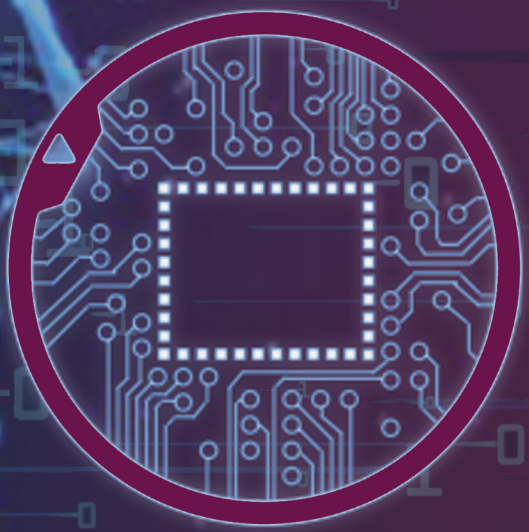
VOLUME-4
ISSUE-1
2024

e-ISSN: 2791-8335

JOURNAL OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE



İZMİR KÂTİP ÇELEBİ UNIVERSITY
Artificial Intelligence and Data Science
Research and Application Center

JAIDA

HTTPS://DERGIPARK.ORG.TR/PUB/JAIDA

JOURNAL OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

Volume 4 – Issue 1 – June 2024

E-ISSN: 2791-8335

Aim & Scope

The Journal of Artificial Intelligence and Data Science (JAIDA) is an international, scientific, peer-reviewed, and open-access e-journal. It is published twice a year and accepts only manuscripts written in English. The aim of JAIDA is to bring together interdisciplinary research in the fields of artificial intelligence and data science. Both fundamental and applied research are welcome. Besides regular papers, this journal also accepts research field review articles. Paper submission/processing is free of charge.

Contact

Web site: https://dergipark.org.tr/pub/jaida E-mail: ikcujaida@gmail.com

Phone: +90 (232) 329 35 35/3731/ 1072

Fax: +90 (232) 325 33 60

Mailing address: İzmir Katip Çelebi Üniversitesi, Yapay Zeka ve Veri Bilimi Uygulama ve Araştırma Merkezi, Balatçık Kampüsü, Çiğli Ana Yerleşkesi, 35620, İzmir, TÜRKİYE

ÖNSÖZ

Yapay Zeka ve Veri Bilimi alanındaki teknolojik ve bilimsel gelişmeler; Yapay Zekanın endüstri, sağlık, otomotiv, ekonomi, eğitim gibi bir çok farklı alanda uygulanmasına imkan sağlamıştır. Ülkemiz Ulusal Yapay Zeka Stratejisinde; yeni bir çağın eşiğine gelindiği, yapay zekayla üretim süreçleri, meslekler, gündelik yaşam ve kurumsal yapıların yeni bir dönüşüm sürecine girdiği vurgulanarak, Yapay Zekanın öneminden bahsedilmiştir.

Sayın Cumhurbaşkanımızın da belirttiği gibi ülkemiz adına insan odaklı yeni bir atılım yapmanın zamanının geldiğine inanıyoruz. Yapay zeka çağına geçiş noktasında Türkiye'nin lider ülkelerden biri olması motivasyonu ile üniversitemizde yapay zeka teknolojilerinin kullanıldığı projeler gerçekleştirmekte, kongreler ve bilimsel etkinlikler düzenlemekteyiz.

Günümüz dünyasına rengini veren dijital teknolojilerin odağındaki ana unsurun yapay zeka teknolojilerinin olduğu düşüncesi ile yola çıkarak hazırlamış olduğumuz Yapay Zekâ ve Veri Bilimi Dergisinin, Ülkemiz Ulusal Yapay Zeka Stratejisinde belirtilen "Dijital Türkiye" vizyonu ve "Milli Teknoloji Hamlesi" kalkınma hedefleri doğrultusunda katkı sağlayacağı inancındayız.

Dergimizin hazırlanmasında emeği geçen üniversitemiz Yapay Zekâ ve Veri Bilimi Uygulama ve Araştırma Merkez Müdürü, Baş Editör Prof. Dr. Ayşegül ALAYBEYOĞLU'na, Editör ve Danışma kurulu üyelerine, akademik çalışmaları ile sağladıkları destek için tüm yazarlara, hakem olarak görev alan değerli bilim insanlarına teşekkür eder, dergimizin yeni sayısının ülkemize hayırlı olmasını dilerim.

Prof. Dr. Saffet KÖSE, Rektör

Dergi Sahibi

PREFACE

Technological and scientific developments in Artificial Intelligence and Data Science enabled the application of Artificial Intelligence in many different fields such as industry, health, automotive, economy and education. In our country's National Artificial Intelligence Strategy; the importance of Artificial Intelligence was mentioned by emphasizing the transformation process of production processes, occupations, daily life and corporate structures with artificial intelligence.

As stated by our President, we believe that the time has come to make a new human-oriented breakthrough on behalf of our country. With the motivation of Turkey being one of the leading countries at the point of transition to the age of artificial intelligence, we realize projects in which artificial intelligence technologies are used, and organize congresses and scientific events at our university.

We have prepared the Journal of Artificial Intelligence and Data Science with the idea that the main element in the focus of digital technologies that color today's world is artificial intelligence technologies, and we believe that our journal will contribute to the development goals of the "Digital Turkey" vision and "National Technology Move" stated in the National Artificial Intelligence Strategy of our country.

I would like to thank Prof. Dr. Ayşegül ALAYBEYOĞLU, the Director of Artificial Intelligence and Data Science Application and Research Center of our university. I would also like to thank to Editor and Advisory Board members, to all authors for their supports with their academic studies and to reviewers for their contributions to the preparation of our journal. I wish the new issue of our journal to be beneficial for our country.

Prof. Dr. Saffet KÖSE, Rector

Privilege Owner

BAŞ EDİTÖR'DEN

Değerli Araştırmacılar ve Dergi Okuyucuları;

İzmir Kâtip Çelebi Üniversitesi Yapay Zekâ ve Veri Bilimi Uygulama ve Araştırma Merkezi olarak Rektörümüz Prof. Dr. Saffet Köse sahipliğinde Yapay Zekâ ve Veri Bilimi Dergisinin 4. cilt 1. sayısını sizlerle buluşturmanın gururunu yaşamaktayız.

İzmir Kâtip Çelebi Üniversitesi Yapay Zekâ ve Veri Bilimi Uygulama ve Araştırma Merkezi olarak hedefimiz; Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı ve Sanayi ve Teknoloji Bakanlığı tarafından hazırlanan "Ulusal Yapay Zekâ Stratejisi" hedefleri doğrultusunda dergi, kongre, eğitim, bilimsel etkinlikler ve proje faaliyetleri gerçekleştirerek ülkemizin yapay zekâ alanındaki gelişim sürecine katkı sağlamaktır.

Farklı üniversitelerden, bilimsel disiplinlerden ve alanlardan değerli araştırmacıların İngilizce dilinde hazırlamış oldukları 6 adet araştırma bu sayı kapsamında sunulmaktadır. Siz değerli araştırmacılarımızın destekleri ile kaliteyi daha da arttırarak en kısa sürede ulusal ve uluslararası indekslerde daha çok taranan bir dergi olmayı hedeflemekteyiz.

Dergimizin yayın hayatına başlaması ve tüm merkez faaliyetlerinde büyük desteklerini gördüğümüz başta Rektörümüz Prof. Dr. Saffet KÖSE olmak üzere; dergimize olan destekleri için tüm yazarlara, dergimizin yayına hazırlanmasında heyecanla çalışan ve çok büyük emek harcayan Baş Editör Yardımcılarına, Editör ve Danışma kurulu üyelerimize, hakem olarak görev alan tüm değerli bilim insanlarına en derin şükranlarımı sunarım.

Saygılarımla,

Prof. Dr. Ayşegül ALAYBEYOĞLU

Baş Editör

LETTER FROM THE EDİTOR-IN-CHIEF

Dear Researchers and Readers of the Journal,

As İzmir Katip Çelebi University Artificial Intelligence and Data Science Application and Research Center, we are proud to present you the volume 4 issue 1 of the Journal of Artificial Intelligence and Data Science (JAIDA), hosted by our Rector Prof. Dr. Saffet Köse.

As İzmir Katip Çelebi University Artificial Intelligence and Data Science Application and Research Center, our goal is; to contribute to the development process of our country in the field of artificial intelligence by carrying out journals, congresses, education, scientific events and project activities in line with the objectives of the "National Artificial Intelligence Strategy" prepared by the Digital Transformation Office of the Presidency of Türkiye and the Ministry of Industry and Technology.

6 research articles prepared by valuable researchers from different universities, scientific disciplines and fields are presented within the scope of this issue. With the support of esteemed researchers, we aim to increase the quality even more and become a journal that is scanned in national and international indexes more as soon as possible.

I would like to express my deepest gratitude to Our Rector, Prof. Dr. Saffet KÖSE, who supported the publication of our journal and the research center's activities; to all the authors for their support to our journal; to our Associate Editors, who worked enthusiastically and put great efforts into the preparation of our journal; to our Editorial and Advisory Board members, and all esteemed scientists who served as reviewer.

Best Regards,

Prof. Dr. Ayşegül ALAYBEYOĞLU

Editor-in-Chief

CONTENTS

Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 4, No. 1, pp. 1-10, 2024
Research Article
Received: 15/05/2024 Accepted: 24/06/2024

Predictive Modeling of Endovenous Laser Ablation Treatment Outcome in Varicose Veins

Steve CHUNG [1,*], Serin ZHANG [2], Sanjay SRIVATSA[3]

Abstract

Varicose veins afflict a significant portion of adults, with approximately 30% experiencing this condition, which often necessitates medical treatment like endovenous laser ablation (EVLA). EVLA has emerged as a highly effective and minimally invasive treatment. However, despite its efficacy, there is a lack of literature on predictive modeling of EVLA treatment outcomes considering both surgical settings and patient characteristics. In this study, we present a comprehensive analysis employing logistic regression under both maximum likelihood (ML) and Bayesian frameworks, as well as support vector machine (SVM) regression. Our results indicate that Bayesian logistic regression with uniform prior demonstrates superior performance. Furthermore, through repeated random sub-sampling validation, we confirm the robustness of our models in predicting successful EVLA outcomes. These findings provide the potential of machine learning techniques in augmenting predictive capabilities in medical decision-making. Our study contributes to the burgeoning literature on predictive modeling in medical contexts, offering insights into the optimization of EVLA treatment outcomes.

Keywords: Bayesian logistic regression; Endovenous laser ablation; Logistic regression; Support vector machine; Varicose veins.

1. Introduction

Varicose veins, although often considered a cosmetic concern, can lead to serious complications such as blood clots and leg ulcers if left untreated, emphasizing the importance of research and intervention. Varicose veins occur with the breakdown of the veins one-way valves causing the vein to dilate. The most common treatment for varicose veins is endovenous laser ablation (EVLA). EVLA has been proven to be the most effective and least invasive treatment for varicose veins. This procedure is done by inserting a laser fiber into the vein, typically near the knee or ankle. The laser is then retracted through vein, causing the vein to occlude. According to Mundy [6], EVLA treatment success rates have been over 89%. EVLA was approved by the National Institute for Health and Clinical Excellence (NICE) in March 2004. This procedure has become very common since it avoids the complications of open surgeries while still having spectacular results which is one factor that is often considered when judging the overall excellence of the surgery. Figure 1 provides an illustration of EVLA treatment.

Although the success rate of EVLA treatment is fairly high, there hasn't been any literature on modeling the EVLA based on both the settings of the surgery and patients' conditions. Some have examined the surgery settings has predictors of the success rate of EVLA treatment. Mordon [7] considered a mathematical model that was produced to determine significant variables in EVLA treatment and optimal levels of endovenous laser Treatment variables that result in minimal vein damage and side effects. They considered both the pulsed and continuous mode of the laser treatment. For each mode, the model determined the optimal linear endovenous energy density

*Corresponding author

Steve CHUNG*; California State University Fresno, Mathematics Department, USA; e-mail: schung@csufresno.edu; 0000-0001-7255-7244

Serin ZHANG; California State University Fresno, Information Systems and Decision Sciences Department, USA; e-mail: serin20@gmail.com; 0009-0003-8745-9534

Sanjay SRIVATSA; Heart Artery Vein Center of Fresno, USA; e-mail: drsanjay@gmail.com; 0000-0002-8661-3187

(LEED) (J/cm) for both 3mm and 5mm vein diameter. They also concluded that pullback distance and laser wavelength do not significantly affect treatment outcome.

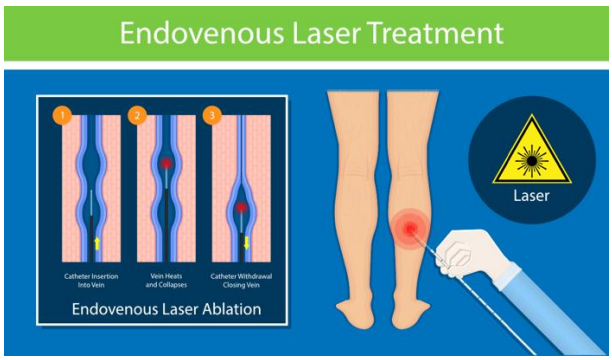


Figure 1. *An illustration of EVLA.*

While the success rate of EVLA treatment is notably high, there remains a gap in the literature regarding the modeling of EVLA outcomes based on both surgical settings and patients' conditions. Some researchers have delved into the predictive value of surgical settings as indicators of EVLA treatment success. For instance, in [7], a mathematical model was developed to identify significant variables in EVLA treatment and determine optimal levels of endovenous laser treatment variables to minimize vein damage and associated side effects. The study explored both pulsed and continuous modes of laser treatment, aiming to establish the optimal linear endovenous energy density (LEED) (J/cm) for vein diameters of 3mm and 5mm. Notably, the researchers concluded that variables such as pullback distance and laser wavelength did not exert a significant impact on treatment outcomes. Cowpland [3] have examined the factors affecting optimal linear endovenous energy density for endovenous laser ablation in incompetent lower limb veins. The findings indicate that the ideal LEED for endovenous laser ablation of the great saphenous vein lies between 80 J/cm and 100 J/cm to achieve optimal closure rates while minimizing side effects and complications. Longer wavelengths, which target water, may have a lower optimal LEED. Conversely, a LEED below 60 J/cm shows reduced efficacy regardless of the wavelength used.

In this study, our initial focus is on modeling the relationship between EVLA treatment outcomes and both surgical settings and patient characteristics using logistic regression under the frequentist approach. We aim to elucidate the factors influencing the success or failure of EVLA procedures. Following this and more importantly, we compare the predictive capabilities of this logistic regression model with Bayesian logistic regression and support vector machine (SVM) regression. These modeling techniques have been extensively utilized in the medical field to predict the success rates of various treatments. For instance, Yussuff [11] employed logistic regression to predict breast cancer based on mammogram results, identifying mass, architectural distortion, skin thickening, and calcification detection as significant predictors. Similarly, Chadwick [1] utilized univariate and multivariate logistic regression analyses to differentiate between dengue fever and other febrile illnesses, achieving a sensitivity of 74% and a specificity of 79%. More recently, Srivatsa [9] investigated the relative contributions of power output, linear endovenous energy density, and pullback rate using logistic regression, highlighting the significance of power output and LEED.

Moreover, researchers have explored Bayesian and SVM approaches in medical prediction tasks. Zhou [12] introduced a Bayesian approach to identifying important genes in cancer classification, while Riaz [8] developed an adaptive SVM regression model to predict the motion of lung tumors, demonstrating its superiority in accuracy compared to traditional methods. Verplankcke [10] investigated the use of SVM models in predicting mortality of critically ill patients with hematological malignancies, finding comparable predictive power to multiple linear regression. Furthermore, Cheng [2] utilized SVM incorporating protein structure and sequence information to predict changes in protein stability following single amino acid mutations. Additionally, Gelman [4] compared four regression methods, including SVM, in tracking lung tumors, with the artificial neural network regression performing slightly better in terms of mean tracking error.

These studies collectively demonstrate the utility of logistic regression, Bayesian logistic regression, and SVM regression in medical prediction tasks, providing valuable insights into treatment outcomes and disease prognosis. By leveraging these diverse modeling techniques, our study aims to enhance the predictive accuracy of EVLA treatment outcomes, contributing to improved patient care and clinical decision-making.

The paper is organized as follows. Section 2 provides a discussion on the data and methodology. It gives the background of the dataset and it provides methodology describing logistic regression, Bayesian logistic

regression, and support vector machine. In Section 3, we provide the results from the models. Section 4 concludes.

2. Data and Methodology

2.1 Dataset

The dataset was obtained from the Heart, Artery, and Vein Center of Fresno, a local clinic in Fresno, CA. It contains information on 359 veins treated using EVLA. Institutional review board (IRB) approval was not needed since the data had already been collected prior to the study. The clinic obtained all patients' consents, and all patient identifications were masked and not revealed to the researchers. Due to missing observations, the complete case data consists of 272 observations from 2015 to 2017. Table 1 presents the descriptions of the variables and Table 2 provides the summary statistics for all the variables.

Table 1. *Variable definitions and summary statistics.*

Variables	**Definitions**	Min	Q1	Median	Mean	Q1	Max	Std. Dev
Age	age in years	26	56.75	65	63.51	72	91	12.41
Height	height in inches	53	63	65.5	65.58	68	74	3.94
Weight	weight in lbs	100	163.8	186	201.9	234	460	60.25
BMI	body mass index	16.8	27.4	30.85	32.75	36.7	62.4	8.33
Power	power setting in watts	6	9	10	9.871	12	12	1.76
Time	time of ablation in seconds	9	58	94	97.95	130	258	49.6
Length	length of treated vein in cm	2.5	28	45	42.68	56	92	17.8
Energy	energy used in joules	11	527	839	968.1	1318	2581	588.4
LEED	linear endovenous energy density in J/cm	1	17.75	21.5	21.91	27	41	7.15
Pullback	pullback rate in mm/s	2.756	3.781	4.485	4.695	5.252	14.655	1.38

Table 2. *Variable definitions and frequencies.*

Variables	Definitions	Yes	No	Total
DM	dementia	107	165	272
HTN	hypertension	195	77	272
Hyperlip	hyperlipidemia	184	88	272
Renal	renal disease	20	252	272
CHD	congenital heart defects	33	239	272

2.2 Methodology

In this section, we present the methods and models that we have used in our study. In the frequentist approach, we assume a specific probability distribution such as a normal distribution and then, estimate the parameters in the model. Maximum likelihood estimation has been one of the most popular estimation methods in a frequentist setting due to its efficiency, consistency, and asymptotic normality. In the Bayesian approach, estimation of parameters involves treating them as random variables rather than fixed quantities, allowing for the incorporation of prior knowledge and uncertainty into the modeling process. Unlike the frequentist approach, which relies solely on observed data to estimate parameters, Bayesian inference combines prior beliefs about the parameters with likelihood functions derived from the data to obtain posterior distributions. These posterior distributions represent updated beliefs about the parameters after observing the data, reflecting both the information contained in the data and the prior knowledge. Bayesian data analysis has become a well-established component of modern applied statistics and machine learning terminology. However, there is no universal consensus on which approach provides better results. This study aims to compare these approaches, along with a machine learning technique, support vector machine, to find an optimal predictive model for EVLA outcome.

2.2.1 Logistic Regression

Logistic regression is a most commonly used model for predicting a binary response variable. Let Y_i be the binary response variable with the conditional probability $p_i = P(Y_i = 1|X = x)$ where the event $\{Y_i = 1\}$ denotes the success of an outcome for the i-th observation. The logistic regression model has the form

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_1 x_{ik}. \quad (1)$$

Solving for p_i from above gives

$$p_i = \frac{1}{1+\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_1 x_{ik})}, \quad (2)$$

which is the probability of the success after observing $x_1, x_2, \ldots, x_k$. The parameters $\beta_0, \beta_1, \ldots, \beta_k$ can be estimated from the maximum likelihood estimation. That is, since *Y* is an independent binary random variable the likelihood function is defined as

$$L(\beta|y) = \prod_i^n \frac{\exp(y_i(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}))}{1+\exp(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik})^{n_i}}, \quad (3)$$

where n_i is the total number of *i*-th trial. Differentiating above equation with respect to β gives k+1 equations and solving for β gives the estimated parameters. However, solving this system of nonlinear equations is not easy since the solution cannot be derived algebraically. A numerical method such as Newton's method is often used to obtain the solution.

2.2.2 Bayesian Logistic Regression

In Bayesian analysis, the posterior distribution $p(\beta|y)$ is obtained from the likelihood function $L(\beta|y)$ and a prior distribution $p(\beta)$. That is,

$$p(\beta|y) = \frac{L(\beta|y)p(\beta)}{\int L(\beta|y)p(\beta)d\beta}. \quad (4)$$

Choosing an appropriate prior distribution is critical in Bayesian setting because the posterior heavily depends on it. For instance, if the prior distribution is chosen to be a beta distribution, then it can be easily be shown that the posterior distribution belongs to a class of beta distributions. Of course, a beta prior distribution may be subjective since the parameter space for β lie in the whole real, whereas a beta distribution only takes the values in (0,1). On the other hand, if the prior distribution is chosen as a normal then there is no closed form for the posterior distribution (unless the likelihood function is normal) and hence, sampling from this posterior is not easy.

For simplicity, we assume that the parameters $\beta_0, \beta_1, \ldots, \beta_k$ are independent. We used two prior distributions in our work. They are

$$\beta_j \sim N(\mu, \sigma^2) : p(\beta_j) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(\frac{-(\beta_j-\mu)\hat{}2}{2\sigma^2}\right) \text{ and } \beta_j \sim Unif(a,b): p(\beta_j) = 1. \quad (5)$$

A normal prior puts a heavy weight near μ while the uniform prior assigns equal weight and hence, it is called a non-informative prior. As mentioned previously, sampling from a posterior is difficult in many cases. For instance, under a normal prior distribution, the posterior is

$$p(\beta|y) \propto \prod_i^n \frac{\exp(y_i(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}))}{1+\exp(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik})^{n_i}} \; exp\left(\frac{-\sum(\beta_j-\mu)^2}{2\sigma^2}\right). \quad (6)$$

However, there is no closed form for this posterior distribution and sampling and making inference from this posterior is not straightforward. Therefore, we resort to the random walk metropolis algorithm which is ubiquitous tool for producing dependent simulations from an arbitrary distribution. The reader is referred to [4] for details on this algorithm. In the Appendix, trace plots and density plots are shown to verify the convergence of the algorithm for each parameter β_j.

2.2.3 Support Vector Machine Regression

Support Vector Machine (SVM) regression is a supervised learning algorithm used for regression tasks, where the goal is to predict outcomes. Unlike traditional regression methods that minimize error directly, SVM regression aims to fit a "tube" around the data points, with the goal of including as many points as possible within the tube while minimizing the margin violations (points outside the tube). SVM regression aims to find a hyperplane that best fits the data points while maximizing the margin, subject to a tolerance ϵ. This hyperplane is used to predict the target values for new data points. The objective of SVM regression is to minimize the following function:

$$\frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{n}(|y_i - w\,\phi(x_i) - b|-\epsilon)_+, \tag{7}$$

where w is the weight, b is the bias term, ϵ is the tube radius (tolerance), C is the regularization parameter (trade-off between maximizing the margin and minimizing the errors. The optimization problem involves finding the optimal values for w and b that minimize the objective function while satisfying the margin constraints. This is be formulated as a quadratic programming problem and solved using optimization technique.

3. Results

The primary objective of this paper is to identify the optimal model for predicting successful outcomes of EVLA treatment. To achieve this goal, we explore frequentist, Bayesian, and machine learning approaches. Our aim is to determine the most effective approach, laying the groundwork for the development of superior models in the future. Model assessment relies on repeated random sub-sampling validation, commonly referred to as Monte Carlo cross-validation. We randomly partition the dataset into four subsamples, with three utilized as training data and the remaining subsample serving as validation data for testing the model. This process is repeated 50 times to ensure robust evaluation.

Given that this is a binary classification problem, each case in the validation set is classified as either a correct or incorrect prediction. We define success as $Y_i = 1$ if the estimated probability is greater than or equal to 0.90. This criterion is referred to as accuracy (ACC) in our results. Sensitivity (TPR), specificity (TNR), and precision were calculated based on the following formulas.

$$Accuracy = ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$Sensitivity = TPR = \frac{TP}{TP+FN} \tag{9}$$

$$Specificity = TNR = \frac{TN}{TN+FP} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

where TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives.

First, we fitted a logistic regression to the whole dataset using the maximum likelihood estimation method as presented earlier. This approach provides parameter estimates that can be easily interpreted, along with the significance of each input variable in our model. Table 3 provides the results of this frequentist approach. Power emerges as the most significant variable, along with Energy and LEED. This result is consistent with the findings of [9].

Table 3. *Logistic regression analysis output.*

Variable	Estimate	Std. Error	Z value	P-value	Significance
Intercept	24.309	14.407	1.687	0.092	.
Age	-0.001	0.020	-0.049	0.961	
Height	-0.305	0.217	-1.407	0.159	
Weight	0.056	0.034	1.639	0.101	
BMI	-0.385	0.209	-1.836	0.066	.
Power	0.576	0.158	3.641	0.000	***
Time	-0.019	0.017	-1.083	0.279	
Length	-0.083	0.043	-1.943	0.052	.
Energy	0.006	0.002	2.558	0.011	*
LEED	-0.222	0.107	-2.081	0.037	*
Pullback	-0.385	0.244	-1.576	0.115	
DM	0.132	0.478	0.275	0.783	
HTN	-0.322	0.550	-0.584	0.559	
Hyperlip	0.729	0.556	1.311	0.190	
Renal Disease	-0.685	0.753	-0.910	0.363	
CHD	1.041	0.726	1.435	0.151	

Table 4. *Logistic regression analysis from the backward selection process.*

Variable	Estimate	Std. Error	Z value	P-value	Significance
Intercept	21.434	13.240	1.619	0.105	
Height	-0.311	0.200	-1.551	0.121	
Weight	0.055	0.031	1.801	0.072	.
BMI	-0.379	0.190	-1.997	0.046	*
Power	0.497	0.147	3.383	0.001	***
Length	-0.080	0.037	-2.192	0.028	*
Energy	0.004	0.002	2.088	0.037	*
LEED	-0.120	0.083	-1.45	0.147	
CHD	1.045	0.702	1.489	0.136	

Subsequently, we applied backward stepwise logistic regression to fit a subset model, selecting the variables Height, Weight, BMI, Power, Length, Energy, LEED, and CHD for classification purposes. The results, displayed in Table 4, indicate that Power is the most significant variable, alongside BMI, Length, and Energy. In contrast, LEED is non-significant in this model, possibly due to its correlation with other variables such as Power and Length.

