

JSTER

JOURNAL OF SCIENTIFIC, TECHNOLOGY AND ENGINEERING RESEARCH



Bilim, Teknoloji ve Mühendislik Arařtırmaları Dergisi
ISSN: 2717-8404

Cilt/Volume: 2 Sayı/Issue: 2 Yıl/Year: 2021

JOURNAL

SCIENTIFIC | TECHNOLOGY | ENGINEERING

Bilim ⊕ Teknoloji ⊕ Mühendislik





KÜNYE

Teknik Editör

Editor in Chief

Dr. Mehmet BULUT

YAYIN DANIŞMA KURULU

Editorial Advisory Board

Prof.Dr. Ali KARA

Gazi Üniversitesi

Prof. Dr. Ayhan İSTANBULLU

Balıkesir Üniversitesi

Prof.Dr .Osman TAYLAN

King Abdulaziz University, KSA

Prof. Dr. İbrahim TÜRKOĞLU

Fırat Üniversitesi

Dr. Hakan TORA

Atılım Üniversitesi

Dr. Şenol PAZAR

Biruni Üniversitesi

Dr. İbrahim ÖZ

Türksat Genel Müdürlüğü

EDİTÖRLER KURULU /

Editorial Board / Section Editors

Prof. Dr. A. Beril TUGRUL

İstanbul Teknik Üniversitesi

Prof. Dr. Enrique Herrera VIEDMA

University of Granada, SPAIN

Prof. Dr. Yunus ÇENGEL

Aydın Adnan Menderes Üniversitesi

Prof. Dr. Hüsamettin BULUT

Harran Üniversitesi

Prof. Dr. Hamit ADİN

Batman Üniversitesi

Doç. Dr. Metin VARAN

Sakarya Uygulamalı Bilimler Üniversitesi

Asst. Prof. Dr. Hasan TİRYAKİ

İstanbul University - Cerrahpaşa

Dr. Yalçın BULUT

MATRİSEB Engineering Consultancy, TURKEY

Dr. Chitaranjan PANY

Structural Engineering Entity, VSSC, INDIA

Dr. Hijaz AHMAD

International Telematic University, ITALY

Dr. Mahdi HOSSEINI

Jawaharlal Nehru Technological University,

Hyderabad, INDIA

Uluslararası Danışma Kurulu

International Advisory Board

Prof. Dr. Hafedh BELMABROUK

Majmaah University, KSA

Prof. Dr. Ali ALLAHVERDİ

Kuwait University, KUWAIT



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

e-mail : jster.editor@gmail.com

EDİTÖR NOTU / Editor Note

Journal of Scientific, Technology and Engineering Research, mühendislik uygulamalarının tüm akademik ve teknoloji odaklı taraflara ulaştırılması amacıyla yönelik olarak yayınlanan dergimiz ikinci yılını da tamamlamış olup, **110 farklı ülkeden araştırmacı tarafından ziyaret edilmiş ve farklı ülkelerden makale başvurusu olarak uluslararası akademik dergi** olduğunu perçinlemiştir. göstermiştir.

JSTER, yerli ve yabancı araştırmacılara, bilim, teknoloji ve mühendislik alanlarındaki çalışmalarını ve emeklerini göstermeleri açısından kendilerini ifade etme şansı veren bir platform olmuştur. Bilim, Teknoloji ve Mühendislik Araştırmaları Dergisi, hedeflediği misyonu doğrultusunda kalitesinden ödün vermeden yoluna devam etmeyi sürdürecektir.

Journal of Scientific, Technology and Engineering Research has completed its first year published for the purpose of delivering engineering applications to all academic and technology-oriented parties, and its website has been visited by researchers from 78 different countries and has shown that it has attracted close attention.

JSTER has been a platform that gives local and foreign researchers the chance to express themselves in terms of showing their work and efforts in the fields of science, technology and engineering. Journal of Science, Technology and Engineering Studies will continue on its way without compromising its quality in line with its mission.

Dr. Mehmet BULUT

Kapsam / Scope

Dergide yayınlanan makaleler, izin alınmaksızın hiçbir şekilde başka bir yerde yayımlanamaz veya konferans, *seminer*, kongre gibi yerlerde bildiri olarak sunulamaz. Makalelerin bir kısmı veya tamamı, ancak dergimiz kaynak gösterilerek kullanılabilir.

Articles published in the journal may not be published elsewhere without permission or can not be presented as a paper in conferences, seminars, congresses. Some or all of the articles can only be used by citing our journal.

Yayın Dili / Publishing Language

Dergide yayınlanacak makalelerin yayın dili olarak, Türkçe ve İngilizce dillerinde yazılmış makaleleri yayınlamaktadır.

As the publication language of the articles published in the journal, it publishes articles written in Turkish and English.

İlim alçakta olanları yükseltir, ilimsizlik de yüksektekileri alçaltır.

Knowledge uplifts the low, ignorance lowers the high.

Hz. Ali (r.a)



İÇİNDEKİLER / CONTENTS

İnceleme Makalesi / Review Article

Makale Başlığı / Article Name

Dil/Lang. Sayfa / Pages

Covid-19 Virüsü Salgını İle Mücadelede Büyük Veri Çalışmaları: Çin Örneği
Big Data Studies in Combating the Covid-19 Virus Outbreak: The Case of China
Ugur ERTOY, Muammer AKÇAY

TUR 4-14

Araştırma Makalesi / Research Article

Contemplation of the Photonics Beam to Complement the Water Cycle in Gulf Region
Körfez Bölgesi'ndeki Su Döngüsünü Tamamlayacak Fotonik Işının İzlenmesi
Imadeldin ELMUTASİM, İzzeldin I. MOHD, Khalid Hamid BİLAL

ENG 15-21

A Reliability Assessment of an Industrial Communication Protocol on a Windows OS Embedded PC for an Oil Rig Control Application
Windows Gömülü Bilgisayar üzerinde Petrol Sondaj Kulesi için Endüstriyel İletişim Güvenirliliği
Ongun YUCESAN, Altan ÖZKİL, Efe ÖZBEK

ENG 22-30

Endüstriyel Robot Hareket Planlama Algoritmaları Performans Karşılaştırması
Industrial Robot Motion Planning Algorithms Performance Benchmarking
Uğur YAYAN, Alim ERDOĞMUŞ

TUR 31-45

Comparative Analysis of Machine Learning Algorithms Based on Variable Importance Evaluation
Değişken Önemi Değerlendirmeye Dayalı Makine Öğrenme Algoritmalarının Karşılaştırmalı Analizi
Hasan YILDIRIM

ENG 46-53

On modeling of a recurrent neural network from neural spiking data
Öz yinelemeli sinirsel ağların sinirsel atım verilerinden yararlanılarak modellenmesi hakkında
Özgür DORUK, Mohammed AL-AKAM

ENG 54-66

Genetik Algoritma Yaklaşımıyla Öznitelik Seçimi Kullanılarak Makine Öğrenmesi Algoritmaları ile Kalp Hastalığı Tahmini
Heart Disease Prediction with Machine Learning Algorithm Using Feature Selection by Genetic Algorithm
Berna VATANSEVER, Hakan AYDIN, Ali ÇETİNKAYA

TUR 67-80

Classification of Blood Cells from Blood Cell Images Using Dense Convolutional Network
Dense Evrimsel Ağ Kullanarak Kan Hücreleri Görüntülerinden Kan Hücrelerinin Sınıflandırılması
Ferhat BOZKUR

ENG 81-88

Infusing Energy Efficient Illumination Design to Retrofit Existing Infrastructures - A case of energy efficient illumination design of Multipurpose Hall at Jigme Namgyel Engineering College
Hemlal BHATTARAI, Abhishek POKHREL, Sarda CHUWAN, Abiskar CHHETRI, Chimi DEM

ENG 89-100

2012-2019 Yılları Arasında Maden Sektöründe Yaşanan İş Kazalarının Analizi ve ÇKKV Yöntemlerinin Katkısı / *Analysis of Occupational Accidents in the Mining Industry Between 2012-2019 and Contribution of MCDM Methods*
Tuğçe ORAL

TUR 101-109

RETRACTION: DSTATCOM Based on Artificial Neural Networks and Particle Swarm Optimization for Voltage Profile Improvement
Fouad ZARO

ENG 110





REVIEW ARTICLE

Big Data Studies in Combating the Covid-19 Virus Outbreak: The Case of China

* Uğur ERTOY, ¹ Muammer AKÇAY

*Kutahya Dumlupınar University, Computer Engineering Dept., Kutahya, Turkey. ugur.ertoy@gmail.com, Orcid.0000-0001-6747-9309

¹Kutahya Dumlupınar University, Computer Engineering Dept., Kutahya, Turkey. muammer.akcay@dpu.edu.tr, Orcid.0000-0003-0244-1275

HIGHLIGHTS

- This article is a compilation of AI studies applied in China against Covid-19
- Studies in this topic has been gathered and compiled differently.
- The portion of big data infrastructure and AI apps in China
- This article is a good source as a compilation of the literature regarding China case

GRAPHICAL ABSTRACT

The coronavirus epidemic has greatly affected lives by causing radical changes all over the world. It has caused changes in a variety of areas, from the way of working and education, communication and interaction methods, shopping behavior of people. China has actively used smart technologies such as big data, artificial intelligence, cloud computing, blockchain, 5G to fight the coronavirus. Companies made their algorithms public, researchers shared data. Companies have also increased access to important video streaming tools for educators and remote workers. While its effects are continuing under various restrictions all over the world and the effects of the second wave are heavy in many countries, the epidemic has been brought under control in countries such as South Korea, Singapore, Taiwan and especially China, and life almost completely returned to normal. In this research, it is aimed to investigate how the big data and artificial intelligence studies which have been applied in China are used to fight against the pandemic and the situation is taken under control by the installation of digital infrastructure, the analysis and monitoring of the data provided by the sensors and cameras.

Keywords:

- Coronavirus
- Covid-19
- Epidemic
- Big data studies
- Artificial intelligence



Figure 3. The people who wears masks during Spanish Flu [10]

Article Info:

Received : 02.03.2021

Accepted : 24.04.2021

Published : 21.12.2021

DOI:

10.5281/zenodo.4718426

*Correspondence:

Uğur Ertoy

ugurertoy@gmail.com

+90 555 792 0770.

Aim of Article: It is aimed to research how China prevented Covid19 viruses' spread in just a few months by artificial intelligence using big data.

Theory and Methodology: The experimental method or theory of this study is mixed methodology which includes case studies, participant and non-participant observations, observational trials.

Findings and Results: China has succeeded in collecting big data, processing the collected data with the help of artificial intelligence and producing meaningful interpretations, creating a technological ecosystem by keeping a whole structure consisting of drones, autonomous vehicles, health institutions, cameras and sensors in communication thanks to the internet of things.

Conclusion: China is ahead of almost all countries in the world in the war with the pandemic. Therefore, other countries have a lot to learn from China's data-driven crisis management assessment.



İNCELEME MAKALESİ

Covid-19 Virüsü Salgını ile Mücadelede Büyük Veri Çalışmaları: Çin Örneği

* Uğur ERTOY, ¹ Muammer AKÇAY

*Kütahya Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Böl., Kütahya, Türkiye. ugur.ertoy@gmail.com, Orcid.0000-0001-6747-9309

¹Kütahya Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Böl., Kütahya, Türkiye. muammer.akcay@dpu.edu.tr, Orcid.0000-0003-0244-1275

Alıntı / Citation :

Ertoy, U., Akçay, M. (2021). Covid-19 Virüsü Salgını ile Mücadelede Büyük Veri Çalışmaları: Çin Örneği, Journal of Scientific Technology and Engineering Research, 2(2): 4-14. DOI: 10.5281/zenodo.4718426

ÖNE ÇIKANLAR

- Bu makale Covid-19'a karşı Çin'deki yapay zeka uygulamalarını içeren bir derlemedir.
- Bu konuda yapılmış çalışmalar derlendi ve farklı bir şekilde yorumlandı.
- Bu makalede Çin'deki büyük veri altyapısının yapay zekayı besleyerek salgınla mücadelede payına yer verilmiştir.
- Bu makale Çin örneği için iyi derlenmiş bir kaynaktır.

Makale Bilgileri /Article Info

Geliş Tarihi : 02.03.2021

Kabul Tarihi : 24.04.2021

Yayın Tarihi: 12.21.2021

DOI: 10.5281/zenodo.4718426

*Sorumlu Yazar /

Corresponding Author:

Uğur Ertoy,

ugurertoy@gmail.com

Tel: +90 555 7920770

ÖZET

Koronavirüs salgını, bütün dünyada köklü değişikliklere sebep olarak bireylerin hayatını büyük ölçüde etkiledi. Çalışma ve eğitim şekillerinden, iletişim ve etkileşim yöntemlerine, alışveriş davranışlarına kadar çeşitli alanlarda değişikliklere sebep oldu. Çin, koronavirüsle savaşmak için büyük veri, yapay zekâ, bulut bilişim, blok zinciri, 5G gibi akıllı teknolojileri aktif olarak kullandı. Şirketler algoritmalarını halka açık hale getirdi, araştırmacılar verileri paylaştı. Şirketler, eğitimciler ve uzaktan çalışanlar için önemli video yayın araçlarına erişim olanakları artırıldı. Etkileri hala bütün dünyada çeşitli kısıtlamalar altında devam ederken ve birçok ülkede ikinci dalganın etkileri ağır bir şekilde seyrederken, Güney Kore, Singapur, Tayvan ve özellikle Çin gibi ülkelerde salgın kontrol altına alındı ve neredeyse hayat tamamen normale döndü. Bu çalışmada, Çin'de uygulanan büyük veri ve yapay zekâ çalışmalarının salgınla mücadelede nasıl kullanıldığı ve durumun dijital altyapı kurulumu, algılayıcıların (sensörlerin), kameraların sağladığı verileri analizi ve takibi ile kontrol altına alındığı incelenmektedir.

I. GİRİŞ [INTRODUCTION]

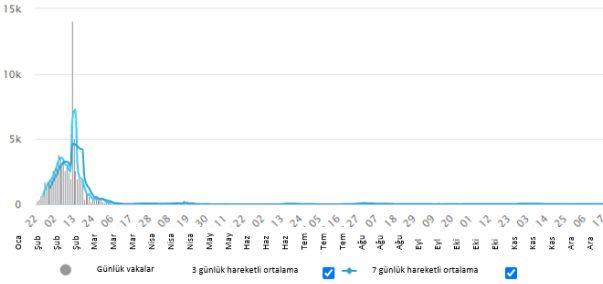
Yeni tip koronavirüs vakası, ilk olarak Aralık 2019'da Çin'in Hubei (Wuhan) kentinde ortaya çıkmıştır. COVID-19 salgını dünyanın 214 ülkesine ve bölgesine yayıldı ve günlük hayatımızın her yönünü önemli ölçüde etkiledi. Bu çalışmanın yapıldığı tarih itibarıyla, [1] dünya genelinde 76.4 milyon pozitif vaka, 1.688.574 ölüm raporlandı. Bu makalede, Çin devletinin COVID-19 salgınıyla mücadelede ne gibi yapay zekâ uygulamalarından faydalandığı, büyük veriden nasıl faydalandığı incelenmiştir. Yapay zekâ ve büyük verinin önemine vurgu yapılan bu çalışmada, COVID-19'a karşı savaşmayı amaçlayan uygulamalardan bahsedilmiş, son teknoloji

çözümlerle ilişkili zorluklar ve sorunlar belirtilmiştir.

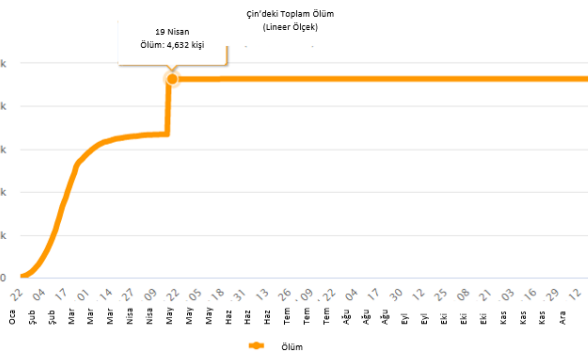
Yeni tip koronavirüsün neden olduğu COVID-19, dünyayı sadece sağlık hizmetleri alanında değil, aynı zamanda eğitim, ulaşım, politika, tedarik zinciri gibi insan hayatının birçok alanında da önemli ölçüde değiştirdi. Yeni koronavirüse karşı savaşın liderleri olarak Dünya Sağlık Örgütü (WHO) ve Hastalık Kontrol ve Önleme Merkezleri (CDC) bir dizi kamu tavsiyesi ve teknik kılavuz yayınladı [2] [3]. Ulusal hükümetler ve büyük şirketler arasındaki iş birliği ve çabaların COVID-19 salgınının yayılmasından kaynaklanan riskleri önemli ölçüde azaltması bekleniyor. Örneğin, Google, koronavirüs haritası, en son istatistikler ve sık sorulan sorular gibi yararlı bilgilerin paylaşıldığı bir

COVID-19 portalını [4] kullanıma sundu. Aynı şekilde Microsoft da benzer bir veri tabanı [5] hizmete sundu. Diğer bir örnek, IBM, Amazon, Google ve Microsoft'un Beyaz Saray ile birlikte koronavirüsle ilgili araştırmalar için bir süper hesaplama sistemi geliştirmesidir [6]. Pandemiye yanıt olarak, bazı yayıncılar artık COVID-19 benzeri virüsle ilgili makalelere, teknik standartlara ve diğer belgelere ücretsiz erişim sunarken, arXiv, medRxiv ve bioRxiv gibi web arşiv hizmetleri de COVID-19 ile ilgili tüm belgelere ulaşılması için hızlı bir bağlantı oluşturmuştur [7]. Yapay zekâ (AI) ve büyük veri, şimdiye kadar bilgisayar bilimi, bankacılık, tarım ve sağlık hizmetleri gibi çeşitli alanlarda birçok uygulamada bulundu. Çin'de 19 Aralık 2020 tarihi itibarıyla 86.806 pozitif vaka, 4.634 ölüm 81.865 iyileşen hasta tespit edildi. Salgın başlangıcından sonra 14.108 pozitif vaka ile pik yaptığı tarih 12 Şubat 2020'dir (Şekil 1). Ölümler ise 4.632 ölüm ile 19 Nisan 2020 tarihinde pik yapmıştır (Şekil 2). Makalenin yazıldığı bu tarihe kadar ise Çin'de yeni vaka ve ölüm göz ardı edilecek kadar az olduğu için yok sayılmaktadır. Çin devletinin, ilk vaka Wuhan'da çıktığı halde bütün dünyadan önce bunu nasıl başardığı ve 2.dalgaya dahi yakalanmadan süreci yönettiği merak konusu olmuştur.

Günlük Yeni Vakalar



Şekil 1. Çin'de Günlük Vakalar (19 Aralık 2020) [1]



Şekil 2. Çin'de Görülen Ölümler (19 Aralık 2020) [1]

II. LİTERATÜR ÇALIŞMASI [LITERATURE STUDY]

A. Tarihte Salgın Hastalıklar

İnsan veba vakaları çok eski zamanlardan beri bilinmektedir. İlk veba kaydı, İncil'de (I Samuel, V ve VI) anlatılan MÖ 1320'de Filistliler arasında bir salgındır. Son iki bin yılda, veba yaygın hale geldi ve birçok pandemi sırasında çoğu kıtada çok sayıda ülkeyi etkiledi. Justinianus salgını olarak bilinen ilk veba salgını MS 542 ile MS 546 arasında meydana geldi ve Asya, Afrika ve Avrupa'da salgınlara neden oldu. Yaklaşık 100 milyon mağdur olduğu tahmin ediliyor.

İkinci veba salgını, on dördüncü yüzyılın ortalarında ortaya çıkan (1347–1350) "Kara Ölüm" olarak tanınan vebadır. Bu salgın, sonraki yüzyıllarda Avrupa ve Afrika'yı kasıp kavuran bir dizi veba salgınının başlangıcıydı.

Üçüncü salgın 1894'te Kanton ve Hong Kong'da başladı ve ticaret filolarındaki yavaş hareket eden yelkenli gemilerin yerini alan daha hızlı buharlı gemilerdeki fareler tarafından hızla tüm dünyaya yayıldı. 10 yıl içinde (1894–1903) veba, beş kıtada 77 limana girdi. Veba birçok ülkede yaygınlaştı. Hindistan'da 1898'den 1908'e kadar 6 milyondan fazla ölüm oldu [8].

20. yüzyılda ortaya çıkan ve Covid-19 virüsünün temelini oluşturduğu belirtilen, İspanyol Gribi 1918-1919 yıllarında milyonlarca ölüme sebep oldu. İspanyol gribi, tarihteki pandemilerin en büyüğü olarak bilinmektedir. 1918 yılında 1. Dünya Savaşı etkisiyle yayılımını artırmıştır [9]. Şekil 3'te İspanyol Gribi dönemindeki maske takan bir grup insan görülmektedir.

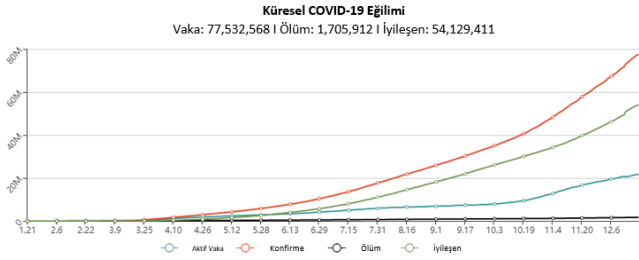


Şekil 3. İspanyol Gribi döneminde maske kullanımı [10]

B. Koronavirüs (Covid-19)

COVID-19 bir betakoronavirüs olan şiddetli akut solunum sendromu yeni tip koronavirüsten (SARS-CoV-2)

kaynaklanır. Yeni vaka sayısı hala çok büyük, günde yaklaşık +440.000 yeni vaka ve +7.000 ölü ile ölüm oranı% 2.2'dir (22 Aralık 2020 itibariyle). Küresel eğilim Şekil 4'te görülmektedir. COVID-19'un ciddi durumu nedeniyle DSÖ (Dünya Sağlık Örgütü), COVID-19 risk değerlendirmesini en yüksek seviyeye çıkardı ve pandemi olarak ilan etti.



Şekil 4. Küresel COVID-19 eğilimi (22 Aralık 2020) [11]

COVID-19'un dünya üzerindeki çarpıcı etkisi olarak COVID-19 salgınına karşı savaşmak için birçok çözüm girişiminde bulunuldu. Hükümetler temel olarak pandemiye durdurılmaktan sorumludur, örneğin enfeksiyonun yayılmasını sınırlamak için kısıtlamaların ilan edilmesi, sağlık sisteminin salgınla baş edebilmesini sağlamak ve ulusal ekonomi üzerindeki etkileri hafifletmek için kriz paketi sağlamak ve COVID-19 durumuna göre uyarlanabilir politikalar benimsemesi gibi topluma maddi ve manevi katkı sağlayacak aksiyonlar almak durumundadır. Aynı zamanda, halka açık yerlerde maske takılması, ellerin sık yıkanması, sosyal mesafenin korunması ve en son bilgilerin bölge sağlık merkezine bildirilmesi için halk sağlıklı kalmaya ve başkalarını korumaya teşvik edildi. Öte yandan, COVID-19 ile ilgili araştırma ve geliştirmeye artık öncelik verilmiştir. Hükümetler, endüstriler ve akademi gibi çeşitli paydaşlardan özel ilgi görmüştür. Örneğin, yapılan araştırmalar COVID-19 salgınının küresel tedarik zinciri üzerindeki büyük etkilerini gösterdi ve sürdürülebilirlik, istikrar, sağlamlık ve esneklik dahil olmak üzere tedarik zincirlerinin farklı yönlerini değerlendirdi. COVID-19 koronavirusu için etkili bir aşı ve tıbbi tedavi geliştirmeye yönelik küresel girişimin yanı sıra, bilgisayar bilimi araştırmacıları COVID-19 ile mücadele için ilk çabaları gösterdi. Yapay zekânın ve çeşitli alanlardaki büyük verinin muazzam başarısından motive olarak, COVID-19 koronavirus hastalığının üstesinden gelmek için yapay zekâyâ ve büyük verilere dayanan son teknoloji çözümler ve yaklaşımlar sunulmaya başlandı [12].

C. Büyük Veri

Büyük veri kavramı 1990 yılında ortaya atılmış ve bu tarihten itibaren sanayideki ve akademideki giderek artan bir eğilim izlemiştir. Nesnelere internetinin ve teknolojinin son yıllardaki hızlı büyümesi ile birlikte hem yazılımsal gelişimin hem de donanımsal gelişimin hayatımızın her alanında kullanılması dünya üzerinde toplanan büyük verinin hacminin artmasında büyük rol oynamaktadır. Büyük verinin hızlı büyümesi depolama, işleme, analiz etme ve yorumlama gibi problemleri de beraberinde getirmiştir. Bu bağlamda büyük veri kavramı hem veriyi hem de beraberinde getirdiği teknolojiyi, sorunları ve yöntemleri de barındırır. Büyük veri tanımını yaparken beş farklı karakteristik özellikten bahsetmek gerekmektedir [13].

- Hacim: Bu özellik, verinin büyüklüğünü, miktarını ifade eder. Genel olarak terabayttan eksabayta kadar değişebilen birimlerle ifade edilmektedir.
- Çeşitlilik: Bu özellik, büyük verinin çeşitliliğini ifade eder. Örneğin, giyilebilir cihazlardan sağlanan sağlık verileri, mobil cihazlardan sağlanan veriler, sağlık sektörü çalışanlarının girdileri, internete bağlı olan tıbbi cihazlardan sağlanan veriler, sosyal medya verileri, otonom sistemlerden sağlanan veriler, sokak kameraları ve termal kameralardan sağlanan veriler gibi birçok farklı alandan heterojen bir veri akışı bulunmaktadır. Veriler, yapılandırılmış veya yapılandırılmamış türlerde bulunmaktadır.
- Hız: Bu özellik, veri üretme hızını ifade eder. Veriler sürekli akmaktadır ve akan verilerin değerlendirilerek yorumlanması gerekmektedir. Örneğin, sağlık sektöründeki veri üretme hızı, sağlık durumunu takip etme veya teşhisin hızlı konması gibi durumlar için büyük önem taşımaktadır.
- Doğruluk: Bu özellik verinin kalitesini, veri üzerindeki şüpheyi ifade eder. Veri kalitesinin yüksek oranda olmaması yanlış kararlar verilmesine yol açabilir.
- Değer: Büyük veriler sonuç olarak bir katma değere sahip olmalıdır. Ortaya çıkacak verilerin yukarıda belirtilen özelliklerle birlikte geliştirilebilir olması beklenmektedir. Buna göre kurulacak altyapı ve donanım yatırımlarına karar verilebilmektedir.



1) Koronavirüs ile Mücadelede Büyük Veri

Büyük verilerin koronavirüs gibi bulaşıcı hastalıklarla mücadelede yardımcı olmaktadır. Yapay zekâ analitiği ile birleşen büyük veri, salgının tahmin edilmesi ve takibi, virüsün yapısı, hastalığın tedavisi, vakaların izlenebilirliği, bulaş takibi ve aşı üretimi açısından koronavirüsün anlaşılmasına yardımcı olmuştur. Örneğin, yapay zekâ tabanlı donanımlarla ilişkili büyük veriler, salgın tahmini için koronavirüs veri akışlarını kullanarak karmaşık benzetim modelleri oluşturabilir. Bu, sağlık kuruluşlarının koronavirüs yayılmasını izlemesine ve daha iyi önleyici tedbirler hazırlamasına yardımcı olmuştur. Bu modeller erken tespit için büyük miktarda veriden yararlanma konusundaki veri toplama yeteneği sayesinde gelecekte koronavirüs salgını tahminini de desteklemektedir. Aynı zamanda bütün dünyanın veri tabanlarını birbirine açarak güvenilirlik oranı yüksek kapsamlı tedavi çözümleri geliştirmek için yapay zekâ geliştirilerek koronavirüs araştırmalarının verimliliğinin ve hızının artmasını desteklemektedir. Büyük veri aynı zamanda sağlık sektöründeki teşhislerin daha doğru yapılması ve aşı/ilaç geliştirilmesi ve tedavilerin daha doğru uygulanması için virüs gelişimini anlamada yardımcı olmaktadır [13].

Büyük veri analitiğinin, koronavirüs veri keşfi gibi diğer bilgileri ve faydalı gizli kalmış bilgileri keşfetmek için büyük hacimli veri kümelerini toplama ve analiz etme süreci olduğu vurgulanabilir. Yapay zekâ koronavirüs belirti sınıflandırmaları gibi olası sonuçları öğrenmek, sınıflandırmak ve tahmin etmek için bir makinenin işlevini taklit edebilen insan zekâsı oluşturmak amacıyla mantık ve akıl yürüterek uygulamayı amaçlamaktadır. Koronavirüs salgınıyla mücadelede her teknolojinin potansiyel uygulamaları bir dizi pratik kullanım örneği aracılığıyla aşağıdaki bölümlerde açıklanacak ve tartışılacaktır.

2) Koronavirüs ile Mücadelede Yapay Zekâ

Yapay zekâ, bilgisayar bilimi başlığı altında toplanan, makine öğrenmesi ve derin öğrenmeyi kapsayan ana bilim dalıdır. Yapay zekâ günümüzde birçok alanda birbirinden farklı uygulama ile sürekli gelişmeye devam etmektedir. Günümüzde var olan teknolojilerden bahsetmek gerekirse, otomotiv endüstrisinde otonom araçlar, sağlık sektöründe tıbbi teşhis ve teletıp uygulamaları, siber güvenlik sistemleri, finans sektöründe yazılımlar, görüntü işleme ve sesli komutların yazıya aktarılması da birer yapay zekâ örneğidir. Makine öğrenmesi, sistemlerin algoritmalar ve veri kümeleri dâhilinde kalıpları tanımasını ve çözüm kalıpları geliştirmesidir. Makine öğrenmesinde yapay

bilgi deneme yoluyla üretilir yani sistem öncelikle insanlar tarafından algoritmalar ve veriler ile beslenmelidir. Derin öğrenme ise insan sinir sisteminden ilham alarak yapay sinir ağları ile örnekler üzerinden insana en yakın öğrenmeyi sağlayan yapay zekâ yaklaşımıdır. Daha basit temsillerden öğrenerek karmaşık sistemleri çözebilir [13]. Örneğin, bilim adamları “ilacın yeniden kullanılması” için mevcut ve ticari ilaçları belirlemek yani pozitif hastalara hemen uygulanabilecek mevcut ilaçları kullanarak hızlı bir ilaç stratejisi bulmak için bir derin öğrenme modeli geliştirdi. Bu çalışma, yeni geliştirilen ilaçların piyasaya sürülmeden önce başarıyla test edilmesinin genellikle yıllar alması gerçeğinden kaynaklanmaktadır. Bu çalışmadaki bulgular şu anda klinik olarak onaylanmamış olsa da koronavirüs hastalığıyla mücadelede yeni veriler oluşturmaktadır. Bilgisayarlı tomografi (CT) görüntü işleme için derin öğrenme kullanılan çalışmada, 499 CT hacmi ile eğitim ve 131 CT hacmi üzerinde test yaparken önerilen derin öğrenme modelinin 0,840 pozitif tahmin değeri ve 0.982 negatif tahmin değeri ile 0,901 doğruluk değeri elde edebileceği gösterilmiştir. Bu çalışma sayesinde vakaları zamanında teşhis ve takip ederek karantinaya alma ve tıbbi tedaviye başlama süresini hızlandırmıştır. Bir diğer örnek ise, Çin’deki koronavirüs salgınının gerçek zamanlı tahmin dinamiklerinde (ör. Enfekte vakaların sayısı, sona erme süresi ve salgının seyri) yapay zekânın (yani otomatik kodlayıcı tabanlı bir yöntem) kullanılmasıdır [13]. Önerilen yapay zekâ tabanlı yaklaşımın yüksek doğruluğu, COVID-19 salgınının izlenmesiyle birlikte sağlık ve politika stratejilerinin iyileştirilmesine yardımcı olur. Bahsedilen uygulamaların yanı sıra COVID-19 virüsünün araştırma ve geliştirmesini hızlandırmak için araştırmacılar, doktorlar ve bilim adamlarının etkin bir şekilde desteklenebilmesi için dev teknoloji şirketlerinin katkısına ihtiyaç vardır. Son zamanlarda IBM, COVID-19 ile ilgili araştırma makalelerinin bir koleksiyonu olan COVID-19 açık veri kümesi (CORD-19) üzerinde eğitilmiş bulut tabanlı bir araştırma kaynağı sağladığını duyurdu [14]. Ayrıca IBM, ilaç keşfi için önerdiği yapay zekâ teknolojisini benimsemiştir ve bu teknolojiden 3000 yeni COVID-19 molekülünün elde edildiği resmi olarak rapor edilmiştir [15]. Diğer bir destek ise COVID-19 ile ilgili araştırma önerilerine açık olan COVID-19 HPC Konsorsiyumunun [16] geliştirilmesiyle Beyaz Saray Bilim ve Teknoloji Politikası Ofisi, ABD Enerji Bakanlığı ve IBM’den gelmektedir. Diğer bir örnek ise COVID-19’a karşı aşılarda ve terapötik çözümler geliştiren bir grup bilim insanı olan Koronavirus Uluslararası Araştırma Ekibi (COV-IRT) [17].

D. Koronavirüs ile Mücadelede Çin'de Büyük Veri ve Yapay Zekâ Uygulamaları

Çin, veriye dayalı çözümleri hızla ve Covid-19 salgınının başlarında devreye aldı. Halihazırda büyük miktarda veri üreten önceden var olan kapsamlı bir dijital altyapıya güvenerek pandemiyle savaşmaya başladı. Çin hükümeti mevcut altyapıyı kolaylıkla dönüştürerek koronavirüsle mücadele stratejisinin ayrılmaz bir parçası haline getirdi. Mevcut dijital teknoloji ekosisteminin bileşenlerini yenileyip geliştirdiler. Yüz tanıma sistemleri ve WeChat gibi uygulamalar vardı. Hükümet ve iş dünyası arasındaki köklü bağlantılar, Pekin'in büyük miktarda kullanıcı verisinden, genellikle gerçek zamanlı olarak yararlanmasını sağladı. Cep telefonu takip verilerinden üretilen araçlar Çin'in Salgın Önleme ve Kontrol (Epidemic Prevention and Control) yaklaşımının merkezinde yer almaktadır. Karantina kısıtlamalarının hızlı bir şekilde kaldırılmasını sağlayan QR kodu sağlık uygulamalarını geliştirmek için girdiler sağladılar.

Potansiyel enfekte olan kişileri belirlemek için Çin hükümeti sıcaklık sensörleri ve kızılötesi tanımlama çözümleri içeren geliştirilmiş yüz tanıma teknolojisi kullandı. Hastaneler ve doktorlar, büyük veri ve yapay zekâyâ dayalı hastalık izleme, teşhis ve kaynak yönetimi sistemleri için dijital platform çözümlerini ve ücretsiz çevrimiçi sağlık danışmanlıkları kullandı. 6 Mayıs 2020 itibarıyla Çin'de koronavirüse karşı yapay zekâ ve büyük veri teknolojileri kullanılarak farklı disiplinlerde uygulanmış 543 uygulama mevcuttur. Bu 543 uygulamanın 98 tanesini büyük veri akıllı analiz sistemleri, 90 tanesini sıcaklık ölçümü sağlayan akıllı tanı sistemleri, 69 tanesini akıllı eğitim ve çevrimiçi ofis uygulamaları, 63 tanesini akıllı danışmanlık, akıllı rehber uygulamaları, 42 tanesini medikal robot uygulamaları, 40 tanesini medikal asistan tanı robotları ve diğerleri oluşturmuştur [17].

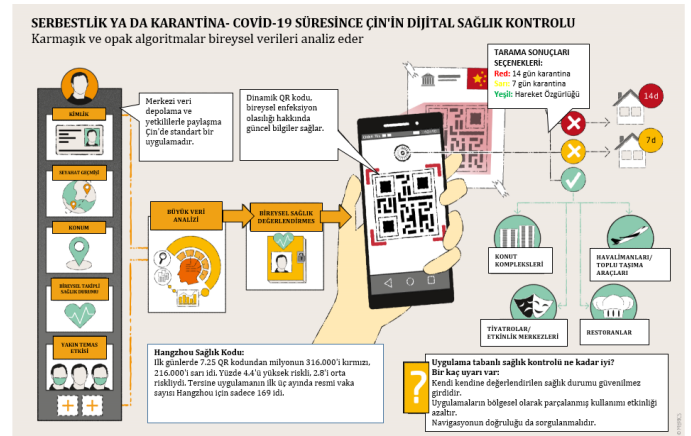
1) İnsansız Hava Aracı ve Otonom Araç Kullanımı

Çin devletinin ilk uygulamaya koyduğu işlerden birisi hâlihazırda belirli bir kitlenin kullanmakta olduğu maske ve tıbbi malzemelerin dağıtılması için insansız hava araçlarını kullanmaya başlamıştır. Tıbbi malzemelerin sağlık kurumları arasında transferlerinin sağlanması için kullanılmıştır. Kamusal alanlarda insansız hava araçlarının gezdirilerek, yüz tanıma sistemleri ve termal algılayıcılar sayesinde, hem karantina kurallarını ihlal eden insanların tespit edilmesi hem de olası yüksek ateşli insanların acilen karantinaya alınması için bilgi toplanması sağlanmıştır. İlaç ve gıda malzemelerinin sağlık kurumları arasında ulaştırılmasında sokak temizliği

ve hastane dezenfeksiyonu için otonom araçlar kullanılmıştır.

2) Akıllı Temas Takibi

Temas takibi büyük verilerin nasıl geniş ölçekte kullanılabilceğinin ve mevcut teknolojilerinin geliştirilerek neler yapılabileceğinin en güzel örneklerinden birisidir. Çoğu uygulamanın merkezinde mobil cihaz sahibinin önceki 14 gün içindeki seyahat geçmişine, coğrafi konuma ve kişisel kimliğe dayalı olarak bir kişinin virüse yakalanma olasılığını ölçen algoritmalar bulunmaktadır. Kullanıcı tarafından gönderilen bilgiler, enfekte bir kişiyle teması ve sağlık durumunun kendi kendine değerlendirilmesini içerebilir. Bu kişiselleştirilmiş dijital sağlık değerlendirme sertifikaları QR kodlarında saklanır. QR kod tarandığında üç sonuçtan birini verir. Yeşil "serbestçe hareket etmesine izin verildi", turuncu "yedi günlük karantina için önerilir" veya kırmızı "14 günlük karantina için önerilir". Çoğu toplu taşıma, süpermarket ve konut kompleksi artık bireylerden önceden verilen kâğıt sertifikalar yerine bir QR kodu göstermelerini talep ediyor. Çoğu QR uygulaması, tarandıktan sonra tek tek bilgileri merkezi bir veri tabanından alan dinamik kodlar üretmek için tasarlanmıştır. Statik QR kodlarındaki bilgiler kodun içine kalıcı olarak yerleştirilmiştir (Şekil 5). Benzer bir uygulama Türkiye'de de HES uygulaması ile pandeminin erken safhalarında hayata geçirilerek uygulamaya konmuştur.



Şekil 5. Büyük Veri ve Yapay Zekâ ile Temas Takibi [19]

Çin devletinin tüm dünyadan çok daha hızlı bir şekilde harekete geçerek 11 Mart'ta DSÖ (Dünya Sağlık Örgütü) tarafından küresel pandemi ilan edilmesinden dahi önce büyük veri ve yapay zekâ destekli uygulamaları hayata geçirmiş oldukları kronolojik olarak gözlemlenmektedir. 2019 Aralık ayında ilk vaka ortaya çıkmıştır. 12 Ocak



2020'de DSÖ 2019-nCoV olarak virüsün adını duyurmuştur. 20 Ocak 2020'de Çin başkanı insandan insana bulaşan salgın olduğunu duyurmuştur ve ilk büyük veri çalışma grubu kurularak ülke genelinde ulusal sağlık örgütüyle birlikte çalışmalara başlanmıştır, 23 Ocak 2020'de 639 vakanın ortaya çıkmasıyla kısıtlamalar başlatılmıştır. 30 Ocak 2020'de Baidu firması yeni bir algoritma ile virüs RNA'sının analizinin hızlandırılması, 2 Şubat 2020'de hastane çalışanları ve kaynak yönetimi platformu, 8 Şubat 2020'de yakın temas takibi, 9 Şubat 2020'de sağlık kodu uygulaması, 15 Şubat 2020'de vücut sıcaklık tespit sistemleri, 15 Şubat 2020'de maskeli yüzlerin tanınmasını sağlayan sistemler, 29 Şubat 2020'de 3 büyük telekom şirketi tarafından büyük veri seyahat geçiş kodu uygulaması, Tencent firması tarafından 20 Mart 2020'de eğitim sektörüne ait sağlık kodu uygulaması, 21 Nisan 2020'de ise büyük veri seyahat geçiş kodu uygulaması iyileştirilerek uluslararası seyahatleri de kapsayacak düzeyde antipandemik sağlık bilgi kodu olarak hayata geçirilmiştir [18].

Akıllı temas takibi yetkililerin olası bir tehdidi belirlemek, virüsün aktarıcılarını izole etmek ve tehlikede olabilecek herkesi bilgilendirmek için tepki süresini kısaltır. Herkese açık uygulama ayrıca her kullanıcıyı, onaylanan hastalar tarafından hangi alanların ziyaret edildiğini, hangi yerlerin karantina altında olduğunu ve potansiyel olarak etkilenebilecekleri bilgilendirir. Sistem aynı zamanda virüsün aktivitesini de takip ederek hastanelerin ve kliniklerin bir sonraki dalga durumuna daha iyi hazırlanmasına yardımcı olur.

3) Test ve Teşhis

Son dijitalleşme çabaları nedeniyle Çin'in sağlık sektörü veriye dayalı araçları kullanmaya hazırdır. 2018'den beri Danıştay ve Ulusal Sağlık Komisyonu tarafından yürütülen ulusal bir girişim, bilgi kanallarının sayısallaşmasını ilerletti. Bunlar, tıbbi kayıtlar, laboratuvar değerlendirmeleri ve görüntü arşivlemeyi, bireysel verilerin ve teşhis araçlarının çakıştığı alanları içerir. Bu girişim platform tabanlı erken uyarı sistemleri, hastalık izleme ve kaynak yönetimi sistemlerini uygulamak için büyük veri ve yapay zekânın kullanımını teşvik etti. Yararlanılacak alanlar arasında halk sağlığı ve tıp alanında araştırma ve geliştirme (Ar-Ge) bulunmaktadır. Örneğin, PingAn Technology tarafından geliştirilen bir tahmin modeli, Chongqing ve Shenzhen şehirlerindeki resmi sağlık kuruluşlarının salgınları yüzde 90'dan fazla doğruluk oranlarıyla tahmin etmelerine yardımcı olmaktadır. Shenzhen'deki The National supercomputing Center, Sensetime gibi şirketlerle iş

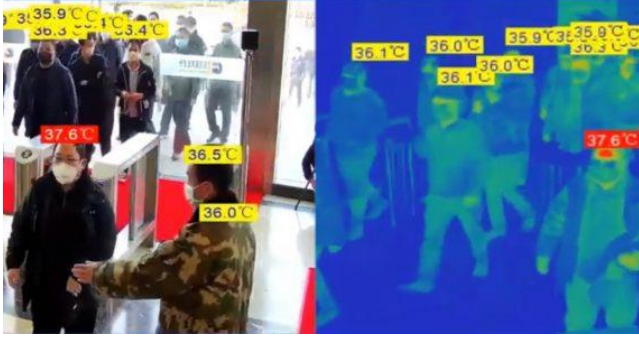
birliği içindedir, potansiyel ilaçların büyük ölçekli taramasını yapan ve virüs mutasyonlarını tahmin etmek için çalışan araştırmacılara yüksek performanslı bilgi işlem desteği sağlamaktadır. Benzer çabalar tomografi (Computer Tomography) taramalarını değerlendirmek veya SARS-CoV-2'nin genetik RNA yapısını analiz etmek için gereken zamanı büyük ölçüde azaltmada başarılı oldu. Çin'in yerli Beidou navigasyon sistemindeki uydular Çin hükümetinin altyapı kurulumunu ve kaynak dağıtımını geliştirmesini sağladı. Örneğin Wuhan Üniversitesi devasa derme çatma hastaneler için en iyi noktaları bulmak için çoklu veri akışını toplayıp analiz edebildi. Yüksek çözünürlüklü dünya gözlem uyduları daha sonra inşaatlarını izledi. E-ticaret şirketi JD ayrıca, Beidou'da çalışan dronlar ve robotlarla Wuhan'ın uzaktaki hastanelerine tıbbi ekipman sağladı.

Şirketler bireysel davranışın kapsamlı bir şekilde izlenmesini sağlamak için yüz tanıma teknolojilerini hızla geliştirdiler. Seyahat geçmişi, karantina ve maske takma uyumluluğunu (dronlar aracılığıyla) izlemenin yanı sıra vücut sıcaklığı testi de önemli bir yenilik olmuştur. Geliştiriciler, sıcaklık izleme için yüzde doksan dokuzun üzerinde ve maske takan kişilerin yüz tanımları için yüzde doksandandan fazla doğruluk seviyeleri ($\pm 0,3^\circ$ Celsius hata oranı ile) olduğunu iddia ediyorlar. Ancak doğruluklar iddia edilenden daha düşük olabilir. Aralık 2019'da yapılan bir anket Çinli yanıt verenlerin yaklaşık yüzde 60'ının konut kompleksleri ve toplu taşımayı içeren durumlarda yüz tanıma teknolojisi (FRT) ile ilgili sorunlarla karşılaştığını gösterdi.

Mobil cihazlardan GPS verileri, kredi kartı işlemleri, kamu alanlarındaki kameralar ile yüz tanıma sistemlerinin kullanılması gibi konum verileri ile temas takipleri mümkün kılınmıştır. Hatta hastalık özelinde getirilen kısıtlamaları ve belirtileri tespit edebilmek amacıyla yüz tanıma sistemlerine insanların maske giyip giymediğinin ve vücut sıcaklıklarının kontrolü de entegre edilmiştir. Termal görüntüleme teknolojisi, bir insan kalabalığı arasında kimlerin yüksek vücut ısısına sahip olduğunu belirlemeye yardımcı olup ateşi yüksek olan kişileri ayıklamak için kullanılır. Şekil 6'da yer alan görsel bu uygulamaya bir örnek olarak gösterilmektedir.

Hasta olarak tespit edilmiş bir hastanın hareketinin kapsamlı bir resmi kamuya açık olarak paylaşılır. Uygulama bireyin geçmişi, zaman çizelgesi ve ziyaret ettiği yer hakkında bilgi verir. Restoran ve otel ziyaretleri, hangi ulaşım aracını kullandıkları, hangi hat veya otobüse bindikleri, hatta sinemadaki koltuk tam olarak takip

edilmektedir. CCTV kameralar kalabalık kitlelerin vücut sıcaklıklarını, virüs aktarımını izlerken ve analiz ederken de yoğun bir şekilde kullanılır. 2020 itibarıyla Çin’de 626 milyon civarında CCTV kamera olduğu hesaplanmaktadır, bu sayı her geçen gün artmakta olup ABD’de genelindeki kamera sayısının yaklaşık 12 katıdır. Çin devleti 2022 yılına kadar 2 kişiye 1 kamera düşecek şekilde kamera sayılarının ülke genelinde artırmayı planlamaktadır [17].



Şekil 6. Termal kamera ya yansıyan insanların görüntüsü [20]

E. Sağlık Sektörüne Özel Çözümler

Çevrimiçi platformlar birçok tıp kurumunun Covid-19 ile ilgili zorluklara etkili yanıtlar geliştirmesi için temel bir destek olmuştur. Çin Sanayi ve Bilgi Teknolojileri Bakanlığı’ndan (MIIT) alınan hükümet verileri ülke çapında 190’dan fazla kamu tıp kurumunu ve yaklaşık 100 kurumsal internet hastanesinin şu anda geleneksel kamu hastaneleri üzerindeki baskıyı kısmen hafifleten ücretsiz çevrimiçi danışmanlık sağladığını göstermektedir. Good Doctor Online, Chunyu Doctor, Ping An Good Doctor ve diğer kuruluşlar, hastalara ücretsiz danışmanlık sağlamak için solunum, bulaşıcı hastalıklar ve iç hastalıkları alanlarında 10.000’den fazla tıbbi uzmanın kaynaklarını bir araya getirdi. Bu platformlar ayrıca kaynak yönetiminin iyileştirilmesine de yardımcı oldu; örneğin, “Huoshayun Sağlık Yönetimi Platformu” artık Fujian yerel yönetimine, kaynakların dağıtımını koordine etmek ve hastane personelini rahatlatmak için sağlık sektöründeki 4.500 işletmeden 60.000 çalışandan oluşan bir veri tabanı sağlamaktadır.

Büyük veri ve yapay zekâ tabanlı çözümler sağlık hizmetleri sektöründe kaynak yönetimini ve hasta tedavisini destekledi. Çözümler arasında personel ve yatakların hastane çapında yönetimi ve hastaneler arasında yapay zekâ tabanlı uzaktan çevrimiçi danışmanlıklar,

yerinde teşhis ve hasta bakımı yer almaktadır. Örneğin, Jiangsu Sağlık Komisyonu için çevrimiçi danışmanlıklara izin vermek ve tıbbi tanı formlarını (ERM) standartlaştırmak için inşa edilen bir sağlık platformu, 68 hastane tarafından benimsenmiş ve salgındaki diğer 20 sağlık hizmeti platformuyla ilişkilendirilmiştir [18].

Çin’deki doktorlar COVID-19 hastalarını teşhis etmek için bilgisayarlı tomografi (CT) taramalarına ve kan testlerine odaklandılar. Bir yapay zekâ algoritması geliştirmek için tomografi taramalarından gelen verileri klinik bilgilerle ve kullanıcıların tıbbi verileriyle entegre ederek yapay sinir ağları ve makine öğrenmesi yöntemleriyle yazılımı eğitmeyi başardılar. Geliştirilen bu yapay zekâ bir doktorun COVID-19’u teşhis etmek için kullandığı iş akışını taklit edip pozitif veya negatif tanı için nihai bir tahmin vermektedir. Yapay zekâ ayrıca bilgisayarlı tomografi taramalarına klinik verilere ve her ikisine birden dayalı olarak COVID-19 pozitif olma olasılıkları da üretmektedir.

1) Ülkeler Arasındaki Vaka Durumunun Karşılaştırılması

Tablo 1’de yer alan nüfus büyüklüklerine göre ilk 24 ülke değerlendirildiğinde dünyanın en büyük nüfusuna sahip olan Çin, 14.02.2021 itibarıyla ABD ve Hindistan’ın arkasından en çok test uygulayan üçüncü ülke olarak görünmektedir. Çin, Dünyanın en büyük nüfusuna sahip olmasına rağmen nüfusunun sadece %11’e karşılık gelen kısmına test uygulamıştır. Dünyanın süper gücü olan ABD, nüfusuna göre %100’ü geçen bir oranda test uygulaması yapmıştır. Türkiye, nüfusuna göre %36’lık bir test uygulaması yapmıştır. İngiltere, nüfusuna göre %118’lik bir test uygulaması yapmıştır.

Vaka oranlarına baktığımızda en düşük vaka oranlarında bu listede 3. olarak karşımıza çıkmaktadır. ABD, en çok vakanın yaşandığı ülke olarak milyonda 84877 kişide virüse rastlanmıştır. İngiltere ise milyonda 59128 vaka ile en çok vaka görülen 2.ülke olarak görünmektedir. Türkiye’de ise milyonda 30387 kişide virüse rastlanmıştır. Ölüm oranlarına bakıldığında Çin bu 24 ülke arasında en az kaybı veren ülkelerden 2. sırada yer almaktadır. Milyonda 3 kişiyi kayıp olarak vermiştir. İngiltere en çok ölüm görülen ülke olup milyonda 1717 kişiyi kaybetmiştir. Türkiye’de ise milyonda 322 kişi hayatını kaybetmiştir. ABD verilerine baktığımızda ise en çok ölüm görülen ikinci ülke olarak milyonda 1493 kişi kaybedilmiştir.



Tablo 1
Covid-19 Ülkelere Vaka Durumu Tablosu [1] (19 Aralık 2020)

#	Ülke	Toplam Vaka Sayısı	Nüfusa Göre Toplam Vaka Oranı	Milyonda Vaka Sayısı	Toplam Ölüm Sayısı	Nüfusa Göre Toplam Ölüm Oranı	Milyonda Ölüm Sayısı	Toplam Test Sayısı	Nüfusa Göre Toplam Test Oranı	Milyonda Test Sayısı	Nüfus
1	Çin	89.763	0,0062%	62	4.636	0,0003%	3	160.000.000	11,1163%	111.163	1.439.323.776
2	Hindistan	10.904.940	0,7854%	7.854	155.673	0,0112%	112	206.230.512	14,8536%	148.536	1.388.419.695
3	ABD	28.197.074	8,4877%	84.877	496.065	0,1493%	1.493	334.805.401	100,7815%	1.007.815	332.209.157
4	Endonezya	1.217.468	0,4422%	4.422	33.183	0,0121%	121	10.061.418	3,6545%	36.545	275.316.836
5	Pakistan	563.029	0,2519%	2.519	12.307	0,0055%	55	8.434.098	3,7730%	3.773	223.537.866
6	Brezilya	9.811.255	4,5955%	45.955	238.647	0,1118%	1.118	28.600.000	13,3959%	133.959	213.497.994
7	Nijerya	145.664	0,0696%	696	1.747	0,0008%	8	1.398.630	0,6683%	6.683	209.280.632
8	Bangladeş	540.592	0,3262%	3.262	8.274	0,0050%	50	3.848.116	2,3222%	23.222	165.706.987
9	Rusya	4.071.883	2,7895%	27.895	80.126	0,0549%	549	106.800.000	73,1640%	73.164	145.973.468
10	Meksika	1.988.695	1,5324%	15.324	173.771	0,1339%	1.339	5.085.315	3,9186%	39.186	129.772.637
11	Japonya	413.154	0,3273%	3.273	6.849	0,0054%	54	7.539.403	5,9725%	59.725	126.234.697
12	Etiyopya	145.704	0,1248%	1.248	2.181	0,0019%	19	2.040.290	1,7481%	17.481	116.714.297
13	Filipinler	549.176	0,4971%	4.971	11.515	0,0104%	104	8.288.845	7,5023%	75.023	110.483.925
14	Mısır	173.202	0,1673%	1.673	9.935	0,0096%	96	1.000.000	0,9659%	9.659	103.525.295
15	Vietnam	2.228	0,0023%	23	35	0,0000%	0,4	1.529.651	1,5627%	15.627	97.882.052
16	Kongo	2.424	0,0027%	266	693	0,0008%	8		0,0000%		91.229.525
17	Türkiye	2.579.896	3,0387%	30.387	27.377	0,0322%	322	31.412.611	36,9988%	369.988	84.901.668
18	İran	1.518.263	1,7934%	17.934	58.945	0,0696%	696	10.048.871	11,8699%	118.699	84.658.640
19	Almanya	2.336.905	2,7837%	27.837	65.415	0,0779%	779	41.758.675	49,7419%	497.419	83.950.702
20	Tayland	24.571	0,0351%	351	80	0,0001%	1	1.217.873	1,7421%	17.421	69.909.040
21	İngiltere	4.027.106	5,9128%	59.128	116.908	0,1717%	1.717	81.044.007	118,9938%	1.189.938	68.107.747
22	Fransa	3.448.617	5,2761%	52.761	81.647	0,1249%	1.249	48.207.178	73,7525%	737.525	65.363.463
23	Tanzanya	509	0,0008%	8	21	0,0000%	0,3		0,0000%		60.777.837
24	İtalya	2.710.819	4,4876%	44.876	93.356	0,1545%	1.545	35.962.939	59,5350%	59.535	60.406.354

Bu tablodan da anlaşılacağı üzere 1,5 milyarlık Çin, nüfusuna göre çok küçük oranlarda seyreden vaka ve ölüm oranlarıyla virüsü Nisan 2020 sonunda büyük veriyle beslenen, bu verileri analiz eden, akıllı sistemlerden oluşan teknolojik altyapısı sayesinde başarılı bir şekilde yönettiği gözlemlenmektedir.

2) Koronavirüsün Ülkelere Maliyeti

12 Haziran 2020 itibarıyla, dünya çapındaki ülkeler yaklaşık 11 trilyon ABD doları tutarında mali önlemler almıştır. Bu harcamaların 8 trilyon ABD doları Nisan ayına kadar yapılmıştır. Mali destekler hem bütçesel hem de bütçe dışı likidite önlemlerinden oluşmaktadır. ABD 3 trilyon ABD doları, Japonya 1,7 trilyon ABD doları ve Almanya 1,5 trilyon ABD doları harcamıştır. İlk üçte yer alan bu ülkelerden sonra Çin 400 milyar ABD doları civarında bir harcama yapmıştır. GSMH oranlarına göre büyükten küçüğe sıralandığında Almanya GSMH'dan %40,9 harcaman yaparken, İtalya %37,5, Japonya %35,4 harcama yapmıştır. Türkiye, bu sıralamada 21.sırada

GSMH'nın %9,4'ünü harcamıştır. Çin 26.sırada yer alarak GSMH'nın %4,6'sını harcamıştır [21].

III. SONUÇ VE DEĞERLENDİRME [CONCLUSION AND EVALUATION]

Çin'in Covid-19'a yönelik veriye dayalı yönetim yaklaşımının etkinliğine ilişkin değerlendirme karışık bir tablo ortaya çıkardı. Teknolojik adaptasyonun ve benimsemenin hızı ve kapsamı Pekin'in acil politik ihtiyaçları ve bazı kamu ihtiyaçlarını karşılamak için teknik çözümleri teşvik etme yeteneğini göstermiştir. Dijital önlemler Çin vatandaşları arasında kamu güvenliği algısının artmasına katkıda bulundu. Ancak tüm bu teknolojik iyileştirme ve geliştirmeler kişisel verilerin korunması pahasına meydana geldi. Kişisel verilerin korunması söz konusu değildir. Aynı zamanda teknik işlevsellikteki zayıflıkları iyileştirmeden ve kişisel veri korumasını büyük ölçüde sağlamadan merkezi ve yerel yetkililerin tam kamu desteği sağlama noktasında kısıtları olduğunu da ortaya çıkardı. ABD ve Avrupa ülkeleri ise yönetim anlayışları gereği vatandaşlarının tepkisini



almamak üzere hâlihazırda ellerinde olan verileri özgürce kullanmadılar.

Çin pandemi ile savaşta Dünya'daki nerdeyse bütün ülkelerin önündedir. Bu nedenle diğer ülkelerin Çin'in veriye dayalı kriz yönetimi değerlendirmesinden öğrenebilecekleri çok şey vardır. Çin büyük veriyi toplamayı, topladığı verileri yapay zekâ yardımıyla işlemeyi ve anlamlı yorumlar üretmeyi, insansız hava araçları, otonom araçlar, sağlık kuruluşları, kameralar, sensörlerden oluşan bütün bir yapıyı nesnelere interneti sayesinde iletişim halinde tutarak adeta bir teknolojik ekosistem oluşturmayı başardı. Çin'de yaşanan bu deneyimlerin gösterdiği gibi sağlık hizmetleri sektöründeki dijital çözümlerin, hastane yönetimi ve teşhis uygulamalarında faydalı olduğu gösterilmiştir.

Öte yandan veriye dayalı dijital çözümler, örneğin mobil cihazlarla izleme verileri ve akıllı şehir altyapı uygulamaları, yüksek hata payları ve kötüye kullanım riskleriyle birlikte gelmektedir. Bu nedenle veri güvenliği ve gizliliği politikalarına daha fazla önem verilmesi gerekmektedir. Her gün logaritmik olarak artan kişisel verilerin korunması, merkezi otoriteler, özel şirketler ele geçirilmesi birçok sorunu beraberinde getirecektir. Veri depolamaya odaklanırken veri paylaşımına ilişkin şeffaflık ve kullanıcı onaylarının alınmasına da odaklanılmalıdır. Çin'deki gelişmelerden elde edilen kanıtlar sabit bir veri paylaşımının veri ihlallerinin temel kaynaklarından biri olduğunu göstermektedir. Bu ülkelerde uygulama kullanımı genellikle zorunludur. Ancak yasal kısıtlamalar uygulandı ve çevrimiçi/çevrimdışı önlemler aktif ve şeffaf bir şekilde iletildi.

TEŞEKKÜR

Makale çalışması sırasında bilgi ve tecrübeleriyle desteklerini esirgemeyen değerli hocam Dr. Durmuş Özdemir'e teşekkürlerimi sunarım.

ÇIKAR ÇATIŞMASI

Yazarlar arasında ve ilgili kurumları arasında herhangi çıkar çatışması olmadığını bildirmişlerdir.

ETİK KURALLARA UYGUNLUK

Yazarlar bu makalenin etik kurul onayı veya herhangi bir özel izin gerektirmediğini beyan ederler.

