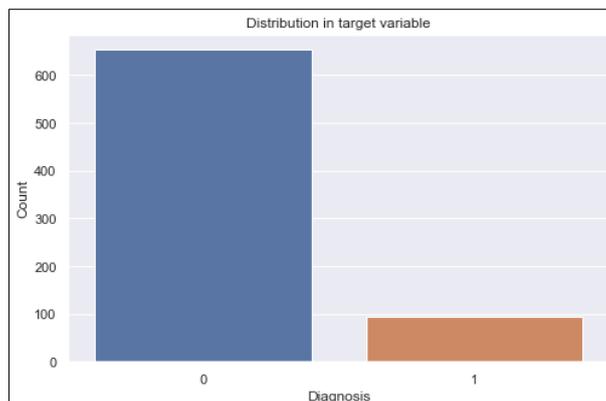


Figure 14. Class Distribution of Target Variable

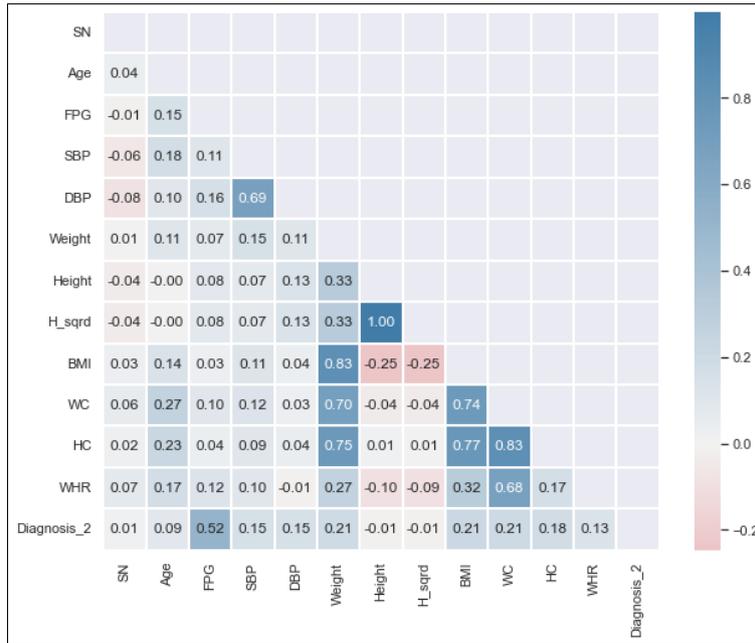


Figure 15. Correlation between feature and target variable in the dataset

Data Preprocessing Results

We go one step further in this phase to fix the discrepancies we found in our dataset. Firstly, after identifying that only 1.2% present of our data is missing (that is 7 records out of 753), the records with the discovered missing data were dropped out of the dataset. This was done using Python’s pandas dropna() method to remove all rows with missing data. The result of the dataset without missing values is shown in Figures 16 and 17.

Secondly, the dataset was balanced only on the training dataset after splitting the data into training and test sets. The SMOTE technique from the Imbalanced Learn Over Sampling library was used to balance our training dataset with both positive and negative classes having the same number of samples (588 each) as shown in Figure 18.

Thirdly, to standardize the data, feature scaling was also carried out on each feature variable from the training dataset using the standard scalar function of the scikit-learn preprocessing module. Finally, the most important features from the dataset were selected by using a recursive elimination technique from the Scikit Learn Feature Selection Library. After discovering these features, we used the PCA method from the Scikit Learn Decomposition Library, by specifying the number of important features earlier identified.

