

The cover features a blue background with a faint, stylized illustration of a human brain. The brain is composed of a network of interconnected nodes and lines, representing neural connections. The top half of the cover is a solid blue color, while the bottom half is a darker blue with a subtle grid pattern. The title is centered in a white, bold, sans-serif font.

**Journal of Soft Computing
and
Artificial Intelligence**

ISSN: 2717-8226



Journal of Soft Computing and Artificial Intelligence

Journal homepage: <https://dergipark.org.tr/en/pub/jscai>

International
Open Access 

Volume 03
Issue 01

June, 2022

Journal of Soft Computing and Artificial Intelligence (JSCAI) is an international peer-reviewed journal that publishes integrated research articles in all areas of soft computing and artificial intelligence. The aim of the JSCAI journal is to provide a platform for researchers, professionals, and academicians around the world to combine and exchange new developments and their applications in various areas of soft computing and artificial intelligence. Journal of Soft Computing and Artificial Intelligence (JSCAI) is an international peer-reviewed journal that publishes integrated research articles in all areas of soft computing and artificial intelligence. The journal covers all branches of engineering, including mechanics, computer science, electronics, energy, aerospace engineering, materials science, nuclear engineering, systems analysis, alternative technologies, etc.

JSCAI publication, which is open access, is free of charge. There is no article submission and processing charges (APCs).

JSCAI is indexed & abstracted in:

Crossref (Doi beginning: 10.55195/jscai..xxxxxx)

Directory of Research Journals Indexing (DRJI)

Google Scholar

Index Copernicus (ICI Journal Master List)

OpenAIRE

Asos Index

Directory of Open Access scholarly Resources (ROAD)

Authors are responsible from the copyrights of the figures and the contents of the manuscripts, accuracy of the references, quotations and proposed ideas and the Publication Ethics (<https://dergipark.org.tr/en/pub/jscai/policy>)

Journal of Soft Computing and Artificial Intelligence (JSCAI) allows the author(s) to hold the copyright of own articles.

©
JSCAI
25 June 2022



This work is licensed under a Creative Commons Attribution 4.0 International License.

Table of Contents

■	Classification of Unwanted E-Mails (Spam) with Turkish Text by Different Algorithms in Weka Program Hüseyin ŞİMŞEK, Emrah AYDEMİR	Page: 1 - 10
■	Applying Machine Learning Prediction Methods to COVID-19 Data Faruk SERİN, Adnan KEÇE, Yiğit ALIŞAN	Page: 11 - 21
■	Investigation of Albedo Factor Parameters in Some Selected Sn Compounds Ahmet TURŞUCU	Page: 22 - 27
■	Hybrid experimental investigation of MR damper controlled tuned mass damper used for structures under earthquakes Hüseyin AGGÜMÜŞ, Rahmi GÜÇLÜ	Page: 28 - 33
■	Time Series Cleaning Methods for Hospital Emergency Admissions Yiğit ALIŞAN, Olcay TOSUN	Page: 34 - 40



Journal of Soft Computing and Artificial Intelligence

Journal homepage: <https://dergipark.org.tr/en/pub/jscai>

International
Open Access 

Volume 03
Issue 01

June, 2022

Research Article

Classification of Unwanted Emails (Spam) with Turkish Text by Different Algorithms in the Weka Program

Hüseyin Şimşek¹ , Emrah Aydemir^{2*} 

^{1,2} Management Information Systems, Institute of Business, Sakarya University, 54000, Sakarya, Turkey

ARTICLE INFO

Article history:

Received April 17, 2022

Revised April 25, 2022

Accepted May 1, 2022

Keywords:

E-mail

Classification Algorithm

Spam email

Weka

ABSTRACT

Recently, with the widespread use of the internet, electronic communication tools have also been widely used. One of these tools is emails. Emails are easy to use and provide the opportunity to reach thousands of people at the same time. This advantage causes some bad uses. Email users are faced with dozens of unsolicited emails (spam) against their will. In this study, 1017 mails collected from about 20 different Gmail and Hotmail accounts were classified as spam or regular email using the algorithms in the Weka program, and the success of the algorithms was compared. In the study, 45 different algorithms were tested. The highest classification success was obtained with the Naive Bayes Multinomial and Naive Bayes Multinomial Updateable algorithms with 94.7886% correct classification. Among other classifier algorithms, Random Forest algorithm 93.6087%, Multi-Class Classifier and SGD 92.4287%, SMO 91.7404%, Random Committee 91.0521%, Naive Bayes and Naive Bayes Updateable 90.3638% classification success.