In Figure 2, a ROC curve displays the Area Under the Curve (AUC) derived from one of the 50 Monte Carlo cross-validations conducted. Table 5 presents the average accuracy, sensitivity, specificity, precision, and AUC across all 50 cross-validations, with the values in parentheses indicating the corresponding standard deviations. Concerning accuracy, sensitivity, specificity, and precision, the Bayesian model with a uniform prior exhibits slightly superior performance compared to (frequentist) logistic regression. Moreover, the Bayesian model with a normal prior outperforms both logistic regression and the Bayesian model with a uniform prior, albeit marginally.

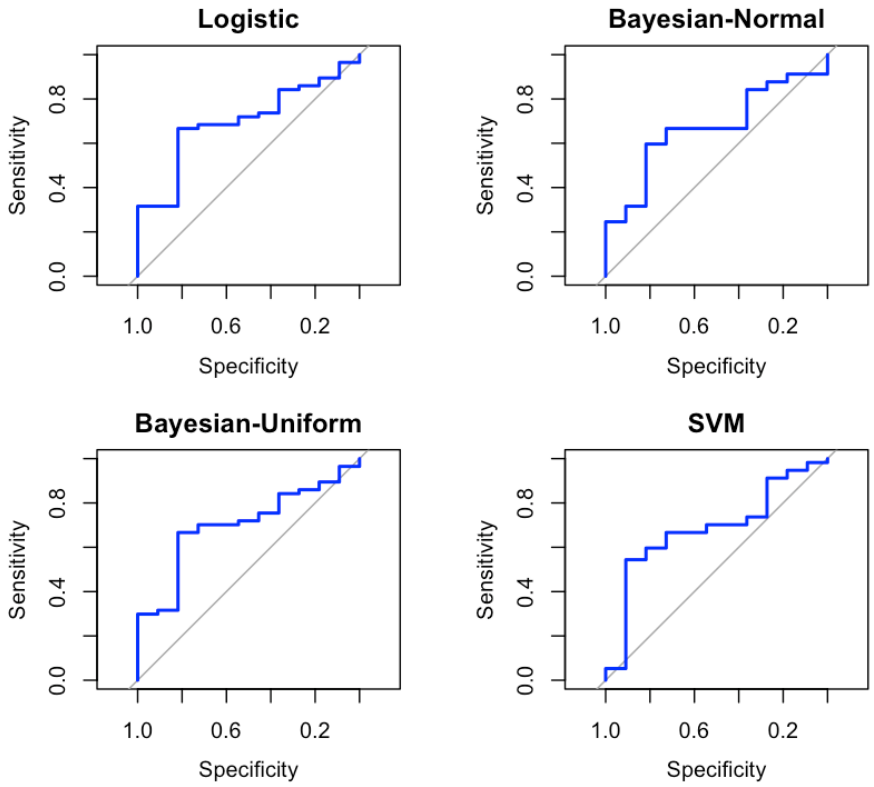


Figure 2. *A sample ROC curve showing the area under the curve (AUC).*

Table 5. *A comparison of different models using accuracy, sensitivity, specificity, precision, and AUC.*

	Logistic	Bayesian (Normal)	Bayesian (Uniform)	SVM
Accuracy	0.861	0.856	0.862	0.816
	(0.040)	(0.038)	(0.039)	(0.036)
Sensitivity	0.102	0.079	0.113	0.074
	(0.092)	(0.100)	(0.102)	(0.088)
Specificity	0.984	0.983	0.985	0.937
	(0.018)	(0.019)	(0.019)	(0.028)
Precision	0.579	0.446	0.585	0.150
	(0.401)	(0.424)	(0.396)	(0.159)
AUC	0.687	0.688	0.687	0.583
	(0.066)	(0.070)	(0.068)	(0.070)

4. Conclusions

Based on the analysis conducted, several key findings emerge regarding the predictive modeling of endovenous laser ablation (EVLA) treatment outcomes:

(a) Logistic Regression Analysis: Initial logistic regression analysis identified significant predictors for EVLA outcomes, notably power, energy, and linear endovenous energy density (LEED). This finding aligns with previous research that highlights the importance of these variables in treatment success.

(b) Comparison of Modeling Approaches: The study compared three modeling approaches: logistic regression, Bayesian logistic regression, and support vector machine (SVM) regression. Across these methods, Bayesian logistic regression with a uniform prior demonstrated slightly superior performance in terms of accuracy, sensitivity, specificity, precision, and AUC compared to logistic regression and Bayesian with a normal prior. SVM regression, while providing acceptable accuracy, showed comparatively lower performance in terms of sensitivity and precision.

(c) Clinical Implications: The findings suggest that Bayesian logistic regression, particularly with a uniform prior, may offer improved predictive capabilities for EVLA treatment outcomes compared to traditional logistic regression. This insight can inform clinical decision-making by providing clinicians with a more accurate assessment of the likelihood of treatment success.
(d) Limitations and Future Directions: While Bayesian logistic regression shows promise, further research is warranted to validate and refine the model. Additionally, exploring additional variables or incorporating advanced machine learning techniques may enhance predictive accuracy further. Furthermore, external validation using data from diverse clinical settings would strengthen the generalizability of the findings.

In conclusion, leveraging Bayesian logistic regression models, particularly with a uniform prior, holds potential for enhancing the prediction of EVLA treatment outcomes. By refining predictive models, clinicians can better tailor treatment strategies, ultimately improving patient care and outcomes in the management of varicose veins.

Declaration of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Steve Chung: the corresponding author, write-up, model fitting, programming. Serin Zhang: write-up, modeling fitting, programming. Sanjay Srivatsa: providing the data, review and editing.

References

[1] D. Chadwick, B. Arch, A. Wilder-Smith, and N. Paton, "Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: Application of logistic regression analysis," Journal of Clinical Virology, vol. 35, no. 2, pp. 147–153, 2006.

[2] J. Cheng, A. Randall, and P. Baldi, "Prediction of protein stability changes for single-site mutations using support vector machines," Proteins: Structure, Function, and Bioinformatics, vol. 62, no. 4, pp. 1125–1132, 2006.

[3] C. A. Cowpland, A. L. Cleese, and M. S. Whiteley, "Factors affecting optimal linear endovenous energy density for endovenous laser ablation in incompetent lower limb truncal veins–A review of the clinical evidence," Phlebology, vol. 32, no. 5, pp. 299-306, 2017.

[4] A. Gelman and D. B. Rubin, "Avoiding model selection in Bayesian social research," Sociological Methodology, vol. 25, pp. 165-173, 1995.

[5] T. Lin, L. I. Cervino, X. Tang, N. Vasconcelos, and S. B. Jiang, "Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy," Physics in Medicine and Biology, vol. 54, no. 4, pp. 981–992, 2009.

[6] L. Mundy, T. L. Merlin, R. A. Fitridge, and J. E. Hiller, "Systematic review of endovenous laser treatment for varicose veins," British Journal of Surgery, vol. 92, no. 10, pp. 1189–1194, 2005.

[7] S. Mordon, B. Wassmer, and J. Zemmouri, "Mathematical modeling of 980-nm and 1320-nm endovenous laser treatment," Lasers in Surgery and Medicine, vol. 39, no. 3, pp. 256–265, 2007.

[8] N. Riaz, P. Shanker, R. Wiersma, O. Gudmudsson, W. Mao, B. Widrow, and L. Xing, "Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression," Physics in Medicine and Biology, vol. 54, no. 19, pp. 5735–5748, 2009.

[9] S. S. Srivatsa, S. Chung, and V. Sidhu, "The relative roles of power, linear endovenous energy density, and pullback velocity in determining short-term success after endovenous laser ablation of the truncal saphenous veins," Journal of Vascular Surgery: Venous and Lymphatic Disorders, vol. 7, no. 1, pp. 90-97, 2019.

[10] T. Verplancke, S. Van Looy, D. Benoit, S. Vansteelandt, P. Depuydt, F. De Turck, and J. Decruyenaere, "Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies," BMC Medical Informatics and Decision Making, vol. 8, no. 56, pp. 1–8, 2008.

[11] H. Yussuff, N. Mohamad, U. K. Ngah, and A. S. Yahaya, "Breast cancer analysis using logistic regression," IJRRAS, vol. 10, January 2012.

[12] X. Zhou, K. Y. Liu, and S. T. Wong, "Cancer classification and prediction using logistic regression with Bayesian gene selection," Journal of Biomedical Informatics, vol. 37, no. 4, pp. 249–259, 2004.

Appendix

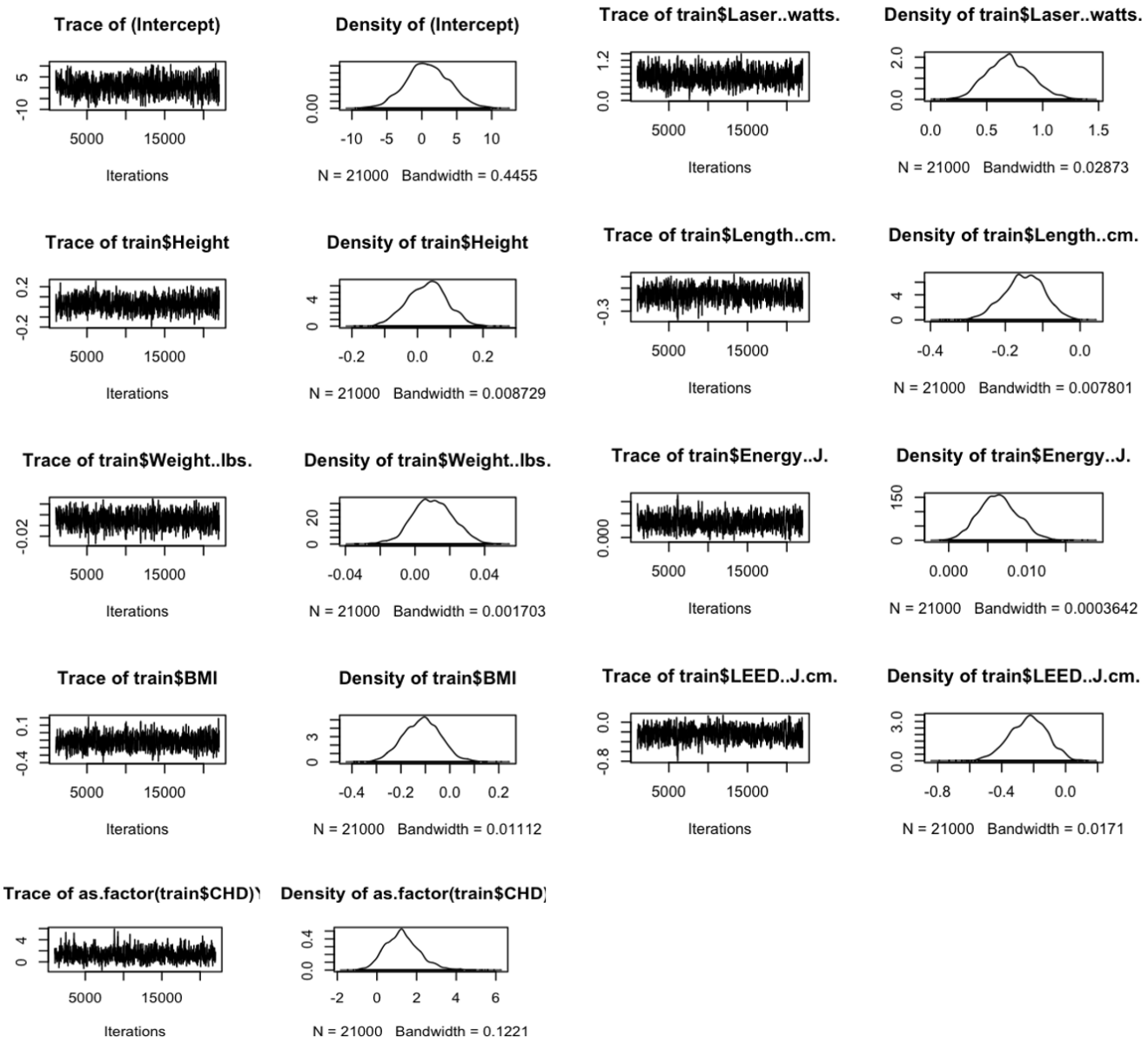


Figure 3. *Trace and density plots of the chains under Bayesian with normal prior.*

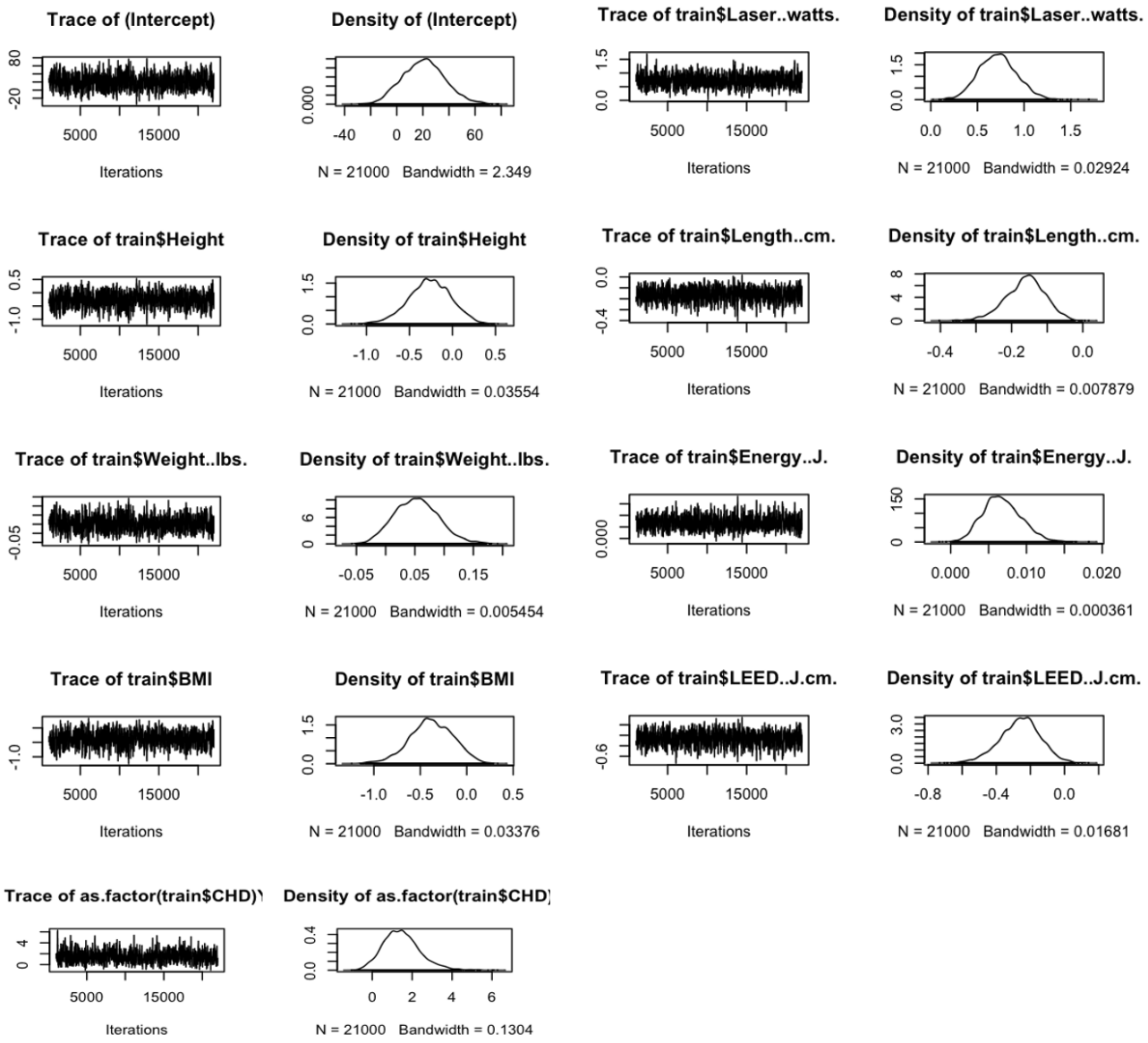


Figure 4. *Trace and density plots of the chains under Bayesian with uniform prior.*

Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 4, No. 1, pp. 11-21, 2024
Research Article
Received: 15/03/2024 Accepted: 17/06/2024

The Age of Digitalization in Industry: From Digital Twins to Digital Product Passport

Mehmet Erdal ÖZBEK*

Abstract

In the age of digitalization, new tools are emerging everyday accommodating Metaverse. Integration of the digital world to the real-world requires a flexible transition framework demonstrated by digital twins (DTs). By definition, DTs constitute a foundation enabling seamless connection using the existing and upcoming emerging technologies. The industry is not exempt from this shift as the digital transformation of industry is already on the way of updating from Industry 4.0 to 5.0. Currently, any transformation attempt is tightly associated with the sustainable development goals. The circular economy requirements challenge the efficiency of the ongoing production mechanisms phrased as smart manufacturing. Therefore, the objective of this paper is to bring to the fore the digital product passport (DPP) that envisioned being a way of pursuing the digitized information through the Internet of Things technology in a production life cycle chain. In this work, a guideline of digital transformation from DTs to DPP is provided on a smart manufacturing process implemented in a laboratory. The method displays the connection between DTs and DPP for a factory model, while discussing possible avenues as well as challenges for further development of the industrial Metaverse. As a result, this study serves as a foundation for companies under digital transformation in achieving better understanding of DT and DPP for a greener, sustainable circle economy.

Keywords: Factory Model; Industrial Metaverse; IoT; Smart Manufacturing; Sustainability.

1. Introduction

The Metaverse is expanding [1, 2]. Its evolution is directly related to the increasing number of user interactions, diverse implementations and proliferating applications. The emerging hardware and software components help to create a better experience with various contents generated in the virtual world. The digitalization effort becomes more apparent with the advanced augmented reality (AR), virtual reality (VR), and extended reality (XR) tools [3]. Previously, Metaverse was known for the composition of the virtual world itself. Today, it is often expressed as a medium for massive interaction centered on the content [4]. The connection of the already existing real-world to a generated virtual/digital world is satisfied through cyber-physical systems. Thus, Metaverse is becoming a foundation for information exchange among humans and those cyber-physical systems.

The undergoing digital transformation envisages shaping the future, embracing machines and humans. It already has significant influence in technological, sociological, and even ideological aspects. This domination will occur not only for building up digitized versions of big physical systems and environments, but also for assembling many small “things” that can be connected in a digital world. The Internet of Things (IoT) technology has enabled this connection through the computer and communication networks almost performing independent of time and physical localization. Due to the standardized enhancements, the IoT technology has now spread to general and everyday use, along with the increasing number of young generations who were born in a comparably higher level of digitalized world [5].

The IoT inevitably has found many extension areas in Industrial IoT (IIoT) where a digital transformation from Industry 4.0 to 5.0 is on the way with the use of artificial intelligence (AI) that particularly depends on machine learning (ML) and deep learning (DL) solutions, handling big amount of data retrieved from sensors through networking [6]. This transition may not be achieved with only a leap in technology. Aside from that it requires multidisciplinary, multiscale processes with coordinated cross functions considering available resources [7]. The digital transformation affects hugely in increasing the resource efficiency of the processes when handled with AI-connected IIoT solutions. It is identified as smart manufacturing that involves all product processes from procurement to recycling in a sustainable way. From this point of view, the manufacturing processes should not only be regarded as finding better efficient solutions for producing products, but also, they have to be designed to pay utmost attention to the environmental and social impacts of all production activities [8].

*Corresponding author

Mehmet Erdal ÖZBEK*; Izmir Katip Çelebi University Faculty of Engineering and Architecture, Türkiye; e-mail: merdal.ozbek@ikcu.edu.tr;

0000-0001-5840-7960

The United Nations has announced the sustainable development goals and related supporting activities to be implemented in all member country actions [9]. The sustainability is also fostered by the European Union (EU) together with resilient and human-centered production for Industry 5.0 [10]. In compliance with both demands, circular manufacturing provides the means for increasing resource efficiency and reducing the use of natural resources. Moreover, it will aid in transition to a circular economy where the life cycle of products is extended using the recycled materials from a discarded product that retain their original quality [11, 12].

Digital twin (DT) [13] situated at the center of this digital transformation provides a gate for Metaverse in order to reach to the physical world [3]. Beginning with a virtual model of a product, it became a pillar technology enabling seamless connection of virtual and physical worlds in Metaverse platforms with IoT [14]. In the case of manufacturing, a DT may simulate any real-world object, which is not necessarily just being the product itself. It may simulate all other actions of production related processes and services. Moreover, DT does not only determine but also proposes solutions related to the real-world system [15]. Thus, the processes of manufacturing may have separate DTs or DT stages for consolidation subsequently in several contexts. Each stage then exploits the product's life cycle beginning with design, then following with prototyping, testing, production, usage, and recycling steps [16].

Therefore, creation of a system that can store and share all relevant information throughout a product's life cycle is an evident requirement to follow-up the digitized information attached to the product. The so-called digital product passport (DPP) introduced by the European Commission (EC) is a digital document that may be regarded as a CV of a product, which will help to acknowledge information of the product from design to the end of life. Thus, being a dynamic information structure, a DPP shall enable information capturing and data sharing in a standardized manner among the actors involved.

The importance of this work resides in managing the product life cycle with a DT connected with an informative acknowledgement mechanism of DPP. Therefore, the purpose of this study addresses digital transformation in industry by taking DT at the core and using DPP to reveal information for screening. While digitization might seem like a straightforward process, e.g. converting some real-world data to digital counterparts, there are many challenges that need to be considered for a product and related manufacturing processes. The accessibility of various sensors for manufacturing processes limits the digital transformation while the production phases are ongoing. To alleviate this problem this study offers an in-laboratory factory model applicable to pursue data flow with DT in order to demonstrate and validate both existing and forthcoming new manufacturing designs. This work also proposes to combine DT with DPP for further upcoming regulations for sustainability requirements.

1.1. Challenges and Motivations

In this work, some of the major challenges are addressed as following:

A DT is not limited to achieving a digital transfer of a product. DT also helps to acknowledge requirements for sustainability concepts. When combined with a new way of carrying product information in the life cycles of products, i.e. DPP, enabling a circular economy and greener impact becomes verifiable.

The scientific literature on those concepts is still limited. Unprecedented ideas and application areas are required while establishing new usages of DT and DPP. Standardization efforts help to reveal the relationship on identification of knowledge to be used and then to generalize them for further innovations. Therefore, companies may benefit from this work for a better understanding of DPP. Besides, they will literate digital transformation while implementing the digital technologies embedded in the DPPs.

1.2. Contributions

Our contributions are listed below, taking into account the above considerations:

- A brief summary of DTs and DPPs, including their relations for sustainable, green manufacturing and circular economy.
- Recommendations of using DTs with DPP within three major phases based on our in-laboratory factory model introduced to overcome the difficulties in obtaining data from a real factory.
- Disclosing the dissemination and standardization efforts, identification of some follow-up keywords and discussing opportunities for future research.

1.3. Outline

The rest of the paper is organized as follows: Section 2 presents the related preliminary concepts of DT and DPP with a discussion about the benefits of a DPP for a sustainable circular economy. Section 3 introduces the proposed method, an observable way of combining DTs with DPP in a laboratory model. It also lists requirements for achievement and realization. Finally, Section 4 concludes this study and presents some of the future ideas and keywords.

2. Related Work

2.1. Digital twin

2.1.1. *Concept*

Grieves provided the first characterization of the concept by defining the constituents of DT as real-world, imaginary (virtual) world, and the data/information exchange flow connecting these two worlds [13]. The concept offered for product lifetime management has been extended throughout the years to include not only the product itself but also all physical systems, their environment, and related processes [17]. It was one of the emerging technologies that became at the peak of the expectations in the Gartner curve in 2018 [18].

The ever-increasing value of DT after commencement within smart systems is due to its capability to comply with the digitization efforts. The real-time monitoring and control of the real-world from a distance may be the trivial use of DTs. Beginning from the Computer Aided Design (CAD), Computer Aided Manufacturing (CAM) tools; creation, analysis, manufacturing processes of products have already been computerized and thus related information has been transformed into digital domain. Besides, DT also includes scenario and risk assessment referring to the virtual world where many possible solutions may come up and tested for predictive purposes. One of the major outcomes of using DTs is to enable more efficient, safe, and informed decision support systems built upon considering many scenarios and what-if actions. Personalization and better documentation capabilities resulted with wide applicability of DTs in various areas, suited for diversified purposes [19]. A visualization of the real state of a manufacturing system could substantially improve benefits such as reliability and maintenance while preventing faults through simulations leading to cost effective solutions [7].

Although there is not yet a consensus on the organization of a DT, the building blocks of a DT can be formed in groups as: IoT or IIoT solutions for retrieving data from various physical product or processes; cyber-physical systems for monitoring, controlling and transferring physical sources into digital domain; solutions for dense computations performed with cloud or edge computing, either locally or globally; and solutions incorporating all aspects of data with ML and AI techniques [20]. Those communication, control, computing, and cooperation aspects (i.e., 4Cs) combined in a DT reveal the importance of DT and places it at the center of all schemes.

2.1.2. *Extensions*

The great breakthrough of DT models appeared when a manufacturing industry extended to include operations and services. They do not focus only to create a virtual representation of a prototype, a product, or a physical system but they rather target a more general concept of physical/virtual reality combination where physical systems, environment and related processes are considered together [17]. This reliable bridge of physical and digital domains can completely describe any behavior by means of data, thus it may provide all influencing factors and relationships that are yet unrevealed.

The role of DTs in digitalization helps to widen the perspective of production together with circular economy and sustainability concepts, using the popularized term “green” that is added to all relevant terminology, as in green DTs [21]. In theory, smart manufacturing should be green and sustainable [8]. Consequently, DTs improve the shortcomings of traditional manufacturing and maintain a digitizing platform converting conventional manufacturing to a smarter and hopefully a greener one.

As a result, there have been efforts to standardize DTs under the International Organization for Standardization (ISO). The ISO 23247, defines a DT framework for manufacturing to support the creation of DTs of observable manufacturing elements including personnel, equipment, materials, manufacturing processes, facilities, environment, products, and supporting documents [22]. It is based on the IoT reference architecture and extends it further to define entity-based and domain-based models with a functional view for DTs in manufacturing. When

the standardization activities from other bodies such as ETSI and IEEE join together, a generally accepted standard will be available.

2.2. Digital product passport

2.2.1. Concept

Circular manufacturing is providing the means for increasing resource efficiency and reducing the use of natural resources. The information sharing mechanisms among the business and industry stakeholders, public authorities and consumers will increase awareness. It is indisputable that the digitalization positioned within DTs accelerates the shift to a more sustainable circular economy with ecological green designs. With these ideas, the EU as a regulatory body promotes the use of DPPs by its Eco-design for Sustainable Products Regulation (ESPR) [23, 24].