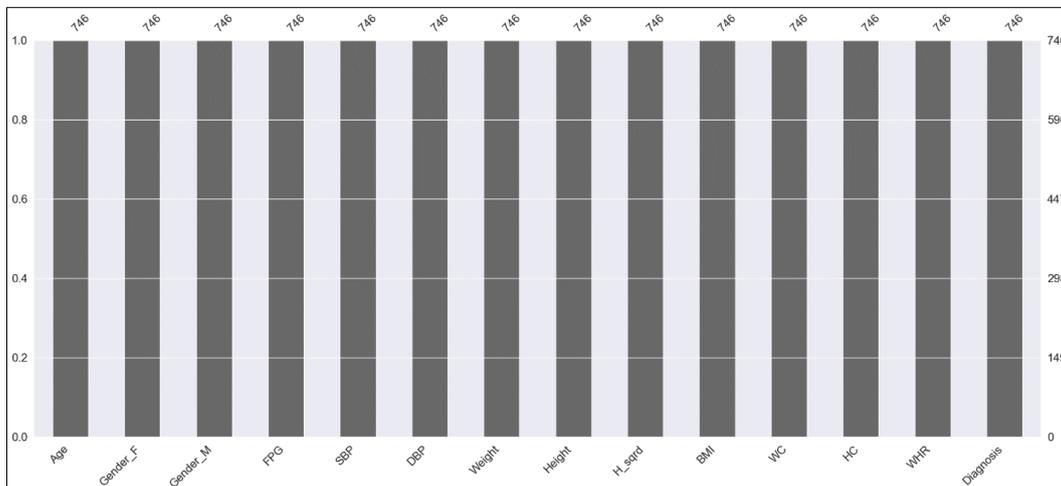


Figure 16. Visualization of Dataset after Removing Missing Values.

```
In [70]: data_copy.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 746 entries, 0 to 745
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Age         746 non-null    int64
1   Gender_F    746 non-null    uint8
2   Gender_M    746 non-null    uint8
3   FPG         746 non-null    int64
4   SBP         746 non-null    int64
5   DBP         746 non-null    int64
6   Weight      746 non-null    int64
7   Height      746 non-null    float64
8   H_sqrd     746 non-null    float64
9   BMI         746 non-null    float64
10  WC          746 non-null    int64
11  HC          746 non-null    int64
12  WHR         746 non-null    float64
13  Diagnosis_2 746 non-null    int64
dtypes: float64(4), int64(8), uint8(2)
memory usage: 71.5 KB
```

Figure 17. Result after Removing Missing Values

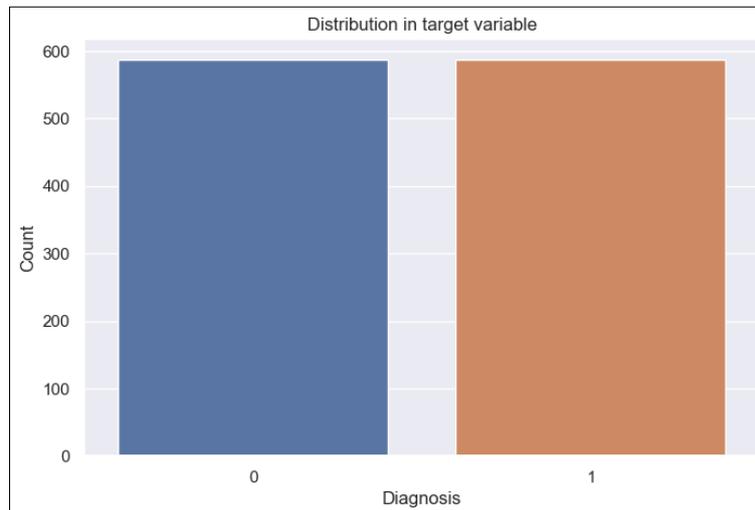


Figure 18. Result after using SMOTE to Balance the Target Class on the Training Dataset.

Setup for the Experiment

The essential experiments at different stages of our suggested technique were carried out in this study using various tools and frameworks. Conda, a powerful tool used to manage all other tools and frameworks, is the primary environment and package manager.

Python-based libraries utilized in the Conda environment include Pandas, Numpy, Matplotlib, and Scikit Learn. Jupyter notebook is an interactive web-based tool, also known as a computational notebook, that was used to build the machine learning algorithms and the libraries for data analysis, visualization, and machine learning stated above.

The Datacamp workspace, which includes an online Jupyter lab for data science and machine learning tasks, is another alternative and helpful web environment that was used for this research. The benefits of employing a cloud-based data science platform for machine learning tasks are demonstrated in this workspace. Many pre-installed modules are accessible for usage, and one may access his or her work remotely from any device without worrying about dependency difficulties.

After preprocessing our data, we performed several experiments utilizing all five machine-learning techniques. We divided our data into training and testing sets, using the Stratified K-fold Cross-validation technique from

the Scikit Learn Model Selection Library. This method was used to split our dataset and to also perform a 10 Cross-Fold validation on the dataset. In addition to using this method, we specified a random state value of 529 to allow us to reproduce the results repeatedly. Finally, we sent a Numpy array of our preprocessed data to our algorithm.