1. Introduction

One of the basic needs of people is communication. Communication is the sharing of feelings, thoughts, ideas, and information between people. Today, new communication tools have emerged with the development of knowledge and internet technologies. One of them is emails that provide electronic communication and communication. Email is the adaptation of classical mailboxes to the electronic environment. The electronic mail system is inspired by correspondence, one of the communication tools used in the past and today and is reflected in the electronic environment with the development of today's internet technology [1]. An email address can be personal or corporate. Email service providers such as Gmail, Hotmail, Yahoo are available. An email address is created in the format "nickname@domainname" [2]. Text, audio, visual,

video, file, etc. contents can be shared easily with emails. They are easy to use and meager cost. In addition, it is a great convenience that content can be transmitted to thousands of different people or ¹ institutions simultaneously.

The ease of use of emails and the ability to reach thousands of people simultaneously has brought some disadvantages. At the top of these disadvantages are unsolicited (spam) messages. The abuse of electronic messaging systems to send random, unsolicited emails is called spam [3]. Thanks to the cost and speed advantage, emails are used for purposes such as advertising, promotion, marketing, creating public opinion, sharing inappropriate content, and obtaining personal information by sending malicious software, and dozens of spam emails fall into their mailboxes every

* Corresponding author

e-mail: emrahaydemir@sakarya.edu.tr

DOI: 10.55195/jscai.1104694

day. According to the results of a study, approximately 269 billion emails were sent and received worldwide in 2017, 281 billion emails in 2018, and 293 billion emails in 2019 [4]. While this causes a waste of time and effort for users, it also causes unnecessary occupation of network traffic. In addition, from the point of view of enterprises, it is seen that it causes enormous financial losses.

Many different methods and techniques are used to filter unwanted emails, and successful results are obtained. Despite this, they continue to use email systems by developing new strategies to overcome the filters applied to spam email users. For this reason, it is essential to carry out recent studies in this field, develop different methods and techniques, create different data sets, and support the analyses.

This study aims to contribute to the spam filtering studies and the literature by identifying the most successful algorithms in the Bayes, Trees, Meta, Lazy, Functions, Misc., Rules classifiers in the WEKA program by using a Turkish data set collected from different email addresses.

2. Mail System and Spam Mail

Electronic mail (email) is the name given to an electronic message, usually in the form of a simple text message, that a user writes on a computer system and transmits to another user who can read it over a computer network [5]. Email messages consist of a header and a body. The title contains the sender (From), recipient user's ID (To), subject header (Subject), date (Date), received (Received), and content number (Message-ID). There is the content of the message in the body part and the part where attachments (Attachment) will be made [2]. Simple Mail Transport Protocol-SMTP protocol is used for the transmission of emails.

Spam emails are messages sent in bulk by people or bot accounts that are not known. These can also be defined as messages sent to the accounts against the will of the person. Unwanted emails are used for purposes such as advertising, promotion, and propaganda. When we open email addresses, we come across dozens of advertising messages every day, and most of them come from addresses we do not know. In addition, some spam messages can send viruses to capture our personal information and bank account information. They can get our information by copying trusted web addresses and making us trust

them. Another reason why we are faced with spam messages today is due to the email trade. Email addresses belonging to thousands of people are marketed to different businesses, and they cause us to receive spam messages from companies we do not know. While businesses are always looking for ways to communicate with their customers more accessible, cheaper, or faster, the internet offers all three [6]. In this case, the marketing of email addresses is one of the reasons for the increase in the number of spam emails.

When examining spam messages, we can list some of their features as follows [2]. The same content is sent to multiple recipients.

- They are sent for promotional purposes.
- Often their content is misleading.
- They may talk about religious beliefs or human feelings and may want the email to be forwarded to many people.
- Address information such as sender, who is not in a proper format, and letter errors are standard as random fakes are usually produced.
- Email message header information is destroyed, making it difficult to trace back.
- Recipients do not have a valid or functional return address to indicate that they do not wish to receive email from this distribution.
- In general, their content is up to date

Today, many different methods are used, and new methods are being developed to filter unsolicited (spam) emails. Some of these methods are Word filtering, Rule-Based Filtering, Blacklists, DNS MX Record Lookup, Reverse DNS Lookup, So Reverse DNS Lookup Honeypots (Honeypots), Hashing Systems, Antivirus Scanning, Fingerprinting (fingerprint), Challenge-Response (challenge) systems and Bayesian filters [5].

3. Aim and Contribution

Today, although technological developments bring great convenience to our lives, they also bring some negativities along with these conveniences. The email has entered our lives with the development of internet technology and has brought a different dimension to communication. Many data such as interpersonal information, documents, pictures, and audio files can be shared quickly and inexpensively via emails. Since emails are a fast and low-cost communication tool, we encounter unsolicited emails, and we are faced with situations such as

endangering people's time, workforce, or personal information. Unnecessary occupancy of network traffic is another problem.

To avoid these problems, it is of great importance to develop new methods, test existing methods with different data sets, and determine successful strategies. This study it is aimed to test different classification algorithms in the Weka program and to choose the most successful classification method by using a data set with Turkish content that has not been used before. In addition, it is thought that it is essential that the more different data are used, the more successful the fight against spam will be.