A product is simply composed of many components and subcomponents that are referred to as ingredients. In the perspective of sustainability, the product life cycle considers the product as a whole and thus it comprises of four action phases. It begins from the initial idea and design phase with provisioned requirements. Following is the production phase where the product is manufactured with fewer possible sources with highest efficiency, including resourcing of ingredient materials. Third stage is the operations phase, where the product is operated or used by the concerned parties including end-users while assuring materials to have longer duration in their best form. Final point is the disposal phase where no use of the product is available afterwards, additional actions should be incorporated [16].

A simplified view of the product life cycle conceptualized with the 10-R strategies framework [11] is presented in Figure 1. It describes strategies beginning from the smarter use of products and their manufacturing. Higher levels deal with extending the life of the product and its subparts. At the highest level the objective reaches utmost in achieving recycling and further possible usage of subcomponents, ingredients, or materials.

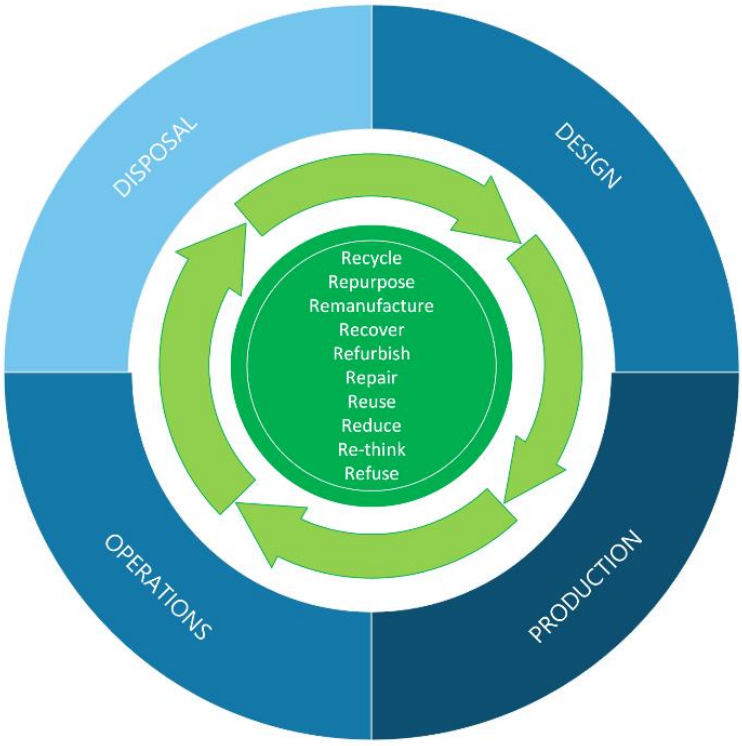


Figure 1. *Product life cycle with 10-R strategies framework.*

A DPP is a digital information document of a product that provides knowledge on the given four levels of life cycle stages considering environmental sustainability. The aim is to establish an extensive and reliable knowledge transfer through the value chain. Providing easily accessible data by scanning a data carrier, such as a watermark

or a quick response (QR) code, it helps competent businesses and consumers to make informed choices, while public authorities are able to perform better regulation controls [24, 25].

A representation of DPP stakeholders is illustrated in Figure 2. The manufacturing industry and business part (shown on the upper left) manufactures their products by complying with the rules of public authorities (shown on the upper right) and serves them to users (shown below). The other actors supporting each stakeholder such as the material suppliers and service providers can be included in the industry part, while other bodies such as standardization and task groups may support the authorities. The corresponding digital information is then available to all stakeholders through a smart phone with an accessible QR code. It serves not only to any production or legislative stakeholders but also to any end-user requiring information about the product to check its quality, compliance with the standards, etc. In brief, DPP lies at an intersection of all.

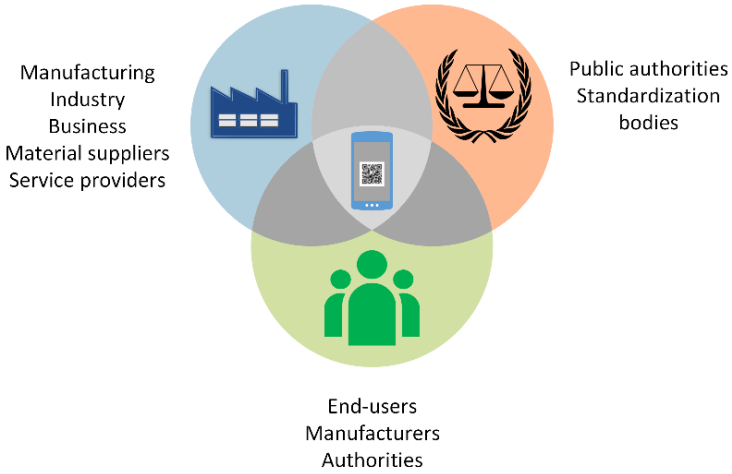


Figure 2. *An illustration of DPP stakeholders.*

2.2.2. Envisaged main benefits of DPPs

Based on the requirements of a more sustainable circular economy with green or eco-design impact, there are many benefits of using DPPs. In compliance with the design of a DPP, it increases transparency such that each collaborator will trace unified information reliably at each stage of the value chain loop. The access and sharing of key product related information would have an impact in transition to a circular economy [26]. Although DPP is still in the pre-conceptual phase and it will more likely display static information, the DPP tool will enable it to follow many eco-design requirements not limited to the product itself but also extended to the sustainability of the product and related resources [27].

The awareness invoked by issuing DPPs will bring advantages on product durability and reliability, possibilities of evoking some of the R strategies from the 10-R model, displayed at the center in Figure 1. Expected major outcomes will be due to carbon footprint reduction and environmental impact tracked by product environmental footprint (PEF) indicators [28], boosting material and energy efficiency by optimizing product design [23]. Moreover, when served as a Product-as-a-Service (PaaS) activity, it will help consumers in making sustainable choices while allowing authorities to verify compliance with legal obligations including trade rules, tariffs, and taxes [29].

On the other hand, DPP offers to be a tool for companies to monitor reaching and administering their sustainability levels, to develop business models enabling resource optimization as well as energy efficiency strategies extending their product lifetimes. The viability of DPPs promotes removing barriers between product manufacturers and end-users, making nimble decisions for better economic and environmental impact in a dynamic digital world.

2.2.3. DPP implementation efforts

As being a new tool under development, there are many efforts in realization of DPPs while overcoming their challenges. First of all, different circular economy activities have potential conflicts [30]. Some of the product related information might be unavailable to the stakeholders [25]. Therefore, it requires an orchestration of DPPs within and between industrial ecosystems while sharing pre-defined digitized information [31].

Secondly, a unified approach for each industry may not be possible or feasible. The duration and updating schedule about the collected data of a product's life cycle is still not clearly defined. Besides, there will always be confidential business information which should be addressed in security and privacy aspects of DPPs [29]. As this task is performed by information technologies using the Internet, solutions require deployment of digital solutions such as IIoT, Blockchain, and Distributed Ledger Technologies (DLT), etc., supported with a Metaverse interface [25, 32].

A recent work [33] lists the consolidated requirements for DPP systems based on system and software quality assessment, structured into eight sections: Legal obligations, functional suitability, security, accessibility, interoperability, modularity and modifiability, availability, and portability. It should be acknowledged that the ESPR provides a comprehensive overview of the requirements of the upcoming DPPs, both in the main text and in the annexes. It will also empower consumers in green transition [24]. Considering DPP as a complex socio-technical system of systems, a DPP Ecosystem is recently defined in order to describe the network of organizations and technologies [34].

Furthermore, a solution underway comes with the CIRPASS project [35], funded by the EC under the Digital Europe Programme. It aims to help in creating a concept for the DPP, demonstrating benefits and roadmaps for its deployment. An understanding of cross-sectoral DPP will build a common source of information for benchmarking purposes.

3. Proposed Work

It is obvious that there are many challenges in representing a real-world system into a digital system. Metaverse is handling some of the interaction issues especially among humans and the environment. Interactions within machines can be both controlled and visualized using DTs. Besides, any action in the digital world or in Metaverse has a consequence in the real-world. Thus requirements in either world must meet with limited resources. The actions should not be performed for today's short-sighted analog-to-digital conversion vision but they have to be over the horizon where the world will definitely be born-digital.

In terms of product manufacturing, digitalization offers a wide range of benefits linked with the circular economy with green or eco-friendly designs acquiring best performance in a sustainable way. Thus, the application of DT and DPP is envisaged to handle the main part in product activities from the design to the end of life. Therefore, they will become a foundation of major business interactions in the digital world when digital transactions, tokens or crypto-currency circulations are considered.

Consequently, one of the purposes of this work is to increase awareness of DT and DPP which will jointly become the leading actors in a digitized industry. However, implementation of these two concepts is not straightforward. Both concepts rely on the information gathered from the real system, particularly from the product itself or from the processes of manufacturing. One of the crucial points is how to access to the real-world data in an operating manufacturing factory. The sensors play the main job capturing data from the real system, however the working environment may cause many difficulties in retrieving data. The collected relevant information needs to be transferred, preferably by the locally displaced IoT devices. Then they become available to be used irrespective of location. Thus, the communication issues should be solved with local, edge, and global computational sources with storage capacity assessments. The other important bottleneck comes with the security issues which business specific information needs to be supervised, either publicly shared, or highly secured.

In order to investigate and offer solutions to overcome those limitations, we propose to use a smart model of a factory constructed in a laboratory while we are building up DT and DPP tools for an under-inspection product and processes. The main idea behind this solution is the ease of accessing the sensors and data in a laboratory environment. The simulation and validation of processes can be quickly demonstrated and verified. Also it brings a dynamic, extensible and adaptive framework where updates of DT and DPP can be performed easily, recalling that the standards and benchmarking are still under development. An illustration of our proposed scheme is presented in Figure 3. The real-world information can be directly transferred to DPP, or through the model factory and its DT while the information flow is bi-directional among these blocks.

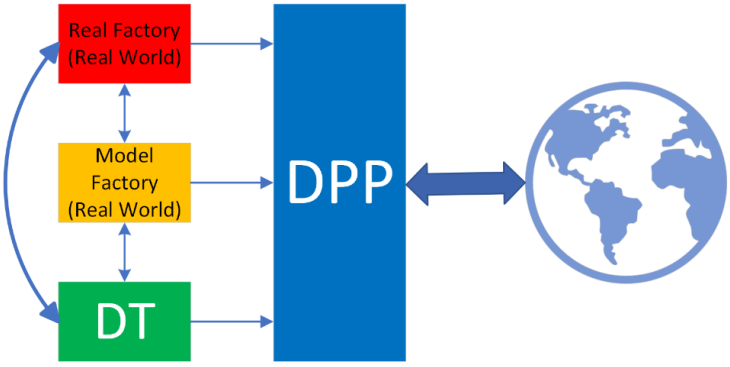


Figure 3. *Information flow transfer from the factory, its model, its DT to DPP.*

In order to implement a DPP of a smart model factory imitated in a laboratory, a holistic three-phase flow is envisaged based on the aforementioned literature screening with DT residing at the core. In each recommended phase, the main motivation is the assignment of DTs that are able to deliver key information to a DPP.

3.1. Recommendations

3.1.1. *Phase 1: Design to manufacture*

The objective of this phase is to create design models of products and manufacturing systems using different types of software programs. Beginning from the CAD, CAM tool drawings, the material information building up the product is required such as the origin and physical properties of the raw material(s). This information is also the first entry for DPP to follow up the product ingredients, materials, usage, and life cycle.

The product manufacturing processes and information flow for building up DTs are also considered in this stage. The detailed drawings are indispensable for building smart manufacturing systems. Initialized with battery, textile, electronics being among the dedicated focus areas [28], extension to other sectors that have high impact on resources with high circularity potential is foreseen. Recently, an example considered in our model factory laboratory for production is marble stones based on a real marble producer production line. The oven models of the production line will enable data acquisition and tracking, such as temperature and humidity of the environment and the product. A selection of main processes or detailed product information for various stakeholders may become available to be released for its DPP. A selection of some utilities or a simplified version of a DPP tool might be considered for the end-users as well.

3.1.2. *Phase 2: Digital twining*

This phase considers building DTs of each production step including machines and supporting system environments. As DT is a software, it is crucial to determine the type, capability, interoperability of the software selected. Steering sensor information into DT and reciprocally sending control commands for actuators in the factory model is the main concern of this phase. Initially defining and further following the key performance indices (KPIs) and PEF indicators become feasible. The ease of definition and calculation of KPIs will help in building scenarios for time, job, human, and energy resource management, as well as reporting with respect to each contributor. Performing what-if actions and exploring the outcomes with ML and AI techniques are the main advantages of a DT. Using the DT-DPP connection, any process specific data or any of the resource management issues are shared for legal regulation controls while they enable access for certain stakeholders. This access is foreseen to be realized with Metaverse connectivity using AR, VR, and XR user-interfaces. On the other hand, Industrial Metaverse (IM) [36, 37] is already available to inform workers on the machines that they use, or to train them with those interfaces before beginning to work in a real environment. Therefore, this stage covers IM that will be extended to integrate content generation with the physical industrial economy [38]. Models with a service-oriented DT within Metaverse-as-a-Service (MaaS) framework can be one of the DT-based solutions [37].

3.1.3. Services, consumption, recycling

This phase deals with the usage of the product till its disposal. This stage collects information from end-users in order to maintain each of the R framework concepts. The duration of the product or any parts supporting the product is important to determine the end of life and then initiate recycling processes. Tracking codes given to the product, displaying necessary time and place information, the profit of using the product denoted by KPI and PEF indicator levels should be displayed by DPP to inform users for increasing awareness in product sustainability. With AI support, all after-production phases can be handled with online-offline scenarios to be performed afterwards.

The proposed phases are organized in the Table 1 according to the general requirements, exemplar DPP display information and associated examples for further clarification.

Table 1. *Recommendations for implementing a DPP*.

Phase	Requirements	DPP Information	Examples
Design to manufacture	Digitized designs; manufacturing system; process information	Product and ingredients; life duration; key production processes; designer, producer, brand information; relevant codes	Origin, physical properties; picture or drawing; brand logo; NACE code; product code, certificates
Digital twining	DT software with options of AR, VR, XR, Metaverse interfaces (e.g., IM)	Product-based information; relevant sensor information; process information	Battery (type, chemicals, metal plate, dimensions, pH)
Services, consumption, recycling	DT software augmented with ML and DL for tracking of data, scenario design	Tracking code; regulations, standards; important dates and conditions; cost; energy savings; services	QR code; standard number; production and best before date; end of life; recovered or recycled material ratio; level of energy saved, carbon footprint; transportation and retail services

3.2. Technical Challenges and Discussions

The phases in Table 1 summarizes the main stages expected in the life cycle of a product. However, there remain many technical problems to tackle for each requirement. The major challenges are grouped and recapitulated as follows.

The first challenge is to obtain the drawings of the product and manufacturing system in a digital format. As there are many CAD/CAM models initiating the design step of a product, convertibility/operability of software files is crucial. Visualization and rendering of 3D animations for DT purposes with various software tools needs to be handled properly. This must be followed by a production system digitalization. Likewise, the production system models are required with individual part drawings to reflect their counterparts in DT. Creating a platform independent tool is foreseen as the best way of overcoming this challenge. Open-source software might be combined in a suitable framework [39].

The second challenge is to obtain data from the real factory. DTs bring their advantages enabling data from sensors then transferring them to the IoT devices. However, real data acquisition is always cumbersome. It requires proper sensors fitted for the purposes. Nevertheless, a sensor model and its DT should be attainable. There comes our proposal that may help to determine which of the data is essential and which can be negligible for the ease of achieving identifiable targets. By the model in a laboratory, these demands can be more easily pursued with a nearby DT. Establishing the connection of manufacturing system data to DT remains mostly production-oriented. Therefore, the solution of this challenge resides in describing the best or preferably the most useful model. Granularity is the key concept to determine the level of accuracy for sufficient and required resolutions. Digitized data always includes rounding errors that should be properly handled.

The storage of data in the cloud systems must be handled where information safety is also entangled to accessing the relevant information with separation of anonymous and authorized users. An additional challenge comes from these security issues that should be considered through all phases. The so-called cyber-security is also one of the major challenges of the digital world including Metaverse. Blockchain and other technologies such as DLT, offer a traceable and verifiable way of managing security issues as well as privacy or sharing options.

The next biggest challenge is due to the usage of the product. The consumption, recycling or PaaS activities may not be fully realized in a laboratory environment. However, MaaS activities will help to embrace DTs and Metaverse which enhanced creation of further digital clones. Then, the sustainability concepts can be tracked through the consumption amount values, time or other KPI information within DT and displayed through DPP.

Nevertheless, with the capability of creating and simulating the scenarios supported by open-software and AI tools [40], DTs are able to handle those challenges. The only challenge left here is to elaborate more realistic scenarios or what-if actions fitted for the real-world.

The final challenge is to incorporate and display data in a DPP such that all the stakeholders can reach appropriate information. As of now, a simple QR code supplying an entry of information coupled within smartphone applications seems to be the common solution.

As a summary, the abovementioned challenges grouped in four D-domain sections, i.e., design, data, deploy, and DPP, are illustrated in Figure 4 with respect to their easiness of achievability. DPP user interface with QR code access is relevantly fulfilled while deploying the sustainability concepts with AI tools requires much more effort.

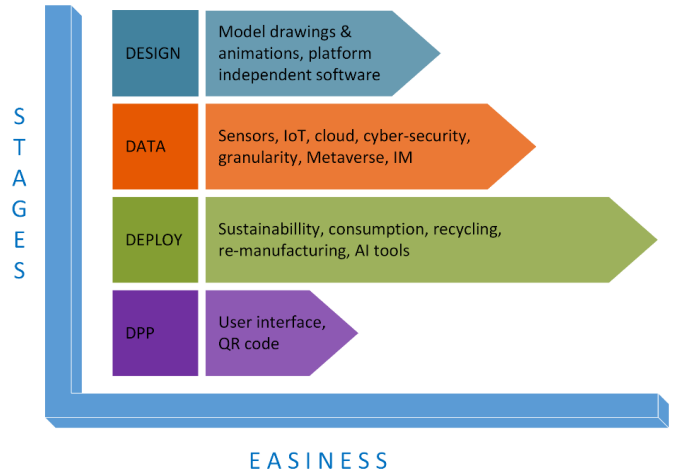


Figure 4. *Technical challenges attempting to satisfy the recommendations of DPP.*

4. Conclusions

The digital transformation has many aspects. In terms of manufacturing products, it may vary from simple digital twinning of a product to a more general benefit of circular and green economy targets with DPP attached to the product and related product-spaces. While the DTs effect in digitalization is obvious, this study demonstrates that DPP becomes inseparable from DTs from a circular economy and sustainability point of view. Therefore, by combining DTs with DPPs, digitalization efforts will be expected to be eco-friendlier or greener, while economic and social benefits will increase in due time.

As summarized in the previous sections, the challenges remain to be solved. Beginning with the digitalization of the manufacturing system, the technical achievements in image processing, computer vision, and related software tools will lead to a better representation of a real-world, thus increasing Metaverse experience. Meanwhile, an indispensable outcome of this digitalization relies on the data acquired from the sensors. The gathering, processing, and storing data in an efficient way with IoT will become the main road of next generation industrial systems with the transformation of systems compatible with the AI technologies.

Therefore, this study attempts to offer solutions by bringing those problems in a laboratory. With a nearby solution where sensor-based information is reliably controlled and where DT is built as close as to the model framework, digitalization aims will be more fulfilled. Besides, any data-oriented solution can be applied such as KPIs where ML and DL techniques aid. Our solution is of course neither covering all issues nor complete in realization of a fully functional DT serving for a DPP. However, it will live and grow up with the uprising of those techniques. Beginning from a simple interface of demonstrating the product origin information, it will cover all necessary and relevant information agreed upon with the upcoming standards scheme. It will always be possible to make updates, extensions or modifications in this framework. The potential avenues for further investigation of concern is relatively wide, ranging from a simple IIoT sensor model representation with its DT to a general system block analysis for prevention of system errors and prediction of failures.

While the main objective of this study is to underline the importance of DT for DPP, the limitations bounded the work in some aspects. First of all, this study remarks that the literature on DT and DPP is rather new. Standards, benchmarks and use-cases are still at their primitive stages and necessitates more groups to gather efforts to predict

and manage those shortcomings. As there are many existing technologies available to be used in digitalization, orchestration is required. Moreover, the technological readiness levels for each technique are different. However, it is expected to be merged to higher levels in the following years while some of the standards will be set. This inevitably includes the use of IM and related technologies of Metaverse. This study proposing a laboratory-based solution contributes to the wider discussion on the digitalization specified to DT and DPP, as the future is ineluctable digital.

Declaration of Interest

The authors declare that there is no conflict of interest.

Acknowledgements

The authors thank Smart Factory Systems Application & Research Center (AFSUAM). This study was supported by Scientific and Technological Research Council of Turkey (TUBITAK) under the Grant Number 123E288. The authors thank to TUBITAK for their supports.

References

[1] H. Wang, H. Ning, Y. Lin, S. Dhelim, F. Farha, J. Ding, and M. Daneshmand, "A survey on the Metaverse: The state-of-the-art, technologies, applications, and challenges", IEEE Internet of Things Journal, vol. 10, no. 16, pp. 14671-14688, 2023.

[2] K. G. Nalbant and S. Aydın, "Development and transformation in digital marketing and branding with artificial intelligence and digital technologies dynamics in the Metaverse universe", Journal of Metaverse, vol. 3, no. 1, pp. 9-18, 2023.

[3] K. Li, Y. Cui, W. Li, T. Lv, X. Yuan, S. Li, W. Ni, M. Simsek, F. Dressler, "When Internet of things meets Metaverse: Convergence of physical and cyber worlds", IEEE Internet of Things Journal, vol. 10, no. 5, pp. 4148-4173, 2023.

[4] S. -M. Park and Y. -G. Kim, "A Metaverse: Taxonomy, components, applications, and open challenges", IEEE Access, vol. 10, pp. 4209-4251, 2022.

[5] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research", IEEE Access, vol. 8, pp. 108952-108971, 2020.

[6] C. Yang, W. Shen, X. Wang, "The Internet of things in manufacturing: Key issues and potential applications", IEEE Systems, Man, and Cybernetics Magazine, vol. 4, no. 1, pp. 6-15, 2018.

[7] Z. Lv, "Digital twins in Industry 5.0", Research, vol. 6, pp. 0071, 2023.

[8] L. Li, B. Lei, and C. Mao, "Digital twin in smart manufacturing", Journal of Industrial Information Integration, vol. 26, pp. 100289, 2022.

[9] United Nations, "Sustainable development goals", [Online] Available: https://sdgs.un.org/goals, [Accessed 15 February 2024].

[10] M. Breque, A. De Nul, L. Petridis, "Industry 5.0 - Towards a sustainable, human-centric and resilient European industry", Policy brief, Publications Office of the European Union, European Commission, Directorate-General for Research and Innovation, 2021.

[11] J. Potting, M. Hekkert, E. Worrell, and A. Hanemaaijer, "Circular economy: Measuring innovation in the product chain", Policy Report, PBL Netherlands Environmental Assessment Agency, The Hague, 2017.

[12] C. Turner, J. Oyekan, W. Garn, C. Duggan, and K. Abdou, "Industry 5.0 and the circular economy: Utilizing LCA with intelligent products", Sustainability, vol. 14, pp. 14847, 2022.

[13] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication. White paper, 2014.

[14] R. Asif and S. R. Hassan, "Exploring the confluence of IoT and Metaverse: Future opportunities and challenges, IoT, vol. 4, pp. 412-429, 2023.

[15] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data", Int J Adv. Manuf. Technol, vol. 94, pp. 3563-3576, 2018.

[16] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models", Proceedings of the IEEE, vol. 108, no. 10, pp. 1785-1824, 2020.

[17] E. VanDerHorn and S. Mahadevan, "Digital twin: Generalization, characterization and implementation", Decision Support Systems, vol. 145, pp. 113524, 2021.

[18] Gartner press release, "Gartner identifies five emerging technology trends that will blur the lines between human and machine", [Online] Available: https://www.gartner.com/en/newsroom/press-releases/2018-08-20-gartner-identifies-

five-emerging-technology-trends-that-will-blur-the-lines-between-human-and-machine. [Accessed 15 February 2024].

[19] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective", IEEE Access, vol. 8, pp. 21980-22012, 2020.

[20] L. Lattanzi, R. Raffaeli, M. Peruzzini, and M. Pellicciari, "Digital twin for smart manufacturing: A review of concepts towards a practical industrial implementation", International Journal of Computer Integrated Manufacturing, vol. 34, no. 6, pp. 567-597, 2021.

[21] J. Deuse, D. Wagstyl, V. H. Moreno, M. Polikarpov, R. Wöstmann, and F. Hoffmann, "Green digital twins in the product life cycle", Schriftenreihe der Wissenschaftlichen Gesellschaft für Arbeits- und Betriebsorganisation (WGAB) e. V., pp. 167-186, 2023.

[22] ISO 23247, "Automation systems and integration - Digital twin framework for manufacturing", 2021.

[23] T. Götz, H. Berg, M. Jansen, T. Adisorn, D. Cembrero, S. Markkanen, and T. Chowdhury, "Digital product passport: The ticket to achieving a climate neutral and circular European economy?", University of Cambridge Institute for Sustainability Leadership (CISL) and the Wuppertal Institute, Cambridge, UK: CLG Europe, 2022.

[24] European Commission, "Proposal for a regulation of the European Parliament and of the Council establishing a framework for setting ecodesign requirements for sustainable products and repealing directive 2009/125/EC", European Commission: Brussels, Belgium, 2022.

[25] S. Nowacki, G. M. Sisik, and C. M. Angelopoulos, "Digital product passports: Use cases framework and technical architecture using DLT and smart contracts", 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Pafos, Cyprus, pp. 373-380, 2023.

[26] S. F. Jensen, J. H. Kristensen, S. Adamsen, A. Christensen, and B. V. Waehrens, "Digital product passports for a circular economy: Data needs for product life cycle decision-making", Sustainable Production and Consumption, vol. 37, pp. 242-255, 2023.

[27] T. Adisorn, L. Tholen, and T. Götz, "Towards a digital product passport fit for contributing to a circular economy", Energies, vol. 14, pp. 2289, 2021.

[28] L. Saari, J. Heilala, T. Heikkilä, J. Kääriäinen, A. Pulkkinen, and T. Rantala, "Digital product passport promotes sustainable manufacturing: Whitepaper", VTT Technical Research Centre of Finland, 2022.