Confusion Matrix Result

We successfully built a model from our training data and then utilized our testing data to assess how well each of the algorithms in this research performed. All performance matrices were computed from the confusion matrix for each model.

Figure 19 and 20 displays the findings for the confusion matrix for the Support Vector Machine and Stacked Ensemble Model. Each of its characteristics has a value of $TP = 64$, $TN = 8$, $FP = 1$, and $FN = 1$. The performance assessment metrics that produced the greatest outcomes in this study were computed using the values that the Support Vector Machine and Stacked Ensemble Model generated.

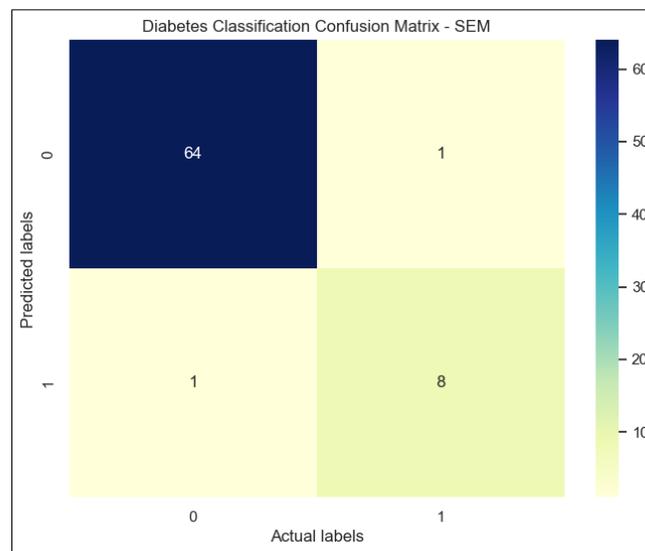


Figure 19. Confusion Matrix for Diabetes Classification with SEM

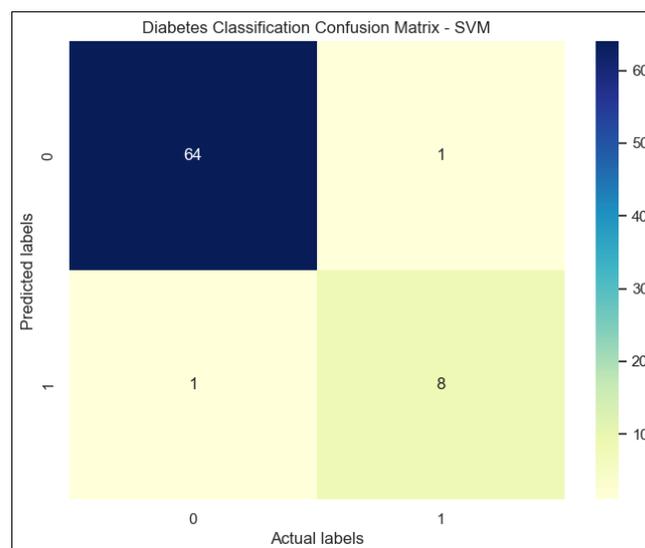


Figure 20. Confusion Matrix for Diabetes Classification with SVM

Evaluation Metrics Results

The values produced by the confusion matrices in Figures 19 and 20 were used to compute the output for the evaluation metrics in Figures 21, 22, 23, and 24. The outcome shows that, compared to all the other evaluation metrics employed in this study, the Accuracy and ROC AUC score assessment metrics had a more significant percentage. Furthermore, all results were computed using 10-fold cross-validation on all evaluation metrics. The final score for each measure is the mean of the total number of predictions. Cross-validation gives a more reliable result to check against overfitting. Finally, Figure 25 shows the graphical representation of the Stacked Ensemble Method model in a pipeline with Logistic Regression as the final estimator.

```

for train_idx, val_idx in sk.split(X, y):      Show code
Our accuracy on the validation set is 0.9733 and AUC is 0.9369
===== Fold 3 =====
Our accuracy on the validation set is 0.9733 and AUC is 0.9848
===== Fold 4 =====
Our accuracy on the validation set is 0.9733 and AUC is 0.9846
===== Fold 5 =====
Our accuracy on the validation set is 0.9467 and AUC is 0.9269
===== Fold 6 =====
Our accuracy on the validation set is 0.9600 and AUC is 0.9769
===== Fold 7 =====
Our accuracy on the validation set is 0.9595 and AUC is 0.8333
===== Fold 8 =====
Our accuracy on the validation set is 0.9189 and AUC is 0.7624
===== Fold 9 =====
Our accuracy on the validation set is 0.9730 and AUC is 0.8889
===== Fold 10 =====
Our accuracy on the validation set is 0.9730 and AUC is 0.9368
Our out of fold AUC score is 0.9161
Our out of fold ACC score is 0.9611
Our out of fold Precision score is 0.8463
Our out of fold Sensitivity score is 0.8567
Our out of fold MCC score is 0.8259