KAYNAKLAR [REFERENCES]

- [1] <https://www.worldometers.info/coronavirus/>, Erişim Tarihi: 19 Aralık 2020.
- [2] Coronavirus Disease (COVID-19) Pandemic, WHO, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, Erişim Tarihi: 19 Aralık 2020.
- [3] Coronavirus (COVID-19), Hastalık Kontrol ve Önleme Merkezleri (CDC), 2020. <https://www.cdc.gov/coronavirus/2019-nCoV/index.html>, Erişim Tarihi: 19 Aralık 2020.
- [4] <https://www.google.com/covid19/>, Erişim Tarihi: 01 Ocak 2021
- [5] <https://www.bing.com/covid/>, Erişim Tarihi: 01 Ocak 2021
- [6] <https://www.whitehouse.gov/briefings-statements/white-house-announces-new-partnership-unleash-u-s-supercomputing-resources-fight-covid-19/>, Erişim Tarihi: 19 Aralık 2020.
- [7] <https://blog.arxiv.org/2020/03/30/new-covid-19-quick-search/>, 19 Aralık 2020.
- [8] <https://www.who.int/csr/resources/publications/surveillance/plague.pdf>, 20 Aralık 2020
- [9] Parıldar, H., "Tarihte Bulaşıcı Hastalık Salgınları", *Tepecik Eğitim ve Araştırma Dergisi*, 30:19-26, 2020.
- [10] <https://tr.euronews.com/2020/07/07/fotograflarla-1918-ismanyol-gribi-abd-salg-nlara-kars-kat-edilen-mesafeyi-sorguluyor>, 24 Mart 2021
- [11] <https://coronaboard.kr/en/>, Erişim Tarihi: 24 Mart 2021
- [12] Q. Pham, D. C. Nguyen, T. Huynh-The, W. Hwang and P. N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts", in *IEEE Access*, vol. 8, pp. 130820-130839, 2020, doi: 10.1109/ACCESS.2020.3009328.
- [13] Salman Y. G., "Transforming laparoendoscopic surgical protocols during the COVID-19 pandemic; big data analytics, resource allocation and operational considerations", *International Journal of Surgery*, Volume 80, Pages 21-25, ISSN 1743-9191, 2020, <https://doi.org/10.1016/j.ijso.2020.06.027>.
- [14] Z. Hu, Q. Ge, S. Li, L. Jin and M. Xiong, "Artificial intelligence forecasting of COVID-19 in China", *arXiv:2002.07112*, 2020,
- [15] IBM Releases Novel AI-Powered Technologies to Help Health and Research Community Accelerate the Discovery of Medical Insights and Treatments for COVID-19, 2020. <https://www.ibm.com/blogs/research/2020/04/ai-powered-technologies-accelerate-discovery-covid-19/>, Erişim Tarihi: 01 Ocak 2021
- [16] <https://covid19-hpc-consortium.org/>, Erişim Tarihi: 01 Ocak 2021
- [17] <https://www.cov-irt.org/>, Erişim Tarihi: 01 Ocak 2021
- [18] <https://web.archive.org/web/20191009032639/https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/>, Erişim Tarihi: 01 Ocak 2021
- [19] <https://merics.org/en/report/tracing-testing-tweaking>, Erişim Tarihi: 25 Aralık 2020



- [20] <https://www.cbronline.com/news/china-to-roll-out-temperature-taking-infrared-cameras>, Erişim Tarihi: 01 Ocak 2021
- [21] <https://www.imf.org/en/Topics/imf-and-covid19/Fiscal-Policies-Database-in-Response-to-COVID-19>, Erişim Tarihi: 01 Ocak 2021
- [22] Demirdöğmez, M., Taş, H., Gültekin, N., “Koronavirüs’ün (Covid-19) E-Ticarete Etkileri”, *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 16 (29), 1907-1927, 2020. DOI: 10.26466/opus.734477
- [23] <https://business.blogthinkbig.com/the-big-data-and-iot-applications-fighting-coronavirus/>, 02 Aralık 2020
- [24] <https://towardsdatascience.com/coronavirus-a-big-data-lesson-from-south-korea-5bb703b8b0ae>, 02 Aralık 2020
- [25] Rastogi, Y.R., Sharma, A., Nagraik, R., Aygün, A., Şen, F., “The Novel Coronavirus 2019-Ncov: Its Evolution and Transmission into Humans Causing Global Covid-19 Pandemic”, *International Journal of Environmental Science and Technology*, 17, 4381–4388., 2020. doi:10.1007/s13762-020-02781-2
- [26] A. A. Hussain, O. Bouachir, F. Al-Turjman and M. Aloqaily, “AI Techniques for COVID-19”, *IEEE Access*, vol. 8, pp. 128776-128795, 2020, doi: 10.1109/ACCESS.2020.3007939.
- [27] Robinson, K., “A False Promise of Covid-19 ‘Big’ Health Data? Health Data Integrity and The Ethics and Realities of Australia’s Health Information Management Practice”, *Health Information Management Journal*, 50, 9–12., 2021. doi:10.1177/1833358320941190
- [28] P. Yu, Z. Xia, J. Fei and S. K. Jha, “An Application Review of Artificial Intelligence in Prevention and Cure of Covid-19 Pandemic”, *Computers, Materials & Continua*, vol. 65, no.1, pp. 743–760, 2020.
- [29] L. Akin and M. G. Gözel, “Understanding Dynamics of Pandemics”, *Turkish Journal of Medical Sciences*, vol. 50, no. SI–1, pp. 515–519, 2020.
- [30] J. Fan, B. D. Hambly, and S. Bao, “The Epidemiology of COVID-19 in the Gansu and Jinlin Provinces, China”, *Frontiers in Public Health*, vol. 8, 2020.
- [31] T. Alamo, D. Reina, M. Mammarella, and A. Abella, “Covid-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic”, *Electronics*, vol. 9, no. 5, p. 827, 2020.
- [32] Y. Wang, J. Li, X. Zhao, G. Feng, and X. (Robert). Luo, “Using Mobile Phone Data for Emergency Management: A Systematic Literature Review”, *Information Systems Frontiers*, vol. 22, no. 6, pp. 1539–1559, 2020.
- [33] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, et al., “Deep Learning-Based Detection For COVID-19 From Chest CT Using Weak Label”, *medRxiv*, 2020.
- [34] A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev, D. S. Bezrukov, et al., “Potential 2019-nCoV 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches”, *Insilico Med. Hong Kong Ltd A*, vol. 307, no. 2, pp. E1, 2020.
- [35] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)”, *IEEE Access*, vol. 6, pp. 52138-52160, 2018.
- [36] <https://www.weforum.org/agenda/2020/04/how-next-generation-information-technologies-tackled-covid-19-in-china/>, Erişim Tarihi: 01 Ocak 2021
- [37] <https://www.bbc.com/turkce/haberler-dunya-52638919>, Erişim Tarihi: 01 Ocak 2021

RESEARCH ARTICLE

Contemplation of the Photonics Beam to Complement the Water Cycle in Gulf Region

*  Imadeldin Elsayed Elmutasim,¹  Izzeldin I. Mohd ²  Khalid Hamid Bilal

¹University Malaysia Pahang ,26600 Pekan, Malaysia. emadcts@yahoo.com, Orcid. 0000-0002-4984-3522

²University Malaysia Pahang, 26600 Pekan, Malaysia. izzeldin@ump.edu.my, Orcid.0000-0003-0348-6811

³University of Science and Technology, Omdurman, Sudan, khalidhamidk9@gmail.com, info@ust.edu.sd

HIGHLIGHTS

- The significant of the based on the fresh look to utilise the laser beam.
- The capabilities of the laser provies a potential to solve the environmental issue.
- The main contribution is a novel study that open the door for the companies to to think through unprecedented technical wise.
- To evaluate the potential of suitable wavelengths that deal with the absorption in the atmosphere. study and obtained results why is important.

Keywords:

- Elevated duct
- Refractive profiles
- Propagation facto
- Fading
- Loss

Article Info:

Received : 01.06.2021

Accepted : 28.06.2021

Published : 21.12.2021

DOI:

10.5281/zenodo.5037867

*Correspondence:

Imadeldin Elsayed Elmutasim

emadcts@yahoo.com

+97455960722.

GRAPHICAL ABSTRACT

The unparalleled laser beam provides promising avenues of future research and continuation solutions concerning due to its unimaginable capabilities. One of the great advantages that could be achieved is complemented the water cycle particularly when gaseous factories waste concentrate in the manufacturing countries atmosphere where the high contamination reach to works as a layer in front of the normal water life cycle, thus the perforation process become indispensable using a specific instrument such as the laser beam.

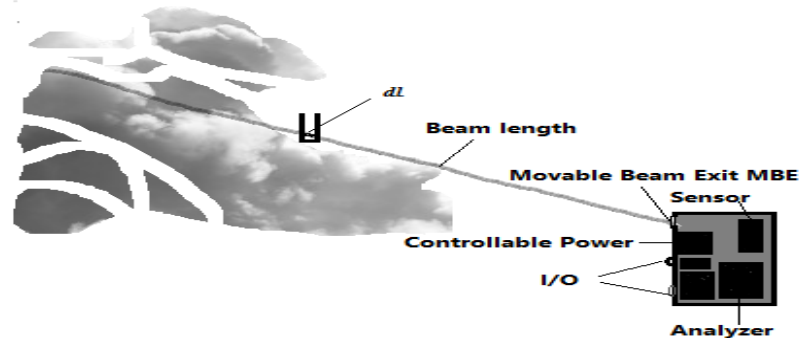


Figure 1. Optical beam station model to complement the water life cycle

Aim of Article: Enhance the water cycle in the Gulf region and mitigate impact rain outage through providing a technical applicable solution.

Theory and Methodology: The optical beam terminal put forward to shoot the loaded beam at a specific area to pass through the accumulation of pollution layers in the air. The beam initially controlled via parameters such as the power to produce a suitable heat among the layers, and the sensor to sense the layer contents, thus select a proper wavelength to avoid attenuation due to absorption and scattering.

Findings and Results: optical beam station proposed and discussed the capability to control, sense, and analyze in keeping with the atmospheric weather circumstance, the wavelength absorption gives the partial stability between 610 nm to 700 nm indicated that easy to see less disturbance in a visible wave then start to rise after that up to reach the peak at 1100 nm to 1150 nm.

Conclusion: The proposal gives a fresh look to the laser beam to provide a solution that could contribute to complete the water cycle through laser carrier carry the evaporation gases via pollution layer to reach the condensation layer and complete the process. The result showed the lowest wavelength prone to absorption between 610 nm up to 700 nm and from 1550 nm until 1600 nm. While the highest severely to the phenomenon around 1100 nm up to 1150 nm.



RESEARCH ARTICLE

Contemplation of the Photonics Beam to Complement the Water Cycle in Gulf Region

* Imadeldin Elsayed Elmutasim,¹ Izzeldin I. Mohd ² Khalid Hamid Bilal

¹University Malaysia Pahang ,26600 Pekan, Malaysia. emadcts@yahoo.com, Orcid. 0000-0002-4984-3522

²University Malaysia Pahang, 26600 Pekan, Malaysia. izzeldin@ump.edu.my, Orcid.0000-0003-0348-6811

³University of Science and Technology, Omdurman, Sudan, khalidhamidk9@gmail.com, info@ust.edu.sd

Citation:

Elmutasim, I.E., Mohd, I.I., Bilal, K.H. (2021). *Contemplation of the Photonics Beam to Complement the Water Cycle in Gulf Region*, Journal of Scientific Technology and Engineering Research, 2(2): 15-21. DOI:10.5281/zenodo.5037867

HIGHLIGHTS

- The significant of the based on the fresh look to utilise the laser beam.
- The capabilities of the laser provides a potential to solve the environmental issue.
- The main contribution is a novel study that open the door for the companies to think through unprecedented technical wise.
- To evaluate the potential of suitable wavelengths that deal with the absorption in the atmosphere. study and obtained results why is important.

Article Info

Received : 01.06.2021

Accepted : 28.06.2021

Published : 21.12.2021

DOI:

10.5281/zenodo.5037867

*Corresponding Author:

Imadeldin Elsayed

Elmutasim

emadcts@yahoo.com

+97455960722.

ABSTRACT

Photonic beam has attracted massive researchers' attention from the time of invention in the earlier decades and would continue due to unimaginable characteristics that contributed potentially to reach the highest data transmission rate and remarkable bandwidth in addition to the less susceptible to the environmental parameters via hundreds of kilometers. This advantage gives to think about it to address the water cycle issue in the shortage water region such as Gulf countries. The proposal considers the controllable beam properties that could carry the evaporation gas to pass through the air pollution layers and complete the condensation process hence reinforces an increase in rainfall chances, and lead to reducing the desalination process as well cost-effectiveness. Accordingly, the photonics beam station could sense and analyze the pollution layer to make a proper decision based on the data variability. The paper takes into account the matter via suggesting a new model that accommodates the significant environmental parameters in the desert area particularly in Qatar which is severely prone to the water lake due to salinity and stark weather in addition to the oil and gas manufacturing, which add a heavy layer in the atmosphere. So, this proofing would help the government sector strikingly the meteorology, energy, and resources to mitigate the impact of water shortage and provide a technical solution that would reflect in the whole countries economy. The work also attention to the absorption phenomenon for the most significant available wavelengths such as 850,1330 and 1550 nm and the results demonstrated the lowest wavelength prone to absorption between 610 nm up to 700 nm and from 1550 nm until 1600 nm.

Keywords: Elevated duct , refractive profiles, propagation factor, fading , loss

I. INTRODUCTION

In the recent decade, revolution growth has been noticed in the advanced communication technologies due to increasing usage of cloud, live streaming, and video conferencing as a whole represents a part of the

high-speed internet service. The customer demands need endless capacity and massive bandwidth to fulfill modern life needs. The limitation of the Radio Frequency RF carrier in addition to exceedingly challenges even when thinking about a rising to higher



frequencies such as millimeter-wave [1,2], which could provide a vast bandwidth but with susceptible to the climate atmospheric, hence to accommodate that, they give a line of thinking to shift to the optical carrier instead of Rf. However, the optical carrier has various advantages such as spectrum restriction and licensing in addition to high capacity and inconceivable bandwidth to carry the data, even the power consumption is quite better than the RF [3]. For these features and more, the narrow directivity beam promising to do lots in the upcoming digitalization and beyond. Our study based on the Photonic Beam PhB which is distinguished from others by the beam coherent passing long distances. Thus, this characteristic gives the edge to play a prime service to solve various challenges in many sectors such as natural resource particularly in the countries located in the desert area and suffering from the water shortage [4]. The industrial countries consider the water resources availability as part of national security, while the scarcity of natural water could push the governments to consume a huge financial to provide it. This work proposed laser photonics to accommodate the shortage of water as anticipated in the Gulf countries and particularly in Qatar which is a small peninsula located in the Arabian Gulf one of the highest seawater salinity globally [5] surrounded by water from three directions with stark desert climate led to totally depends on the desalination process when considering the water availability, considered a horrible situation to the government financially and environmentally. In addition to that, oil and gas manufacturing is taking into account the main contributor to air pollution and presents a negative impact on utterly environmentally as well as healthy [2,6]. This various pollution could increase the atmospheric layers and lead to blocking the water evaporation and hence outage in the water life cycle. The rare rain in the country becomes inconvenient for the farmers, birds, and a whole wildlife. This situation gives a unique opportunity to open a new chapter to use the proposal to accommodate the issue and provide a solution without compromise an environment.

Different examinations based on the optical link have been proposed to equalize the environmental challenges, for instance in [7] the authors measure atmospheric turbulence using a relay mirror for the three different laser beam that radiator for 350 km from the ground to the satellite, whereas in [8] the laser transmitter station from California and New Mexico

demonstrated distortion of the laser beam via atmospheric due to turbulence. Also, the comparison between theoretical and experimental has been achieved utilizing argon-ion laser [9] and the result has shown lognormal distribution in the received signal, whereas the adaptive optics was used to decrease the atmospheric turbulence through correct the leaving beam in single pulse that contains more than 100 photons [10]. Continuously to review direction, the laser has been used for the military purpose to provide data up to 30 Mbps using 5 W at 1064 nm between Earth and Mars [11].

These literature reviews open the door to focus on the laser beam capability to enhance the water cycle in the Gulf region and mitigate impact rain outage, in addition, to give thought to the desert climate fluctuation and providing a technical applicable solution. Furthermore, the various gas accumulation and highest salinity seawater with the elevated humidity due to rising temperatures provide a unique distinguishing aspect. The rest of the paper is organized as follows; in the following section, the optical beam characteristics in water cycle to describe the proposal thoughts. While the section 3 demonstrates the model approach; and in section 4 the analysis and discussion to obtain outcomes, then finally, summarize the work with recommendation for future work.

II. PHOTONIC BEAM PROPERTIES

The proposal of laser beam to implement the evaporation process and complete the significant missing stage in the water cycle based on the beam properties and environmental parameters for the specific area. Accordingly, the study focuses on the Arabian Gulf area; particularly in Qatar state which is considered one of the fastest-growing countries globally with the most significant oil and gas industry. Although it which located in a peninsula surrounded by water from three directions the woefully insufficient annual rainfall consider notable yearly [4]. In addition to that, the highest seawater salinity universally gives dramatically different environment. Further, the desert weather among the salt seawater there is fully qualified to achieve the task through the unique phenomenon. Whereas air pollution eliminates the water cycle process due to accumulation layers. The high-temperature whether by heat or electricity stimulating the atom to generate a charge and emerges the light, while the laser generates same the light through stimulating photonics of certain gases such as neon,



argon, and helium. Despite the laser beam is quite dangerous as it could be loaded to carry a massive amount of energy and become harmful, while the useful applications are undeniable starting from the highly accurate measurements, alloys, welding, and cutting iron until medicine and operations without bleeding. Whereas the famous revolution laser used lay in fiber optics telecommunication field. The high bandwidth, focusing, unlicensed spectrum, and far up security gives the laser beam utterly and unmistakably advantages from others. For instance, the sharp directivity laser beam in 10 kW is quite enough to cut a heavy glass plate [12,13] from laser beam consequences point of view. When the controllable beam heat interacts with the weather factors such as fog, it will give a positive action through a strike and passes the photonics heat hence make a duct depends on the beam directivity bandwidth [14]. The unparalleled beam properties could comfortably hit where the target is, and reaching out loaded by the various energy, data, and gases. The evaporation process needs an open water surface with a high temperature to change the state of the matter from the liquid to the gas [15]. Whereas to complement the water cycle, the beam should carry this gas and pass into the condensation layer in the cloud, that is based on the thickness of the gasses layers in the air which are work as the blocker. The next section will denote the manageable optical beam terminal invention which is contributing to the research in different areas such as meteorological, manufacturing, and ecology.

III. THE MODEL APPROACH

For a comprehensive understanding, the optical beam terminal put forward to shoot the loaded beam at a specific area to pass through the accumulation of pollution layers in the air. The beam initially controlled via parameters such as the power to produce a suitable heat among the layers, and the sensor to senses the layer contents, thus select a proper wavelength to avoid attenuation due to absorption and scattering [16]. The Beam specification B in frequency f_p loaded by energy to pass the pollution layer in the air vd , hence from the Doppler effect law [17]:

$$B = f_p \left(\frac{vp + vd}{vp} \right) \left(\frac{vp}{vp - vd} \right) \cdot j \quad (1)$$

and after calculations equation 2 has been obtained below:

$$B = f_p \left(\frac{vp + vd}{vp - vd} \right) \cdot j \quad (2)$$

where f_p is photon frequency, vp photon velocity in the beam, vd layer movement via the air, and j is model

factor which is equal to $n\varphi$ where n represents the refractive index when the beam transfers between different weather mediums and equal to the ratio between the speed of light in vacuum to the speed of light in medium and give clear result that higher reflective index of the medium at the nominator leads to light travel more slower in the medium, and φ is an atmospheric coefficient depends on the various weather parameters [18].

Due to distinguish weather aspects in the Gulf region, j should consider as the key role to accurate the model. The stark humidity and unimaginably temperature mixed with factories cement, oil, and gas in the global dust belt, can form a tremendous air layer blockage leads to prohibitive the evaporation gas pass to higher. If consider the pollution layer at the θ , and the probability to pass through it σ equal to:

$$\sigma = \theta n\varphi \quad (3)$$

$$n\varphi = \sigma / \theta \quad (4)$$

where σ is the chance to push through, θ air pollution thickness layer, and $n\varphi$ equal to j which is the model factor. Hence, the equation could be valid for the turbulence weather with consider:

$$B = \frac{\sigma f_p}{\theta} \left(\frac{vp + vd}{vp - vd} \right) \quad (5)$$

while from the Column law which describe the charge q along the beam length l [19].

$$\lambda = q / l \quad (6)$$

where λ is a linear density that will help to calculate to the small beam photon

$$dq = \lambda dl \quad (7)$$

The electricity filed E that could strike the layer θ from the charge (photon) dq can calculate is:

$$dE = n\varphi dq / l^2 \quad (8)$$

Hence

$$= n\varphi \lambda dl / l^2 \quad (9)$$

$$= \int n\varphi \lambda dl / l^2$$

$$= n\varphi \lambda \int l^{-1} dl$$

$$= n\varphi \lambda / l$$

where $n\varphi$ (j) is the constant model factor and represents the refractive index in atmospheric weather, dq is one charge (photon) pick up from the laser beam and the l

is the length of the beam. While in agreement with the charge and electricity field could calculate the force F based on the Newton law [20].

To express in a certain manner, the next figure 1 denotes the mini directional optical beam terminal when hit an accumulation layer in the atmosphere.

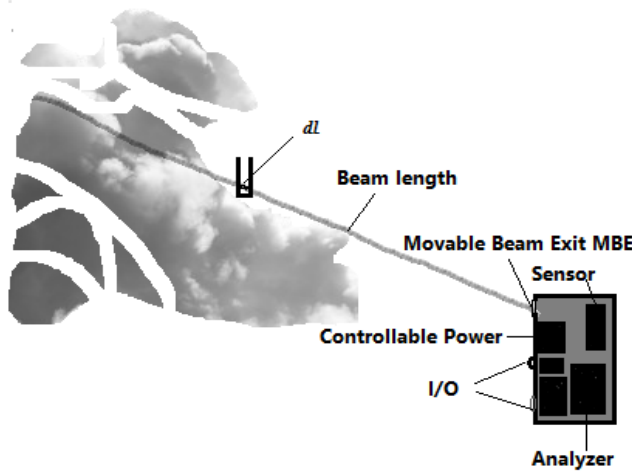


Figure 1. Optical beam station proposed for water life cycle

From equation 5 prospective, the momentum inversely proportional with the wavelength which leads to increasing the energy that increase the frequency and decreasing the wavelength, thus raise the loaded photonics energy as well as ability to pass a cross the air pollution layers, while in equation 9 the relation between electricity field E and the beam length l demonstrates also the inversely correlation when hit the layer in the atmosphere.

As reported in the beam terminal parts which is consist of the:

- a- Analyzer:
It works like a microcontroller in the electronic devices through makes an order such as increase or decrease the power after analysis the operations.
- b- Sensor:
To sense the gases and other air pollution accumulation in the air layers and provides the feedback to take the decision which it could be contents of Light Detection and Ranging (LiDAR) to capable the vision from the air environment.
- c- Movable Beam Exit:
Through controllable movement, the beam carries the evaporation gases and hit the air

layer via appropriate carrier. It might be directional and concentrated on the specific point to achieve the target based on the wavelength and the aperture diameter.

- d- Controllable Power:
Supplies the power depends on the layer formation in the air; as well as through it could be mitigating the noise via provide an appropriate narrow beam power.
- e- I/O:
To transfer the data when make connection with the other peripheral devices.

Concerning the atmospheric parameter, the figure 2 shows the significant beam wavelength characteristic from the absorption prospective [10], which is assist the photonics engineers to select the proper wavelength and get the flawless system, while table 1 below demonstrated the calculation correlation of the absorption phenomenon for given wavelengths.

Table 1: Relation between wavelength and Absorption

Wavelength in nm	Absorption dB/Km
500	0.15
550	0.13
600	0.03
650	0
700	0.01
750	0.1
800	0.21
850	0.41
900	0.60
950	0.70
1000	0.92
1050	1.02
1100	1.1
1150	1.1
1200	1.01
1250	0.9
1300	0.78
1350	0.58
1400	0.38
1450	0.2
1500	0.1
1550	0.01
1600	0

Accordingly, Matlab software has been used to simulate the absorption phenomenon in the light wave range applied, whereas the next section will discuss the result obtained.

V. CONCLUSION

This work presented to accommodate the outage of the water resource in Gulf region and Qatar particularly. The proposal gives a sound to the laser beam station to provide a solution that could contribute to complete the water cycle through laser carrier carry the evaporation gases via pollution layer to reach the condensation layer and complete the process. Due to that, study the light wavelengths consider extremely important to avoid the attenuation phenomenon such as absorption. The result showed the lowest wavelength prone to absorption between 610 nm up to 700 nm and from 1550 nm until 1600 nm. While the highest severely to the phenomenon around 1100 nm up to 1150 nm. Whereas, due to various climates and environmental, the model factor considered crucial to the effectiveness as it considers the refractive index and atmospheric coefficient. The future work will cover the Ozone layer influences when depletion due to manufactured chemicals

CONFLICTS OF INTEREST

The authors of the study have no any conflicts of interest to declare.

RESEARCH AND PUBLICATION ETHICS

In the studies carried out within the scope of this article, the rules of research and publication ethics were followed.

REFERENCES

- [1] I. E. Elmutasim and I. I. Mohd, "Examination rain and fog attenuation for path loss prediction in millimeter wave range," in *Lecture Notes in Electrical Engineering*, 2021, vol. 666, pp. 935–946, doi: 10.1007/978-981-15-5281-6_67.
- [2] L. Sfică, M. Bulai, V. A. Amihăesei, C. Ion, and M. Ștefan, "Weather conditions (With focus on uv radiation) associated with covid-19 outbreak and worldwide climate-based prediction for future prevention," *Aerosol and Air Quality Research*, vol. 20, no. 9, pp. 1862–1873, 2020, doi: 10.4209/aaqr.2020.05.0206.
- [3] S. Pan and Y. Zhang, "Microwave Photonic Radars," *Journal of Lightwave Technology*, vol. 38, no. 19, pp. 5450–5484, 2020, doi: 10.1109/JLT.2020.2993166.
- [4] I. E. Elmutasim and I. I. Mohd, "Investigate the Electromagnetic Waves to Desalinate Gulf Water and beyond," in *2020 7th International Conference on Frontiers of Industrial Engineering, ICFIE 2020*, 2020, pp. 119–122, doi: 10.1109/ICFIE50845.2020.9266726.
- [5] K. A. K. Charfeddine Lanouar, Afnan Yousef Al-Malk, "Air Pollution in Qatar: Causes and Challenges," *White Paper*, vol. 1, no. 3, pp. 1–7,

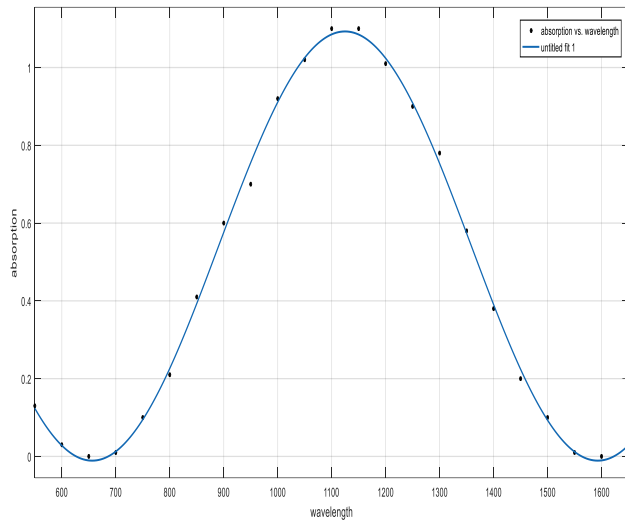


Figure 2: Absorption phenomenon in most common waves

IV. RESULTS AND DISCUSSION

In line with a scientific method, the optical beam station proposed and discussed the capability to control, sense, and analyze in keeping with the atmospheric weather circumstance. While the prediction of the electrical field effects to estimate the force which could apply on the respective air layer.

The specific photon beam has been chosen because of the availability as well as accommodating the most significant wavelengths which are used in remarkably light techniques such as 850, 1310, and 1550 nm. From the beam wavelength analytical point of view, The first slop appeared which indicates the lower absorption from the start of the wavelength range which is 500 nm up to reach 700 nm, while getting undeniable raise from 750 nm which is 0.1 dB continuously until reaching the peak at 1100 to 1150 nm which is 1.1 dB, then starts to slow down with remarkable decreasing into arriving in 1600 nm which is 0 dB and represents the same absorption value at 650 nm, while mathematically the value doesn't mean that is no absorption at all.

The partial stability between 610 nm to 700 nm indicated that easy to see less disturbance in a visible wave. The stability comes again from 1550 nm up to 1600 nm, which could propose a new wavelength to attention. The properties of the blockage layer determine the beam specifications such as power and wavelengths while the electric field surrounding the charge would generate the force that could pass through the layers and hence the accumulation of pollutants gases to reach the condensation layer and complete the process.



- 2016.
- [6] D. R. Roshan, M. Koc, A. Abdallah, L. Martin-Pomares, R. Isaifan, and C. Fountoukis, "UV index forecasting under the influence of desert dust: Evaluation against surface and satellite-retrieved data," *Atmosphere*, vol. 11, no. 1, 2020, doi: 10.3390/ATMOS11010096.
- [7] P. A. Lightsey "Scintillation in ground-to-space and retroreflected laser beams," vol. 33. no. 8, pp. 2535–2543, 1994.
- [8] K. E. Wilson and J. R. Lesh, "An overview of galileo optical experiment (GOPEX)." Tech Report: TDA progress Report 42-114, Communication Systems Research Section, NASA, 1993.
- [9] K. E. W. M. Jeganathan, "Preliminary analysis of fluctuations in the received uplink-beacon-power data obtained from the GOLD experiments." [Online]. Available: <https://www.toptechnoynews.com/>. [Accessed: 21-Mar-2021].
- [10] H. Kaushal and G. Kaddoum, "Optical Communication in Space: Challenges and Mitigation Techniques," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 1, pp. 57–96, 2017, doi: 10.1109/COMST.2016.2603518.
- [11] D. B. A. Biswas, "Mars laser communication demonstration: What it would have been." [Online]. Available: <https://www.toptechnoynews.com/?camp=AACV&on=&no=1&c=1>. [Accessed: 21-Mar-2021].
- [12] G. Wang *et al.*, "Highly efficient optical beam steering using an in-fiber diffraction grating for full duplex indoor optical wireless communication," *Journal of Lightwave Technology*, vol. 36, no. 19, pp. 4618–4625, 2018, doi: 10.1109/JLT.2018.2832200.
- [13] X. Sun *et al.*, "Field Demonstrations of Wide-Beam Optical Communications through Water-Air Interface," *IEEE Access*, vol. 8, pp. 160480–160489, 2020, doi: 10.1109/ACCESS.2020.3020878.
- [14] A. Garufo *et al.*, "Pulsed Photoconductive Connected Slot Array Operating at the Sub-mm Wavelength Band," *EuMIC 2019 - 2019 14th European Microwave Integrated Circuits Conference*, pp. 266–269, 2019, doi: 10.23919/EuMIC.2019.8909489.
- [15] W. Tang, S. Yueh, D. Yang, A. Fore, and A. Hayashi, "Investigating the utility and limitation of SMAP sea surface salinity in monitoring the arctic freshwater system," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2018-July, pp. 5647–5650, 2018, doi: 10.1109/IGARSS.2018.8519602.
- [16] E. A. Velichko, "Resonance scattering and absorption of light by a silver tube of large diameter and nanoscale thickness," *2017 IEEE International Young Scientists Forum on Applied Physics and Engineering, YSF 2017*, vol. 2017-January, pp. 104–107, 2017, doi: 10.1109/YSF.2017.8126599.
- [17] A. Semenova and A. Tikhomirov, "Feedforward carrier estimation method for bpsk with large Doppler shift," *BIHTEL 2014 - 10th International Symposium on Telecommunications, Proceedings*, 2014, doi: 10.1109/BIHTEL.2014.6987643.
- [18] W. Li, Y. Liang, and M. Fu, "The radio propagation affected by variation of refractive index," *Proceedings - 2009 3rd IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, MAPE 2009*, pp. 560–563, 2009, doi: 10.1109/MAPE.2009.5355602.
- [19] M. Zahn, "Derivation of the korteweg-helmholtz electric and magnetic force densities including electrostriction and magnetostriction from the quasistatic poynting's theorems," *Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, CEIDP*, pp. 186–189, 2006, doi: 10.1109/CEIDP.2006.312092.
- [20] M. Raya-Armenta, E. Alvarado-Mendez, and J. F. Gomez-Aguilar, "Dynamics of a charged particle in a ramp magnetic field," *Conference Proceedings - 12th International Conference on Laser and Fiber-Optical Networks Modeling, LFNM 2013*, pp. 95–96, 2013, doi: 10.1109/LFNM.2013.6644848.



RESEARCH ARTICLE

A Reliability Assessment of an Industrial Communication Protocol on a Windows OS Embedded PC for an Oil Rig Control Application

* Ogun Yücesan, ¹ Altan Özkil, and ² Mehmet Efe Özbek

*Atılım University, Modes, Ankara, Turkey, ogunyucesan@gmail.com, Orcid.0000-0003-2263-6803,

¹Atılım University, School of Civil Aviation, Ankara, Turkey, altan.ozkil@atilim.edu.tr, Orcid.0000-0001-8136-6087,

²Atılım University, School of Nat. & App. Sci., EEE, Ankara, Turkey, efe.ozbek@atilim.edu.tr, Orcid.0000-0001-5216-7062

HIGHLIGHTS

- Investigates Reliability vs Load in Window Computer
- Processor consumption been investigated in literature but “Reliability vs. Load” is rarely investigated.
- Measures reliability of Data History Polls and monitoring
- Such a reliability vs. load is observed on physical mockup, findings are hard to predict and close to actual scenarios.

Keywords:

- Plug and Produce
- OPC UA
- Industry 4.0
- M2M Communications.
- MTBF

GRAPHICAL ABSTRACT

Reliability is an important criterion for the applications, especially when there are humans involved in the activity. There are a limited number of works addressing the reliability concerns of a digital industrial plant scenario with machine to machine (M2M) communications. Results on a windows embedded platform are rare as well. In this paper, we investigate the suitability of such devices for presenting the data of a digital oil rig to the humans within the scope of the TÜBİTAK KAMAG Project. There can be two types of reads. A simple one regarding current values of observed variables. Also, a period of their occurrences can be recalled, referred to as a Historic read. An algorithm can consider both techniques. Our observations indicate: the simple reads are more resilient than the historical ones that represent past data of a readable variable.

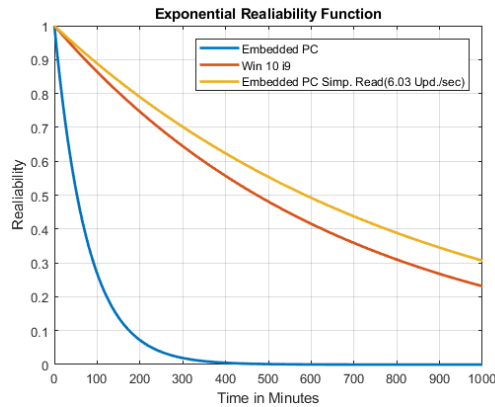


Figure A. The name figure or table about given info and results

Article Info:

Received : 14.07.2021

Accepted : 08.08.2021

Published : 21.12.2021

DOI:

10.53525/jster.971534

*Correspondence:

Dr. Ogun Yücesan
ogunyucesan@gmail.com

Aim of Article : Article illustrates a Mean Time Between Failures (MTBF) based method can be used to identify more reliable choice of design element.

Theory and Methodology: The techniques employed include development of an industrial communication client using a Prosys Open Platform Communications Unified Access (OPC UA). MTBF based method used to eliminate failures. To choose among the suitable techniques of accessing data, experiments were conducted on a physical connection between the server and a desktop computer.

Findings and Results: Polling past history of a data parameter is less reliable than monitoring it.

Conclusion : The algorithms used during development of the data monitoring activity need to keep the past of parameters at the server side not allowing the depletion of precious CPU resources and taking advantage of more reliable less burdensome simple data transfers.



RESEARCH ARTICLE

A Reliability Assessment of an Industrial Communication Protocol on a Windows OS Embedded PC for an Oil Rig Control Application

* Ogun Yücesan, ¹ Altan Özkil, and ² Mehmet Efe Özbek

*Atılım University, Modes, Ankara, Turkey, ongunyucesan@gmail.com, Orcid.0000-0003-2263-6803,

¹Atılım University, School of Civil Aviation, Ankara, Turkey, altan.ozkil@atilim.edu.tr, Orcid.0000-0001-8136-6087,

²Atılım University, School of Nat. & App. Sci., EEE, Ankara, Turkey, efe.ozbek@atilim.edu.tr, Orcid.0000-0001-5216-7062

Citation:

Yücesan, O., Özkil, A., Özbek, M.E., (2021). *A Reliability Assessment of an Industrial Communication Protocol on a Windows OS Embedded PC for an Oil Rig Control Application*, Journal of Scientific Technology and Engineering Research, 2(2): 22-30.
DOI: 10.53525/jster.971534

HIGHLIGHTS

- Investigates Reliability vs Load in Windows Computer
- Processor consumption been investigated in literature but “Reliability vs. Load” is rarely investigated
- Measures reliability of Data History Polls and monitoring
- Such a reliability vs. load is observed on physical mockup, findings are hard to predict and close to actual scenarios

Article Info

Received : 14.07.2021
Accepted : 08.08.2021
Published : 21.12.2021

DOI:

10.53525/jster.971534

*Corresponding Author:

Dr. Ogun Yücesan
ongunyucesan@gmail.com

ABSTRACT

Reliability is an important criterion for the applications, especially when there are humans involved in the activity. There are a limited number of works addressing the reliability concerns of a digital industrial plant scenario with machine to machine (M2M) communications. Results on a windows embedded platform are rare as well. Extension to existing literature includes reliability measurements with respect to processor load increase. In this paper, we investigate the suitability of such devices for presenting the data of a digital oil rig to humans. There can be two types of reads. A simple one regarding current values of observed variables. Also, a period of their occurrences can be recalled, referred to as a Historic read. For the review made by Geological teams for residues of ore, an algorithm can consider both techniques. Our observations indicate: the simple reads are more resilient than the historical ones that represent a data past of a readable variable.

Keywords: Plug and Produce, OPC UA, MTBF

I. INTRODUCTION

In Industrial Automation, Internet of Things (IoT), and Machine to Machine (M2M) applications, special purpose machines need to collaborate and communicate with each other. The success of these applications relies on Embedded or Industrial PCs.

Among the previous studies about performance bottleneck identifications, Burger et. al. [1] indicates the CPU resources deplete before memory or other resources. This work is extended by considering inner server load increases rather than load coming from network with employment of Confidence intervals. Investigation of the reliability comparison with respect to processor load can be seen as an extension to this work. For this purpose, Mean Time between Failure



(MTBF) is employed. This is inversely proportional to the failure rate which is a fundamental parameter in reliability calculations. Cavalieri (et.al) [2] studies the end-to-end delay performance of OPC UA client-server applications, where Transport Control Protocol (TCP) and Simple Object Access Protocol (SOAP/HTTP) are interchangeably used with and without the employment of encryption and security certificates. One of their main conclusions is that security comes at a cost of higher delay. Eckhardt et al. [3] studies the round-trip time of OPC UA Server and client architecture running on Xilinx Embedded PC boards with IEEE Time Sensitive Network (TSN) capabilities. Their results show that milliseconds RTTs can be achieved, so that delay contributions of the sender, network and receiver are comparable with the same order of magnitude. Cenedese et. al. [4], Morato et.al.[24], [22] compares the RTT performances of open source OPC UA servers and clients, where open62541 is identified as more efficient compared to C++ and Python based implementations of the same free OpcUa software. Later work includes more varieties, including a result with Prosys Java OPC UA implementation. Kim et al. [5] indicates that demanding more data by decreasing query interval below a certain value does not further improve the amount of queried and read data per second. Unix, Linux and Windows seem to be equally open to the Denial-of-Service attacks (DoS) which have been studied by Neu et. al. [6] for OPC UA. Garcia et. al. [7] work on incorporating OPC UA to IEC 61499 which is a standard promoting interoperability for distributed industrial automation systems. Their work is concentrated on Oil and Gas industries indicating the need for reliable communication in a harsh environment of this type. [23] Yucesan et. al. includes some generic testing methodologies incorporating MTBF.

Elfeham et. al. [8], Zhang et. al. [9], Hoffmann et. al. [10], Grüner et. al. [11] are among those who make fundamental extensions to OPC UA, Horrmann et. al. [12], Tu et. al. [13], Oksanen et. al. [14], Jo et. al. [15], Kim et. al. [16], Latif et. al. [17], Garcia et. al. [18], Wen et. al. [19] studied field integration of OPC UA, Cho et. al. [20] are among studies where the performance and power consumption are investigated, Kim et. al. [16] investigates backward compatibility of OPC UA.

These works open the way forward for OPC UA. They all make the OPC UA more understandable and usable. This study involves a continuous block of data

transfers rather than a single value, which is called Historic Data Access (HDA). Apart from the above studies, we develop a reliability model of HDA on COTS Windows OPC UA platform. Comparisons to a single read and a better hardware platform for the software type HDA experiments are conducted based on reliability as well. The results obtained allows for design of more reliable systems. TÜBİTAK KAMAG project is a project about construction and development of control mechanisms for a Native Oil Rig. Obtained results will be useful in development of monitoring and Human Machine Interface (HMI) software.

The document follows with Section II, description of methodologies followed. In Section III Results are presented. Section IV includes the reliability model estimation. Section V Conclusion includes discussion and final remarks.

II. METHODOLOGY

The control application in the form of PLC codes is implemented offline using an automation software development kit (SDK) and loaded into SDK Embedded PC run-time environment. The PLC implementation is performed using limited license *MATLAB Simulink PLC Target* and *Structured Text* script for additional load generation. A real-time test-bed consisting of a COTS Windows OPC UA Server connected through Ethernet to a desktop computer hosting an OPC UA client is utilized for the experiments. For safe and successful operation of the rig, the hardware and software employed for the control application should not have a failure during the operation.

The number of variables which are monitored by the control application is an important parameter determining the amount of the CPU power consumption. By varying the number of variables active on the OPC UA Server, its effects on CPU utilization, Memory utilization and Reliability are studied. The number of successful read/write experiments to the first failure from the beginning of an experiment is recorded for analysis. The mean of these recorded observations represents MTBF, which is a fundamental parameter for reliability predictions. In addition to reading/writing a single value of a parameter, there is a capability that allows one to read the historical values of any parameter. Single reads/writes are observed to be more resilient than historical reads indicating that the reliability bottleneck

is the historical read activity. OPC UA client SDK is modified to provide a Human Machine Interface for reading variables from the OPC UA Server. The number of variables observed from the client side is kept constant as two (2) while the number of variables on the OPC UA Server is varied from 12 to 152 for a controlled experimentation.

A. Test Bed

The Test-bed includes one client and one server connected through 100 Mbps Local Area Network of the university. Here, the server is relatively low-capacity Intel 1.3 GHz Single Celeron Processor, 2GB Ram, *Windows 7 Embedded Real Time Operating system*, passive cooled "Embedded PC". A COTS Windows based OPC UA server runs on this platform and communicates through the Ethernet ports. In test-bed, there is also a "Desktop computer" hosting the modified Prosys OPC UA client software. The desktop computer has Intel quad core i7, 3.4 GHz clock rated processor with 16GB RAM. This computer also keeps the Programmable Logic Controller (PLC) code which is loaded to Embedded PC run-time on the server. The PLC code defines the variables to be observed. To increase the testing coverage, a limited licensed version of a MATLAB Simulink target is employed to convert Simulink blocks to PLC code.

The Software infrastructure as depicted in Fig. 1 consists of COTS lightweight components for Embedded PC while all processing intensive user interface for coding and/or Simulink is performed on Win 10 Desktop. When you need to run the developed code you can transfer it to the embedded platform saving from the processing power.

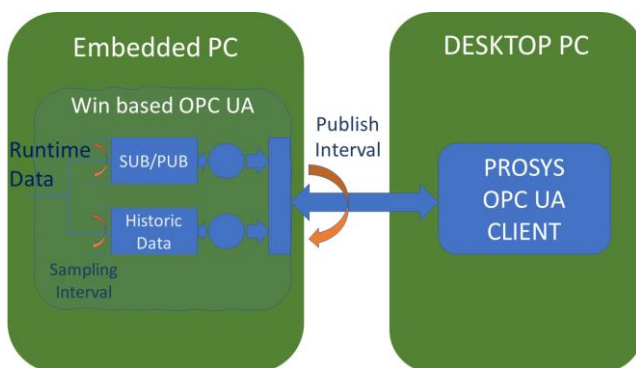


Figure. 1. The Experimental Software Setup.

Once a variable is subscribed, it is read at a regular interval called *publish interval* which is determined by the client every time subscription is made. It is also possible to monitor the working flow of the PLC code

variables from Desktop while they are running on Embedded PC. This brings an extra load on the Network, Memory, and CPU, from which we observe their effects in our experiments. This capability can be turned on and off with respective referring names as "logged in" and "logged out" activity. After the operation is started, the PLC run-time exchanges information with the COTS OPC UA server on Embedded PC which can be connected from the OPC UA client. By increasing the number of "integer variables" from the set {12, 52, 102, 152}, defined in the PLC code, the CPU and Memory are overloaded for experimentation. To test the effect of Embedded PC Server hardware on reliability, we utilize Intel i9 platform @2.9 GHz with 32 GB RAM with 64 bit Windows 10 Home OS while Prosys Client code runs on the same Desktop.

B. Sample Size

During the initial phases of the study, the limitations and resources, represented in the sense of a metric that can be employed both for guiding the programming effort and purchase of the equipment were not that much present. A windows tool with research results is also rare in the literature. Our aim was to identify worst reasonable performing number of variables as a metric representing the load on Embedded PC. This knowledge would be a hypothetical threshold for number of variables, which should not be surpassed or if had to be surpassed hardware upgrades or algorithmic improvements should be made. A coarse increment of counters is handy for such knowledge's understanding. For this aim, Double Integers (DINT) counters of Structured Text coding is employed. These are 32-bit registers. The lowest value they can represent is -2147483648 and the highest is 2147483647. As per computer CPU hardware, an update in value of a counter means first data is taken to the register then value updated on CPU, later re-sent back to memory location. Therefore, counters continuously are upgraded, which we hope will fully represent worst case, in the sense of heaviest data update scenarios. Our counters are mostly changing between -15000 to 25000. They are incremented with different values from 1 to 11 and sometimes 100 as well. This way, simultaneous logic operations are prevented presenting a more average case scenario.

At first, we started from small steps trying to see the limitations. Activities were conducted in "logged-off" manner. Considering the Antivirus burden of the CPU



no scan activity is allowed during experimentation, which is started after upgrades and scan activities were completed.

In the case of expeditions, gradually starting from 12 variables, which we increased observing the effects till 152 variables. CPU utilization wasn't too different however at 152 variables Embedded PC started to demonstrate certain lags and delays where we stopped further increasing our number of counters or variables.

C. Sample Characteristics

During Read Count observations, we observed that the experiments conducted by repeating the read experiments repetitions were a little low at first. However, during this phase we notice that the CPU on a passively cooled Embedded PC is affected by the room temperature in read counts sense. Oil rigs do have air conditioned, a high volume of cooled air flowing cabinets. We try to represent this by employing two fans. One is sucking hot air coming from the Embedded PC and the other is pouring fresh air of room temperature of roughly 22 to 24 degrees Celsius into the PC to the best of our abilities.

The CPU utilizations are irrelevant to the heat of CPU, however all Read counts and reliability observations are made under cooled environment. The simple read count is obtained during a non-cooled experiment yet still representative of its worst case, since if cooled would perform to a longer duration. It's *publish interval* is 10 ms, slightly above the single cycle completion time of a loop with two (2) variables on our Embedded PC, which we observed to be roughly around 5 to 7 ms. Therefore, it is representing its worst-case scenario with at best a marginal difference. By observations of worst-case scenarios, it is possible to find a more reliable method based on calculation. Scaling the results to certain usage profiles, it can highlight some expected reliabilities. In an Oil rig Control software monitoring scenario, under conditions with depleted CPU resources, the control software run can be interrupted causing interruption to crane controls. This can cause dangerous accelerations for crew working under heavy equipment. Whenever HMI for monitoring conditions have failed, it may cause certain parameters to go high without alerts. These may cause fires, explosions, catastrophic failures. Also, precious ore can be missed.

III. RESULTS

To model an oil rig environment one server on an Embedded PC and a client modified to needs on a Desktop PC was considered. One can read data from the server using two different methods: subscription based and historic data access readings. The latter is, as its name indicates, the history of the subscribed values.

A. Central Processing Unit (CPU) Utilization for Historic Data Access Reads

History reads/queries are used for monitoring the progress of an activity in an Oil Rig environment. For the conduct of the experiments, the history data is kept in Embedded PC Hard-disk with 4 samples/second and 10000 queue size. 7.5 minutes window of data is queried by Prosys Client. The number of variables observed in OPC UA Server is changed from 12 to 152. As the number of variables approaches 150, our tiny 1.4 GHz Celeron presents effects, like mouse pointer not moving smoothly and fluently. Averages and maximum utilization are taken over a 120 second interval while historic access queries are conducted with a 10 second period. This experiment is repeated five (5) times for statistical accuracy. All data is presented with 90% confidence intervals marking the uncertainty on them. The x-axis on figures presents the number of variables included in experimentation representing the load, while y-axis is the representative of the CPU utilization that is occurring with that load. In Figure 2, we observe average CPU utilization, where "logged in activity" resides slightly higher over "logged out activity". However, as the number of variables increases, the CPU utilization figures for logged out and logged in cases approach each other.

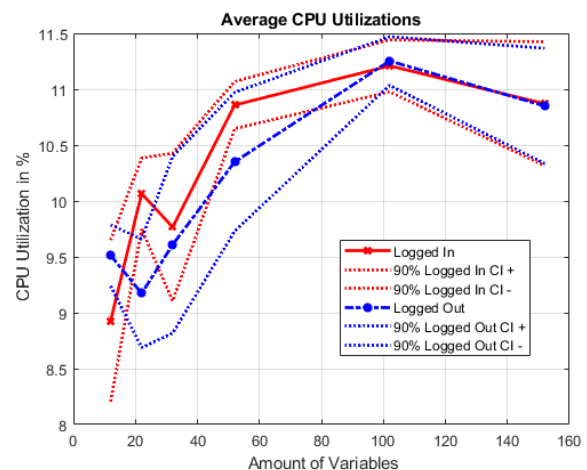


Figure 2. Average Utilization, both logged in and out.

The load created by historic data access is low considering the Ram capacity of the Embedded PC. The utilization figures reflect the load increases, and all of the average utilization measurements are within the 3% margin. The associated 90% confidence intervals are tight as depicted in the figure. The oscillations are similar for both logged in and out activities. These also are indicative of the lack of wide register resources of the Celeron CPU. There is a chance the tasks were not served on time but had to wait or were dropped in Random Access Memory (RAM) in a random manner.

The results were of the average utilization. As an outline of other observations regarding the hardware resources: The maximum CPU utilization is always within 4% margin and reaches to 100% levels when the number of the variables is around 50. This situation is visible in Figure 3. However, "Logged in" maximum utilization does not reach 100% level. These can be due to overwhelming job demands on a single core CPU with no waiting queue. The Confidence intervals in the figure can give clue for how much extra CPU demand exists. Even though, it is not possible to have more than 100% utilization, the overshoots indicate variabilities. In load levels, where CPU is more utilized, success rate of the tasks at hand should be high. It is because of the lesser variance in these regions. The levels, where CPU is not well utilized, are accompanied with high variances. It is reasonable to assume there are more off times with respect to rejected tasks in these loads. Therefore, overshooting confidence interval boundaries indicate higher chance for interruption to services.

The levels for memory, both logged in and out activity, are consistent with findings in literature in the sense that they are relatively flat. Logged in case requires slightly less memory, this may be due to the job demands returning to the RAM, to be accessed back in a random time. In cases where CPU utilization approaches 100%, there can be interruptions to other functions on the server.

B. Effect of Number of Variables on Read Counts

In this section, we report the history read counts, which are defined as the number of successful read attempts from the beginning of an experiment.

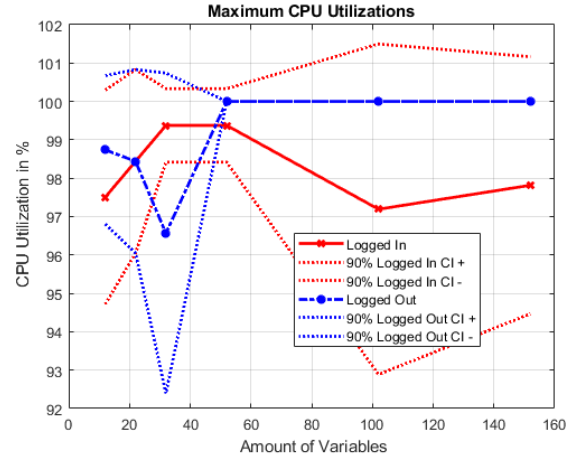


Figure 3. Average Utilization, both logged in and out.

Figure 4, shows the history read counts by varying the number of variables when the query period is set to 15 seconds. This selection of query period gives a chance for the client to present data on time, yet it is still among the lowest observed duration of reads from the previous section, namely worst-case. The mean of history read counts increases as the number of variables changes from 12 to 152. The COTS OPC UA server employed includes two program pieces to achieve its functions. One is called Direct Access (DA) Server and other is called Unified Access (UA) Server. Both are needed to achieve operation. DA server for OPC framework uses TCP for data transmission. UA server is the new server architecture that allows for various transmission control protocols. Even if there is no increase in network demand, internal communication between these two program pieces should be using TCP for communicating. TCP is known to have better latency performance under heavier load because of protocol overheads. Increasing trend can be because of such a situation. The standard deviation σ and confidence interval boundaries keep widening indicating the highest uncertainty at 152 variables. This result is intuitive due to the modest embedded computing resources.

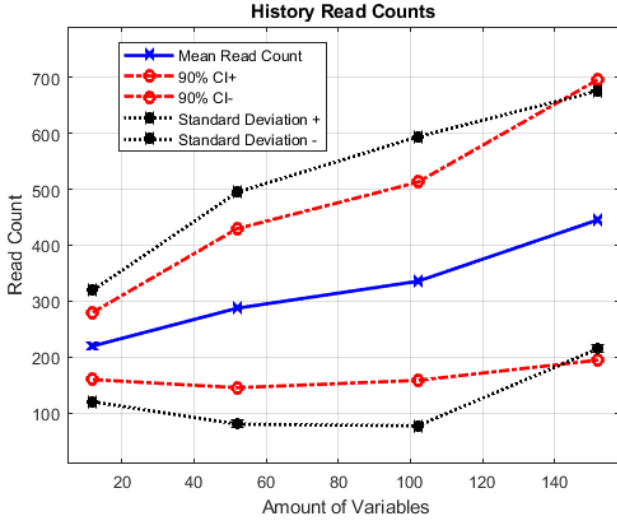


Figure 4 Effect of Increasing variables count on reliability

We carry out another experiment by moving the OPC UA server to a stronger hardware while keeping the Prosys client on the same desktop. The new server hardware is a 64bit Win10 OS working on i9 Intel platform. We observe that the stronger hardware provides 2732 successful history reads until its termination while Embedded PC provides lower than 500 read counts for all experiments.

C. Reading and Writing Data with Client/Server Subscription/Publishing

Up to this point, we have reported the results of Historic Data Access read. In the subscription-based method, a variable in the server side needs to be "subscribed" by the client for monitoring its updated values. For experiment in this section Publish Interval is set to 10 ms. The number of data read and write counts is reported as 1.163 Million and 764K, respectively while the PLC Runtime/OPC UA server is running on the Embedded PC. With respect to the observations made for the historic read counts, 1.163 Million is significantly higher. In another set of observations, we have observed 210675, 326690, 381121 where the experiment was terminated at the end of a duration. 6.03 updates are made every second during this set of experiments. The publish interval was again selected as 0.01 corresponding to 10ms. The average obtained from the second set is 306162 single value updates.

IV. RELIABILITY MODEL ESTIMATION

The observed read counts from the experiments are translated into time by multiplying with the query period Δ , where Δ is 15 seconds. This choice of the

query period is due to its closeness to the worst case read count achieved in time perspective. It also allows avoiding unnecessary processing delays in the client side. By using all the values obtained from Figure 4, the sample mean of history read counts is calculated as $\mu = 304$ Reads. This is translated into time as follows $\mu_t = \mu \times \Delta = 304 \times 15 \text{ seconds} = 4560 \text{ seconds} = 76 \text{ Minutes}$. In this study, we employ the exponential reliability function. Using the expectation of $E[X_{\text{Exponential}}] = 1/\lambda = \mu_t$ as defined in the book by Trivedi [21], the exponential reliability function can be expressed as in equation (1).

$$R(t_{\text{Exponential}}) = e^{-\lambda.t} \quad (1)$$

The same calculation for historic read count on OPC UA Server on Win 10 i9 PC yields the result of $\mu_{t-\text{win10}} = \mu \times \Delta = 2732 \times 15 \text{ seconds} = 40980 \text{ seconds} = 683 \text{ Minutes}$ in order to serve as a basis for a comparison.

We translate the second set of simple read counts into time by calculations based on updates per interval, this value is according to the observations made during, second set of values collection $\mu_{t-\text{sub}} = \mu \times \Delta = 306162 / 6.03 \text{ upd./sec.} = 50773.13 \text{ seconds} = 846.22 \text{ Minutes}$.

The estimated exponential reliability functions obtained from the above calculated values are shown in Figure 5. Highest reliability of history read count values come from strong i9 platform as depicted with a yellow line. Observation from subscription reads with orange lines are residing in the highest range. Reliability of the historic read counts is the lowest curve drawn with blue lines.

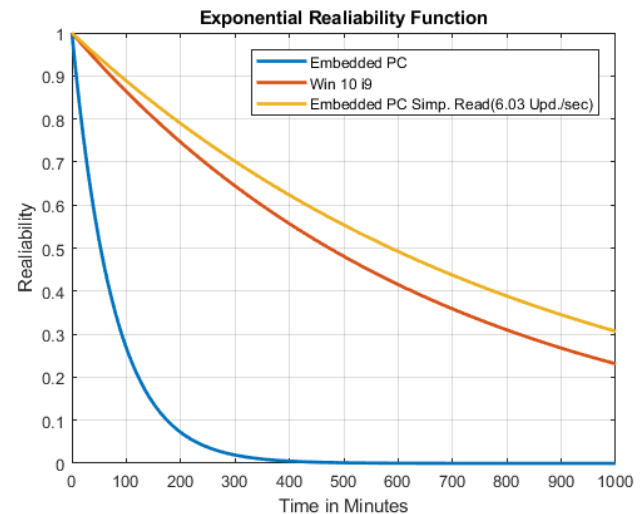


Figure 5. Reliability Estimate of the Embedded PC with 15 second Query Period



We can see that our model using historic read counts is a lower bound for the system reliability compared to subscription reliability model. From the figure, we can also observe that our software platform is capable of achieving higher reliability levels on stronger hardware compared to those observed in a limited resourced Embedded PC indicating correct functioning given broad resources.

V. CONCLUSION

In this paper, an Embedded PC, as an OPC UA communication server is considered. There are three scenarios. These consist of the historic access; the subscription reads and History Access over a stronger Intel i9 platform. The MTBF figures are employed in the classical exponential reliability models. The results indicate that the HDA activity has a lower chance of success. The same software on powerful i9 hardware yields better reliability.

Over the TUBİTAK KAMAG project, the results been reported to implementation team. The purchase of a stronger hardware was addressed as the writing date of the article. Reports to the development team indicating chances for overload with HDA access been provided to be considered during still ongoing effort. The outcomes are indicative of benefits of making physical experiments rather than using available predictions. The oscillations of CPU utilizations are characteristic with Celeron Processors, whereas increasing trend in MTBF vs load figure may be due to TCP achieving better with increasing load. Widening confidence intervals are useful in representing increasing uncertainties.

As a future work, further investigation of statistical distributions underlying the failure behavior can be considered.

CONFLICTS OF INTEREST

They reported that there was no conflict of interest between the authors and their respective institutions.

RESEARCH AND PUBLICATION ETHICS

In the studies carried out within the scope of this article, the rules of research and publication ethics were followed.

ACKNOWLEDGMENT

This research is supported by TUBİTAK (KAMAG) under the grant no: 115G007. Research is conducted for the needs of Turkish Petroleum Anonymous Partnership (TPAO) and Turkish Petroleum International Company (TPIC).