[29] J. Walden, A. Steinbrecher, and M. Marinkovic, "Digital product passports as enabler of the circular economy", Chemie Ingenieur Technik, vol. 93, pp. 1717-1727, 2021.

[30] T. Rantala, L. Saari, M. Jurmu, K. Behm, J. Heikkilä, A. Jokinen, R. Palmgren, E. Paronen, K. Rainio, K. Valtanen, M. Vierimaa, and M. Ylikerälä, "Digital technologies for circular manufacturing", VTT Technical Research Centre of Finland, VTT White Paper, 2023.

[31] D. J. Langley, E. Rosca, M. Angelopoulos, O. Kamminga, and C. Hooijer, "Orchestrating a smart circular economy: Guiding principles for digital product passports", Journal of Business Research, vol. 169, pp. 114259, 2023.

[32] S. B. Far and A. I. Rad, "Applying digital twins in Metaverse: User interface, security and privacy challenges", Journal of Metaverse, vol. 2, no. 1, pp. 8-15, 2022.

[33] M. Jansen, T. Meisen, C. Plociennik, H. Berg, A. Pomp, and W. Windholz, "Stop guessing in the dark: Identified requirements for digital product passport systems", Systems, vol. 11, pp. 123, 2023.

[34] M. R. N. King, P. D. Timms, and S. Mountney, "A proposed universal definition of a digital product passport ecosystem (DPPE): Worldviews, discrete capabilities, stakeholder requirements and concerns", Journal of Cleaner Production, vol. 384, pp. 135538, 2023.

[35] Cirpass project, [Online] Available: https://cirpassproject.eu. [Accessed 15 February 2024].

[36] C. Chi, D. Lin, R. Ramadoss, Y. Yuan, Z. Yin, C. Luo, W. Chen, B. Yang, L. Wei, R. Ma, "White paper - The industrial Metaverse report", The Industrial Metaverse Report, IEEE, pp. 1-20, 2023.

[37] D. Mourtzis, "Digital twin inception in the era of industrial Metaverse", Frontiers in Manufacturing Technology, vol. 3, pp. 1155735, 2023.

[38] T. Bohné, C. Li, and K. Triantafyllidis, "Exploring the industrial Metaverse: A roadmap to the future", World Economic Forum, Briefing Paper, 2023.

[39] J. A. Fortoul-Diaz, L. A. Carrillo-Martinez, A. Centeno-Tellez, F. Cortes-Santacruz, I. Olmos-Pineda, and R. R. Flores-Quintero, "A smart factory architecture based on Industry 4.0 technologies: Open-source software implementation", IEEE Access, vol. 11, pp. 101727-101749, 2023.

[40] J. Soldatos (Editor), Artificial Intelligence in Manufacturing, Enabling Intelligent, Flexible and Cost-Effective Production Through AI, Springer, 2024.

Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 4, No. 1, pp. 22-32, 2024
Research Article
Received: 17/04/2024 Accepted: 03/06/2024

Classification of Fake News Using Machine Learning and Deep Learning

Muhammed Baki ÇAKI [1], Muhammet Sinan BAŞARSLAN [2,*]

Abstract

The rapid advancement of technology has led to an increase in the spread of fake news, which has a detrimental effect on people in various fields, particularly in their daily lives. The negative impacts of fake news can be mitigated through the use of artificial intelligence. The development of AI technologies has made the detection of fake news a prominent area of research within natural language processing. This study explores style-based fake news detection using machine learning and deep learning methods. The texts were processed using natural language processing techniques and investigated with different models on the open-source ISOT dataset. The models utilised text processing, text representations (TF-IDF, word2Vec), and different machine learning (ML) methods (K-Nearest Neighbor, Naïve Bayes, Logistic Regression) as well as Long Short-Term Memory (LSTM). The performance of the models was evaluated using accuracy (Acc), precision (P), recall (R), and F1-score. Among the tested models, the LSTM model demonstrated the highest performance, with an accuracy of 99.2%. The development of state-of-the-art methods for text representation and classification, including preprocessing in text classification, and the application of these methods in practical settings can significantly reduce the prevalence of fake news.

Keywords: Deep learning, Fake news detection, Machine learning, Style based detection.

1. Introduction

Artificial Intelligence (AI) is a field that is divided into many sub-headings with its potential and is the subject of many researches in order to produce better solutions to our problems. Natural Language Processing (NLP), Machine Learning (ML) and Deep Learning (DL) are the main sub-topics of AI. Fake news is one of the problems we want to solve. The rapid spread of false content produced for various reasons causes social and economic damage to individuals, organizations and societies. This problem is growing with the increasing speed of communication. Misinformation and disinformation have negative effects on society. Therefore, new and effective methods are needed to detect and prevent fake news.

The main purpose of our study is to contribute to existing studies to find solutions to this problem with AI. In order to classify and distinguish between fake and real news, linguistic features of news texts are processed and analyzed with NLP. Then ML and DL models are built. After the models are trained, prediction is made for the given news text to be real or fake. In this study, various models are built using different NLP techniques and ML algorithms and the results obtained are analyzed. The results of the study show that NLP and ML models have a significant potential in fake news detection.

In the second part of the study, similar studies in the literature are presented. The third section discusses the dataset, preprocessing, vectorization, ML, DL and performance criteria. The fourth section describes the experimental setup. Section five presents the experimental results. The discussion and conclusion in sections six and seven provide an overall assessment and future works.

2. Related Works

Similar studies on fake news detection in the studied ISOT dataset will be described in this section. In their study, the researchers created models with Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF) and deep neural network. They achieved 91% Acc with neural [1] . After GloVe, the best performance with 92% Acc was obtained by using Linear Support Vector Machine (LSVM) as a classifier [2]. After vectorization with Term Frequency - Inverse Document Frequency (TF-IDF) on ISOT dataset, classifier

*Corresponding author

Muhammed Baki ÇAKI; Istanbul Medeniyet University, Faculty of Engineering and Architecture, Computer Engineering Department, Türkiye; e-mail: muhammedbakicaki@gmail.com; 0009-0005-2651-4047

Muhammet Sinan BAŞARSLAN; Istanbul Medeniyet University, Faculty of Engineering and Architecture, Computer Engineering Department, Türkiye; e-mail: muhammet.basarslan@medeniyet.edu.tr; 0000-0002-7996-9169

models were created for fake news detection with various ML algorithms. Among these models, the best result was obtained with Decision Tree (DT) with 96.8% Acc [3]. They created models with ML methods such as SVM, LSVM, K-Nearest Neighbor (KNN), DT on the data of fake news collected by ISOT and themselves. LSVM gave the highest Acc result with 92% [4]. After vectorization with the Word2Vec method called Maithi-Net, they obtained 97.28% Acc result in fake news detection with this method [5]. They obtained 74% Acc with NB after CBOW [6]. Word2Vec obtained 82.67% Acc with conditional random fields (CRF) classifier after CBOW [7]. They obtained 99% Acc in a study on fake news classification with ensemble learning after TF-IDF [8]. If we look at other fake news studies other than this dataset; They obtained 99.10% Acc in fake news classification model by hybridizing Recurrent Neural Network (RNN) and LSTM on Liar dataset [9]. In their study on the detection of false news in the pandemic, they obtained 96.19% Acc and 95% F1 with the Convolutional Neural Network (CNN) they proposed by optimizing hyperparameters after embedding methods such as GloVe [10].

It is seen that ML methods are frequently used after TF-IDF, Word2Vec text representation methods on ISOT dataset [2] and similar content data related to fake news. After the text representation methods TF-IDF and Word2Vec, which are frequently studied in the literature, ML (KNN, NB, LR), and DL (LSTM) models were created for the classification of fake news.

The contribution of this study to the detection of fake news on ISOT, an open source shared and balanced dataset, is listed below:

- It is investigated which of the popularly preferred text representation methods such as TF-IDF and Word2Vec has more impact on the performance of the models.
- The holdout discrimination results of the models built with classical ML (KNN, SVM, NB, LR) and DL (LSTM) are investigated.

3. Materials and Methods

In this section, dataset, preprocessing, vectorization, ML, DL, and performance criteria will be explained. In this study, a text analysis-based approach is adopted for automatic fake news detection. In this approach, the characteristics of the texts are extracted and processed by NLP methods, then modelled and predicted by ML methods. The field of study is in the combination of ML, DL, and NLP fields in Figure 1.

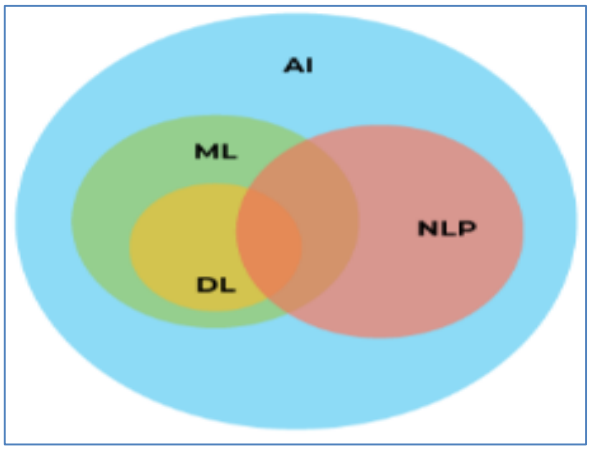


Figure 1. *Fields of AI*

The experimental steps carried out in the study are given in Figure 2.

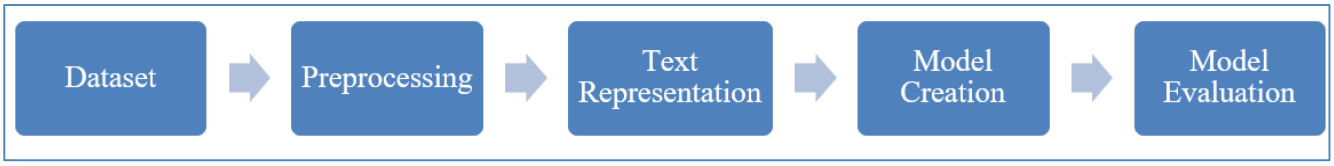


Figure 2. *Study Pipeline*

As seen in Figure 2, the dataset was preprocessed and then subjected to text representation (TF-IDF, Word2Vec). Then the model was created with LSTM, KNN, NB, LR with 75%-25% training-test separation. F1, P, R, Acc were used to evaluate the models.

3.1. Dataset

In the study, ISOT Fake News Dataset [2], which consists of fake and real news data created by the researchers with news collected from the internet between 2016-2017, was used.

The distribution of the dataset, which consists of 44,898 news in total, is given in Table 1. According to the researchers, news with true content was collected from the reuters.com website, while news with false content was collected from various websites marked as unsafe by Polifact [2]. Table 1 provides information about the content of the dataset.

Table 1. *Dataset*

News	Number of articles	Subjects	
		Type	**Articles size**
Real	21417	Government-News	1570
		Politics-News	11272
		Type	**Articles size**
Fake	23481	US News	783
		Left-news	4459
		Politics	6841
		News	9050

Figure 3 shows a sample image from the dataset.

	title	text	subject	date	label
4528	EPA chief says Paris climate agreement 'bad de...	The United States should continue to be "engag...	politicsNews	April 2, 2017	1
10310	BREAKING NEWS: President Trump Announces Major...	President Trump just tweeted out a new policy ...	politics	Jul 26, 2017	0
10937	Trump says New Hampshire win not necessary to ...	U.S. Republican presidential candidate Donald ...	politicsNews	February 7, 2016	1
13470	Kremlin: U.S. sanctions aimed at turning busin...	The Kremlin said on Thursday it was confident ...	worldnews	November 30, 2017	1
19397	MUST WATCH: Kellyanne Conway PUNCHES BACK Afte...	Kellyanne Conway s response to Williams criti...	left-news	Dec 27, 2016	0

Figure 3. *Summary Visualization of the Dataset*

Figure 3 shows the title of the news item, the text of the news item, the date of publication of the news item and the class label of whether it is fake or real. Figure 4 shows the graph of class distribution.

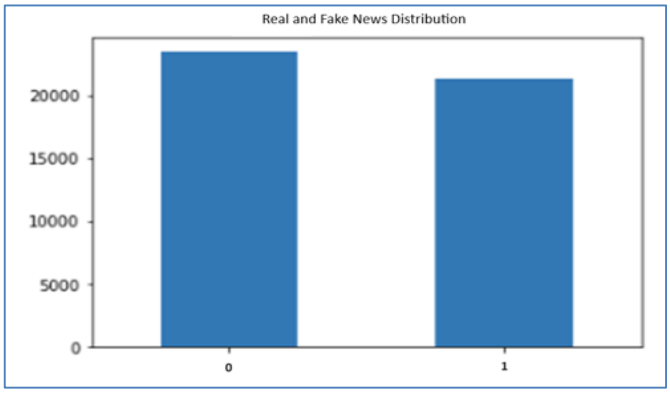


Figure 4. *News Class Distribution*

In Figure 4, according to the class distribution of the dataset; we see that the skewness coefficient is calculated as -0.08 and the kurtosis coefficient is calculated as -1.18. Since the skewness is very close to zero, we can accept the distribution as symmetric. In the light of this result, the dataset is balanced.

3.2. Text Representation

In this section you will find information about text representation.

3.2.1. Term Frequency Inverse Document Frequency

The method called Term Frequency - Inverse Document Frequency is based on the principle of extracting the attributes of the text by weighting each word in the text according to its importance. In this method, the importance of words is determined by analyzing how many times they occur in the examined text and how many times they occur in other texts. The TF-IDF representing the term in sentence t, document d is given in equation (1) [8].

$$TF(t,d) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d}{Total\ number\ of\ terms\ in\ document\ d} \quad (1)$$

D is the collection of all documents (corpus), the addition of 1 to the denominator is to prevent the term from dividing by zero if it is not found in any document IDF is given in equation (2) [8].

$$IDF(t,d) = \frac{Total\ number\ of\ documents\ in\ the\ corpus\ N}{Number\ of\ documents\ containing\ term\ t+1} \quad (2)$$

The TF-IDF in the document is given in Equation (3) [9].

$$TF - IDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (3)$$

According to this equation, frequent occurrence of a word in the relevant document increases its importance. If it is a common word in other documents, it decreases its importance. In this way, stopwords in documents also become unimportant.

3.2.2. Word2Vec

Word2Vec is a method of converting words into vectors of real numbers using artificial neural networks. Words with close meaning are also numerically close in vector representation. In this way, the semantic proximity and context information of the words are kept [11].

3.3. Machine Learning

In this section, ML methods are described.

3.3.1. K-Nearest Neighbors

KNN algorithm is a lazy learning algorithm used in classification and regression problems in ML. In the space where the data points are represented, prediction is made based on the distance of the relevant point to other points. For the classification task, the distances to the k nearest points are calculated. It is predicted as belonging to the class with the least total distance. The reason why it is categorized as a lazy learning algorithm is that there is no learning phase before the data to be predicted arrives. It takes two basic parameters, 'number of neighbours' and 'distance metric' [12].

The number of neighbours is the value 'k', which is also in the name of the algorithm. Distance is calculated with the k nearest neighbours. The distance metric is the algorithm to be used to measure the distance. The most commonly used equation for distance calculation is given in (Euclidean distance) equation (4).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

3.3.2. Naïve Bayes

NB Classifier is a probability-based prediction method that uses a simplified version of Bayes' Theorem in probability. It is often used in classification tasks. Bayes' Theorem allows the calculation of the probability of event A occurring when event B occurs; when the probabilities of event A occurring, event B occurring, event B occurring when event A occurs are known. Its formula is given in equation (5) [13]:

$$P(B) = \frac{P(A) * P(B|A)}{P(B)} \tag{5}$$

In the NB Classifier, the denominator part of the Bayes equation is ignored since the aim is to find the class with high probability instead of finding the exact value. For a two-class classification task, the probabilities of the data belonging to classes X and Y are calculated with the help of the equation. Whichever class the probability of belonging to is calculated to be higher, is predicted to belong to that class.

3.3.3. Logistic Regression

LR is an algorithm frequently used in classification problems in ML. It is based on probability-based class prediction by fitting the data to the logistic function. It is more suitable for binary classification task. Multi-class classification can also be performed. Figure 5 shows a visual of LR [14].

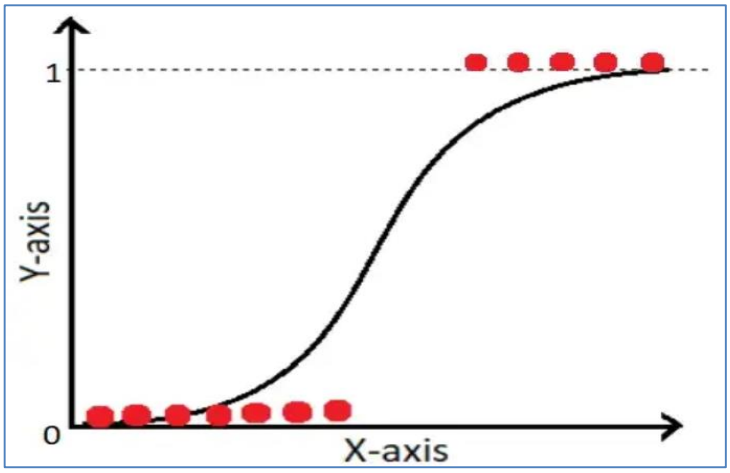


Figure 5. *LR [14].*

3.4. Deep Learning

In this section, LSTM, one of the DL methods used in this study, will be explained.

3.4.1. Long Short-Term Memory

Long short-term memory is a DL architecture that is an advanced version of the RNN model. In RNN, as each output affects the next input, a memory structure is formed. This memory is short term. In long inputs, the effect of past data on the new input decreases rapidly and disappears (gradient vanishing problem). In LSTM architecture, input, output and forget gates are used in addition to RNN. In this way, by creating a short-term and long-term memory structure at the same time, context information can be preserved in long inputs such as paragraphs. LSTM architecture is frequently used in the development of sequence-based prediction systems such as anomaly detection and time series [15]. Figure 6 shows the LSTM architecture.

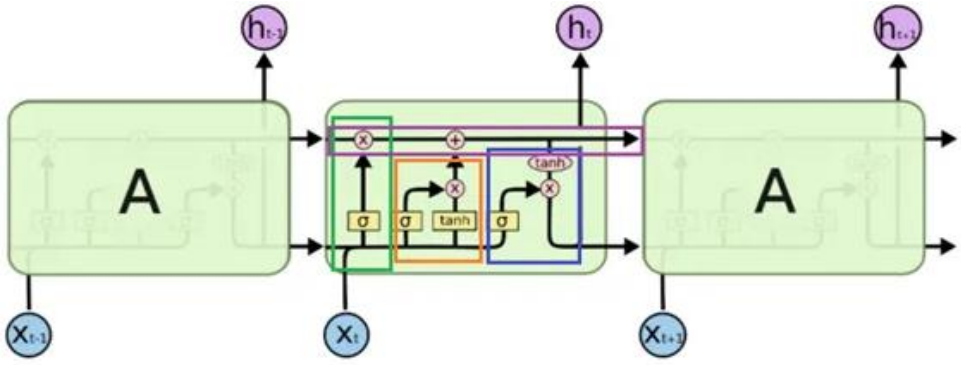


Figure 6. *LSTM Architecture* [15].

3.5. Performance Criteria

In this section, the metrics used in the evaluation of the performance criteria used for model evaluation and the confusion matrix that is used to obtain this metrics are explained.

3.5.1. Confusion Matrix

The confusion matrix is constructed by comparing the predicted and actual values of the test data. In each cell, the total number of samples belonging to that cell is recorded. In this way, the test result of the model can be analysed on a single table [16].

- True Positive (TP) if the true value of the data is positive and predicted as positive,
- False Negative (FN) when the true value is positive and estimated as negative,
- False Positive (FP) when the true value is negative and estimated as positive,
- A negative true value and a negative predicted value constitute True Negative (TN) cases.

Figure 7 shows a visualization of the confusion matrix.

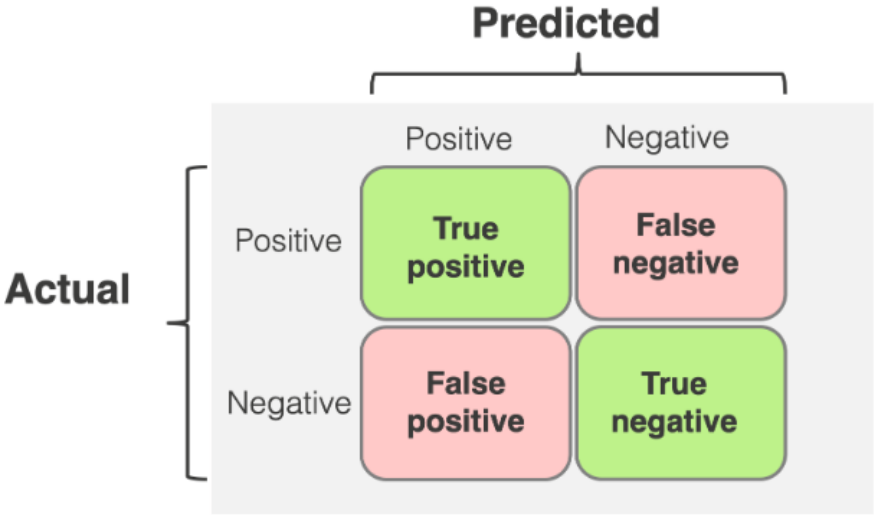


Figure 7. *Confusion Matrix* [17]

Various metrics are calculated to measure how well the created models make predictions. The predictions made by the model are processed to the relevant part in the confusion matrix. Metrics are calculated using the relevant fields in the confusion matrix. The four most important metrics used for evaluating classification tasks are explained

Accuracy:

It is the ratio of the model's correct predictions to all predictions [18-19]. This ratio is given in Equation (6).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision:

It is the metric that measures the success ratio of the model when it predicts the outcome as positive [19]. This metric is given in Equation (7).

$$P = \frac{TP}{TP+FP} \quad (7)$$

Recall:

It is the metric that measures the extent to which the model can accurately detect situations that are actually positive [20]. This metric is given in Equation (8).

$$R = \frac{TP}{TP+FN} \quad (8)$$

F1:

In cases such as measuring high Acc and low P values in an unbalanced dataset, the Acc value may be misleading about the success of the model. F1 metric is obtained by the harmonic mean of P and R values and indicates the stability of the prediction [21]. This metric is given in Equation (9) [18].

$$F1 = 2 * \frac{P*R}{P+R} \quad (9)$$

4. Experimental setup

The work was done in Google Colab [22], which allows the Python programming language [23] to run on a Jupyter notebook. In this chapter, preprocessing of texts, TF-IDF and Word2Vec followed by modelling with ML and DL will be explained.

4.1. Preprocessing

The operations performed within the scope of the study will be explained in this section. Attributes other than text and label have been removed. Label attribute is a binary data type containing 0 for false news and 1 for true news. The preprocessing processes are listed below:

1. The news source at the beginning of real news has been removed.
2. Letters and characters other than the '@' sign were removed from the news texts.
3. News texts were converted to lower case.
4. Stopwords in the news texts were removed.
5. The words in the news texts were stemmed.

4.2. Text Representation

Since the models to be created cannot process text data, the text must be matrixised and given as input to the model. For this reason; TF-IDF with its implementation in sci-kit learn library, Word2Vec text representation methods were used with its implementation in the gensim library. The Word2Vec parameters used in the study are presented in Table 2.

Table 2. *TF-IDF and Word2Vec parameters*

	Parameters	Value
Word2Vec	Vector size	200
	Window	5
	Min count	5
	Methods (sg)	CBOW (0)
	Workers	4

4.3. Model Creation

In the study, a total of seven models were created with TF-IDF and Word2Vec text representation method, KNN, NB, LR, and LSTM. The data to be used to train and test these models are divided into 75% training and 25% test data. In the study, sklearn library was used to create models with KNN, LR, and NB, which are traditional ML methods, and keras library was used for LSTM, which is a DL model.

4.3.1. Machine learning model

In the study, for the creation of ML models, Multinomial NB Classifier and LR Classifier implementations in the sci-kit learn library were used to create models with default parameters. For the KNN model, the implementation of the same library was used and the grid search algorithm was used for hyperparameter optimization.

The TF-IDF vectorization method was tested with the grid search algorithm for the number of neighbours (k) parameter of the generated KNN model for values (1-10) and the optimal value was observed to be 1. It was observed that the optimum k value for the model created with Word2Vec vectorization method was 5. The distance metric parameter was chosen as "euclidean".

4.3.2. Deep learning model

In the study, an LSTM model was created using the LSTM module in the Keras library. The model was optimized by creating and comparing models with different hyperparameters. Used and preferred hyperparameters are shown in table 3.

Table 3. *TF-IDF and Word2Vec parameters*

Output Dimension	Neuron	Dropout	Epochs	Batch size	Loss Function	Opitimizer	Activation Function
100, 200	10-**100**	**0.2**, 0,3	5, 8, **10**	32, **64**	binary crossentropy	adam	sigmoid

**Optimum values in bold*

Confusion matrix and score metrics obtained from confusion matrix were used to measure the model performance. The loss functions and Acc values of the training and validation data were monitored to observe that overfitting/underfitting situations do not occur in the training phase.

4. Experimental Results

Dataset is splitted as 75% and 25% for training and testing respectively. The results of models created after text representation with TF-IDF and Word2Vec are given in Table 4 and Table 5.

Table 4. *Results with TF-IDF*

	KNN	NB	LR
Acc	79.3	92.3	**97.4**
P	72.8	91.4	**97.5**
R	95.2	93.8	**97.5**
F1	82.5	92.6	**97.5**

As seen in Table 4, the model created with LR is ahead of the other models in Acc, P, R, and F1 in 75%-25% hold-out separation after TF-IDF.

Table 5. *Results with Word2Vec*

	KNN	NB	LR	LSTM
Acc	94.3	89.2	97.1	**99.2**
P	96.0	88.2	97.5	**98.8**
R	92.8	91.3	96.8	**99.6**
F1	94.4	89.7	97.2	**99.2**

As seen in Table 5, the model created with LSTM is ahead of the other models in A, P, R, and F in 75%-25% hold-out discrimination after Word2Vec. When Table 4 and Table 5 are evaluated together, the results obtained with Word2Vec in KNN are ahead. However, in LR and NB, the difference between TF-IDF and Word2Vec is not much compared to KNN.

LSTM was the model that gave the best results among all models. The graphs of train-validation Acc and loss values of the LSTM neural network are given in Figure 8.