```

Figure 21. Performance Evaluation Metrics Results for SVM

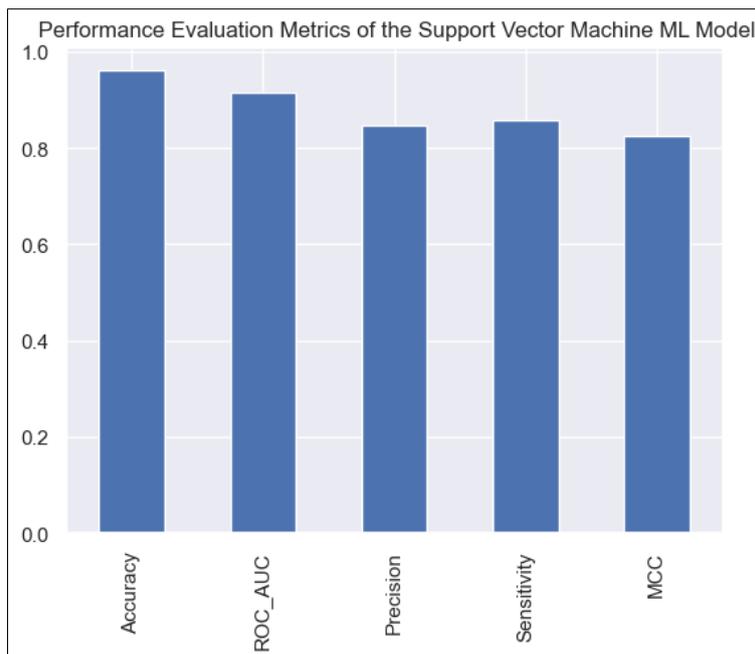


Figure 22. Comparing all Five Evaluation Metrics for the Support Vector Machine ML Model

```

for train_idx, val_idx in sk.split(X, y):
    Show code

===== Fold 4 =====
Our accuracy on the validation set is 0.9600 and AUC is 0.9938
===== Fold 5 =====
Our accuracy on the validation set is 0.9200 and AUC is 0.9600
===== Fold 6 =====
Our accuracy on the validation set is 1.0000 and AUC is 1.0000
===== Fold 7 =====
Our accuracy on the validation set is 0.9459 and AUC is 0.9795
===== Fold 8 =====
Our accuracy on the validation set is 0.9189 and AUC is 0.9470
===== Fold 9 =====
Our accuracy on the validation set is 0.9730 and AUC is 0.9932
===== Fold 10 =====
Our accuracy on the validation set is 0.9730 and AUC is 0.9966
Our out of fold AUC score is 0.9847
Our out of fold ACC score is 0.9571
Our out of fold Precision score is 0.8605
Our out of fold Sensitivity score is 0.7844
Our out of fold MCC score is 0.7944

```

Figure 23. Performance Evaluation Metrics Results for the Stacked Ensemble Model

Figure 24 graph shows that the area under the receiver operating characteristic curve has a value range significantly closer to 1 than 0.5.