4. Literature Review

When the literature on filtering and classification of spam is examined, it is seen that methods such as Spam classification with Machine Learning and Word Set technique, Phishing email detection with Deep Learning Models, text mining applications, Decision trees, Bayesian Classifiers, artificial immune system, and spam filtering are examined. In addition, a new approach based on Binary Patterns, filtering methods such as Word2Vec, Support Vector Machines are used. It has been observed that generally successful results have been obtained in the studies carried out.

In the study called Filtering Spam Emails Using the Bayesian Method in 2006, 2387 emails with Turkish content were used. Two different models were tested with the Bayesian method, and it was seen that the first model was classified as spam emails at a rate of 81%, 92%, and 84%, and 93.2%, respectively [5]. In the study conducted within the scope of SMTP Protocol and Spam Mail Problem, DNSBL technique was used, and it was seen that many mails could escape from DNSBL [3]. Tekeli and Aşhyan analyzed a data set in the UCI machine learning repository in the Weka program in their studies on the detection of spam emails with the multi-layered Perceptron, KNN, and C4.5 Methods and obtained a 92.8% successful classification with the C4.5 algorithms [7].

Cahide Ünal and İsmail Şahin designed a rule-based expert system for the detection of unsolicited emails and aimed to detect spam emails over content and IP addresses. In the study, a data set consisting of 4601 emails obtained from the Hewlett-Packard laboratory was used, and a total of 57 features were extracted from this data set. The developed Expert system examines the emails according to these 57 features and gives feedback to the user about whether

the email is a slur [8]. Nazlı Nazlı tested the Word2Vec vector and SVM(Poly) algorithm on a dataset of 300 emails in her Machine Learning-Based Spam Filtering Techniques study and achieved 98.33% successful results [9].

A new spam filtering approach, using binary patterns obtained by comparing the UTF-8 values of characters with each other by Kaya and Özdemir, who tried to detect spam with a new system based on scrolling binary patterns, shifted one-dimensional local binary patterns has been suggested. A proposed C1W-LBS method is a statistical approach based on low-level information obtained because of comparisons of each value on the signal with its neighbors. A benchmark (spamassian) and a dataset created by us were used to test our method. According to the results obtained, it has been seen that the proposed method is a successful method for feature extraction from text-based emails. 92.34% success was achieved in the filtering performed using the Weka Program [10].

Çıtlak, Dođru, and Dörterler In a Spam Detection System Study with Short Links, it has been determined that websites marked as spam in the Google Safe Browsing database can hide by using short link services. In the study, temporary link addresses were first listed, and then software was developed that converts quick link addresses to long web addresses. In the software made, these addresses are automatically checked whether they are spam or not spam in the Google Safe Browsing data set [11].

In the study titled an Analysis of Various Algorithms for Text Spam Classification and Clustering Using RapidMiner and Weka, NB, SVM, KNN algorithms were tested using Weka and RapidMiner programs. In the study, in the tests made using 5572 messages in the UCI Machine Learning database, the NB algorithm in the Weka Program was 94.56%, the SVM algorithm 98.21, the KNN algorithm 94.80% accurate classification success, while the RapidMiner program made the NB algorithm 84.79, SVM algorithm 96.64 and KNN algorithm 94. It was observed that 74 successful classifications were made. The study shows that the Weka Program makes better predictions than RapidMiner [12].

In 2017, the algorithm's success was tested using the Naive Bayes algorithm using two different data sets in Malaysia. In the study, 9324 Spam datasets collected from various email addresses and SPAMBASE dataset consisting of 4601 emails taken from UCi machine learning database were used. Five hundred features were extracted for Spam dataset, and 58 features were extracted for Spabase dataset.

In the results of the analysis, it was seen that the Naive Bayes algorithm made 91.13% correct predictions for the Spam dataset and 82.54% for the Spambase dataset, and it was seen that the collection of the data set from many different sources increased the percentage of correct predictions [13].

Eryılmaz and Kılıç examined the methods used for the detection of spam emails, basically examined artificial intelligence-based and non-artificial intelligence-based filtering techniques, and that non-artificial intelligence-based filtering methods (blacklist, whitelist, gray list, content review, etc.) can be passed, they stated that they have negative sides such as constant updating. Previously, non-AI-based methods would have been effective. But spam detection has improved as a result of the increase in machine learning algorithms. Thus, artificial intelligence-based systems have become more used [4].

Aman Kumar used 4601 email datasets from the UCI machine learning database to compare algorithms for spam filtering. 1813 pieces of data constitute spam emails. The data has a total of 58 features, 57 continuous and one nominal. As a result of the test, it was seen that the J48 Algorithm was the highest correct classification with 92.7624%, and the other algorithms were followed by CART with 92.632%, ADTree with 90.915%, and ID3 with 89.111%, respectively. According to the test results, it has been determined that the J48 algorithm is more successful than CART, ADTree, and ID3 in spam classification [14]. Again, in a similar study, Naive Bayes, Bayesnet, J48, and LAZY-IBK algorithms were compared, and it was seen that the J48 Algorithm was more successful than other algorithms, with a success rate of 85.06% [15].