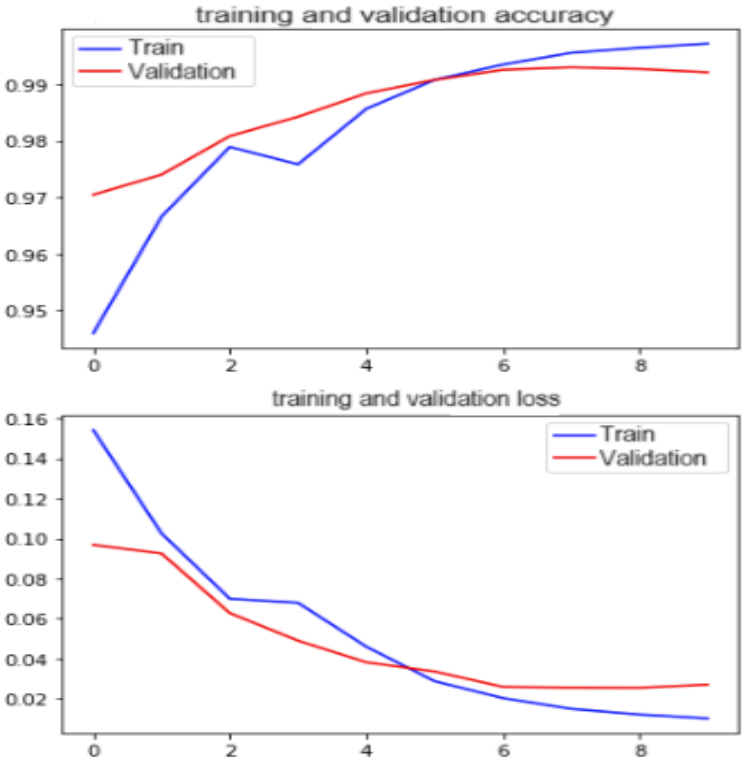


Figure 8. *Confusion Matrix*

As predicted, the success rate of the LSTM model exceeded the classical ML algorithms. However, it was observed that the training time of the model was considerably high compared to others.

5. Conclusion and Discussion

In the study where various traditional ML, DL models and text representation methods were compared for the Fake News Detection task, it was observed that the best result was obtained with the DL Model, the performance of the KNN model was more affected by the vectorization method, and NB and LR models obtained close and good results in both vectorization methods. In particular, the LR Model was found to be the best model in terms of efficiency for this binary text classification study. Similar studies on the same dataset are given in Table 6. Acc results were used because the dataset is balanced.

Table 6. *Previous studies on the ISOT dataset*

Authors	Text Representation	Model	Acc (%)
[2]	GloVe	LSVM	92.0
[3]	TF-IDF	DT	96.8
[4]	BOW	LSVM	92.0
[5]	Word2Vec	CNN	97.28
[6]	BOW	NB	74.0
[7]	Word2Vec	CRF	82.6
[8]	TF-IDF	Ensemble learning (DT)	99.0
This study	Wod2Vec	LSTM	**99.2**

In the models created for the classification of fake news with ML (KNN, NB, LR) and DL (LSTM) after TF-IDF and Word2Vec on the fake news dataset (ISOT), the best result was obtained with Word2Vec LSTM with 99.2% ACC. The model created in the study is a model that competes with the literature as seen in Table 6. The models created with Word2Vec are more successful than the models created with TF-IDF in most cases. This situation will be investigated in future studies by working on different datasets.

In the future, the process can be repeated with different data sets and the hyperparameters of the LSTM model can be further optimized. In addition, newer and successful state-of-the-art models such as BERT, RoBERTa can be used. As these models have been pre-trained with very large datasets, they have brought great advances in the fields of NLP and ML. For this reason, it is predicted that the success of the study will increase.

Declaration of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Muhammed Baki ÇAKI; data analysis, experiments and evaluations, manuscript draft preparation. Muhammet Sinan BAŞARSLAN; defining the methodology, evaluations of the results, and original draft, supervision.

References

[1] C. K. Hiramath and G. C. Deshpande, “Fake News Detection Using Deep Learning Techniques,” in 2019 1st International Conference on Advances in Information Technology (ICAIT), IEEE, Jul. 2019, pp. 411–415. doi: 10.1109/ICAIT47043.2019.8987258.

[2] H. Ahmed, I. Traore, and S. Saad, “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques,” 2017, pp. 127–138. doi: 10.1007/978-3-319-69155-8_9.

[3] F. A. Ozbay and B. Alatas, “Fake news detection within online social media using supervised artificial intelligence algorithms,” Physica A: Statistical Mechanics and its Applications, vol. 540, p. 123174, Feb. 2020, doi: 10.1016/j.physa.2019.123174.

[4] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning--based Text Classification,” ACM Comput Surv, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.

[5] D. Muduli, S. K. Sharma, D. Kumar, A. Singh, and S. K. Srivastav, “Maithi-Net: A Customized Convolution Approach for Fake News Detection using Maithili Language,” in 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3), IEEE, Jun. 2023, pp. 1–6. doi: 10.1109/IC2E357697.2023.10262664.

[6] M. Granik and V. Mesyura, “Fake news detection using naive Bayes classifier,” in 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), IEEE, May 2017, pp. 900–903. doi: 10.1109/UKRCON.2017.8100379.

[7] A. Priyadarshi and S. K. Saha, “Towards the first Maithili part of speech tagger: Resource creation and system development,” Comput Speech Lang, vol. 62, p. 101054, Jul. 2020, doi: 10.1016/j.csl.2019.101054.

[8] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, “Fake News Detection Using Machine Learning Ensemble Methods,” Complexity, vol. 2020, pp. 1–11, Oct. 2020, doi: 10.1155/2020/8885861.

[9] A. K. Shalini, S. Saxena, and B. S. Kumar, “Automatic detection of fake news using recurrent neural network—Long short-term memory,” Journal of Autonomous Intelligence, vol. 7, no. 3, Dec. 2023, doi: 10.32629/jai.v7i3.798.

[10] M. Akhter et al., “COVID-19 Fake News Detection using Deep Learning Model,” Annals of Data Science, Jan. 2024, doi: 10.1007/s40745-023-00507-y.

[11] J. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, “Distributed representations of words and phrases and their compositionality,” in Advances in Neural Information Processing Systems, 2013. [Online]. Available: https://proceedings.neurips.cc/paper

[12] R. Ahmed, M. Bibi, and S. Syed, “Improving Heart Disease Prediction Accuracy Using a Hybrid Machine Learning Approach: A Comparative study of SVM and KNN Algorithms,” International Journal of Computations, Information and Manufacturing (IJCIM), vol. 3, no. 1, pp. 49–54, Jun. 2023, doi: 10.54489/ijcim.v3i1.223.

[13] T. Öztürk, Z. Turgut, G. Akgün, and C. Köse, “Machine learning-based intrusion detection for SCADA systems in healthcare,” Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 11, no. 1, p. 47, Dec. 2022, doi: 10.1007/s13721-022-00390-2.

[14] H. Canlı and S. Toklu, “Design and Implementation of a Prediction Approach Using Big Data and Deep Learning Techniques for Parking Occupancy,” Arab J Sci Eng, vol. 47, no. 2, pp. 1955–1970, Feb. 2022, doi: 10.1007/s13369-021-06125-1.

[15] R. Vankdothu, M. A. Hameed, and H. Fatima, “A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method,” Computers and Electrical Engineering, vol. 101, p. 107960, Jul. 2022, doi: 10.1016/j.compeleceng.2022.107960.

[16] H. Canli and S. Toklu, “Deep Learning-Based Mobile Application Design for Smart Parking,” IEEE Access, vol. 9, pp. 61171–61183, 2021, doi: 10.1109/ACCESS.2021.3074887.

[17] M. Z. Khaliki and M. S. Başarslan, “Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN,” Sci Rep, vol. 14, no. 1, p. 2664, Feb. 2024, doi: 10.1038/s41598-024-52823-9.

[18] S. N. Başa and M. S. Basarslan, “Sentiment Analysis Using Machine Learning Techniques on IMDB Dataset,” in 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/ISMSIT58785.2023.10304923.

[19] F. Kayaalp, M. S. Basarslan, and K. Polat, “TSCBAS: A Novel Correlation Based Attribute Selection Method and Application on Telecommunications Churn Analysis,” in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), IEEE, Sep. 2018, pp. 1–5. doi: 10.1109/IDAP.2018.8620935.

[20] Öztürk, T., Turgut, Z., Akgün, G. et al. Machine learning-based intrusion detection for SCADA systems in healthcare. Netw Model Anal Health Inform Bioinforma 11, 47 (2022). https://doi.org/10.1007/s13721-022-00390-2

[21] Ardaç, H.A., Erdoğmuş, P. Question answering system with text mining and deep networks. Evolving Systems (2024). https://doi.org/10.1007/s12530-024-09592-7

[22] Google LLC, “Colab.” https://colab.research.google.com/. Accessed 1 Feb 2023

[23] Python, “Python.” https://www.python.org/downloads/. Accessed 1 Feb 2023

Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 4, No. 1, pp. 33-38, 2024
Research Article
Received: 08/05/2024 Accepted: 30/05/2024

Microstrip Antenna Design for 2.4 GHz RF Energy Harvesting Circuits with Artificial Neural Networks

Burak DÖKMETAŞ [1], Mehmet Ali BELEN [2,*]

Abstract
This study explores the synthesis of microstrip antennas designed for 2.4 GHz RF energy harvesting circuits through the integration of artificial neural networks (ANNs). Utilizing a 3D electromagnetic (EM) simulation tool, extensive datasets were generated for training and testing the ANN model. A meticulous trial-and-error process was employed to optimize critical hyperparameters, including the number of hidden layers, neurons per layer, and activation function types. The outcome of this process was the identification of an optimal ANN model, proficient in accurately capturing complex relationships between antenna design parameters and energy harvesting efficiency. The integration of the 3D EM simulation tool and the tuned ANN model facilitated a computationally efficient approach to antenna optimization, reducing reliance on resource-intensive simulations. This research contributes to the advancement of RF energy harvesting systems, showcasing the potential of artificial intelligence in streamlining the design process for optimal microstrip antennas in 2.4 GHz applications. The demonstrated methodology provides insights into the future of computational design, offering a swift and efficient path for meeting the evolving demands of wireless communication and sensor technologies.

Keywords: Microstrip Antenna; Artificial Neural Network; Optimization; 2.4 GHz; Energy Harvester.

1. Introduction

In recent years, the increasing demand for wireless communications and the expansion of Internet of Things (IoT) devices have focused researchers on this area [1]. Concurrently, the importance of efficient and sustainable energy sources is becoming increasingly critical. Traditional power solutions face limitations in terms of size, weight, and environmental impact, making Radio Frequency (RF) energy harvesting a promising alternative for powering these autonomous and energy-limited devices. RF energy harvesting systems utilize electromagnetic radiation from the environment and convert it into electrical power for a variety of applications, from wireless sensor networks to wearable devices [2-4]. The electromagnetic spectrum, particularly the 2.4 GHz frequency band, has gathered significant interest due to its widespread use in various communication standards such as Wi-Fi and Bluetooth. This frequency range not only provides extensive RF energy in urban and industrial environments but also aligns with the operating frequencies of many electronic devices, making it an ideal candidate for energy harvesting applications.

Especially, the design and optimization of microwave antennas for RF energy harvesting circuits present a challenging problem in the field of wireless communication and sensor technologies [5]. The pursuit of antennas with enhanced performance characteristics such as increased efficiency, compact size, and adaptability to various operating conditions has become critically important in meeting the demands of modern electronic devices. As the demand for wireless communication systems continues to rise, addressing the challenges associated with microwave antenna design is crucial for advancing the capabilities of these technologies.

*Corresponding author

BURAK DÖKMETAŞ: Kafkas University, Faculty of Engineering and Architecture, Electrical Electronic Engineering Department, Türkiye; e-mail: burak.dokmetas@kafkas.edu.tr; 0000-0001-5900-6691

MEHMET ALİ BELEN*; Iskenderun Technical Universtiy, Faculty of Faculty of Engineering and Natural Sciences, Electrical Electronic Engineering Engineering Department, Türkiye; e-mail: mali.belen@iste.edu.tr 0000-0001-5588-9407

One of the key challenges in antenna design lies in achieving optimum performance at specific frequency bands such as the commonly used 2.4 GHz range. The complexity of microwave propagation, material considerations, and the interaction of antenna parameters require a comprehensive approach to design and optimization. Traditional methods typically involve a time-consuming and iterative process that relies heavily on manual tuning and extensive simulations to achieve the desired characteristics [6].

To overcome these challenges and accelerate the design process, artificial intelligence (AI) techniques, especially artificial neural networks (ANNs), have emerged as effective tools for antenna optimization [7-11]. Inspired by neural networks in the human brain, ANNs excel at learning complex patterns and relationships within data, which makes them well-suited to handle the complex and nonlinear nature of antenna design problems [12-13]. ANNs can acquire knowledge from large datasets containing antenna performance measurements, simulation results, and design parameters, allowing them to create models that capture the inner relationships among these variables. Researchers can leverage the computational power of ANNs to efficiently explore the design space, identify optimal configurations, and quickly converge on antenna geometries that meet the desired specifications [14].

This study addresses the application of artificial neural networks for the design and optimization of a microstrip antenna adapted for RF energy harvesting circuits operating at 2.4 GHz. Integrating ANNs into the design process not only facilitates the optimization task, but also opens up possibilities for novel antenna geometries that may be difficult to explore with traditional methods. The flow chart of the proposed work is presented in Fig. 1.

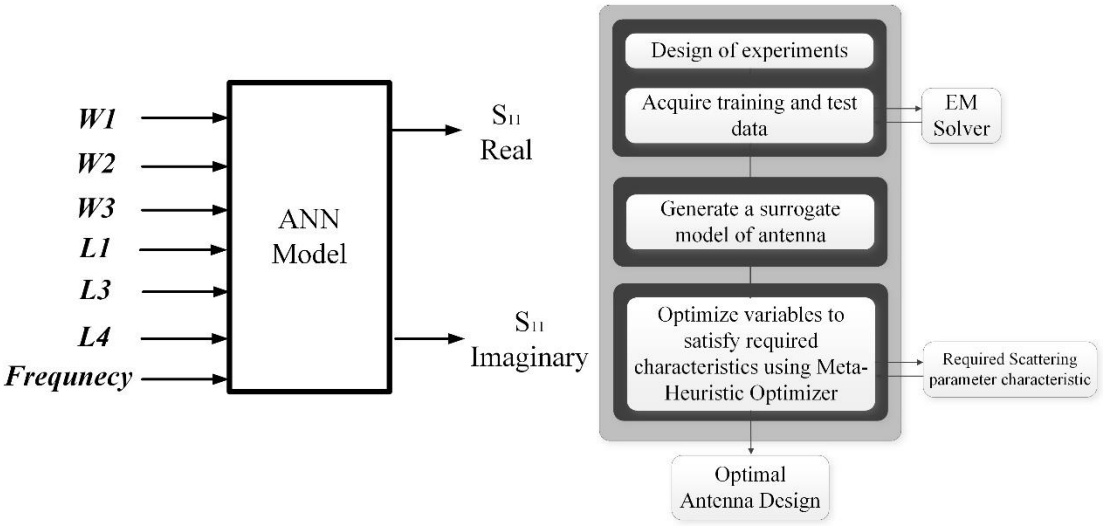


Figure 1. The flow chart of the proposed work

2. Antenna Design and Data Preparation

In this part of the study, a dataset has been prepared according to the variables defined in Table 1 for the microstrip antenna structure presented in Figure 2. According to the ranges provided in Table 1, a different antenna geometry is formed for each variable set. Thus, when the data obtained is presented to the ANN model as training data, the model will establish a relationship between the variables of the antenna geometry and the target output of the problem, which is the return loss (S_{11}). To verify the accuracy of this relationship, an additional dataset (test set) has been prepared. Here, a total of 800 data points have been prepared for training, and 200 data points for testing, with the support of 3D simulators. The frequency range has been set as 0.1-6 GHz. The Latin-Hyper-Cube sampling method has been used as the data sampling technique. In the other part of the study, ANN models related to the antenna have been prepared with the obtained datasets, and an optimization has been conducted with this model to achieve an optimal antenna design for the 2.4 GHz band. The EM simulations are done using CST microwave suit.

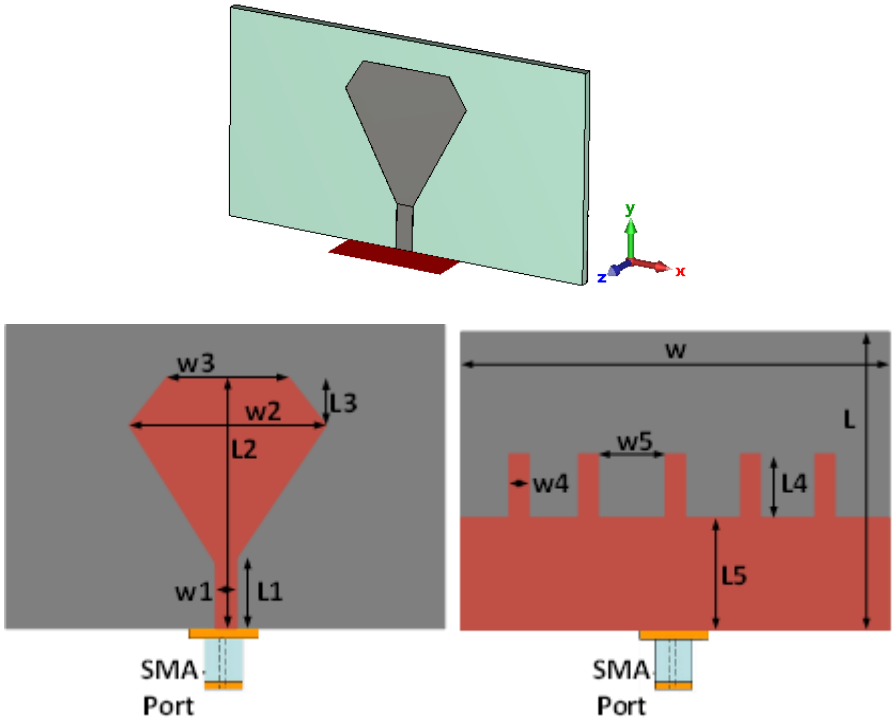


Figure 2. Images of the Proposed Microstrip Antenna

Table 1. *Variable ranges of the proposed microstrip antenna in [mm] (W= 70, L = 40, L5 = 1.5 *L4, W4 = W1, W5 = 4 * W1, L2 = L1 * 4)*

W1	1-5	L1	5-10
W2	15-30	L3	3-10
W3	10-20	L4	3-10

3. Modeling and Simulation Results

The selection of hyperparameters plays a critical role in determining the performance, efficiency, and generalization capabilities of Artificial Neural Networks [15]. Hyperparameters can generally be defined as the number of hidden layers, the number of neurons in each layer, and the type of activation function used. Careful tuning of hyperparameters is crucial to achieve optimal model performance and to overcome specific challenges associated with different tasks [16].

The number of hidden layers in an ANN affects its capacity to capture complex patterns and relationships within the data. As the depth of the network increases, it gains the ability to model complex hierarchical features. However, a deep network can be susceptible to overfitting, especially when dealing with limited data [17]. Adjusting the number of hidden layers to find the right balance is very important to prevent overfitting and allow the network to learn meaningful representations [18].

Activation functions introduce non-linearity to the network, enabling it to learn complex mappings between inputs and outputs. The choice of activation function affects the network's modeling and generalization capabilities [19]. Different activation functions are suitable for different tasks. For instance, the Rectified Linear Unit (ReLU) is known for its simplicity and effectiveness in many scenarios, while in certain cases sigmoid or hyperbolic tangent functions are preferred. The choice depends on the nature of the problem and experimentation is necessary to determine the most suitable activation function.

In this study, experiments were conducted for 1, 2, 3, 4 hidden layers, neuron numbers of 8, 16, 32, 64, 128, and activation functions tanh, sigmoid, and ReLu. To present the results of the study in a more convenient way and easy to understand, the results of the model with the lowest test error from all these experiments have been used. Here more than 100 different models (only 15 different designs for single layer model 5 different neurons size x 3 different activation function) had been tested using Relative Mean Error (RME) Eq. 2 [20]. The targeted value (T) is the S_{11} value provided form the test data set which is a complex number, while the predicted value (P) is provided by the ANN model for each of the test samples over the operation frequency. As it can see from

figure 1, the ANN model predict real and imaginary parts of the S_{11} separately and combine them as complex number during the test evaluation. From all these evaluated models, the model with three-layer model with neuron numbers set to 16-32-64 and equipped with the ReLu activation function found to be the optimal design for studied problem. The test performance of the model has been obtained as 4.6% in terms of RME. In a similar approach to the study presented in [21], the ANN model obtained was run with an optimization algorithm aimed at achieving the desired antenna design. The cost function used for this purpose has been presented in equation (2). The design variables obtained at the end of this optimization process are provided in Table 2.

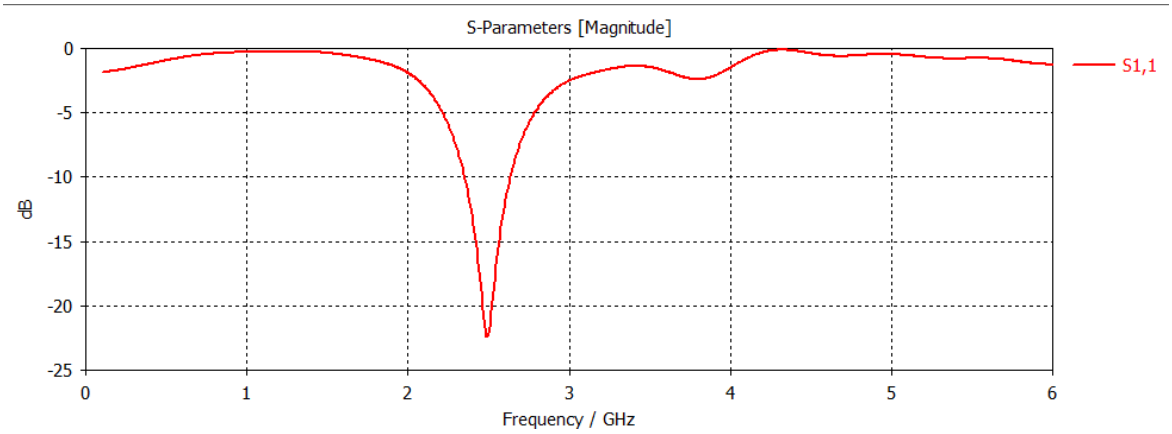
$$RMEt = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_i - P_i|}{|T_i|} \quad (1)$$

$$Cost = \sum_{f_{min}}^{f_{max}}\frac{1}{|S_{11}(f)|} \quad (2)$$

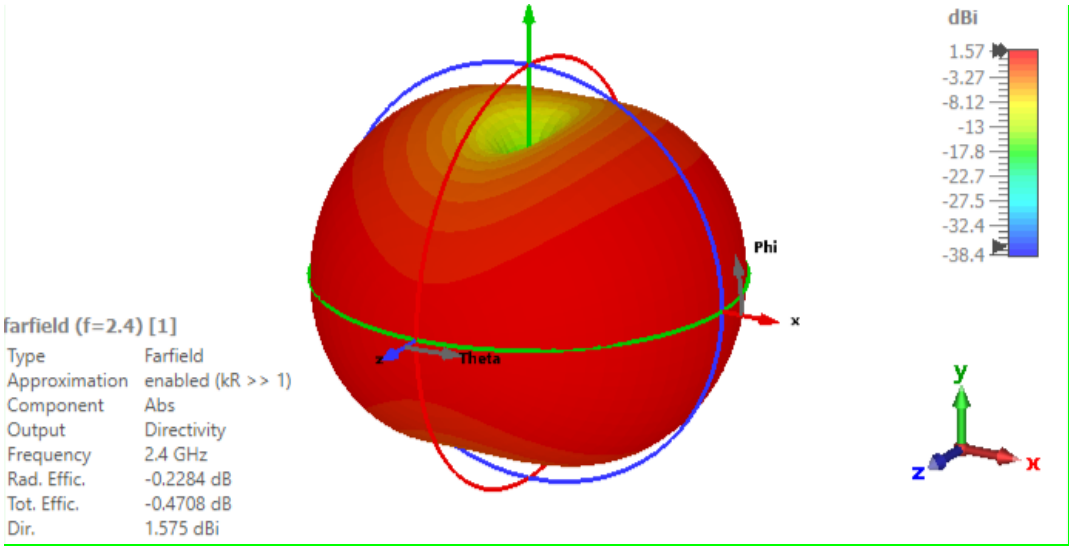
Table 2. *Optimum design variables obtained with ANN-Based optimization in [mm].*

W1	2	L1	8
W2	22.5	L3	5.3
W3	16	L4	8

To verify the accuracy of the obtained optimum antenna design values, the results were inputted into a 3D simulator tool. Figure 3(a) presents the graph of the S_{11} formed based on these data. Figure 3(b) shows the radiation pattern of the antenna at 2.4 GHz.



(a)



(b)

Figure 3. Simulated antenna with optimum variable values showing (a) Return loss S_{11}, (b) 3D radiation pattern

4. Conclusion

In this study, a microstrip antenna design for RF energy harvesting systems at the 2.4 GHz frequency band has been presented using artificial neural networks (ANNs). A 3D electromagnetic (EM) simulation tool was utilized to generate the training and testing data for the ANN model. A series of trial and error iterations and hyperparameters were used to develop the ANN model. These hyperparameters, which are crucial in shaping the learning and optimization capabilities of the neural network, include the number of hidden layers, the number of neurons in each layer, and the type of activation function. The hyperparameter tuning process resulted in the definition of an optimal ANN model. The fine-tuned ANN model through hyperparameter optimization served as a proxy for traditional and time-consuming design processes. This not only accelerated the design phase but also demonstrated the potential of artificial intelligence to efficiently navigate the multidimensional parameter space inherent in antenna optimization. Consequently, the combination of 3D EM simulation tools and artificial neural networks has been proven to be a synergistic and effective approach in the search for optimum microstrip antenna design. This study contributes to the advancement of RF energy harvesting systems by providing a glimpse into the future of computational design methodologies that leverage the power of artificial intelligence to meet the evolving demands of wireless communication and sensor technologies.

Declaration of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Conceptualization, BD MAB; methodology, MAB ; data generation, BD; investigation, BD; designing BD; writing—original draft preparation, BD, MAB; writing—review and editing, MAB; visualization, BD; supervision, MAB; project administration, MAB. All authors reviewed the manuscript.

References

[1] B. Dokmetas, and G.O. Arican, “Design of dual-band SIW antenna for millimeter-wave communication,” 31st Telecommunications Forum (TELFOR), pp. 1-4, November, 2023.

[2] W. A. Khan, R. Raad, F. Tubbal, P. I. Theoharis, and S. Iranmanesh. “RF energy harvesting using rectennas: A comprehensive survey,” IEEE Sensors Journal (2024).

[3] J. Szut, P. Piątek, and M. Pauluk, “RF energy harvesting Energies,” vol. 17, no. 5, 2024.