REFERENCES

- [1] A. Burger, H. Koziolok, J. Ruckert, M. Platenius-Mohr, and G. Stomberg, "Bottleneck identification and performance modeling of opc ua communication models," in Proceedings of the 2019 ACM/SPEC International Conference on Performance Engineering, ser. ICPE '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 231–242.
- [2] S. Cavalieri and F. Chiacchio, "Analysis of opc ua performances," *Computer Standards and Interfaces*, vol. 36, no. 1, pp. 165 – 177, 2013.
- [3] A. Eckhardt and S. Muller, "Analysis of the round-trip time of opc ua and tsn based peer-to-peer communication," in 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2019, pp. 161–167.
- [4] A. Cenedese, M. Frodella, F. Tramarin, and S. Vitturi, "Comparative assessment of different opc ua open-source stacks for embedded systems," in 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2019, pp. 1127–1134.
- [5] W. Kim and M. Sung, "Opc-ua communication framework for plc-based industrial iot applications: Poster abstract," in Proceedings of the Second International Conference on Internet-of-Things Design and Implementation, ser. IoTDI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 327–328.
- [6] C. V.Neu, I. Schiering, A. Zorzo. Simulating and Detecting Attacks of Untrusted Clients in OPC UA Networks. CECC 2019: Proceedings of the Third Central European Cybersecurity Conference. 2019, p.1-6.
- [7] M. V. Garcia, E. Irisarri, F. Perez, E. Estévez, and M. Marcos, "An openccps automation architecture based on iec-61499 over opc-ua for flexible manufacturing in oilgas industry", 20th IFAC World Congress. IFAC-PapersOnLine, vol. 50, no. 1, pp. 1231 – 1238, 2017
- [8] H. Elfaham, F. Palm, S. Gruner, and U. Epple, "Full integration of matlab/simulink with control application development using opc unified architecture," in 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), July 2016, pp. 371–376.
- [9] L. Zhang, "Specification and design of cyber physical systems based on system of systems engineering approach," in 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Oct 2018, pp. 300–303.
- [10] M. Hoffmann, C. Buscher, T. Meisen, and S. Jeschke, "Continuous integration of field level production data into top-level information systems using the opc interface standard," *Procedia CIRP*, vol. 41, pp. 496 – 501, 2016,
- [11] S. Gruner, J. Pfrommer, and F. Palm, "Restful industrial communication with opc ua," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1832–1841, Oct 2016.
- [12] R. Hormann, S. Nikelski, S. Dukanovic, and E. Fischer, "Parsing and extracting features from opc unified architecture in industrial environments," in Proceedings of



the 2nd International Symposium on Computer Science and Intelligent Control, ser. ISCSIC '18. New York, NY, USA: Association for Computing Machinery, 2018.

[13] N. T. T. Tu, N. D. Cuong, V. V. Tan, and H. Q. Thang, "Research and development of opc client-server architectures for manufacturing and process automation," in Proceedings of the 2010 Symposium on Information and Communication Technology, ser. SoICT '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 163–170.

[14] T. Oksanen, P. Piirainen, and I. Seilonen, "Remote access of iso 11783 process data by using opc unified architecture technology," *Comput. Electron. Agric.*, vol. 117, no. C, p. 141–148, Sep. 2015.

[15] G. Jo, S.-H. Jang, and J. Jeong, "Design and implementation of cpps and edge computing architecture based on opc ua server," *Procedia Computer Science*, vol. 155, pp. 97 – 104, 2019.

[16] J. Kim, G. Jo, and J. Jeong, "A novel cpps architecture integrated with centralized opc ua server for 5g-based smart manufacturing," *Procedia Computer Science*, vol. 155, pp. 113 – 120, 2019.

[17] H. Latif, G. Shao, and B. Starly, "Integrating a dynamic simulator and advanced process control using the opc-ua standard," *Procedia Manufacturing*, vol. 34, pp. 813 – 819, 2019.

[18] M. V. Garcia, E. Irisarri, F. Perez, E. Estévez, D. Orive, and M. Marcos, "Plant floor communications integration using a low cost cpps architecture," in 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2016, pp. 1–4.

[19] X. Wen, H. Chen, B. Wen, J. Liu, Y. Li, and N. Xi, "Conceptual framework of smart factory based on opc ua and lstm encoder-decoder," in 2018 IEEE 1st International Conference on Micro/Nano Sensors for AI, Healthcare, and Robotics (NSENS), Dec 2018, pp. 44–48.

[20] H. Cho and J. Jeong, Performance Evaluation of Industrial OPC UA Gateway with Energy Cost-Saving: Third International Conference, SmartCom 2018, Tokyo, Japan, December 10–12, 2018, Proceedings, 12 2018, pp. 55–66.

[21] K. S. Trivedi, *Probability and Statistics with Reliability, Queuing and Computer Science Applications*, 2nd ed. GBR: John Wiley and Sons Ltd., 2001.

[22] A. Morato, S. Vitturi, F. Tramarin and A. Cenedese, "Assessment of Different OPC UA Implementations for Industrial IoT-Based Measurement Applications," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021

[23] O. Yucesan, A. Ozkil, "Time Complexity Comparison of Stopping at First Failure and Completely Running the Test." *J Electron Test* vol. 36, pp. 409–417, 2020.

[24] A. Morato, S. Vitturi, F. Tramarin and A. Cenedese, "Assessment of Different OPC UA Industrial IoT solutions for Distributed Measurement Applications," 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 2020, pp. 1-6.



Industrial Robot Motion Planning Algorithms Performance Benchmarking

¹ Alim Kerem Erdoğan, ^{2*} Uğur Yayan

¹Inovasyon Mühendislik Ltd. Şti., R&D, Eskişehir, Turkey, kerem.erdogmus@inovasyonmuhendislik.com Orcid. 0000-0001-5111-5965
²Eskişehir Osmangazi University, Software Engineering Department, Eskişehir, Turkey, ugur.yayan@ogu.edu.tr Orcid. 0000-0003-1394-5209

HIGHLIGHTS

- Effect and important of this article in literature
- Exchange between sources in related subjects of this article
- Contribution and strongest impact on the related subject of this article
- Examined study and obtained results why is important

Keywords:

- Robotics
- Verification&Validation
- Trajectory Planning
- Industrial Quality Control

Article Info:

Received : 07.08.2021

Accepted : 17.09.2021

Published : 21.12.2021

DOI:10.53525/jster.979689

*Correspondence:

Uğur Yayan

ugur.yayan@ogu.edu.tr

Tel: +90 222 239 37 50

GRAPHICAL ABSTRACT

In robotics studies, motion and trajectory planning for industrial robot systems is one of the most actively studied topics. "Automated Robot Inspection Cell for Quality Control of Automotive Body-in-White (ROKOS)" has been defined as the Use-case Scenario within the scope of the VALU3S project, where studies are carried out for the verification and validation (V&V) of autonomous systems. In this study, ROKOS system, developed in the real environment, was transferred to the GAZEBO simulation environment for the V&V of the defined use case scenario, and SRVT (Simulation Based Robot Verification Testing Tool) was revealed and simulation-based tests of the evaluation scenarios created within the scope of the project were carried out on this system. Figure A shows the results of the Full Test with Reset.

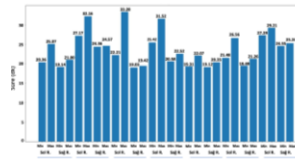


Figure A. Full test with reset results

Aim of Article: In this study, within the scope of "Safety Trajectory Optimization", which is one of the evaluation scenarios, it is aimed to verify and validate the software developed by creating trajectories in which the robot arms of the ROKOS system will move safely according to the determined evaluation criteria for the quality control of the bus chassis transferred to the simulation environment.

Theory and Methodology : The OMPL and EST trajectory planning algorithms of the ROS MoveIt tool were studied, and the pros and cons of these algorithms were determined. In order to verify these algorithms in the simulation environment, detailed tests were carried out on the usage scenario and the results were analyzed. Tests were conducted for 3 different scenarios (Speed Test, Full Test with Reset (FTR) and Full Test without Reset (FToR)) and validation activities were improved in terms of time and cost for the existing ROKOS system transferred to SRVT.

Findings and Results: Within the scope of these studies, more than 400 tests were carried out on many motion planning algorithms of the OMPL planner, and the most optimal planning algorithm was tried to be found. At the end of the study, it was determined that this planning algorithm should be BiEST or PRM and the tests were completed.

Conclusion : The SRVT system was created ROS tools, Gazebo and ROKOS. The study was carried out on the examination and comparison of the planning performances of the motion planning algorithms used through the software. In order to reduce the task completion times of the ROKOS integrated into the SRVT ecosystem, it has been focused on determining the most suitable trajectory planning algorithm for this system. Before moving on to these applications, many basic motion planning algorithms in the OMPL library were examined, and the operability and effectiveness of the motion planning mentality of these algorithms in the SRVT integration of the ROKOS system were studied. In the light of research and tests, it has been seen that the task completion times of the ROKOS system can be considerably reduced when appropriate algorithms are detected.



ARAŞTIRMA MAKALESİ

Endüstriyel Robot Hareket Planlama Algoritmaları Performans Karşılaştırması

¹ Alim Kerem Erdoğan, ^{2*} Uğur Yayan

¹Inovasyon Mühendislik Ltd. Şti., Ar-Ge Bölümü, Eskişehir, Türkiye. kerem.erdogmus@inovasyonmuhendislik.com, Orcid.0000-0001-5111-5965
²Eskişehir Osmangazi Üniversitesi, Yazılım Mühendisliği Bölümü, Eskişehir, Türkiye., ugur.yayan@ogu.edu.tr, Orcid.0000-0003-1394-5209

Alıntı :

Erdoğan, A.K., Yayan, U., (2021). Endüstriyel Robot Hareket Planlama Algoritmaları Performans Karşılaştırması, *Journal of Scientific Technology and Engineering Research*. (2021) - 2(2) : 31-45. DOI: 10.53525/jster.979689

ÖNE ÇIKANLAR

- Emniyetli Yörünge Optimizasyon kapsamında, simülasyon ortamına aktarılmış olan otobüs şasesinin kalite kontrolü
- ROKOS sistemine ait robot kollarının belirlenen değerlendirme kriterine göre emniyetli bir şekilde hareket edecekleri yörüngelerin oluşturulması
- ROS MoveIt aracına ait OMPL ve EST yörünge planlama algoritmalarının artı ve eksi yönlerinin belirlenmesi
- İncelenen algoritmaların simülasyon ortamında doğrulanmaları için kullanım senaryosu üzerinde detaylı testler

Makale Bilgileri

Geliş Tarihi : 07.08.2021

Kabul Tarihi: 17.09.2021

Yayın Tarihi: 21.12.2021

DOI:10.53525/jster.979689

*Sorumlu Yazar:

Uğur Yayan,

ugur.yayan@ogu.edu.tr

Tel: +90 222 239 37 50

ÖZET

Robotik çalışmalarında, endüstriyel robot sistemleri için hareket ve yörünge planlama, aktif olarak çalışılan konuların başında gelmektedir. Otonom sistemlerin doğrulanması ve onaylanması (V&V) için çalışmalar yürütülen VALU3S projesi kapsamında Kullanım Senaryosu olarak "Araç Şase Kalite Kontrolü için Otonom Robot Denetleme Hücresi (ROKOS)" tanımlanmıştır. Bu çalışmada, tanımlanmış kullanım senaryosunun doğrulanması ve onaylanması amacıyla gerçek ortamda geliştirilmiş ROKOS sistemi, GAZEBO simülasyon ortamına aktarılarak SRVT (Simülasyon Tabanlı Robot Doğrulama Test Sistemi) ortaya çıkarılmış ve bu sistem üzerinde ise proje kapsamında oluşturulan değerlendirme senaryolarının simülasyon tabanlı testleri yapılmıştır. Bu çalışmada, değerlendirme senaryolarından "Emniyetli Yörünge Optimizasyon" kapsamında, simülasyon ortamına aktarılmış olan otobüs şasesinin kalite kontrolü için ROKOS sistemine ait robot kollarının belirlenen değerlendirme kriterine göre emniyetli bir şekilde hareket edecekleri yörüngelerin oluşturularak geliştirilen yazılımların doğrulama ve onaylama işlemlerinin yapılması hedeflenmiştir. Bu kapsamda, ROS MoveIt aracına ait OMPL ve EST yörünge planlama algoritmaları üzerinde çalışılmış, bu algoritmaların artı ve eksi yönleri belirlenmiştir. Bu algoritmaların simülasyon ortamında doğrulanmaları için kullanım senaryosu üzerinde detaylı testler yapılmış ve sonuçları analiz edilmiştir. Testler 3 farklı senaryo (Hızlı test, Resetli Tam Test ve Resetsiz Tam Test) için yapılmış ve SRVT'ye aktarılmış mevcut ROKOS sistemi için doğrulama faaliyetleri zaman ve maliyet açısından iyileştirilmiştir.

Anahtar Kelimeler: Robotik, Doğrulama&Onaylama, Yörünge Planlama Algoritmaları, Endüstriyel Kalite Kontrol

I. GİRİŞ

Son yıllarda insan gücü ve aklı ile ilerlemekte ve gelişmekte olan endüstriyel çalışmalar için kullanılmaya başlanmış basit robotlar yerini karmaşık otonom robotlara bırakmakta, endüstrinin yanı sıra sağlık, uzay ve askeriye gibi kritik alanda da bu tarz robotlar ve robotik

uygulamaların desteklendiği ve yaygınlaştığı görülmektedir [19-22]. Bu çok hızlı gelişen yayılımın bir nedeni, robotların mükemmel çok yönlülüğü ve esnekliğidir. Bu esneklik onları farklı görevleri yerine getirmek için çok uygun hale getirmektedir [16]. Endüstriyel robotlarla ilgili olarak, bunları bir manipülatör ve eklemlerle [17] birbirine bağlanan bir dizi sert koldan oluşan mekanik bir yapı olarak tanımlamak mümkündür.

Hareket kabiliyeti sağlayan bir yapı, el becerisi veren bir bilek ve robotun kullanıldığı görevi yerine getiren bir uç efektör belirlenmesi ile endüstriyel alanda çalışan kompleks yapıda bir robotun temel altyapısı tanımlanmış olur.

Konu endüstriyel robotik sistemlere geldiğinde, sahadaki görevlerin tamamlanmasında zaman çok önemli bir faktördür. Sistemlere zaman kazandıracak ve görevleri sorunsuz gerçekleştirmelerini sağlayacak unsurlar da robotik kol sistemleri için kritik önem taşıyan yörünge planlama algoritmalarıdır. Yörünge planlama, hesaplama açısından veri yoğunluğu oluşturabilen ve robotikte zaman alıcı olabilen bir konudur [18]. Bir robotun, bir ortam hakkında bilgi edinmek için sensörlerden gelen bilgilere erişmesi ve bunları işlemesi gerekir. Yörünge planlamanın zorluğu, en uygun yolu bulmayı ve hassas manipülasyonları yönetmeyi içerir. Bir yörünge planlama algoritması için tüm olası yörüngelerde bir arama işlenirken, bir kullanıcının robottan yanıt almadan önce prosedürün tamamlanmasını beklemesi gerekir. Bu tarz prosedürler, robotikte hareket planlamanın performansının, yapılan işin kalitesi ve zamanlaması konusundaki hassasiyetini arttırmaktadır. Bu durum da planlama algoritmalarının sisteme uygun bir şekilde seçilmesinin önemini vurgulamaktadır.

Robotik sistemlerde yörünge planlamaları için farklı yörünge planlama algoritmalarından faydalanılmakta ve bu sistemler için yeni birçok algoritma geliştirilmektedir. Gasparetto ve ekibinin [23] yaptığı çalışmalar, doğrudan robotikteki yol bulma ve yörünge belirleme algoritmalarından bazıları üzerinde yoğunlaşmıştır. Bu çalışmadaki planlama çeşitleri, robot kollarının buldukları konumlar ile ortam engelleri arasındaki mesafelerin referans alındığı yol haritası (roadmap) teknikleri, hücre ayrışması (cell decomposition) algoritmaları ve yapay potansiyel (artificial potential) metotları olarak üç farklı çeşitte incelenmiştir. Bu metotların hepsi, ortamın geometrik açıdan sınıflandırılması ile geliştirilen farklı yaklaşımları ele alır. Robotik kollar için yörünge planlaması için geliştirilen yaklaşımlardan biri Streinu'ya aittir [24]. Yaptığı çalışmada robot kollarının hareket kinematiklerini ve terminolojilerini belirlemiş, bu bağlamda robot kolu için geliştirilecek algoritmalara yol gösteren bir çalışma gerçekleştirmiştir. Moll ve ekibinin çalışmalarında [36], Streinu'nun çalışmasındakinden farklı mantalitede bir değerlendirme (benchmark) yapısı geliştirmişlerdir. Bu yapı, kullanıcı tanımlı hareket planlama algoritma sorunlarını çözme odaklıdır. Cohen ve ekibinin

çalışmalarında [37], hareket planları ile ilgili farklı bir sorun olan hareketli robotik kolların, robot hareketi ile kol hareketi hızlarının uyumsuzluğundan kaynaklanan planlama sıkıntısına değinilmiştir. Bu sorunu çözmek için yine farklı planlama algoritmalarını incelemişler ve bir değerlendirme sunucusu oluşturarak bir yaklaşım geliştirmişlerdir. Liu ve ekibinin çalışmaları [38], robotik yörünge planlama optimizasyonu testlerine bir alternatif olacak şekildedir. Tünel, kütüphane ve endüstriyel alanlar gibi farklı ortamlarda, farklı planlama algoritmaları ile yapılan testler, algoritmaların planlama yapma süreleri baz alınarak gerçekleştirilmiştir. Testlerde algoritmaların başarı yüzdeleri, yolu pürüzsüz geçiş süreleri gibi parametreler de baz alınmıştır.

Yörünge planlama için kullanılan algoritmalarından birisi olan STOMP (Stochastic Trajectory Optimization) [7], genel kısıtlamalarla başa çıkabilen bir hareket planlama yaklaşımıdır. Bu yaklaşım, bir dizi gürültülü yörüngeden istifade ederek stokastik yörünge optimizasyonunu kullanmaktadır. Her tekrar uygulamada, bir dizi yörünge üretilir. Oluşturulan yörüngeler, daha sonra aday çözümünü güncellemek için kullanılan maliyetleri belirlemek amacı ile simüle edilmektedir. Bu süreçte gradyan bilgisi gerekmediğinden, genel kısıtlamalar ve ek sorunsuz maliyetler optimize edilebilir. Diğer bir planlama algoritması olarak CHOMP (Covariant Hamiltonian Optimization) [6], hareket planlaması için iyi yörüngeler elde etmek amacıyla oluşturulmuş basit bir varyasyon stratejisidir. CHOMP, dışbükey olmayan maliyet fonksiyonlarında yüksek boyutlu uzaylarda optimizasyon problemlerini ele almaktadır. Son olarak planlama algoritmalarının en yaygın tarafında ise OMPL (Open Motion Planning Library) kütüphanesi ve bu kütüphane bünyesindeki algoritmalar bulunmaktadır. Bu algoritmalar ROS temelli robotik uygulamalarda kullanılabilirliği yüksek, yazılım geliştirmelerine açık algoritmalar olduğu için, günümüzde çok daha yaygın kullanılmaya başlamıştır [1]. Xinyu ve ekibi [25] OMPL'in algoritmalarından RRT* algoritmasının çift yönlü hareketteki etkisi üzerine çalışmıştır. Çalışmasında RRT* algoritmasının oldukça optimize olduğunu ancak işlem süresinin uzun ve yavaş olduğuna dikkat çekmiştir. LaValle ve ekibinin çalışmasında [8], RRT (Rapidly-Exploring Random Trees) algoritmasının 12 eksene kadar yörünge planlaması yapabilecek kabiliyette olduğu, engel tespitleri ve dizayn parametrelerinin, ideal bir hareket planlayıcı algoritmasında olması gerektiği kalitede olduğunu çeşitli örneklerle açıklamışlardır. Kuffner ve ekibi [9], daha sonra RRT-Connect algoritması ile RRT'deki yörünge planlama menziline arttıracak yeni

düzenlemeler geliştirmişlerdir. Karaman ve ekibi [10] de RRT ve PRM (Probabilistic Roadmaps) algoritmalarının RRT* ve PRM* yükseltmeleri üzerinde çalışmalar yapmışlardır. Bu geliştirmeler de yine RRT-Connect örneğinde olduğu gibi algoritmaların planlama zamanlarını kısaltmak ve ortam menzillerini genişletmek üzerine tasarlanmıştır. EST (Expansive Space Trees) algoritması, Hsu ve ekibi [34] tarafından ortaya çıkarılmış bir hareket planlama algoritmasıdır. Bu algoritma temelinde, rastgele örneklenmiş kilometre taşlarından oluşan bir yol haritasıyla bağlantısı etkin bir şekilde yakalanabilen bir robot konfigürasyon alanı üzerine odaklanmıştır. Bu alan ailesini karakterize etmek için genişleme kavramı oluşturmak üzerine kurulmuş bir algoritma ortaya çıkarmışlardır.

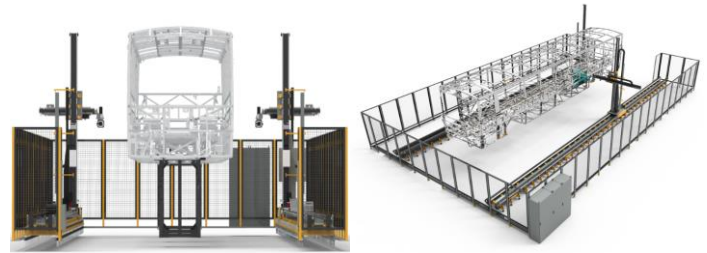
Bu çalışmada, tanımlanmış kullanım senaryosunun doğrulanması ve onaylanması amacıyla gerçek ortamda geliştirilmiş ROKOS sistemi GAZEBO simülasyon ortamına aktarılarak SRVT ortaya çıkarılmış ve bu sistem üzerinde ise proje kapsamında oluşturulan değerlendirme senaryolarının simülasyon tabanlı testleri yapılmıştır.

Bu çalışmada, değerlendirme senaryolarından “Emniyetli Yörünge Optimizasyon” kapsamında, simülasyon ortamına aktarılmış olan otobüs şasesinin kalite kontrolü için ROKOS sistemine ait robot kollarının belirlenen değerlendirme kriterine göre emniyetli bir şekilde hareket edecekleri yörüngelerin oluşturularak geliştirilen yazılımların doğrulama ve onaylama işlemlerinin yapılması hedeflenmiştir. Bu kapsamda, ROS MoveIT aracına ait OMPL ve EST yörünge planlama algoritmaları üzerinde çalışılmış, bu algoritmaların artı ve eksi yönleri belirlenmiştir. Bu algoritmaların simülasyon ortamında doğrulanmaları için kullanım senaryosu üzerinde detaylı testler yapılmış ve sonuçları analiz edilmiştir. Testler 3 farklı senaryo (Hızlı test, Resetli Tam Test ve Resetsiz Tam Test) için yapılmış ve SRVT’ye aktarılmış mevcut ROKOS sistemi için doğrulama faaliyetleri zaman ve maliyet açısından iyileştirilmiştir.

Makale akışı olarak, Bölüm 2’de ROKOS sistemi ve bu sistemin işlevi, Bölüm 3’te bu sistemin entegre edildiği SRVT’yi oluşturan yazılımlar ile ilgili detaylar açıklanmıştır. Bölüm 4’te yörünge planlaması için test edilecek OMPL planlama algoritmaları incelenmiş, Bölüm 5’te ise bu algoritmaların performans kıyaslamaları, ROKOS sisteminin test süresi tabloları üzerinden incelenmiştir. Çalışmanın sonuçları ise Bölüm 6’da verilmiştir.

II. ROKOS (ROBOT KONTROL SİSTEMİ)

ROKOS sisteminin kullanım senaryosu, yenilikçi görsel kontrol teknikleriyle ürünün kalite kontrol süresinin kısaltarak otobüs karoserisindeki parçaların varlık-yokluk kontrolünü daha hassas yapmaya odaklanmaktadır. Bu kullanım senaryosunun temeli, daha iyi kalite kontrolü elde etmek için hataya dayanıklılık açısından daha iyi performanslı bir üretim sistemi sağlamaktır. (Şekil 1). 2500-3000 gövde parçasının varlığının kontrolünün kartezyen robot ve kamera sensör sistemi ile tam otomatik olarak yapılması planlanmaktadır. OTOKAR tarafından geliştirilen ve uzaktaki sunucuda çalıştırılan dijital ikiz yazılımı, tüm parçaların varlığını kontrol etmek için güvenli robot yörünge noktalarını belirler. Sunucudaki veri tabanında depolanan her robot yörüngesi noktası için yazılım, sanal bir kamerayı CAD (Computer-aided Design) ortamında konumlandırır ve sanal iki boyutlu görüntüler oluşturur. Sunucudan gelen yörünge noktalarını girdi olarak işleyen ve sistem bilgisayarında çalıştırılan yazılım, robot eksen motorlarını PLC ve sürücüler aracılığıyla kontrol ederek Otokar’ın geliştirdiği kartezyen robotları gerçek üretim ortamında konumlandırır. Yörüngedeki her konum için kamera sensörleri, gerçek üretim ortamından 2B görüntüler yakalar. Bu konumlar için CAD verilerinden elde edilen sentetik 2B veriler ile kamera sisteminden elde edilen gerçek 2B görüntü verileri karşılaştırılarak parça varlık-yokluk kontrolü yapılır. Sistem durumu ve kalite kontrol raporları, kalite personeline arayüzlerle gösterilir ve depolanır.



Şekil 1. Kalite kontrol için robot denetleme sistemi

Şekil 2’de gösterildiği gibi, hedeflenen sistem topolojisi aşağıdaki gibidir: Sistem uzaktaki bir sunucuya gerçek zamanlı veri ileten ve giriş verileri alan iki güvenli bölgeden oluşmaktadır. Birinci güvenli bölgede sistem bilgisayarları, 2 adet Time of Flight (ToF) kamera ve 2 adet 2B kamera bulunmaktadır. Kameralar sistem bilgisayarına görüntü verilerini direk göndermektedir.

VALU3S projesi kapsamında belirlenmiş Değerlendirme

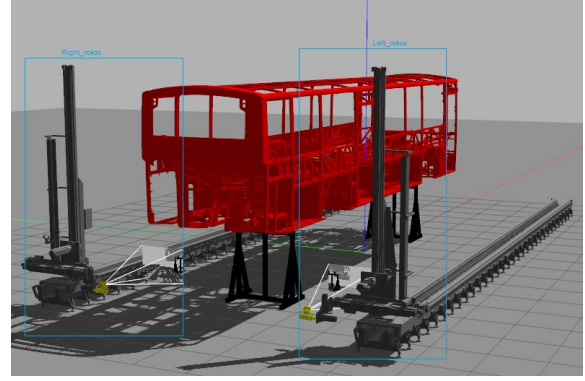
Senaryoları ve Kriterleri (Evaluation Scenarios and Criteria) kapsamında Otokar tarafından sağlanan bu sistemin, “Güvenli Yörünge Optimizasyonu (Safety Trajectory Optimization)” alanında çalışmalar yapılmıştır. Bu değerlendirme senaryosu, robotun ve robot aparatlarının güvenliğini, çalışma alanındaki statik nesnelere de otomatik olarak kapsayan robot yörünge noktaları oluşturmak, robotun güvenli bir hareket rotası çizimini planlamayı amaçlar. Yapılan bu çalışmada kullanılan iki ROKOS Kartezyen robot kolunun hareketli tüm eklemleri ayrı ayrı sisteme tanımlanmakta ve kontrol edilmektedir. Bu kontrol, ROS üzerinde geliştirilen yazılımlar vasıtasıyla gerçekleştirilmekte, çeşitli kontrol parametreleri doğrultusunda gerçekçi robot hareketleri elde edilmektedir.

III. YAZILIMLAR VE SİMULASYON ORTAMI

Bu bölümde, ROKOS sisteminin entegre edildiği SRVT ekosistemi ve bu ekosistemi oluşturan yazılımlar hakkında temel bilgilere yer verilmiştir.

Robotik İşletim Sistemi (ROS) [1],[2] olarak tabir edilen sistem, açık kaynak kodlu olarak paylaşılan, robotlar için donanım/yazılım kontrolü, servisler ve iletişim protokolleri gibi birçok robotik aracı sağlayan bir arakattir. Robotik sistemleri simülasyon ortamlarında modellemeyi, kontrol etmeyi ve yeni yazılım araçları entegre ederek geliştirmeyi mümkün kılmaktadır.

Gazebo [3], robot modellerini simülasyon ortamında, kinematik özelliklerini gerçeğiyle birebir modellemenin mümkün olduğu, gerçekçi bir ortam simülasyon yazılımıdır. ROS destekli çalışabilmekte, ROS üzerinde Gazebo için geliştirilmiş paketler ile kullanılabilir. Gazebo'nun sağladığı simülasyon ortamında, gerçeğe çok yakın ortamlar oluşturmak ve robotları bu ortamlarda gerçekçi fizik kuralları altında kullanmak mümkündür. Bu çalışmada ROKOS sistemi ve bu sistemin kullanıldığı otobüs şasesi, gerçeklerine birebir uygun ölçü ve atalet momentleri değerleri ile Gazebo ortamına uyarlanmıştır (Şekil 2).

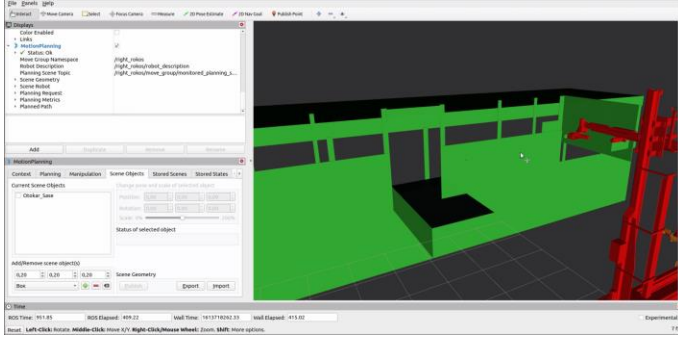


Şekil. 2. Gazebo simülasyon ortamına modellenmiş ROKOS sistemi ve otobüs şasesi

SMACH, hızlı bir şekilde karmaşık robot davranışları oluşturmak için görev düzeyinde bir mimaridir. SMACH özünde, hiyerarşik durum makineleri oluşturmak için ROS'tan bağımsız bir Python kitaplığıdır. SMACH, bakımı yapılabilen ve modüler kod ile hızlı bir şekilde sağlam robot davranışı oluşturmak için çok eski konseptlerden yararlanan yeni bir kütüphanedir. Bu kütüphane kullanılarak, robot kolların çeşitli görev durumlarının kolaylıkla kontrolü mümkün olmaktadır.

Moveit! [4],[15] ROS'un hareket planlama yazılımıdır (Şekil 3). Robot, robot kolu ve dronlar gibi robotik araçların hareket kontrollerini ve planlamalarını sağlayan bir ortam sağlar. Bu planlama sırasında robotların sensörlerini, ortam haritalarını ve hareket planlama algoritmalarını kullanır. Robotun URDF modellerini yeniden düzenleyerek, bu sistem içinde çalışabilir SRDF tanımlama dosyasına çevirir ve planlayıcı algoritmalarını işletir. Bu dosya, ilgili robotun hareketli ya da hareketsiz parçalarının tanımlanmasını ve bu parçaların gerekli parametrelerinin tanımlanmasında kullanılır.

Moveit!, ROS'ta hareket planlama ve mobil manipülasyon için birincil yazılım aracıdır ve PR2, Robonaut ve DARPA'nın Atlas robotu dahil olmak üzere birçok robotla başarılı bir şekilde entegre edilmiştir. Moveit! tamamen C++ ile yazılmıştır, ancak Python bağlantılarını da içerir. Moveit! varsayılan olarak çekirdek ROS derleme ve mesajlaşma sistemlerini kullanır. Bileşenleri kolayca değiştirebilmek amacıyla eklentileri kullanır: hareket planlama eklentileri (örn. OMPL), çarpışma algılama (örn. Hızlı Çarpışma Kitaplığı (FCL)), kinematik eklentileri (örn. OROCOS Kinematik ve Dynamics Library (KDL)). Moveit!'in hedef uygulaması endüstriyel, ticari ve araştırma ortamlarında manipülasyonaa yönelik geliştirilmiştir [35].



Şekil. 3. MoveIt! arayüzü

Moveit! yazılımı, içerisinde OMPL, CHOMP ve STOMP planlayıcıları ve bu planlayıcıların hareket planlama algoritmaları bulundurmaktadır. CHOMP (Covariant Hamiltonian Optimization for Motion Planning) kütüphanesi, robotik sistemlerdeki rotasyon optimizasyonunu sağlayabilmek için yeniden parametrelendirme yapılabilmesini sağlayan yöntemler oluşturan algoritmaları barındıracak bir kütüphane olarak tasarlanmıştır [6]. Bu kütüphane, bir başlangıç yörüngesinin kalitesini tekrarlı olarak iyileştirmek için işlevsel gradyan adı verilen teknikleri kullanır. STOMP ise, daha düşük maliyetle oluşturulmuş bir yörünge oluşturmak için birleştirilen ilk (muhtemelen mümkün olmayan) bir yörünge etrafındaki alanı keşfetmek için, gürültü eklenmiş yörüngeler oluşturmaya dayanan bir teknik kullanır. Her tekrarda engel ve pürüzsüzlük maliyetinin bir kombinasyonunu temel alan bir maliyet işlevi optimize edilerek çalışır [7]. Bu çalışmada OMPL algoritmaları üzerinde çalışılmıştır.

IV.OMPL HAREKET PLANLAMA ALGORİTMALARI

OMPL (Open Motion Planning Library), birçok son teknoloji planlama algoritmasının uygulamalarını içeren, örneklemeye dayalı hareket planlaması için geliştirilmiş bir yazılım kütüphanesidir [5]. Bu kütüphane, kullanıcının çeşitli karmaşık hareket planlama problemlerini minimum girdi ile kolayca çözmesine olanak tanıyacak şekilde tasarlanmıştır. OMPL, yeni hareket planlama algoritmalarının eklenmesini kolaylaştırmakta ve diğer yazılım bileşenleriyle uygun bir şekilde arayüzlenebilmektedir.

OMPL, bünyesinde RRT (Rapidly-exploring Random Trees) [8], RRT-Connect (Rapidly-exploring Random Trees - Connect) [9], RRT* (Rapidly-exploring Random Trees - Star) [10], PRM (Probabilistic Road Map) [11], PRM* (Probabilistic Road Map - Star) [12], EST (Expansive Space Trees) [13], KPIECE (Kinodynamic

Motion Planning by Interior-Exterior Cell Exploration) [14] ve BKPIECE (Bidirectional implementation of KPIECE) [14] gibi birçok algoritma bulundurmaktadır. Bu çalışmada verilen örneklerdeki algoritmalarından RRT, RRT*, RRTConnect, PRM, PRM*, EST (Expansive Space Trees) ve BiEST (Bidirectional Expansive Space Trees) algoritmaları kullanılmıştır. Şekil 4'te bu algoritmaların bazılarının kompleksitesi ile alakalı bir tablo verilmiştir.

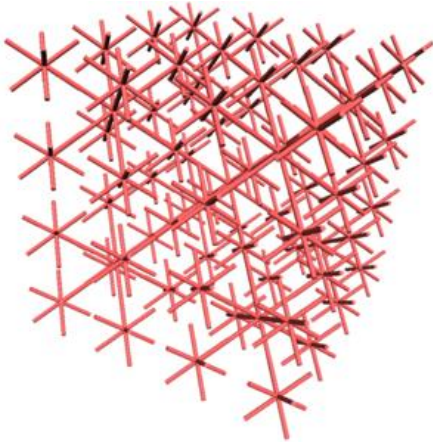
	Algorithm	Probabilistic Completeness	Asymptotic Optimality	Monotone Convergence	Time Complexity		Space Complexity
					Processing	Query	
Existing Algorithms	PRM	Yes	No	Yes	$O(n \log n)$	$O(n \log n)$	$O(n)$
	sPRM	Yes	Yes	Yes	$O(n^2)$	$O(n^2)$	$O(n^2)$
	k-sPRM	Conditional	No	No	$O(n \log n)$	$O(n \log n)$	$O(n)$
	RRT	Yes	No	Yes	$O(n \log n)$	$O(n)$	$O(n)$
Proposed Algorithms	PRM*	Yes	Yes	No	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$
	k-PRM*						
	RRG	Yes	Yes	Yes	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$
	k-RRG	Yes	Yes	Yes	$O(n \log n)$	$O(n)$	$O(n)$
	RRT*						
k-RRT*							

Şekil. 4. OMPL algoritmaları

Aşağıda, bu çalışmada teste tabi tutulan algoritmaların temel özellikleri ve diğer algoritmalarla benzerlik ve farklılıkları incelenmiştir.

A. RRT Algoritmaları (RRT, RRT-Connect, RRT*)

RRT algoritması LaValle ve Kuffner'in çalışmalarında [29,30] ortaya çıkarılmış bir yörünge planlama algoritmasıdır. Boşluk dolduran bir ağaç yapısını (space-filling tree) [31] rastgele oluşturarak, konveks olmayan ve yüksek boyutlu olan alanlar içerisinde, verimli bir yörünge aramak için tasarlanmış bir algoritmadır (Şekil 6). Bu ağaç yapısı, arama alanından rastgele alınan örneklerden aşamalı olarak oluşturulur ve doğası gereği, uygun yörüngeyi bulabilmek için algoritmanın incelemediği geniş alanlara doğru büyüyerek tarama yapmaya meyillidir. Engeller ve diferansiyel kısıtlamalarla (holonomik olmayan ve kinodinamik) sorunları kolayca çözmekte ve otonom robotik hareket planlamasında yaygın olarak kullanılmaktadır.



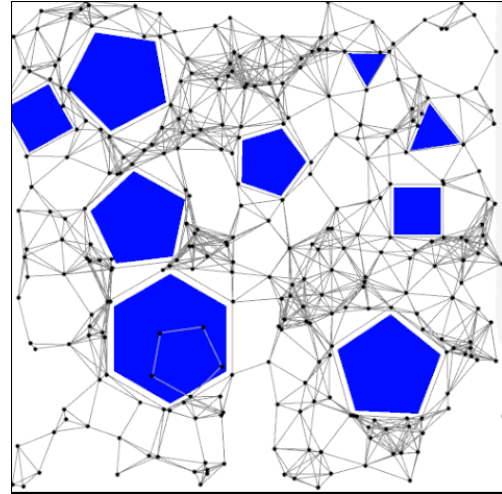
Şekil 5. Küp şeklinde örnek bir "space-filling tree" yapısı

Probabilistik yol haritası (probabilistic roadmap) methodundan esinlenilerek oluşturulmuş bir sistem olan RRT'nin, en belirgin avantajı, holonomik olmayan ve kinodinamik planlamaya doğrudan uygulanabilmeleridir. Bu avantaj, RRT'lerin konfigürasyon çiftleri (veya durumlar) arasında herhangi bir bağlantı yapılmasını gerektirmemesinden, olasılıklı yol haritalarının ise tipik olarak on binlerce bağlantı gerektirmesinden kaynaklanmaktadır. Özetle RRT'ler, holonomik yörünge planlaması için temel bir olasılıklı yol haritasından daha verimli bir algoritma özelliği taşımaktadır [29,30].

Daha sonraki yıllarda RRT algoritmasının bazı eksikliklerini (planlama hızı ya da uygulama hızı sorunları gibi) gidermek için, RRT algoritmasından türetilmiş yeni algoritmaları ortaya çıkarılmıştır. Bunlardan kısaca bahsetmek gerekirse:

RRT-Connect: Yine Kuffner ve LaValle tarafından oluşturulmuş, RRT'nin hareket planlama menzilini arttırmak için geliştirilmiş RRT algoritmasıdır [9].

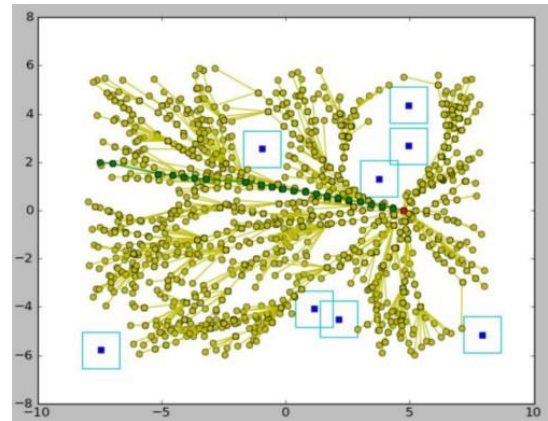
RRT*: RRT*, çalışılan düğüm sayısının sonsuza yaklaştığı durumlarda, hedefe ulaşmak için mümkün olan en kısa yolu verecek şekilde revize edilmiş bir RRT algoritmasıdır. Karaman ve ekibi tarafından oluşturulmuş olan bu algoritma [10, 32], gerçekçi olarak sonsuza yakın düğüm ihtimali mümkün olmasa da algoritmanın en kısa yolu geliştirmek için çalıştığını öne sürmektedir. RRT*'ın temel prensibi RRT ile aynıdır, ancak algoritmaya yapılan iki önemli ekleme önemli ölçüde farklı sonuçlar elde edilmesini sağlar. Bu farklılıklar, RRT*'ın her bir tepe noktasının ana tepe noktasına göre kat ettiği mesafeyi kaydetme ve bunun üzerinde çalışma yapacak şekilde tanımlanması ve boşluk dolduran ağacın yeniden bağlanması ve yapılandırılmasının gerçekleştirilmesidir [10, 29, 30, 32] (Şekil 6).



Şekil 6. RRT* yörünge planlaması

B. PRM Algoritmaları (PRM, PRM*)

Kavraki ve ekibi tarafından ortaya çıkarılan PRM yörünge planlama algoritması [11] çarpışmalardan kaçınırken robotun bir başlangıç konfigürasyonu ile bir hedef konfigürasyonu arasında bir yol belirleme sorununu çözen robotikte bir hareket planlama algoritmasıdır. PRM algoritmasının arkasındaki temel fikir, robotun hareket edeceği sahadan rastgele örnek noktalar alarak bu noktaların boş alanda kalıp kalmadıklarını sınamak ve bu noktaları yakındaki diğer noktalara bağlamak için yerel bir planlayıcıdan istifade etmeye dayanır (Şekil 7).



Şekil 7. Bir dizi poligonel engel etrafında uygun yolları araştıran bir PRM algoritması örneği.

PRM algoritması çalışırken, başlangıç ve hedef noktaları eklenir ve bu noktalar arasında bir yol belirlemek için ortaya çıkan grafiğin üzerinde bir de grafik arama algoritması uygulanır. Bu çift planlama ile optimal bir planlama algoritması oluşturulmaya çalışılmıştır.

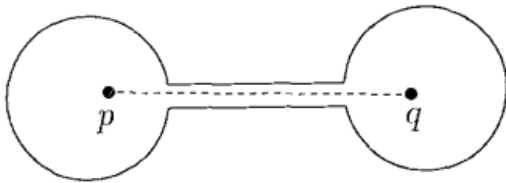
PRM algoritmasının planlama işlemi iki aşamadan oluşur: Oluşturma aşaması ve Sorgulama aşaması. Oluşturma

aşamasında, çevrede yapılması mümkün olan hareketlere ya da yönelimlere yakın bir yol haritası (grafik) oluşturulur. Bu grafik önce rastgele bir konfigürasyon temelinde meydana gelir. Daha sonra genelde ya en yakın komşulara ya da önceden belirlenmiş bir mesafeden daha az olan tüm komşulara bağlanır. Yol haritası yeterince yoğun olana kadar grafiğe konfigürasyonlar ve bağlantılar eklenir. Sorgulama aşamasında ise, başlangıç ve hedef konfigürasyonları grafiğe bağlanır ve yol Dijkstra'nın en kısa yol sorgusu [33] ile elde edilir.

PRM'nin avantajı genel ve neredeyse her tür holonomik robota uygulanabilir bir algoritma yapısına sahip olmasıdır. Ayrıca, kolaylıkla özelleştirilebilir. Bu özelleştirme, yerel planlayıcı gibi yöntemin genel yapılarının, dikkate alınan ortam özelliklerine daha iyi uyan daha spesifik yapılarla değiştirilmesi ile mümkündür [11]. Özelleştirilmiş uygulama, çok sayıda serbestlik derecesi içeren robot için çok zor yörünge planlama sorgularını, birkaç dakikalık bir öğrenme süresinden sonra bir saniyeden kısa bir sürede çözebilmektedir.

C. EST Algoritmaları

EST kütüphanesi ile geniş hareket alanlarının analizi yapan, yeni bir dinamik planlama algoritmaları tasarlanmıştır (Şekil 8). Bu algoritmalar, konfigürasyon alanının sadece mevcut sorgu ile ilgili kısmını örneklemeye çalışır ve tüm konfigürasyon alanı için bir yol haritasını önceden hesaplamının maliyetini ortadan kaldırır. Bu nedenle, belirli bir ortam için tek bir sorgunun gönderildiği problemler için çok uygundur [34].



Şekil 8. Dar bir geçidin varlığı nedeniyle rastgele örnekleme yoluyla bağlanabilirliği yakalanması zor olan bir konfigürasyon alanı

V. HAREKET PLANLAYICILARDA PERFORMANS KARŞILAŞTIRMASI

Bu kısımda, SRVT sisteminin bünyesinde bulunan hareket planlayıcılarının içerisinde yer alan OMPL ve EST planlayıcı algoritmalarının performanslarının karşılaştırılması ve teknik bazı özellikler açıklanmıştır.

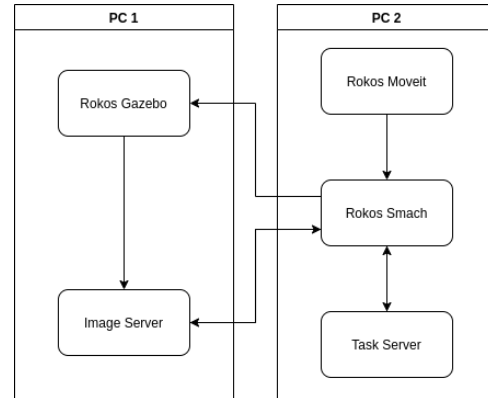
A. Test Donanımlarının Özellikleri

ROKOS sisteminin simule edildiği SRVT sistemi, yüksek performanslı işlemci gücü gerektiren bir görev planlama, uygulama ve simule etme sistemi olduğundan, çalışmalar sırasında bu sistem iki ayrı PC'den kullanıldı. Tablo 1'de bu iki PC'nin donanım özellikleri verilmiştir.

Tablo I. Test Donanımlarının Özellikleri

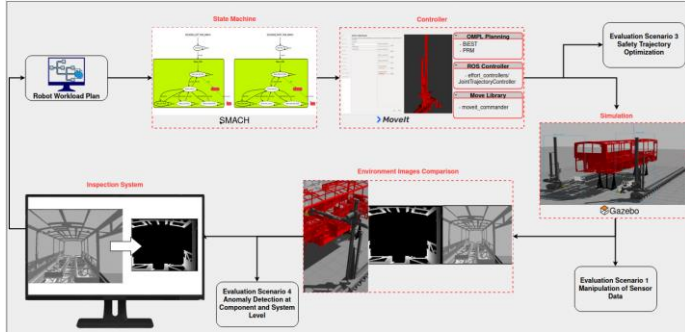
Özellikler	PC1	PC2
Platform	Linux x86_64	Linux x86_64
Platform Dist.	Ubuntu 20.04.02 LTS	Ubuntu 20.04.02 LTS
Kernel	5.8.0-50-generix	5.8.0-50-generix
Release		
ROS Dist.	ROS Noetic	ROS Noetic
CPU Model	Intel® Core(TM) i7-7700HQ CPU	Intel® Core(TM) i7-8750H CPU
CPU Core	8	12
CPU Speed	2.80 GHz	2.20 GHz
RAM	16 GB	8 GB
Graphic Card	NVIDIA GeForce GTX 1050	NVIDIA GeForce GTX 1050

Yukarıda donanım özellikleri verilen PC1, SRVT sisteminin simülasyon ortamının çalıştırıldığı donanımdır, PC2 bu simülasyon ortamındaki görev atamalarını ve planlamalarını gerçekleştiren donanımdır. Bu iki donanımda çalışan ROS düğümlerinin oluşturduğu sistem mimari diyagramı Şekil 9'da gösterilmiştir. Sistemin genel çalışma diyagramı da Şekil 10'da gösterildiği gibidir.



Şekil 9. SRVT ROKOS entegrasyonunun ROS sistem mimarisi

SRVT sistemi, Bölüm 2'de anlatıldığı gibi iki kartezyen robotun aynı anda çalıştığı, belirli görevleri icra ettiği bir işlemdir. Bu işlem süreci, iki ayrı robota, belirlenen algoritmaların ya da komutların uygulandığı iki ayrı kod sisteminin uyumlu çalışması ile mümkün olmaktadır. Şekil 10'da verilen SRVT çalışma diyagramında, bu iki ayrı robot kolu için işletilen SMACH yapısı görülebilir.



Şekil 10. SRVT çalışma diyagramı

B. Görev Listesinin Oluşturulması

ROKOS sistemi, kendisine gönderilen görev listelerinde belirtilen koordinatlara, verilen eksen hareket sıralarına uyacak şekilde hareketler ile intikal etme üzerine tasarlanmıştır. Bu görev listeleri, SRVT sisteminin ROS paketlerinde ROKOS kollarına ait konfigürasyon dosyaları içerisinde yer alan parametre klasörlerinde saklanmaktadır. Tablo 2’de bu parametre dosyalarından birinin içeriği görülebilir.

Tablo II. Rokos sol kol görev parametresi içerik kesiti

```

---
Left_Rokos:

- Task:
  Task_ID: 35694 %Görev Numarası

  Vehicle_CODE: '13M38_1' %Şase Kodu

  Mode: 1 %Kamera Kayıt Modu

  Tag: 'ON' %Şaseye Göre Görev Konumu

  Position: %Robot Kol Koordinatları
  X: 6.079
  Y: 2.840
  Z: 1.603
  C_AX: 10.000
  C_AZ: 170.000

  Queue: %Hareketin Uygulanma Sırası
  X: 1
  Y: 3
  Z: 2
  C_AX: 4
  C_AZ: 5
  ...

```

Tablo 2’de bir kesiti yer alan .yaml uzantılı görev listesi dosyası, ROKOS kollarının ROS paketleri çalıştığında sistem tarafından istenir, bu verilere uygun olarak robot kolları sırasıyla görevleri gerçekleştirir. Kesitte yorum satırı olarak vurgulanmış detaylara göre sınıflandırılmış görevler, SRVT sistemi tarafından uygulanmakta, sistem

çıkartılarının kayıtları bu detaylara göre tutulmakta, robot kol hareketleri ilgili koordinatlar ve bu koordinatlara intikal için gerekli yörünge hareket sıralamalarına uygun hareketler yaparak görevleri tamamlamaktadır.

Tablo III. Gerçek ortamdaki görev koordinatlar listesi

Y	35269	35270
Araç_Kodu	13M38_1	13M38_1
Robot_NO	1	1
Kamera_X	1750	1150
Sıra_X	3	3
Kamera_Y	1082	1682
Sıra_Y	2	2
Kamera_Z	50	50
Sıra_Z	1	1
Kamera_AX	10	10
Sıra_AX	4	4
Kamera_AZ	0	0
Sıra_AZ	5	5
Durum	1	1
Created_at	2020-10-18 2:09:57.000	2020-10-18 02:09:57.000
Etiket	ARKA	ARKA

Tablo 3’te SRVT ortamına göre uyarlanmış ROKOS görev bilgilerinin bulunduğu dosya içeriğinden bir kesit görülmektedir. Yeni bir görev oluşturulurken ya da görevlerde değişiklik yapıldığında bu değişikliklerin “.yaml” uzantılı görev dosyalarına çevrilmesi için bir ROS düğümü geliştirilmiştir. Bu düğüm, hazırlanan .csv uzantılı görev tablolarını otomatik olarak .yaml uzantılı konfigürasyon dosyalarına çevirmektedir. Bu şekilde ROKOS sistemi için görev dosyaları oluşturulmuş olur.

ROKOS sistemi için tasarlanmış görev listelerinde, robotun yörünge planı olmaksızın engellere çarpmadan hareketi şeklinde bir çalışma mantalitesi benimsendiğinden, belirli konumlar arasına yerleştirilmiş reset noktaları bulunmaktadır. Bu noktalar sayesinde robot, intikal etmesi gereken iki nokta arasında ilerlerken önce belirlenen reset konumlarına ulaştığından engellere çarpmadan hareketler gerçekleştirebilmektedir. Toplamda iki ROKOS kolu için 330 görev noktası bulunmaktadır ve bu noktaların 58’i reset noktalarıdır. ROKOS sisteminde ise çarpışmasız dinamik yörünge planlama olduğu için bu reset noktalarının kullanılmasına gerek kalmamaktadır. Testler gerçekleştirilirken hem eski görev listesi hem de reset noktaları çıkarılmış yeni görev listesi kullanılarak süreler elde edilmiştir.

C. Test Yöntemlerinin Belirlenmesi

Hızlı test, görev listesinden rastgele seçilen 15’er görev ile oluşturulan kısa görev listesini tamamlama sürelerinin hesaplanmasını içerir. Belirlenen noktalar, tam görev listesinin içerisinde, ilerleme sırasına göre rastgele



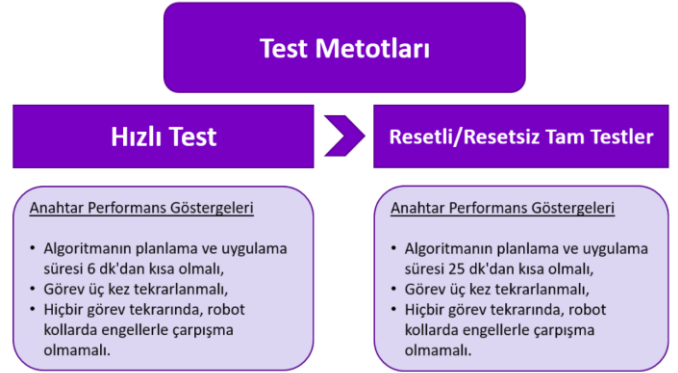
seçilmiştir. Bu da robotun tüm görevleri yapmasa da hareket güzergahı boyunca hareket etmesini sağlaması, testin robot kollarının menzillerini de hesaba katabilmesi için belirlenmiştir. Resetli Tam Test, gerçek ortamdaki ROKOS sistemi için oluşturulmuş, görev sırasında çarpışmaları engellemek için bazı reset noktaları belirlenmiş 330 görev koordinatından oluşan listeye uygulanan testtir. Resetsız Tam Test ise bu reset noktaları çıkarılarak, görev listesinin sayısının 272 noktaya indirilmesiyle oluşturulmuş yeni listeye uygulanan testtir. Bu testler sonucunda elde edilecek süreler her iki kol aynı anda çalışırken, ayrı ayrı hesaplanır ve verilen görev listesini üçer defa tamamlanması ile elde edilen veriler üzerinden gerçekleştirilmiştir. Testlerin başarı kriterleri Şekil 10'daki diyagramda verilmiştir. Ayrıca Tablo 4'te bu görev listelerinin dosya yapısı gösterilmiştir.

Tablo IV. ROKOS görev listeleri dosya yapısı

```
---
Rokos_Task:
# None = Full Task

Right_Rokos: None
#- 35272
#- 35277
#- 35279
#- 35291
#- 35310
#- 35320
#- 35330
#- 35340
#- 35349
#- 35360
#- 35374
#- 35380
#- 35390
#- 35414
#- 35432

Left_Rokos: None
#- 35702
#- 35705
#- 35700
#- 35710
#- 35730
#- 35749
#- 35760
#- 35770
#- 35780
#- 35790
#- 35818
#- 35825
#- 35830
#- 35840
#- 35875
```



Şekil. 11. Test yöntemleri ve anahtar performans göstergeleri diyagramı

Şekil 11'de verilen test yöntemlerindeki başarı kriterleri, "Güvenli Yörünge Optimizasyonu" değerlendirme senaryosu kapsamında yer alan kriterlere göre oluşturulmuştur.

Testler sol ve sağ ROKOS kolları için ayrı ayrı süreler alınmak suretiyle uygulanmaktadır. Bunun sebebi her iki robot kolu için planlama algoritmalarının ayrı ayrı çalıştırılıyor olmasıdır. Ayrıca iki kol için de anlık görev koordinatları ve yörüngeler farklılık göstermektedir. Bu farklılıklarla birlikte aslında tek bir robotik sistemde iki farklı robot kolunun ayrı ayrı planlama ve planı uygulama sürecinden geçirilmesi durumunu oluşturur. Bu nedenle testler esnasında her kol için algoritma performansları ayrı ayrı değerlendirilmekte ve bu sonuçların ortalamaları alınarak sonuçlar analiz edilmektedir.

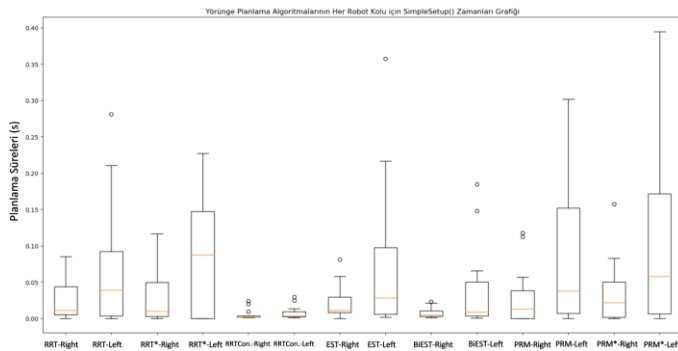
D. Algoritma Testleri

Yörünge planlama algoritmaları için belirlenen test yöntemlerinin uygulamasından önce, bu algoritmaların planlama ve planı uygulama çözümlerini bulma sürelerinin de testleri gerçekleştirilmiştir (Planlama Kurulum ve Çözüm Bulma değerleri). Algoritmaların bu zaman değerlerinin ortalama ve standart sapma değerleri, Tablo 5'te gösterilmiştir.

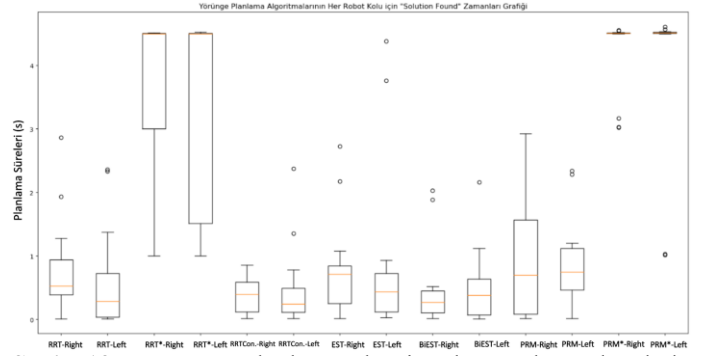
Tablo V. OMPL Algoritmaları Uygulanan Rokos Kollarının PlanlamaKurulum (SimpleSetup) ve ÇözümBulma (SolutionFound) Zamanları Tablosu

Planlayıcılar&Rokos Kolları	Simple Setup Ortalama (sn)	Simple Setup Standart Sapma (sn)	Solution Found Ortalama (sn)	Solution Found Standart Sapma (sn)
RRT-Sol	0.027837	0.030814	0.829886	0.788524
RRT-Sağ	0.070042	0.083717	0.676665	0.844881
RRT*-Sol	0.027837	0.030814	0.829886	0.788524
RRT*-Sağ	0.095220	0.09871	3.371867	1.564152
RRTCon.-Sol	0.005815	0.007541	0.399085	0.282422
RRTCon-Sağ	0.008482	0.008846	0.504321	0.640946
EST-Sol	0.025089	0.024726	0.817884	0.761156
EST-Sağ	0.080437	0.101916	0.905613	1.332297
BIEST-Sol	0.008851	0.007632	0.527579	0.648691
BIEST-Sağ	0.041339	0.056627	0.531412	0.579818
PRM-Sol	0.031796	0.040521	1.074248	1.030772
PRM-Sağ	0.100877	0.104994	0.921887	0.756369
PRM*-Sol	0.037562	0.042537	4.229041	0.599214
PRM*-Sağ	0.115702	0.121598	4.059866	1.233106

Yukarıda verilen tabloda yer alan SimpleSetup zamanları, algoritmaların planlamayı gerçekleştirdiği süreleri vermektedir. SolutionFound süreleri ise bu planları, robot kollarına uygulattıkları süredir. Bu süreler, robot kollarının bir görev boyunca yaptıkları tüm hareket planlarından elde edilen zamanların analizi ile hesaplanmıştır. Bu sürelerle elde edilen süre grafikleri de Şekil 12 ve Şekil 13'te verildiği gibidir.



Şekil. 12. Yörünge planlama algoritmalarının her robot kolu için SimpleSetup zamanları grafiği



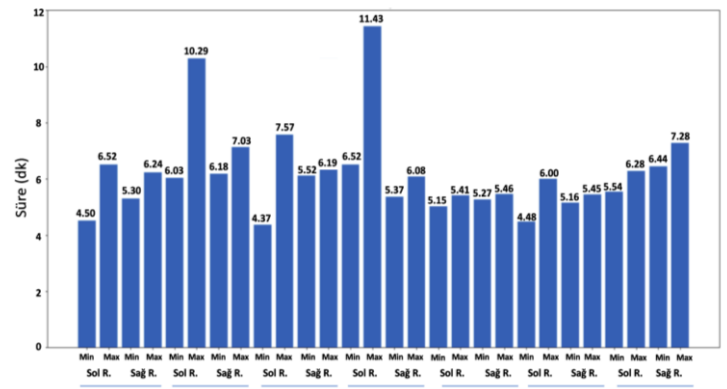
Şekil. 13. Yörünge planlama algoritmalarının her robot kolu için SolutionFound zamanları grafiği

E. Test Sonuçları

Hızlı test sonucunda elde edilen veriler, Tablo 6 ve Şekil 14'te gösterilmiştir.