While many different algorithms are being developed for spam blocking, spammers continue to send spam with new solutions. One of the methods used for spam is to send the message by embedding the content in the picture. When the content is embedded in the image, it does not get caught in the

content-based spam filters, and the users continue to receive unwanted messages. In his study, which used 180 data sets, 60 from Google, 60 from Flickr, and 60 from spam, for the detection of spam messages containing images during the daytime, he extracted the histograms of the images and classified the spammy images at a rate of 81%. When the histogram of spammy images is examined, it has been determined that they usually contain few colors, and the value of '0' is relatively high for colors that are not used in the histogram [16].

Looking at the studies, it is clearly seen that artificial intelligence-based machine learning algorithms give successful results in classifying spam emails. Despite this, it is seen that spammers continue to send unsolicited emails with new solutions. In addition, it is seen that the studies are generally studied on English data sets. For this reason, the creation of new data sets and the use of data sets in different languages are important in combating spam.

5. Material and Method

5.1 Data Collection

A total of 1017 emails were collected to be tested in this study. While 502 of the emails were regular mails, 517 of them were spam emails. The data set does not consist of any readymade data set but consists of emails sent to and from 20 different email addresses. Only the title and content parts of the emails with Turkish content were included in the data set, and each email was recorded separately in a text file. For regular mails, the names are norm1.txt, norm2.txt, norm502. spam1, spam2,, spam517 filenames are saved as spam. The dataset here has been uploaded publicly to its address so that it can be used by other researchers (<https://www.kaggle.com/datasets/emrahaydemr/turkish-mail-dataset-normalspam>).

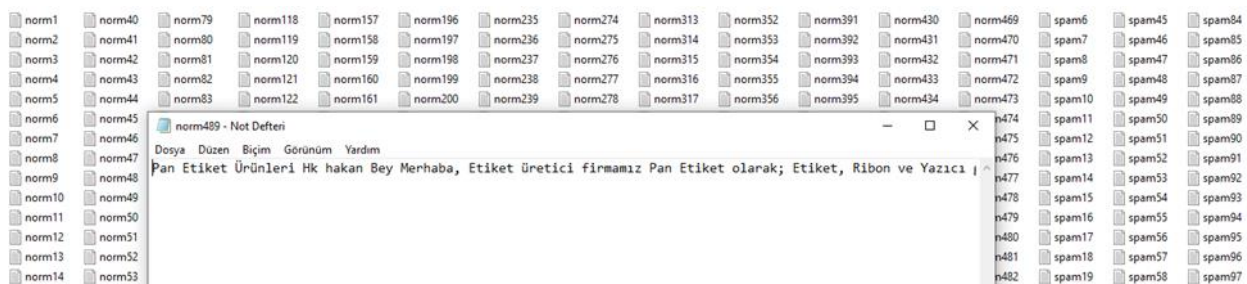


Figure 1 Created Dataset Pool

5.2 Data Analysis

During the analysis of the data, some operations were carried out in order to process the data in the Weka program. At the beginning of these processes, the data set collected in separate txt files was converted into a single norm_spam.txt file using the

Python programming language, and punctuation marks, memorable characters, and numbers in the text were cleaned from the text. The created text file has been converted into a format that the Weka program can analyze as an arff file.

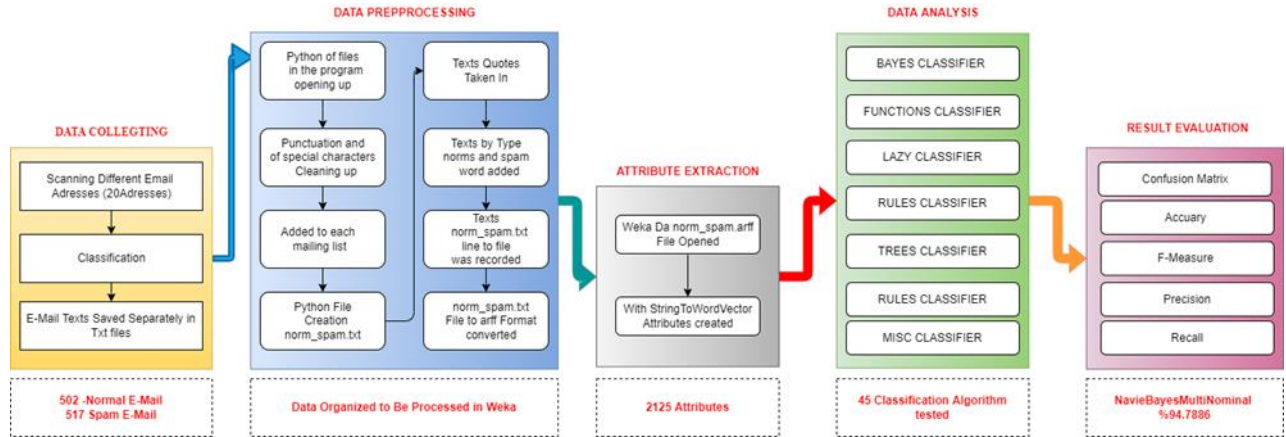


Figure 2 Email Analysis Workflow Chart

In the continuation of these processes, the norm_spam.arff file we created first was opened in the Explorer window of the Weka program with the