[4] V. Kuhn, C. Lahuec, F. Seguin and C. Person, “A multi-band stacked RF energy harvester with RF-to-DC efficiency up to 84%,” IEEE Transactions on Microwave Theory and Techniques, vol. 63, no. 5, pp. 1768-1778, May 2015.

[5] X. Zhang, J. Grajal, M. López-Vallejo, E. McVay, and T. Palacios, “Opportunities and challenges of ambient radio-frequency energy harvesting,” Joule, vol. 4, no. 6, pp. 1148-1152, 2020.

[6] P. Mahouti, “Design optimization of a pattern reconfigurable microstrip antenna using differential evolution and 3D EM simulation-based neural network model,” International Journal of RF and Microwave Computer-Aided Engineering, vol. 29, no. 8, 2019.

[7] S. Roshani, S. Koziel, S. I. Yahya, M. A. Chaudhary, Y. Y. Ghadi, S. Roshani, and L. Golunski, “Mutual coupling reduction in antenna arrays using artificial intelligence approach and inverse neural network surrogates,” Sensors, vol. 23, no. 16, 2023.

[8] P. Pragya, and J. S. Sivia, “Design of minkowski curve-based slotted microstrip patch antenna using artificial neural network,” Journal of The Institution of Engineers (India): Series B, vol. 104, no. 1, pp. 129-139, 2023.

[9] M. Mahouti, N. Kuskonmaz, P. Mahouti, M. A. Belen, and M. Palandoken, “Artificial neural network application for novel 3D printed nonuniform ceramic reflectarray antenna,” International journal of numerical modelling: electronic networks, devices and fields, vol. 33, no. 6, 2020.

[10] D. Prabhakar, P. Karunakar, S. V. Rama Rao, and K. Srinivas, “Prediction of microstrip antenna dimension using optimized auto-metric graph neural network,” Intelligent Systems with Applications, vol.21, 2024.

[11] R. Ramasamy and M. A. Bennet, “An efficient antenna parameters estimation using machine learning algorithms,” Progress In Electromagnetics Research C, vol. 130, pp. 169-181, 2023.

[12] S. Koziel, N. Çalık, P. Mahouti and M. A. Belen, “Accurate Modeling of Antenna Structures by Means of Domain Confinement and Pyramidal Deep Neural Networks,” IEEE Transactions on Antennas and Propagation, vol. 70, no. 3, pp. 2174-2188, March 2022.

[13] A. Papathanasopoulos, P. A. Apostolopoulos and Y. Rahmat-Samii, “Optimization assisted by neural network-based machine learning in electromagnetic applications," IEEE Transactions on Antennas and Propagation, vol.72, no.1, pp.160-173, 2023.

[14] S. Koziel, “Fast simulation-driven antenna design using response-feature surrogates,” International Journal of RF and Microwave Computer-Aided Engineering, vol. 25, no. 5, pp. 394-402, 2015.

[15] S. Koziel, N. Çalık, P. Mahouti and M. A. Belen, "Reliable computationally efficient behavioral modeling of microwave passives using deep learning surrogates in confined domains," IEEE Transactions on Microwave Theory and Techniques, vol. 71, no. 3, pp. 956-968, March 2023, doi: 10.1109/TMTT.2022.3218024.

[16] B. Si, Z. Ni, J. Xu, Y. Li, and F. Liu, "Interactive effects of hyperparameter optimization techniques and data characteristics on the performance of machine learning algorithms for building energy metamodeling," Case Studies in Thermal Engineering, 2024.

[17] X. Ying, "An overview of overfitting and its solutions," In Journal of physics: Conference series, vol. 1168, IOP Publishing, 2019.

[18] S. Koziel, N. Çalık, P. Mahouti, and M. A. Belen, "Accurate modeling of antenna structures by means of domain confinement and pyramidal deep neural networks," IEEE Transactions on Antennas and Propagation, vol. 70, no. 3, pp. 2174-2188, 2021.

[19] N. Calik, F. Güneş, S. Koziel, A. Pietrenko-Dabrowska, M. A. Belen, and P. Mahouti, "Deep-learning-based precise characterization of microwave transistors using fully-automated regression surrogates," Scientific Reports, vol. 13, no. 1, 2023.

[20] N. G. Reich, J. Lessler, K. Sakrejda, S. A. Lauer, S. Iamsirithaworn, and D. AT Cummings. "Case study in evaluating time series prediction models using the relative mean absolute error," The American Statistician, vol. 70, no. 3, 2016.

[21] P. Mahouti, A. Belen, O. Tari, M. A. Belen, S. Karahan, and S. Koziel, "Data-driven surrogate-assisted optimization of metamaterial-based filtenna using deep learning," Electronics, vol. 12, no. 7, 2023.

Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 4, No. 1, pp. 39-43, 2024
Research Article
Received: 07/06/2024 Accepted: 24/06/2024

Aortic Coarctation Diagnosis on Echocardiography Images

Yaren ENGIN [1], Omer Pars KOCAOGLU [2,*]

Abstract

Aorta is the main artery that carries clean highly-oxygenated blood pumped by the heart. Aortic coarctation is a congenital heart disease that restricts blood flow in this artery and it is a condition that can be difficult to diagnose in the fetus or newborns. This difficulty stems from a narrow seperation occurring between the aorta and the ductus arteriosus, often near the origin of the left subclavicular artery. The ductus arteriosus usually closes spontaneously after birth, when aortic coarctation starts being detectable. Echocardiography is the preferred method of diagnosis of aortic coarctation in newborns, which is performed by the physician examining the image. Such diagnosis has proven to be a difficult with low success rates. The aim of this study is to measure aortic diameter of newborns using echocardiography image processing techniques for better, faster and more objective diagnosis of aortic coarctation.

Keywords: Aortic coarctation, echocardiography, image processing.

1. Introduction

Aortic coarctation (CoA) in newborns (the first 28 days [1]) exhibits itself as the narrowing of the aortic lumen making it difficult for blood flow from the heart to the upper body. Aortic coarctation can be a difficult to diagnose in the fetus or newborns. The difficulty in diagnosis stems from the existence of a vessel called the ductus arteriosus, a vascular connection that provides passage between the pulmonary artery and the aorta in the fetus, that helps maintain blood flow in the aorta. A narrowing occurs between the aorta and the ductus arteriosus, often near the origin of the left subclavicular artery. The ductus arteriosus usually closes spontaneously after birth, when aortic coarctation becomes detectable. Aortic coarctation is the most common congenital heart disease of ductal-dependent systemic circulation [2]. The neonatal period is characterized by changes in organ functions. The neonatal myocardium is less able to tolerate increased preload and has a lower response to increased post load. Therefore, early and accurate diagnosis of aortic coarctation is crucial for preventing acute deterioration of cardiac functions in newborns. Persistent ductus arteriosus (PDA) in the fetus and newborns can change the anatomy of the aortic lumen, making it difficult to evaluate the degree of narrowing of the aortic isthmus [2].

Symptoms of aortic coarctation vary depending on the degree of stenosis. In mild cases, there may be no symptoms. In moderate and severe cases, tiredness, shortness of breath, chest pain, dizziness, fainting and swelling in the legs may be observed [2]. In newborns, obstruction due to coarctation may occur a few days after birth and in rare cases it can be detected using pulse oximetry in newborn screening [3]. Recently, cardiac ultrasound has become a more commonly used test for early diagnosis of CoA, used to determine the extent and location of aortic narrowing. Advancements in cardiac ultrasound imaging technology have the potential to improve prenatal diagnosis of CoA [4].

Surgical intervention of CoA involves widening or repairing the stenosis in the aorta. This procedure is usually performed as open-heart surgery during infancy. In older children and adults, less invasive procedures such as stent placement or balloon dilation may also be used. Surgery is arguably the treatment of choice for native aortic coarctation in newborns. In older babies, balloon coarctoplasty also has good early and mid-term outcomes and acceptable reintervention rates [6]. Therefore, it is also considered as an alternative to open-heart surgery, especially in critically ill patients with high surgical risk [6]. The best approach to treat CoA in the pediatric population is a highly debated topic. However, stent implantation has been shown to result in shorter hospital stays and fewer acute complications compared to surgical repair or balloon angioplasty. Additionally, covered stents appear to be more protective than bare stents [7].

*Corresponding author

OMER PARS KOCAOGLU*; Izmir Katip Celebi University, Faculty of Engineering and Architecture, Biomedical Engineering Department, Türkiye; e-mail: omerpars.kocaoglu@ikcu.edu.tr; 0000-0002-9660-9448

YAREN ENGIN; Izmir Katip Celebi University, Faculty of Engineering and Architecture, Biomedical Engineering Department, Türkiye; e-mail: yarengin98@gmail.com; 0009-0009-0412-6069

Critical aortic coarctation can be fatal if left untreated. Therefore, it is important for infants with this condition to be diagnosed and treated rapidly after birth. People with aortic coarctation still need regular checkups throughout life. These checks help prevent the stenosis from progressing or complications from occurring [7]. Regardless of the treatment type, aortic coarctation is a disease that can be treated. However, it can be fatal if it is not diagnosed in a timely manner or if a false negative diagnosis occurs. Current diagnosis method requires: an echocardiography taken by a pediatric cardiologist, ductus arteriosus being closed at the time of imaging, and aorta diameter measurements made manually.

Recent studies from Scandinavia found that at least 50% of newborns with aortic coarctation were not diagnosed accurately within five days after birth [8]. In a study conducted in California, 27% of patients with aortic coarctation died before diagnosis at an average age of 17 days [8]. Ward *et al.* observed that infants with symptomatic aortic coarctation appeared between five and 14 days after birth. Coarctation can be detected by fetal echocardiography, but this method is generally used by specialists and is not intended for a wider spread general use [8].

Since medical images vary widely and has complexities due to different imaging modalities, image quality, patients, and other factors, it is difficult to create analytical solutions or simple equations to represent objects such as lesions and anatomical structures. Machine learning is a powerful tool that can address these differences and complexities in medical images. Machine learning algorithms are designed to learn from examples and represent data on their own allowing a potential for more objective, accurate and reliable results for diagnostic tasks in medical imaging. Machine learning is currently used in a variety of medical imaging applications for computer-aided diagnosis, image segmentation, tissue classification, pathology detection, and more [9, 10].

Marrow *et. al.* have investigated differences in aortic size between newborns with and without coarctation of the aorta. In comparison, they found that the aortic ring, the ring at the base of the aorta, was smaller in newborns with CoA than in healthy newborns. Their study also revealed that transverse aortic arch and isthmus parts of the aorta were narrower in newborns with CoA than in healthy babies [11].

In this project, to address the need for a better and timely diagnostic procedure for diagnosis of CoA in newborns, we aim to utilize image processing on echocardiography images for recognizing the aorta and measuring its diameter to detect any narrowing or widening when compared to healthy newborns' aorta dimensions. Such automated diagnosis has the potential for use in situations even when an expert interpreter of echocardiography images is not available for consulting, for example in the emergency rooms where time savings increase newborns' chances of survival. Additionally, this type of diagnosis will help minimizing subjectivity in measurements.

2. Methodology

This study was conducted by using the echocardiography images of healthy newborns for training and comparing them with the ones diagnosed with aortic coarctation. Images were also pre conditioned by basic image processing methods for facilitating of accurate detection of CoA. The method explained here was intended for a decision support system that includes the processes of identifying life-threatening aortic stenosis, making the necessary measurements for its treatment through a visual model, and determining non-standard values.

2.1. Ultrasound Images

Eechocardiography images of 55 healthy newborns and 4 newborns with aortic coarctation were obtained from Erzurum Atatürk University, Faculty of Medicine, Department of Pediatric Cardiology and Kastamonu Private Anadolu Hospital Neonatal Intensive Care Service after the related ethical and regulatory permissions were obtained. Number of images containing CoA patients were limited to 4 as it is a rare condition which presents challenges for higher number of samples collection.

2.2. Image Processing and Measurement Tool

We describe a decision support system that can help diagnosing CoA even at the emergency departments by utilizing a quick evaluation tool for echocardiography images of infants. First, the aortic vessel echocardiography images were introduced to a machine learning algorithm for training purposes. A software using CvZone and NumPy libraries on the Python-PyCharm platform over existing echocardiography images of healthy newborns

was used. As a result, measurement system that can evaluate cardiac ultrasound images with different reference values were generated.

Once the echocardiography images were loaded, the software flow incorporates following steps of processing:

1. Scaling: Echocardiography images can have reference values of 6 cm or 8 cm. Thus, as a first step he appropriate reference values were entered to the software manually by the user for accurate scaling and diameter reporting.
2. Gray scale optimization: The image was transformed into a study-standardized and previously conditioned 8-bit grayscale image, which is optimized for clear visibility and identification of the aorta.
3. Gaussian Blur: A Gaussian blur that was designed to preserve the imaging device's theoretical resolution is applied to the image to reduce salt-and-pepper noise and pixelation degrading the appearance of cardiovascular structures. Thus, further improving the visibility of aorta.
4. Image segmentation: A section of the image where the overall heart structures were visible was segmented out of the complete image for further processing.
5. Feature detection: Using the segmented image, general features of the heart were marked manually by a trained professional for locating the area of interest (AOI).
6. Training: Images with marked features together with the acquisition parameters of the images were used for training the algorithm for three-dimensional appearance of the heart and aorta.
7. Detection of aorta diameter: In the final step, the algorithm detects a circular region using the Hough circle transform where the aorta region exists. Then, by changing the circular regions' radius using parameters withing the previously set minimum and maximum radii limits and step sensitivity, a resulting green circle was drawn on the original ultrasound image.

As a result, the software reports the measurement of the aorta diameter in millimeters (mm) or centimeters (cm) on the user interface. At the end of this process, the images with labelled aortas and related diameter measurements were saved in to a selected folder.

3. Results

All 59 echocardiography images obtained from newborns were used as inputs to the software tool for automated detection and reporting of the aorta diameter. For 55 healthy newborns, an average of 8,09±1,18 mm was reported, consistent with those previously reported [3, 8, 11]. On the other hand, the average of 4 CoA patients' diameter measurements reported by the software were lower at 2,83±0,55 mm, a statistically significant difference (unpaired T-test two-tailed P value = 0,001). Figure 1A, show an exemplary echo image and corresponding aorta diameter measurement of an infant (6,71 mm) reported as "healthy" by the software. Figure 1B show an exemplary echo image and corresponding aorta diameter measurement (2,12 mm) of a CoA patient infant reported as "not normal" by the software.

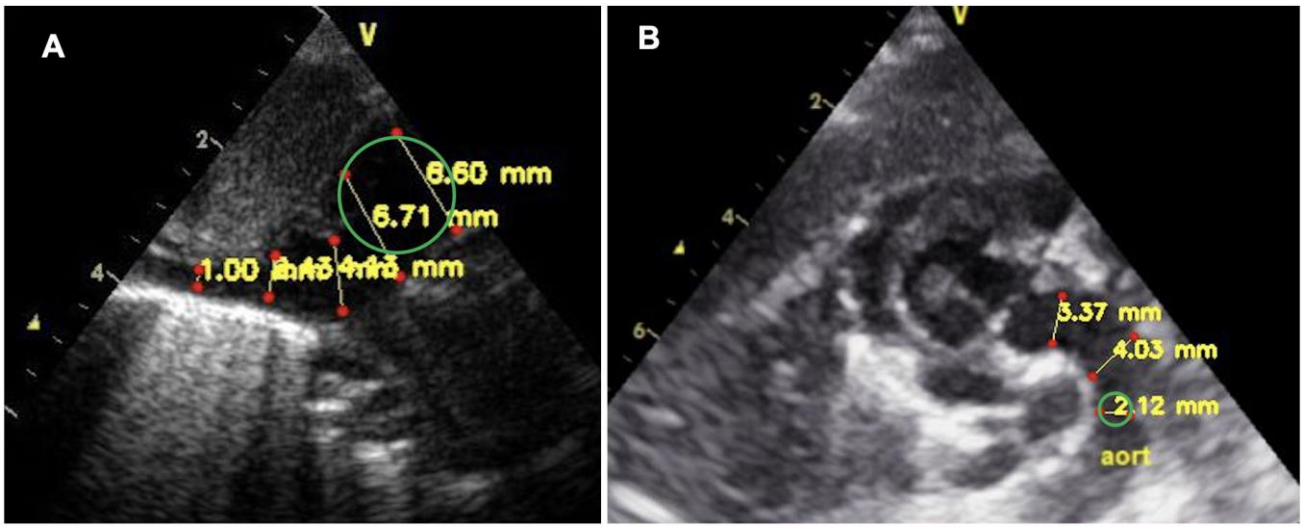


Figure 1. ***A.*** *Echo image of newborn aortic diameter measured at* 6,71 mm *mm by the software.* ***B.*** *Echo image of newborn aortic diameter measured at* 2,12 mm *mm by the software.*

Although with limited number of samples, due to difficulty of obtaining echo images of infants with CoA, this study demonstrated the feasibility of objectively diagnosing newborn CoA patients over a group 59 samples. Albeit the small sample size, such demonstration helps us understand the potential of machine learning based software development in medical imaging applications. This type of software solutions, after receiving required validations and experienced physician confirmations, could allow minimizing the undesired effects specialists being inaccessible at remote locations and hospitals with insufficient resources to get an alternative and objective diagnosis possibility. It also helps making more optimal intervention decisions by physicians.

4. Discussion

Specialists' CoA diagnosis success rates may vary depending on doctors' personal experience levels, evaluation consistency and perspectives, which inserts an undesired variability into diagnosis and treatment processes. Future machine learning studies can contribute mitigating problems that may stem from such inconsistencies and have potential for obtaining more objective and consistent measurements that can result in more accurate diagnoses with lower false negative or false positive rates. However, it is important to understand that these algorithms require a large amount of data to be accurate and effective. In this study, the small data size prevents us from drawing conclusive outcomes and demonstrating the full potential of our processing technique in medical ultrasound images. Nevertheless, even with limited data size we demonstrate its feasibility. In the future, we aim to increase the data entry to improve automated diagnosis potential.

In aortic coarctation, narrowing occurs between the aorta and the ductus arteriosus, usually near the origin of the left subclavicular artery. This is a situation in which we have a high detection rate when we image the part defined as the aortic arch. However, for now, the developed software can only evaluate transthoracic echocardiography images and take circular measurements. The measurement reliability was limited by the software's occasional section recognition issues, image acquisition angle related variances, and the fact that only one single section was being used. In the future, we aim to improve the software by defining the aortic arch and take measurements in cross-sections with high lateral sampling. Such improvements would help reducing the possibility of an aorta narrowing being overlooked by the algorithm [12]. Machine learning models can identify complex patterns and trends that may escape human eyes, which could help diagnose aortic stenosis more accurately and minimize misdiagnoses while also allowing cost savings.

In summary, albeit the small data size, our software was able to recognize the transthoracic echocardiography images and detect aorta narrowing. With improvements like introduction of aortic arch and larger data sets, the presented software has the potential for higher accuracy results and better performance in a bigger variety of image angles and qualities. Consequently, it can help faster, more accurate and less costly diagnosis of infant CoA.

Declaration of Interest

The authors declare that there is no conflict of interest.

Acknowledgements

The authors acknowledge the support of Erzurum Ataturk University Faculty of Medicine, Department of Pediatric Cardiology doctors, who supported this project by sharing anonymous newborn echocardiography images. We also thank Dr. Naci CEVİZ and Specialist Dr. Serkan ORDU from Kastamonu Private Anadolu Hospital Cardiology Department. We also thank Drs. Sertac ESIN and Dr. Hacer Inan GUNGOR for supporting this project with her knowledge and experience.

Author Contributions

Yaren Engin carried out the image collection, image processing software development, measurements and organizing of the results shown in this study. Omer Pars Kocaoglu defined the project criteria, interpreted results and wrote the manuscript.

References

[1] WHO, https://www.who.int/westernpacific/health-topics/newborn-health

[2] S. E. Rusu, D. Toma, C. Blesneac, L. Matei, C. Ghiragosian, R. Togănel, "Diagnosis of the aortic coarctation in the neonatal period — a critical condition in the emergency room," Journal Of Cardiovascular Emergencies. 3(3): 128-132, 2017.

[3] N. K. Goyal, “The newborn infant,” *Nelson Textbook of Pediatrics*. 21st ed. Philadelphia, PA: Elsevier; 2020.

[4] P. J. Rozance, C. J. Wright, “The neonate,” *Gabbe's Obstetrics: Normal and Problem Pregnancies*. 8th ed. Philadelphia, PA: Elsevier; 2021.

[5] W. R. Thompson, S. M. Mehrotra, “Cardiac History and Physical Examination, “ Heart Diseases in Children. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-7994-0_1, 2011.

[6] S. Sen, S. Garg, S. G. Rao, s. Kulkarni, “Native aortic coarctation in neonates and infants: immediate and midterm outcomes with balloon angioplasty and surgery,” Annals Of Pediatric Cardiology, (11): 261-266, 2018.

[7] L. Yu, J. Chan, F. Cetta, N. W. Taggart, J. H. Anderson, “Transcatheter treatment of native coarctation of the aorta in a child using a balloon-expandable endoprosthesis,” JACC: Case Reports. (14): 2023.

[8] J. I. Hoffman, “The challenge in diagnosing coarctation of the aorta,” Cardiovascular Journal Of Africa.; 29(4): 252-255, 2017

[9] P. Y. Shen, “Machine learning in medical imaging,” Machine Vision And Applications, 1327-1329, 2013;

[10] K. Spanos, P. A. D. Giannoukas, G. Kouvelos, I. Tsougos, A. Mavroforou, “Artificial intelligence application in vascular diseases. Journal Of Cardiovascular Surgery,” 76(3); 615-619, 2022.

[11] W. R. Morrow, J. C. Huhta, D. J. Murphy Jr., D. G. McNamara, “Quantitative morphology of the aortic arch in neonatal coarctation,” Journal of the American College of Cardiology, 8(3): 616-620, 1986.

[12] T. Yan, J. Qin, Y. Zhang, Q. Li, B. Han, X. Jin, ”Research and application of intelligent image processing technology in the auxiliary diagnosis of aortic coarctation,” Frontiers in Pediarics. 11: 1-13, 2023.

Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. , No. 1, pp. -58, 202
Research Article
Received: / / Accepted: / /

Dual-Class Stocks: Can They Serve as Effective Predictors?

Veli SAFAK

Abstract

This paper investigates the three stocks of Kardemir Karabuk Iron Steel Industry Trade & Co. Inc. (Kardemir), the 24th largest industrial company in Turkey, listed on the Borsa Istanbul under tickers KRDMA, KRDMB, and KRDMD. Despite sharing identical attributes except for voting power, these stocks have displayed notable price divergences over an extended period from January 2001 to July 2023. Through an extensive analysis, this paper identifies and quantifies the price divergence patterns, revealing a compelling arbitrage opportunity through pair trading with a maximum potential gain of 361.21%. Employing wavelet coherence analysis, this study documents a strong coherence among the stock prices for the majority of the analysis period. Additionally, it demonstrates that the use of a sliding-window approach in selecting the training set significantly improves predictive performance based on 2,408 long short-term memory (LSTM) models. Notably, the fixed-horizon approach is found to lead to a statistically significant underestimation of future prices. Finally, the empirical findings emphasize that even when predictors exhibit strong coherence with the target variable, they may adversely impact predictive performance.

Keywords: Dual-class stock; long short-term memory; stock price prediction; wavelet analysis

1. Introduction

The establishment of Kardemir, Turkey's inaugural integrated iron and steel factory, dates back to September 10, 1939, when it was initiated by İsmet İnönü, who served as Prime Minister during that time. This significant step was part of the broader national industrialization efforts championed by the republic's founder, Mustafa Kemal Atatürk. The main activity subject of the company is the production and sale of all kinds of crude iron and steel products, coke, and coke by-products. It was listed in the Borsa Istanbul on Jun 1, 1998, with 3 stocks: group A (ticker: KRDMA), group B (ticker: KRDMB), and group D (ticker: KRDMD). The Group A shareholders have the right to elect 4 members to the Board of Directors, the Group B shareholders have the right to elect 2 members to the Board of Directors, and the Group D shareholders have the right to elect 1 member to the Board of Directors. Apart from this voting privilege, there are no other privileges.

This stock structure with different voting privileges is known as dual-class stock structure. There is significant cross-country evidence suggesting that investors pay a premium for stocks with voting privileges. The pioneering empirical investigation in this domain was conducted by Lease, McConnell, and Mikkelson [37], who demonstrated that higher vote shares in the United States are associated with a premium of approximately 5%. Horner [28] examined dual-class stocks in Switzerland and observed a voting premium of merely about 1%. Zingales [26] identified a substantial premium of roughly 80% in Italy. Smith and Amoako-Adu [3] detected a premium of around 19% in Canada during the period 1988-1992, which closely resembles the premium documented in Sweden by Rydqvist [22] at 15%. Additionally, Megginson [48] provided evidence of a premium of around 13% in the United Kingdom.

Voting premium is not the only interesting phenomenon about the dual-class structures. There is also evidence suggesting that prices of dual-class stock also tend to exhibit high co-integration [1]. Since 2014, the relative price ratio of GOOG (without voting power) and GOOGL (with voting power) ranged between 0.9459 and 1.05. Wu [19] used the co-integration between GOOG and GOOGL and designed a pair trading strategy. Pair trading constitutes a market-neutral tactic centered on the selection of stock pairs grounded in their relative prices or alternative indicators. The primary objective is to pinpoint pairs that exhibit a substantial level of correlation or cointegration, indicative of their tendency for synchronized price movements. This strategy finds prevalent usage among hedge funds and can be further refined through the assimilation of supplementary insights, such as volatility, anti-persistence, or qualitative information derived from financial reports. Diverse methodologies, spanning statistical assessments, machine learning algorithms, and genetic programming, can be employed to unearth lucrative pairs and formulate trading cues. The efficacy of pair trading extends across various asset categories and market conditions, with certain investigations intimating heightened effectiveness in periods of market decline [18], [25], [9], [11], and [5].

The academic interest in dual-class stock structure stems from the curiosity of understanding the stock prices. Forecasting stock prices is a classic problem laying in the intersection of finance, computer science, and economics. Various methods have been developed and used to forecast future stock prices. Fundamental analysis based

*Corresponding author

Veli SAFAK*; e ie e o U i e i , ; e-mail: e .cm .e ;

on companies' financial statements and technical analysis based on various indicators formulated as functions of past price action are cornerstones in quantitative finance.

In parallel to reductions in computation cost and increase in the volume of accessible data, researchers have developed more sophisticated methods to forecast price action in foreign exchange and stock markets. In recent years, deep learning algorithms have gained attraction among researchers. There are four major types of deep learning algorithms: convolutional neural network, deep neural network, recurrent neural network, and long short-term memory. While this paper exclusively employs LSTM models, it provides a brief overview of the other methods. Additionally, it presents a list of studies utilizing methods not emphasized in this paper for reference. For a more in-depth literature review, readers may refer to [47] and [54].