Tablo VI. OMPL Algoritmaları Uygulanan Rokos Kollarının Hızlı Test Tamamlama Süreleri ve Ortalaması

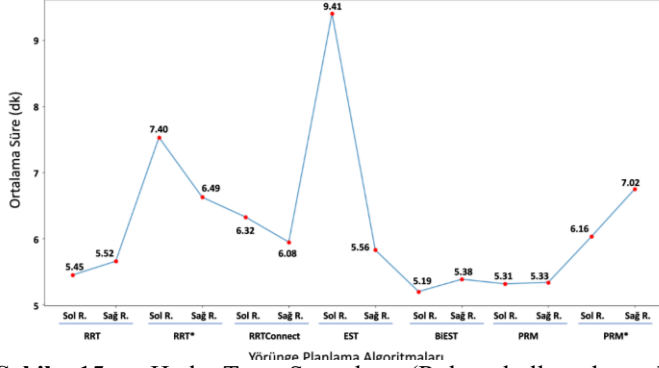
Planlayıcılar&Rokos Kolları	Tur 1	Tur 2	Tur 3	Ortalama Süre
RRT-Sol	06:52.97	04:50.54	05:31.53	05:45.01
RRT-Sağ	06:24.42	05:42.52	05:30.48	05:52.47
RRT*-Sol	10:29.36	06:03.42	06:27.28	07:40.05
RRT*-Sağ	07:13.12	06:56.35	06:18.12	06:49.20
RRTCon.-Sol	07:03.23	04:37.18	07:57.60	06:32.67
RRTCon-Sağ	06:19.93	06:12.19	05:52.78	06:08.30
EST-Sol	11:43.22	06:52.42	10:28.21	09:41.28
EST-Sağ	05:37.37	06:02.84	06:08.21	05:56.14
BIEST-Sol	05:15.27	05:41.59	05:02.73	05:19.86
BIEST-Sağ	05:41.93	05:46.64	05:27.91	05:38.83
PRM-Sol	06:00.56	04:48.77	05:45.50	05:31.61
PRM-Sağ	05:16.33	05:45.56	05:39.12	05:33.67
PRM*-Sol	05:54.98	06:27.28	06:28.12	06:16.79
PRM*-Sağ	07:28.44	06:53.26	06:44.68	07:02.13



Şekil. 14. Hızlı Test Sonuçları (Robot kolları bazında algoritmaların en iyi ve en kötü süreleri)



Şekil 14'teki grafikte, hızlı test sonucunda her algoritmanın her robot kolundaki en iyi ve en kötü görev tamamlama süreleri verilmiştir.

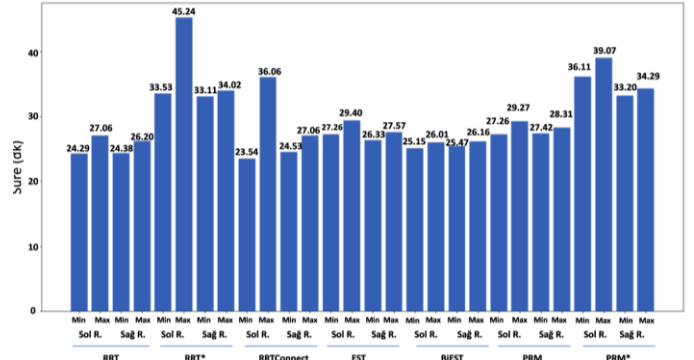


Şekil 15. Hızlı Test Sonuçları (Robot kolları bazında algoritmaların ortalama süreleri)

Şekil 15'deki grafikte ise hızlı test sonucundaki her algoritmanın her robot kolundaki ortalama görev tamamlama süreleri verilmiştir. Bu tablodan ve grafiklerden de anlaşılacağı üzere RRT, PRM ve BiEST algoritmaları, hızlı test sonucunda iyi süreler çıkarmıştır.

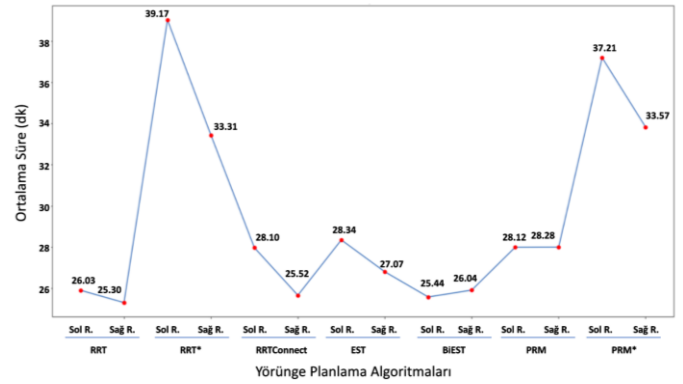
Tablo VII. OMPL Algoritmaları Uygulanan Rokos Kollarının Tam Test Tamamlama Süreleri ve Ortalaması

Planlayıcılar&Rokos Kolları	Tur 1	Tur 2	Tur 3	Ortalama Süre
RRT-Sol	26:34.13	27:06.65	24:29.95	26:03.58
RRT-Sağ	26:20.79	25:31.29	24:38.43	25:30.17
RRT*-Sol	45:24.15	38:35.71	33:53.49	39:17.78
RRT*-Sağ	34:02.28	33:11.62	33:19.71	33:31.20
RRTCon.-Sol	23:54.40	36:06.81	24:30.71	28:10.64
RRTCon.-Sağ	24:53.51	27:06.96	25:37.17	25:52.55
EST-Sol	27:26.21	29:40.67	28:37.24	28:34.71
EST-Sağ	27:57.94	26:49.73	26:33.45	27:07.04
BiEST-Sol	26:01.29	25:15.89	25:55.85	25:44.34
BiEST-Sağ	26:08.30	25:47.53	26:16.96	26:04.26
PRM-Sol	29:27.44	27:26.18	27:43.53	28:12.38
PRM-Sağ	28:25.82	28:31.74	27:42.99	28:28.78
PRM*-Sol	36:11.85	36:43.32	39:07.91	37:21.03
PRM*-Sağ	34:00.81	34:29.87	33:20.94	33:57.21



Şekil 16. Tam Test Sonuçları (Robot kolları bazında algoritmaların en iyi ve en kötü süreleri)

Şekil 16'daki grafikte, tam test sonucunda her algoritmanın her robot kolundaki en iyi ve en kötü görev tamamlama süreleri verilmiştir.



Şekil 17. Tam Test Sonuçları (Robot kolları bazında algoritmaların ortalama süreleri)

Şekil 17'deki grafikte ise tam test sonucundaki her algoritmanın her robot kolundaki ortalama görev tamamlama süreleri verilmiştir. Verilen tablo ve grafiklerde yer alan verilere göre BiEST ve RRT algoritmaları, anahtar performans göstergelerine göre tam testlerde de yeterli bir performans göstermiştir. Hızlı testlerde yeterli performans gösteren PRM algoritması tam testlerde aynı performansını gösterememiştir. Ortalamada yeterli bir performans gösterse de RRT algoritmasının sol ROKOS kolu için en kötü görev turu tamamlama süresine sahip olması nedeniyle çok da stabil bir çalışma göstermediği görülmektedir. Bu anlamda BiEST algoritması oldukça stabil görev tamamlama süreleri ortaya koymuştur.

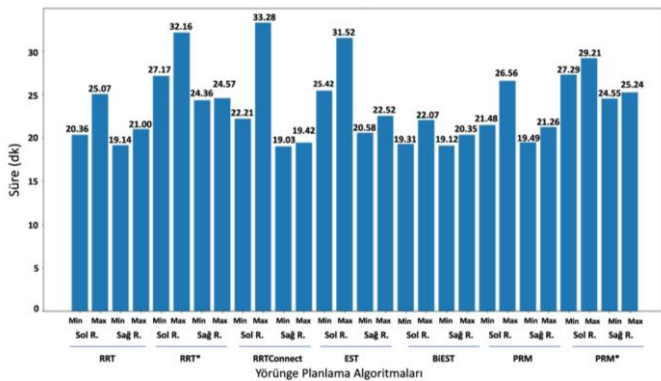
Gerçekleştirilen tam testin görev listesi Bölüm V.B'de bahsi geçen reset noktalarına sahip olan eski görev listesi üzerinde gerçekleştirilmiştir. Testin son bölümünde, reset noktaları çıkarılarak 330'dan 272 noktaya indirilmiş yeni

görev listeleri üzerinde (sağ ve sol ROKOS görev listeleri) tam testin tekrar edilmesi gerçekleştirilmiştir. Tablo 8’de güncel görev listesi üzerinde yapılan tam testin sonuçları ve önceki listeye nazaran elde edilen süre kazancı yüzdeleri görülebilir.

Tablo VIII. OMPL Algoritmaları Uygulanan Rokos Kollarının Güncel Görev Listesinde Tam Test Tamamlama Süreleri, Ortalaması ve Eski Görev Listesi Sürelerinden Kazanç Oranları

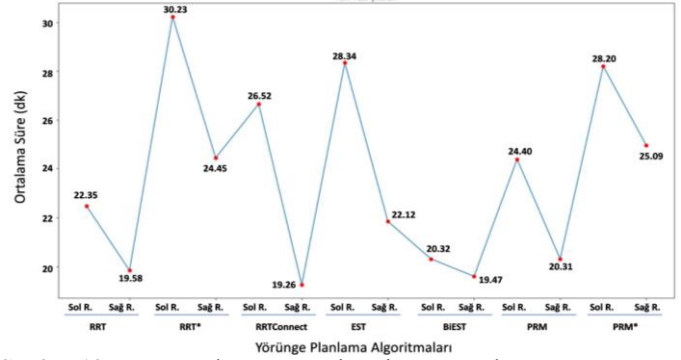
Planlayıcılar & Rokos Kolları	Tur 1	Tur 2	Tur 3	Ortalama Süre	Süre Kazancı
RRT-Sol	20:36.31	22:03.76	25:07.26	22:35.78	13.29%
RRT-Sağ	19:14.61	21:00.89	19:38.84	19:58.11	21.7%
RRT*-Sol	31:35.66	27:17.49	32:16.33	30:23.16	22.67%
RRT*-Sağ	24:42.85	24:57.86	24:36.64	24:45.78	26.12%
RRTCon.-Sol	24:47.58	22:21.49	33:28.19	26:52.42	4.63%
RRTCon.-Sağ	19:42.34	19:33.87	19:03.21	19:26.47	24.87%
EST-Sol	28:08.15	25:42.31	31:52.57	28:34.34	0.02%
EST-Sağ	20:58.26	22:52.11	22:45.94	22:12.10	18.13%
BiEST-Sol	22:07.90	19:31.22	19:57.43	20:32.18	20.21%
BiEST-Sağ	20:35.92	19:12.75	19:34.17	19:47.61	24.08%
PRM-Sol	21:48.94	26:56.35	25:14.94	24:40.08	12.54%
PRM-Sağ	19:49.56	21:26.52	20:19.86	20:31.98	27.9%
PRM*-Sol	29:21.22	27:29.25	28:10.61	28:20.36	24.13%
PRM*-Sağ	25:08.47	25:24.70	24:55.36	25:09.51	25.9%

Şekil 18’deki grafikte, tam test sonucunda her algoritmanın her robot kolundaki en iyi ve en kötü görev tamamlama süreleri verilmiştir.



Şekil. 18. Güncel Görev Listesine Uygulanan Tam Test Sonuçları (Robot kolları bazında algoritmaların en iyi ve en kötü süreleri)

Şekil 19’deki grafikte ise tam test sonucundaki her algoritmanın her robot kolundaki ortalama görev tamamlama süreleri verilmiştir.



Şekil. 19. Güncel Görev Listesine Uygulanan Tam Test Sonuçları (Robot kolları bazında algoritmaların ortalama süreleri)

Verilen tablo ve grafiklerde yer alan verilere göre güncel görev listesinde de BiEST ve RRT algoritmaları, anahtar performans göstergelerine bakıldığında başarılı bir performans göstermiştir. Görev listelerinden planlayıcı için gereksiz olabilecek noktalar çıkarıldığında görev tamamlama yüzdelerinin %27.9’a kadar arttığı gözlemlenmiştir. Ayrıca görev tamamlama sürelerinin 20 dakikanın altına indirilebildiği de görülmüştür. Buradan yola çıkılarak ROKOS sistemine uygulanan dinamik planlama sisteminin, az konum noktasında daha etkili çalıştığı sonucuna ulaşılmıştır.

Bu testlerle birlikte, ROKOS kolları çok tekrarlı bir şekilde görev listesinin ya bir kısmıyla ya da tamamıyla sınanmıştır. Hızlı test ile diğer yandan alternatif görev listelerinde robotların vereceği tepkiler de gözlenmiştir.

Bu çalışmalar kapsamında, OMPL planlayıcısına ait birçok hareket planlama algoritması üzerinde 400’ü aşkın, saatlerce süren testler gerçekleştirilmiş, en optimal planlama algoritması bulunmaya çalışılmıştır. Çalışma sonunda bu planlama algoritmasının BiEST olması gerektiği tespit edilerek testler tamamlanmıştır.

VI. SONUÇ VE DEĞERLENDİRME

Bu çalışma ROS araçlarından ve Gazebo simülasyon ortamının entegrasyonu ile ortaya çıkarılan SRVT sistemine, ROKOS robot kol sisteminin entegrasyonu ile SRVT içerisindeki bir diğer unsur olan Moveit! yazılımı aracılığıyla kullanılan hareket planlama algoritmalarının planlama performanslarının incelenmesi ve kıyaslanması üzerine gerçekleştirilmiştir. SRVT ekosistemi içerisine entegre edilmiş ROKOS robot kol sisteminin, görev tamamlama sürelerinin düşürülmesi için, bu sisteme en uygun yörünge planlama algoritmasının belirlenmesine odaklanılmıştır. Bu uygulamalara geçilmeden önce OMPL kütüphanesi içindeki birçok temel hareket planlama algoritması incelenmiş, bu algoritmaların hareket



planlama mantalitesinin, ROKOS sisteminin SRVT entegrasyonunda çalıştırılabilirliği ve verimliliği üzerinde çalışma yapılmıştır. Araştırmalar ve yapılan testler ışığında, ROKOS sisteminin görev tamamlama sürelerinin, uygun algoritmalar tespit edildiğinde oldukça düşürülebildiği görülmüştür.

Gelecekte ROKOS'un görev tamamlama sürelerinin 25 dakikanın altında stabil hale getirilmesi için algoritma test çeşitliliğinin artırılması, farklı görev listelerinde testler yapılması gibi optimizasyon çalışmalarına odaklanılması planlanmaktadır.

BİLGİLENDİRME [ACKNOWLEDGMENT]

The research leading to this paper has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876852. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Czech Republic, Germany, Ireland, Italy, Portugal, Spain, Sweden, Turkey. The views expressed in this document are the sole responsibility of the authors and do not necessarily reflect the views or position of the European Commission.

Bu çalışma İnovasyon Mühendislik tarafından yürütülen 120N803 nolu "Otomatik Sistemlerin Emniyet ve Güvenliğinin Doğrulanması ve Geçerlenmesi" 1071 projesi kapsamında Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenmiştir.

ÇIKAR ÇATIŞMASI [CONFLICTS OF INTEREST]

Yazarlar arasında ve ilgili kurumları arasında herhangi çıkar çatışması olmadığını bildirmişlerdir.

ETİK KURALLARA UYGUNLUK [RESEARCH AND PUBLICATION ETHICS]

Yazarlar bu makalenin etik kurul onayı veya herhangi bir özel izin gerektirmediğini beyan ederler.

KAYNAKLAR

- 1 Ragel, R., Maza, I., Caballero, F., & Ollero, A. (2015, November). Comparison of motion planning techniques for a multi-rotor UAS equipped with a multi-joint manipulator arm. In 2015 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS) (pp. 133-141). IEEE.
- 2 Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., ... & Ng, A. Y. (2009, May). ROS: an open-source Robot Operating System. In ICRA workshop on open source software (Vol. 3, No. 3.2, p. 5).
- 3 Gazebo website. [Online]. Available: <http://gazebo.org/>, (2021)
- 4 Chitta, S., Sucas, I., & Cousins, S. (2012). Moveit![ros topics]. IEEE Robotics & Automation Magazine, 19(1), 18-19.

- 5 Sucas, I. A., Moll, M., & Kavraki, L. E. (2012). The open motion planning library. IEEE Robotics & Automation Magazine, 19(4), 72-82.
- 6 Zucker, M., Ratliff, N., Dragan, A. D., Pivtoraiko, M., Klingensmith, M., Dellin, C. M., ... & Srinivasa, S. S. (2013). Chomp: Covariant hamiltonian optimization for motion planning. The International Journal of Robotics Research, 32(9-10), 1164-1193.
- 7 Kalakrishnan, M., Chitta, S., Theodorou, E., Pastor, P., & Schaal, S. (2011, May). STOMP: Stochastic trajectory optimization for motion planning. In 2011 IEEE international conference on robotics and automation (pp. 4569-4574). IEEE.
- 8 LaValle, S. M., Kuffner, J. J., & Donald, B. R. (2001). Rapidly-exploring random trees: Progress and prospects. Algorithmic and computational robotics: new directions, 5, 293-308.
- 9 Kuffner, J. J., & LaValle, S. M. (2000, April). RRT-connect: An efficient approach to single-query path planning. In Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065) (Vol. 2, pp. 995-1001). IEEE.
- 10 Karaman, S., & Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. The international journal of robotics research, 30(7), 846-894.
- 11 Kavraki, L. E., Svestka, P., Latombe, J. C., & Overmars, M. H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. IEEE transactions on Robotics and Automation, 12(4), 566-580.
- 12 Marble, J. D., & Bekris, K. E. (2013). Asymptotically near-optimal planning with probabilistic roadmap spanners. IEEE Transactions on Robotics, 29(2), 432-444.
- 13 Plaku, E., Bekris, K. E., Chen, B. Y., Ladd, A. M., & Kavraki, L. E. (2005). Sampling-based roadmap of trees for parallel motion planning. IEEE Transactions on Robotics, 21(4), 597-608.
- 14 Sucas, I. A., & Kavraki, L. E. (2011). A sampling-based tree planner for systems with complex dynamics. IEEE Transactions on Robotics, 28(1), 116-131.
- 15 Görner, M., Haschke, R., Ritter, H., & Zhang, J. (2019, May). Moveit! task constructor for task-level motion planning. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 190-196). IEEE.
- 16 Gasparetto, A., Boscariol, P., Lanzutti, A., & Vidoni, R. (2012). Trajectory planning in robotics. Mathematics in Computer Science, 6(3), 269-279.
- 17 Sciavicco, L., Siciliano, B., Villani, L., Oriolo, G.: Robotics. Modelling, Planning and Control. Springer, London (2009)
- 18 Krishnaswamy, K., Sleeman, J., & Oates, T. (2011, May). Real-time path planning for a robotic arm. In Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments (pp. 1-4).
- 19 Roy, R., Mahadevappa, M., & Kumar, C. S. (2016). Trajectory path planning of EEG controlled robotic arm using GA. Procedia Computer Science, 84, 147-151.
- 20 Hao, W. G., Leck, Y. Y., & Hun, L. C. (2011, May). 6-DOF PC-Based Robotic Arm (PC-ROBOARM) with efficient

- trajectory planning and speed control. In 2011 4th International Conference on Mechatronics (ICOM) (pp. 1-7). IEEE.
- 21 Xie, B., Zhao, J., & Liu, Y. (2011, June). Human-like motion planning for robotic arm system. In 2011 15th International Conference on Advanced Robotics (ICAR) (pp. 88-93). IEEE.
- 22 Iqbal, J., Islam, R. U., & Khan, H. (2012). Modeling and analysis of a 6 DOF robotic arm manipulator. *Canadian Journal on Electrical and Electronics Engineering*, 3(6), 300-306.
- 23 Gasparetto, A., Boscaroli, P., Lanzutti, A., & Vidoni, R. (2015). Path planning and trajectory planning algorithms: A general overview. *Motion and operation planning of robotic systems*, 3-27.
- 24 Streinu, I. (2000, November). A combinatorial approach to planar non-colliding robot arm motion planning. In *Proceedings 41st Annual Symposium on Foundations of Computer Science* (pp. 443-453). IEEE.
- 25 Xinyu, W., Xiaojuan, L., Yong, G., Jiadong, S., & Rui, W. (2019). Bidirectional potential guided RRT* for motion planning. *IEEE Access*, 7, 95046-95057.
- 26 Savsani, P., Jhala, R. L., & Savsani, V. J. (2013, April). Optimized trajectory planning of a robotic arm using teaching learning based optimization (TLBO) and artificial bee colony (ABC) optimization techniques. In *2013 IEEE International Systems Conference (SysCon)* (pp. 381-386). IEEE.
- 27 Cousins, S. (2010). Ros on the pr2 [ros topics]. *IEEE Robotics & Automation Magazine*, 17(3), 23-25.
- 28 Wurm, K. M., Hornung, A., Bennewitz, M., Stachniss, C., & Burgard, W. (2010, May). OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems. In *Proc. of the ICRA 2010 workshop on best practice in 3D perception and modeling for mobile manipulation (Vol. 2)*.
- 29 LaValle, S. M. (1998). Rapidly-exploring random trees: A new tool for path planning.
- 30 LaValle, S. M., & Kuffner Jr, J. J. (2001). Randomized kinodynamic planning. *The international journal of robotics research*, 20(5), 378-400.
- 31 Sagan, H. and J. Holbrook: "Space-filling curves", Springer-Verlag, New York, 1994.
- 32 Karaman, S., & Frazzoli, E. (2010). Incremental sampling-based algorithms for optimal motion planning. *Robotics Science and Systems VI*, 104(2).
- 33 Sniedovich, M. (2006). Dijkstra's algorithm revisited: the dynamic programming connexion. *Control and cybernetics*, 35(3), 599-620.
- 34 Hsu, D., Latombe, J. C., & Motwani, R. (1997, April). Path planning in expansive configuration spaces. In *Proceedings of International Conference on Robotics and Automation (Vol. 3, pp. 2719-2726)*. IEEE.
- 35 Coleman, D., Sucan, I., Chitta, S., & Correll, N. (2014). Reducing the barrier to entry of complex robotic software: a moveit! case study. *arXiv preprint arXiv:1404.3785*.
- 36 Moll, M., Sucan, I. A., & Kavraki, L. E. (2014). An extensible benchmarking infrastructure for motion planning algorithms. *arXiv preprint arXiv:1412.6673*.
- 37 Cohen, B., Sucan, I. A., & Chitta, S. (2012, October). A generic infrastructure for benchmarking motion planners. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 589-595). IEEE.
- 38 Liu, S., & Liu, P. (2021). Robot Motion Planning Benchmarking and Optimization using Motion Planning Pipeline.
- 39



RESEARCH ARTICLE

Comparative Analysis of Machine Learning Algorithms Based on Variable Importance Evaluation

^{1,*} Hasan YILDIRIM

^{1,*} Karamanoğlu Mehmetbey University, Kamil Özdağ Science Faculty, Mathematics, Karaman, Turkey,
hasanyildirim@kmu.edu.tr, Orcid: 0000-0003-4582-9018.

HIGHLIGHTS

- Machine learning algorithms may provide useful insights on variable importance via effective tuning.
- The cubist algorithm can present predictive results on property value assessments.
- The cascade of machine learning and statistical methods are powerful tools for the property valuation.

Keywords:

- Cubist
- Random Forest
- Machine Learning
- MARS
- Variable Importance

GRAPHICAL ABSTRACT

One of the main goals in machine learning studies is to determine the most significant variables on a specific research problem. Various algorithms have been developed to achieve this goal. Random forest, Cubist, and MARS algorithms are the most common ones among these algorithms. Although classical statistical algorithms have been useful to obtain the importance level of the effective variables on the output in a certain amount, the machine learning algorithms may provide clearer and more precise results. In this study, the estimation results of Random Forest, Cubist, and MARS algorithms have been presented comparatively in terms of some performance criteria like mean squares error, the coefficient of determination, and mean absolute error by using a real data set. The results show that the performances of Random Forest and Cubist are similar amongst themselves but better than MARS. Additionally, the rank of the most important variables varies according to the type of algorithm. The concordance between algorithms is investigated from a statistical perspective and found satisfactory. Consequently, Random Forest, Cubist, and MARS can be considered effective and reasonable algorithms for both estimation performance and variable importance evaluation.

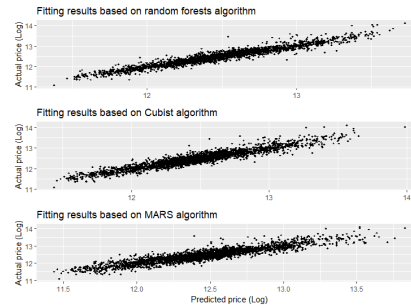


Figure A. Fitting results of whole data set based on each algorithm

Article Info:

Received : 2021-08-30
Accepted : 2021-10-06
Published : 2021-12-21

DOI:
10.53525/jster.988672

*Correspondence:

Hasan YILDIRIM
hasanyildirim@kmu.edu.tr
0 (338) 226 2000 - 3766.

Aim of Article: The main goals are (1) to compare the most effective machine learning algorithms by using a real and original estate data set and (2) to present the variable importance abilities.

Theory and Methodology: Based on a solid experimental setting and hyper-parameter tuning, including k-fold cross-validation, Cubist, Mars, and Random Forest algorithms have been compared to each other on real estate data set. Also, the Kendall W coefficient as a statistical significance test has been added to check the concordance of all algorithms.


Findings and Results: The main results show that all algorithms present reasonable results, but the Cubist is found as the best algorithm. The concordance between algorithms is significant, according to the Kendall W coefficient.

Conclusion: Machine learning algorithms may provide better insights in terms of variable importance evaluation than conventional methods like correlational or exploratory methods. With careful hyper-parameter tuning, each algorithm has both generalizability and the ability of variable importance determination.



RESEARCH ARTICLE

Comparative Analysis of Machine Learning Algorithms Based on Variable Importance Evaluation

1,*  Hasan YILDIRIM

1, * Karamanoğlu Mehmetbey University, Kamil Özdağ Science Faculty, Mathematics, Karaman, Turkey,
hasanyildirim@kmu.edu.tr, Orcid: 0000-0003-4582-9018.

Citation:

Yıldırım, H. (2021). *Comparative Analysis of Machine Learning Algorithms Based on Variable Importance Evaluation*, Journal of Scientific Technology and Engineering Research, 2(2): 46-53. DOI: 10.53525/jster.988672

HIGHLIGHTS

- Machine learning algorithms may provide reasonable insights on variable importance via effective tuning.
- The cubist algorithm can present predictive results on property value assessments.
- The cascade of machine learning and statistical methods are powerful tools for the property valuation.

Article Info

Received : 2021-08-30

Accepted : 2021-10-06

Published : 2021-12-21

DOI:

10.53525/jster.988672

*Corresponding Author:

Hasan YILDIRIM

hasanyildirim@kmu.edu.tr

0 (338) 226 2000 - 3766

ABSTRACT

One of the main goals in machine learning studies is to determine the most significant variables on a specific research problem. Various algorithms have been developed to achieve this goal. Random forest, Cubist, and MARS algorithms are the most common ones among these algorithms. Although classical statistical algorithms have been useful to obtain the importance level of the effective variables on the output in a certain amount, the machine learning algorithms may provide clearer and more precise results. In this study, the estimation results of Random Forest, Cubist, and MARS algorithms have been presented comparatively in terms of some performance criteria like mean squares error, the coefficient of determination, and mean absolute error by using a real data set. The results show that the performances of Random Forest and Cubist are similar amongst themselves but better than MARS. Additionally, the rank of the most important variables varies according to the type of algorithm. The concordance between algorithms is investigated from a statistical perspective and found satisfactory. Consequently, Random Forest, Cubist, and MARS can be considered effective and reasonable algorithms for both estimation performance and variable importance evaluation.

Keywords: Cubist, Random Forest, Machine Learning, MARS, Variable Importation

I. INTRODUCTION

The amount of data to be considered in research has been increased exponentially with rapid advances in technology. Besides the possibility of getting new insights by using big data, there have been some challenges including processing and storing data effectively. The necessity of reducing processing time increased the interest of researchers from different disciplines [1],[2] in this field. Although various analysis techniques have been proposed for especially high dimensional data set, it may be wise to reduce the size of the data set. There are two common ways for reducing the dimension of a data set: (1) to use some

dimensional reduction techniques (such as principal component analysis, locally linear embedding, t-SNE, etc.) and (2) to determine the most important attributes by using some techniques including intrinsic measurements (such as Random forest, partial least squares, etc.). In the first case, the idea is to obtain new attributes representing the actual ones to be explained the data set as much as possible. Thus, the dimension is reduced in a preprocessing step. The idea in the second case is to omit the least important attributes from the data set. During this process, the redundant or irrelevant attributes are determined and excluded. The decision of exclusion is researcher-dependent. The purpose of this process is to help the researcher to filter or determine the most important determinants in



measures and take a decision wisely. Herewith feature extraction provides some advantages such as better learning performance, lower computational cost, smaller memory necessity, and models with better generalization performance [3].

There have been many studies on feature selection, feature exclusion, and feature extraction. Most studies focus on a particular task, such as regression, classification, or clustering. Although the main idea is to reach the advantages mentioned in Li et al. [3], different approaches may be adopted. Hall and Smith [4] proposed a filter approach based on correlation and used it as a wrapper feature selection technique. Guyon and Elisseeff [5] carried out a review study on feature selection and feature extraction in a broad context, including feature construction, feature ranking, multivariate feature selection, efficient search algorithms, and feature validity assessment algorithms. The filter, wrapper, and embedded algorithms for feature selection were summarized in Saeys et al. [6] study. Alelyani et al. [7] focused on feature selection algorithms for clustering tasks only. In Chandrashekar and Şahin's [8] study, the comparative results using well-known benchmark data sets to conduct variable elimination processes were presented. Tang et al. [9] summarized the feature selection algorithms for classification tasks only. Li et al. [3] suggested a data perspective process on feature selection and gave a comprehensive overview of recent advances in this field. From the perspective of big data analytics, El-Hasnony [10] proposed some heuristic algorithms to improve the feature selection performance of machine learning algorithms. Karasu et al. [11] developed a wrapper-based method based on particle swarm optimization to achieve a more predictive regression model. More detailed explanations related to heuristic feature selection algorithms can be found in the study of Sharma and Kaur [12]. Feature selection methods can be classified into three parts as in most of the literature: (1) Filter algorithms, (2) Wrapper algorithms, and (3) Embedded algorithms. The embedded methods correspond to the algorithms having intrinsic measurements for determining the most important features.

In this study, we focused on some embedded algorithms, including, Random forest, multivariate adaptive regression splines, and the cubist, and compared them in terms of the performance on variable selection. The following sections are organized as follows: In Section 2, the details of algorithms have

been presented. The results of the performance comparison have been given in Section 3. The conclusions have been summarized in Section 4.

II. ALGORITHMS

In this section, the details of algorithms used in this study were given in brief.

A. *Random forest*

Random forest is a very well-known machine learning algorithm and proposed by Leo Breiman [13]. The idea behind the Random forest algorithm is to create decision trees as independently as possible and combined them to obtain a single and strong learner. Random forest is a very powerful and computationally efficient algorithm in consequence of considering only a fraction of feature space for each split in a decision tree. This property makes Random forest superior because of the reduction in the variance of the estimates. The main reason for the high variance in estimation results is to have some correlated results which commonly occur during the usage of all possible features for building trees and the possibility of being dominant for some important features in the whole process. As a single tree, i.e., base learner, CART, or conditional inference trees are used commonly. For more detailed theoretical explanations about this algorithm, James et al. [14] and Hastie et al. [15] are suggested to the reader.

B. *Multivariate Adaptive Regression Splines*

Multivariate adaptive regression splines (MARS) is proposed by Friedman [16]. MARS algorithm is a non-parametric regression and highly effective algorithm for especially high dimensional settings. Unlike classical regression algorithms, it does not need any assumption for the underlying distribution of data sets.

The main idea behind MARS is to model the data by using a set of surrogate features instead of the original measurements. These new features are the hinge functions of the original data. A hinge function whose cutting (or threshold) value is equal to c , is usually expressed as $h(x - c)$ and $h(c - x)$. Based on the value of the feature, different parameter estimation for each possible hinge condition is obtained and added to the main regression equation. In other words, it creates a piecewise linear model by determining cutting points for a feature and getting the parameter estimates between the feature and a dependent variable.



The cutting point is found in each feature and the new model is created depending on this point. The model error based on each new model is calculated and the model and corresponding cutting point with the lowest error is used for the rest of the estimation process. The MARS algorithm has some superiority such as interpretability, being fast, suitable for the high dimensional task, conducting automatic feature selection process, and robustness to outliers [17]. For more detailed theoretical explanations about this algorithm, James et al. [14] and Hastie et al. [15] are suggested to the reader.

C. Cubist

Cubist is a rule-based and complement algorithm of C5.0 that is used for classification. It is based on different approaches proposed in Quinlan [18-20]. Cubist has some properties unlike other opponents: (1) different types of pruning, smoothing and creating rules process, (2) an optional boosting procedure, and (3) adjustable estimation with the possibility of choosing nearby units for the training data set [17]. The process of building a tree is similar to other decision algorithms, but Cubist carried out a different pruning by considering a weighted linear combination of two trees including an actual tree and the parent of it. The weights of each tree are calculated using a criterion based on the covariance of the tree residuals and the variance of the difference between the residuals. The

model with lower error has a larger weight compared to the other one. After determining the weights of each model, the adjusted error rate is calculated by removing each rule from the rule-set. If the adjusted error rate is increased when a rule is deleted, that rule is omitted from the set. For more detailed theoretical explanations about of this algorithm, James et al. [14] and Hastie et al. [15] are suggested to the reader.

III. EXPERIMENT

A. Description of Data Set

The data set belongs to a common task in machine learning field [21] and has been retrieved from a popular real estate website [22] by using entries in January & February 2018. The size of the data set is 3102 and the number of variables is 11. Seyhan, Çukurova, Yüreğir, and Sarıçam which are the most developed and crowded districts have been considered in the study. The name variables in the data set are location, age, credit status, size, distance to the city center, type of heating system, floor location, dues, number of rooms, number of bathrooms, the number of floors. The output (i.e., dependent) variable is the price of a house. The descriptive statistics for qualitative and quantitative variables are given in Table I and Table II, respectively. The number of houses and their percentages corresponding to each category of variables are presented in Table I.

Table I. Descriptive Statistics (Qualitative Variables)

	Categories	N	Percent
Location	Çukurova	1371	44.2
	Seyhan	1261	40.7
	Sarıçam	395	12.7
	Yüreğir	75	2.4
Credit status	Yes	2951	95.1
	No	151	4.9
Age	0-5	2079	67.0
	6-10	263	8.5
	11-15	377	12.2
	16-20	306	9.9
	21-25	57	1.8
	26-30	14	0.5
	31-35	3	0.1
	36-40	3	0.1
Heating system	Combi boilers	1801	58.1
	Air conditioning	974	31.4
	Central heating	302	9.7
	Stove	19	0.6
Dues	Floor heating	6	0.2
	Yes	3021	97.4
	No	81	2.6
	Total	3102	100.0



Table II. Descriptive Statistics (Quantitative Variables)

	Min	Max	Mean	SD
Size	30	400	160.99	45.730
# Rooms	1	8	3.81	0.797
# Bathrooms	0	6	1.56	0.521
# Floors	1	20	10.65	3.474
Floor location	1	20	5.27	3.623
Distance (km)	1.7	27.2	10.855	3.6785
Rental price (TL)	215	3480	1120	370
Price (TL)	65000	135000	282412	125684.89
Price (Log)	11.0822	14.1157	12.5511	11.7416

B. Experimental Settings

The data set is randomly split into training and testing sets with ratios of %70 and %30, respectively. The tuning parameters are determined by using a 10-fold cross-validation approach based on only training data set are listed as follows:

Random forest: The number of selected predictors for each split(m).

-(m): {2, 3, ... ,23}

Cubist: The number of committees (C) and the number of neighbors(n).

-(C): {10, 20, ... ,100} and (n): {2, 3, ... ,9}

MARS: The level of degree(d) and the number of prune (p)

-(d): {1,2} and (p): {2, 3, ... ,20}

The results are obtained via R software packages including caret[23], earth [24], Cubist [25], randomForest [26]. The natural logarithm of the output variable is modeled in this study. The basic CART learner is used as a base learner in the Random forest algorithm. RMSE, R^2 and MAE are used as the performance criteria.

Table III.

Comparison of training and testing performance

Algorithm	Split	Size	RMSE	R^2	MAE
Random forest	Train	2171	0.1612	0.8347	0.1190
	Test	931	0.1554	0.8639	0.1158
Cubist	Train	2171	0.1611	0.8352	0.1211
	Test	931	0.1552	0.8626	0.1167
MARS	Train	2171	0.1708	0.8148	0.1295
	Test	931	0.1668	0.8409	0.1278

C. Performance Comparison

In this part, the performance of Random forest, Cubist, and MARS algorithms for training and testing data is presented and given in Table III. According to the results of Table III, the performance of Random forest and Cubist algorithms are quite similar for both training and testing data. Based on RMSE and R^2 performance criteria, it is shown that Cubist is slightly better than random forest. On the other side, the MARS algorithm produces poorer results than these algorithms for all criteria.

The fitting results using best tuning parameters for Random forest, Cubist, and MARS algorithms are given in Fig.1. The relative superiority of Random forest and Cubist against MARS can be seen visually. The approximation performance for the higher and lower house prices via the MARS algorithm is more unstable than the opponents. Also, the linearity corresponding to the closeness between the estimated and actual prices of houses for Random forest and Cubist is more explicit than the MARS algorithm. Random forest, Cubist and MARS algorithms have intrinsic measurements for determining the most important features (i.e., variables). After fitting each algorithm using the best tuning parameters which are found via cross validation approach, the importance level and rank of each feature have been obtained. In Table IV, the results are given separately for each algorithm.

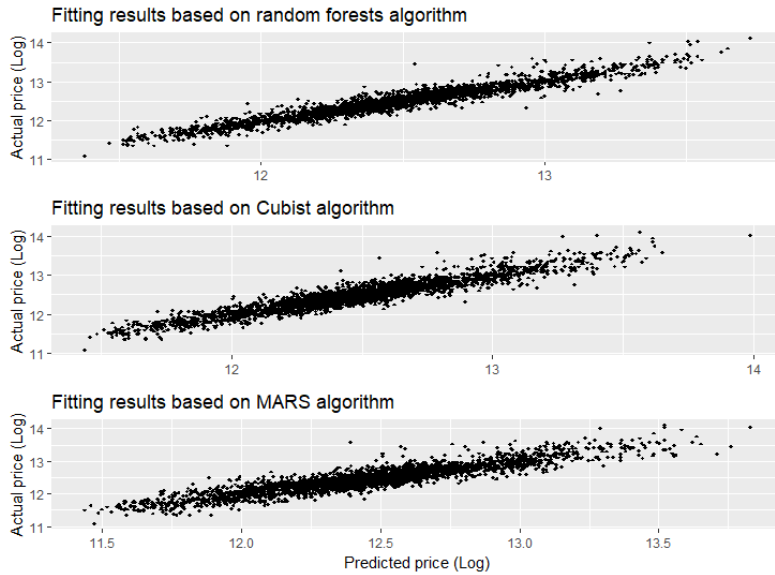


Figure 1. Fitting results of the whole data set based on each algorithm

According to Table IV, rental price, distance to the city center, the number of baths, the number of floors are commonly found the most important features for the price of a house by all algorithms. District (2) and district (4) have better rank values. These districts correspond to Seyhan and Yüreğir, respectively. Seyhan is the biggest district in Adana. On the other side, the economic level of Yüreğir is the lowest compared to the other districts. The reason for higher ranks may be the demand for Seyhan and the

purchasing power for Yüreğir. Heating (3) has the highest importance value among other categories. This corresponds the central heating. The top ten most important features for each algorithm are quite similar. It is shown that the MARS algorithm is more parsimonious on scoring importance compared to the Random forest and Cubist. Most dummy variables created for categorical features have no importance value according to the MARS algorithm.

Table IV. The results of variable importance evaluation based on each algorithm

Random Forest			Cubist			Mars		
Value	Variable	Rank	Value	Variable	Rank	Value	Variable	Rank
100	Distance	1	100	Rental price	1	100	Rental price	1
92,7	Baths	2	94,4	Distance	2	40,1	Baths	2
88,4	Number of Floors	3	83,9	Size	3	35,5	District(4)	3
87	Rental price	4	70,6	Baths	4	30,9	Rooms	4
76,4	District(2)	5	69,9	Number of floors	5	28,8	Floor	5
73,5	Size	6	58	Rooms	6	26,6	Distance	6
72,8	District(4)	7	57,3	District(2)	7	22,6	District(3)	7
63,8	Rooms	8	50,3	District(4)	8	18,5	Number of floors	8
52	Heating(3)	9	39,2	Age(4)	9	16,4	District(2)	9
46,9	Floor	10	37,1	Heating(3)	10	14,3	Heating(3)	10
44,8	District(3)	11	36,4	Floor	11	10	Size	11
32,6	Age(4)	12	32,9	Age(3)	12	0	Credit(2)	12
30,1	Age(3)	13	26,6	Age(2)	13	0	Age(2)	13
20,4	Credit(2)	14	25,2	District(3)	14	0	Age(3)	14
20,1	Age(2)	15	11,9	Age(5)	15	0	Age(4)	15
11,5	Heating(2)	16	7,7	Heating(2)	16	0	Age(5)	16
9,2	Heating(5)	17	2,8	Credit(2)	17	0	Age(6)	17
9	Dues(2)	18	2,1	Age(6)	18	0	Age(7)	18
8,8	Age(6)	19	0	Age(7)	19	0	Age(8)	19
6,8	Age(7)	20	0	Age(8)	20	0	Heating(2)	20
5,3	Heating(4)	21	0	Heating(4)	21	0	Heating(4)	21
2,7	Age(5)	22	0	Heating(5)	22	0	Heating(5)	22
0	Age(8)	23	0	Dues(2)	23	0	Dues(2)	23



Although Table IV provides very insightful and clear results on determining a house price, it may be useful to investigate the concordance between algorithms. The importance value and corresponding ranks are more similar between Random forest and Cubist, unlike MARS. The statistical examination of concordance is carried out by using Kendall's W coefficient of statistics. According to Kendall W statistics, Random forest, Cubist, and MARS algorithms statistically significantly matched in scoring ranks of important features ($W = .913, p < 0.0001$).

IV. CONCLUSION

In this study, we investigated the performance of Random forest, Cubist, and MARS algorithms in terms of accuracy and variable importance evaluation. The tuning parameters for each algorithm have been obtained via 10-fold cross-validation. Furthermore, the concordance between estimated ranks of features has been examined using Kendall's W coefficient. The results show that Random forest and Cubist present similar performance with each other but better than MARS. Based on overall performance, Cubist is found as the best among these algorithms. On the other hand, the top ten most important features are more common for all algorithms. The agreement of three algorithms on ranking features is found satisfactory.

Consequently, Random forest, Cubist and MARS algorithms produce not only good estimation performance, but also clear insights for variable importance evaluation. These kinds of algorithms can provide some useful insights from the inferential perspective and can be seen as functional tools from the pre-processing steps of all information-based applications in any area. The main advantage of this approach is to improve computational cost with a better understanding for particularly expert systems.

V. LIMITATIONS AND FUTURE STUDIES

One of the main limitations of this study is to be based on only non-heuristic feature selection algorithms. Because of that, a wide range of machine learning algorithms has feature selection properties, a comprehensive comparison study can be conducted. In future studies, we consider developing some novel algorithms that combined both wrapper-like and heuristic approaches via variable dimensions of data and tasks.

CONFLICTS OF INTEREST

No conflict of interest or common interest has been declared by the authors.

RESEARCH AND PUBLICATION ETHICS

In the studies carried out within the scope of this article, the rules of research and publication ethics were followed.

REFERENCES

- [1] Ertoý, U. & Akçay, M. (2021). Covid-19 Virüsü Salgını İle Mücadelede Büyük Veri Çalışmaları: Çin Örneği . *Journal of Scientific, Technology and Engineering Research* , 2 (2) , 4-14 . DOI: 10.5281/zenodo.4718425.
- [2] Pazar, Ş. , Bulut, M. & Uysal, C. (2020). Yapay Zeka Tabanlı Araç Algılama Sistemi Geliştirilmesi . *Journal of Scientific, Technology and Engineering Research* , 1 (1) , 31-37 . DOI: 10.5281/zenodo.3922425
- [3] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6), 1-45. DOI: 10.1145/3136625.
- [4] Hall, M. A., & Smith, L. A. (1999, May). Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In *FLAIRS conference (Vol. 1999, pp. 235-239)*. DOI: 10.5555/646812.707499.
- [5] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182. DOI: 10.5555/944919.944968.
- [6] Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19), 2507-2517. DOI: 10.1093/bioinformatics/btm344.
- [7] Alelyani, S., Tang, J., & Liu, H. (2018). Feature selection for clustering: A review. *Data Clustering*, 29-60. DOI: 10.1201/9781315373515-2.
- [8] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28. DOI: 10.1016/j.compeleceng.2013.11.024.
- [9] Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37. DOI: 10.1201/b17320.
- [10] El-Hasnony, I. M., Barakat, S. I., Elhoseny, M., & Mostafa, R. R. (2020). Improved feature selection model for big data analytics. *IEEE Access*, 8, 66989-67004. DOI: 10.1109/ACCESS.2020.2986232.
- [11] Karasu, S., Altan, A., Bekiros, S., & Ahmad, W. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy*, 212, 118750. DOI: 10.1016/j.energy.2020.118750
- [12] Sharma, M., & Kaur, P. (2021). A Comprehensive Analysis of Nature-Inspired Meta-Heuristic Techniques for Feature Selection Problem. *Archives of*



- Computational Methods in Engineering, 28(3). DOI: 10.1007/s11831-020-09412-6.
- [13] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32. DOI: 10.1023/A:1010933404324.
- [14] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *Statistical learning*. In *An introduction to statistical learning* (pp. 15-57). Springer, New York, NY. DOI: 10.1007/978-1-0716-1418-1_2.
- [15] Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer. DOI: 10.1007/978-0-387-84858-7.
- [16] Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1-67. DOI: 10.1214/aos/1176347963.
- [17] Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer. DOI: 10.1007/978-1-4614-6849-3.
- [18] Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, 27(3), 221-234. DOI: 10.1016/S0020-7373(87)80053-6.
- [19] Quinlan, J. R. (1992, November). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348). DOI: 10.1142/9789814536271
- [20] Quinlan, J. R. (1993, June). Combining instance-based and model-based learning. In *Proceedings of the tenth international conference on machine learning* (pp. 236-243). DOI: 10.5555/3091529.3091560.
- [21] Yıldırım, H. (2019). Property value assessment using artificial neural networks, hedonic regression and nearest neighbors regression methods. *Selcuk University Journal of Engineering, Science and Technology*, 7(2), 387-404. DOI: 10.15317/Scitech.2019.207.
- [22] Zingat. An online real estate website. <https://www.zingat.com>. (Last Access: March, 2018).
- [23] Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software*, 28(1), 1-26. DOI: 10.18637/jss.v028.i05.
- [24] Milborrow, S. (2019). earth: Multivariate Adaptive Regression Splines. R package version 5.1.1.
- [25] Kuhn, M., Weston, S., Keefer, C., Coulter, N., & Quinlan, R. (2014). Cubist: Rule-and instance-based regression modeling, R package version 0.0. 18.
- [26] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22. DOI: 10.1021/ci034160g.



RESEARCH ARTICLE

On modeling of a recurrent neural network from neural spiking data.

* Mohammed Al-AKAM, ¹ R. Ozgur DORUK

* Modeling and Design of Engineering Systems PhD Programme Student at Atılım University, 06830, Ankara, TURKEY
 Orcid. 0000-0003-4774-2645

¹ Department of Electrical and Electronics Engineering, Atılım University, 06830, Ankara, TURKEY
resat.doruk@atilim.edu.tr Orcid. 0000-0002-9217-0845

HIGHLIGHTS

- A fusion of statistical analysis, modeling, and biology
- Guides the computational neuroscientists in modeling of computational features of the neuron models from realistic neural spiking data.
- Assessment of the estimation accuracy under the existence of temporal only data (no amplitude)

GRAPHICAL ABSTRACT

We present a theoretical and computational work aiming at the estimation of an excitatory and inhibitory recurrent neural network from realistic stimulus-response data. The neural network incorporates firing rates of the excitatory and inhibitory units as state variables. The stimulus and response recordings are taken from a previous study which performs a measurement on the H1 neurons of the order Diptera flies. The parameter estimation is performed by maximum likelihood method. As the stimulus-response data is a single recording of 20 minutes, it is segmented, and individual segments are superimposed on each other to increase the statistical content of information. The true values of the model parameters are unknown as we are not using synthetic data. Because of this fact, two sample Kolmogorov-Smirnov test is applied to compare the interspike intervals of the recorded and model responses. Estimation and analysis results are presented in tabular and graphical forms. In addition, a comparison with previous research employing a modified Fitzhugh-Nagumo model is made.

Keywords:

- Firing rate based neural networks
- Excitatory-inhibitory neurons
- Neural spiking,
- Kolmogorov-Smirnov tests
- Interspike intervals

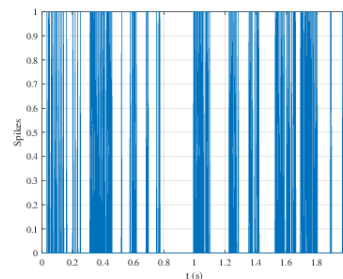


Figure A. The response of the H1 neurons of true flies of order Diptera to white noise visual stimulus

Article Info:

Received : 22.09.2021
 Accepted : 20.10.2021
 Published : 21.12.2021

DOI: 10.53525/jster.999008

*Correspondence:

R. Özgür DORUK
resat.doruk@atilim.edu.tr

Aim of Article : A theoretical study aiming at the modeling of an excitatory-inhibitory neural network from neural spiking data.

Theory and Methodology: The methodology is based on maximum likelihood method where the likelihood function is derived from the probability mass function of an Inhomogeneous Poisson Process. The optimization algorithm is of constrained type based on interior point methods. The measurement data sets are the neural spiking instants of the **H1 neurons of the order Diptera flies**.

Findings and Results: As one does not know the true parameters of the model in consideration, the accuracy of the findings are assessed using two sample Kolmogorov-Smirnov tests.

Conclusion : The results suggests that the p-Value obtained from Kolmogorov-Smirnov tests has a dependence of the number of segments and the individual segment lengths. The estimated parameters appear to be closer to each other when the segment size is kept in the segment lengths stay in the range (10,25) seconds.



RESEARCH ARTICLE

On modeling of a recurrent neural network from neural spiking data.

* Mohammed Al-AKAM, ¹ R. Ozgur DORUK

* Modeling and Design of Engineering Systems PhD Programme Student at Atılım University, 06830, Ankara, TURKEY
Orcid. 0000-0003-4774-2645

¹ Department of Electrical and Electronics Engineering, Atılım University, 06830, Ankara, TURKEY
resat.doruk@atilim.edu.tr Orcid. 0000-0002-9217-0845

Citation:

Al-Akam M., Doruk R. O. (2021). *On modeling of a recurrent neural network from neural spiking data.*, Journal of Scientific Technology and Engineering Research, 2(1): 54-66. DOI: 10.53525/jster.999008

HIGHLIGHTS

- A fusion of statistical analysis, modeling, and biology
- Guides the computational neuroscientists in modeling of computational features of the neuron models from realistic neural spiking data.
- Assessment of the estimation accuracy under the existence of temporal only data (no amplitude)

Article Info

Received : 22.09.2021
Accepted : 20.10.2021
Published : 21.12.2021

DOI: 10.53525/jster.999008

*Corresponding Author:

R. Özgür DORUK
resat.doruk@atilim.edu.tr
Phone: +90 312 5868000

ABSTRACT

We present a theoretical and computational work aiming at the estimation of an excitatory and inhibitory recurrent neural network from realistic stimulus-response data. The neural network incorporates firing rates of the excitatory and inhibitory units as state variables. The stimulus and response recordings are taken from a previous study which performs a measurement on the H1 neurons of the order Diptera flies. The parameter estimation is performed by maximum likelihood method. As the stimulus-response data is a single recording of 20 minutes, it is segmented, and individual segments are superimposed on each other to increase the statistical content of information. The true values of the model parameters are unknown as we are not using synthetic data. Because of this fact, two sample Kolmogorov-Smirnov test is applied to compare the interspike intervals of the recorded and model responses. Estimation and analysis results are presented in tabular and graphical forms. In addition, a comparison with previous research employing a modified Fitzhugh-Nagumo model is made.

Keywords: Firing rate based neural networks, Excitatory-inhibitory neurons, Neural spiking, Kolmogorov-Smirnov tests, interspike intervals

I. INTRODUCTION

A. Literature survey

The science related to neuron modeling dates well back to 1950s. The related developments led to an emerging field called as “theoretical and computational neuroscience”. The research community classified these models into three types namely the compartmental, cascade, and black box models. The compartmental models may involve single or multiple subsystems and thus are relatively complicated. They are

used when one needs to simulate one or more biophysical features of realistic neurons. Compartmental neural modeling is appeared after the development of the well known Hodgkin-Huxley model [23] in 1952. This was a highly nonlinear fourth order differential equation aiming at the description of quantitative features such as membrane potentials and ion channel conductances Following that, simpler models such as Fitzhugh-Nagumo [18], Morris-Lecar [35] and Hindmarsh-Rose [22] models appeared. Also models with more emphasis on specific subsystems of a



neuron such as dendrites, soma or axon are also met (such as [3]).

The cascade models on the other hand are less complex and place lesser emphasis on the computational details. These models are chosen when biophysical characteristics are not important. They generally consist of a linear filter and a nonlinearity that include some dynamical features [13]. Examples of studies involving such models are [30,24,38,1]. These are specifically related to visual subsystem.

The third type of modeling that can be met in computational neuroscience is based on a black box approach. In this case, the models focus on the ability of the neurons to process signals and they have a reasonable set of statistical features. An example is the representing the neuron's response by a probability distribution when the neuron is triggered by stimulus. Associated examples are [4,2,17,20].

Efficiency and applicability of a suggested model depends explicitly on the researcher's aim and the related mechanism of data collection. In the case of *in vivo* experiments, it is difficult to measure membrane potentials directly by placing an electrode on the membrane of the neuron under examination. That may cause a damage of the neuron or change in its activities. Those will obviously lead to incorrect measurements [11]. Alternatively, it is possible to collect data from an examined cell without touching it by placing an electrode in an area surrounding the cell membrane. Although this approach does not give us the correct amplitudes of the membrane potential, it is likely to detect the bursts of successive action. In other words, that yields information as a time series of successive peaks rather than a value of potential levels. This set of data is called as a neural spike train (or trains) which is similar to bursts and idle times that would be formed based on the neuron's activity. Bursts could be defined as closely spaced spikes, whereas the idle periods can be represented by separate longer-term transactions. Although they may seem meaningless at a first look, they include valuable information about the activity of the investigated neuron. The work in [39] suggests that a sudden jump in data during an instant discovers what actually happened inside a neural cell. In addition, the random distribution of those spikes within a specific time interval largely obeys an Inhomogeneous Poisson Process (IPP). Therefore, it is possible to apply likelihood function derived from probability mass function of a Poisson distribution [36]. The studies by [5, 37] proposed a system identification approach that estimates internal parameters using maximum likelihood for integrating and firing neuron models. Firing probabilities are used to derive likelihood function based on a local Bernoulli approximation. Similarly, a study in [7] used maximum likelihood estimation to characterize functionality between neurons rather than study neurons individually. The study [40] presented a data analysis problem that deals with the stimulus-response experiments

in neurophysiology. When the stimulus is known and the experiment is under the control, it is possible to record the spiking activities of the neurons. The study in concern proposed a state-space model to analyze response when the stimulus is implicit. Conditional intensity function is estimated based on general point processes that characterize spiking activities of neurons. This study also applied maximum likelihood algorithm to optimize parameters. The research in [21] claimed that the electrical signals in a biological neurons are random due to the chemical reactions in the ion channels and synaptic processes. The noise generated from the ion channels were described by mathematical models in some recent studies [33, 26, 41]. Other studies [28, 27, 42] stated that the movement of the ions in the channels produces electromagnetic fields that have influences on the membrane potential. However, the authors in [42] claimed that the operation of the heart can be determined through a magnetic fields produced by its electrical system.

Recently, the researchers have given more attention to the modeling of biological neurons by using artificial neural networks. For example, static feedforward neural networks are used by the researchers in some studies [10, 9, 8] to model auditory cortex. However, this type of neural networks does not sufficiently characterize the time dependent features of biological neurons such as action potentials and refractory periods. Therefore, dynamical neural networks seem to be better in the characterization of signal processing mechanisms of the biological neurons.

The literature regarding computational neuroscience has witnessed more attention to identify parameters of the neuron(s). Simple techniques such as minimum mean square estimation and spike synchronization could be utilized to calculate potential of membrane or firing rate when they are measurable. These methods become inapplicable when the collected data is discontinuous and there is no information about amplitude of amplitudes or rate of firing. Some studies such as [29] applied synchronization technique to spiking neurons. However, this technique requires a synchronization of two successive set of spikes that needs a reducing in inter-arrival time of spikes. The main problem in this method is the additional computational overhead introduced by the evaluation of the distance between two successive spikes. That will add extra complexity to the system.

B. About this work

As a difference from the studies performed by [15, 11, 13, 14] we will attempt to fit a firing rate based recurrent neural network model to a realistic stimulus-response data. One similar research is done by [12] where the response of H1 neurons of true flies of order Diptera [19] against a white noise visual stimulus is used. The data is available as a result of a research by [25]. In the aforementioned work ([12]), the model fitted to the H1 neuron stimulus-response pair is a modified Fitzhugh-Nagumo model [18] where the output

membrane potential is mapped to a firing rate variable by a specially formed sigmoid function. In this work, we will use an excitatory-inhibitory neural network model proposed by [34]. The model is formed by two differential equations representing the dynamics of the firing rates of excitatory and inhibitory units. The excitatory neuron's firing rate will be taken into consideration. The stimulus and neural spiking data will be the same as that of [12]. This will allow us to perform a comparison in the estimation performance of the two approaches. As we are not employing synthetically generated data from models with known parameters, no true parameter information will be available. After the estimation process is finalized, the model can be simulated with the estimated parameters, a firing rate profile and associated spike timings will be generated. The spike generation will make use of Inhomogeneous Point Process simulation with event rate being the firing rate of the excitatory neuron. The interspike intervals of the simulated spikes and the ones available from the data will be compared using Two Sample Kolmogorov-Smirnov statistical test [32, 31]. The p -values obtained from the tests will be presented in graphical forms. These results will be compared to the ones in [12].

II. MATERIALS AND METHODS

A. The neuron model

In this research, we will have a second order recurrent neural network model in the following form:

$$\begin{aligned} \dot{r}_e &= \beta_e \left[-r_e + \frac{F_e}{1 + e^{-a_e(w_{ee}r_e - w_{ei}r_i + I)}} \right] \\ \dot{r}_i &= \beta_i \left[-r_i + \frac{F_i}{1 + e^{-a_i(w_{ie}r_e - w_{ii}r_i + I)}} \right] \end{aligned} \quad (1)$$

In the above the subscripts (e) and (i) represent the association of the related variables with the excitatory and inhibitory units respectively. The model have some differences from the one presented in [15]. Main difference appears in the choice of the state variables. [15] models the dynamics of the membrane potentials (v_e and v_i) whereas (1) presents the dynamics of the firing rates (r_e and r_i) directly. The discussion on both types of models can be found in the reference [34]. The definitions of the parameters in (1) is available in Table 1.

Table 1: The definitions and nominal values of the parameters of the neural network model in (1)

Parameter	Definition
β_e	The reciprocal time constant of the excitatory neuron
β_i	The reciprocal time constant of the inhibitory neuron
w_{ee}	Weight representing the strength of excitatory autaptic input

w_{ei}	Weight representing the strength of inhibitory synaptic input to excitatory unit
w_{ie}	Weight representing the strength of excitatory synaptic input to inhibitory unit
w_{ii}	Weight representing the strength of inhibitory autaptic input
F_e	Maximum firing rate of excitatory unit
F_i	Maximum firing rate of inhibitory unit
a_e	Excitatory slope
a_i	Inhibitory slope

B. Stimulus and response data

In this research, we will apply the same stimulus and response pair as in [12]. That is a white noise based visual stimulus and the associated response obtained from H1 neurons. This data set is a 20 minute recording of the experimental measurements on the H1 neurons of true flies [19, 25]. For convenience, a plot of the stimulus and associated response can be seen in **Figures 1** and **2** respectively.

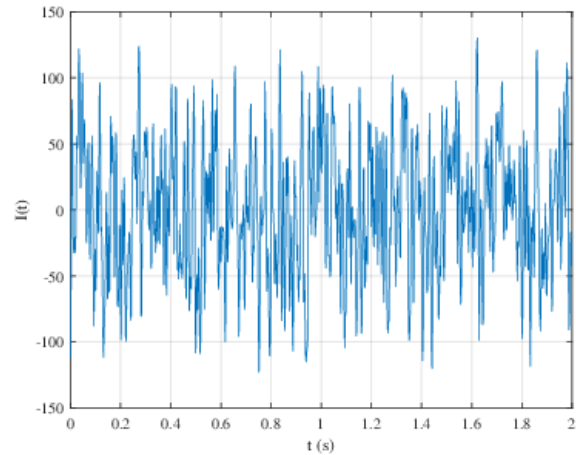


Figure 1: The white noise visual stimulus applied to the H1 neurons of true flies of order Diptera by [25]. Only the first two seconds are shown here to for the sake of proper displaying of its variation. Here the bin size is $\delta t = 2$ ms.