OpenFile tab, and the attributes of the data were extracted with the StringToWordVector filter from the filter tab. A total of 2125 word vectors belonging to the data set were extracted as features.

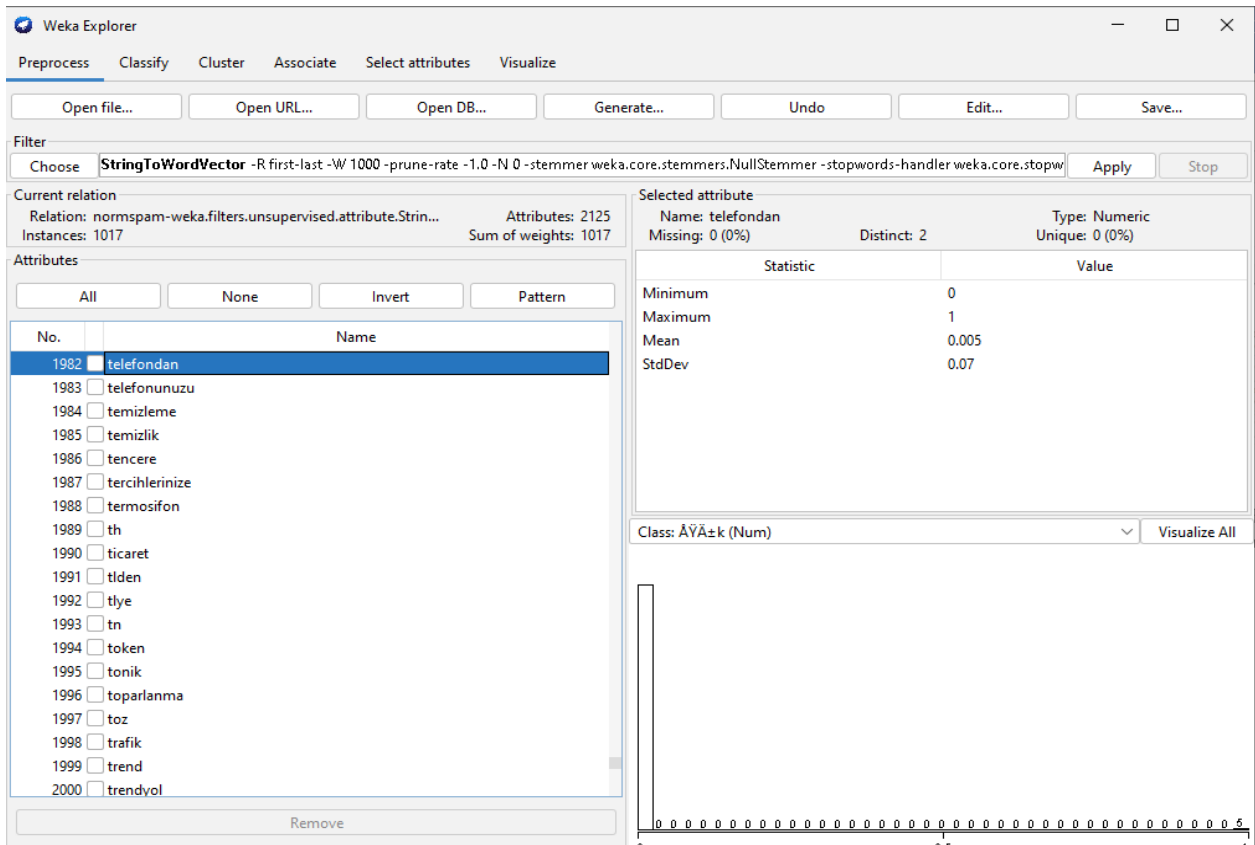


Figure 3 Attribute Extraction Screen

5.2.1 Weka Program

Aydemir introduced the Weka Program in the Artificial Intelligence Book with Weka as follows. Weka is a software developed at the University of Waikato in New Zealand and is licensed under the GNU and GPL. That is, it is software that is available to the public, free and open source. It is named after the initials of the phrase "Waikato Environment for Knowledge Analysis". The name of this Java-based program is also the name of a flightless and endangered bird found in the islands of New Zealand. This software, which is used for data mining and machine learning applications, contains almost all frequently used algorithms. It has been produced in order to be able to try on datasets quickly by using the existing methods themselves. In addition to this feature, it allows the analysis of the results. Through the program, basic data mining operations such as classification, clustering, and association can be performed in general. It can run on all systems, especially Linux, Windows, and Macintosh [17].