Convolutional neural networks (CNNs) are a class of deep neural networks specifically designed for processing and analyzing visual data, such as images and videos. CNNs are characterized by their ability to automatically learn hierarchical representations of features from input data. The key innovation of CNNs lies in the use of convolutional layers, which apply convolutional operations to input data. These operations involve small, learnable filters that scan the input in a systematic way, capturing local patterns and spatial relationships. This enables the network to recognize low-level features, such as edges and textures, and progressively build more abstract and complex representations through subsequent layers. Typically, CNN architectures consist of convolutional layers followed by pooling layers, which downsample the spatial dimensions of the data, reducing computational complexity. Fully connected layers are then employed to make predictions or classifications based on the learned features. The strength of CNNs lies in their ability to automatically extract relevant features from raw input data, making them highly effective in tasks such as image classification, object detection, and image segmentation. The hierarchical feature learning in CNNs mimics the human visual system, contributing to their success in various computer vision applications.

Numerous studies have explored CNNs in stock market prediction and explored various aspects, including model comparisons, graph theory integration, technical indicator application, multi-indicator feature selection, ensemble models, event-driven prediction, and unique architectural approaches. These studies contribute to advancing the utilization of deep learning techniques for enhanced stock market forecasting [10], [12], [13], [14], [16], [20], [29], [30], [34], [35], [40], [42], [44], [49], [51], and [53].

Deep neural networks (DNNs) represent a class of artificial neural networks characterized by their depth, involving multiple layers of interconnected nodes or neurons. These networks are designed to automatically learn hierarchical representations of features from input data, allowing them to capture intricate patterns and relationships. The architecture of DNNs typically consists of an input layer, one or more hidden layers, and an output layer. Each layer contains nodes that process information and pass it to subsequent layers, with weighted connections determining the strength of these interactions. The depth of DNNs facilitates the extraction of complex and abstract features from raw input, enabling them to effectively model intricate relationships in data. Training DNNs involves adjusting the weights of connections through backpropagation, where the network learns by minimizing the difference between predicted and actual outputs. This iterative learning process enhances the network's ability to generalize and make accurate predictions on new data. DNNs have demonstrated considerable success in various domains, including image and speech recognition, natural language processing, and reinforcement learning. Their capacity to automatically learn hierarchical representations contributes to their effectiveness in capturing intricate patterns and solving complex tasks, making them a prominent tool in machine learning research.

Various studies have investigated the application of deep neural networks (DNNs) in stock market prediction. One study utilized a DNN model with novel input features and a plunge filtering technique, demonstrating notable profitability [52]. Another proposed a DNN model using the Boruta feature selection technique, outperforming some other machine learning models [32]. Additionally, a study employed boosted approaches in a DNN model to predict stock market crises, highlighting their relevance in price prediction [43]. Another research revealed the superiority of DNNs over shallow neural networks and representative machine learning models [27]. Lastly, a study on a deep factor model suggested a nonlinear relationship between stock returns and factors, outperforming linear models and other machine learning methods [21].

Recurrent neural networks (RNNs) constitute a category of artificial neural networks specifically designed to process sequential data by incorporating temporal dependencies. Unlike traditional feedforward neural networks, RNNs possess internal memory mechanisms, allowing them to retain information about previous inputs and use it to influence subsequent predictions. The architecture of RNNs includes recurrent connections that form loops, enabling information to persist within the network over time. This inherent memory capacity makes RNNs well-suited for tasks involving sequential patterns, such as natural language processing, time series prediction, and speech recognition. However, traditional RNNs suffer from challenges like the vanishing gradient problem, which hinders their ability to effectively capture long-range dependencies in sequential data. To address this limitation,

variants like long short-term memory (LSTM) networks and gated recurrent unit (GRU) networks have been developed. These architectures incorporate specialized memory cells and gating mechanisms, allowing for improved information retention and flow through the network. The success of RNNs lies in their ability to model and comprehend sequential dependencies, making them valuable tools in diverse applications. Despite their effectiveness, ongoing research aims to further enhance their capabilities and address remaining challenges to advance the field of sequential data analysis.

Several studies have explored RNNs for financial prediction. One introduced a CRNN model combining convolutional neural network and recursive neural network and demonstrated its outperformance relative to LSTM and CNN models in forecasting Forex pairs' prices [24]. Another proposed a multi-task RNN model with Markov Random Fields (MRF), employing a multi-multilayer perceptron (MMLP) for feature extraction without reliance on technical indicators [6]. A study presented an RNN-Boost model incorporating technical indicators, sentiment features, and Latent Dirichlet allocation (LDA) features, demonstrating superior performance over a single-RNN model [46]. In a different approach, a Deep and Wide Neural Network (DWNN) model integrated CNN's convolution layer into the RNN's hidden state transfer process, achieving a 30% reduction in prediction mean squared error compared to a general RNN model [38]. Additionally, an Attention-based RNN (ARNN) with wavelet denoised input was proposed, combining autoregressive integrated moving average (ARIMA) and a RNN model output for enhanced forecasting [55].

Overall, this paper presents results of three primary analyses. It begins by examining the historical voting premium within Kardemir stocks. This analysis represents the first documentation of enduring discrepancies among Kardemir stocks. The identification of such consistent differences provides traders with the potential to strategically employ pair trading techniques in a profitable manner. The second analysis examines the coherence between their daily returns through wavelet coherence analysis. While Kardemir stocks generally exhibit strong coherence, there were prolonged instances where the coherence among them weakened. Finally, a comparative analysis of LSTM models with different specifications is provided. The variations in specifications arise from three sources: (i) the decision to use or not use dual stock prices as predictors, (ii) the number of lags employed as predictors, and (iii) the selection of the training set. This final analysis does not intend to compare LSTM models with various other machine learning models. Instead, it focuses on addressing two key questions:
Question 1: Can past dual-class stock prices effectively forecast future prices of each other?
Question 2: Does the use of longer price lags enhance forecast performance in LSTM models?

2. Methodology

2.1. Data and Variables

The data used in this study is based on daily high and low prices (in Turkish liras) for Kardemir stocks (tickers: KRDMA, KRDMB, KRDMD) between Jan 2001 and July 2023. The sole data source is Bloomberg. To capture the most likely intraday price, this study uses daily mid-prices. The daily mid-prices are calculated as follows:

$$HL = 0.5(high + low) \tag{1}$$

The daily premium of Group *i* over Group *j* is calculated as the percentage difference between daily mid-prices for Group *i* and daily mid-prices for Group *j*.

$$v_{ij} = HL_i/\mathrm{HL}_j - 1 \tag{2}$$

2.2. Wavelet Coherence Analysis

This paper uses the continuous wavelet transform (CWT) to quantify the magnitude, direction, and lead-lag effects between Kardemir stocks. This approach has a number of advantages. First, it uncovers the dynamic relationship between these stocks, allowing me to distinguish between periods at which prices are linked. Secondly, using the CWT, it is possible to identify changes in the direction of the relationship over time. Finally, the CWT provides insights about the relationship between these stocks at different time horizons simultaneously.

According to Torrence and Campo [7], the wavelet coefficients $W_{\varepsilon,\tau}$ associated with a time series $f(t)$ are calculated as:

$$W_{\varepsilon,\tau} = \sum_{t=1}^{n} f(t)\,\psi^{*}\left[\frac{t-\tau}{\varepsilon}\right] \tag{3}$$

where * represent the complex conjugate, $\varepsilon > 0$ is the scale associated with the wavelet and $\tau \in [-\alpha, \alpha]$ is the window location and $1/\varepsilon$ is the normalization factor. Here, the Morlet wavelet with wave number $\omega_0 = 6$ is used following Grinsted et al. [2]. More specifically, the Morlet wavelet is formulated as:

$$\psi(t) = \pi^{0.25} e^{i\omega_0 t} e^{\frac{-t^2}{2}}. \quad (4)$$

The cross-wavelet power spectrum is calculated as the product of two wavelet coefficients and represents the common variation between two time series over time and scale. It is formulated as:

$$W_{\varepsilon,\tau}(f,g) = W_{\varepsilon,\tau}(f) W^*_{\varepsilon,\tau}(g). \quad (5)$$

Like the correlation, the wavelet squared coherency is defined by normalizing the smoothed cross-wavelet power spectrum by the smoothed wavelet power spectrum associated with the individual time series:

$$\rho^2_{\varepsilon,\tau} = \frac{\left|Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f,g)\right)\right|^2}{\left|Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f)\right)\right|^2 \left|Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(g)\right)\right|^2} \quad (6)$$

where Q is the smoothing operator. By construction, $\rho^2_{\varepsilon,\tau}$ takes values between 0 and 1. It implies no comovement when $\rho^2_{\varepsilon,\tau} = 0$, and perfect comovement when $\rho^2_{\varepsilon,\tau} = 1$. To identify statistically significant squared coherency regions, the study uses a Monte-Carlo method with 1,000 iterations.

To uncover lead-lag effects, the following wavelet multi-scale phase is used:

$$\theta_{\varepsilon,\tau}(f,g) = tan^{-1}\left(\frac{\mathcal{I}\left(Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f,g)\right)\right)}{\mathcal{R}\left(Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f,g)\right)\right)}\right). \quad (7)$$

Here, $\mathcal{I}$ and $\mathcal{R}$ represent the imaginary and real components of the wavelet coefficients. Phase arrows are utilized within wavelet coherence plots to depict the direction of simultaneous movement and the effects of leading or lagging. Arrows pointing east (west) signify being in (out of) sync, while arrows pointing north (south) indicate that one time series leads (lags) the other. When the phase arrow points in a northeast (southeast) direction, it means that the two series are in sync, but the second one (or first one) leads the first one (or second one). Differing outcomes are conveyed by arrows facing northwest and southwest.

2.3. Long Short-Term Memory (LSTM)

While training a recurrent neural network, each iteration receives an update proportional to the partial derivative of the error function with respect to its current weight. When the gradient is vanishingly small, the training may slow and, in some cases, stops [39]. The long short-term memory technique [41] is developed as a potential solution for the vanishing gradient problem. The LSTM approach is widely used in predicting stock prices because of its capacity to recognize patterns and generate more accurate predictions compared to other methods [36], [17], [33], [23], and [31].

An LSTM unit consists of a cell, and within this cell, there are three gates that manage the movement of information and regulate the cell state. These gates include an input gate, an output gate, and a forget gate. The LSTM units are then interconnected, forming a chain where each individual cell acts as a memory module within the LSTM architecture. Figure 1 illustrates a standard LSTM architecture and Figure 2 shows a standard LSTM cell architecture.

In Figure 2, f_t, i_t, and o_t respectively represent the forget gate, input gate, and output gate. Also, X_t is the input, h_t is the output, C_t is the cell state, and $\hat{C}_t$ is the internal cell state. Based on the input, previous output, and previous cell state (X_t, h_{t-1}, and C_{t-1}); f_t, i_t, o_t, $\hat{C}_t$, C_t, and h_t are calculated as follow:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, X_t] + b_f\right) \quad (8)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (9)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (10)$$

$$\hat{C}_t = tanh(W_C \cdot [h_{t-1}, X_t] + b_C) \quad (11)$$

$$C_t = i_t \cdot \hat{C}_t + f_t \cdot C_{t-1} \quad (12)$$

$$h_t = o_t \times tanh(C_t) \quad (13)$$

Here, σ represents the sigmoid function and $tanh$ represents the hyperbolic tangent function.

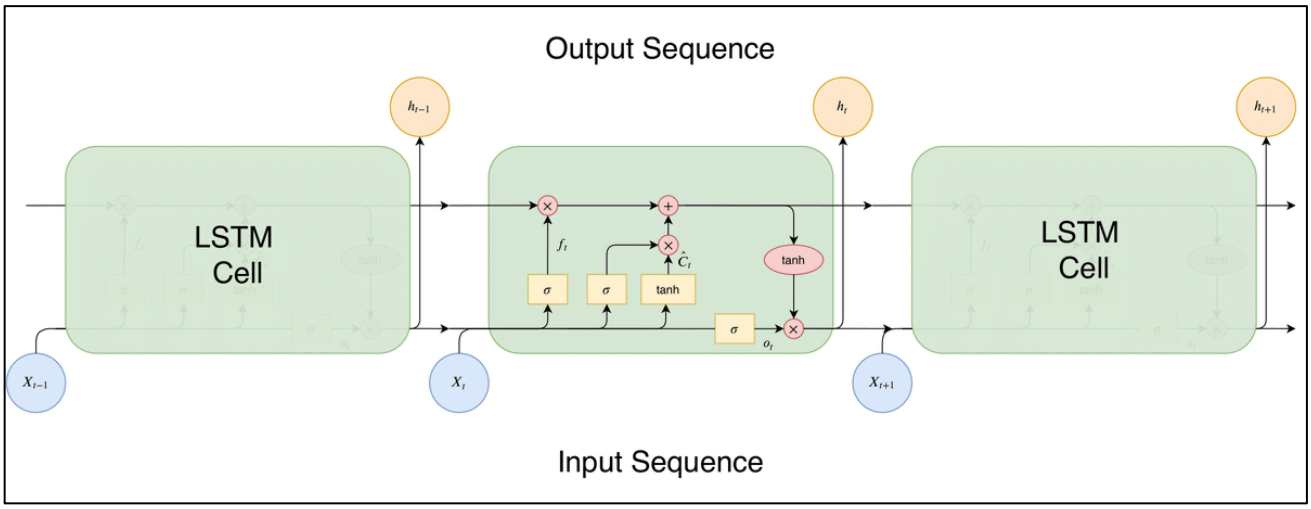


Figure 1: LSTM Architecture (source: [45])

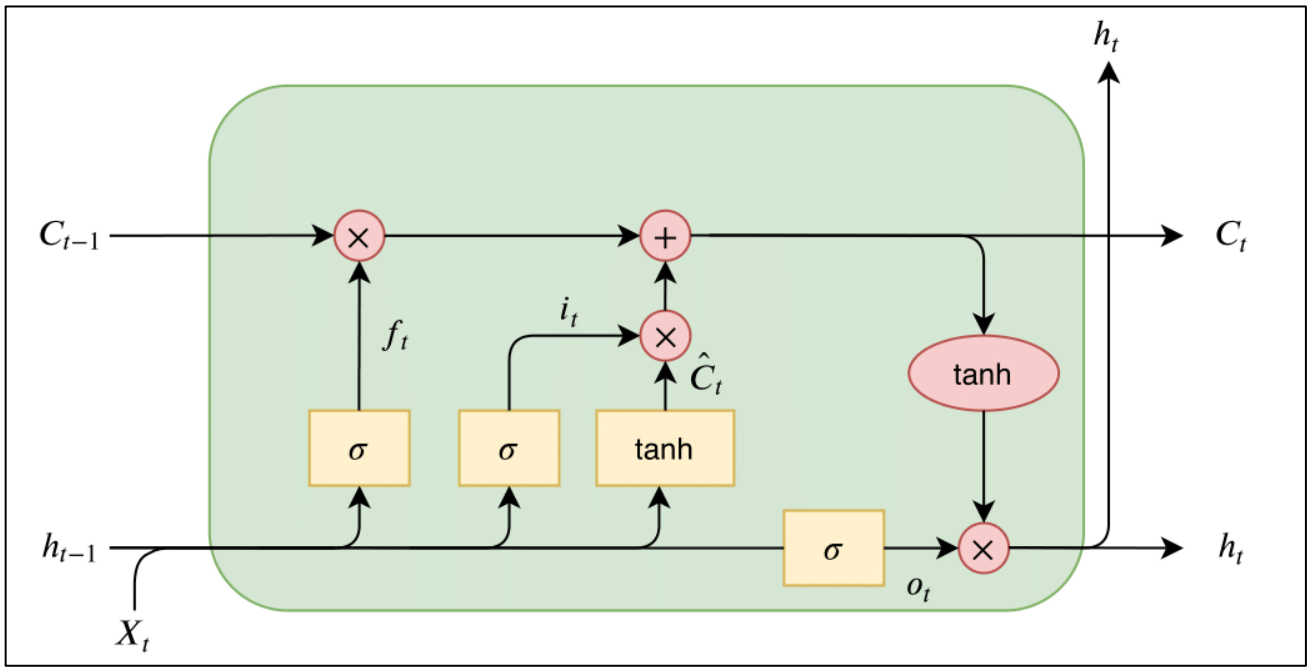


Figure 2: LSTM Cell Architecture (source: [45])

An LSTM model, operating as a black-box method, has the potential to exhibit overfitting issues, diminishing its effectiveness when applied to new, untrained data. To gauge the extent of overfitting, data is typically divided into two subsets in the realm of machine learning: the training set and the validation set. The training set is utilized for model development, while the validation set remains untouched during the training phase and serves to evaluate the model's predictive performance.

Traditionally, practitioners have favored training their models on large datasets with numerous observations, a practice grounded in the law of large numbers. However, this study posits that this conventional approach may not be well-suited for forecasting financial variables. In essence, it argues that a model trained on the daily prices of an asset spanning long periods (several lagged prices) may not perform as effectively as a model trained solely on the more recent daily prices.

One potential reason for this discrepancy lies in the fact that a long-range training set encompasses both downward and upward market trends. Such training data may not yield accurate predictions when applied to data sampled during a trend in a single direction. Consequently, utilizing more recent data points as the training set may lead to superior predictive performance.

In this paper, the use of rolling training sets is proposed for predicting the next observation. In this scenario, the 5,300th, 5,301st, 5,302nd, 5,303rd, and 5,304th observations serve as the training set to forecast the 5,305th observation when the training window is set to 5. This approach ensures that every observation is predicted based on the most recent price action, rather than relying on price action from hundreds of days ago.

To test the effectiveness of this new approach, 2 training set rules are used:

Approach 1 (Fixed horizon): The entire dataset is divided into two mutually exclusive and collectively exhaustive sets, namely a training set and a test set. The first 5,282 observations are used as the training set and the remaining 300 observations are used as the test set. In this case, all observations among the last 300 observations are used for forecasting based on a single model developed by using the first 5,282 observations.

Approach 2 (Sliding window): For every observation among the last 300 observations, the prior 5, 10, 20, and 50 observations are used as the training set. In this approach, a model is trained to forecast the next observation for each training window.

In total, 2,408 LSTM models with configurations above are trained to forecast the last 300 observations in the sample. As a preprocessing step, a transformation on the daily mid-prices by subtracting 100 from each value and then scaling the results by a factor of 1/100 is performed, resulting in the formula x/100-1. This scaling operation has the effect of confining all observations within the range of -1 to 1, consistent with the range of tanh function used in LSTM models. It's important to note that this scaling choice is based on the assumption that the daily mid-prices will consistently remain below 100 Turkish Lira (TRY). The selection of this threshold is based on the observation that all data points within the fixed-range training set are significantly lower than the chosen threshold. Consequently, the study has intentionally refrained from constraining the model to only produce forecasts that surpass the maximum value observed in the training set.

Finally, a deliberate decision was made to avoid using the conventional min-max scaling method. This choice was driven by the understanding that min-max scaling assumes prior knowledge of the range of daily mid-prices in the test set. However, in this context, the range of these mid-prices is considered unknown since they are the very values we aim to forecast.

3. Results

3.1. Historical Premiums

Figure 3 shows that KRDMA was traded at a premium relative to KRDMB for 1,171 days out of a total of 5,582 days. In 2020, the premium of KRDMA over KRDMB was the strongest when KRDMA predominantly traded at a premium for most of that year. Conversely, KRDMD consistently saw substantial discounts relative to both KRDMA and KRDMB. For 4,360 days, KRDMA was traded at a premium. Similarly, KRDMB was traded at a premium status for 4,318 days. Figure 1 also demonstrates a substantial decrease in premiums paid for KRDMA and KRDMB over KRDMD, starting from 2018. A reversion occurred starting in 2021 for KRDMD discounts. Since 2021, both KRDMA and KRDMB have consistently been traded at a discount relative to KRDMD.

Summary statistics of these premiums are given in Table 1. It clearly shows that substantial premiums paid for KRDMA and KRDMB over KRDMD between 2001 and 2023. At their heights these premiums reached 235.87% and 361.21% respectively for KRMDA and KRDMB.

Table 1. Summary Statistics of Premiums

	KRDMA over KRDMB	*KRDMA over KRDMD*	*KRDMB over KRDMD*
Minimum	-59.70%	-33.08%	-33.21%
1st Quartile	-19.10%	5.99%	12.98%
Median	-7.71%	35.47%	57.82%
Mean	-10.30%	49.44%	75.73%
3rd Quartile	0.00%	73.51%	128.13%
Maximum	57.06%	235.87%	361.21%

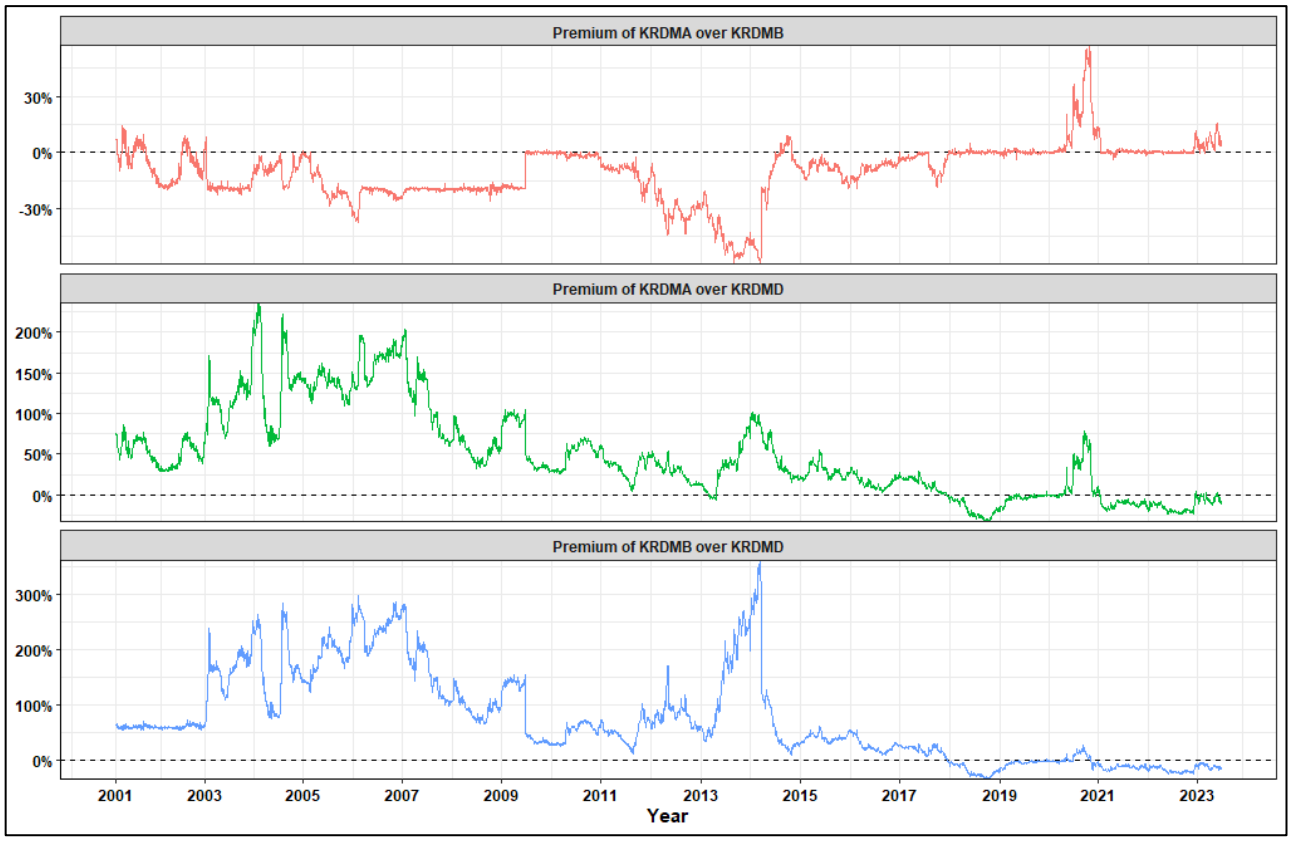


Figure 3: Historical Premiums

3.2. Wavelet Coherence Analysis

This section presents the dynamic relationship between daily returns (calculated as percentage change in daily mid-prices) of Kardemir stocks by using the wavelet coherence technique explained above.

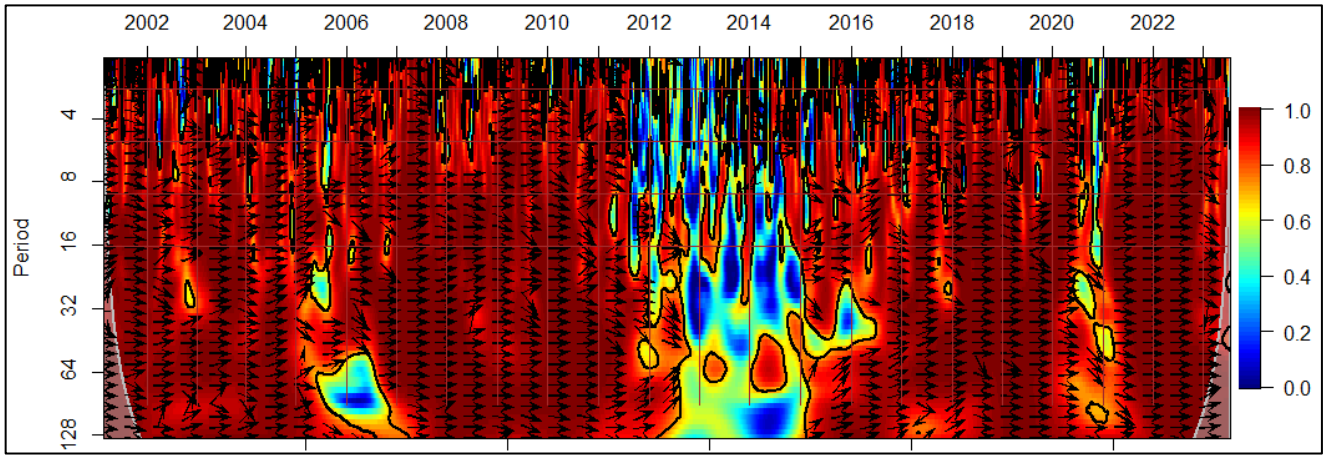


Figure 4: Wavelet Coherence between KRDMA and KRDMB Daily Returns

As depicted in Figure 4, the coherence between the daily returns of KRDMA and KRDMB remained consistently strong throughout the analyzed period, with only a few exceptions. The first notable deviation occurred in 2006, spanning periods 64 to 128, during which there was a clear lack of coherence between the daily return patterns of the two stocks. The second significant divergence surfaced in 2012 and persisted for over two years. During this period, the daily returns of KRDMA and KRDMB exhibited noticeable discrepancies. It's worth highlighting that over this time frame, the premium of KRDMA over KRDMB reached its lowest point, showing a substantial decline of -59.64%. Importantly, the figure also emphasizes the absence of a substantial cause-and-effect relationship between these two series. Instead, their temporal progression displayed synchronized movements, without any prominent identifiable temporal precedence or lag.