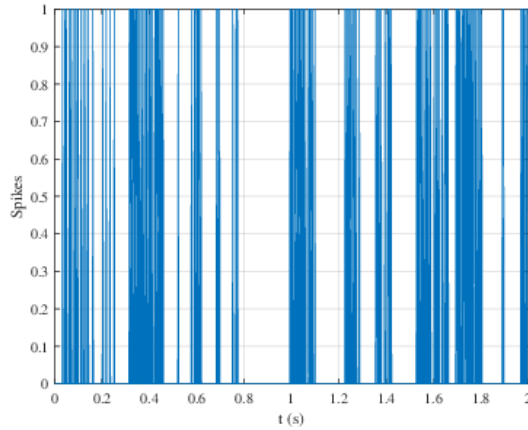


Figure 2: The response of the H1 neurons of true flies of order Diptera to white noise visual stimulus. The data is provided by [25]. Only the first two seconds are shown here to for the sake of proper displaying of its variation. Here the response is shown as spikes and each occurrence of a spike is shown in a binary fashion. i.e. A **1** for an existing spike and a **0** for no spike.

C. Inhomogeneous Poisson Processes

While introducing this study, it is stated that the neural spiking process largely obeys the Inhomogeneous Poisson Processes. In its simplest sense, it is a discrete random process characterized by an event rate λ and has a probability mass function defined by:

$$Prob [N(t + \Delta t) - N(t) = k] = \frac{e^{-\lambda} \lambda^k}{k!} \quad (2)$$

where k is the number of events that occur in the interval $[t, t + \Delta t)$. In homogeneous versions of the Poisson processes the λ is constant in that interval. In neural operation on the other hand, the process is much more complex and assuming a constant event rate will mostly be not sufficient and thus a time varying rate is needed. That actually is equivalent to the firing rate $r(t)$ of the neuron under examination. One can refer to r_e or r_i in (1). This will be an Inhomogeneous Poisson Point Process with the event rate λ replaced by a mean firing rate defined by:

$$\lambda = \int_t^{t+\Delta t} r(\tau) d\tau \quad (3)$$

where $r(t)$ will be $r_e(t)$ if the neural spiking of the excitatory neuron is considered. For the inhibitory neuron $r(t)$ should be replaced by $r_i(t)$. Now the term k represents

the number of spikes in the interval $[t, t + \Delta t)$. That is statistically related to the firing rate $r(t)$ and λ will represent the mean spike count corresponding to the time dependent firing rate $r(t)$. In (2), $N(\tau)$ stands for the cumulative total number of spikes up to time τ , thus making $N(t + \Delta t) - N(t)$ equal to the number of spikes for the time interval $[t, t + \Delta t)$.

Now, suppose that one has a spike train (t_1, t_2, \dots, t_K) in the time interval $(0, T)$. Here, $0 \leq t_1 \leq t_2 \leq \dots \leq t_K \leq T$, and so t and Δt become 0 and T . The spike train can be considered as a time series (including the time stamps for K spikes). So a likelihood function associated with any spike train (t_1, t_2, \dots, t_K) can be derived from an Inhomogeneous Poisson Process [16, 6] in the following way:

$$p(t_1, t_2, \dots, t_K) = \exp\left(-\int_0^T r(t, x, \theta) dt\right) \prod_{k=1}^K r(t_k, x, \theta) \quad (4)$$

The function yields the likelihood of a given spike train (t_1, t_2, \dots, t_K) that occurs with the rate function $r(t, x, \theta)$ which obviously is relying mainly upon model parameters and the stimulus.

An important point that is to be discussed here is the generation of test spikes for the statistical analysis. As we don't have any firing rates available from experimental measurements we will need to apply a statistical analysis on the spike timings. The estimated model's firing rate output should be converted to spike timings. In order to obtain those, one needs to perform an Inhomogeneous Poisson Process simulation using the firing rate output(s) of the model in (1) with the estimated parameters. The stimulus should be the same as the one used in the generation of the experimental data (provided by [25]). The simulation can be performed using the following scheme [16]:

- (1). Given the firing rate of a neuron as $r(t)$,
- (2). Find the probability of firing at time t_i by evaluating $p_i = r(t_i) \delta t$ where δt is the integration interval. It should be a small real number such as one milliseconds,
- (3). Draw a random number $x_{rand} = U[0,1]$ which is uniformly distributed in the interval $[0,1]$. Here, U stands for a uniform distribution,
- (4). If $p_i > x_{rand}$ fire a spike at $t = t_i$, else do nothing,
- (5). Collect spikes as $S = [t_1, \dots, t_{N_s}]$ where N_s will be the total number of spikes collected from one simulation.

Note that the above scheme requires sufficiently small bin sizes like 1 or 2 milliseconds.

D. Maximum Likelihood Approaches

The parameters in consideration is vectorized as follows:

$$\theta = [\beta_e, \beta_i, w_{ee}, w_{ei}, w_{ie}, w_{ii}, \alpha_e, \alpha_i, F_e, F_i] \quad (5)$$

The maximum probability here relies on the function proposed in (4) and includes each spike timing as well. Estimation theory asserts that determining maximum likelihood is asymptotically effective and goes as far as the Cramér-Rao bound when the data size increases. Thus, in order to expand the likelihood function in (4) to further cover settings with numerous spike trains initiated by numerous stimuli, a series of M stimuli should be assumed. Take the m -th stimulus ($m = 1, \dots, M$) to initiate a spike train containing K_m spikes in the time window $[0, T]$, and the spike timings are given by $S_m = (t_1^{(m)}, t_2^{(m)}, \dots, t_{K_m}^{(m)})$. By (4). According to (4), the probability function for the spike train S_m can be determined as:

$$p(S_m|\theta) = \exp\left(-\int_0^T r^{(m)}(t)dt\right) \prod_{k=1}^{K_m} r^{(m)}(t_k^{(m)}) \quad (6)$$

in which $r^{(m)}$ represents the firing rate against the m -th stimulus. Here it should be noted that the rate function $r^{(m)}$ entirely relies on the parameters related to neuron parameters θ and the applied stimulus. Supposing the stimulus and its elicited responses in each m^{th} trial are independent one can derive a joint likelihood function as:

$$L(S_1, S_2, \dots, S_M|\theta) = \prod_{m=1}^M p(S_m|\theta) \quad (7)$$

To improve its convexity, we can make use of natural logarithm and derive a log likelihood function as shown below:

$$l(S_1, S_2, \dots, S_M|\theta) = -\sum_{m=1}^M \int_0^T r^{(m)}(t)dt + \sum_{m=1}^M \sum_{k=1}^{K_m} \ln r^{(m)}(t_k^{(m)}) \quad (8)$$

Finally, the maximum likelihood estimates of the parameter vector θ is obtained by:

$$\hat{\theta}_{ML} = \operatorname{argmax}_{\theta} [l(S_1, S_2, \dots, S_M|\theta)] \quad (9)$$

E. Analysis approaches

As the true values of the parameters in Table 1 are not available to us, the only viable approach seems to perform a statistical analysis on the measured and simulated spike trains. One such approach is the Kolmogorov-Smirnov test [32, 31] which assesses whether two samples are drawn from the same distribution or not. MATLAB's `[h,p,ks2stat] = kstest2(x1,x2)` routine can perform this test for two arrays x_1 and x_2 respectively. The routine tests the null hypothesis that the data in vectors x_1 and x_2 are from the same continuous distribution using the two-sample Kolmogorov-Smirnov test. The alternative hypothesis is that x_1 and x_2 are from

different distributions. MATLAB routine can return both the h value and p value at 5% significance level. Here h value appears to be **1** when the test rejects the null hypothesis. The p value behaves reversely as it is the probability of observing a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis. As a result it will be closer to **1** when the null hypothesis is not rejected.

In this research, the test is performed on the interspike intervals (ISI) of the measured and simulated neural spike trains.

(1). After the parameter identification by maximum likelihood method completed, the model in (1) is simulated with the estimated parameters against the stimuli used in the estimation process (i.e. the one in **Figure 1**).

(2). The firing rate is recorded and an Inhomogeneous Poisson Process simulation is performed using the method introduced in **Section 2.C**.

(3). The interspike intervals (ISI) of both simulated spikes obtained in **Step 2** and the measured ones available from the data provided by [25] are evaluated.

(4). The interspike intervals are provided to two sample Kolmogorov-Smirnov test and the p -values are recorded.

(5). During the tests, the stimulus-response pair is segmented and the measured and simulated responses are superimposed. Different number of segments are examined. This is performed as a single train of spikes may not be statistically sufficient to reflect the statistical content of the response.

(6). In `[h,p,ks2stat] = kstest2(x1,x2)`, x_1 corresponds to ISI of measured spikes and x_2 corresponds to the ISI of the simulated spikes obtained in **Step 2**. The p values are taken into consideration in this research.

III. RESULTS

In this study, we attempt to estimate the parameters of an firing rate based neural network model representing excitatory and inhibitory behaviors. The mathematical model of this network is presented in (1) and the parameters to be estimated are available in Table 1. The stimulus-response pair is a realistic data obtained as a result of an experiment performed by [25]. Here the stimulus is a visual input of white noise type recorded for 20 minutes (First two seconds is available in **Figure 1**). The response is the firing instants received from the H1 neurons of the true flies of order Diptera (First two seconds is available in **Figure 2**). As we are not using synthetically generated data, we have no information on true parameters. The performance of the estimation is assessed using two sample Kolmogorov-Smirnov test on the interspike intervals of the recorded spiking data and responses of the model with estimated parameters.



The parameters are estimated by maximizing the joint likelihood function in (8). Optimization will be performed by MATLAB's `fmincon` routine. This allows the user to define constraints and we apply a lower bound of zero to ensure all the parameters remain in positive region. This will both ensure the stability of the algorithm and also the excitatory-inhibitory nature of the model in (1) is not disturbed. We only have a single recording of a relatively

longer duration. Thus, one will need to partition the stimulus and response into segments of predetermined length. The segmentation will be beneficial as responses falling into individual segments can be superimposed to increase statistical content of information. One can observe in **Table 2** the results obtained from maximum likelihood estimation of parameters.

Table 2: The estimated parameters θ_{ML} and their dependence to the length of the segments (T_{seg}). The number of segments are indicated by N_{seg} .

T_{seg}	N_{seg}	$\hat{\beta}_e$	$\hat{\beta}_i$	w_{ee}	w_{ei}	w_{ie}	w_{ii}	a_e	a_i	F_e	F_i
600.0	2	188.50	279.52	315.24	64.68	271.54	236.12	153.84	60.42	44.67	289.09
400.0	3	23.54	260.10	178.55	16.63	457.98	90.23	243.11	294.44	44.67	103.13
300.0	4	181.63	352.44	415.08	19.04	121.46	75.13	411.93	186.02	44.67	294.13
200.0	6	15.11	154.32	226.83	13.54	210.41	56.20	86.12	197.90	44.68	136.20
150.0	8	111.80	117.90	243.51	350.43	474.50	170.36	446.64	160.09	44.67	19.42
120.0	10	16.70	498.80	79.12	431.33	90.73	123.66	120.03	129.57	89.40	265.75
100.0	12	164.52	117.64	243.50	350.60	486.09	170.23	447.39	159.94	44.67	16.24
80.0	15	93.89	298.27	418.81	19.21	274.78	233.81	128.49	29.33	44.67	300.95
60.0	20	132.89	345.36	433.38	52.32	409.45	338.66	464.99	421.40	44.68	378.80
50.0	24	105.35	243.59	183.07	13.42	379.81	265.20	353.56	29.84	44.68	169.42
40.0	30	68.27	116.33	422.60	41.86	403.69	469.93	76.69	284.45	44.68	258.11
30.0	40	102.12	293.87	322.42	24.72	274.87	233.75	125.00	24.04	44.68	301.16
25.0	48	111.04	17.59	405.97	482.82	150.12	352.77	340.92	312.97	44.68	52.77
20.0	60	95.52	21.35	480.33	188.89	476.42	415.38	410.33	231.41	44.69	466.87
15.0	80	83.48	19.78	485.35	188.49	483.46	415.68	410.61	231.39	44.69	467.84
12.0	100	117.87	21.53	480.75	188.56	476.67	415.39	410.34	231.41	44.69	466.91
10.0	120	104.81	25.79	486.65	186.30	485.18	415.65	410.59	231.39	44.71	467.76
8.0	150	94.14	116.14	461.20	77.55	226.71	467.42	158.95	119.61	44.72	212.01
6.0	200	135.70	29.57	336.38	169.75	125.37	223.65	256.82	16.09	44.71	331.76
5.0	240	90.23	70.67	450.14	62.48	269.97	192.39	360.96	277.86	44.75	37.13
4.0	300	126.92	244.80	122.57	13.70	380.86	265.27	354.33	17.67	44.75	197.02
3.0	400	122.45	357.09	418.17	3.05	411.64	339.47	475.62	424.06	44.77	380.12
2.0	600	124.77	164.32	150.83	17.40	353.63	263.02	40.09	81.77	44.84	168.31
1.0	1200	110.37	7.46	340.41	165.29	349.10	388.43	499.99	155.79	45.04	278.53
0.5	2400	121.30	343.46	484.86	68.36	18.48	13.92	228.91	318.02	45.52	239.72

For the Kolmogorov-Smirnov analysis one can refer to the **Figures 3 - 17**. The figure explanations comment on the variation of p value with the super imposed number of samples (or samples N_{seg}). The test results show that regardless of the length of the segments (T_{seg}), the p value exceeds $p = 0.95$ level when number of samples are at least equal to $N_{seg} = 40$. When the length of the segments are restricted to a minimum of $T_{seg} = 5$ seconds, most of the cases yield $p > 0.95$ when we have a sample size of $N_{seg} = 30$ or more.

IV. CONCLUSION

A. Summary

In this work, we presented a framework for estimation of the parameters of a firing rate based excitatory-inhibitory dynamical neural network from a realistic set of stimulus-response data. The model is of second order type and has a total of 10 parameters that are to be estimated (1). The stimulus is a 20 minute recording of white noise visual excitation. The response is a binary array of neural spikes

that are received from H1 neurons of true flies of order Diptera [19, 25]. As a result, only the temporal locations of the spikes are available to us. Thanks to the evidence that the neural spiking phenomenon largely obeys the Poisson processes, we can employ a point process likelihood function such as (7) to perform an estimation of our model parameters. Contrary to the studies like [15, 13, 11, 14] we do not know the true value of our parameters (as we do not employ synthetic data here) and thus needed to apply two sample Kolmogorov-Smirnov statistical test to compare the interspike intervals of the recorded response and the one obtained after the simulation of the model in (1) with the estimated parameters.

B. Discussion of the results

In the analysis by Kolmogorov-Smirnov method, the resultant p value which shows the probability that the interspike interval distribution comes from two similar distributions. In the case that 5% significance level is sufficient, $p = 0.95$ level should be achieved to conclude that our spiking procedures are coming from two point

processes with the same event (thus firing) rate. If the significance level is pushed to 10%, then $p = 0.90$ should be reached. However, higher percentages for the significance level should be avoided.

Based on the results in **Figures 3-17**, one can note that the p value obtained from the two sample Kolmogorov-Smirnov analysis stays in ranges higher than $p = 0.95$ if the sample size (number of segments) remains at least $N_{seg} = 40$ for all cases. When the segment size is larger than $T_{seg} = 5$ seconds, the required number of samples is even smaller i.e. $N_{seg} = 30$. In **Table 2**, parameters appear to be closer to each other when the segment size is kept in the range $10 \leq T_{seg} \leq 25$ seconds. Interesting coincidence is that, the minimum required sample size is also lower than $N_{seg} = 40$.

When the results of [12] is examined, it appears that the required sample size to have p values higher than $p = 0.95$ requires larger number of samples. In **Table 3**, one is able to see some results from [12] which bears the segment sizes T_{seg} in seconds and required number of samples N_{seg} to have p values higher than $p = 0.95$. It is pretty obvious that, the model in [12] requires more samples to yield p values higher than $p = 0.95$. The minimum required sample size is $N_{seg} = 50$ and obtained when the segment size is $T_{seg} = 0.5$ seconds. For larger segment sizes, the required number of samples is even larger. So one can say that, the model in (1) has a better modeling capacity for H1 neurons of Diptera flies.

Table 3: The required sample size N_{seg} v.s. the segment size T_{seg} which yields p values higher than $p = 0.95$ for the model in [12].

T_{seg} (s)	N_{seg}
0.5	50
1	170
2	130
3	140
4	130
6	150

CONFLICTS OF INTEREST

They reported that there was no conflict of interest between the authors and their respective institutions.

RESEARCH AND PUBLICATION ETHICS

In the studies carried out within the scope of this article, the rules of research and publication ethics were followed.

FIGURES

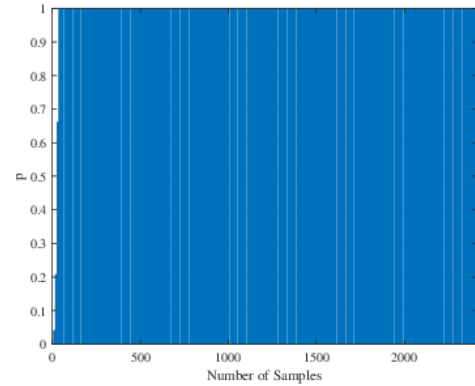


Figure 3: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 0.5$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 40$ or larger.

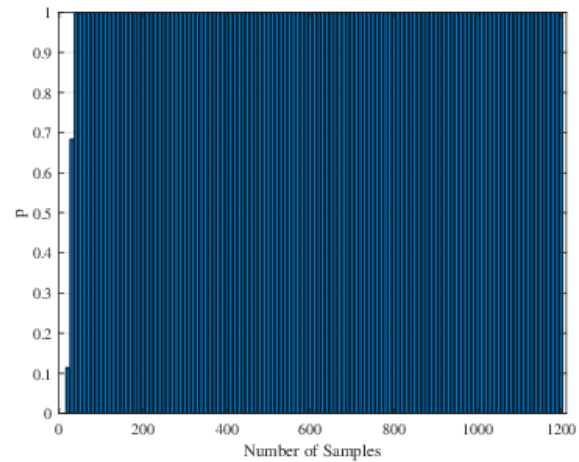


Figure 4: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 1$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 40$ or larger.

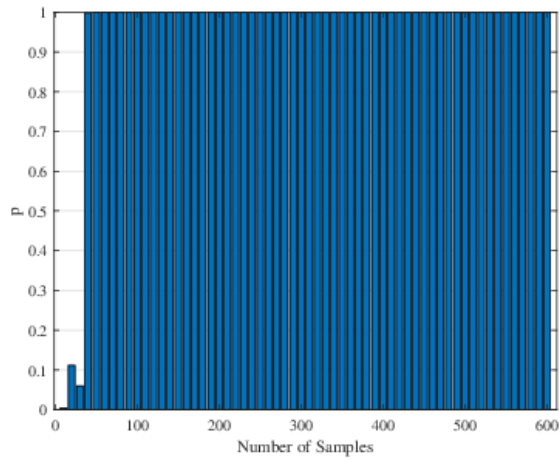


Figure 5: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 2$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 40$ or larger.

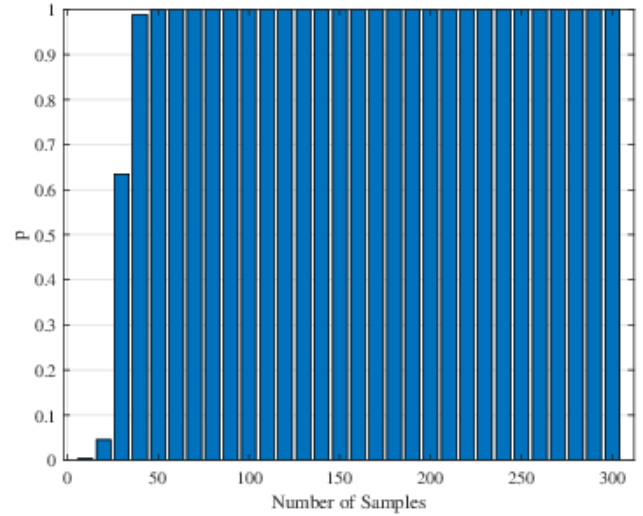


Figure 7: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 4$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 40$ or larger.

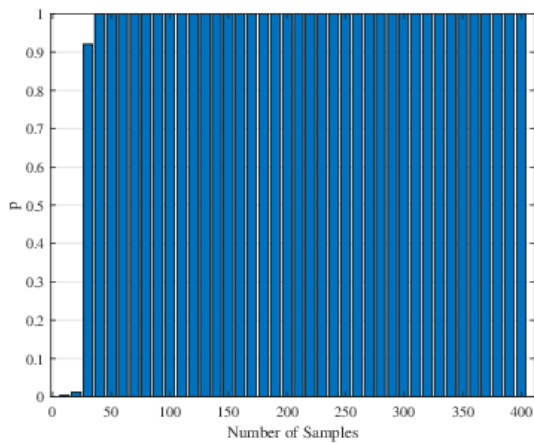


Figure 6: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 3$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 40$ or larger. Considering 10% significance p value reaches $p = 0.90$ when number of segments reach $N_{seg} = 30$.

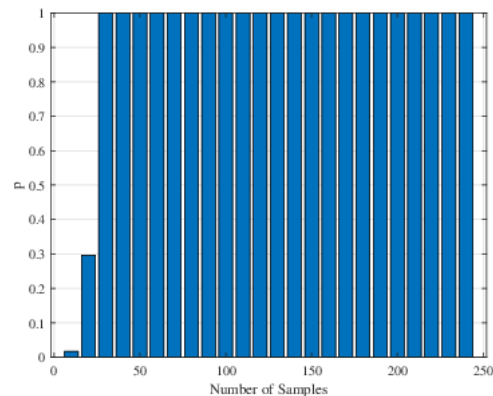


Figure 8: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 5$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

value reaches $p = 0.90$ when number of segments reach $N_{seg} = 30$.

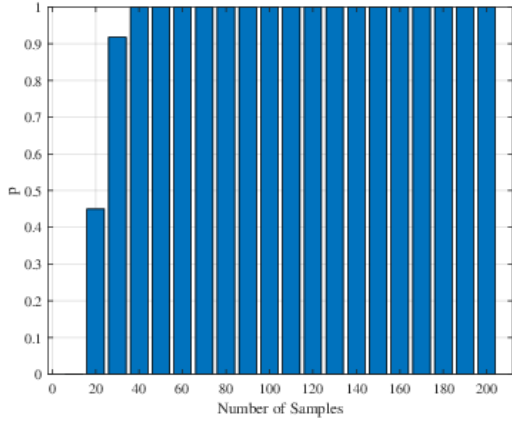


Figure 9: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 6$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

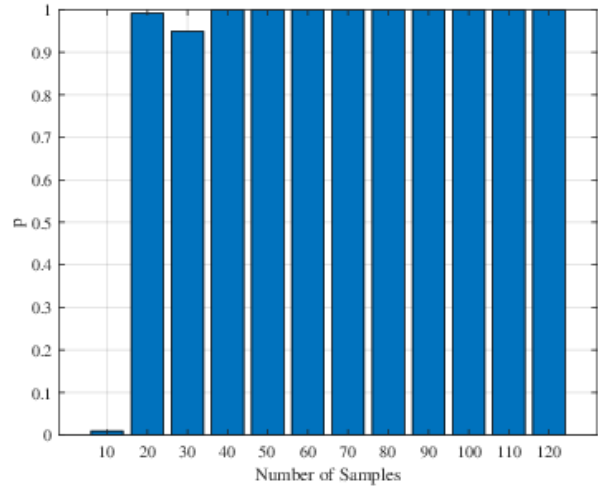


Figure 11: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 10$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 20$ or larger.

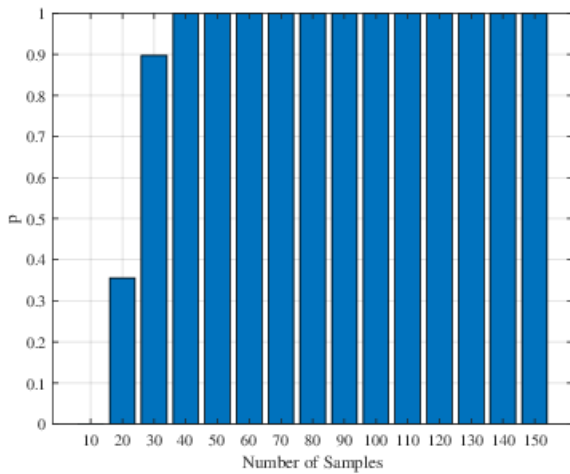


Figure 10: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 8$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 40$ or larger. Considering 10% significance p

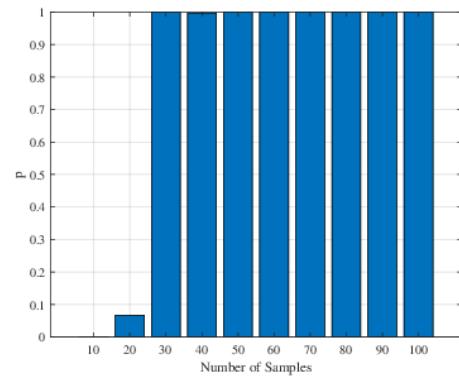


Figure 12: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 12$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

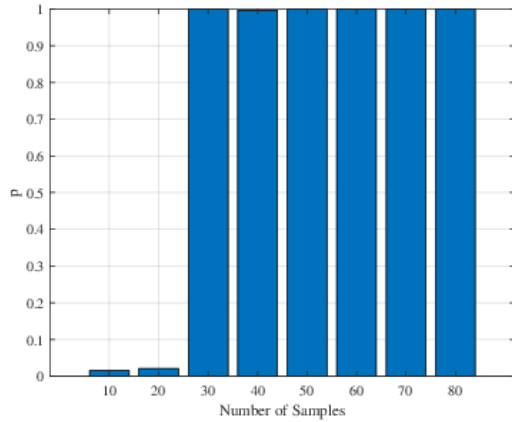


Figure 13: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 15$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

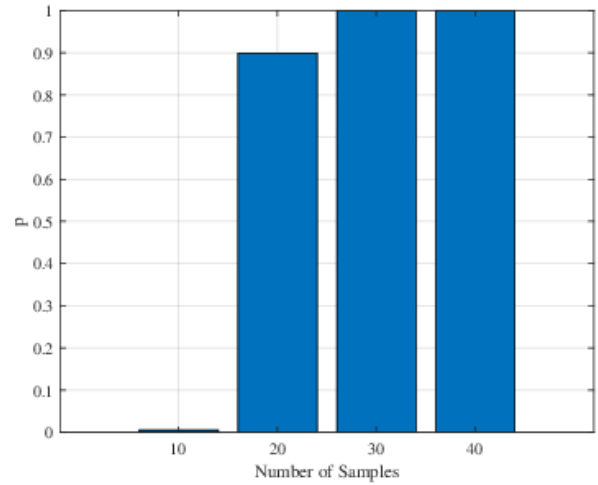


Figure 15: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 25$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger. Considering 10% significance p value reaches $p = 0.90$ when number of segments reach $N_{seg} = 20$.

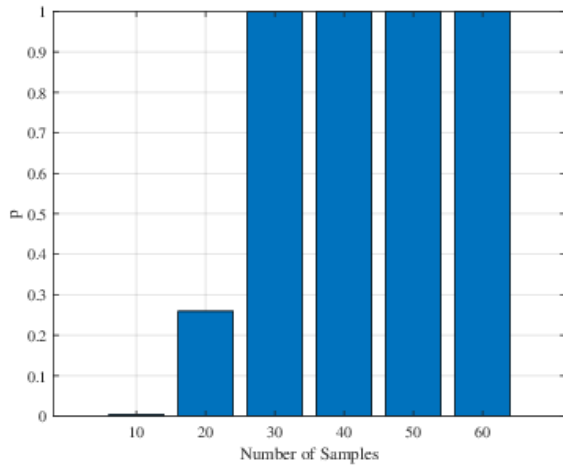


Figure 14: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 20$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

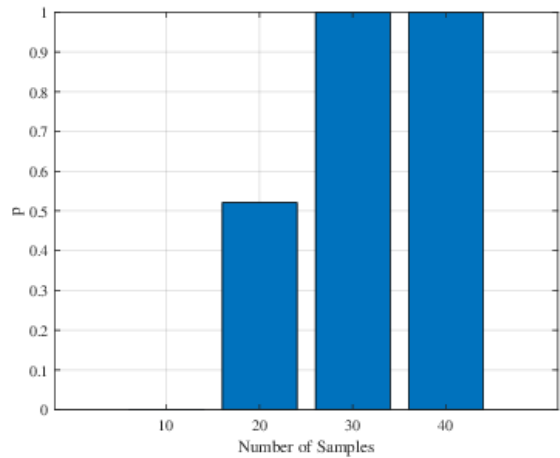


Figure 16: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 30$ seconds. The p value exceeds $p = 0.95$ level (5%

significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

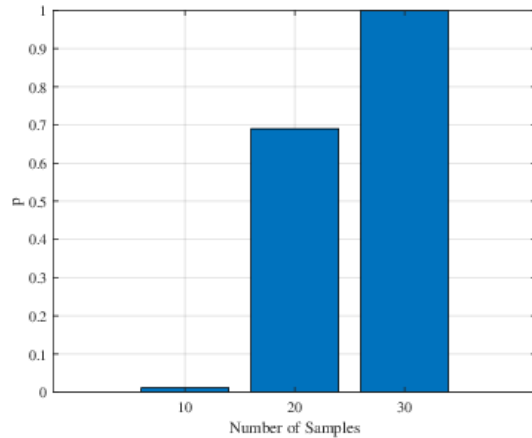


Figure 17: The variation of the p value for two sample Kolmogorov-Smirnov test when the segment size is $T_{seg} = 40$ seconds. The p value exceeds $p = 0.95$ level (5% significance level) when number of samples or segments reach $N_{seg} = 30$ or larger.

REFERENCES

[1] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J Physiol-London*, vol. 117, no. 4, p. 500, 1952. [Online]. Available: <https://doi.org/10.1113%2Fjphysiol.1952.sp004764>

[2] R. FitzHugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophys J*, vol. 1, no. 6, pp. 445–466, 1961.

[3] C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," *Biophys J*, vol. 35, no. 1, pp. 193–213, 1981.

[4] J. L. Hindmarsh and R. Rose, "A model of neuronal bursting using three coupled first order differential equations," *Proc R Soc Lond B*, vol. 221, no. 1222, pp. 87–102, 1984.

[5] V. Booth, J. Rinzel, and O. Kiehn, "Compartmental model of vertebrate motoneurons for Ca^{2+} -dependent spiking and plateau potentials under pharmacological treatment," *J Neurophysiol*, vol. 78, no. 6, pp. 3371–3385, 1997.

[6] R. O. Doruk, "Neuron modeling: estimating the parameters of a neuron model from neural spiking data," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, no. 5, pp. 2301–2314, 2018.

[7] V. Mante, R. A. Frazor, V. Bonin, W. S. Geisler, and M. Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nat Neurosci*, vol. 8, no. 12, p. 1690, 2005.

[8] T. Hosoya, S. A. Baccus, and M. Meister, "Dynamic predictive coding by the retina," *Nature*, vol. 436, no. 7047, p. 71, 2005.

[9] N. C. Rust, O. Schwartz, J. A. Movshon, and E. P. Simoncelli, "Spatiotemporal elements of macaque v1 receptive fields," *Neuron*, vol. 46, no. 6, pp. 945–956, 2005.

[10] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *JOSA A*, vol. 2, no. 2, pp. 284–299, 1985.

[11] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nat Neurosci*, vol. 2, no. 11, p. 947, 1999.

[12] H. Barlow, "Possible principles underlying the transformation of sensory messages. in. W. Rosenblith (ed.) *Sensory communication* (pp. 217–234)," 1959.

[13] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. R. d. R. van Steveninck, "Efficiency and ambiguity in an adaptive neural code," *Nature*, vol. 412, no. 6849, p. 787, 2001.

[14] B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der zeit-reihenfolgen-und vorzeichenbewertung bei der bewegungsperzeption des rüsselkäfers chlorophanus," *Z Naturforsch B*, vol. 11, no. 9–10, pp. 513–524, 1956.

[15] O. R. Doruk and K. Zhang, "Adaptive stimulus design for dynamic recurrent neural network models," *Frontiers in neural circuits*, vol. 12, p. 119, 2019.

[16] M. N. Shadlen and W. T. Newsome, "Noise, neural codes and cortical organization," *Curr Opin Neurol*, vol. 4, no. 4, pp. 569–579, 1994.

[17] I. J. Myung, "Tutorial on maximum likelihood estimation," *J Math Psychol*, vol. 47, no. 1, pp. 90–100, 2003.

[18] D. R. Brillinger, "Maximum likelihood analysis of spike trains of interacting nerve cells," *Biol Cybern*, vol. 59, no. 3, pp. 189–200, 1988.

[19] L. Paninski, "Maximum likelihood estimation of cascade point-process neural encoding models," *Network-Comp Neural*, vol. 15, no. 4, pp. 243–262, 2004.

[20] E. Chornoboy, L. Schramm, and A. Karr, "Maximum likelihood identification of neural point process systems," *Biol Cybern*, vol. 59, no. 4, pp. 265–275, 1988.

[21] A. C. Smith and E. N. Brown, "Estimating a state-space model from point process observations," *Neural Comput*, vol. 15, no. 5, pp. 965–991, 2003.

[22] A. V. Herz, T. Gollisch, C. K. Machens, and D. Jaeger, "Modeling single neuron dynamics and computations: a balance of detail and abstraction," *Science*, vol. 314, no. 5796, pp. 80–85, 2006.

[23] J. Ma and J. Tang, "A review for dynamics in neuron and neuronal network," *Nonlinear Dynamics*, vol. 89, no. 3, pp. 1569–1578, 2017.

[24] D. Linaro, M. Storace, and M. Giugliano, "Accurate and fast simulation of channel noise in conductance-based



model neurons by diffusion approximation,” *PLoS Comput Biol*, vol. 7, no. 3, p. e1001102, 2011.

[25] J. A. White, J. T. Rubinstein, and A. R. Kay, “Channel noise in neurons,” *Trends in neurosciences*, vol. 23, no. 3, pp. 131–137, 2000. Data available in: <https://github.com/phvu/theoretical-neuroscience/tree/master/exercises/c2/data..>

[26] M. Lv, C. Wang, G. Ren, J. Ma, and X. Song, “Model of electrical activity in a neuron under magnetic flow effect,” *Nonlinear Dynamics*, vol. 85, no. 3, pp. 1479–1490, 2016.

[27] M. Lv and J. Ma, “Multiple modes of electrical activities in a new neuron model under electromagnetic radiation,” *Neurocomputing*, vol. 205, pp. 375–381, 2016.

[28] F. Wu, C. Wang, Y. Xu, and J. Ma, “Model of electrical activity in cardiac tissue under electromagnetic induction,” *Scientific reports*, vol. 6, no. 1, pp. 1–12, 2016.

[29] C. DiMattina and K. Zhang, “Adaptive stimulus optimization for sensory systems neuroscience,” *Front Neural Circuit*, vol. 7, 2013.

[30] —, “Active data collection for efficient estimation and comparison of nonlinear neural models,” *Neural Comput*, vol. 23, no. 9, pp. 2242–2288, 2011.

[31] —, “How to modify a neural network gradually without changing its input-output functionality,” *Neural Comput*, vol. 22, no. 1, pp. 1–47, 2010.

[32] E. P. Lynch and C. J. Houghton, “Parameter estimation of neuron models using in-vitro and in-vivo electrophysiological data,” *Frontiers in neuroinformatics*, vol. 9, p. 10, 2015.

[33] R. O. Doruk and K. Zhang, “Fitting of dynamic recurrent neural network models to sensory stimulus-response data,” *J Biol Phys*, vol. 44, no. 3, pp. 449–469, jun 2018. [Online]. Available:

<https://doi.org/10.1007%2Fs10867-018-9501-z>

[34] R. O. Doruk, “Fitting a recurrent dynamical neural network to neural spiking data: tackling the sigmoidal gain function issues,” *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 27, no. 2, pp. 903–920, 2019.

[35] R. O. Doruk and L. Abosharb, “Estimating the parameters of fitzhugh–nagumo neurons from neural spiking data,” *Brain sciences*, vol. 9, no. 12, p. 364, 2019.

[36] M. A. Frye and M. H. Dickinson, “Fly flight: a model for the neural control of complex behavior,” *Neuron*, vol. 32, no. 3, pp. 385–388, 2001.

[37] G. Lewen, W. Bialek, and R. Steveninck, “Neural coding of naturalistic motion stimuli,” *Network: Computation in Neural Systems*, vol. 12, no. 3, pp. 317–329, 2001.

[38] K. D. Miller and F. Fumarola, “Mathematical equivalence of two common forms of firing rate models of neural networks,” *Neural Comput*, vol. 24, no. 1, pp. 25–31, 2012.

[39] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[40] G. Marsaglia, W. W. Tsang, J. Wang et al., “Evaluating kolmogorov’s distribution,” *Journal of statistical software*, vol. 8, no. 18, pp. 1–4, 2003.

[41] U. T. Eden, “Point process models for neural spike trains,” *Neural Signal Processing: Quantitative Analysis of Neural Activity*, pp. 45–51, 2008.

[42] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank, “The time-rescaling theorem and its application to neural spike train data analysis,” *Neural Comput*, vol. 14, no. 2, pp. 325–346, 2002.



RESEARCH ARTICLE

Heart Disease Prediction with Machine Learning Algorithm Using Feature Selection by Genetic Algorithm

¹ Berna Vatansver, ² Hakan Aydın, * Ali Çetinkaya

^{1,2} Istanbul Gelisim University, Computer Eng., Istanbul, Turkey
 bernavatansver1999@hotmail.com orcid.0000-0001-6441-6660
 haaydin@gelisim.edu.tr orcid.0000-0002-0122-8512

*Istanbul Gelisim University, TTO UAM R&D Center, Istanbul, Turkey
 alcetinkaya@gelisim.edu.tr orcid.0000-0003-4535-3953

HIGHLIGHTS

- Predicting heart disease by analyzing the factors that affect it.
- Utilizing Machine Learning (ML) algorithms by selecting dataset features for prediction processes
- Demonstrating that the prediction results obtained after feature selection using the Genetic Algorithm (GA) method and machine learning methods have a higher success rate.

Keywords:

- Machine Learning Algorithms
- Genetic Algorithms
- Heart Disease Prediction
- Feature Selection

Article Info:

Received : 07.10.2021
 Accepted : 06.11.2021
 Published : 21.12.2021

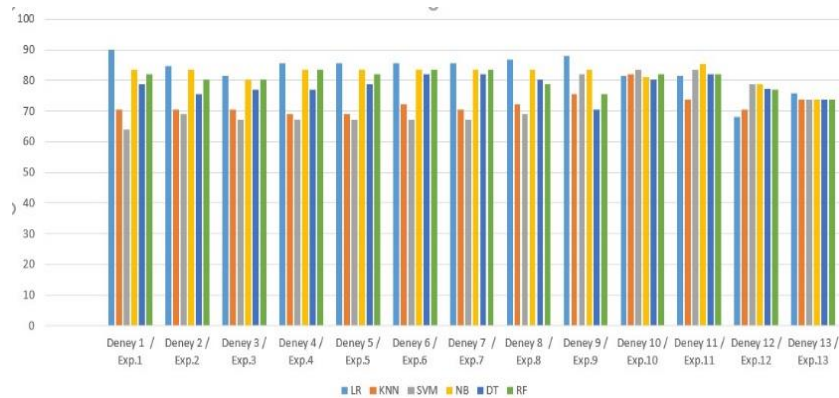
DOI:
 10.53525/jster.1005934

*Correspondence:

Ali Çetinkaya
alacetinkaya@gelisim.edu.tr
 +90 212 422 70 00 / 7187

GRAPHICAL ABSTRACT

The aim of this study was determined as the prediction of heart disease by analyzing the factors that affect this disease with Machine Learning (ML) algorithms using the feature selection with the Genetic Algorithm (GA) approach. In this context, K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB) and Support Vector Machine (SVM) algorithms were used in the study. The graph of the success rates obtained in the experiments conducted in the first group using these ML algorithms within the scope of the study is presented below.



Aim of Article : The purpose of this study is to predict heart disease with ML algorithms using feature selection with GA approach.

Theory and Methodology : In this study, ML methods and GA are used.

Findings and Results: In the study, the importance of feature selection with GA was emphasized by proving that the accuracy of heart disease prediction was higher after feature selection was made with GA in the data set.

Conclusion : As a result of the study, it has been shown that heart disease can be predicted with higher success rate and accuracy with ML algorithms using feature selection with the GA approach.



Genetik Algoritma Yaklaşımıyla Öznitelik Seçimi Kullanılarak Makine Öğrenmesi Algoritmaları ile Kalp Hastalığı Tahmini

¹ Berna Vatansver, ² Hakan Aydın, * Ali Çetinkaya

^{1,2} İstanbul Gelişim Üniversitesi, Bilgisayar Müh. Bölümü, İstanbul, Türkiye
bernavatansver1999@hotmail.com, orcid.0000-0001-6441-6660,
haaydin@gelisim.edu.tr orcid.0000-0002-0122-8512

*İstanbul Gelişim Üniversitesi, TTO UAM R&D Center, İstanbul, Türkiye
alcetinkaya@gelisim.edu.tr orcid.0000-0003-4535-3953

Alıntı / Citation :

Vatansver, B., Aydın, H., Çetinkaya, A. (2021). Genetik Algoritma Yaklaşımıyla Öznitelik Seçimi Kullanılarak Makine Öğrenmesi Algoritmaları ile Kalp Hastalığı Tahmini, *Journal of Scientific Technology and Engineering Research*, 2(2):67-80. DOI: 10.53525/jster.1005934

ÖNE ÇIKANLAR / HIGHLIGHTS

- Kalp hastalığını etkileyen faktör ve özelliklerin analiz edilerek bu hastalığın tahmin edilmesi
- Geleceğe dönük tahmin işlemleri için veri seti öznitelik seçimi yapılarak Makine Öğrenmesi (MÖ) algoritmalarından yararlanılması
- Genetik Algoritma (GA) metodu ve makine öğrenmesi yöntemleri kullanılarak öznitelik seçimi yapıldıktan sonra elde edilen tahmin sonuçlarının daha yüksek başarı oranına sahip olduğunun gösterilmesi

Makale Bilgileri

Geliş Tarihi : 07.10.2021
Kabul Tarihi : 06.11.2021
Yayın Tarihi : 21.12.2021

DOI:

10.53525/jster.1005934

*Sorumlu Yazar:

Ali Çetinkaya
alcetinkaya@gelisim.edu.tr
+90 212 422 70 00 / 7187

ÖZET / ABSTRACT

Günümüzde Yapay Zeka (YZ) uygulamalarından sağlık sektöründe pek çok hastalığın teşhis ve tedavisinde giderek artan oranlarda faydalanılmaktadır. Bu çalışmanın amacı Genetik Algoritma (GA) yaklaşımı ile öznitelik seçimi yaparak Makine Öğrenmesi (MÖ) algoritmalarıyla kalp hastalığını tahmin etmek olarak belirlenmiştir. Bu kapsamda çalışmada MÖ algoritmalarından K-En Yakın Komşu (K-EYK), Lojistik Regresyon (LR), Karar Ağacı (KA), Rastgele Orman (RO), Naive Bayes (NB) ve Destek Vektör Makinesi (DVM) algoritmaları ile 3 (üç) farklı grupta toplamda 28 (yirmi sekiz) deney gerçekleştirilmiştir. Çalışma sonucunda en yüksek başarı oranı GA ile yapılan deneylerde %93.44 ile elde edilmiştir. Çalışmamızda yapılan diğer deneylerde ise LR ile %90.16, K-EYK ile %81.96, DVM ile %83.60, NB ile %85.24, KA ile %81.96 ve RO ile %83.60 oranında başarı oranı elde edilmiştir. Çalışmada ABD Cleveland eyaletindeki bireylerin verisini içeren veri seti kullanılmıştır. Araştırmamızda elde edilen sonuçlarda kalp hastalığı tahmininin GA yaklaşımı ile öznitelik seçimi yapılması durumunda daha yüksek doğruluk oranının elde edildiği görülmüştür. Çalışmamızın GA ile öznitelik seçimi yapılarak MÖ ile kalp hastalığının tahmin edilmesinde sağlık alanındaki literatüre katkı sağlayacağı değerlendirilmektedir.

Anahtar Kelimeler / Keywords: Makine Öğrenmesi Algoritmaları, Genetik Algoritmalar, Kalp Hastalığı Tahmini, Öznitelik Seçimi

I. GİRİŞ [INTRODUCTION]

Günümüzde Yapay Zekâ (YZ) uygulamaları sağlık sektöründe pek çok hastalığın teşhis ve tedavisinde giderek artan oranlarda kullanılmaktadır. Son yıllarda sağlık sektöründe pek çok hastalığın teşhis ve tedavisinde YZ tabanlı sistemlerin kullanıldığı görülmektedir. YZ

teknik ve yöntemleri ile sağlık alanında birçok hastalığın tespiti, erken teşhisi ve tedavisi mümkündür. YZ yöntemlerinin sağlık alanında kullanımı hastalıkların başta hastalık tanılarının insan hataları olmak daha yüksek doğruluk ve oranda yapılabilmesi olmak üzere hastalıkların tanı ve teşhisinde pek çok imkân sunmaktadır. Kalp hastalıkları da bunlardan bir tanesidir.



Bu hastalık dünyanın dört bir tarafında yaşanan ölümlerin en büyük sebeplerinden birisi olarak öne çıkmaktadır. Kalp hastalığı, Dünya’da ve Türkiye’de ölüm sebebi olarak birinci sırada gösterilmektedir [1,2]. Kalp hastalıkları kadın ve erkeklerde dünyanın en yaygın hastalıkları arasındadır. Gelecekte bu hastalık nedeniyle ölüm sayısının daha da artacağı öngörülmektedir. Bu hastalık aynı zamanda pek çok başka hastalığın ortaya çıkmasına da sebep olmaktadır. Son yıllarda yapılan araştırmalar ani olarak hayatını kaybeden bireylerin yüzde 46,2’sinin kalp ve damar hastalıkları nedeniyle hayatını kaybettiklerini, bu nedenle bu hastalığın erken teşhisinin ve tedavisinin çok önemli olduğunu ortaya koymaktadır [3]. Bulanık Mantık, Doğal Dil İşleme, Konuşma Tanıma, Örüntü Tanıma, Yapay Sinir Ağları (YSA), Makine Öğrenmesi (MÖ), Genetik Algoritmaları (GA) YZ kapsamında sağlık sektöründe uygulama imkânı bulmaktadır. MÖ öğrenmesi ile örnek veri setlerine dayanan sınıflandırma veya tahmin işlemleri yapılabilmektedir. MÖ öğrenmesi ile tahmin işlemlerinde veri setinin başarısını etkileyen birçok öznelik bulunmaktadır. Veri setinde yer alan öznelikler tahmin başarı oranlarını olumlu veya olumsuz düzeyde etkileyebilmektedir. Bu nedenle veri setindeki tüm özneliklerin hepsini birebir alıp kullanmak yerine, veri setini daha belirgin olarak betimleyebilecek, veri setini en doğru temsil edebilecek nitelikli öznelik alanlarının belirlenmesi önem kazanmaktadır. Bu şekilde belirlenecek olan öznelik verileri sınıflandırma yöntemlerinin tahmin başarı oranlarını artırabilecektir. GA yöntemi de bu bağlamda özneliklerin belirlenmesinde kullanılan yöntemlerden bir tanesidir. GA ile gerçekleştirilen uygulamalarda analiz yapılmasında yüksek doğruluk oranına sahip başarılı sonuçlar elde edilebilmektedir. [4].

MÖ çeşitli görevlerin öğrenilmesi, mantıksal ve ikili çıkarımlar yoluyla otomatik hesaplama yöntemlerini kapsayan bir süreç olarak kullanılmaktadır. MÖ ile pek çok hastalığın tanısı yüksek doğruluk ile yapılabilmektedir. Literatürde hastalık tanı tahminlerinin en yüksek doğruluk oranı ile alınabilmesi ile MÖ kullanılarak çeşitli çalışmalar yapıldığı görülmektedir. En yüksek doğruluk oranı ile YZ teknikleriyle hastalıkların tahmin edilmesinin hastalıklarının tedavisine büyük katkılar sağlayacağı bir gerçektir. Bu çalışmanın amacı GA yaklaşımıyla öznelik seçimi kullanılarak MÖ algoritmaları ile kalp hastalığını tahmin etmek olarak belirlenmiştir. Bu bağlamda çalışmada YZ’nin bir alt dalı olan MÖ öğrenmesi ile daha yüksek doğruluk oranına sahip tahmin sonuçları elde edilmesinde GA ile öznelik

seçiminin önemini belirlenmesi hedeflenmiştir. Bu bağlamda çalışmada öznelik seçimi için GA’dan faydalanılmış ve bu yöntemin başarı oranına etkisi ortaya konmaya çalışılmıştır. Çalışmada kullanılan veri seti üzerinde farklı deneyler yapılmış ve bu deneyler neticesinde elde edilen başarı oranları değerlendirilmiştir. Deneylerde veri setinde öznelik seçim yöntemlerinden YZ tabanlı GA kullanılarak daha belirgin öznelikler seçilmiş ve MÖ sınıflandırma algoritmaları ile kalp hastalığının tahmini gerçekleştirilmiştir.

Bu makale şu şekilde düzenlenmiştir: Çalışmanın ikinci bölümünde sağlık alanında hastalıkların tanımlanmasında literatürde yer alan çalışmalar incelenmiştir. Üçüncü bölümünde yapılan deneyler anlatılmış ve elde edilen sonuçlar belirtilmiştir. Son bölümde ise çalışma neticesinde ulaşılan sonuçlar değerlendirilmiştir.

II. İLGİLİ ÇALIŞMALAR [RELATED STUDIES]

Sağlık alanında YZ yönteminin kullanılması ile ilgili çalışmalar, son yıllarda giderek artan bir hızla yaygınlaşmaya başlamıştır. Sağlık alanına yönelik olarak MÖ olmak üzere YZ teknik ve yöntemleri kullanılarak pek çok çalışmanın bulunduğu, ayrıca bu çalışmaların giderek arttığı görülmektedir. Biyomedikal veri kümesi üzerinde YSA algoritması dahil edilerek yapılan çalışmada elde edilen başarı oranları karşılaştırıldığında en yüksek doğruluk oranının K-EYK algoritması ile olduğu görülmüştür [5]. Karaciğer hastalığı üzerine denetimli makine öğrenimi algoritmalarını ile tahmin elde eden bir çalışmada LR, RO, KA, DVM, K-EYK ve NB için %75, %74, %69, %64, %62 ve %53 doğruluk değeri bulunmuştur [6]. MÖ ve Veri Madenciliği (VM) teknikleriyle öğretmenlerin psikososyal risk düzeylerinin tahmin edilmesi amacıyla yapılan çalışmada kullanılan veri seti üzerinde Elektrokardiyografi (EKG), Elektromiyografi (EMG) ve Elektroensefalografi (EEG) gibi fizyolojik faktörler MATLAB ortamında test edilmiş ve MÖ algoritmalarından KA, K-EYK, DVM karşılaştırılmış ve %99,9 doğruluk oranına ulaşılmıştır [7]. MÖ algoritmaları kullanılarak klinik tabanlı veriler üzerinden gerçekleştirilen çalışmada radyoloji verileri ile bütünlük sağlanarak kanser hastaları için kişisel tedavi aracı geliştirilmesi amaçlanmıştır [8]. Kalp hastalığı tahminin tespiti amacıyla yapılan diğer bir çalışmada ise RO yöntemi ve DVM kullanılmış ve çalışmalar neticesinde %87,87 tahmin doğruluk oranına ulaşılmıştır [9]. Sağlık hizmetleri alanında yorumlanabilir makine öğrenmesi modellerinin kullanımı üzerine yapılan çalışmada ihtiyaç duyulan MÖ algoritmasının sisteme nasıl ekleneceği araştırılmıştır [10]. Diyabet teşhisi



konmasında hastalığın tahmin edilebilmesine yönelik olarak YZ teknikleri kullanılarak yapılan çalışmada optimizasyon algoritmaları kullanılmış ve optimizasyon algoritmalarının kullanımının tahmin sonucu başarı oranını artırabileceği gösterilmiştir [11]. Araç algılama sistemi geliştiren bir çalışmada YZ temelli çalışan bir video algılayıcı sistem kullanılarak ortalama hız, araç sınıflandırma vb. araç takibi bilgileri üretilir [12].

Öznetelik seçimiyle kanser sınıflandırması yapılan bir çalışmada sinyal görüntü oranı, K-EYK, DVM yararlanılmıştır. K-EYK ile %99.3 doğruluk oranı elde edilmiştir [13]. DVM algoritması ile tahmin işlemleri amacıyla yapılan çalışmada Veri Madenciliği (VM) ve R programlama dili kullanılmıştır [14]. Literatürde içinde bulunduğumuz Pandemi dönemine ilişkin olarak da çalışmalar yapıldığı görülmektedir. Yüksek risk altındaki hastaları, ölüm oranlarını ve diğer anormallikleri belirlemek için yapılan çalışmada MÖ öğrenmesi ile risk faktörlerini yaş, sosyal alışkanlıklar, konum ve iklim faktörlerinin hastalığa olan etkisi araştırılmıştır [15]. Ortaöğretim 6, 7 ve 8. sınıf öğrencilerine, belirli ölçütlere göre 24 sorudan oluşan anket uygulanarak yapılan bir diğer çalışmada anket sonuçları dikkate alınarak, MÖ öğrenmesiyle Türkçe, Matematik dersleri ve dönem sonu genel başarı puan ortalamalarının regresyon / çok sınıflı MÖ modelleri oluşturulmuştur [16]. Cilt kanserinin YZ teknik ve yöntemleri ile tahmin edilmesi konusunda yapılan çalışmada %94 ile doğruluk oranına ulaşılmıştır [17]. Kalp hastalıklarının tahmin edilmesi amacıyla yapılan bir diğer çalışmada VM sınıflandırma teknikleri kullanılmış ve çalışma neticesinde kalp hastalıkları tahminlerinde risk faktörlerinin tahmin oranlarının yüksek veya düşük olarak belirlenmesine etki ettiği ortaya konmuştur [18]. Kalp ve meme kanseri hastaları için tahmini için NB, KA, Simple Logistic, J48 algoritmalarının kullanıldığı çalışmada çapraz doğrulama yöntemleriyle algoritmaların performansları karşılaştırılmış ve Naive Bayes'in en başarılı sonuç verdiği anlatılmıştır [19]. DVM, NB, J48, RO, Adaboost, LR, KA, Tek ve Çok Katmanlı Perceptron ile yapılan çalışmada kalp hastalığı tahmini yapılması amaçlanmıştır [20]. MÖ öğrenmesi yöntemleriyle kalp hastalıklarının tahmin edilmesi için Anfis (Adaptive Neuro-Fuzzy Inference System) ve Bulanık K Ortalamalar yöntemleri kullanılmış ve K kat çaprazlama yöntemi ile %98,35 başarı oranı elde edilmiştir [21]. VM, YSA ve GA kullanılarak Matlab ortamında kalp hastalığı riskinin tahmini amacıyla yapılan çalışmada %89 başarı oranı elde edilmiştir [22]. Diğer bir çalışmada Koroner Arter Hastalığı riski tespitinde bir sınıflama modeli %83,498 başarıyla Çok katmanlı YSA sınıflandırma yöntemi ile geliştirilmiştir [23]. Kalp hastalarının yaşam kalitesini artırmak

hastaneye yatışını azaltmak ve iyileştirmek için kanıta dayalı tedavi kapsamında yapılan çalışmada bilimsel teknikler ve ölçümler kullanılmıştır [24]. "Kardiyovasküler Hastalık Risk Faktörleri Bilgi Düzeyi (KARRİF-BD)" ölçeği ile verilerin toplandığı çalışmada ise t testi, Kruskal Wallis Varyans Analizi, Mann Whitney U testi ve Spearman korelasyon analizi kullanılmış, çalışma neticesinde Göğüs ağrısı tipi, egzersize bağlı anjin ve egzersize bağlı ST depresyonu nitelikleri yüksek oranda birbiriyle ilişkili olduğu gösterilmiştir [25]. Kalp hızı ve kalp hızı değişkenliğinin kardiyovasküler risk faktörleri ile ilişkisinin araştırıldığı çalışmada başta kalp hızı ölçümü olmak üzere farklı değerler kullanılmıştır [26]. Yaşa bağlı olarak yapısal değişimlerin gen ifade düzeyleriyle olan ilişkisinin araştırıldığı çalışmada da yaş arttıkça gen ifade düzeylerindeki azalmanın görüldüğü ortaya onmuştur [27]. Obezite ile birçok hastalığın meydana geldiği düşüncesinden hareketle yapılan bir diğer çalışmada obezitenin kalp hastalığı oluşumunda etken faktör olduğu gösterilmiştir [28]. Amerikan Kalp Derneği (AHA) tarafından yapılan çalışma kalp hastalığı, felç ve kardiyovasküler risk faktörleri ile ilgili en güncel istatistiklere katkı sağlamak amacıyla yapıldığı görülmektedir [29]. Temel bileşen analizi ile hipertansiyon hastalığının tahmininin amaçlandığı çalışmada Katmanlı Algılayıcı Ağ (ÇKA), Karar Tablosu ve C4.5 sınıflandırma algoritmaları kullanılmış ve çalışma neticesinde en başarılı sonucun %88 ile NB algoritması ile alındığı görülmektedir [30]. Kalp hastalarının sosyo-demografik hastalığa ilişkin özelliklerin yaşam kalitesi ve sağlıklı yaşam biçimi davranışlarına etkisinin araştırıldığı çalışmada; erkek, 49-60 yaş arası, evli, lisans ve üstü mezunu, sosyo-ekonomik durumu iyi, çalışan, ek hastalığı olmayan bireylerin hastalıklarıyla ilgili eğitim alanlarının risk faktörü araştırılmıştır [31]. Kardiyovasküler hastalıkların yaşla ilişkisinin araştırıldığı çalışmada Kardiyovasküler hastalıklardan ölümlerin ve majör komplikasyonların %78'i, 75 yaş üzerindeki hastalarda görüldüğü çalışma sonucu olarak belirtilmiştir [32]. NB, Destek Vektör Makineleri, Lineer Diskriminant Analiz Algoritması, Yapay Sinir Ağları, C4.5, C5.0, K-Nearest Neighbor algoritmaları ile yapılan çalışmada bu algoritmaların başarı oranları ortaya konmuştur [33]. Diğer bir çalışmada MÖ algoritmalarının sağlık alanında hastalık biyolojisi, ilaç keşfi, mikrobiyom araştırmaları ve sentetik biyoloji gibi birçok alanda kullanılabileceği araştırılmıştır [34]. Benzer şekilde başka bir çalışmada aynı araştırmanın ZeroR, OneR, NB, J48 Karar Ağacı, RO, Çoklu Algılayıcılar, K-EYK, LR, DVM gibi sınıflandırma algoritmaları ile WEKA programında yapıldığı görülmektedir [35]. MÖ yöntemleriyle telekomünikasyon sektöründeki müşteri kaybı tahmini üzerine çalışmada Yeniden Örnekleme, Sentetik Azınlık

Aşırı Örneklemeye Yöntemi (SMOTE), Uyarlanabilir Sentetik Örneklemeye (ADASYN), Rastgele Aşırı Örneklemeye (ROSE) yöntemleri uygulanmış, sınıflandırma problemi için NB, KA, RO, Yapay Sinir Ağları, LR, DVM ve K-EYK yöntemleri 10 kat Çapraz Geçerlemeyle araştırılmıştır [36]. Futbol takımların hücumu dayalı ve savunmaya dayalı yetenekleri dikkate alınarak maçın sonucunu ne kadar başarılı tahmin edebildiği konusunda yapılan çalışmada 16 futbol liginde yapılan testler sonucu Karar Tablosu çoğunlukla en başarılı algoritma olduğu anlatılmaktadır [37]. Fiber optik dağıtılmış akustik algılama yapan bir çalışmada sınıflandırma bölümünde iki farklı derin öğrenme algoritmasından yararlanılmıştır [38]. MÖ ve VM tahmin ve modelleme ile ilişkisinin araştırıldığı çalışma, tahmine dayalı modelleme, finansal hizmetler, sağlık hizmetleri, telekomünikasyon vb. alanları üzerine yapılmış olan bir çalışmadır [39]. Koroner arter hastalığı teşhisi için MÖ ve VM teknikleri konusunda yapılan çalışmada, çalışma kapsamında geliştirilen veri tabanı sunulmaktadır [40].