5.2.2 String to Word Vector

Word vectors simply focus on the relationships between words. Semantic analyzes are made based on the relationships of these words [18]. StringToWordVector Generates a numeric attribute showing the frequencies of the words in the String data type [17]. For natural language processing algorithms to understand text, texts must be represented as numbers. There are standard methods for this. The StringToWordVector filter used in the Weka program also converts texts into numerical vectors with techniques such as TF-IDF (Term Frequency — Inverse Document Frequency) and n-gram. With the help of this filter, the number of terms in the text or data set is counted and their frequency is revealed. For example, if a word occurs 10 times in the text and the dataset contains 1000 texts, then the value of $10/1000$, which is 0.01, is created for this word. In addition, the reverse document frequency determines how vital the searched word is. In other words, if the number of data in which the number of words searched is 10, and it is included in a total of 3 data, then the $\log(10/3)$ value of 0.52 is obtained. With this filter in Weka, word count, rooting, converting to lowercase, etc. It also provides features.

5.2.3 Algorithms Used in Data Analysis

For the analysis of the data, 45 different algorithms in the Classify window of the Weka program were tested, and the results were given in the findings section. The information on the classification algorithms used before proceeding to the findings is as follows.

- *Bayes Classifier:* Bayesian classifier is a classification algorithm that is used a lot in machine learning studies because it is fast and has a high success rate. Bayesian classifier is based on Bayes' theorem introduced by Thomas Bayes in 1763 [19]. This theorem, which deals with the events to be classified independently, predicts which class the data belong to [19]. Bayesian filters are one of the widely used methods of spam classification. To determine the probability that an email is a spam, filters use Bayesian analysis to compare the frequency of words or phrases in the email in previous (regular and spam) emails of the relevant user [5].

- *Naive Bayes Algorithm:* The logical foundations of the Naive Bayes algorithm are based on the approaches introduced by Thomas Bayes in the 18th century [20]. It is possible to have an idea about the direction of their content by analyzing the numerous news in the media or social media through text mining. Naive Bayes algorithm is one of the algorithms that can be used for this purpose [20].

- *Decision Trees:* Decision Trees in data mining are one of the most preferred methods because they are cheap to create, can be easily integrated with data systems, are safe, easy to interpret, and have a high comprehensibility [21]. When we look at the structure of decision trees, they consist of roots, branches, and leaves [21]. It resembles a tree with its structure. Decision trees that start with the root node divide many datasets into small groups and branches as they go down [21]. In decision trees, the first node is called the root node, the other nodes are called the leaf node, and the last is the decision node [22].

- *Lazy (Lazy Algorithms):* Lazy classifiers store the training samples and do no real work until it is time to classify [23]. The simplest lazy learning algorithm is the k-nearest neighbor classifier called IBk[17].

- *Meta Heuristics Algorithms:* The high-level heuristic approach includes methods that perform a probabilistic but conscious search in the solution space. These methods produce new solutions based on the solution set created at each step. Thus, by doing searches at the points close to the most suitable

one in the search space, it is tried to reach the most appropriate solution by getting rid of the local best point selection [24]. Rules Algorithms: The rule inference system (RULES) family is an inductive learning family that includes several overlay algorithms. This family is used to construct a predictive model based on the given observation. It works based on the concept of dividing and conquers to create rules and knowledge pools directly from a specific training set [25].

5.2.4 Success Criteria

The table where we can interpret the classification successes of the algorithms, we used in the research in an understandable way is the confusion matrix. We can compare the actual values with the estimated values with the confusion matrix. In the data set we used, we divided regular emails into two separate classes as norm spam and spam. The confusion matrix will allow us to interpret the results of the algorithm in 4 classes.

Table 1 Confusion Matrix

		Prediction	
		Regular	Spam
Real	Regular	TP	TN
	Spam	FN	FP

- **True Positive (TP):** Indicates the number of correctly classified emails that are actually regular emails.
- **True Negative (TN):** The number of emails classified as spam even though it is actually regular email.
- **False Negative (FN):** Number of spam emails classified as regular email even though they are actually spam.
- **False Positive (FP):** Shows the number of emails that are actually spam emails and are correctly classified as spam by the algorithm used.

6. Findings

After the dataset's attributes we used were extracted with StringToWordVector in the Weka Program Preprocess window, 45 algorithms were tested with the generally accepted 10-fold cross-validation (Cross-Validation Folds 10) method in classification processes in the Classify window, and the findings are given in the tables below.