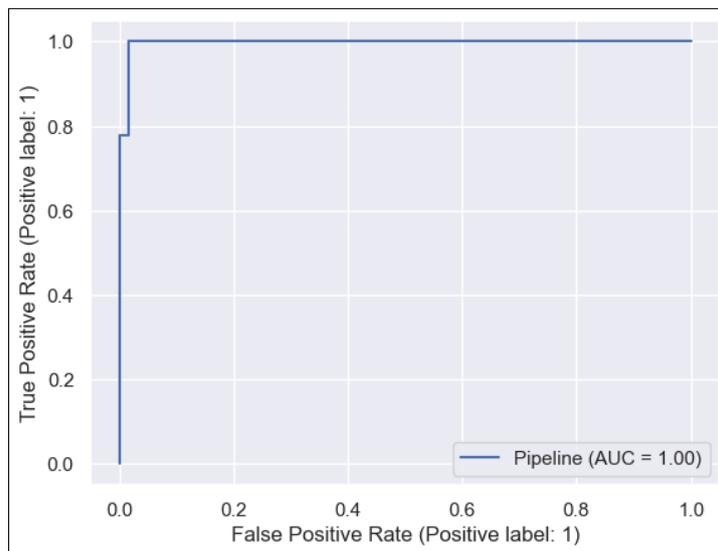


Figure 24. ROC_AUC Graph for our SEM Model

Comparison of the ML algorithms Employed in this Research

Table 2 compares the findings of the five evaluation metrics that were utilized in this study, and it shows that our Stacked Ensemble Method was the best model based on an ROC AUC score of **98.47 %**. However, SVM performed best in all other evaluation metrics with an Accuracy score of **96.11 %**, Precision score of **91.61%**, Sensitivity Score of **85.67%**, and Matthew's Correlation Coefficient score of **82.59 %**. The Stacked Ensemble Method model was the second best in all other evaluation metrics. Finally, Table 3 compares the findings of the combinations of different stacked models. The first combination was the stacking of the trained Support Vector Machine and Random Forest models (SEM_2). These were the two best algorithms from Table 2 without considering the stacked ensemble model column results reported. The next combination was the stacking of Support Vector Machine, Random Forest, and Logistic Regression models (SEM_3). Lastly, we stacked all five trained models (SEM_5) used in this study which produced the best results when compared to all other stacked ensemble methods employed.

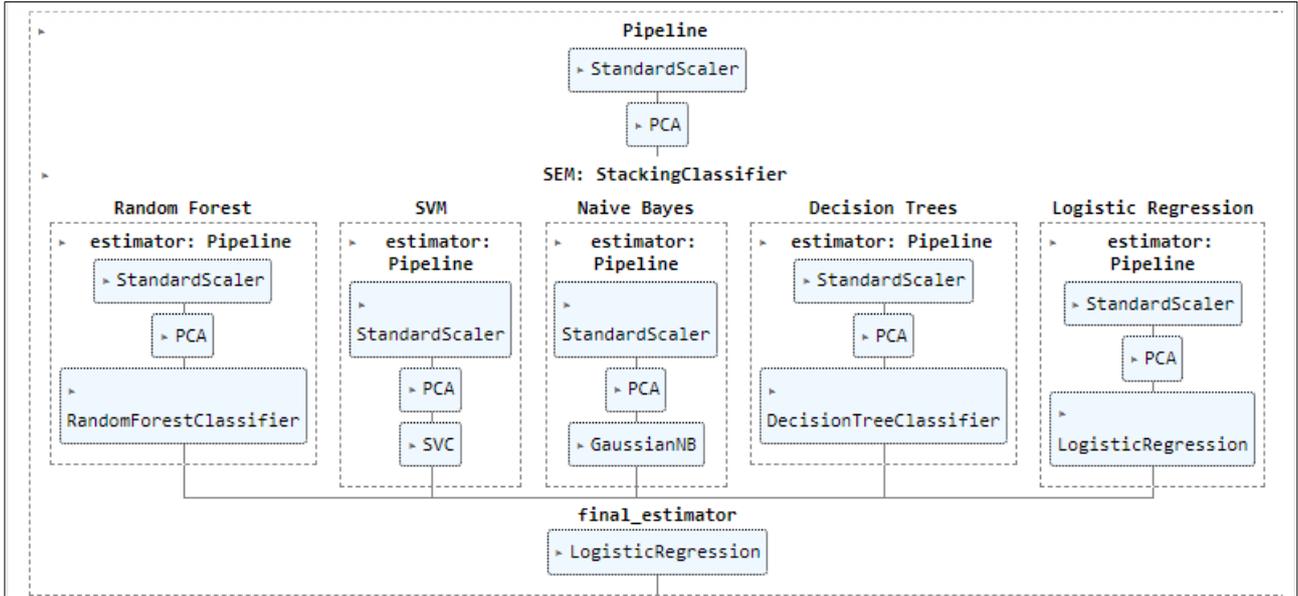


Figure 25. SEM Model in a Pipeline

Table 2. Comparative Performance Analysis of the Five (5) ML Algorithm used in this Study