Biyomedikal veri kümeleri ile MÖ algoritmaları üzerine yapılan bir çalışmada YSA, K-EYK ve Lineer Diskriminant Analiz (LDA) algoritmaları ile EEG veri kümesi üzerinde sınıflandırma işlemleri sonucunda epileptik veriler başarı ile sınıflandırılmıştır [41]. MÖ yöntemleriyle kredi kartı dolandırıcılığına karşı yapılan bir çalışmada elde edilen en yüksek başarı sonucu Naive Bayes algoritmasında %97.92 olarak bulunmuştur [42]. Kredi kartı dolandırıcılığını MÖ yöntemleri kullanılarak tahmin eden bir başka çalışmada ise en yüksek başarı oranı çok katmanlı yapay sinir ağları ile %99,943'tür. Naive Bayes algoritması ile %98,207 başarı sonucu elde edilmiştir [43]. Değişken önem değerlendirmesine dayalı MÖ algoritmalarını karşılaştırmalı analizini yapan bir çalışmada, Random Forest, Cubist ve MARS, hem tahmin performansı hem de değişken önem değerlendirmesi için etkilidir sonucu çıkarılmıştır [44].

Çok katmanlı algılayıcı (ÇKA) sinir ağı modeli ile DVM yöntemleri kullanılarak yapılan çalışmada buzlanmadan kaynaklı kazaların önlenmesine yönelik %87.26 doğruluk oranına sahip bir buzlanma tahmin algoritması ve mobil uygulama geliştirilmiştir [45]. EEG sinyallerinden epileptik nöbet sınıflandırılmasının yapıldığı çalışmada DVM kullanıldığı görülmektedir [46]. Bölgelere göre itfaiye istasyonu ihtiyacının MÖ kullanılarak tahmin sınıflandırılmasında en başarılı algoritmanın %93,84 ile RO algoritması olduğu yer almaktadır [47]. Sağlık hizmeti üzerine veri madenciliği ile yapılan çalışmada NB, KA ve DVM algoritmaları karşılaştırılmıştır [48]. Kayısı çekirdeklerinin tatlı veya acı

olarak sınıflandırılması üzerine yapılan çalışmada K-EYK, DVM, KA, Rasgele Orman, NB ve Çok Katmanlı Algılayıcı algoritmaları kullanılmıştır [49]. Sosyal medya paylaşımları ile markaların itibarını analiz etmesini ve tüketiciler ile ilişkisini yöneten bilgi sistemi geliştirilmesinin hedeflendiği çalışmada MÖ yöntemleriyle sosyal medya marka itibar analizi alanında yazılım yapılmış ve çalışma neticesinde modeller arasından %90 F-ölçütü başarı skoruna sahip NB seçilmiştir [50]. Diğer bir çalışmada ise veri kümelerinin MÖ ile analiz edilmesi araştırılmıştır [51]. Duygu analizi konusunda yapılan çalışmada kümeleme, sınıflandırma, YSA ve VM teknikleri kullanılarak araştırma yapılmıştır [52]. Doğal dil işleme ve MÖ kullanılarak Twitter üzerinden iletilen mesajların hastalık bilgisinin içerme durumuna göre hastalık yayılım hızı analizi gerçekleştirilmiştir [53].

Geliştirilen MÖ uygulamasında yapılan deneylerin üzerinde GA ile öznitelik seçimi yapılarak başarı oranının arttığı belirtilmiştir [54]. Genetik algoritma kullanarak MRI tabanlı beyin tümörü Sınıflandırması yapan bir çalışmada %92,03 doğruluk elde edilmiştir [55]. GA ile çözüme ulaşılabilecek evrim süreci bilgisayar ortamında taklit edilmektedir. Karmaşık çok boyutlu arama uzayında bütünsel en iyi çözümü aramaktadır [56]. GA, çok sayıda sınırlama içeren ve karmaşık optimizasyon sorunlarını yazılım yöntemleriyle çözerler. GA ve sınıflandırıcı yöntemler kullanarak kanseri tahmin eden bir çalışmada en başarılı sonuç akciğer kanserine ait verilerde %94.09 olarak DVM algoritmasında elde edilmiştir [57]. Amacı tek bir çözüm değil çözümler kümesi ortaya çıkarmaktır. Hatta çözümsüz problemlere çözüm üretebilmektir. Bir problemin çözümü birden fazla olabilir. Bu çözüm kümesine nüfus adı verilir. Bireye ait her bir elemana gen denir. Algoritmada yapılan işlemler genetik operatör olarak adlandırılır. Genetik Operatörler seçme (selection), çaprazlama (crossover) ve mutasyon (mutation)'dur. Mutasyon sayesinde genetik çeşitlilik artar [58]. Bu sayede daha çok çeşit ortaya çıkar. Elde edilen bu nüfus ileriki nesillerde kullanılır. GA'da Mutasyon ile, kayıp veya keşfedilmemiş genetik materyaller bulunur ve erken yakınsamayı önler, yerel optimum seviyede kalır [59]. Her nesil için bu adımlar tekrarlanır.

Çalışmamızda elde edilen başarı oranı ile literatürde yer alan benzer çalışmaların başarı oranları Tablo 1'de sunulmuştur. Çalışmamızın sonuçlarının araştırılan diğer çalışmalarla benzer sonuçlar içerdiği ve çalışmalarda belirtilen referans aralığında başarı sonucu gösterdiği görülmektedir.



Tablo 1. Literatür taramasının değerlendirilmesi

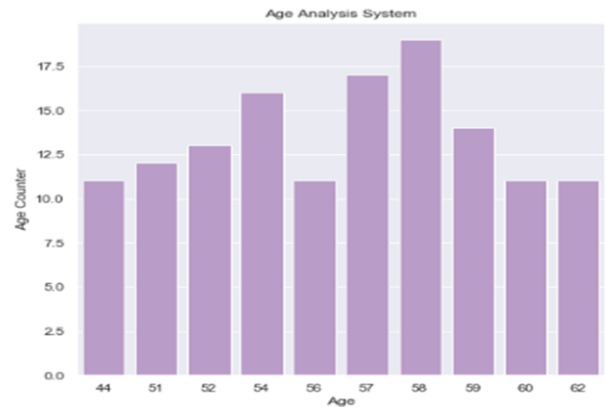
Literatür Adı	Kullanılan Algoritma Tipi	Veri Seti	Başarı Sonucu	Hata Sonucu
Polatgil, 2020	Anfis ve Bulanık K Ortalamalar	Kalp	%98.35	%1.65
Ekrem, 2020	Rastgele Orman	Kalp	%87.87	%12.13
Ünver, 2020	Naive Bayes	Hipertansiyon	%88	%12
Kabaoğlu, 2021	ÇKA	2017 yılı kış ayının hava tahmin verileri	%87.26	%12.74
Selimoğlu, 2021	Çok katmanlı yapay sinir ağları	Kredi Kartı Dolandırıcılık Teşhis	%99.943	%0.057
Mishra, 2011	K-EYK Algoritması	Lösemi	%99.3	%0.7
Rahman, 2019	Logistic Regresyon	Hint karaciğer hastaları	%75	%25
Çalışmamız	Genetik Algoritma	Kalp	%93.44	%6.56

III. TEORİ VE YÖNTEM [THEORY AND METHODOLOGY]

A. Veri Seti (Dataset)

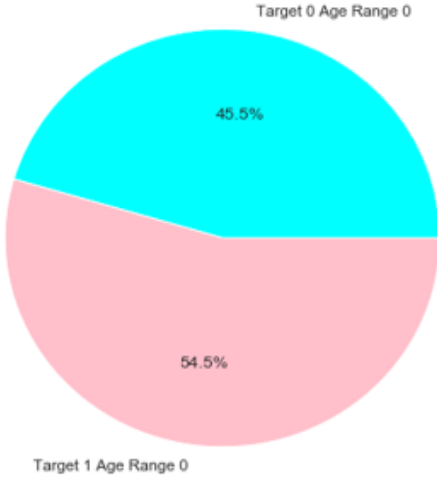
Bu çalışmada Kaggle'dan elde edilen kalp veri seti kullanılmıştır. Bu veri seti ABD'nin Cleveland eyaletindeki bireylerin verisidir. Kalp veri seti; Andras JANOSI, William STEINBRUNN, Matthias PFISTER ve Robert DETRANO tarafından oluşturulmuştur [60]. Kullanılan veri setinde yaralanma öznelik bilgileri 14 sütun ve 303 satırdan oluşmaktadır. Veri setinde hasta bireyler 1 rakamı, sağlıklı bireyler ise 0 rakamı ile ifade edilmiştir. Sütunlarda yer alan tüm veriler sayısal olarak gösterilmiştir. Veri setinde yer alan 14 farklı öznelik değerleri şu şekildedir: Yaş, cinsiyet (0=kadın,1=erkek), göğüs ağrısı tipi (0=tipik angina, 1=atipik angina, 2=anginal olmayan ağrı, 3=asemptomatik angina), dinlenme kan basıncı (tansiyon), kolesterol, açlık kan şekeri (120'den büyükse 1,değilse 0), dinlenme EKG ölçümü (0=normal,1=ST-T dalga anormalliği, 2=Sol ventrikül hipertrofisi), maksimum kalp atış hızı, egzersiz kaynaklı ağrı (1=evet,0=hayır), Egzersizin neden olduğu ST depresyonu, ST segmentinin eğimi (1=yukarı eğimli,2=düz,3=aşağı eğimli), ana damar sayısı(0-3), Talesimi (0=talasemi hastalığı yok, 1=sabit kusur, 2=normal kan akışı, 3=geri dönüştürülebilir kusur). Veri setinde boş değer (null) bulunmamaktadır. Hedef sütunu y, diğer tüm sütunlar x olarak ayrılmıştır. Hold-out yöntemiyle veriler %20 test ve %80 eğitim (train) verisi olarak ayrılmıştır. Bu kapsamda çalışmada kullanılan toplamda 303 adet veriden 242 adedi eğitim verisi ve 61 adedi ise test verisi olarak kullanılmıştır. Verilerin kalitesini artırmanın yolu verileri ön işlemden geçirmek suretiyle veri kalitesinin artırılması yoluyla

gerçekleşebilir. Bu maksatla veri setimiz üzerinde veri ön işleme adımları gerçekleştirilmiştir. Veri setimizdeki verilerin tamamlanmamış, tutarsız ya da güncelliğini kaybetmiş olup olmadığı kontrol edilmiş ve kaliteli veri için veri ön işleme adımları uygulanmıştır. Bu bağlamda null değere sahip veriler sonuca dahil edilmeden önce veri setinden çıkarılmıştır. Veri ön işleme aşamasında eksik, gürültülü ve tutarsız olan veriler iyileştirilmiştir. Veri setimizde yer alan bireylerin yaşlarını gösteren sütun grafik Şekil 1'de gösterilmektedir. Bu grafiğe göre, örneğin 58 yaşında olan birey sayısı 19 kişidir.



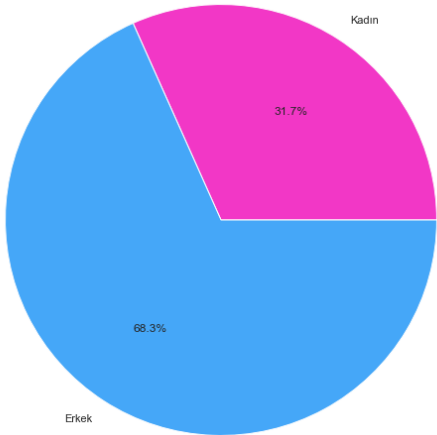
Şekil 1. Yaşa göre sütun grafiği

Şekil 2'de genç yaşta bulunan bireylerin hasta ve sağlıklı olma durum oranları sunulmuştur. Söz konusu şekilde hasta bireyler 1 rakamı ile ve sağlıklı bireylerde 0 rakamı ile gösterilmiştir.



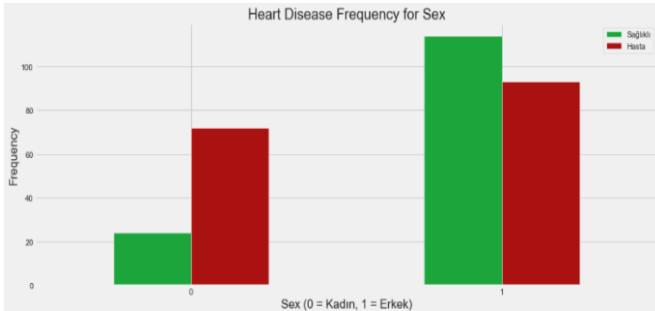
Şekil 2. Gençlerin kalp hastası/sağlıklı oranı

Şekil 3'te ise veri setinde bulunan bireylerin cinsiyet dağılımına göre pasta grafiği oluşturulmuştur. Grafiğe göre veri setindeki bireylerin %31.7' si kadın, %68.3' ü erkeklerden oluşmaktadır.



Şekil 3. Cinsiyet dağılımına göre pasta grafiği

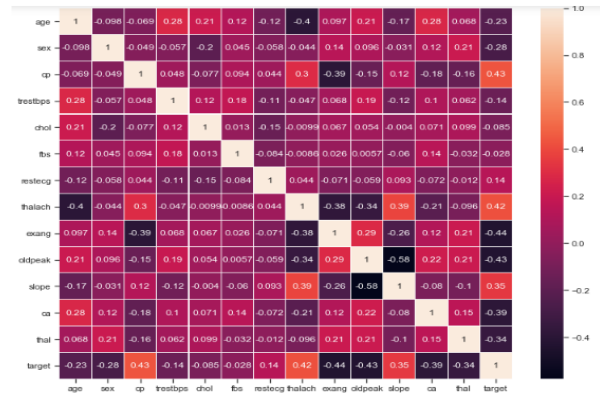
Şekil 4'te kalp hastalığının cinsiyete bağlı görülme sıklığı sunulmuştur. Bu grafiğe göre cinsiyeti kadın olan 23 sağlıklı ve 70 hasta birey yer alırken cinsiyeti erkek olan toplamda 100 sağlıklı ve 90 hasta birey yer almaktadır.



Şekil 4. Kalp hastası-cinsiyet dağılımı

Şekil 5'de veri setindeki özneliklerin korelasyon matrisi sunulmuştur. Bu matriste kullanılan korelasyon (r) kavramı ile özelliklerin birbiriyle ilişkisini gösterilmiştir. Bu bağlamda -1'e olan yakınlık değişkenler arasındaki negatif ilişkiyi, +1'e olan yakınlık ise pozitif ilişkiyi ve 0'a yakınlık da birbiriyle ilişkisi olmadığını göstermektedir. Bu duruma ilişkin durum Denklem (1)'de formülize edilmiştir.

$$-1 \leq r \leq +1 \quad (1)$$



Şekil 5. Korelasyon matrisi

Şekil 5 incelendiğinde bir bireyin kalp hastası olmasına göğüs ağrısı tipi, egzersize kaynaklı ağrı ve egzersiz neden olduğu, ayrıca ST depresyonu arasında pozitif ilişki bulunduğu görülmektedir. Bu durum kalp hastalığı tanısının konmasında etkindir. Ayrıca cinsiyet ve yaş öznelikleri de kalp hastalığı tanısında etkilidir. Cinsiyeti erkek olan bireylerin ve yaşlıların kalp hastası olma ihtimalinin daha yüksek olduğu görülmektedir.

B. Makine Öğrenmesi (MÖ) Algoritmaları

MÖ algoritma yöntemleri denetimli öğrenme, yarı denetimli öğrenme ve denimsiz öğrenme olmak üzere üç başlık altında gösterilebilir. K-EYK, NB, KA, RO, LR ve DVM algoritmaları bu kapsamda en çok kullanılan algoritmalar arasında sayılabilir.

LR, çeşitli tahmin değişkenleri ile ikili sonuç arasındaki ilişkiyi değerlendiren istatistiksel bir tekniktir [61, 62]. LR analizinin sağlık hizmetleri alanında sıkça kullanıldığı görülmektedir [63]. Bu algoritma bir sınıflandırma algoritmasıdır. Bu algoritmanın formülü Denklem (2)'de sunulmuştur.

$$\text{logic}(p) = \ln\left(\frac{p}{1-p}\right) \quad p = \text{probability} \quad (2)$$

K-EYK algoritması da sık kullanılan sınıflandırma algoritmaları arasındadır. Bu algoritma parametrik bir algoritma değildir. Tahmin işlemlerinde veri seti içinde belirlenen K değeri hesaplanacak en yakın komşuların eleman sayısını ifade eder. Örneğin; K değeri 5 ise en yakın 5 komşunun elde edilen değer ile arasındaki uzaklık hesaplanır. Uzaklık hesaplar Kenan Denklem (3) Öklid, Denklem (4) Manhattan ve Denklem (5) Minkowski standart mesafe ölçümü denklemlerinden yararlanır.

$$\text{Öklid} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (3)$$

$$\text{Manhattan} = \sum_{i=1}^k |x_i - y_i| \quad (4)$$

$$\text{Minkowski} = \left(\sum_{i=1}^k (|x_i - y_i|^q) \right)^{1/q} \quad (5)$$

MÖ yöntemlerinden DVM algoritması sınıflandırma ve regresyon için verilerin analizinde kullanılan denetimli öğrenme modelidir. Bu algortmada amaç, iki sınıf arasındaki uzaklığın maksimum olduğu durumun bulunmasıdır. Vektör uzayı tabanlı olan bu makine öğrenmesi algoritması parametre almaz. DVM algoritması Denklem (6)'da formül olarak gösterilmiştir.

$$\{x_i, y_i\} | i = 1 \dots L, y_i \in \{-1, 1\}, x \in R^D \quad (6)$$

NB algoritması, sınıflandırma algoritmalarındandır. Bu algoritmanın çalışma mantığı her bir eleman için tek tek her durumun olasılığını hesaplayarak en yüksek olasılık değerini temel alarak sınıflandırma yapmaktadır. NB algoritması Denklem (7)'de göre formül olarak gösterilmiştir.

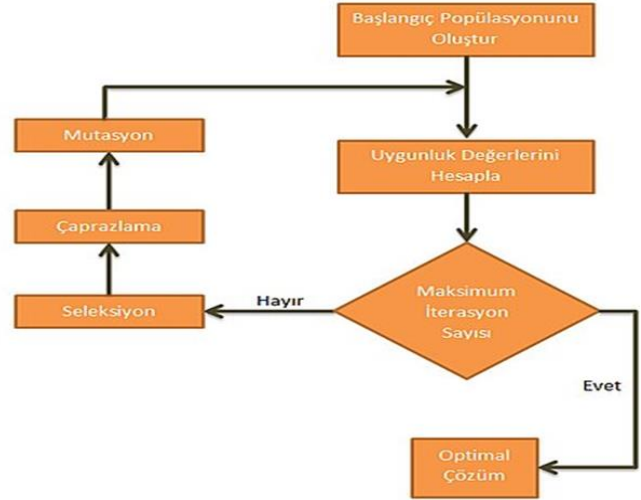
$$P\left(\frac{\text{class}}{\text{data}}\right) = \frac{p(\text{data/class}) \times p(\text{class})}{p(\text{data})} \quad (7)$$

KA algoritması, veri kümesindeki çok büyük verileri belirli kurallar ile daha küçük kümelerle ayıran sınıflandırma algoritmasıdır. RO algoritması denetimli sınıflandırma algoritmalarındandır. Büyük verileri kümeleyerek küçük ağaçlar oluşturur. Bu algortmada ağaç sayısı artarsa kesinlik artar. Bu algoritmanın KA algoritmasından farkı kök düğüm bulma ve düğümlere bölme işleminin rastgele çalışmasıdır.

C. Genetik Algoritmalar (GA)

Teknolojinin gelişmiş algoritmalarından biri olan GA çalışma mantığı insan genlerinin nesiller arası geçiş sürecine benzetilmektedir. GA, doğal genetik ve biyolojik evrimin mekaniklerine dayanan fonksiyon optimizasyonu için kullanılan stokastik bir yöntemdir. Genetik

algoritmanın çalışma prensibi Şekil 6'da gösterilmiştir [64].



Şekil 6. Genetik algoritma akış diyagramı

GA çalışma adımlarında ilk olarak “nüfus” bilgisi bulunmaktadır. Bu bilgi ile verilere uygun popülasyon rastgele olarak başlatılır. İkinci aşama “uygunluk” değeridir. Sonuca ulaştıran her bir çözümün (kromozom) uygunluk değeri bulunur. Üçüncü aşamada gelecek nesil ile gen geçişi arasında yeni bir popülasyon olması için ebeveyn olarak en uygun kromozomlar seçilir. Dördüncü aşamada çaprazlama işlemi yapılmaktadır. Ebeveynler birleştirilip yeni kromozom seti oluşturulur. Daha sonra yeni oluşan popülasyona eklenir. Beşinci aşamada oluşan popülasyon setindeki kromozomun gen değişimleri yapılır. Bu işleme Mutasyon denilmektedir.

D. Sistemin Akış Diyagramı

Çalışmada Şekil 7 üzerinde belirtilen akış diyagramı kullanılmıştır. Akış diyagramına uygun olarak öncelikle Jupyter notebook yazılım programı başlatılmaktadır. Tahmini ve analizi yapılacak olan veri seti çalışmaya eklenmiştir. Kullanılacak algoritmalar, istatistikler, görselleştirme-animasyonlar için gerekli olan kütüphaneler yazılıma eklenmiştir. Elde edilen istatistiklerin görsel şekiller çıkarılmıştır. Veri setindeki bireylerin yaş dağılımı, cinsiyet ayrımı, öznitelik gruplandırma vb. istatistiksel sonuçları daire, satır, sütun şekilleri ile gösterilmiştir. Gerekli analizler yapılır. Daha sonrasında öznitelik ayrımı yapılır. Bu ayrım sonucu hedef dışındaki tüm veriler giriş(x), hedef özniteliği ise çıkış(y) olarak ayrılmıştır. Hold out yöntemiyle %20 test ve %80 eğitim verisi olarak veri seti ayrımı yapılır. LR, KA, NB, K-EYK, RO, DVM algoritmaları olan 6 makine öğrenme algoritması sırasıyla programa eklenir.

Sonrasında başarı ve hata oranları hesaplanır. Bu temel MÖ algoritmaları sütun grafiği ile karşılaştırması yapılır. Öznitelik seçimi olarak adlandırılan 24 farklı özniteliklerin algoritmalarındaki başarı sonuçları hesaplanır. Yapılan tahminlerin her algoritma için sonucu karşılaştırılmıştır. Çalışma kapsamında toplamda 28 adet farklı deney yapılmıştır. Çalışmada kümesi içinde hangi

özelliklerin sonuç üzerinde ne kadar etkili olduğunu belirlemek amacıyla öznitelik seçimi (feature selection) yöntem, özellik seçimi işlemi sistemin daha hızlı çalışmasını sağladığından dolayı uygulanmıştır.



Şekil 7. Çalışmanın Akış Diyagramı

IV. BULGULAR VE TARTIŞMA [RESULT AND FINDINGS]

Çalışma kapsamında 3 farklı grupta toplamda 28 adet deney gerçekleştirilmiştir. İlk grupta 13 adet deney yapılmıştır. Bu grupta yapılan deneylerin amacı veri setindeki öznitelik sayısının her bir deneyde farklılaştırılarak başarı oranının ortaya komasıdır. İkinci grupta 10 adet deney yapılmıştır. Bu grupta yapılan deneylerin amacı, birinci grup deneyler sonucunda ortaya çıkan modellerin test verisi ile test edilmesidir. Üçüncü grup deneyde ise 5 adet deney yapılmıştır. Bu grupta yapılan deneylerin amacı GA ile öznitelik seçimi yapılarak başarı oranını ortaya koymaktır. Deneylerde veri setinde bulunan bireylerin hasta veya sağlıklı olduğunun tahmini, çalışmadaki en başarılı sonuç %93.44 ile GA olmuştur.

Diğer algoritmalar için en başarılı sonuçlar; LR=%90.16, K-EYK=%81.96, DVM=%83.60, NB=%85.24, KA=%81.96, RO=83.60 olarak hesaplanmıştır. Çalışmanın başarı sonuçları %68.13 ile %93.44 arasındadır. İlk grup 13 deneyde giriş verilerinin ve MÖ algoritmalarının değişikliğine göre başarı sonuçları değişmiştir. Bu deneylerde elde edilen başarı sonuçları Tablo 2’de, hata sonuçları ise Tablo 3’de gösterilmiş, ayrıca algoritmalarından elde edilen başarı sonuçları karşılaştırılmıştır.

Tablo 2. Öznitelik seçimi ile MÖ algoritmalarının başarı sonuçları

Deney No.	LR (%)	K-EYK (%)	DVM (%)	NB (%)	KA (%)	RO (%)
1.	90.16	70.49	63.93	83.60	78.68	81.96
2.	84.61	70.49	68.85	83.60	75.40	80.32
3.	81.31	70.49	67.21	80.32	77.049	80.32
4.	85.71	68.85	67.21	83.60	77.049	83.60
5.	85.71	68.85	67.21	83.60	78.68	81.96
6.	85.71	72.13	67.21	83.60	81.96	83.60
7.	85.71	70.49	67.21	83.60	81.96	83.60
8.	86.81	72.13	68.85	83.60	80.32	78.68
9.	87.91	75.40	81.96	83.60	70.49	75.40
10.	81.31	81.96	83.60	81.24	80.32	81.96
11.	81.31	73.77	83.60	85.24	81.96	81.96
12.	68.13	70.49	78.68	78.68	77.049	77.049
13.	75.82	73.77	73.77	73.77	73.77	73.77

Tablo 2 ve 3 üzerinde yapılan 13 deney kapsamında yaş, cinsiyet, göğüs ağrısı tipi, dinlenme kan basıncı, kolesterol, açlık kan şekeri, dinlenme ekg ölçümü, en fazla kalp atış hızı, egzersize bağlı ağrı, dinlenmeye göre egzersizin neden olduğu ST depresyonu, ST segmentinin eğimi, ana damarların sayısı, talasemi sütunları x, hedef sütunu y olarak temel MÖ algoritmalarının öznitelik seçimi yapılarak başarı oranı değerleri hesaplanmıştır.



Tablo 3. Öznitelik seçimi ile MÖ algoritmalarının hata sonuçları

Deney No.	LR (%)	K-EYK (%)	DVM (%)	NB (%)	KA (%)	RO (%)
1.	09.84	29.51	36.07	16.40	21,32	18.04
2.	15.39	29.51	31.15	16.40	24.60	19.68
3.	18.69	29.51	32.79	19.68	22.951	19.68
4.	14.29	31.15	32.79	16.40	22.951	16.40
5.	14.29	31.15	32.79	16.40	21.32	18.04
6.	14.29	27.87	32.79	16.40	18.04	16.40
7.	14.29	29.51	32.79	16.40	18.04	16.40
8.	13.19	27.87	31.15	16.40	19.68	21.32
9.	12.09	24.60	18.04	16.40	29.51	24.60
10.	18.69	18.04	16.40	18.76	19.68	18.04
11.	18.69	26.23	16.40	14.76	18.04	18.04
12.	31.17	29.51	21.32	21.32	22.951	22.951
13.	24.18	26.23	26.23	26.23	26.23	26.23

Bu deneylere ilişkin elde edilen sonuçlar aşağıda yer almaktadır. Yapılan deneylerde genel olarak her algoritma için veriler %20 test ve %80 eğitim verisi olarak ayrılmıştır.

- LR algoritması kullanılarak yapılan deneylerde LR aşamasında algoritma 109 defa eğitilerek %90.16 başarı sonucuna ulaşılmıştır. Hata oranı %9.84'tür.
- K-EYK algoritmasında 3 komşu dikkate alınmıştır. Komşu sayısının tek sayı verilmesi optimum sonuç olmasına yardımcı olacağından hareketle komşu sayısı belirlenmiştir. Bu algoritma 41 defa eğitilerek %70.49 başarı sonucuna ulaşılmıştır. Hata oranı %29.51'dir.
- DVM algoritması 100 defa eğitilerek %63.93 başarı sonucuna ulaşılmıştır. Hata oranı da %36.07'dir.
- NB algoritması 41 defa eğitilerek %83.60 başarı sonucuna ulaşılmıştır. Hata oranı %16.40'tır.
- KA algoritması 109 defa eğitilerek %78.68 başarı sonucuna ulaşılmıştır. Hata oranı da 21.32'dir.
- RO algoritması 41 defa eğitilerek %81.96 başarı sonucuna ulaşılmıştır. Hata oranı %18.04'tür.

Tablo 2 ve 3 üzerindeki deneylerden 7 numaralı deneyde; dinlenme EKG ölçümü, en fazla kalp atış hızı,

egzersize bağlı ağrı, dinlenmeye göre egzersizin neden olduğu ST depresyonu, ST segmentinin eğimi, ana damarların sayısı, Talasemi sütunları x, hedef sütunu y olarak temel MÖ algoritmalarının başarı oranı değerleri hesaplanmıştır. Veri setinde bulunan yaş, cinsiyet, göğüs ağrısı tipi ve dinlenme kan basıncı, kolesterol, açlık kan şekeri hariç tüm sütunlar dahil edilmiştir.

Bu deneylere ilişkin elde edilen sonuçlar aşağıda yer almaktadır:

- LR algoritmasında %30 test ve %70 train olarak ayrılmıştır. Bu algoritma 41 defa eğitilerek %85.71 başarı sonucuna ulaşılmıştır. Hata oranı %14.29'dur.
- K-EYK algoritmasında 3 komşu dikkate alınmıştır. Komşu sayısının tek sayı verilmesi optimum sonuç olmasına yardımcı olur. Bu algoritmada %20 test ve %80 train olarak ayrılmıştır. Algoritma 41 defa eğitilerek %70.49 başarı sonucuna ulaşılmıştır. Hata oranı %29.51'dir.
- DVM algoritması 100 defa eğitilerek %67.21 başarı sonucuna ulaşılmıştır. Hata oranı da %32.79'dur.
- NB algoritması 41 defa eğitilerek %83.60 başarı sonucuna ulaşılmıştır. Hata oranı %16.40'tır.
- KA algoritması 109 defa eğitilerek %81.96 başarı sonucuna ulaşılmıştır. Hata oranı da 18.04'tür.
- RO algoritmasında %20 test ve %80 train olarak verisi ayrılmıştır. Bu algoritma 41 defa eğitilerek %83.60 başarı sonucuna ulaşılmıştır. Hata oranı %16.40'tır.

Çalışmada yapılan ikinci grup deneylerde veri setinde yer alan hasta ve sağlıklı olan toplam 10 bireyin verileri MÖ yöntemleri ile tahminleri yapılarak başarı ve hata sonuçları listelenmiştir. Örneğin; Deney 1'de tahmin edilen birey hastadır ve değeri "1"dir. LR algoritması ile tahmin edilmesi istendiğinde %90.16 başarı oranıyla bireyin hasta olduğu tahmin edilmiştir.

Tablo 4'de görüldüğü gibi bu tahminin hata oranı 0.1764'dür. Hasta birey: Hedef 1 ve Sağlıklı birey: Hedef 0 olarak tanımlanmıştır. Tahminler; %20 test, %80 train olarak ayrılarak 100 defa eğitilerek sonuçlandırılmıştır.

Tablo 4. Veri setinde bulunan 10 bireyin tahmini ve algoritmaların başarı sonuçları

Deney No.	Gerçek Veri	LR (%)	K-EYK (%)	DVM (%)	NB (%)	KA (%)	RO (%)
1.	1	%90,16 (1)	%70,49 (0)	%63,93 (1)	%83,60 (1)	%75,40 (0)	%81,96 (0)
2.	0	%90,16 (0)	%70,49 (1)	%63,63 (0)	%83,60 (0)	%75,40 (0)	%81,96 (0)
3.	1	%90,16 (1)	%70,49 (0)	%63,93 (1)	%83,60 (1)	%75,40 (1)	%81,96 (1)
4.	0	%90,16 (0)	%70,49 (1)	%63,93 (1)	%83,60 (1)	%75,40 (0)	%81,96 (0)
5.	0	%90,16 (0)	%70,49 (0)	%63,93 (0)	%83,60 (0)	%75,74 (0)	%81,96 (0)
6.	1	%90,16 (1)	%70,49 (1)	%63,93 (1)	%83,60 (1)	%75,40 (1)	%81,96 (1)
7.	1	%90,16 (1)	%70,49 (1)	%63,93 (1)	%83,60 (1)	%75,40 (1)	%81,96 (1)
8.	0	%90,16 (0)	%70,49 (0)	%63,93 (0)	%83,60 (1)	%75,40 (0)	%81,96 (0)
9.	0	%90,16 (0)	%70,49 (0)	%63,93 (0)	%83,60 (0)	%72,13 (0)	%81,96 (0)
10.	1	%90,16 (0)	%70,49 (1)	%63,93 (1)	%83,60 (0)	%73,77 (1)	%81,96 (1)

Tablo 5. Veri setinde bulunan bireylerin GA ile tahmini

Deney No.	Gerçek Veri	GA ile Tahmin Veri
1.	1	1
2.	1	1
3.	0	0
4.	0	1
5.	0	0

Yapılan üçüncü aşamadaki deneyde ise GA ile veri setindeki bir bireyin değerleri girilerek hasta veya sağlıklı olup olmadığı tahmin edilmiştir. Tablo 5 üzerinde veri seti içerisinde bulunan 5 adet bireyin GA ile hasta veya sağlıklı olma tahmini yapılmıştır. Bu çalışma kapsamında GA kullanılarak %93.44 başarı ve %6.56 hata sonucu elde edilmiştir. Tablo 5 üzerinde “0” bilgisi sağlıklı bireyi, “1” bilgisi ise hasta bireyi temsil etmektedir.

Tablo 6. Genetik algoritma parametreleri

Parametreler	Değerler
Birey Sayısı	303
Birey Seçim Yöntemi	Sıralama
Çaprazlama Olasılığı	1
Mutasyon Olasılığı	0.1

Tablo 6 üzerinde GA ile analizi yapılan işlemlerde kullanılan parametreler verilmiştir. Bu parametrelerden “Birey Sayısı” ile GA’da hesaplanacak çalışmanın popülasyonuna ait birey sayısını belirlenmiştir. “Birey Seçim Yöntemi” parametresi ile bireylerin çaprazlaması için hangi yöntemin seçeceğini belirlenmiştir. “Çaprazlama Olasılığı” parametresi ile yeni nesli oluştururken seçilen bireylerin ne oranda çaprazlama yapacağını belirten sayıdır. “Mutasyon Olasılığı” parametresi ise mutasyon yapılırken hangi oranda yapılacağını belirleyen olasılık parametresidir.

Tablo 7 üzerindeki veriler ise Tablo 5 üzerindeki 1 numaralı deneydeki bireyin detaylı bilgileri ve GA sonucu tahmin verisi verilmiştir.

Tablo 7. Veri setinde bulunan bir bireyin detay bilgileri

Tahmini yapılacak bireyin özellikleri	Değerler
Yaş	37
Cinsiyet	Erkek
Göğüs ağrısı tipi	2
Dinlenme kan basıncı	130
Kolesterol	250
Açlık kan şekeri	0
Dinlenme ekg ölçümü	1
Maksimum kalp atış hızı	187
Egzersize bağlı ağrı	0
Dinlenmeye göre egzersizin neden olduğu ST depresyonu	3.5
ST segmentinin eğimi	0
Ana damar sayısı	0
Talasemi	2
Hedef	1(Hasta)

Şekil 8’de GA ile bireyin tahmin sonucunu gösteren kod gösterilmiştir.

```
data=data[data["Ca"]==0]
data=data[data["Thal"]==2]
tahmin=logmodel.predict([[37,1,2,130,250,0,1,187,0,3.5,0]])
print(tahmin,accuracy_score(y_test,predictions))
#target 1
```

[1] 0.9344262295081968

Şekil 8. GA ile hastalık tahmini kod parçası

Genetik algoritma (GA), çok boyutlu uzayda maliyet fonksiyonuna uygun olacak şekilde iterasyonlar yaparak optimum çözümü ararlar. LR ise bağımlı ve bağımsız



değişken arasındaki ilişkiyi en az sayıda değişken ile en iyi uyuma sahip olacak şekilde tahmin etmeyi amaçlar. Farklı sayıda değişken içeren regresyon modellerinden başarı sonucu yüksek olacağından genellikle Genetik Algoritma kullanılması tavsiye edilir. Bu sebeple LR başarısı yüksek çalışmaların GA başarısı da yüksektir. Bu çalışmada da LR yöntemi başarı sonucu % 90.16 olarak elde edilmiştir. Çalışmada GA temel MÖ yöntemleri ile karşılaştırılmış ve daha yüksek başarı sonucunu göstermiştir.

V. SONUÇ VE DEĞERLENDİRME [CONCLUSION]

Günümüzde YZ'nin bir alt dalı olan MÖ algoritmaları sağlık alanında çeşitli hastalıklara ilişkin teşhislerde uygun maliyet, hızlı ve daha doğru sonuçlar elde ederek tahminler yapmak mümkündür. Dünyada kalp hastalığı sonucu ölümlerin ilk sırada olması, aynı zamanda her geçen gün kalp hastalığı nedeniyle ölüm sayısının artması bu hastalığa ilişkin teşhislerde doğru tahminler yapılmasını gerektirmektedir. MÖ algoritmaları ile eğer bu hastalık yüksek başarı oranları ile tahmin edilebilirse erken teşhis ve önlemler sayesinde ölüm oranları azaltılabilir. Bu çalışmada kalp hastalığının GA yaklaşımıyla öznitelik seçimi kullanılarak MÖ algoritmaları ile tahmini gerçekleştirilmiştir. Bu kapsamda MÖ yöntemleri ve GA'dan faydalanılmıştır. Çalışmada sınıflama işlemi için hangi özniteliklerin önemli hangilerinin ise daha az önemli olduklarının belirlenmesi aşamasında öznitelik seçimi işlemi için GA metodundan faydalanılmıştır. Çalışmada ABD'nin Cleveland eyaletindeki bireylerin verisinden yararlanılarak Andras Janosi, William Steinbrunn, Matthias Pfister ile Robert Detrano tarafından oluşturulan ve Kaggle'dan [47] alınan veri seti kullanılmıştır. Çalışmada öncelikle kullanılan veri setinde GA'dan faydalanılmış ve MÖ yöntemlerinde öznitelik seçimi yapılmış, daha sonra belirlenen MÖ algoritmaları kalp hastalığı tahmininin yüksek doğruluk oranı ile yapılabileceği ortaya konmuş, böylelikle veri setinde GA ve MÖ yöntemlerinde öznitelik seçiminin tahmin doğruluk oranını artırdığı gösterilmiştir. Çalışma kapsamında kullanılan veri seti üzerinde K-EYK, LR, KA, RO, NB, DVM ve GA algoritmaları ile 3 grup deney gerçekleştirilmiştir. Farklı aşamada toplam 28 (yirmi sekiz) deney gerçekleştirilmiştir.

Çalışma kapsamında yapılan deneylerde alınan başarı sonuçları sırasıyla LR=90.16, K-EYK=81.96, DVM=83.60, NB=85.24, KA=81.96, RO=83.60 olarak ölçülmüştür. Elde edilen hata oranlarına bakıldığında en başarılı sonucun GA yaklaşımıyla yapılan tahminde %93.44 başarı oranı elde edildiği görülmektedir. Çalışmada GA ile tahmin yapıldıktan sonra elde edilen başarı oranının daha yüksek olduğu gözlemlenmektedir. Çalışmada elde edilen sonuçlar GA yaklaşımı ve MÖ

yöntemlerinde öznitelik seçimi kullanılarak kalp hastalığının daha yüksek başarı oranı ve doğrulukla tahmin edilebileceğini göstermektedir. Bu çalışma bize göstermiştir ki ilgili algoritma yapıları üzerinde veri sayısı ve öznitelik sayısı değişimi algoritmalar arasında başarı oranını etkilemektedir.

GA ve MÖ yöntemleri ile geliştirilen bu çalışmanın sağlık alanındaki literatüre katkı sağlaması amaçlanmıştır.

ÇIKAR ÇATIŞMASI [CONFLICTS OF INTEREST]

Yazarlar arasında ve ilgili kurumları arasında herhangi çıkar çatışması olmadığını bildirmişlerdir.

ETİK KURALLARA UYGUNLUK [RESEARCH AND PUBLICATION ETHICS]

Yazarlar bu makalenin etik kurul onayı veya herhangi bir özel izin gerektirmediğini beyan ederler.

KAYNAKLAR [REFERENCES]

- [1] M. Pavithra, A. M. Sindhana, T. Subajanaki, and S. Mahalakshmi, 2021. "Effective Heart Disease Prediction Systems Using Data Mining Techniques", *Annals of the Romanian Society for Cell Biology*, 6566-6571.
- [2] L. Tokgozoglu, M. Kayikcioglu, and B. Ekinici, 2021. "The Landscape of Preventive Cardiology in Turkey: Challenges and successes", *American Journal of Preventive Cardiology*, 100184.
- [3] İ. Ozcan, B. Tasar, A. B. Tatar, and O. Yakut, 2019. "Destek Vektör Makinesi Algoritması ile Kalp Hastalıklarının Tahmini", *Bilgisayar Bilimleri*, 4(2), 74-79.
- [4] A. Coşkun, and N. Arıcı, 2006. "Genetik Algoritma Kullanılarak Deriden Geçiş Özelliklerine Göre Kimyasal Maddelerin Molekül Yapılarının Bulunması", *Politeknik Dergisi*, 9(4), 255-261.
- [5] M. Karakoyun, and M. Hacıbeyoğlu, 2014. "Biyomedikal Veri Kümeleri ile Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatistiksel Olarak Karşılaştırılması", *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 16(48), 30-42.
- [6] A. K. M. Rahman, F. M. J. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, 2019. "A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms", *International Journal of Scientific & Technology Research*, 8(11), 419-422.
- [7] Şengür, D., 2021. "Eeg, Emg And Ecg Based Determination of Psychosocial Risk Levels in Teachers Based on Wavelet Extreme Learning Machine Autoencoders", *Politeknik Dergisi*, 1-1.
- [8] M. Nazari, I. Shiri, and H. Zaidi, 2021. "Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients", *Computers in Biology and Medicine*, 129, 104135.



- [9] Ö. Ekrem, O. K. M. Salman, B. Aksoy, and S. A. İnan, 2020. "Yapay Zekâ Yöntemleri Kullanılarak Kalp Hastalığının Tespiti", *Mühendislik Bilimleri ve Tasarım Dergisi*, 8(5), 241-254.
- [10] M. A. Ahmad, C. Eckert, and A. Teredesai, 2018. "Interpretable Machine Learning In Healthcare", *In Proceedings Of The 2018 Acm International Conference On Bioinformatics, Computational Biology, And Health Informatics*, 559-560.
- [11] U. Köse, 2019. "Zeki Optimizasyon Tabanlı Destek Vektör Makineleri ile Diyabet Teşhisi", *Politeknik Dergisi*, 22(3), 557-566.
- [12] Ş. Pazar, M. Bulut, and C. Uysal, 2020. "Yapay Zeka Tabanlı Araç Algılama Sistemi Geliştirilmesi", *Journal of Scientific, Technology and Engineering Research*, 1(1), 31-37.
- [13] D. Mishra, and B. Sahu, 2011. "Feature Selection for Cancer Classification: a Signal-To-Noise Ratio Approach", *International Journal of Scientific & Engineering Research*, 2(4), 1-7.
- [14] E. Çağlayan Akay, 2020. "Ekonometride Büyük Veri ve Makine Öğrenmesi: Temel Kavramlar", *Der Yayınları*.
- [15] S. Kushwaha, S. Bahl, A. K. Bagha, K. S. Parmar, M. Javaid, A. Haleem, and R. P. Singh, 2020. "Significant Applications Of Machine Learning For Covid-19 Pandemic", *Journal of Industrial Integration and Management*, 5(4).
- [16] M. Gök, 2017. "Makine Öğrenmesi Yöntemleri ile Akademik Başarının Tahmin Edilmesi", *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 5(3), 139-148.
- [17] C. Akyel, and N. Arıcı, 2020. "Cilt Kanserinde Kıl Temizliği ve Lezyon Bölütlemesinde Yeni Bir Yaklaşım", *Politeknik Dergisi*, 23(3), 821-828.
- [18] A. Abdullah Sheik, and R. R. Rajalaxmi, 2012. "A Data Mining Model For Predicting The Coronary Heart Disease Using Random Forest Classifier", *In International Conference in Recent Trends in Computational Methods, Communication And Controls*, 22-25.
- [19] V. Chaurasia, and S. Pal, 2014. "Performance Analysis Of Data Mining Algorithms For Diagnosis And Prediction Of Heart And Breast Cancer Disease", *Review of Research*, 3(8).
- [20] Ö. Özmen, A. Khdr, E. Avcı, 2018. "Sınıflandırıcıların Kalp Hastalığı Verileri Üzerine Performans Karşılaştırması", *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 30(3), 153-159.
- [21] M. Polatgil, 2020. "Anfis ve Bulanık K Ortalamalar ile Kalp Hastalığının Tespit Edilmesi", *Bilişim Teknolojileri Dergisi*, 13(4), 443-449.
- [22] S. Amin Umar, K. Agarwal, and R. Beg, 2013. "Genetic Neural Network Based Data Mining In Prediction Of Heart Disease Using Risk Factors", *In 2013 IEEE Conference On Information and Communication Technologies*, 1227-1231.
- [23] Ş. Cihan, B. Karabulut, G. Arslan, and G. Cihan, 2018. "Koronar Arter Hastalığı Riskinin Veri Madenciliği Yöntemleri ile İncelenmesi", *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 10(1), 85-93.
- [24] M. Pagnesi, M. Adamo, and M. Metra, 2021. "March 2021 at a glance: focus on epidemiology, prevention and COVID-19", *European Journal of Heart Failure*, 23(3): 347-349.
- [25] E. M. Gökaş, and M. Yağanoğlu, 2020. "Veri Bilimi Uygulamalarının Hastalık Teşhisinde Kullanılması: Kalp Krizi Örneği", *Journal Of Information Systems And Management Research*, 2(2), 26-32.
- [26] A. Çil, E. Edem, and Ö. Göldedi, 2018. "Hipertansif Hastalarda Kalp Hızı Değişkenliği ile Kardiyovasküler Risk Faktörlerinin İlişkisi", *Sakarya Tıp Dergisi*, 8(2), 303-310.
- [27] Y. Olgar, D. Billur, and B. Turan, 2021. "Yaşa Bağlı Kalp Fonksiyon Değişiklikleri ve miRNA'lar", *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 74(2):239-244.
- [28] I. Kalan, and Y. Yeşil, 2010. "Obezite ile İlişkili Kronik Hastalıklar", *Diyabet ve Obezite*, 78.
- [29] S. S. Virani, A. Alonso, H. J. Aparicio, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, ... and American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee, 2021. "Heart disease and stroke statistics—2021 update: a report from the American Heart Association", *Circulation*, 143(8), e254-e743.
- [30] M. H. Ünver, Y. Kökver, and A. Çifci, 2020. "Hipertansiyon Tahmini İçin Temel Bileşen Analizinin Kullanımı", *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 12(3), 42-51.
- [31] N. Küçükberber, K. Özdu, and H. Yorulmaz, 2011. "Kalp Hastalarında Sağlıklı Yaşam Biçimi Davranışları ve Yaşam Kalitesine Etki Eden Faktörlerin Değerlendirilmesi", *Anadolu Kardiyol Derg*, 11, 619-26.
- [32] M. İyigün, 2021. "Kardiyovasküler ve Torasik Cerrahi", *In Geriatrik Anestezi*, 283-294. *Ankara: Güneş Kitabevi*.
- [33] A. Dar Masroof, and G. Atul, 2019. "Performance Analysis Of Data Mining Algorithms", *Journal Of Computational And Theoretical Nanoscience*, 16(9), 3849-3853.
- [34] M. D. Camacho, M. K. Collins, K. R. Powers, C. J. Costello, and J. J. Collins, 2018. "Next-Generation Machine Learning For Biological Networks", *Cell*, 173(7), 1581-1592.
- [35] M. E. Taşçı, and R. Şamlı, 2020. "Veri Madenciliği ile Kalp Hastalığı Teşhisi", *Avrupa Bilim ve Teknoloji Dergisi*, 88-95.
- [36] A. M. Aydın, 2021. "Müşteri Kaybı Tahmininde Sınıf Dengesizliği Problemi", *Politeknik Dergisi*, 1-1.
- [37] B. Karaoğlu, 2015. "Makine Öğrenmesi ile Spor Karşılaşmalarının Modellenmesi", *Emo Bilimsel Dergi*, 5(9), 1-5.
- [38] E. Uzundurukan, and A. Kara, 2020. "Deep Learning Based Threat Classification for Fiber Optic Distributed Acoustic Sensing Using SNR Dependent Data Generation", *Journal of Scientific, Technology and Engineering Research*, 1(2), 4-12.



- [39] D. Sarkar, R. Bali, and T. Sharma, 2018. "Forecasting Stock And Commodity Prices in Practical Machine Learning With Python, *Apress*, 467-497.
- [40] R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, A. Khosravi, M. Panahiazar, A. Koohestani, F. Khozeimeh, S. Nahavandi, and N. Sarrafzadegan, 2019. "A Database For Using Machine Learning And Data Mining Techniques For Coronary Artery Disease Diagnosis", *Scientific Data*, 6(1), 1-13.
- [41] B. Çağlıyan and U. Köse, "Epilepsi EEG Verilerinin Makine Öğrenmesi Teknikleriyle Sınıflandırılması", *Avrupa Bilim ve Teknoloji Dergisi*, no. 23, pp. 163-172.
- [42] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, 2017. "Credit card fraud detection using machine learning techniques: A comparative analysis", *In 2017 International Conference on Computing Networking and Informatics (ICCNi)*, *IEEE*, 2017, 1-9.
- [43] M. Selimoglu, and A. Yılmaz, 2021. "Kredi Kartı Dolandırıcılık Tespitinin Makine Öğrenmesi Yöntemleri ile Tahmin Edilmesi", *Beykent Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 13(2), 28-33.
- [44] H. Yıldırım, 2021. "Comparative Analysis of Machine Learning Algorithms Based on Variable Importance Evaluation", *Journal of Scientific, Technology and Engineering Research*, 2(2), 46-53.
- [45] H. Kabaoglu, E. Uçar, and F. Duran, 2021. "Buzlanma Tahmini Yapan Mobil Uygulama Geliştirilmesi", *Politeknik Dergisi*, 1-1.
- [46] E. Tuncer, and D. E. Bolat, 2021. "Destek Vektör Makinaları ile EEG Sinyallerinden Epileptik Nöbet Sınıflandırması", *Politeknik Dergisi*, 1-1.
- [47] C. Aydın, 2018. "Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması", *Avrupa Bilim ve Teknoloji Dergisi*, (14), 169-175.
- [48] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, 2021. "Multiple disease prediction using Machine learning algorithms", *Materials Today: Proceedings*.
- [49] A. F. Şenel, 2020. "Makine Öğrenmesi Algoritmaları Kullanılarak Kayısı İç Çekirdeklerinin Sınıflandırılması", *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 9(2), 807-815.
- [50] Ş. B. Bozyiğit, and Ç. Tarhan, 2020. "Makine Öğrenmesi Yöntemleri Kullanılarak Sosyal Medyada Marka İtibar Analizi", *Yönetim Bilişim Sistemleri Dergisi*, 6(2), 57-76.
- [51] M. Libbrecht, W. Noble, 2015. "Machine Learning Applications in Genetics And Genomics", *Nature Reviews Genetics*, 16(6), 321-332.
- [52] M. Atalay, and E. Çelik, 2017. "Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları", *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.
- [53] A. Öztürk, Ü. Durak, and F. Badıllı, 2020. "Twitter Verilerinden Doğal Dil İşleme ve Makine Öğrenmesi ile Hastalık Tespiti", *Konya Mühendislik Bilimleri Dergisi*, 8(4), 839-852.
- [54] A. E. Keleş and M. Kaya Keleş, "Genetik Algoritma ile Öznitelik Seçimi Kullanılarak Yapım Yönetiminde Çalışan Liderlik Algısının Tahmini", *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 33, no. 4, pp. 97-110.
- [55] N.B. Bahadure, A. K. Ray, and H. P. Thethi, 2018. "Comparative Approach of MRI-Based Brain Tumor Segmentation and Classification Using Genetic Algorithm", *Journal of digital imaging*, 31(4), 477-489.
- [56] S. Katoch, S. S. Chauhan, and V. Kumar, 2021. "A review on genetic algorithm: past, present, and future", *Multimedia Tools and Applications*, 80(5), 8091-8126.
- [57] H. Candan, A. Durmuş, and G. Harman, 2019. "Genetik Algoritma ve Sınıflandırıcı Yöntemler ile Kanser Tahmini", *Veri Bilimi*, 2(1), 30-34.
- [58] S. Karasu, and Z. Saraç, 2020. "Güç Kalitesi Bozulmalarının Hilbert-Huang Dönüşümü, Genetik Algoritma ve Yapay Zeka/Makine Öğrenmesi Yöntemleri İle Sınıflandırılması", *Politeknik Dergisi*, 23(4), 1219-1229.
- [59] B. C. Kalaycı, O. Ertenlice, H. Akyer, and H. Aygoren, 2017. "A Review On The Current Applications Of Genetic Algorithms in Mean-Variance Portfolio Optimization", *Pamukkale University Journal of Engineering Sciences*, 23(4), 470-476.
- [60] Heart Disease Data Set, Available, <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>, Accessed Time: 15.06.2021
- [61] M. Koklu, S. Sarigil, and O. Ozbek, 2021. "The use of machine learning methods in classification of pumpkin seeds (*Cucurbita pepo* L.)", *Genetic Resources and Crop Evolution*, 1-14.
- [62] P. Ranganathan, S. C. Pramesh, and R. Aggarwal, 2017. "Common Pitfalls In Statistical Analysis: Logistic Regression", *Perspectives in Clinical Research*, 8(3), 148.
- [63] C. P. Austin, and J. Merlo, 2017. "Intermediate And Advanced Topics In Multilevel Logistic Regression Analysis", *Statistics in Medicine*, 36(20), 3257-3277.
- [64] M. Öztürk, 2018. Paksoy T., and Öztürk M., "Genetik Algoritmalar (GA) ile Nüfus Artış Tahmini", *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 11(1), 40-51.



RESEARCH ARTICLE

Classification of Blood Cells from Blood Cell Images Using Dense Convolutional Network

* Ferhat Bozkurt

* Ataturk University, Department of Computer Engineering, Erzurum, Turkey
fbozkurt@atauni.edu.tr, Orcid: 0000-0003-0088-5825

HIGHLIGHTS

- A Dense CNN model is proposed for white blood cell classification
- Comparison in terms of accuracy, precision, recall, and F1 measure with other classifiers
- The overall accuracy rate of 94% was obtained with DenseNet121 model
- Use of publicly available dataset for state-of-the-art comparison with other studies

GRAPHICAL ABSTRACT

This study aims to classify white blood cells from blood cell images. Input image goes along multiple convolutions and acquire high-level features with standard CNN. In Dense Convolutional Network (DenseNet), each layer acquires additional inputs from all previous layers and transfers on its own feature-maps to all next layers as shown in graphic. DenseNet is a dense-connected convolutional network that forward connects each layer to the other layers. DenseNet gives superior performance for various computer vision problems. DenseNet reduces the gradient descent problem, increases feature reuse, and reduces parameter usage. The advantage of DenseNet architectures is to allow feature propagation and feature reuse. Thus it minimizes the number of parameters. The main structure block of DenseNet121 is the Dense Block. These Dense Blocks consist of convolution layers. DenseNet121 contains 121 connected convolutional layers with a fully-connected layer of 1000 units as the final output layer.

Keywords:

- White blood cells
- Classification
- Deep learning
- DenseNet
- Pre-trained model

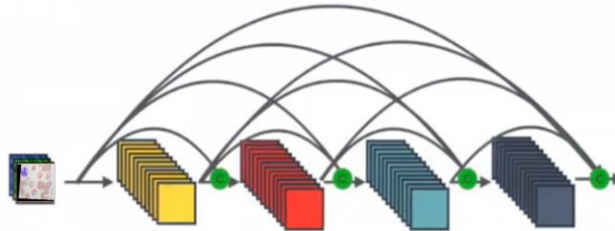


Figure A. Fundamental architecture of one dense block in DenseNet

Article Info:

Received : 24.10.2021
Accepted : 14.11.2021
Published: 21.12.2021

DOI:

10.53525/jster.1014186

*Correspondence:

Ferhat Bozkurt
fbozkurt@atauni.edu.tr
+904422316057.

Aim of Article : This study aims to classify white blood cells from blood cell images.

Theory and Methodology: In this study, blood cells are classified from blood cell images using dense convolutional neural network. This paper intends to utilize a dense convolutional neural network model (DenseNet121) to overcome the blood cell classification problem that is one of the most compelling problems in blood diagnosis.


Findings and Results: The data in the experimental study: 70% is reserved for training, 10% for validation, and 20% for testing. DenseNet121 model is built with batch size 128 and with 100 epochs at training. The total number of parameters is 7,628,484 for this model. DenseNet-121 has achieved the best feature extractor pre-trained model with accuracy rate of 94%.

Conclusion: Considering the high accuracy value obtained with this model, automatic detection of which class the cells belong to will speed up the diagnosis and allow more data to be examined.



RESEARCH ARTICLE

Classification of Blood Cells from Blood Cell Images Using Dense Convolutional Network

*  Ferhat Bozkurt

* Ataturk University, Department of Computer Engineering, Erzurum, Turkey
fbozkurt@atauni.edu.tr, Orcid: 0000-0003-0088-5825

Citation:

Bozkurt, F. (2021). *Classification of Blood Cells from Blood Cell Images Using Dense Convolutional Network*, Journal of Scientific Technology and Engineering Research, 2(2):81-88. DOI: 10.53525/jster.1014186

HIGHLIGHTS

- A Dense CNN model is proposed for white blood cell classification
- Comparison in terms of accuracy, precision, recall, and F1 measure with other classifiers
- The overall accuracy rate of 94% was obtained with DenseNet121 model
- Use of publicly available dataset for state-of-the-art comparison with other studies

Article Info

Received : 24.10.2021
Accepted : 14.11.2021
Published: 21.12.2021

DOI:

10.53525/jster.1014186

*Corresponding Author:

Ferhat Bozkurt
fbozkurt@atauni.edu.tr
Phone: +904422316057

ABSTRACT

Classification of white blood cells plays a significant role in the detection of diseases which are infections caused by abnormalities in the immune system, allergies, anemia, leukemia, cancer, AIDS, etc. In traditional methods, experts manually examine the white blood cells under the microscope, and since this process is tedious, takes more time and can be more error-prone, automated systems have become necessary. Making this classification automatically will help experts in the detection of diseases. In this study, blood cells are classified from blood cell images using dense convolutional neural network. This paper intends to utilize a dense convolutional neural network model to overcome the blood cell classification problem that is one of the most compelling problems in blood diagnosis. In this study, a DenseNet121 model is built to classify blood cell images. Experiments are conducted on an open-access BCCD dataset (12507 white blood cell images that contain four types of white blood cells). Performance evaluations are performed on the accuracy of the techniques, and the results are compared with state-of-the-art deep learning-based approaches as Xception, VGG19, EfficientNetB1. In experimental studies, the highest accuracy (94%) is obtained with the proposed DenseNet121 model. Considering the high accuracy value obtained with this model, automatic detection of which class the cells belong to will speed up the diagnosis and allow more data to be examined by doctors.

Keywords: White blood cells, Classification, Deep learning, DenseNet, Pre-trained model

I. INTRODUCTION

White blood cells (WBC) are one of the most essential cells for the human immune system. They are the cells that protect the human body against various infectious diseases and foreign invaders. While most white blood cells are produced in the bone marrow, some are produced by important glands in the body [1], [2]. Information that will be useful for diagnosing diseases can be learned in the blood, which is examined under the microscope. Recognition and examination of WBC

in peripheral blood can help hematologists in diagnosing diseases like allergies, anemia, leukemia, cancer, AIDS, and blood cancer [3]. Studies in this context have become one of the most critical steps in hematological procedures. Classification of white blood cells plays a significant role in the detection of diseases which are infections caused by abnormalities in the immune system. Making this classification will help experts in the detection of diseases. One of the challenges in medical science is the detection and recognition of white blood cells, since red blood cells



are much more numerous than white blood cells in body [4]. In a healthy adult individual, white blood cells make up about 1% of the total blood volume [5]. Due to this low ratio of white blood cells in the blood, the classification and recognition of white blood cells and especially their subtypes becomes a more difficult problem. A change in the number of any white blood cell subtype could mean that something is wrong with body and body is reacting to some type of pathogen. If we can recognize the white blood cells and their subtypes, we can get an idea about the future course of any disease and take various precautions in advance based on this information. Existing manual methods are error-prone and heavily time-consuming. There is a need for faster and more real solutions with less chance of error and to eliminate these disadvantages [6].

In this study, a DenseNet model is built to classify blood cell images. The rest of the work is organized as follows. In Section 2, the related works in literature are presented. Material and method are explained in Section 3. In this section, the data set used in the experimental study and the DenseNet model applied are explained in detail. In Section 4, experimental results are given, and the discussion and conclusion part of the study is given in Section 5.