Table 2 Findings

Classifier	Algorithm	Confusion Matrix	Accuracy	Precision	TP rate	FP rate	F-Measure	Recall	Class
BAYES	Bayes Net	a b <-- classif: 437 65 a = norm 72 443 b = spam	86,529	0,859 0,872	0,871 0,865	0,140 0,129	0,864 0,866	0,871 0,860	Norm Spam
	Naive Bayes	a b <-- classif 465 37 a = norm 61 454 b = spam	90,3638	0,884 0,925	0,926 0,882	0,118 0,074	0,905 0,903	0,926 0,882	Norm Spam
	Naive Bayes Multinomial	a b <-- classif 477 25 a = norm 28 487 b = spam	94,7886	0,945 0,951	0,950 0,946	0,054 0,050	0,947 0,948	0,950 0,946	Norm Spam
	Naive Bayes Multinomial Text	a b <-- classif 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	Naive Bayes Multinomial Updateable	a b <-- classif 477 25 a = norm 28 487 b = spam	94,7886	0,945 0,951	0,950 0,946	0,054 0,050	0,947 0,948	0,950 0,948	Norm Spam
	Naive Bayes Updateable	a b <-- classif 465 37 a = norm 61 454 b = spam	90,3638	0,884 0,925	0,926 0,882	0,118 0,074	0,905 0,903	0,926 0,882	Norm Spam
FUNCTIONS	Logistic	a b <-- classif: 427 75 a = norm 65 450 b = spam	86,234	0,868 0,857	0,851 0,874	0,126 0,149	0,859 0,865	0,851 0,874	Norm Spam
	Simple Logistic	a b <-- classif: 432 70 a = norm 42 473 b = spam	88,9872	0,911 0,871	0,861 0,918	0,82 0,139	0,885 0,894	0,861 0,918	Norm Spam
	SMO	a b <-- classif 456 46 a = norm 38 477 b = spam	91,7404	0,923 0,912	0,908 0,926	0,074 0,092	0,916 0,919	0,908 0,926	Norm Spam
	Voted Perceptron	a b <-- classif 427 75 a = norm 61 454 b = spam	86,6273	0,875 0,858	0,851 0,882	0,118 0,149	0,863 0,870	0,851 0,882	Norm Spam
	SGD Text	a b <-- classif 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 1,000	0,000 1,000	Norm Spam
	SGD	a b <-- classif: 460 42 a = norm 35 480 b = spam	92,4287	0,929 0,920	0,916 0,932	0,068 0,084	0,923 0,926	0,916 0,92	Norm Spam
LAZY	IBk	a b <-- classif: 373 129 a = norm 109 406 b = spam	76,5978	0,774 0,759	0,743 0,788	0,212 0,257	0,758 0,773	0,743 0,773	Norm Spam
	KStar		80,0393	0,836	0,741	0,142	0,786	0,741	Norm