Algorithm /Evaluation Metrics	Random Forest	Naïve Bayes	Logistic Regression	Support Vector Machine	Decision Tree	Stacked Ensemble Method
Accuracy	93.97 %	89.09 %	92.23 %	96.11 %	86.60 %	95.71 %
Precision	79.02 %	56.03 %	64.23 %	91.61 %	48.17 %	86.05 %
Sensitivity	70.44 %	67.44 %	84.78 %	85.67 %	66.11 %	78.44 %
MCC	70.66 %	54.83 %	69.51 %	82.59 %	48.84 %	79.44 %
AUC_ROC	96.06 %	89.86 %	96.02 %	91.53 %	85.41 %	98.47 %

Table 1. Comparative Analysis of the Three (3) Different Stacked Ensemble Method Combinations.

Stacked Models /Evaluation Metrics	Stacked Ensemble Method (SEM_2)	Stacked Ensemble Method (SEM_3)	Stacked Ensemble Method (SEM_5)
Accuracy	94.63 %	94.77 %	95.71 %
Precision	79.95 %	81.13 %	86.05 %
Sensitivity	75.00 %	76.22 %	78.44 %
MCC	74.16 %	75.23 %	79.44 %
AUC_ROC	97.68 %	97.86 %	98.47 %

4. CONCLUSION

To optimize the insights found to enhance better and more significant outcomes when used in the medical domain, more study is being conducted on the use of machine learning in detecting illnesses, notably diabetes. In this study, examining evaluation metrics other than accuracy alone and comparing them to other metrics yields excellent findings in terms of precision, sensitivity, and area under the receiver operating characteristic curve. These results were obtained by applying data preprocessing techniques since clinical data might be prone to missing data and discrepancies, in addition to understanding the working principles of the algorithms utilized.

Recommendation for Future Research

The classification approaches to predict diabetic disease produced acceptable accuracy, sensitivity, precision, MCC, and ROC AUC values. The following enhancements are recommended for this study's future research:

1. To enhance all measures employed in this study and utilize additional cutting-edge supervised machine learning and deep learning techniques.
2. A web-based or mobile application for diagnosing diabetic illness can use the best models as its backend.
3. To use clinical data with machine learning algorithms to determine the kind of diabetes a patient has (Type I, Type II, or Gestational Diabetes, etc.).

AUTHOR CONTRIBUTIONS

Conceptualization, Nathan Zoakah, Augustine Nsang and Abel Ajibesin; Methodology, Nathan Zoakah and Augustine Nsang; Data Source, Ayuba Zoakah; Software, Nathan Zoakah; Title, Augustine Nsang and Abel Ajibesin; Experiments: Nathan Zoakah, Augustine Nsang; Validation, Augustine Nsang, Ayuba Zoakah; Manuscript-original draft, Nathan Zoakah; Manuscript-review and editing, Abel Ajibesin, Ayuba Zoakah Augustine Nsang; Visualization, Nathan Zoakah; Supervision, Augustine Nsang, Ayuba Zoakah

ACKNOWLEDGEMENT

This study contains clinical data from The Apex Clinic and Diabetes Screening Centre in Jos, Plateau State, Nigeria. Their contribution to the success of this study is gratefully acknowledged.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Armstrong, A. (2022, March 1). Python in Healthcare: AI Applications in Hospitals. https://www.datacamp.com/blog/python-in-healthcare-ai-applications-in-hospitals?utm_medium=email&utm_source=customerio&utm_id=7430059&utm_campaign=dc_insights&utm_term=v2blog
- Bhatia, P. (2019). *Data mining and data warehousing: Principles and practical techniques*. Cambridge University Press.
- Birjais, R., Mourya, A. K., Chauhan, R., & Kaur, H. (2019). Prediction and diagnosis of future diabetes risk: A machine learning approach. *SN Applied Sciences*, 1(9), 1112. <https://doi.org/10.1007/s42452-019-1117-9>
- Choudhary, D. (2021, April 18). Bootstrapping and OOB samples in Random Forests. Analytics Vidhya. <https://medium.com/analytics-vidhya/bootstrapping-and-oob-samples-in-random-forests-6e083b6bc341>