II. RELATED WORKS

Researchers have recently proposed many methods for classifying blood cells because of the importance of blood cell classification in diagnosis. Related studies using similar open-access dataset [7] with this study are summarized as follows. In 2018, Liang et al. [8] proposed a model that assembles features of CNN (used Xception) and RNN (used LSTM). They classified cells using CNN & RNN with an accuracy rate of 90.79%. D. Bani-Hani et al. [9] proposed using CNN for the image classification of four blood cells with optimized CNN's hyperparameters with Genetic Algorithm. They classified cells using CNN & Genetic Algorithm with accuracy rate of 91.00%. In 2019, Şengür et al. [10] used LSTM model which is based on shape and deep feature to classify white blood cells. They achieved 92.89% accuracy with LSTM model. Banik et al. [11] proposed fused CNN model to classify blood cell images. They achieved 90.79 classification accuracy on WBC dataset. In 2020, Patil et al. [12] used to merge CNN and RNN to reinforce the perceiving of image content. Moreover, they used

Canonical Correlation Analysis to overcome problem of multiple cells overlap. They classified cells using this technique with a accuracy rate of 95.48%. In 2020, Mohamed et al. [13] studied the performance ten different pretrained models with six types of machine learning techniques. They reported that DenseNet-169 has achieved the best feature extractor pre-trained model with accuracy rate of 92.00%. In 2020, Dekhil [14], proposed a customized CNN architecture with extracted ROI. This technique achieved rate of accuracy with 92.50%. In 2021, Ghosh and Kundu [15] built a multi-layer network model by combining RNN+CNN to classify different blood cells. They obtained 87.29% accuracy with RNN+CNN model. Ekiz et al. [16] classified cells using the convolutional features in the SVM classifier with an accuracy rate of 85.96%. Other hybrid deep learning methods are summarized as follows. In 2017, X. Li et al. [17] used feature extraction with PCA and then compared the results obtained using an 8-layer CNN architecture with a trained SVM. They presented that CNN was 30% more successful than SVM. Yu et al. [18] classify WBC sub-types by using transfer learning and fine-tuning methods. At the end of their studies, they achieved better results than conventional methods and a standard CNN. Jiang et al. [19] first pre-processed the datasets (data augmentation, PCA, etc.) and then trained the datasets with a 33-layer CNN architecture.

As can be observed in the literature, despite it is possible to achieve high success measures with deep learning, the need for a large amount of data during training process a major problem for researchers working with medical images [20]. For this reason, it is observed that researchers benefit from data augmentation techniques in data sets, thus trying to increase the number of data to an adequate level. Essentially, accuracy can be increased even more with original data. In addition, as observed in the studies, some researchers used transfer learning/fine-tuning methods to overcome with the insufficient data problem, while others used hybrid models in their studies. Also, despite deep learning methods can make extraction directly from the image itself, it has been observed that researchers benefit from pre-processing processes in order to increase the accuracy rate. On the other hand, in cases where the examples in the classes of the data set are insufficient, it has been observed in some studies that some classes are not included in the training.

III. MATERIAL AND METHOD

A. Material

The blood consists of three different components, these are white, red blood cells, and thrombocytes. White blood cells consist of five different types of cells. The cells are exist in the rate of Basophil 0-1%, Monocyte 2-10%, Eosinophil 1-5%, Lymphocyte 20-45%, and Neutrophil 50-70% [21]. Neutrophils are the most abundant type of white blood cell in the body. This cell type protects against infections from toxic substances, deals with viruses, bacteria, fungi, and even cancer cells. Lymphocyte deals with infection and prevent infections. Eosinophil deals with and prevents parasitic and bacterial infections. Eosinophils are white blood cells that control inflammation that causes asthma and allergies. Monocyte removes microorganisms, unknown agents and dead cells. Basophil is a kind of WBC that assists prevent and treat wound infections. It contains suitable substances that help relieve allergies and control blood clotting [22].

Blood cell images in this study are obtained from the BCCD dataset of the popular Kaggle database [7]. This data set consists of images referring to four different classes as Eosinophil, Lymphocyte, Monocyte, and Neutrophil. Fig. 1 shows samples of four cells types in the dataset. Size of all images in the dataset is same and in the dimension of 240x320x3. A total of 12.507 augmented images of blood cells in JPG format are available as 3133 Eosinophil, 3108 Lymphocyte, 3095 Monocyte, and 3171 Neutrophil images. Fig. 2 shows categorical distribution of the blood cells data used in this study as 70% training, 10% validation, and 20% testing.

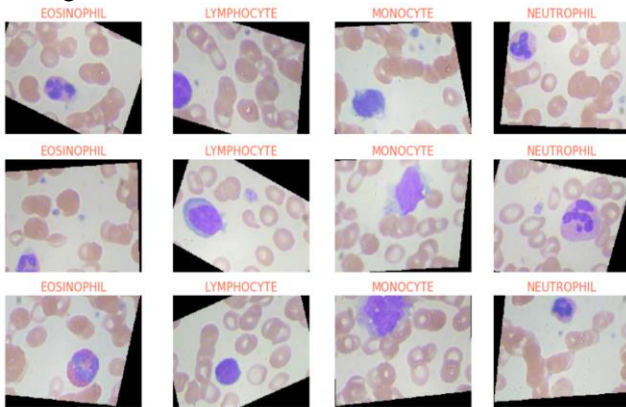


Figure 1. The sample cell types in dataset (4 different cell types are Eosinophil, Lymphocyte, Monocyte, and Neutrophil).

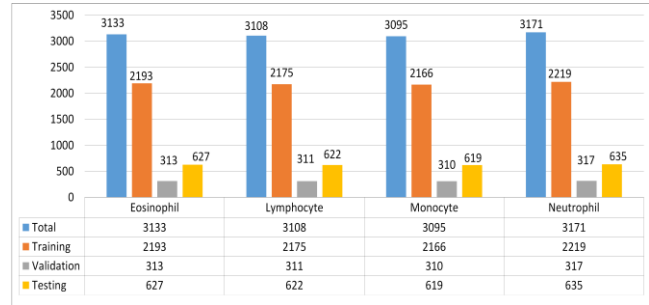


Figure 2. Categorical distribution of the blood cells in the dataset.

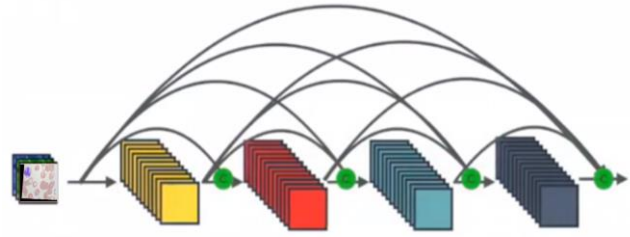


Figure 3. Fundamental architecture of one dense block in DenseNet [23].

B. Method

Deep learning has emerged as an exciting trend in machine learning in recent years. Deep learning is an extension of classical machine learning. This extension is defined as the use of functions that add more depth to the model and allow data to be represented in several hierarchical ways. The high hierarchical structure and wide learning capacity of deep learning models allow classification and predictions to be made with consistency. Deep neural networks are the multi-layered and multi-neuron version of Neural Network. The most important feature is automatic feature extraction in accordance with the problem and these features are obtained by the learning of the network. CNN architecture stands out with its popularity among various algorithms of deep learning. Inspired by the visual cortex of the brain, CNN is mainly used to classify images, signals, cluster similarities and perform object recognition [24], [25]. This architecture consists of convolution layers, pooling layers, nonlinear layers, and fully connected layers by processing the image with various layers. The convolution layer is the layer that the most intensive mathematical processing takes place, and some filters are applied to the image to detect low and high level features in the image. In the nonlinear layer, which is also called the activation layer, nonlinearity is introduced to the system and the ReLu function is used because it usually gives the best results.

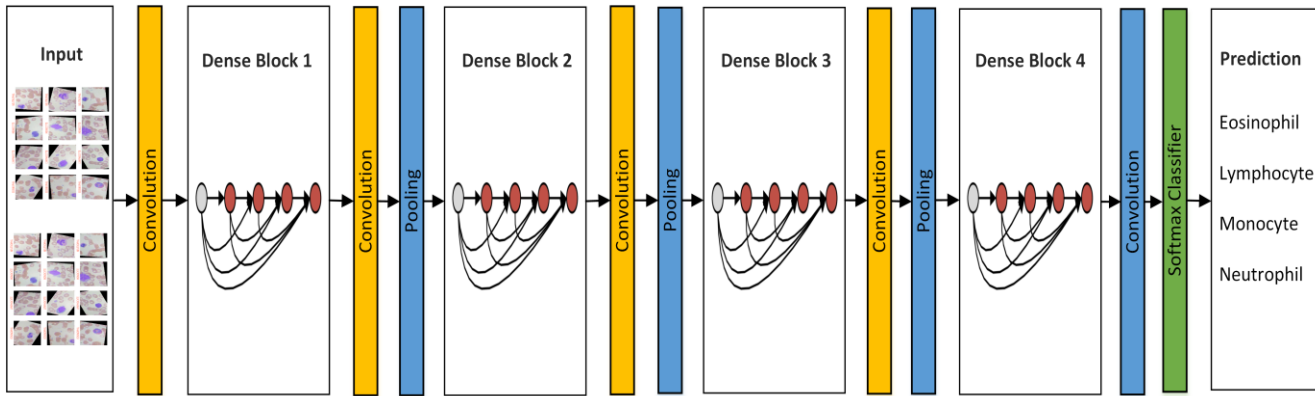


Figure. 4. The workflow of DenseNet121 architecture used in this study to classify blood cells from blood cell images.

The pooling layer reduces the number of weights on the system by reducing the size of the feature maps in order to reduce the parameters in the network. Fully connected layers act as a classifier in the model by establishing a connection between the input and output layers [26].

Input image goes along multiple convolutions and acquire high-level features with standard CNN (ConvNet). In Dense Convolutional Network (DenseNet), each layer acquires additional inputs from all previous layers and transfers on its own feature-maps to all next layers as shown in Fig. 3 [23]. DenseNet is a dense-connected convolutional network that forward connects each layer to the other layers. DenseNet gives superior performance for various computer vision problems. DenseNet reduces the gradient descent problem, increases feature reuse, and reduces parameter usage. The advantage of DenseNet architectures is to allow feature propagation and feature reuse. Thus it minimizes the number of parameters. The main structure block of DenseNet121 is the Dense Block. These Dense Blocks consist of convolution layers. DenseNet121 consists of 121 connected convolutional layers with a fully-connected layer of 1000 units as the final output layer [27]. CNN architectures are generally hierarchical, so feature maps of $l-1$ layer is given as an input to the l .layer. Feed-forward networks just like convolutional networks connect the output of the l .layer as input to the $(l+1)$.layer, that gives source to the sequent layer transition: $X_l = H_l(X_{l-1})$ which l indexes the layer. H_l is a composite function and that contains some of operations as batch normalization, ReLU, pooling, or Convolution. On the other hand in DenseNet, the feature maps of all previous layers are combined and used as input for any specific layer. In other words, DenseNet is a connectivity model that further improves

the flow of information between layers. Direct connections from any layer to all following layers are performed. In addition, its own feature maps are used as input for all subsequent layers. Therefore, the feature maps of all previous X_0, X_1, \dots, X_{l-1} layers are combined and used as input for layer X_l . As a result, the l .layer receives the feature-maps of all preceding layers, X_0, X_1, \dots, X_{l-1} , as input eq. (1) where $[X_0, X_1, \dots, X_{l-1}]$ indicates the integration of the feature-maps produced in layers $0, \dots, l-1$. For ease of fulfillment, the multiple inputs of $H_l(\cdot)$ in eq. (1) are concatenated into a single tensor [27], [28].

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (1)$$

As shown in Fig. 4, DenseNet121 has four Dense blocks and a transition layer between each Dense blocks. Any Dense block contains some convolution layers, and any transition layer contains a batch normalization, a pooling and convolution layer. Activation function (ReLU) is used in DenseNet to increase nonlinearity. In this model, global average pooling layer which is the last layer of DenseNet121, generates features from the input image. Then, obtained features are operated by the classifier to perform the final prediction. Thus, a fully-connected layer is followed by a softmax classifier that uses softmax activation function. A fully connected three-unit layer is used for classification (four classes for blood cells). The output of the fully connected layer is normalized and probability distribution over the predicted output classes is created by softmax activation. This particular model improves the pass of information over the network and decreases the vanishing gradient problem. In addition, DenseNet improves feature reuse and parameter efficiency and provides combined information of the network to each layer [29]. The most significant reason for choosing is

that dense connectivity has a regularized effect and reducing overfitting at training.

IV. EXPERIMENTAL RESULTS

The experimental studies in this study are conducted on a desktop computer with Intel Core (i7) 8700U CPU @ 3.20 GHz, NVIDIA 4 GB GeForce-GTX 1050 Ti graphics and 16GB primary memory. The software was implemented using Python 3.7 and its related libraries. The data in the experimental study: 70% is reserved for training, 10% for validation, and 20% for testing. The input image resolution of the model is (240x320x3). DenseNet121 model is built with batch size (mini-batch) 128 and with 100 epochs at training. The total number of parameters is 7,628,484 for this model. At this stage, the performance evaluation of the method is performed over four performance criteria as Precision, Recall, F1-Score, and Accuracy. These values are calculated over confusion matrix (as shown in Fig. 5) for each class (Eosinophil, Lymphocyte, Monocyte, and Neutrophil). Performance parameters of the experiment are given in Table I. The calculated scores about Precision, Recall, F1-Score for each class and average results are given in this table. According to this experimental study, the accuracy value obtained with the proposed DenseNet121 technique is 94%.

Table I.
Performance parameters of the DenseNet121 model

Accuracy: 94.00%	Precision	Recall	F1-Score
Eosinophil	0.9535	0.8501	0.8988
Lymphocyte	0.9851	0.9598	0.9723
Monocyte	0.9919	0.9855	0.9887
Neutrophil	0.8382	0.9543	0.8925
Average:	0.9422	0.9374	0.9381

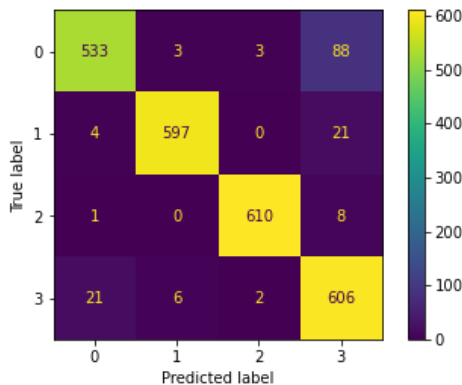


Figure 5. Confusion matrix of DenseNet121 model.

As shown in Table II, transfer learning and CNN-based different popular techniques as Xception,

VGG19, and EfficientNetB1 have been compared with the proposed DenseNet121 technique. It was observed that the DenseNet121 model in this study gives superior results than other models.

Table II.
Comparison of other deep learning techniques with DenseNet121 model

Model	Precision	Recall	F1-Score	Accuracy (%)
Xception	0.7897	0.7869	0.7844	78.59
VGG19	0.8490	0.8478	0.8466	84.70
EfficientNetB1	0.8892	0.8680	0.8668	86.72
DenseNet121	0.9422	0.9374	0.9381	94.00

In Fig. 6, training/test accuracy and training/test loss graphs for 100 iterations of the DenseNet121 model are presented. Both training and test accuracy curves show a rising curve as the number of iterations increases. Moreover, decrease in error rate is shown from loss curve. It shows that training and learning process of the network are at a good learning rate. As shown in Fig. 6, while the loss value decreases in each iteration, the accuracy rate increases with the given training set and learning occurs.

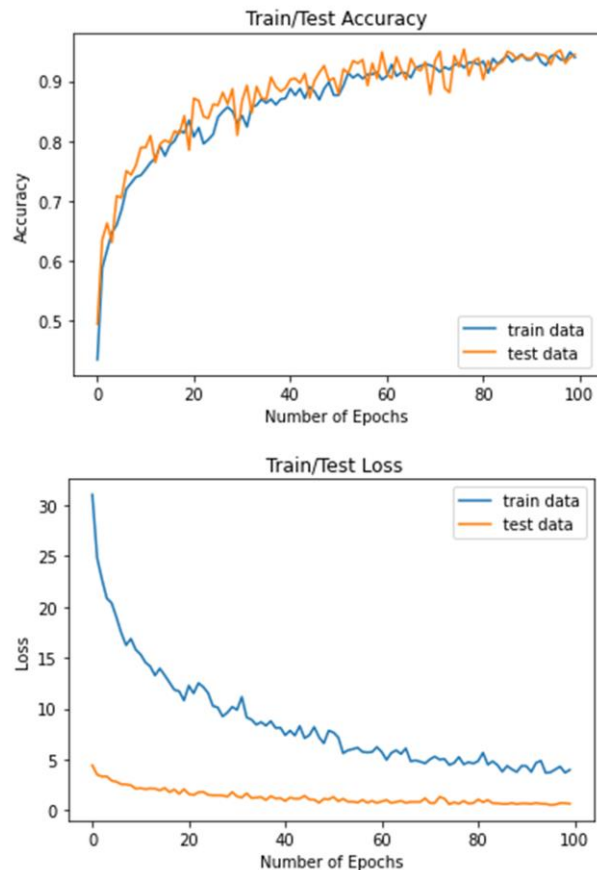


Figure 6. Training/test accuracy and training/test loss graphs of the DenseNet121 model.



Table III.
 Performance comparison of this study with other studies using same BCCD dataset

Study	Year	Method	Accuracy (%)
Liang et al. [8]	2018	CNN + RNN	90.79
Bani-Hani et al. [9]	2018	CNN + Genetic Algorithm	91.00
Şengür et al. [10]	2019	LSTM	92.89
Banik et al. [11]	2019	Fused CNN	90.79
Patil et al. [12]	2020	CNN + RNN	95.48
Mohamed et al. [13]	2020	DenseNet-169	92.00
Dekhil [14]	2020	CNN with extracted ROI	92.50
Ghosh & Kundu [15]	2021	CNN + RNN	87.29
Ekiz et al. [16]	2021	Con-SVM	85.96
This study	2021	DenseNet121	94.00

Performance comparison of the current study with other studies using same BCCD dataset is given as shown in Table 3. In this table, the state-of art studies using same BCCD dataset are discussed. At this stage, we studied the experimental results which divide randomly whole data as training and testing in literature (like this study 70% training, 10% validation, and 20% testing). DenseNet121 model in this study gives the reasonable accuracy score of 94% for BCCD dataset when compared to other techniques. For instance, a similar study by Mohamed et al. [13] about DenseNet-169 was conducted on the BCCD dataset (for approximately 80% training and 20% testing). They achieved the best feature extractor pre-trained model by DenseNet-169 with accuracy rate of 92.00%.

V. CONCLUSION

White blood cells are one of the immune systems that play a role in protecting the body from infectious diseases and foreign invaders. There are different categories of white blood cells, and each category may indicate a disorder of the body. Currently, WBC diagnosis is usually studied manually by a doctor. This process is prone to a lot of time, costs and errors compared to automatic computerization. Therefore, an automated classification technique for WBC images is proposed in this study. In this paper, a dense convolutional neural network model is studied. Parameters and layers of this architecture are specially adjusted to get high performance from this model.

Experimental studies are conducted on an open-access BCCD dataset. In this dataset, there are a total of 12507 images belonging to four classes: Eosinophil, Lymphocyte, Monocyte, and Neutrophil. For this purpose, a DenseNet121 model is applied, which is able to classify blood cell images. The data in the

experimental study: 70% is reserved for training, 10% for validation, and 20% for testing. Performance evaluations are performed on the accuracy of the techniques, and the results are also compared with state-of-the-art (popular) deep learning-based approaches like Xception, VGG19, EfficientNetB1. In experimental studies, the highest accuracy (94%) is obtained with the proposed DenseNet121 model.

According to experimental results, both training and test accuracy curves show a rising curve as the number of iterations increases. Thus, decrease in error rate is shown from loss curve. It shows that training and learning process of the network are at a good learning rate. Considering the high accuracy value obtained with this model, automatic detection of which class the cells belong to will speed up the diagnosis and allow more data to be examined by doctors. In further studies, other deep learning-based techniques and new dataset include abnormal white blood cells (recognizing other kinds of cell types) will be studied and comparisons will be made to complete the diagnostic system. We can also work on making the system more stable, reducing noise and improving overall performance with pre-processing and post-processing stages.

CONFLICTS OF INTEREST

There is no conflict of interest.

RESEARCH AND PUBLICATION ETHICS

The author declare that this article does not require ethics committee approval or any special permission.

REFERENCES

- [1] W. Stock, R. Hoffman, "White blood cells 1: non-malignant disorders," *The Lancet* 355 (2000) 1351–1357, [https://doi.org/10.1016/S0140-6736\(00\)02125-5](https://doi.org/10.1016/S0140-6736(00)02125-5).
- [2] B. Medical, "Medical gallery of blausen medical 2014," *WikiJournal of Medicine*, vol. 1, no. 2, pp. 1–79, 2014.
- [3] Q. Wang, J. Wang, M. Zhou, Q. Li, Y. Wen, and J. Chu, "A 3D attention networks for classification of white blood cells from microscopy hyperspectral images," *Optics & Laser Technology*, vol. 139, Article ID 106931, 2021
- [4] M.S. Blumenreich, "The White Blood Cell and Differential Count," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, H. K. Walker, W. D. Hall, and J. W. Hurst, Eds., ed Boston: Butterworths Copyright © 1990, Butterworth Publishers, a division of Reed Publishing., 1990.
- [5] B.J. Bain, "Diagnosis from the blood smear," *N Engl J Med* 353 (2005) 498–507.



- [6] S. Khan, M. Sajjad, T. Hussain, A. Ullah, and A. S. Imran, "A Review on Traditional Machine Learning and Deep Learning Models for WBCs Classification in Blood Smear Images," *IEEE Access*, vol. 9, Article ID 10657, 2020.
- [7] P. Mooney, "Blood cell images," <https://www.kaggle.com/paultimothymooney/blood-cells>, Accessed 22/10/2020.
- [8] G. Liang, H. Hong, W. Xie, & L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," *IEEE Access*, 6, 36188-36197, 2018.
- [9] D. Bani-Hani, N. Khan, F. Alsultan, S. Karanjkar, & N. Nagarur, "Classification of leucocytes using convolutional neural network optimized through genetic algorithm," In Proc. of the 7th Annual World Conference of the Society for Industrial and Systems Engineering, 2018.
- [10] A. Şengür, Y. Akbulut, Ü. Budak, & Z. Cömert, "White blood cell classification based on shape and deep features," In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1-4, IEEE, 2019.
- [11] P.P. Banik, R. Saha, K. Kim, "Fused Convolutional Neural Network for White Blood Cell Image Classification," in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), 2019, pp. 238-240. Doi: 10.1109/ICAIIIC.2019.8669049.
- [12] A. M. Patil, M. D. Patil, & G. K. Birajdar, "White blood cells image classification using deep learning with canonical correlation analysis," *IRBM*, 2020.
- [13] E. H Mohamed, W. H El-Behaidy, G. Khoriba, & J. Li, "Improved White Blood Cells Classification based on Pre-trained Deep Learning Models," *Journal of Communications Software and Systems*, 16(1), 37-45, 2020.
- [14] O. Dekhil, "Computational techniques in medical image analysis application for white blood cells classification," *Electronic Theses and Dissertations*. Paper 3424, 2020.
- [15] I. Ghosh & S. Kundu, "Combining Neural Network Models for Blood Cell Classification," arXiv preprint arXiv:2101.03604, 2021.
- [16] A. Ekiz, K. Kaplan, & H. M. Ertunç, "Classification of White Blood Cells Using CNN and Con-SVM," In 2021 29th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE, 2021.
- [17] X. Li, W. Li, X. Xu, and W. Hu, "Cell classification using convolutional neural networks in medical hyperspectral imagery," 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2017.
- [18] W. Yu, J. Chang, C. Yang, L. Zhang, H. Shen, Y. Xia, and J. Sha, "Automatic classification of leukocytes using deep neural network," *IEEE 12th International Conference on ASIC (ASICON)*, Guiyang, China, 2017.
- [19] M. Jiang, L. Cheng, F. Qin, L. Du, and M. Zhang, "White blood cells classification with deep convolutional neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, 32 (9): 1857006, 2018.
- [20] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, & M. K. Khan, "Medical image analysis using convolutional neural networks: a review," *Journal of medical systems*, 42(11), 1-13, 2018.
- [21] A.I. Shahin, T. Guo, K. M. Amin, and A. A. Sharawi, "White blood cells identification system based on convolutional deep neural learning networks," *Computer Methods and Programs in Biomedicine*, 168:69-80, 2019.
- [22] K. Throngnumchai, P. Lomvisai, A. Tantasirin and P. Phasukkit, "Classification of White blood cell using Deep Convolutional Neural Network," *Biomedical Engineering International Conference (BMEiCON)*, 12:1-4, 2019.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708, 2017.
- [24] S. Dargan, M. Kumar, M. R. Ayyagari, & G. Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, 27(4), 1071-1092, 2020.
- [25] E. Uzundurukan & A. Kara, "Deep Learning Based Threat Classification for Fiber Optic Distributed Acoustic Sensing Using SNR Dependent Data Generation," *Journal of Scientific, Technology and Engineering Research*, 1(2), 4-12, 2020.
- [26] P. Kim, "Convolutional neural network. In MATLAB deep learning," pp. 121-147, Apress, Berkeley, CA, 2017.
- [27] L. Sarker, M. Islam, T. Hannan, and A. Zakaria, "COVID-DenseNet: A Deep Learning Architecture to Detect COVID-19 from Chest Radiology Images," Preprints, 2020.
- [28] S. Kumar, S. Mishra, and S. K. Singh, "Deep Transfer Learning-based COVID-19 prediction using Chest X-rays," medRxiv, 2020.
- [29] F. Bozkurt and M. Yağanoğlu, "COVID-19 detection from chest X-Ray images using dense convolutional network," *International Symposium on Applied Sciences and Engineering (ISASE2021)*, Erzurum, Turkey, 2021.



Ferhat Bozkurt was born in Erzurum, Turkey in 1982. He received Computer Engineering degree from Dokuz Eylül University, İzmir, in 2007, the M.Sc. degree from Atatürk University, Erzurum, in 2012, and the Ph.D. degree from Karadeniz Technical University, Trabzon, in 2018. From 2007 to

2010, he was a software developer in a company. From 2011 to 2018, he was a research assistant at department of computer engineering, Ataturk University. Since 2019, he has been an Assistant Professor with the department of computer engineering, Ataturk University. His research interests include medical image processing, computer vision, artificial intelligence, machine learning, and data science.

RESEARCH ARTICLE

Infusing Energy Efficient Illumination Design to retrofit existing infrastructures – A case of energy-efficient illumination design of Multipurpose Hall at Jigme Namgyel Engineering College

*Hemlal Bhattarai¹, Abhishek Pokhrel², Sarda Chuwan³, Abiskar Chhetri⁴, Chimi Dem⁵

¹Centre for Lighting and Energy Efficiency Studies (CLEES), Jigme Namgyel Engineering College, Bhutan
b.hemlal@gmail.com 0000-0002-6512-1685

^{2,3,4}BEPE, Jigme Namgyel Engineering College, Bhutan

⁵Faculty, Department of Electrical Engineering, Jigme Namgyel Engineering College, Bhutan

HIGHLIGHTS

- *Illumination requirements need to be considered*
- *A clear understanding of illumination requirements of workspace is vital always*
- *Incorporating mixed design for enhanced comfort as well as productivity*
- *Multiple options need to be explored so as to retrofit.*

Keywords:

- Illumination design
- Energy Efficient Lighting
- Natural Lighting
- Problem Based Learning
- Multipurpose Hall

GRAPHICAL ABSTRACT

As illumination design is crucial for visual requirements and comfort as well as enhancing productivity, its consideration needs to be critical. Due to technological changes as well as the availability of energy-efficient luminaires in the market, the designers and implementers need to take advantage of these and make needful incorporation in new infrastructures or retrofits of the existing infrastructures. The energy consumption from lighting is quite significant and efforts have been in a switch to more energy-efficient luminaires. On top of this, the education campus is one significant contributor to maximum energy consumptions from lighting. As a result, the design needs to holistically look into illumination design based on a standard to meet the illumination requirements of a given workplace keeping in mind of possible incorporation of natural light and energy-efficient luminaires. Substantial illumination especially can be derived from natural light which was not seriously taken into consideration in old infrastructures thus rectifying and retrofits needs to explore best of the best options.

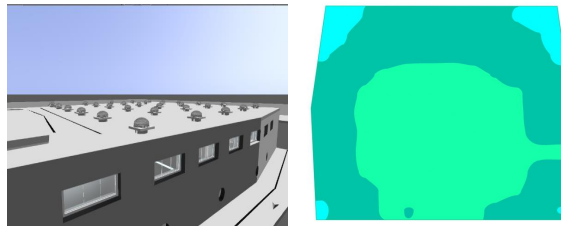


Figure A. a) MPH with 30 light tunnels arrangement, b) Light distribution on auditoria

Article Info:

Received: 13 December 2021
Accepted: 18 December 2021
Published: 21 December 2021

DOI: 10.53525/jster.1035864

*Correspondence:

Hemlal Bhattarai
b.hemlal@gmail.com

Aim of Article: *This research is to understand the existing lighting conditions of the Multipurpose Hall (MPH) of Jigme Namgyel Engineering College, and come up with the findings for the measures that can be derived for interventions.*

Theory and Methodology: *The study incorporated 'Problem Based Learning (PBL)' methodology from problem scanning, problem framing, data collections and analysis along with possible recommendations. The use of primary as well as secondary data in this research which is qualitative in nature is the base for overall research. Also 'Post Occupancy Evaluation (POE)' approaches are incorporated in data collections so that realistic understandings and necessary measures can be taken in the study.*

Findings and Results: *It is evident from the study that the current lighting conditions of MPH under the study are having significance lacking in its illumination requirements. There are immediate needs felt for interventions so as to derive the maximum benefits of the infrastructures without compromising the comfort of the users. The study also explored that there are multiple options that can be incorporated for the needful.*

Conclusion: *Existing infrastructures can be retrofits using multiple options so that visual comforts, as well as the productivity of users, can be enhanced from usages of given working spaces.*



RESEARCH ARTICLE

Infusing Energy Efficient Illumination Design to retrofit existing infrastructures – A case of efficient illumination design of Multipurpose Hall at Jigme Namgyel Engineering College

*Hemlal Bhattarai¹, Abhishek Pokhrel², Sarda Chuwan³, Abiskar Chhetri⁴, Chimi Dem⁵

¹Centre for Lighting and Energy Efficiency Studies (CLEES), Jigme Namgyel Engineering College, Bhutan
b.hemlal@gmail.com 0000-0002-6512-1685

^{2,3,4}BEPE, Jigme Namgyel Engineering College, Bhutan

⁵Faculty, Department of Electrical Engineering, Jigme Namgyel Engineering College, Bhutan

Citation:

Bhattarai, H., Pokhrel, A., Chuwan, S., Chhetri, A., Dem, C., (2021). *The Problems Encountered in the Process of Publishing Articles in the Journal and Investigation of the Solutions*, Journal of Scientific Technology and Engineering Research, 2(2):89-100. DOI: 10.53525/jster.1035864

HIGHLIGHTS

- Illumination requirements need to be considered for the given workplace for visual comfort and productivity
- A clear understanding of illumination requirements of workspace vital always for design and retrofits
- Incorporating mixed design for enhanced visual comfort as well as productivity
- Multiple options need to be explored for design and retrofit for better implementations.

Article Info

Received: 13 December 2021
Accepted: 18 December 2021
Published: 21 December 2021

DOI: 10.53525/jster.1035864

*Corresponding Author:

Hemlal Bhattarai
b.hemlal@gmail.com
Phone: +97517623426

ABSTRACT

Illumination design is a very important part of any infrastructure. The level of illumination for various workspace requirements are different and it is crucial to calculate the required level of illumination for new as well as existing infrastructures. Furthermore, the usage of natural light, as well as energy-efficient luminaries, are considered to be the preferred choices in illumination design. In this study, the illumination design of one of the most frequent infrastructures (i.e. Multipurpose Hall (MPH)) has been studied incorporating the 'Problem Based Learning (PBL)' approaches where actions on problem scanning, working on the problem and coming up with multiple solutions to counter the realistic problem were looked into. The study which is more qualitative in nature is backed by primary as well as secondary data to study the problems and frame multiple options of recommendation for retrofits that can be incorporated in enhancing the illumination level through energy-efficient illumination design as well as integration of natural light where possible.

Keywords: Illumination design, Energy Efficient Lighting, Natural Lighting, Problem Based Learning, Multipurpose Hall

I. INTRODUCTION

Illumination of working space is crucial and achieving the desired level of illumination on the working plane is the most sought aspect in illumination design. In general, the amount of illuminance received on the working plane has a significant impact on productivity. The optimum illumination is met when the right

luminary is placed properly. However, in implementing the design, often, the luminaries chosen while designing are not locally available and/or are too impractical because of its cost and many a time it is ignored to be implemented as designed. Every infrastructure and its amenities will have various levels of illumination level as recommended. A target for the lighting system to become efficient is also a



consideration of daylight along with artificial lighting. The idea behind energy-efficient architecture is to optimize the lighting system by incorporating natural and artificial illumination.

Furthermore, more than any other type of building, daylighting is a critical consideration in the design of educational institutions [1]. Researches working on the field of indoor illumination highlighted that ensuring lighting quality in an educational environment is a rather tedious task particularly due to glare calculation. Glare calculations are an important aspect in terms of visual comfort but are usually avoided in most lighting designs due to their complexity. The lighting performances of different buildings under different conditions, activities on given amenities involving different methods were carried out to understand the lighting distribution. Those researches were carried out also to find how the position and the alignment of a structure, use of different luminaires affect the amount of light received on the working plane. However, not much study was done to the improving the lighting system in any structure by considering the daylight [2]. A study was done on efficient illumination design of some of the most energy-consuming buildings in an educational institute and concluded that the choice of lamp can greatly affect energy consumption for the same illumination [3].

Some critical studies on lighting performances done at various places are;

- The natural lighting for energy is saving and visual comfort in collective housing: A case study in the Algerian building context. In this research, much focus was given to the visual comfort of the occupants. Post Occupancy Evaluation (POE) was employed [4].
- Assessment of natural lighting performance and visual comfort of educational architecture in southern Europe. The case of typical educational school premises in Cyprus. This paper is like the earlier paper. However, it acknowledges the fact that daylighting is more important in an educational facility rather than artificial lighting [5].

A study of Analysis of natural light distribution in Phinisi tower was carried out mainly to study how the position, floor height and window openings of a building affect the distribution of natural light. This research has not used any software instead they have

manually measured the illumination level for different rooms in the building for the different hours of a day for three days. It was found out that the level of daylight depends on the sky condition and it has affected the light distribution into the building. The window opening and the height of the building after the illuminance of a room. And it was found out that the distribution of natural light increases during the day if the sky condition does not change [6]. The influence of natural light on the design of electrical lighting-Taking Liaohe Art Museum as an example. It was concluded that the integration of natural light and artificial light brings better psychological effects. The European Standard EN12665:2011 defines visual comfort as a subjective state of visual well-being caused by the visual environment [7]. Visual comfort depends on (i) the physiology of the human eye, (ii) the physical quantities describing the amount of light and its distribution in space, and (iii) the spectral emission of the light source. It has been commonly studied through the assessment of a series of factors regulating the relationship between the human needs and the light environment, such as (i) the amount of light, (ii) the uniformity of light, (iii) the quality of light, and (iv) the occurrence of glare [8]. According to the current regulations and specifically the internal house wiring standards-safety specifications of Bhutan [9], the recommended values of illumination in lux for relevant areas are as tabled.

Table 1. Required illumination level for various amenities

Area	Illumination in lux
Assembly hall in educational centers	150
Corridor	70
Stairs	100
Stage (Displays)	200-300
Printing press	300

Unified Glare Rating (UGR) is an indicator of direct glare from the luminaries of a lighting system. A UGR of 10 is considered as no glare and a UGR of 30 is considered as unacceptable discomfort glare [10]. Therefore, a UGR of 19-22 can be considered acceptable as achieving no glare is practically impossible.



One indicator of daylight performance is the Daylight Factor (DF). This is the most widely used index which defines the ratio of interior luminance on a horizontal surface to the exterior illuminance on a horizontal surface under an overcast sky. According to BREEAM standards, the minimum DF should be at least 2% for 80% of the space. A Multipurpose Halls (MPH) in the educational campus is the most frequent usage of amenities and is one of the high energy consumers due to numerous activities taking place in it. Researchers pointed out that MPHs in institutes contribute almost 20% to 30% of the total energy consumption where the share of lighting is significant [11].

II. STUDY UNDER CONSIDERATION

A. Study site and the solar context of the location

In understating the illumination design role of natural light plays a significant role. Hence it is crucial to understand the cloud coverage of the location. The cloud coverage of the region as referred from 'Weather Spark' [12] is shown below in figure 1 where their average of 2013-2021 is being reported.

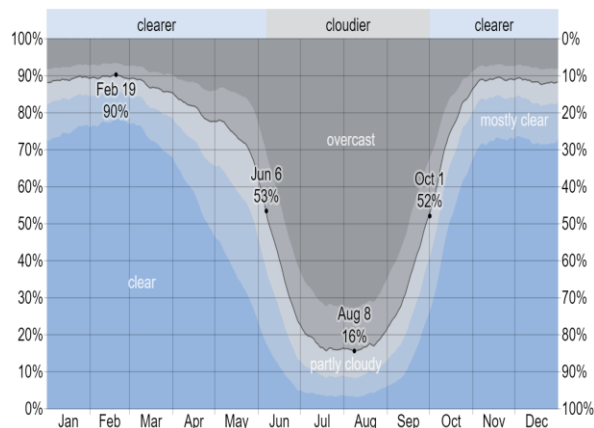


Figure 1. Cloud coverage of the region (Samdrup Jongkhar) [12]

From the figure above, it is evident that;

- The average percentage of the sky covered by clouds experiences extreme seasonal variation over the course of the year.
- The clearer part of the year begins around September 30 and lasts for 8.2 months, ending around June 5. On February 19, the clearest day of the year, the sky is clear, mostly clear,

or partly cloudy 90% of the time, and overcast or mostly cloudy 10% of the time.

- The cloudier part of the year begins around June 5 and lasts for 3.8 months, ending around September 30. On August 7, the cloudiest day of the year, the sky is overcast or mostly cloudy 84% of the time and clear, mostly clear, or partly cloudy 16% of the time.
- The length of the day varies over the course of the year. In 2020, the shortest day is December 21, with 10 hours, 27 minutes of daylight; the longest day is June 21, with 13 hours, 50 minutes of daylight.

B. Study site (Multipurpose Hall of the College)

The MPH understudy of this study consists of three construction moduli. As depicted in Figure 2 below, the three construction moduli as follow:

- Auditoria: an irregular hexagon with an area of 473 m² and a height of 7 meters.
- Stage of area 71 m² raised 0.7 meters from the auditoria area.
- Two galleries at the rear end on the upper floor and various other rooms.

For the modeling of the study site, the use of SketchUp was made so that the layout can be easily understood as well as can effectively used for multiple visual analyses. Furthermore, this software is quite simple and user-friendly so that exploring and working on this software for given work becomes simpler. The 3D model was developed considering the geometry of the hall such that the model would depict the real structure at a high level of accuracy.

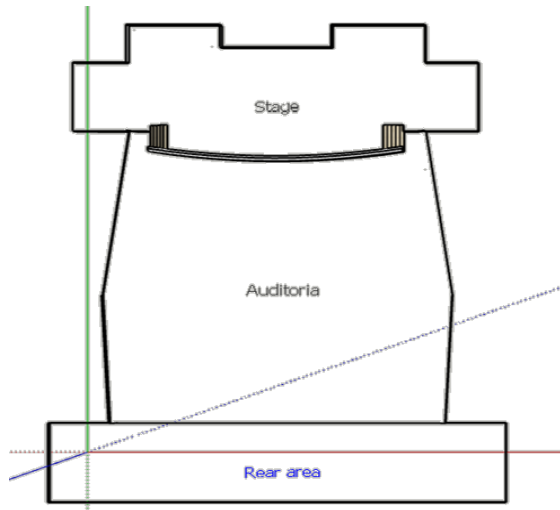


Figure 2. Layout diagram of existing MPH of the college

Understanding from Users Perspectives

POE starts with questionnaires and interviews. Interviews were conducted at different levels in a completely random manner to gather practical insights on the problems of the present lighting system. The survey allowed us to consider the stage as a different area with its illumination requirements. From the interviews with stakeholders such as the estate manager, engineer, technicians and the college administration, valuable data were gained, and it was learned that a renovation is being planned for the near future. The type of lamp that would be used in case of a renovation was also known, much as required by POE. At the end of the research, we would be able to give useful insight on the illumination to the estate manager of JNEC for the upcoming renovation of the hall.

A survey was carried out within a random sample of 105 students. Of the total number of participants, 80.9% of the participants were from the higher class of study who have spent more time on the campus. The same survey was distributed to a sample of 25 students from Cultural-Club members who were frequent users of the MPH, 14 teaching faculties of varying teaching experiences participated in interviews conducted in a random unofficial manner and it was found out that the lighting system in MPH needed some changes for better illumination of the hall.

At the end of the survey, we expect to identify more problems via the established dialogue and be able to work on the project as directed by the survey.

Physical on-site measurement

A Lux meter was used to carry out on-site measurements on 15/12/2020. The equipment used was a Konica Minolta T-10A illuminance meter, with a sensitivity of 0.01 lux to 299900 lux.

The measurements were taken at various points as shown in figure 3 and it was observed that the Illumination level was not met on even one point. The illumination was not uniform. These could be due to the many dysfunctional luminaries.

All required data for design and about luminaries were collected and tabulated.

Rough sketches were made, and measurements were taken using manually using measuring tape to aid in the designing in SketchUp. Measurements were done aiming at a high level of accuracy for the later computer design and simulation. This is one critical aspect of PBL incorporating the beneficiaries (users) being used from the initial stages for coming up with the best of the best understanding of the problems and working to its solution. Keeping the perspective of users for problems insight plays a major role in insightful knowledge and gathering practical situations which are worth for needful actions.

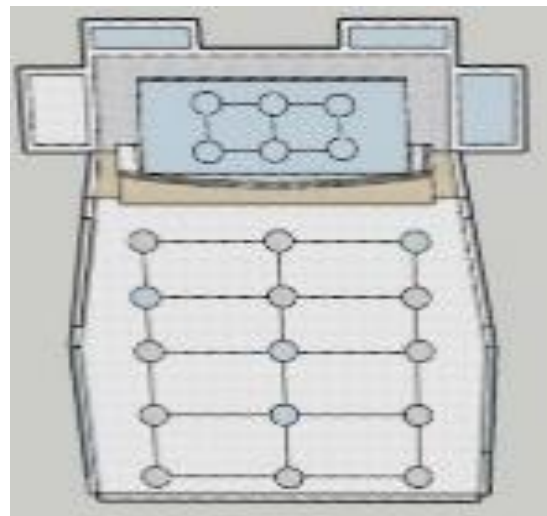


Figure 3. Reading point location on the site plan



The above figure 3 is a planned layout that has been derived taking into consideration the actual reading points for measurement of illumination for the overall analysis of the lighting falling on the working plane.

III. METHOD

The study is based on secondary literature from relevant sources along with primary data that are obtained through the survey, site visit and analysis. The study considered the Problem Based Learning (PBL) approaches in overall problem framing, studies and conclusion. The research is more qualitative in nature and uses ‘Post-Occupancy Evaluation (POE)’ to align for PBL where two crucial investigation techniques of measurement and questionnaire are based for the overall study. Usage of freely available design software ‘Sketchup’ and illumination design software ‘Dialux’ were incorporated in this study for overall design and analysis.

IV. RESULTS AND DISCUSSION

Analysis from the survey

Graphical representations of the survey are given below. Some observations are discussed below.

most of the participants were satisfied. However, 9 % of the participants felt the brightness in the MPH is not all acceptable. On the other hand, 11% of the participants were satisfied with the brightness level in the MPH.

Regarding the uniformity of illumination, 22 % of the participants felt that the illumination is non-uniform. While 9 % were fairly satisfied with the uniformity of illumination. Generally, the majority of the participants were not happy with the uniformity level. As shown in the visual comfort satisfaction chart, it can be concluded that the majority of the participants experienced the glare in the hall

The survey included 3 questions. The participants were asked to rate the brightness, uniformity of light and visual comfort of the MPH on a scale of 0-10. Finally, suggestions were asked to improve the lighting system in the future. The survey is quite critical as it reflected the practical situational analysis and the feedback obtained was more inclined to the need of enhancing the lighting system in MPH under study.

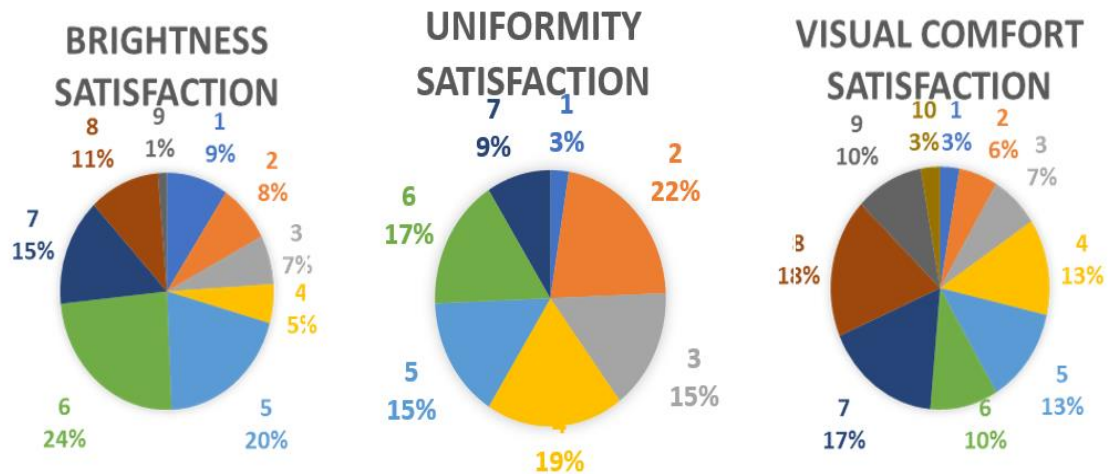


Figure 4. Survey results on the satisfaction level

Figure 4 is derived from the POE interventions three aspects of illuminations were surveyed. These aspects are brightness satisfaction, illumination uniformity satisfaction and visual comfort satisfaction. It is found as shown in figure 4 above that in terms of brightness,

Analysis using standard calculation methods:

Manual calculations were done as tabulated in Table 2 using the empirical formula;

$$Illumination\ in\ lux = \frac{Utilisation\ factor \times Maintenance\ factor \times \sum No.\ of\ lamps \times Lumen\ output}{Floor\ area\ in\ meter\ sq.}$$



Table 2. Energy Consumptions and Illumination level for present conditions and when all luminaires functioning.

Space	Lamp specification		Number of lamps		Total wattage (W)		Illumination (Lux)		Number of lamps required to meet the illumination level
	Lumen output (lm)	Wattage (W)	Working lamps	Existing fixtures	Wattage of working lamps	Wattage if all the existing fixtures had lamps	Illumination of working lamps	Illumination if all fixtures had lamps	
Auditoria	810	9	33	63	297	567	46	97	-
	5100	85	3	8	255	680			-
	1800	18	2	2	36	36			-
Stage	810	9	8	8	72	72	60	60	-
	2400	30	3	3	90	90			-
Back entrance	810	9	1	2	9	18	5	10	20
Left gallery	810	9	1	2	9	18	21.6	43.2	7
Right gallery	810	9	1	2	9	18	21.6	43.2	7
Corridor	810	9	0	3	0	27	0	43.2	7
Women's Public toilet	810	9	2	2	18	18	120	120	4
	1000	100	1	1	100	100			
Men's Public toilet	810	9	2	2	18	18	120	120	4
	1000	100	1	1	100	100			
E-printing	1800	18	1	2	18	36	28.2	56.4	5
Left corridor	810	9	4	4	36	36	40	40	10
Right corridor	810	9	4	4	36	36	40	40	10

It was concluded that the illumination level would not be met even if all the lamps were working. However, since only the lamps of one room can be considered using the formula, the effect of light from other areas does not get accounted for. This can be later confirmed using the simulated results.

Comparisons of lighting performance under present situations and when all lamp:

Lighting performances of the existing system and considering all luminaires functional without the consideration of natural light were simulated in Dialux Evo.

To understand the uniformity, the light distribution of the rooms was considered. Had all luminaires been functioning, the illumination level in the auditoria would be met but would be subjected to high glare levels. Light distribution of the rooms was considered. Had all luminaires been functioning, the illumination level in the auditoria would be met but would be subjected to high glare levels.

From the calculation as well as analysis, it is observed that in the auditorium as well as main hall and stage there are requirements for increasing the illumination level to meet the standard.



Table 3. Comparison of lighting performances of the present situation with that of all luminaries functioning.

Parameter	Area	Existing System	All luminaries functioning
Illumination	Auditoria	65.2	128
	Stage	60	64.1
	E-printing	39.3	79
	Rear lobby	8	16.1
	Left gallery	35.5	76.9
	Right Gallery	35.1	75.8
Light distribution	Auditoria		
	Stage		
	E-printing		
	Rear lobby		
	Left gallery		
	Right Gallery		
UGR	Auditoria	15.4	17.4
	Stage	14.3	13.7
	E-printing	24.3	22.6
	Rear lobby	23.2	23.1
	Left gallery	25	22.5
	Right Gallery	25.1	22.8

V. RECOMMENDATION

Multiple cases can be looked into to make holistic action in incorporating the energy-efficient retrofits in the MPH. A couple of cases are recommended as follow:

Case 1. Enhancing illumination through increased retrofits

The lighting distribution in Stage and Auditoria is shown as shown in figure 5 below.

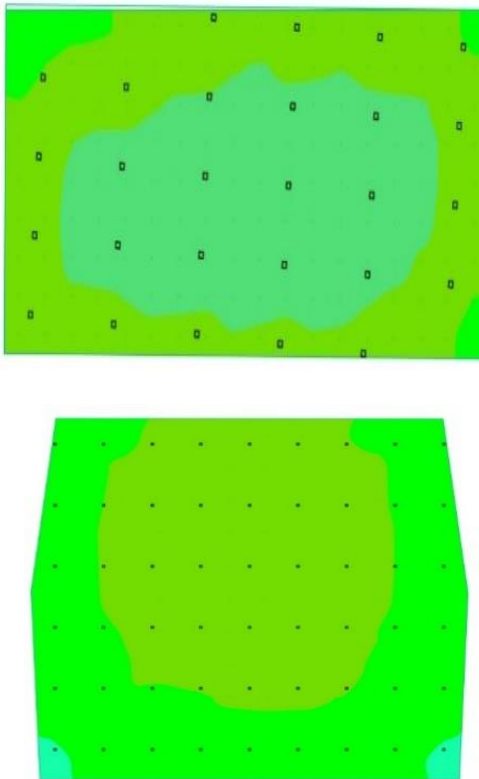


Figure 5. 5a. Lighting Distribution for Stage, 5b. Lighting distribution for auditoria

In an auditorium space, the addition of around 46 lux is required to meet the illumination level. This additional 46 lux could be met by increasing the lumen output by 52000 lumens. 5200 lumens can be achieved by any combination of lamps by adding sums of the product of lumen output and number of lamps.

For times when the hall functions as a badminton court, the illumination required would be met by increasing the lumen output by 34000 lumens. 34000 lumens can

be achieved by placing 8 floodlights of 4200 lumens right about the badminton area, connected in such a way that it is only turned on while playing badminton. For the stage under the present scenario, the illumination is short by 167 lux which can be overcome by increasing the lumen output by 25000 lumens. 25000 lumens can be achieved by using 6 floodlights of 4200 lumens.

Case 2. Enhancing illumination through changes of luminaries

Though the illumination level in the Auditoria, stage and when the auditoria functions as a badminton facility are met when the number of luminaires used is increased. However, it is seen that the glare requirement is not met either in an existing system or when the number of lamps is increased. Therefore, instead of replacing the existing luminaires with luminaries of the same type, if luminaries of 24 W and 1680 Lumen (as suggested by the estate manager) is to be used then Stage and Auditoria if uniformly illuminated and the glare level is within the permissible range.

Case 3. Enhancing illumination through the enhanced window for natural lighting

The existing system with its window size doubled gives the following result. The increasing window size can be an expensive renovation but the new design could have larger windows. Various research done on window size suggests that the windows must at least be 6 % of the floor area and must not exceed 25 %. Having a higher window to floor ratio increases the daylight factor. For a 3% daylight factor, a window to floor ratio (WFR) must be kept at 10 % [15]. The WFR for the MPH under the case study is only 3%.

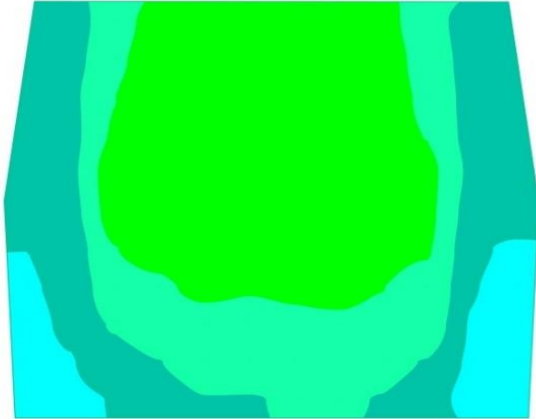


Figure 6. Lighting distribution in auditoria

It was observed from figure 6 above that an average of around 100 lux is met with this proposal. Thus, one-third of lamps can be turned on during the daytime.

Case 4. Enhancing illumination through the use of light domes OR translucent roofs OR light tunnels

The existing system with 30 light domes as shown installed above the auditoria, gives the light distribution as shown in Figure 7b.

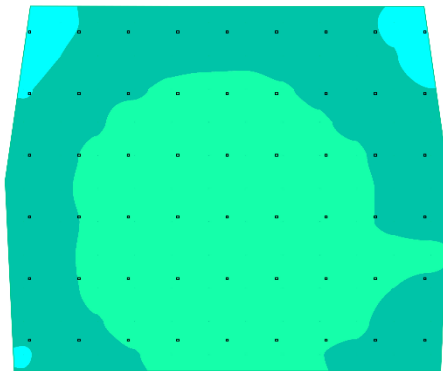


Figure 7. 7a. MPH with 30 domes arrangement, 7b. Light distribution on auditoria

The existing system with 10 translucent roofs installed above the auditoria, gives the following result of lighting distribution as shown in Figure 8b.

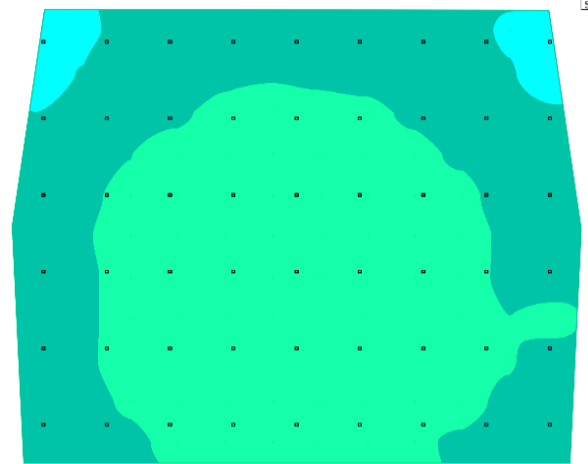
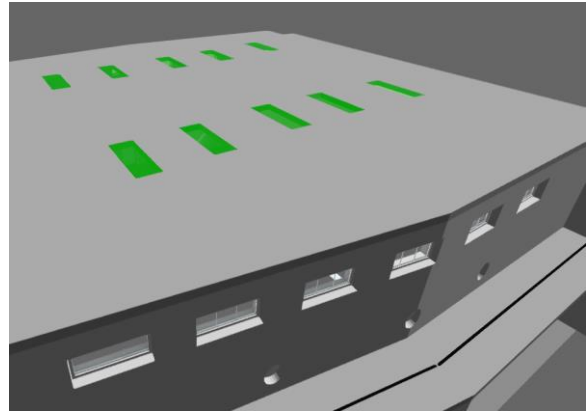


Figure 8. 8a. MPH with 10 translucent sheets arrangement, 8b. Light distribution on auditoria

As shown from the figure 8b above, it is possible to turn on just one-third of the lamps and the average of around 100 lux can be met from the natural lighting arrangements.

The existing system with 30 light tunnels installed above the auditoria, gives the following result of lighting distribution as shown in Figure 9b below. As shown from the figure 9b above, it is also possible to turn on just one-third of the lamps and the average of around 100 lux can be met from the natural lighting arrangements. Some of the essential illumination requirements can be met from the natural light in all of these three options which will result in minimal usages

of artificial light during daytime use of MPH auditoria as it can meet about 100 lux of illumination.

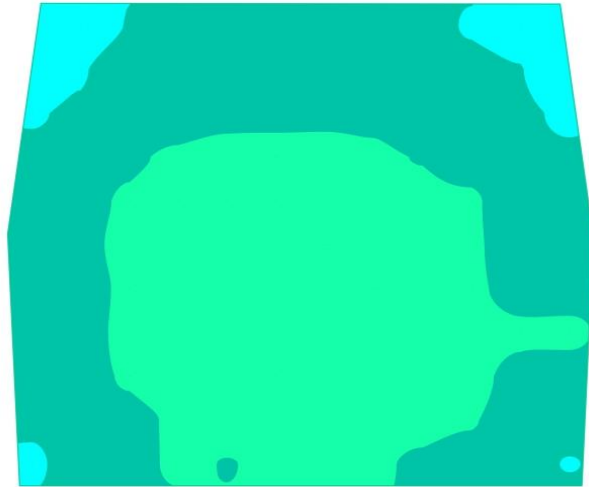
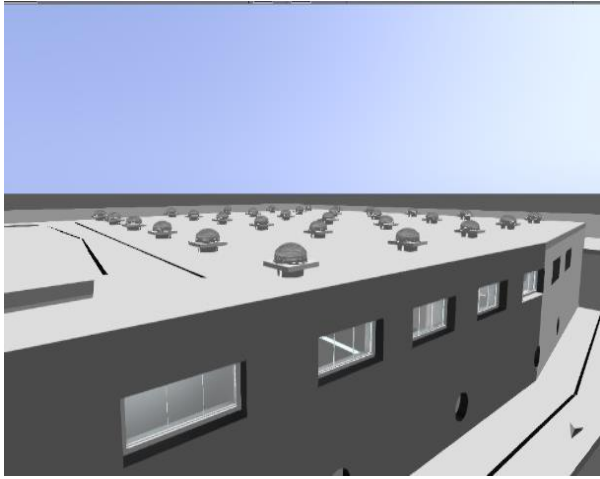


Figure 9. 9a. MPH with 30 light tunnels arrangement, 9b. Light distribution on auditoria

The combination of these approaches can still be more beneficial in trapping the maximum benefits of artificial light. Though such measures in the time of renovations are costly, those options are always beneficial in trapping the natural lighting for efficient illumination design. It needs to be noted that there should always be efficient illumination luminaires installed so that the required illuminations are not deprived during nighttime. These three options as stand-alone or in combinations are only for day usage lighting efficiency options through the utilization of natural light.

VI. CONCLUSION

Starting from the observations that were confirmed by surveying, the focus was given to the glare level, illumination level and the uniformity in the level of illumination in MPH. Firstly, the illumination level would be met if all the lighting fixtures currently present were to be made functional. However, the glare level would be beyond the permissible level. For this case, installing better lamps and using greener alternatives like using light tubes, domes, larger windows, and translucent sheets would be more effective in providing better illumination in the MPH. If a part of MPH is to function as a badminton facility, 8 flood lamps of 4200-lumen output can be used, turning it on as and when required. It was found out that 15 domes or 15 light tubes or 5 translucent sheets produce one-third of the required illumination, i.e., an equivalent of 50 lux. Domes and light tubes are more capital intensive. Translucent sheets are more efficient in terms of cost. A window to floor ratio of 6-25 % is recommended.

Finally, a PBL approach can be used as an efficient tool in terms of realizing the practical problem which can be incorporated with rigor in coming up with multiple options in countering the problems. In the case of efficient illumination design and incorporations, it is always crucial to study illumination fulfillment with the best of the best efficient luminaires available in the market along with tapping of the maximum potential of natural light.

CONFLICTS OF INTEREST

This research has no conflict of interest involved with any individual as well as institution. All the contributions are well referred and acknowledged the sources where necessary.

RESEARCH AND PUBLICATION ETHICS

All contributing authors, as well as contributors, are acknowledged and there is no ethical issues with plant or animals involved in this research work.

ACKNOWLEDGMENT

For this work, the authors would like to acknowledge CLEES for necessary support. Also, the authors would like to acknowledge PBL South Asia project co-funded by the European Union which is underway in the college where the



option of integrating PBL in teaching and learning is incorporated and the teams are able to benefit from it.