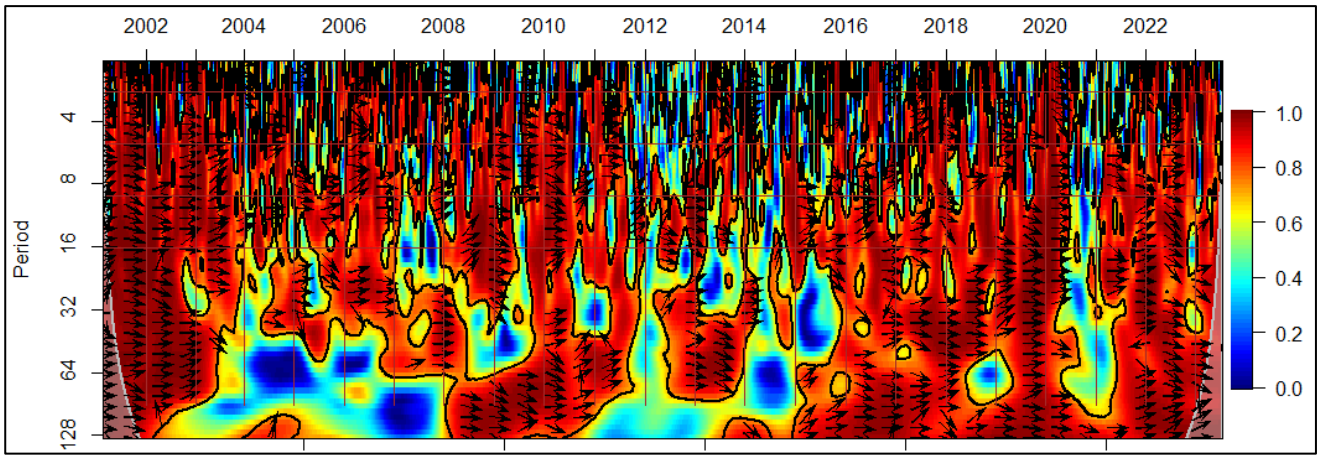


Figure 5: Wavelet Coherence between KRDMA and KRDMD Daily Returns

Figure 5 illustrates the results of the wavelet coherence analysis applied to the relationship between KRDMA and KRDMD. Across most of the time span, from January 2001 to July 2023, the daily returns of these two stocks displayed significant synchronization. It's worth noting that during this period, the coherence between the two stocks was notably strong and consistent, contributing to their aligned behavior.

Valuable insights can be derived from the coherence patterns at various time scales. Specifically, between 2003 and 2008, as well as intermittently in the first quarter of 2012 and throughout 2014, the coherence associated with longer cycles, particularly those spanning from 64 to 128 days, was relatively weak when compared to the coherence observed within shorter cycles, ranging from 8 to 16 days. This observation underscores temporal variations in the degree of synchronization across different scales, highlighting periods of heightened and diminished shared behavior.

Moreover, a clear absence of coherence becomes evident in the latter part of 2020 across various time cycles. During this specific period, the synchronization between the two stocks was notably absent. What's particularly noteworthy is that this timeframe coincided with a substantial premium of over 75% attributed to KRDMA over KRDMD. This convergence of factors highlights the potential interplay between coherence patterns and premium fluctuations, implying intricate dynamics at play in the relationship between these stock returns during this period.

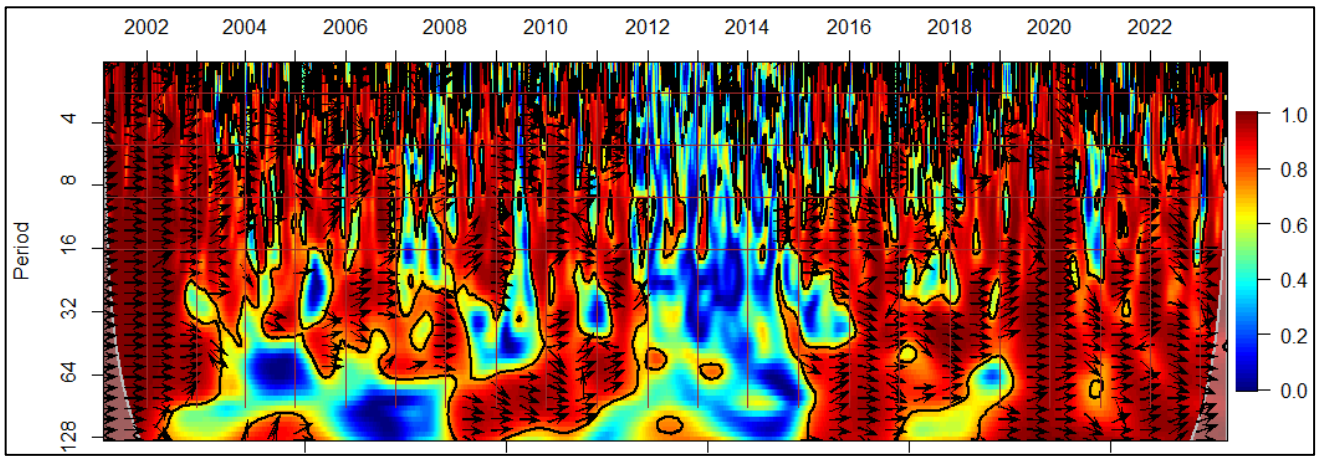


Figure 6: Wavelet Coherence between KRDMB and KRDMD Daily Returns

The final wavelet coherence plot, depicted as Figure 6, reveals the weakest coherence observed among Kardemir stock returns, specifically between KRDMB and KRDMD. Notably, the coherence between the daily returns of KRDMB and KRDMD was particularly weak, primarily spanning the years from 2012 to 2015. This period coincided with a time when the premium paid for KRDMB over KRDMD reached its peak, skyrocketing to an unprecedented level of 361.21%.

Across the broader time frame spanning from January 2001 to 2012, a robust coherence pattern was evident among the daily returns of the three Kardemir stocks across various time cycles. This robust coherence paradigm underwent a transformation, transitioning to a less robust coherence configuration from 2012 to 2015, only to reemerge in 2016. Another episode of coherence weakness emerged towards the latter part of 2020, encompassing

all Kardemir stocks within shorter cycles. Typically, the highest coherence was observed between KRDMA and KRDMB.

Furthermore, the wavelet plots emphasize the absence of a clearly discernible leading or lagging relationship between various Kardemir stocks. Notably, instances of significant divergence in the daily mid-prices of Kardemir stocks coincided with the absence of coherence, particularly within longer cycles. This observation suggests a potential complex interplay between coherence dynamics and price disparities in the long-term context**.**

3.3. Long Short-Term Memory (LSTM) Models

This section presents the predictive performance results obtained from a total of 2,408 LSTM models. These models were trained with the specific goal of forecasting the daily mid-prices of the last 300 observations. To evaluate the effectiveness of dual-class stocks as predictors for each other, two sets of models are employed.

In the first set of models, the prediction models did not incorporate the historical price movements of dual stocks as lagged variables when forecasting future mid-prices. In contrast, the second set of models was designed to include lagged dual-class stock mid-prices as factors for predicting future mid-prices. For these lagged variables, two options are considered: using 4 lags and 9 lags of daily mid-prices. All these models were trained following one of the training-set rules outlined above. In summary, there are three significant distinctions among these specifications:

Criterion 1: Inclusion of dual-class stock prices as predictors (Possible values: Yes or No).
Criterion 2: The choice of the number of lags of previous stock prices as predictors (Possible values: 4 or 9).
Criterion 3: Selection of the training set (Possible values: The first 5,282 observations, the most recent 5 observations, the most recent 10 observations, the most recent 20 observations, or the most recent 50 observations).

To evaluate and compare their predictive performance, three key metrics are employed: root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Lower metric values indicate better predictive performance. These statistics for all models are reported in Tables 3-5. In Table 2, the results of t-tests based on the null hypothesis suggesting that the mean prediction error is zero are presented. The table displays mean prediction errors and p-values (in parentheses) associated with these t-tests.

Table 2 reveals that using a fixed-range training set, as opposed to a sliding window, consistently leads to underprediction, except for KRDMB with 4 lags and dual-class stock prices. Supporting this finding, Tables 3-5 show that the model employing a 5-day training window outperforms others across all predictive performance metrics. This finding aligns with the thesis against extended training periods in financial data forecasting and is consistent with prior research on investor and managerial myopia [15], [4], [8], and [50].

It's worth highlighting that the validation set, comprising the last 300 data points, coincided with a period marked by strong coherence between the daily returns of KRDMA, KRDMB, and KRDMD. Despite this robust coherence, the inclusion of dual stocks as predictors did not yield any noticeable additional insights into future stock prices. In fact, there is often a modest decline in accuracy as illustrated in Tables 3-5.

4. Conclusions

In this study, a case of dual-class stock structure on Borsa Istanbul was examined. This case is marked by a distinctive characteristic: prolonged disparities among three stocks, namely KRDMA, KRDMB, and KRDMD. These disparities reached staggering heights, with differentials soaring as high as 361.21%. To put this into perspective, consider that the most significant divergence observed between GOOG and GOOGL on NASDAQ has been approximately 5% since 2005. This stark contrast highlights that arbitrage opportunities between KRDMA, KRDMB, and KRDMD may present greater profit potential.

Moreover, this study also sheds light on the fact that even when there exists a strong coherence between dual-class stock prices, these prices may not necessarily serve as reliable predictors for future price movements of each other. Lastly, the study offers substantial empirical evidence supporting the practice of favoring shorter training periods over extended ones, contrary to the common practice of employing larger training sets with numerous observations and lags.

Overall, these findings underscore the importance of exercising caution when choosing training sets and predictors while training LSTM models. The inclusion of additional predictors, even those strongly coherent with the target variable, and the extension of training sets to encompass past values in financial time series have the potential to diminish predictive performance, ultimately resulting in poor forecasting.

Table 2. T-test Results

	Lag = 4		**Lag = 9**	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.020 (0.639)	0.034 (0.474)	0.025 (0.595)	0.034 (0.466)
KRDMB	0.021 (0.603)	0.030 (0.487)	0.019 (0.658)	0.015 (0.736)
KRDMD	0.026 (0.573)	0.025 (0.603)	0.030 (0.548)	0.035 (0.484)
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.038 (0.467)	0.031 (0.585)	-0.003 (0.963)	0.451 (0.049)
KRDMB	0.052 (0.294)	0.027 (0.594)	0.017 (0.764)	-0.001 (0.991)
KRDMD	0.048 (0.377)	0.033 (0.584)	0.031 (0.629)	0.008 (0.904)
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.097 (0.149)	0.139 (0.044)	0.009 (0.907)	0.064 (0.451)
KRDMB	0.087 (0.199)	0.041 (0.535)	0.121 (0.118)	0.094 (0.241)
KRDMD	0.137 (0.075)	0.050 (0.544)	0.097 (0.247)	-0.015 (0.857)
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.038 (0.655)	0.117 (0.165)	0.035 (0.710)	0.171 (0.114)
KRDMB	0.123 (0.130)	0.134 (0.105)	0.167 (0.073)	-0.022 (0.815)
KRDMD	0.062 (0.461)	0.215 (0.025)	0.239 (0.014)	0.257 (0.024)
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	-0.122 (0.009)	-0.126 (0.010)	-0.635 (<0.001)	-0.166 (0.005)
KRDMB	-0.167 (<0.001)	0.074 (0.099)	-0.399 (<0.001)	-0.343 (<0.001)
KRDMD	-0.210 (<0.001)	-0.328 (<0.001)	-0.742 (<0.001)	-0.378 (<0.001)

Table 3. KRDMA Predictive Performance Results

	Lag = 4		**Lag = 9**	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.7433	0.8106	0.8006	0.8080
MAE	0.5077	0.5466	0.5578	0.5575
MAPE	3.387	3.6617	3.7375	3.7516
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.9027	0.9698	1.0020	1.1226
MAE	0.6230	0.6745	0.7044	0.7667
MAPE	4.2066	4.5580	4.8132	5.2047
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.1606	1.1915	1.4265	1.4671
MAE	0.8411	0.7863	0.9940	1.0660
MAPE	5.7301	5.3693	6.8049	7.3317
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.4709	1.4571	1.6515	1.8731
MAE	1.1134	1.0881	1.2367	1.3784
MAPE	8.0633	7.7143	8.9525	9.8199
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.8173	0.8510	1.1949	1.0517
MAE	0.5873	0.6132	0.9246	0.7664
MAPE	3.9752	4.1735	6.3747	5.2103

Table 4. KRDMB Predictive Performance Results

	Lag = 4		**Lag = 9**	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.6834	0.7567	0.7440	0.7757
MAE	0.4601	0.4995	0.5047	0.5251
MAPE	3.2067	3.4693	3.5233	3.6607
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.8587	0.8917	1.0084	1.0299
MAE	0.5756	0.6030	0.6862	0.7005
MAPE	4.0155	4.2513	4.8465	4.9389
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.1757	1.1435	1.3400	1.3932
MAE	0.7796	0.7850	0.8755	0.9492
MAPE	5.4603	5.5653	6.2216	6.6879
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.4014	1.4328	1.6128	1.6676
MAE	1.0322	1.0424	1.2265	1.2750
MAPE	7.6067	7.6053	9.0592	9.5743
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.7612	0.7824	0.9884	0.9943
MAE	0.5407	0.5490	0.7306	0.7220
MAPE	3.8301	3.9241	5.2094	5.0590

Table 5. KRDMD Predictive Performance Results

	Lag = 4		**Lag = 9**	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.7910	0.8321	0.8596	0.8564
MAE	0.5608	0.5811	0.6050	0.6117
MAPE	3.4491	3.5723	3.7123	3.7537
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.9408	1.0293	1.1151	1.1721
MAE	0.6584	0.7155	0.7678	0.8209
MAPE	4.0577	4.4043	4.7642	5.0716
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.3288	1.4243	1.4513	1.4485
MAE	0.9194	1.0142	1.0385	1.0465
MAPE	5.6409	6.3443	6.4796	6.5548
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.4489	1.6598	1.6948	1.9712
MAE	1.1061	1.2782	1.2880	1.5104
MAPE	7.1237	8.1918	8.1252	9.7193
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.8730	0.9280	1.2142	1.1088
MAE	0.6322	0.6894	0.9765	0.8100
MAPE	3.9094	4.2723	6.3353	5.0450

Declaration of Interest

The author declares that there is no conflict of interest.

Acknowledgements

The author would like to express sincere gratitude to Ayşegül Alaybeyoğlu, the editor, Muhammet Mustafa Bahşi, the section editor, and the two anonymous referees for their valuable suggestions, feedback, and professionalism.

References

[1] A. Bylund, "What's the difference between Alphabet's stock tickers, GOOG and GOOGL?," The Motley Fool, https://www.fool.com/investing/2022/07/27/whats-the-difference-between-goog-and-googl/ (accessed Aug. 16, 2023).
[2] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to Geophysical Time Series," Nonlinear Processes in Geophysics, vol. 11, no. 5/6, pp. 561–566, 2004. doi:10.5194/npg-11-561-2004
[3] B. F. Smith and B. Amoako-Adu, "Relative prices of dual class shares," The Journal of Financial and Quantitative Analysis, vol. 30, no. 2, p. 223, 1995. doi:10.2307/2331118
[4] C. D. Rio and R. Santamaria, "Stock characteristics, investor type, and market myopia," Journal of Behavioral Finance, vol. 17, no. 2, pp. 183–199, 2016. doi:10.1080/15427560.2016.1170682
[5] C. Erten, N. Chotai, and D. Kazakov, "Pair trading with an ontology of SEC Financial Reports," 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020. doi:10.1109/ssci47803.2020.9308384
[6] C. Li, D. Songand D. Tao, "Multi-task Recurrent Neural Networks and Higher-order Markov Random Fields for Stock Price Movement Prediction: Multi-task RNN and Higer-order MRFs for Stock Price Classification", ACM, Jul. 2019. doi: 10.1145/3292500.3330983.
[7] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," Bulletin of the American Meteorological Society, vol. 79, no. 1, pp. 61–78, 1998. doi:10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2
[8] C. W. Holden and L. L. Lundstrum, "Costly trade, managerial myopia, and long-term investment," Journal of Empirical Finance, vol. 16, no. 1, pp. 126–135, 2009. doi:10.1016/j.jempfin.2008.05.001
[9] C.-H. Chen, W.-H. Lai, and T.-P. Hong, "An effective correlation-based pair trading strategy using genetic algorithms," Computational Collective Intelligence, pp. 255–263, 2021. doi:10.1007/978-3-030-88081-1_19
[10] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," Expert Systems with Applications, vol. 129, pp. 273–285, 2019. doi:10.1016/j.eswa.2019.03.029
[11] E. Tokat and A. C. Hayrullahoğlu, "Pairs trading: Is it applicable to exchange-traded funds?," Borsa Istanbul Review, vol. 22, no. 4, pp. 743–751, 2022. doi:10.1016/j.bir.2021.08.001
[12] H. Maqsood et al., "A local and global event sentiment based efficient stock exchange forecasting using Deep Learning," International Journal of Information Management, vol. 50, pp. 432–451, 2020. doi:10.1016/j.ijinfomgt.2019.07.011
[13] H. S. Sim, H. I. Kim, and J. J. Ahn, "Is deep learning for image recognition applicable to stock market prediction?," Complexity, vol. 2019, pp. 1–10, 2019. doi:10.1155/2019/4324878
[14] H. Yang, Y. Zhu, and Q. Huang, "A multi-indicator feature selection for CNN-Driven Stock Index Prediction," Neural Information Processing, pp. 35–46, 2018. doi:10.1007/978-3-030-04221-9_4
[15] I. Nyman, "Stock market speculation and managerial myopia," Review of Financial Economics, vol. 14, no. 1, pp. 61–79, 2005. doi:10.1016/j.rfe.2004.06.002
[16] J. Eapen, D. Bein, and A. Verma, "Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 2019. doi:10.1109/ccwc.2019.8666592
[17] J. M.-T. Wu et al., "A graphic CNN-LSTM model for stock price prediction," Artificial Intelligence and Soft Computing, pp. 258–268, 2021. doi:10.1007/978-3-030-87986-0_23
[18] J. P. Ramos-Requena, M. N. López-García, M. A. Sánchez-Granero, and J. E. Trinidad-Segovia, "A cooperative dynamic approach to pairs trading," Complexity, vol. 2021, pp. 1–8, 2021. doi:10.1155/2021/7152846
[19] J. Wu, A pairs trading strategy for GOOG/GOOGL using machine learning, https://cs229.stanford.edu/proj2015/028_report.pdf (accessed Aug. 16, 2023).
[20] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial time-series data analysis using deep convolutional neural networks," 2016 7th International Conference on Cloud Computing and Big Data (CCBD), 2016. doi:10.1109/ccbd.2016.027
[21] K. Nakagawa, T. Uchida, and T. Aoshima, "Deep factor model," ECML PKDD 2018 Workshops, pp. 37–50, 2019. doi:10.1007/978-3-030-13463-1_3
[22] K. Rydqvist, "Dual-class shares: A Review," Oxford Review of Economic Policy, vol. 8, no. 3, pp. 45–57, 1992. doi:10.1093/oxrep/8.3.45
[23] Ko, Ching-Ru, and Hsien-Tsung Chang. "LSTM-Based Sentiment Analysis for Stock Price Forecast." PeerJ Computer Science, vol. 7, 11 Mar. 2021, p. e408, https://doi.org/10.7717/peerj-cs.408.
[24] L. Ni et al., "Forecasting of forex time series data based on Deep Learning," Procedia Computer Science, vol. 147, pp. 647–652, 2019. doi:10.1016/j.procs.2019.01.189
[25] L. Zhang, "Pair trading with machine learning strategy in China Stock Market," 2021 2nd International Conference on Artificial Intelligence and Information Systems, 2021. doi:10.1145/3469213.3471353
[26] L. Zingales, "The value of the voting right: A study of the Milan stock exchange experience," Review of Financial Studies, vol. 7, no. 1, pp. 125–148, 1994. doi:10.1093/rfs/7.1.125
[27] M. Abe and H. Nakayama, "Deep learning for forecasting stock returns in the cross-section," Advances in Knowledge Discovery and Data Mining, pp. 273–284, 2018. doi:10.1007/978-3-319-93034-3_22

[28] M. R. Horner, “The value of the corporate voting right,” Journal of Banking and Finance, vol. 12, no. 1, pp. 69–83, 1988. doi:10.1016/0378-4266(88)90051-9
[29] M. U. Gudelek, S. A. Boluk, and A. M. Ozbayoglu, “A deep learning based stock trading model with 2-D CNN trend detection,” 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017. doi:10.1109/ssci.2017.8285188
[30] M. Wen, P. Li, L. Zhang, and Y. Chen, “Stock market trend prediction using high-order information of Time Series,” IEEE Access, vol. 7, pp. 28299–28308, 2019. doi:10.1109/access.2019.2901842
[31] N. Foysal Ahamed, and M. Mahmudul Hasan. “Predicting Stock Price from Historical Data using LSTM Technique.” Journal of Artificial Intelligence and Data Science 3.1: 36-49.
[32] N. Naik and B. R. Mohan, “Stock price movements classification using machine and deep learning techniques-the case study of Indian Stock Market,” Engineering Applications of Neural Networks, pp. 445–452, 2019. doi:10.1007/978-3-030-20257-6_38
[33] Niu, Hongli, et al. “A Hybrid Stock Price Index Forecasting Model Based on Variational Mode Decomposition and LSTM Network.” Applied Intelligence, vol. 50, no. 12, 17 July 2020, pp. 4296–4309, https://doi.org/10.1007/s10489-020-01814-0.
[34] P. Oncharoen and P. Vateekul, “Deep learning using risk-reward function for stock market prediction,” Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, 2018. doi:10.1145/3297156.3297173
[35] P. Patil, C.-S. M. Wu, K. Potika, and M. Orang, “Stock market prediction using ensemble of graph theory, machine learning and Deep Learning Models,” Proceedings of the 3rd International Conference on Software Engineering and Information Management, 2020. doi:10.1145/3378936.3378972
[36] P. Srivastava and P. K. Mishra, “Stock market prediction using RNN LSTM,” 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021. doi:10.1109/gcat52182.2021.9587540
[37] R. C. Lease, J. J. McConnell, and W. H. Mikkelson, “The market value of control in publicly-traded corporations,” Journal of Financial Economics, vol. 11, no. 1–4, pp. 439–471, 1983. doi:10.1016/0304-405x(83)90019-3
[38] S. Basodi, C. Ji, H. Zhang, and Y. Pan, “Gradient amplification: An efficient way to train deep neural networks,” Big Data Mining and Analytics, vol. 3, no. 3, pp. 196–207, 2020. doi:10.26599/bdma.2020.9020004
[39] R. Zhang, Z. Yuan, and X. Shao, “A new combined CNN-RNN model for sector stock price analysis,” 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2018. doi:10.1109/compsac.2018.10292
[40] S. Cai, X. Feng, Z. Deng, Z. Ming, and Z. Shan, “Financial News quantization and Stock Market Forecast Research based on CNN and LSTM,” Lecture Notes in Computer Science, pp. 366–375, 2018. doi:10.1007/978-3-030-05755-8_36
[41] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735
[42] S. Liu, C. Zhang, and J. Ma, “CNN-LSTM neural network model for quantitative strategy analysis in stock markets,” Neural Information Processing, pp. 198–206, 2017. doi:10.1007/978-3-319-70096-0_21
[43] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, “Forecasting stock market crisis events using deep and Statistical Machine Learning Techniques,” Expert Systems with Applications, vol. 112, pp. 353–371, 2018. doi:10.1016/j.eswa.2018.06.032
[44] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, “Stock price prediction using LSTM, RNN and CNN-sliding window model,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017. doi:10.1109/icacci.2017.8126078
[45] Thorir Mar Ingolfsson, “Insights into LSTM architecture,” Thorir Mar Ingolfsson, https://thorirmar.com/post/insight_into_lstm/ (accessed Aug. 16, 2023).
[46] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, “Leveraging Social Media News to predict stock index movement using RNN-Boost,” Data & Knowledge Engineering, vol. 118, pp. 14–24, 2018. doi:10.1016/j.datak.2018.08.003
[47] W. Jiang, “Applications of deep learning in stock market prediction: Recent progress,” Expert Systems with Applications, vol. 184, p. 115537, 2021. doi:10.1016/j.eswa.2021.115537
[48] W. L. Megginson, “Restricted voting stock, acquisition premiums, and the market value of corporate control,” The Financial Review, vol. 25, no. 2, pp. 175–198, 1990. doi:10.1111/j.1540-6288.1990.tb00791.x
[49] X. Ding, Y. Zhang, T. Liuand J. Duan, “Deep learning for event-driven stock prediction”, AAAI Press, Jul. 2015.
[50] X. Sheng, S. Guo, and X. Chang, “Managerial myopia and firm productivity: Evidence from China,” Finance Research Letters, vol. 49, p. 103083, 2022. doi:10.1016/j.frl.2022.103083
[51] Y. Liu, Q. Zeng, H. Yang, and A. Carrio, “Stock price movement prediction from financial news with deep learning and knowledge graph embedding,” Knowledge Management and Acquisition for Intelligent Systems, pp. 102–113, 2018. doi:10.1007/978-3-319-97289-3_8
[52] Y. Song, J. W. Lee, and J. Lee, “A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction,” Applied Intelligence, vol. 49, no. 3, pp. 897–911, 2018. doi:10.1007/s10489-018-1308-x
[53] Y. Zhao and M. Khushi, “Wavelet denoised-resnet CNN and LIGHTGBM method to predict forex rate of Change,” 2020 International Conference on Data Mining Workshops (ICDMW), 2020. doi:10.1109/icdmw51313.2020.00060
[54] Z. Hu, Y. Zhao, and M. Khushi, “A survey of Forex and Stock Price Prediction using Deep learning,” Applied System Innovation, vol. 4, no. 1, p. 9, 2021. doi:10.3390/asi4010009
[55] Z. Zeng and M. Khushi, “Wavelet denoising and attention-based RNN- Arima model to predict forex price,” 2020 International Joint Conference on Neural Networks (IJCNN), 2020. doi:10.1109/ijcnn48605.2020.9206832