- Choudhury, A., & Gupta, D. (2019). *A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques*. In: J. Kalita, V. E. Balas, S. Borah, & R. Pradhan (Eds.), *Recent Developments in Machine Learning and Data Analytics* (Vol. 740, pp. 67-78). Springer Singapore. https://doi.org/10.1007/978-981-13-1280-9_6
- Gandhi, R. (2018, May 17). Naive Bayes Classifier. Medium. <https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c>
- Harrison, G. (2022, February 28). A Deep Dive into Stacking Ensemble Machine Learning—Part I. Medium. <https://towardsdatascience.com/a-deep-dive-into-stacking-ensemble-machine-learning-part-i-10476b2ade3>
- Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10-19. <https://doi.org/10.38094/jastt20179>
- IDF (International Diabetes Federation) (2021). IDF Diabetes Atlas 10th ed.
- Jakkula, V. (2010) Tutorial on Support Vector Machine (SVM).
- Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. *ICT Express*, 7(4), 432-439. <https://doi.org/10.1016/j.icte.2021.02.004>
- Lanhenke, M. (2022, May 1). Implementing Support Vector Machine From Scratch. Medium. <https://towardsdatascience.com/implementing-svm-from-scratch-784e4ad0bc6a>
- Loeber, P. (2019a, September 29). Naive Bayes in Python—Machine Learning From Scratch 05—Python Tutorial—YouTube. <https://www.youtube.com/watch?v=BqUmKsfSWho&list=PLqnsIRFeH2Upcrywf-u2etjdxkL8nl7E&index=5>
- Loeber, P. (2019b, November 22). Decision Tree in Python Part 2/2—Machine Learning From Scratch 09—Python Tutorial. <https://www.youtube.com/watch?v=Bqi7EFFvNOg>
- Loeber, P. (2019c, November 27). Random Forest in Python—Machine Learning From Scratch 10—Python Tutorial. <https://www.youtube.com/watch?v=Oq1cKjR8hNo>
- Maniruzzaman, Md., Rahman, Md. J., Ahammed, B., & Abedin, Md. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health Information Science and Systems*, 8(1), 7. <https://doi.org/10.1007/s13755-019-0095-z>
- Normalized Nerd (Director). (2021, January 13). Decision Tree Classification Clearly Explained! <https://www.youtube.com/watch?v=ZVR2Way4nwQ>
- Pranto, B., Mehnaz, S., Mahid, E. B., Sadman, I. M., Rahman, A., & Momen, S. (2020). Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh. *Information*, 11(8), 374. <https://doi.org/10.3390/info11080374>
- Prasanna, S. (2019). *Machine Learning with Python*. 1, 167.
- Punthakee, Z., Goldenberg, R., & Katz, P. (2018). Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Canadian Journal of Diabetes*, 42, S10-S15. <https://doi.org/10.1016/j.cjcd.2017.10.003>
- Rokach, L. (2009). *Pattern Classification Using Ensemble Methods* (Illustrated edition, Vol. 75). World Scientific Publishing Company.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). *Data mining for business analytics: Concepts, techniques and applications in Python* (1st ed.). John Wiley & Sons.
- Singh, H. (2021, March 30). Variants of Stacking | Types of Stacking—Advanced Ensemble Learning. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/03/advanced-ensemble-learning-technique-stacking-and-its-variants/>
- Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia Computer Science*, 132, 1578-1585.

-
- Sruthi, E. R. (2021, June 17). Random Forest | Introduction to Random Forest Algorithm. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- Swaminathan, S. (2019, January 18). Logistic Regression—Detailed Overview. Medium. <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>
- Tigga, N. P., & Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. *Procedia Computer Science*, 167, 706-716.
- WHO (World Health Organization) (2021, November 10). Diabetes. <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- Zhang, C., & Ma, Y. (Eds.). (2012). *Ensemble Machine Learning: Methods and Applications* (2012th edition). Springer.
- Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked*, 17, 100179.
- Zimmet, P., Alberti, K. G., Magliano, D. J., & Bennett, P. H. (2016). Diabetes mellitus statistics on prevalence and mortality: Facts and fallacies. *Nature Reviews Endocrinology*, 12(10), 616-622.

JOURNAL OF SCIENCE

PART A: ENGINEERING AND INNOVATION



| Correspondence Address |

Gazi University
Graduate School of Natural and Applied Sciences
Emniyet Neighborhood, Bandırma Avenue
No:6/20B, 06560, Yenimahalle - ANKARA
B Block, Auxiliary Building

| e-mail |

gujsa06@gmail.com

| web page |

<https://dergipark.org.tr/tr/pub/gujsa>

e-ISSN 2147-9542