REFERENCES

[1] Costanzo, V., Evola, G., & Marletta, L. (2017). A review of Daylighting strategies in schools: State of the art and expected future trends. *Buildings*, 7(4), 41. doi.org/10.3390/buildings7020041

[2] Yangzom, K., Bhattarai, H., Choki, S., Choden, P., & Yangden, S. (2021). Energy efficient lighting design - A case study of jigme namgyel engineering college. *Journal of Applied Engineering, Technology and Management*, 1(1), 14-21. doi.org/10.54417/jaetm.v1i1.20

[3] Zhao, G. (2020). The influence of natural light on the design of electrical lighting -taking Liaohe art Museum as an example. *Journal of Physics: Conference Series*, 1601(2), 022009. doi.org/10.1088/1742-6596/1601/2/022009

[4] ITeh standards. (2011). iTeh Standards. <https://standards.iteh.ai/catalog/standards/c/en/d7c62c9a-95ac-4ed8-9a40-862805aa5afc/en-12665-2011>

[5] Preto, S. (2019). Dynamic facades: Optimization of natural light at workplaces. *Advances in Intelligent Systems and Computing*, 392-402. doi.org/10.1007/978-3-030-20151-7_37

[6] Jamala, N., Rahim, R., & Mulyadi, R. (2019). Analysis of natural light distribution in the building. *IOP Conference Series: Materials Science and Engineering*, 619(1), 012024. doi.org/10.1088/1757-899x/619/1/012024

[7] Schmal, C., Herzel, H., & Myung, J. (2020). Clocks in the wild: Entrainment to natural light. *Frontiers Physiology*, 11. doi.org/10.3389/fphys.2020.00272

[8] Balocco, C., Cecchi, M., & Volante, G. (2019). Natural lighting for sustainability of cultural heritage refurbishment. *Sustainability*, 11(18), 4842. doi.org/10.3390/su11184842

[9] MoWHS, RGoB. (2013). Bhutan Green Building Design Guidelines. Ministry of Work and Human Settlement. [https://www.mowhs.gov.bt/wp-](https://www.mowhs.gov.bt/wp-content/uploads/2014/05/Bhutan-GREEN-Building-Design-Guidelines-PDF-for-website-FI.pdf)

[content/uploads/2014/05/Bhutan-GREEN-Building-Design-Guidelines-PDF-for-website-FI.pdf](https://www.mowhs.gov.bt/wp-content/uploads/2014/05/Bhutan-GREEN-Building-Design-Guidelines-PDF-for-website-FI.pdf)

[10] Thompson, A. D. (2006). The worth of natural light. *Journal of Building Appraisal*, 2(1), 44-51. doi.org/10.1057/palgrave.jba.2940037

[11] BREEM.(2011). Hea 01, Visual Comfort. https://www.breeam.com/BREEAMUK2014SchemeDocument/content/05_health/hea01_nc.htm

[12] Michael, A., & Heracleous, C. (2017). Assessment of natural lighting performance and visual comfort of educational architecture in Southern Europe: The case of typical educational school premises in Cyprus. *Energy and Buildings*, 140, 443-457.



Bhattarai, H. was born in Bhutan, in 1976. He received the Diploma in Electrical Engineering from Royal Bhutan Polytechnic, B.E. Electrical Engineering from College of Science and Technology, both from Bhutan and M.Tech. in Power and Energy System from National Institute of Technology, Karnataka, India. He is working as faculty of the Department of Electrical Engineering since 2006 which is under Jigme Namgyel Engineering College, Royal University of Bhutan, Bhutan. Currently working as Dean Research & Industrial Linkages. Before working as Dean, he has served as Head of Electrical Engineering. He was also Project Manager of for “Promoting energy-efficient lighting in Nepal and Bhutan (ELNAB)” which was completed in Jan-2019 and also the Project Manager and Advisory Committee of ongoing projects on “Enhancing Problem Based Learning for South Asian Universities” both of which are funded by the Erasmus Project of European Union. He has taught in the Department of Electrical Engineering since 2006 and has mostly been associated with teaching mostly power systems, energy management & auditing, power electronics, renewable energy, high voltage engineering, research methodology and entrepreneurship. Hemlal research area of interests includes power engineering and energy systems with major focus on Renewable Energy and has been associated with many international conferences/publications including IEEE as program committee, keynote speaker and reviewer. He is currently the focal for Thematic Working Group (TWG) on Energy from Bhutan for ICIMOD.



RESEARCH ARTICLE

Analysis of Occupational Accidents in the Mining Industry Between 2012-2019 and Contribution of MCDM Methods

* Tuğçe ORAL

*Uskudar University, Faculty of Health Sciences, Occupational Health and Safety Department, İstanbul, Turkey
tugce.oral@uskudar.edu.tr orcid.0000-0003-1795-1550

HIGHLIGHTS

- Effect and important of this article in literature
- Exchange between sources in related subjects of this article
- Contribution and strongest impact on the related subject of this article
- Examined study and obtained results why is important

Keywords:

- Concrete Reinforcement,
- MCDM Method,
- Occupational Accident Frequency Rate,
- Mining Industry,

GRAPHICAL ABSTRACT

In this study, the mining sector, where occupational accidents are most common, has been examined. In the research, the occupational accident frequency rates for each year were calculated by evaluating the work accident data between the years 2012-2019, which were notified to the SGK and registered with the SSI. When we look at the results of the calculation of the accident frequency rate by years, it is seen that the employees who did not have a work accident in 2012, and those who were not likely to have an accident were 4%, 13% in 2013, 18% in 2014, 22% in 2015, 29% in 2016, 14% in 2017, It was found to be 19% in 2018 and 17% in 2019 (**Figure A**). Considering the current laws and sanctions, and based on the most up-to-date value, an answer has been sought to the question of how OHS practices in the mining industry can be further integrated. For this reason, in the second stage of the research, by making use of multi-criteria decision making methods; It has been emphasized that the necessity of evaluating the contribution to OHS as a criterion in the selection of machinery-equipment, materials products, or services to be used at every stage of the work in the sector. While evaluating alternatives in the use of concrete reinforcement as an exemplary application stage in mines, the OHS criteria were added in addition to the needs due to the conduct of the work, and the most ideal solution was suggested as a substitute method.

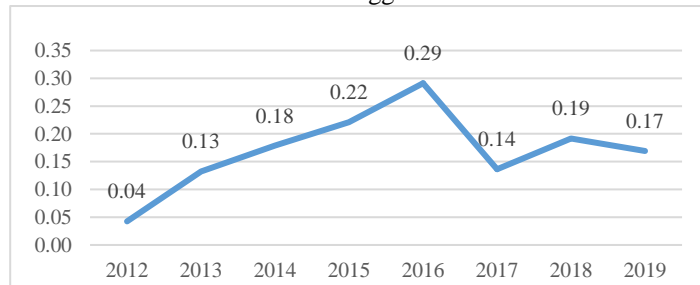


Figure A. Change of Occupational Accident Frequency Rates by Years

Article Info:

Received : 19.11.2021
 Accepted : 20.12.2021
 Published : 21.12.2021

DOI:

10.53525/jster.1026082

*Correspondence:

Tuğçe ORAL
tugce.oral@uskudar.edu.tr
 Tel: +90 216 400 22 22

Aim of Article: To reveal the necessity of deciding on the application method to be determined in cases where the situations that may cause work accidents cannot be eliminated at the source, by using MCDM methods.

Theory and Methodology: In the mining sector, where the probability of an occupational accident is still high, the ideal solution that can be considered both as an alternative product selection and as a substitute method has been calculated by using the BTOPSIS method, which is one of the MCDM methods, in order to offer remedial solutions.

Findings and Results: Considering the development rate of the sector in terms of employment and its contribution to the country's economy, it is observed that this value is remarkable. When the selection of concrete reinforcement, which is considered as an example, is compared; When the cost, engineering and time management contributions between MSF and GD are evaluated, MSF can be recommended as a substitute method in terms of OHS. MSF closeness coefficient value of 0.321 was found to be the most ideal choice. It is recommended to use with high risk.

Conclusion: Especially in large-scale projects such as mines; It has been emphasized that it is necessary to decide on the use of products, machines and materials by considering the OHS criteria at the planning stage of the work. The biggest limitation of this research is that it was applied in a single sample area. With this pioneering study, it is recommended to be applied in different business planning in the mining sector.



2012-2019 Yılları Arasında Maden Sektöründe Yaşanan İş Kazalarının Analizi ve ÇKKV Yöntemlerinin Katkısı

* Tuğçe ORAL¹

*Uskudar University, Faculty of Health Sciences, Occupational Health and Safety Department, İstanbul, Turkey
tugce.oral@uskudar.edu.tr orcid.0000-0003-1795-1550

Alıntı / Citation:

Oral, T., (2021). The Problems Encountered in the Process of Publishing Articles in the Journal and Investigation of the Solutions, *Journal of Scientific Technology and Engineering Research*, 2(2):101-109. DOI:10.53525/jster.1026082

ÖNE ÇIKANLAR / HIGHLIGHTS

- Bu çalışma ile İSG Kanunu'nun uygulanması ile birlikte sektöre etkisi analiz edilmiştir.
- Sektöre ait iş kazası sıklık oranı ve ağırlık hızı hesaplanmıştır.
- Maden sektöründe yaşanan kazaları önlemeye yönelik uygulamalar arasında ÇKKV yöntemlerin tercih gerekliliği vurgulanmıştır.
- Bu makale ile İSG kültürünün kullanılacak malzeme seçiminde bile önemsenmesi gerektiği vurgulanmıştır.
- Bu araştırma ile Ülkemizde Madencilik sektörünün yıllara göre gelişmişlik düzeyi hakkında tespitler sağlanmıştır.

Makale Bilgileri / Article Info

Received : 19.11.2021
Accepted : 20.12.2021
Published : 21.12.2021

DOI: 10.53525/jster.1026082

*Sorumlu Yazar /
Correspondence :

Tuğçe ORAL
tugce.oral@uskudar.edu.tr
Tel: +90 216 400 22 22

ÖZET / ABSTRACT

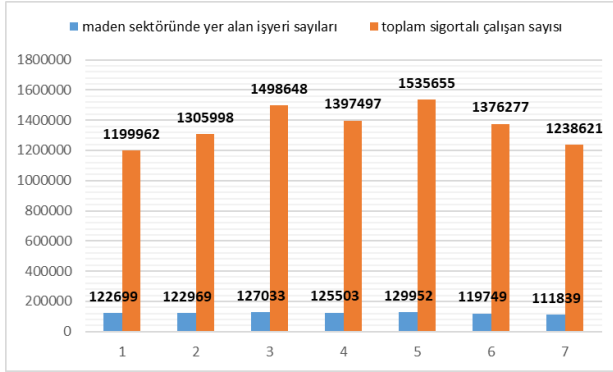
Bu çalışmada, iş kazalarının en sık yaşandığı maden sektörü incelenmiştir. Araştırma da ilk olarak, SGK 'ya bildirisi yapılan ve SGK kayıtlarına geçen 2012-2019 yılları arasındaki iş kazası verileri değerlendirilerek, her yıla ait iş kazası sıklık oranları hesaplanmıştır. Yıllara göre kaza sıklık oranı hesaplama sonuçlarına bakıldığında ise 2012 yılında iş kazası geçirmeyen SGK girişi yapılmış çalışanların kaza geçirme olasılığı geçirmeyenler %4, 2013 yılında %13, 2014 yılında %18, 2015 yılında %22, 2016 yılında %29, 2017 yılında %14, 2018 yılında %19 ve 2019 yılında ise %17 olarak bulunmuştur. Mevcut kanun ve yaptırımlar düşünüldüğünde ve en güncel değer baz alındığında, madencilik sektöründe İSG uygulamalarının daha fazla nasıl entegre edilebilir sorusuna yanıt aranmıştır. Bu nedenle araştırmanın ikinci aşamasında çok kriterli karar verme yöntemlerinden faydalanarak; sektörde işin her aşamasında kullanılacak makine-teçhizat, malzeme, ürün veya hizmet seçiminde İSG'ye katkısında bir kriter olarak değerlendirilmesinin gerekliliği vurgulanmıştır. Madenlerde örnek bir uygulama aşaması olarak beton donatısı kullanımında alternatifler değerlendirilirken, işin yürütümü gereği ihtiyaçlara ek olarak İSG kriteri de eklenmiş ve en ideal çözüm ikame yöntem olarak önerilmiştir.

Anahtar Kelimeler: Beton Donatısı, ÇKKV Yöntemi, İş Kazası Sıklık Oranı, Maden Sektörü,

I. GİRİŞ [INTRODUCTION]

Madencilik sektörü, geniş istihdam imkânı ve ekonomik gelişmelerine katkı sağlaması nedeniyle ülkelerin gelişmişliğini etkileyen önemli sektörlerden biri halindedir. Bu nedenle, toplumların refah ve gelişmişlik düzeyleri ile madencilik faaliyetleri ilişkilendirilmektedir (Özdemir, 2012). Madencilik sektöründe yaşanan iş kazası oranları sadece ülkemize özel olmayıp, tüm dünya genelinde oranların yüksek olduğu işkollarından biri olmaktadır [1]. Çalışma ortamının jeolojik yapısı, yapılan

işin gereği kullanılan iş makineleri, çalışan başına düşen iş yükü, fiziksel risk etmenleri ve çevresel etmenlerin bir arada olması tehlikelere kaynak oluşturmaktadır. Çalışılan ortamda çok fazla risk ve tehlikenin bir arada bulunması ve bunların hepsinin bir arada değerlendirilerek güvenlik tedbirlerinin uygulanması, takip edilmesi çok kolay olmamaktadır. 2012 yılında kabul edilen İş Sağlığı ve Güvenliği (İSG) Kanunu 'nun 2013 yılında yürürlüğe girmesi nedeniyle sektöre ait veriler değerlendirilirken, 2013 yılı ve 2019 yılları arasındaki SGK verileri referans alınmıştır (SGK Verileri, 2013-2019).

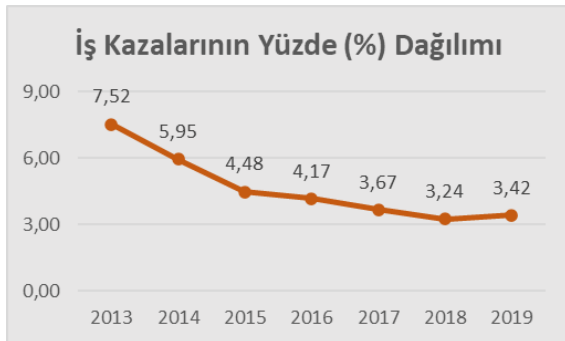


Şekil 1. 2013-2019 Yılları Arasında Maden Sektöründe Yer Alan İşyeri ve Sigortalı Çalışan Sayıları (SGK Verileri, 2013-2019)

Şekil 1'e bakıldığında, maden sektöründe son yıllarda işyeri sayısı ve sigortalı çalışan sayısında %57 'ye yakın bir düşme olduğu görülmektedir.

Maden sektörünün özellikle üretim ve ithalat açısından ülke ekonomisine geniş katkı sağlamasının yanında ocağın kurulduğu bölgede yüksek istihdam imkanı sağlayarak, bölgesel ve kırsal kalkınmayı desteklemektedir. Bu nedenle ülkelerin ekonomik kalkınma hedeflerinin gerçekleşmesine katkı sağlayan maden sektörünün devamlılığının sağlanması için yaşanan iş kazalarını minimize edecek yol haritalarının oluşturulması son derece önem taşımaktadır [2].

Maden sektörü olarak kategorize edilen grupta yer alan çalışma alanlarında 2013-2019 yılları arasında gerçekleşen iş kazalarının, toplam iş kazalarına göre yüzde değerleri hesaplanarak Şekil 2.'de gösterilmiştir.



Şekil 2. Maden Sektöründe Yaşanan İş Kazalarının Tüm Sektörlere Göre Kıyaslanarak Yüzde Dağılımı (SGK Verileri, 2013-2019).

Şekil 2. 'ye bakıldığında sektörde yaşanan iş kazalarının yüzde dağılımı azalan bir eğilim göstermektedir. Fakat bunun nedeni olarak yıllara göre işyeri sayısında ve istihdam edilen sigortalı çalışan sayısındaki azalmadan

kaynaklandığı düşünülmektedir (Şekil 2).

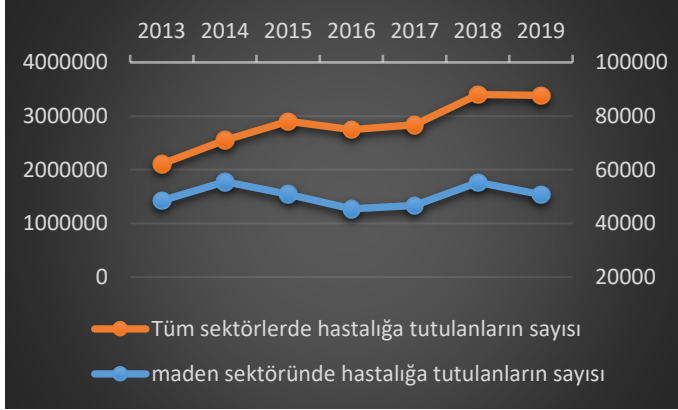
İş kazası ve meslek hastalıklarını önlemek için tüm sektörlerde ait asgari koşulların belirtildiği rehber bir yol gösterici olan İSG Kanunu'nun etkinliği, sektörlerde uygulanabilirlik dereceleri ile paralellik göstermektedir. Bu çalışma ile İSG Kanunu'nu sonrasında sektördeki İSG uygulamalarının etkinliği, yaşanan iş kazalarının sıklık ve ağırlık oranları ile mevcut durum analiz edilmeye çalışılmıştır. Ek olarak, İSG uygulamalarının 2. öncelikli basamağı olan "ikame et" koşulunda, uzmanların ve işverenlerin karar verme tercihlerinde çok kriterli karar verme yöntemlerinden faydalanmasının sağlayacağı katkı araştırılmıştır.

II. MADEN SEKTÖRÜNDE İŞ SAĞLIĞI VE GÜVENLİĞİ

İş sağlığı ve güvenliği uygulamaları işyerlerinin dahil oldukları tehlike grubuna göre değişkenlik göstermektedir [3]. Maden sektörü; çalışma ortamı ve koşulları, kullanılan makine-teçhizatlar ile etkilenebilecek çalışanların iş kazası geçirme ihtimalinin bir arada değerlendirilmesi neticesinde çok tehlikeli sınıfta yer alan işyerleri arasında değerlendirilmektedir.

İş sağlığı ve güvenliği açısından değerlendirildiğinde, işin yürütümünde kullanılan makine-teçhizatlar dışında, sektöre ait başlıca tehlike kaynakları: kazı işi, tozlar, gazlar, tahkimat, malzeme aktarımı, ortam havalandırılması, işin gereği yüksek enerji ihtiyacı ve elektrik, ortamdaki gürültü ve titreşim gibi sıralanabilmektedir. İSG 'nin temelinde iş kazası veya meslek hastalığına kaynak olabilecek tehlikelerin ortadan kaldırılması yer almaktadır. Bu nedenle, çalışma ortamında başlıca tehlike kaynaklarına yönelik asgari İSG tedbirleri mevzuatta açıkça belirtilmiş olmasına rağmen, İSG önlemlerinin sadece bu başlıklarla sınırlandırılmasının yeterli olmadığı SGK verilerinden de anlaşılmaktadır.

2013-2019 yılları arasında yayınlanan SGK verileri incelendiğinde, maden sektöründe istihdam edilen çalışanlar arasında işle ilgili hastalık geçirenlerin sayıları Şekil 3 'te verilmiştir.



Şekil 3. 2013- 2019 Yılları Arasında İstihdam Edilen Toplam Sigortalı Çalışanlar ile Maden Sektöründe Çalışanların Meslek Hastalığı Sayılarının Karşılaştırması (SGK Verileri, 2013-2019).

Şekil 3. incelendiğinde tüm sektörlerde istihdam edilen sigortalı çalışanlar arasında işle ilgili hastalık geçirenlerin sayısını göstermektedir. Tüm sektörlerde istihdam edilenler arasında hastalığa tutuluların; 2013 yılında %1,49 ‘u, 2014 yılında %1,62 ‘si, 2015 yılında %1,64 ‘ü, 2016 yılında % 1,65 ‘i, 2017 yılında % 1,75 ‘i, 2018 yılında % 2,17 ‘si, 2019 yılında % 2,29 ‘u maden sektöründe çalışanlardan oluşmaktadır. Her ne kadar grafik üzerinde azalmaya başlayan bir eğilim gösteriyor gibi olsa da, maden sektöründe işle ilgili hastalığa yakalananların sayısında bir artış olduğu görülmektedir. Bu durum, sektöre ait İSG uygulamalarının etkinliğini arttırmaya yönelik uygulamaların ve denetimlerin gerekliliğini ortaya koymaktadır.

İş kazalarının ve meslek hastalıklarının önlenmesinde işyerlerinde alınacak sağlık ve güvenlik koşullarında işverenler birincil sorumlu tutulmuş olsa da devlet tarafından destekleyici programların oluşturulması pozitif katkı sağlayacaktır. İş sağlığı ve güvenliği aslında işçi, işveren ve devlet olmak üzere üç tarafın ortak katılımı ve katkılarıyla başarı sağlanabilecek bütünsel bir yaklaşımdır. Bu amaçla, maden sektörünün devamlılığı ve sektörde istihdam edilenlerde iş güvenliği kültürü oluşturulması adına T.C. Çalışma ve Sosyal Güvenlik Bakanlığı tarafından 2018 yılında ‘Madencilik Sektöründe İş Sağlığı ve Güvenliğinin Geliştirilmesi’ başlıklı uluslararası proje ile sektörde proaktif yaklaşımla daha iyi çalışma koşullarının oluşturulması amaçlanmıştır [4]. Bu proje ile madencilik sektöründe yer alan tüm paydaşların İSG hakkında bilinçlendirilmesi, bilgi seviyelerinin artırılması, sektördeki proaktif yaklaşım hedeflerinin iyileştirilmesi için gerekli koşulların desteklenmesi amaçlanmıştır.

III. YÖNTEM VE BULGULAR

Ülkemizde İSG Kanunu ile tüm işverenler sigortalı çalışanlarının geçirmiş oldukları iş kazalarını SGK ‘ya 3 (üç) iş günü içerisinde bildirmekle sorumlu tutulmuştur [5]. Böylece, işverenler tarafından SGK ‘ya bildirilen iş kazaları ve meslek hastalıkları sektörlere göre kategorize edilerek kayıt altına alınmakta ve veriler kaynak oluşturacak şekilde SGK tarafından yayınlanmaktadır.

Araştırmada, ülkemiz açısından İSG uygulamalarının sıfır noktası olan 2013 yılı ile son yayınlanan güncel veriler (2019) veri kaynağı olarak kullanılmıştır. Maden sektörüne ait verilerin incelenmesiyle sektöre ait iş kazası sıklık oranı hesaplanmış, böylece mevcut durum sayısal olarak irdelenmeye çalışılmıştır.

İkinci yöntem olarak ise Çok Kriterli Karar Verme Yöntemleri (ÇKKV) arasında, alternatif seçimlerinde en ideal olana karar vermek için en sık kullanılan yöntemlerden biri olan ‘‘Bulanık TOPSİS’’ metodundan faydalanılmıştır. Bulanık TOPSİS verileri toplanması için aktif devam eden 10 farklı maden projesinde yer alan işveren veya işletme müdürleri karar verici olarak belirlenmiştir. Beton donatısı seçiminde kıyaslamayı yapacak karar vericiler alternatifler arasında kıyaslama yapabilecek teknik donanıma sahip kişilerden seçilmiştir. Karar vericiler ile yapılan görüşmeler ile alternatifler arasında kıyaslama yapabilmeyi sağlayacak avantaj kriterleri belirlenmiştir. Karar vericilere yöneltilen anket formları ile alternatiflerin avantaj kriterlerine göre kıyaslaması yapılmış ve veriler toplanmıştır. Araştırmada özellikle ÇKKV yöntemi kullanılmasının nedeni; aynı amaç için kullanılan malzemeler arasında seçim yaparken, tek bir nedene bağlı olmaksızın seçim yapılması vurgulanmak istenmiştir.

A. İş Kazası Sıklık Oranı

Sigortalı çalışanların geçirmiş olduğu iş kazası sayıları referans alınarak, sektöre ait mesai saatlerinin oranlanmasıyla elde edilmektedir [6]. Kısaca, iş kazası sıklık oranıyla bir yıl içerisindeki toplam mesai saatleri içerisinde iş kazası geçirme olasılığı hesaplanmaktadır. Maden sektörü için kömür ve linyit çıkartılması, ham petrol ve doğalgaz çıkartılması, diğer madencilik ve taş ocaklığı, madenciligi destekleyici hizmetler, kok kömürü ve petrol ürünleri imalatı, metal cevheri madenciligi faaliyet kollarına ait iş kazası sayıları veri kaynağı olarak kullanılmıştır (SGK İstatistikleri, 2013-2019). Ek olarak, maden sektörüne ait iş kazası sıklık oranını hesaplamak için haftalık çalışma sürelerinin 37,5 saat ile sınırlı kaldığı, günlük mesai sürelerinin 7,5 saat ve

resmi tatillerde çalışmanın olduğu varsayılmıştır. SGK verileri incelendiğinde 2013 yılında maden sektöründe zorunlu sigortalı çalışan sayısının; 1,199,962 kişi, 2014 yılında 1,305,998 kişi, 2015 yılında 1,498,648 kişi, 2016 yılında 1,397,497 kişi, 2017 yılında 1,535,655 kişi, 2018 yılında 1,376,277 kişi ve 2019 yılında 1,238,621 kişi olduğu saptanmıştır. 2013-2019 yılları arasında gerçekleşen iş kazalarına göre Kaza Sıklık Oranını (KSO) hesaplamak için aşağıdaki formülden faydalanılmaktadır.

$$KSO = \frac{TİKS}{[(TÇG-ÇOG) \times GÇS]} \times 10000 \quad (1)$$

TİKS : Toplam İş Kazası Sayısı

TÇG : Toplam Çalışma Günü

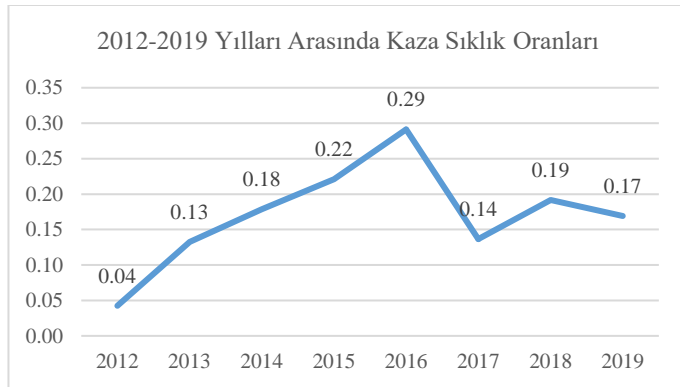
ÇOG : Çalışma Olmayan Gün

GÇS : Günlük Çalışma Süresini, ifade etmektedir.

Tablo 1.
2012-2019 Yılları İş Kazası Verileri ve Sıklık Oranları

YILLAR	2012	2013	2014	2015	2016	2017	2018
TİKS	5867	17113	23880	28747	37738	45412	56023
TÇG-ÇOG	307	287	297	289	288	289	283
GÇS	7,5	7,5	7,5	7,5	7,5	7,5	7,5
Çalışan Sayısı	1083769	1199962	1305998	1498648	1397497	1535655	1376277
saat hesarlama	10000	10000	10000	10000	10000	10000	10000
KSO	0,04	0,13	0,18	0,22	0,29	0,14	0,19
Ortalama Değer	4%	13%	18%	22%	29%	14%	19%

Tablo I. 'de yıllara göre madencilik sektöründe gerçekleşen iş kazası sıklık oranını hesaplamaya yönelik kaynak veriler ve iş kazası sıklık oranları verilmiştir.



Şekil 5. İş Kazası Sıklık Oranlarının Yıllara Göre Değişimi

Şekil 5. incelendiğinde iş kazası sıklık oranlarının 2016 yılına kadar doğrusal bir artış sergilediği görülmektedir. 2016 yılı sonrasında ise artış hızındaki düşüşün nedeni olarak, sektörde gerçekleşen iş kazalarının ve sonuçlarının büyük boyutlara ulaşması nedeniyle oluşan sektörel farkındalık sayesinde; denetimlerin ve yasal yaptırımların artması ve işyeri sayısındaki azalmalardan kaynaklandığı düşünülmektedir.

B. Çok Kriterli Karar Verme (ÇKKV) Yöntemi: BTOPSIS Methodu

ÇKKV yöntemleri son yıllarda alternatifli olan ürün, hizmet veya çalışan istihdamı gibi durumlarda ihtiyaca yönelik, doğru ve maksimum fayda sağlayacak olanı seçmek için kullanılan bir değerlendirme yöntemidir. Karar vermeyi etkileyen tüm etkenlerin; maliyet, performans, kullanım süresi, zaman tasarrufu, uygulama kolaylığı, dayanım, ulaşılabilirlik veya taşınabilirlik gibi alternatiflerin sahip oldukları özelliklere göre belirlenen kriter başlıkları ile sayısal değerlendirmeler yaparak iyinin iyisini veya daha çok faydalı olanı seçmede yardımcı olan bir yöntem olarak tanımlanmaktadır [7].

Araştırmada, Bulanık TOPSIS (BTOPSIS) metodu kullanılarak; iş kazalarının çok sık ve istenmeyen şekilde sonuçlanarak yaşandığı madencilik sektöründe kullanılacak olan makine, malzeme- ekipman veya donanım gibi ihtiyaç listeleri oluşturulurken, işverenlerin sadece maliyet ve performans odaklı olmalarının yanı sıra iş sağlığı ve güvenliği kriteri de eklenerek karar vermelerinin pozitif katkısı, örnek bir uygulama üzerinde açıklanmaya çalışılmıştır.

TOPSIS metodu, 1981 yılında Hwang ve Yoon tarafından geliştirilmiş olup, ÇKKV yöntemleri arasında uygulama kolaylığı açısından yaygın kullanılmaktadır (Doğanalp, 2016). BTOPSIS Metodu özellikle alternatifler arasındaki kıyaslamalarda, hemen hemen, yaklaşık olarak gibi sözel ifadeler kullanılması durumunda doğru seçim için etkili bir yöntem olmaktadır [8]. Bu nedenle, BTOPSIS metodu ile dilsel değişkenler bulanık sayı karşılıkları ile kullanılmasıyla TOPSIS yönteminden farklı olmaktadır.

BTOPSIS metodu, 1992 yılında Chen ve Hwang tarafından bulanık küme teorisinin TOPSIS metoduna entegre edilmesiyle geliştirilmiş bir ÇKKV yöntemidir [9]. TOPSIS yönteminde de olduğu gibi elde edilen çözüm sonucunda negatif çözüme en uzak ve pozitif ideal çözüme en yakın olan alternatif 'ideal seçim' olarak kabul edilmektedir [10].

BTOPSIS yöntemi 8 farklı uygulama aşamasından oluşmaktadır (Chen, 2000, s:6). Birinci aşama: alternatifler arasında kıyaslama yapabilecek düzeyde bilgi veya tecrübeye sahip kişilerin, karar vericilerin, alternatiflerin ve seçim kriterlerinin belirlenmesidir. İkinci aşama: karar vericilerin, alternatifleri seçim kriterlerine göre sözel olarak değerlendirmesidir. Üçüncü aşama: seçim kriterlerinin önem derecelerinin belirlenmesidir. Genellikle önem ağırlıklarının belirlenmesinde Analitik



Hiyerarşi Prosesi (AHP) metodundan faydalanılmaktadır. AHP metodu; ÇKKV yöntemlerinden birisi olup, alternatif seçiminde etkili olan seçim kriterlerinin hiyerarşik olarak sıralanmasında kullanılan bir metod olmaktadır [11]. Bu çalışmada kullanılan örnek uygulamada seçim kriterleri önem sırası AHP metodu ile daha önce belirlenmiş olup, belirlenen önem derecelerine göre BTOPSIS metodu uygulanmıştır. Dördüncü aşamada: Bulanık Karar matrisinin oluşturulması ve normalize edilmesi işlemi yer almaktadır. Beşinci aşamada ise seçim kriterlerinin önem derecelerinin işleme dahil edildiği, ağırlıklandırılmış normalize edilmiş karar matrisinin oluşturulmaktadır. Altıncı aşamada: bulanık ideal çözüm noktaları (pozitif ve negatif uzaklıkların) oluşturulmaktadır. Yedinci aşamada: alternatiflerin ideal çözüm noktalarına olan uzaklıkları hesaplanmaktadır. Sekinci aşamada: alternatiflerin yakınlık katsayıları hesaplanıp, alternatifler sıralanarak, seçim kriterlerinin bileşkesi ile değerlendirilen en ideal seçim belirlenmiş olmaktadır.

C. Madenlerde İkame Ürün Seçiminde ÇKKV Yöntemi Uygulaması

Araştırmada ÇKKV yönteminin özellikle İSG uygulamalarında ikame yöntem, ürün veya hizmete karar vermede kullanılmasının pozitif katkısı ortaya koyulmak istenmiştir.

Türkiye 'de aktif proje olarak devam eden farklı maden ocakları işleten kişiler (işverenler) karar verici olarak atanmıştır. Alternatif olarak, maden ocaklarının oluşturulmasında beton donatısı yani maden galerilerinin çökme gibi etkenlere karşı dayanımını arttırmaya yönelik kullanılan geleneksel donatı uygulaması ile yeni nesil beton katkıları kullanımı arasında kıyaslama yapılmıştır.

İSG açısından, bu uygulama aşamasında çalışanların iş kazası veya meslek hastalığı geçirmesine neden olabilecek durumların ortadan kaldırılması ya da ikame (daha az tehlikeli olanla değiştirmek) edilmesi kanunen gerekmektedir. İşin gereği tehlike kaynağının ortadan kaldırılması mümkün olamayacağı için, ikame yöntemin seçilmesi her zaman işveren ve iş güvenliği uzmanı tarafından tercih edilen öncelikli uygulama olmaktadır. İkame yönteme karar vermede işverenlerin maliyet ve mühendislik katkısına göre seçim yapacağı genel bir durum olduğundan, İSG katkısının da bir kriter olarak göz önünde bulundurulup en ideal alternatifin belirlenebileceği vurgulanmaktadır.

Araştırmada, beton donatısı olarak kullanılacak ürün; geleneksel donatı (GD) mı olmalı? Yoksa, yeni nesil

donatı ürünleri (Makro Sentetik Fiber (MSF)) mi olmalıdır? cevabı aranmıştır. Seçim kriterleri olarak karar vericilerle ve uzmanlarla yapılan görüşmeler sonucu kıyaslama kriterleri olarak: maliyet, mühendislik, zaman yönetimi ve İSG olarak belirlenmiştir. Karar vericilerin yaptıkları kıyaslamalar sonucu elden edilen veriler, yukarıda sıralanan uygulama aşamalarına göre değerlendirilmiştir.

1. Aşama: Karar vericilerin ve alternatiflerin belirlenmesi

Oluşturulan karar verici grubu ile yapılan ön görüşmelere göre alternatifler (Alt.) arasında seçimi etkileyen kriterler belirlenmiştir. Seçim kriterlerine göre alternatifler arasında kıyaslamalar Tablo 2 ve Tablo 3 'de verilen sözel ifadeler karşılık gelecek şekilde yapılmıştır. Daha sonra sözel verilerin analizi için sözel değerlere karşılık gelen sayısal değerlere dönüşümü yapılarak karar matrisi veri kaynakları oluşturulmuştur.

Tablo 2.

Karar Kriterlerinin Belirlenmesinde Kullanılan Sözel İfadeler ve Bulanık Sayı Karşılıkları (Chen, 2000, p: 5).

Kriterlerin Önem Derecesi Belirlemede Kullanılan Sözel İfadeler	
Çok Düşük (ÇD)	(0,0,0.1)
Düşük (D)	(0,0.1,0.3)
Orta Düşük (OD)	(0.1,0.3,0.5)
Orta (O)	(0.3,0.5,0.7)
Orta Yüksek (OY)	(0.5,0.7,0.9)
Yüksek (Y)	(0.7,0.9,1)
Çok Yüksek (ÇY)	(0.9,1,1)

Tablo 3.

Alternatiflerin Karşılaştırılmasında Kullanılan Sözel İfadeler ve Bulanık Sayı Karşılıkları (Chen, 2000, p: 5).

Alternatiflerin Karşılaştırılmasında İfadeler	Kullanılan Sözel
Çok Kötü (ÇK)	(0,0,0.1)
Kötü (K)	(0,0.1,0.3)
Orta Kötü (OK)	(0.1,0.3,0.5)
Orta (O)	(0.3,0.5,0.7)
Orta İyi (Oİ)	(0.5,0.7,0.9)
İyi (İ)	(0.7,0.9,1)
Çok İyi (Çİ)	(0.9,1,1)

2. Aşama: Sözel ifadeleri üçgen bulanık sayılara dönüştürme

Karar vericilerin vermiş oldukları sözel yanıtlar Tablo 2 ve 3 'de verilen sayısal değer karşılıklarına dönüştürüldükten sonra N tane karar vericinin olduğu varsayılarak kriterlerin önem derecesi (W_i) ve alternatiflerin karşılaştırılması

(Xij), Denklem (2) ve (3) 'de verilen formüllerden faydalanılarak hesaplanmıştır (Tekez ve Bark, 2016).

$$\tilde{W}_i = \frac{1}{n} (\tilde{W}_i^1 + \tilde{W}_i^2 + \dots + \tilde{W}_i^n) \quad (2)$$

$$\tilde{X}_{ij} = \frac{1}{n} (\tilde{X}_{ij}^1 + \tilde{X}_{ij}^2 + \dots + \tilde{X}_{ij}^n) \quad (3)$$

Denklem (3)'de verilen eşitlikten faydalanarak değerlendirmeye katılan tüm karar vericiler için alternatif karşılaştırması hesaplanmış ve alternatifler için tek bir bulanık sayı elde edilmiştir. Karar vericilerin her bir kriter için vermiş oldukları yanıtlar üçgen bulanık sayılara dönüştürülmüş, sonra her bir karar vericinin alternatifler için vermiş oldukları yanıtların aritmetik ortalaması alınmıştır.

3. Aşama: Bulanık çok kriterli karar verme matrisi oluşturma

Alternatiflerin karşılaştırılması ve kriterlerin derecelendirilmesi için Tablo 4'deki gibi bir bulanık karar matrisi, seçimi etkileyen kriterlerin önem derecesine hesaplanan Tablo 5'te ki gibi bulanık kriter ağırlıkları matrisi elde edilmiştir.

Tablo 4.
Bulanık Karar Matrisi

	Alt.	K1			K2			K3			K4		
	MSF	0,36	0,47	0,65	0,20	0,38	0,58	0,14	0,28	0,46	0,39	0,57	0,74
GD	0,60	0,78	0,91	0,68	0,85	0,95	0,68	0,85	0,94	0,66	0,82	0,92	
	Alt.	K5			K6			K7			K8		
	MSF	0,80	0,95	1,00	0,12	0,25	0,44	0,26	0,43	0,62	0,72	0,90	0,99
GD	0,72	0,88	0,96	0,78	0,94	1,00	0,78	0,43	1,00	0,64	0,81	0,93	
	Alt.	K9			K10			K11			K12		
	MSF	0,74	0,89	0,96	0,30	0,47	0,65	0,39	0,58	0,77	0,34	0,52	0,71
GD	0,62	0,80	0,92	0,76	0,91	0,97	0,80	0,95	1,00	0,76	0,91	0,98	
	Alt.	K13			K14			K15			K16		
	MSF	0,25	0,25	0,62	0,15	0,29	0,48	0,34	0,54	0,73	0,28	0,46	0,66
GD	0,60	0,60	0,93	0,62	0,81	0,94	0,62	0,80	0,91	0,68	0,85	0,95	
	Alt.	K17											
	MSF	0,16	0,33	0,52									
GD	0,68	0,86	0,96										

Tablo 5.
Bulanık Kriter Ağırlıkları Matrisi

	K1		K2		K3		K4		K5						
	0,42	0,62	0,79	0,78	0,93	0,99	0,7	0,86	0,95	0,37	0,56	0,74	0,68	0,86	0,96
	K6		K7		K8		K9		K10						
	0,62	0,81	0,95	0,82	0,96	1	0,74	0,9	0,98	0,82	0,96	1	0,54	0,73	0,88
	K11		K12		K13		K14		K15						
	0,48	0,68	0,86	0,8	0,95	1	0,56	0,75	0,9	0,64	0,81	0,91	0,72	0,89	0,98
	K16		K17												
	0,52	0,71	0,87	0,7	0,86	0,95									

4. Aşama: Bulanık karar matrisinin normalizasyon işlemi

Tablo 4'te verilen bulanık karar matrisi aşağıdaki denklemlerden faydalanılarak normalize edilmiştir. Tablo 6 'daki normalize edilmiş karar matrisi oluşturulmuştur.

$$\tilde{r}_{ij} = \left(\frac{a_{ij}}{c_j^*}, \frac{b_{ij}}{c_j^*}, \frac{c_{ij}}{c_j^*} \right), \quad j \in B, \quad c_j^* = \max C_{ij} \quad (4)$$

$$\tilde{r}_{ij} = \left(\frac{a_j^-}{c_{ij}^-}, \frac{a_j^-}{b_{ij}^-}, \frac{a_j^-}{a_{ij}^-} \right), \quad j \in C, \quad a_j^- = \min a_{ij} \quad (5)$$

Tablo 6.
Normalize Edilmiş Karar Matrisi

	Alt.	K1			K2			K3			K4		
	MSF	0,55	0,72	1,00	0,34	0,66	1,00	0,30	0,61	1,00	0,53	0,77	1,00
GD	0,66	0,86	1,00	0,72	0,89	1,00	0,72	0,90	1,00	0,72	0,89	1,00	
	Alt.	K5			K6			K7			K8		
	MSF	0,80	0,95	1,00	0,27	0,57	1,00	0,42	0,69	1,00	0,73	0,91	1,00
GD	0,75	0,92	1,00	0,78	0,94	1,00	0,78	0,94	1,00	0,69	0,87	1,00	
	Alt.	K9			K10			K11			K12		
	MSF	0,77	0,93	1,00	0,46	0,72	1,00	0,51	0,75	1,00	0,48	0,73	1,00
GD	0,67	0,87	1,00	0,78	0,94	1,00	0,80	0,95	1,00	0,78	0,93	1,00	
	Alt.	K13			K14			K15			K16		
	MSF	0,40	0,40	1,00	0,31	0,60	1,00	0,47	0,74	1,00	0,42	0,70	1,00
GD	0,65	0,65	1,00	0,66	0,86	1,00	0,68	0,88	1,00	0,72	0,89	1,00	
	Alt.	K17											
	MSF	0,31	0,63	1,00									
GD	0,71	0,90	1,00										

5. Aşama: Ağırlıklı normalize edilmiş karar matrisi oluşturma

Normalize bulanık karar matrisi elemanları ile her bir kriterin ağırlıklarının farklı olması nedeniyle bulanık ağırlıklı karar matrisinin elemanlarının birbiri ile çarpım işlemi Denklem (6)'de verilen işlemler sonucu hesaplanmıştır [12]. Elde edilen yeni verilere göre Tablo 7'de ki ağırlıklı normalize edilmiş karar matrisi oluşturulmuştur.

$$\tilde{V} = [\tilde{V}_{ij}] \text{ mxn}, \quad i = 1, 2, \dots, m \quad j = 1, 2, \dots, n$$

$$\tilde{V}_{ij} = \tilde{r}_{ij} \times \tilde{w}_j \quad (6)$$

Tablo 7.
Ağırlıklı Normalize Edilmiş Karar Matrisi

	Alt.	K1			K2			K3			K4		
	MSF	0,23	0,45	0,79	0,27	0,61	0,99	0,21	0,52	1,00	0,20	0,43	0,74
GD	0,28	0,53	0,79	0,56	0,83	0,99	0,51	0,78	0,95	0,27	0,50	0,74	
	Alt.	K5			K6			K7			K8		
	MSF	0,54	0,82	0,96	0,17	0,46	0,95	0,34	0,67	1,00	0,54	0,82	0,98
GD	0,51	0,79	0,96	0,48	0,76	0,95	0,64	0,90	1,00	0,51	0,78	0,98	
	Alt.	K9			K10			K11			K12		
	MSF	0,63	0,89	1,00	0,25	0,53	0,88	0,24	0,51	0,86	0,38	0,70	1,00
GD	0,55	0,83	1,00	0,42	0,68	0,88	0,38	0,65	0,86	0,62	0,88	1,00	
	Alt.	K13			K14			K15			K16		
	MSF	0,23	0,30	0,90	0,20	0,49	0,91	0,34	0,66	0,98	0,22	0,49	0,87
GD	0,36	0,48	0,90	0,42	0,70	0,91	0,49	0,78	0,98	0,37	0,64	0,87	
	Alt.	K17											
	MSF	0,22	0,55	0,95									
GD	0,50	0,77	0,95										



6. Aşama: İdeal çözüm noktalarının belirlenmesi

Denklem (7) ve (8) 'de verilen eşitliklerden yararlanılarak bulanık pozitif ve negatif çözüm noktaları A^* ve A^- olarak saptanır. Sonuç olarak, pozitif ve negatif çözüm noktalarının elemanları karar kriteri sayısı kadar (1,1,1) ve (0,0,0) değer içermektedir. $\tilde{V}_{ij}^{+*} = (1,1,1)$, $\tilde{V}_{ij}^{-} = (0,0,0)$ olarak kabul edilir (Sorin ve ark, 2016, s: 827; Ünal, 2011, s: 25; Çetin, 2019, s: 56).

$$A^* = [\tilde{V}_1^*, \tilde{V}_2^*, \dots, \tilde{V}_n^*] \quad (7)$$

$$A^- = [\tilde{V}_1^-, \tilde{V}_2^-, \dots, \tilde{V}_n^-] \quad (8)$$

7. Aşama: Alternatiflerin A^* ve A^- noktalarına uzaklıklarının hesaplanması

GD ve MSF arasındaki doğru seçime karar vermek için, A^* ve A^- noktalarına uzaklıkları Denklem (9) ve (10) da verilen eşitlikler yardımıyla hesaplanmalıdır. Hesaplanan değerler d^* ve d^- olarak Tablo 8 'de gösterilmiştir.

$$d^* = \sum_{j=1}^n d(\tilde{V}_{ij}, \tilde{V}_j^*), \quad i=1, 2, \dots, m \quad j=1, 2, \dots, n \quad (9)$$

$$d^- = \sum_{j=1}^n d(\tilde{V}_{ij}, \tilde{V}_j^-), \quad i=1, 2, \dots, m \quad j=1, 2, \dots, n \quad (10)$$

Tablo 8.

Çözüm Noktalarına Uzaklıkları

Alternatifler	d^*	d^-
MSF	8,042	3,794
GD	6,031	4,598

Tablo 8 'e göre alternatiflerin bulanık sayılar arası uzaklığı $d(a,b)$ hesaplamak için Vertex metodundan (11) faydalanılmıştır [13].

Vertex hesaplama formülü aşağıdaki gibidir:

$$d(a,b) = \left[\frac{1}{3} (l_a - l_b)^2 + (m_a - m_b)^2 + (n_a - n_b)^2 \right]^{(1/2)}, \quad d(a,b) \in R^+ \quad (11)$$

8. Aşama: Alternatiflerin yakınlık katsayısını hesaplama

Alternatiflerin yakınlık katsayısı CC_i Denklem (12) 'te verilen eşitlik ile hesaplanmıştır. Yakınlık katsayısı karşılıkları Tablo 9 'dan belirlenmiştir. Böylece, MSF veya GD kullanım tericisinde en ideal seçimin sayısal karşılığı saptanmıştır.

$$CC_i = \frac{d_i^-}{(d_i^* + d_i^-)}, \quad i=1,2, \dots, m \quad (12)$$

Tablo 9.

CC_i Katsayılarına Göre Değerlendirme Durumu (Chen, 2000:6)

Yakınlık Katsayısı CC_i	Değerlendirme Durumu
$CC_i \in [0, 0.2)$	Tavsiye edilmez
$CC_i \in [0.2, 0.4)$	Yüksek risk ile tavsiye edilir
$CC_i \in [0.4, 0.6)$	Düşük risk ile tavsiye edilir
$CC_i \in [0.6, 0.8)$	Kabul edilir
$CC_i \in [0.8, 1)$	Kabul ve tercih edilir

MSF ve GD 'nın ideal çözüm noktalarına göre yakınlık katsayıları: Makro Sentetik Fiber Donatı (yeni nesil beton donatı) = **0,321**, Geleneksel Donatı (Çelik Hasır vb.) = **0,433** olarak hesaplanmıştır.

IV. SONUÇ VE DEĞERLENDİRME [CONCLUSION]

Sonuç olarak, tüm sektörler için kıyaslandığında iş kazalarının yaşanma sıklığının çok daha dikkat çekici olduğu maden sektöründe güncel veriler değerlendirilmiş ve iş kazası sıklık oranları hesaplanmıştır. Madencilik sektörüne ait en son yıla ait duruma bakıldığında ise iş kazası geçirmeyenlerin, her 10000 'lik iş saatinde iş kazası geçirme ihtimallerinin %0,17 olduğu tespit edilmiştir. Özellikle, istihdam açısından sektörün gelişim hızı ve ülke ekonomisine sağladığı katkı düşünülecek olursa, bu değerlerin dikkat çekici olduğu gözlemlenmektedir.

Maden sektörünün kendine özgü ağır çalışma koşulları göz önünde tutulduğunda; iş kazası veya meslek hastalıklarına neden olabilecek faktörlerin sadece çalışma koşulları ile sınırlı olmadığı aşikardır. Bu nedenle, bu çalışmada madenlerde işin herhangi bir aşaması örnek olarak alınmış ve kullanılacak beton donatısı seçiminde bile İSG avantajı bir kriter olarak eklenmesi önerilmiştir. Böylece, madenlerde malzeme veya ürün seçimlerinde İSG kriter olarak eklenmeli ve pozitif katkı düzeyinde bir bileşke olarak değerlendirilmesi gerektiği vurgulanmıştır. Örnek olarak ele alınan beton donatısı seçimi kıyaslandığında; MSF ve GD arasında maliyet, mühendislik, zaman yönetimi katkılarıyla İSG 'ye katkısı da değerlendirildiğinde MSF, İSG açısından ikame yöntem olarak önerilebileceği gibi 0,321 yakınlık katsayısı değeri ile yüksek risk ile kullanımı tavsiye edilen en ideal seçim olarak bulunmuştur. Özellikle iş kazalarının yaşanma olasılığı yüksek olan tüm sektörlerde işin gereği kullanılacak makine-teçhizat, ürün veya hizmet gibi her aşamada karar vermede ÇKKV yöntemlerinden faydalanılmasının hem İSG uygulamaları açısından hem de en iyi seçime karar verme açısından doğru bir tercih olacaktır.



ÇIKAR ÇATIŞMASI [CONFLICTS OF INTEREST]

Yazarlar arasında ve ilgili kurumları arasında herhangi çıkar çatışması bulunmamaktadır.

ETİK KURALLARA UYGUNLUK [RESEARCH AND PUBLICATION ETHICS]

Bu çalışmada BTOPSIS verilerinin toplanmasında Etik Kurula başvurulmuş ve gerekli izinler tamamlandıktan sonra veriler toplanmıştır.

KAYNAKLAR [REFERENCES]

- [1] Bayraktar, B., Uyguçgil, H., Konuk, A. (2018). Türkiye Madencilik Sektöründe İş Kazalarının İstatiksel Analizi, Bilimsel Madencilik Dergisi, 2018, Special Issue, s: 85-90. Erişim Adresi: <http://www.mining.org.tr/en/download/article-file/590279>
- [2] Arslanhan, S., Cünedioğlu, H. E. (2010). Madenlerde Yaşanan İş Kazaları ve Sonuçları Üzerine Bir Değerlendirme, Türkiye Ekonomi Politikaları Araştırma Vakfı, 2010. Erişim Adresi: [https://www.drvehq.com/file/DFPublishFile.aspx/FileID1887633843/Key42eh4wgn4axl/1279030826-2.Madenlerde Yaşanan İş Kazaları ve Sonuçları Üzerine Bir Değerlendirme.pdf](https://www.drvehq.com/file/DFPublishFile.aspx/FileID1887633843/Key42eh4wgn4axl/1279030826-2.Madenlerde%20Yaşanan%20İş%20Kazaları%20ve%20Sonuçları%20Üzerine%20Bir%20Değerlendirme.pdf) .
- [3] Tehlike Sınıfı Tebliği, (2012). İş Sağlığı ve Güvenliğine İlişkin İşyeri Tehlike Sınıfları Tebliği, Resmi Gazete Sayısı: 28509, Resmi Gazete: 26.12.2012. Erişim Adresi: <https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=16909&MevzuatTur=9&MevzuatTertip=5>
- [4] MİSGEP, (2018). Madencilik Sektöründe İş Sağlığı ve Güvenliğinin Geliştirilmesi Projesi, TC. Çalışma ve Sosyal Güvenlik Bakanlığı. Erişim Adresi: <https://www.csgb.gov.tr/isggm/projeler/devam-eden-projeler-ua/> , Erişim Tarihi: 02.08.2021.
- [5] İSGK, (2012). İş Sağlığı ve Güvenliği Kanunu, Resmi Gazete: 28339, Yayın Tarihi: 30/6/2012. <https://www.mevzuat.gov.tr/MevzuatMetin/1.5.6331.pdf>
- [6] Yılmaz Oral, T, Ünal, A. (2020). 2016-2018 Yılları Arasında Seyahat Sektörüne Ait İş Kazası Verilerinin Değerlendirilmesi . Ohs Academy , 3 (2) , 61-72 . DOI: [10.38213/ohsacademy.724873](https://doi.org/10.38213/ohsacademy.724873).
- [7] Baydaş, M., Eren, T. (2021). Finansal Performans Ölçümünde ÇKKV Yöntem Seçimi Problemine Objektif Bir Yaklaşım: Borsa İstanbul'da Bir Uygulama. Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi, 16(3):664-687. Retrieved from: <https://dergipark.org.tr/tr/pub/oguiibf/issue/65252/947593>

- [8] Özçakar, N. & Demir, H. (2018). Bulanık Topsis Yöntemiyle Tedarikçi Seçimi . İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü Yönetim Dergisi , 22 (69) , 25-44 . Retrieved from: <https://dergipark.org.tr/en/pub/iuiieyd/issue/9202/115543>
- [9] Chen, T.Y, Tsao, C.Y., (2008). The Interval-Valued Fuzzy TOPSIS Method And Experimental Analysis, Fuzzy Sets and Systems, 159(11): 1410-1428, DOI : [10.1016/j.fss.2007.11.004](https://doi.org/10.1016/j.fss.2007.11.004).
- [10] Chen, C. (2000), "Extensions of the TOPSIS for Group Decision Making Under Fuzzy Environment", Fuzzy Sets and Systems, 114, 1-9.
- [11] Saaty, T.L. (1991). Some Mathematical Concepts of the Analytic Hierarchy Process. Behaviormetrica, 18(29):1-9.
- [12] Tekez, E, Bark, N. (2016). Mobilya sektöründe bulanık TOPSIS yöntemi ile tedarikçi seçimi. Sakarya University Journal of Science, 20(1):55-63. Retrieved from: <http://www.saujs.sakarya.edu.tr/tr/pub/issue/20706/221207> .
- [13] Oral, T. (2020). Makro sentetik fiber donatıları kullanım avantajlarının iş sağlığı ve güvenliği kriteri dahil edilerek topsis ve AHP yöntemiyle belirlenmesi. Üsküdar Üniversitesi Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- [14] Ecer, F. (2006). Bulanık Ortamlarda Grup Kararı Vermeye Yardımcı Bir Yöntem: Fuzzy TOPSIS Ve Bir Uygulama. Dokuz Eylül Üniversitesi İşletme Fakültesi Dergisi, 7(2):77-96. Retrieved from : <https://dergipark.org.tr/tr/pub/ifede/issue/4601/62873>.
- [15] Özdemir, R. (2012). Maden Sektörü/Aksaray, TC. Ahiler Kalkınma Ajansı Aksaray Yatırım Destek Ofisi, 2010. Erişim Adresi: <https://www.investinaksaray.com/assets/upload/dosyalar/madencilik-sektor-raporu.pdf>
- [16] SGK Verileri, (2013-2019). SGK İş Kazası İstatistik Verileri, Erişim Adresi: http://www.sgk.gov.tr/wps/portal/sgk/tr/kurumsal/istatistik/sgk_istatistik_yilliklari Erişim Tarihi: 06.07.2021.
- [17] Sorin, N., Dzitac, S., Dzitac, I., (2016). Fuzzy TOPSIS: A General View, Procedia Computer Science, 91:823-831. Erişim Adresi: https://www.researchgate.net/publication/305952512_Fuzzy_TOPSIS_A_General_View
- [18] Ünal, Y. (2011). Bulanık Karar Verme Yöntemleri ve Bir Takım Oyunu İçin Oyuncu Seçimi Uygulaması. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Konya.
- [19] Çetin, A., (2019). Bulanık TOPSIS ve AHP Yöntemleri İle Performans Değerlendirmesi: Esenboğa Havalimanında Bir Uygulama, Yüksek Lisans Tezi. Sosyal Bilimler Enstitüsü, Gazi Üniversitesi.



- RETRACTED -

DSTATCOM Based on Artificial Neural Networks and Particle Swarm Optimization for Voltage Profile Improvement

 Fouad Rashed Zaro

Citation:

Zaro, F. Rashed, (2021). Retraction: *DSTATCOM Based on Artificial Neural Networks and Particle Swarm Optimization for Voltage Profile Improvement*, Journal of Scientific Technology and Engineering Research, 2(1): 32-45. DOI: 10.5281/zenodo.4742866

According to Ethical Principles and Publication Policy of JSTER on the <https://dergipark.org.tr/en/pub/jster/policy>

Withdrawal of Published Article (Editors)

If the JSTER Editorial Board detects copyright infringement and plagiarism in a published or early-view work, it will take the following withdrawal and notification procedures within 15 days at the latest. The work in which ethical violations are detected;

1. The phrase “Withdrawn:” is added at the beginning of the title in the electronic display.
2. Instead of the Abstract and Full Text contents in the electronic display, the reasons for the withdrawal of the study and, if there are detailed sources of evidence, are published together with the notifications of the institutions and organizations to which the author(s) are affiliated.
3. Notice of withdrawal is announced on the home page of the journal website.
4. The electronic and hard copy of the first issue to be published as of the date of withdrawal is added to the list of contents as “Withdrawn: Working Title”, the reasons for the withdrawal and the original citations cited as the source are shared with the public and researchers, starting from the first page.

Retraction Decision :

Due to an application made to our journal, we have determined that an article sent by author was sent to both journals at the same time, one week apart, and accepted and published by both journals. <https://dergipark.org.tr/en/pub/jscai/issue/62620/909282>

In addition, it has been understood that there is an article authorship problem between the thesis teacher and his student regarding the copyright of the article.

Thereupon, it was decided to RETRACT the following article published in our JSTER journal by the decision of the board, and a withdrawal note was added to the article. <https://dergipark.org.tr/en/pub/jster/issue/60584/904547>

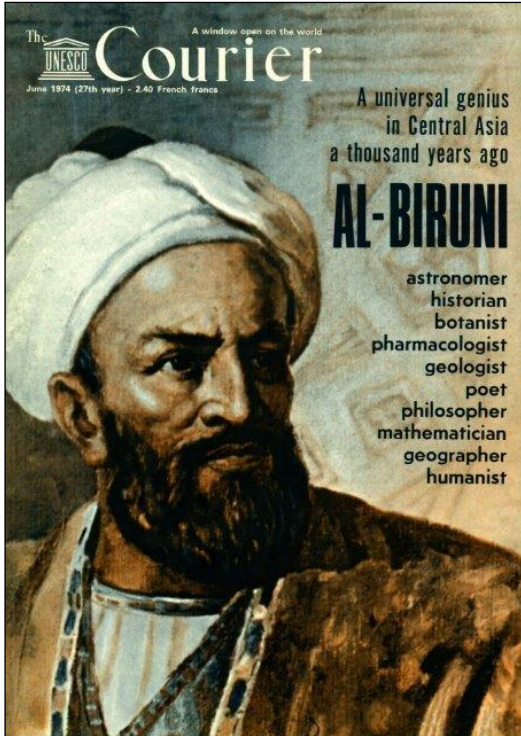


JSTER

JOURNAL OF SCIENTIFIC, TECHNOLOGY AND ENGINEERING RESEARCH
Bilim, Teknoloji ve Mühendislik Araştırmaları Dergisi

ISSN: 2717-8404

El-Biruni



Dünyada bilim sahasına kazandırdığı ilk keşif ve buluşlarla, bilim dünyasına ışık tutan İslam alimi el-Biruni.

Ebu Reyhân Muhammed b. Ahmed el-Biruni h. 3 Zilhicce 362 / m. 4 Eylül 973 tarihinde Harezm bölgesinin merkezi olan Kas'ta dünyaya gelmiştir. el-Biruni'nin vefat tarihi ise tam olarak bilinmemektedir.

Tüm hayatını bilgiye ve bilmediği şeyleri öğrenmeye adayan ve dünyada bilim sahasına kazandırdığı ilk keşif ve buluşlarla, bilim dünyasına ışık tutan müslüman alimi el-Biruni;

fizik, tıp, matematik, astronomi, trigonometri, coğrafya, jeodezi, farmakoloji, jeoloji, filoloji, din gibi konuları incelemiş ve toplamda **113 eser** meydana getirmiştir.

***Islamic scholar al-Biruni**, who sheds light on the world of science with the first discoveries and discoveries he brought to the field of science in the World.*

***Ebu Rayhan Muhammed b. Ahmed al-Biruni** h. 3 Dhul-Hijjah 362 / m. He was born on September 4, 973 in Kas, the center of the Khwarezm region. The date of death of al-Biruni is not known exactly. Muslim scholar al-Biruni, who dedicates his whole life to knowledge and learning what he does not know, and sheds light on the world of science with the first discoveries and discoveries he brought to the field of science in the World.*

*He studied subjects such as **physics, medicine, mathematics, astronomy, trigonometry, geography, geodesy, pharmacology, geology, philology,** religion and produced **113 study** in total.*



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

e-mail : jster.editor@gmail.com / mehmetbulut06@gmail.com