		a b <-- classif 372 130 a = norm 73 442 b = spam		0,773	0,858	0,259	0,813	0,858	Spam
	LWL	a b <-- classif: 496 6 a = norm 418 97 b = spam	58,088	0,543 0,942	0,988 0,188	0,812 0,012	0,701 0,314	0,988 0,188	Norm Spam
RULES	Decision Table	a b <-- classif 428 74 a = norm 171 344 b = spam	75,9095	0,715 0,823	0,853 0,668	0,332 0,142	0,777 0,737	0,853 0,668	Norm Spam
	JRip	a b <-- classif: 326 176 a = norm 73 442 b = spam	75,5162	0,817 0,715	0,649 0,858	0,142 0,351	0,724 0,780	0,649 0,858	Norm Spam
	OneR	a b <-- classif: 454 48 a = norm 396 119 b = spam	56,3422	0,534 0,713	0,904 0,231	0,769 0,096	0,672 0,349	0,904 0,231	Norm Spam
	PART	a b <-- classif: 392 110 a = norm 76 439 b = spam	81,7109	0,838 0,800	0,781 0,852	0,148 0,219	0,808 0,825	0,781 0,852	Norm Spam
	ZeroR	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
			a b <-- classif: 496 6 a = norm 425 90 b = spam	57,6205	0,539 0,938	0,988 0,175	0,825 0,012	0,697 0,295	0,988 0,175
TREES	Decision Stump	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	Hoeffding Tree	a b <-- classif: 381 121 a = norm 81 434 b = spam	80,1377	0,825 0,782	0,759 0,843	0,157 0,241	0,790 0,811	0,759 0,843	Norm Spam
	LMT	a b <-- classif: 433 69 a = norm 37 478 b = spam	89,5772	0,921 0,874	0,863 0,928	0,072 0,137	0,891 0,900	0,863 0,928	Norm Spam
	Random Forest	a b <-- classif: 466 36 a = norm 29 486 b = spam	93,6087	0,941 0,931	0,928 0,944	0,056 0,072	0,935 0,937	0,928 0,944	Norm Spam
	Random Tree	a b <-- classif: 407 95 a = norm 85 430 b = spam	82,3009	0,827 0,819	0,811 0,835	0,165 0,189	0,819 0,827	0,811 0,835	Norm Spam
	REP Tree	a b <-- classif: 376 126 a = norm 64 451 b = spam	81,3176	0,855 0,782	0,749 0,876	0,124 0,251	0,798 0,826	0,749 0,876	Norm Spam
			a b <-- classif: 234 268 a = norm 65 450 b = spam	67,2566	0,783 0,627	0,466 0,874	0,126 0,534	0,584 0,730	0,466 0,874
META	AdaBoostM1	a b <-- classif: 381 121 a = norm 131 384 b = spam	75,2212	0,744 0,760	0,759 0,746	0,254 0,241	0,751 0,753	0,759 0,746	Norm Spam
	Attribute Selected Classifier	a b <-- classif: 401 101 a = norm 47 468 b = spam	85,4474	0,895 0,822	0,799 0,909	0,091 0,201	0,844 0,863	0,799 0,909	Norm Spam
	Bagging	a b <-- classif: 361 141 a = norm 87 428 b = spam	77,5811	0,806 0,752	0,719 0,831	0,169 0,281	0,760 0,790	0,719 0,831	Norm Spam
	Classification Via Regression	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	CV Parameter Selection	a b <-- classif: 393 109 a = norm 54 461 b = spam	83,9725	0,879 0,809	0,783 0,895	0,105 0,217	0,828 0,850	0,783 0,895	Norm Spam
	Filtered Classifier	a b <-- classif: 351 151 a = norm 88 427 b = spam	76,4995	0,800 0,739	0,699 0,829	0,171 0,301	0,746 0,781	0,699 0,829	Norm Spam
	Iterative Classifier Optimizer	a b <-- classif: 351 151 a = norm 88 427 b = spam	76,4995	0,800 0,739	0,699 0,829	0,171 0,301	0,746 0,781	0,699 0,829	Norm Spam
	Logit Boost	a b <-- classif: 427 75 a = norm 65 450 b = spam	86,234	0,868 0,857	0,851 0,874	0,126 0,149	0,859 0,865	0,851 0,874	Norm Spam
	Multi Class Classifier	a b <-- classif: 460 42 a = norm 35 480 b = spam	92,4287	0,929 0,920	0,916 0,932	0,068 0,084	0,923 0,926	0,916 0,932	Norm Spam
	Multi Class Classifier Updateable	a b <-- classif: 467 35 a = norm 56 459 b = spam	91,0521	0,893 0,929	0,930 0,891	0,109 0,070	0,911 0,910	0,930 0,891	Norm Spam
	Random Committee	a b <-- classif: 281 221 a = norm 177 338 b = spam	60,8653	0,614 0,605	0,560 0,656	0,344 0,440	0,585 0,629	0,560 0,656	Norm Spam
	Randomizable Filtered Classifier	a b <-- classif: 429 73 a = norm 39 476 b = spam	88,9872	0,917 0,867	0,855 0,924	0,076 0,145	0,885 0,895	0,855 0,924	Norm Spam
	Random Sub Space	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	Stacking	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	Vote	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	Weighted Instances Handler Wrapper	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
	Multi Schema	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam
MISC	Input Mapped Classifier	a b <-- classif: 0 502 a = norm 0 515 b = spam	50,6391	? 0,506	0,000 1,000	0,000 1,000	? 0,672	0,000 1,000	Norm Spam

7. Conclusion

This study classified regular email and unsolicited email (spam) using various algorithms. 502 regular emails and 517 spam emails collected from different email addresses were tested in the Weka program with 45 different classification algorithms. The highest classification success was obtained with the Naive Bayes Multinomial and Naive Bayes Multinomial Updateable algorithms with 94.7886% correct classification. Among other classifier algorithms, Random Forest algorithm 93.6087%. It is seen from the results that Multi Class Classifier and SGD 92.4287%, SMO 91.7404%, Random Committee 91.0521%, Naive Bayes Updateable 90.3638% classification success. In the study, it was seen that the data set without classification success was clustered in a single class (spam), while the Meta (Stacking, Vote, Weighted Instances Handler Wrapper, Multi Schema, CV Parameter Selection), Hoeffding Tree, Rules Zero R, SGD Text, Naive Bayes Multinomial Text algorithms 50.6391.

When the results are examined, successful results can be obtained by using Random Forest, SMO, Multi Class Classifier Updateable, and Random Committee Algorithms, where bayesian classifiers show higher success in classifying spam. When the studies on filtering spam emails are examined, it is a fact that although successful results have been obtained, spammers are constantly developing new methods. To continue the struggle with this reality, the creation of more Turkish data sets is of great importance for future studies.

References


- [1]. C. Özdemir, M. Ataş, Ve A. B. Özer, “Türkçe İstenmeyen Elektronik Postaların Yapay Bağışıklık Sistemi İle Sınıflandırılması, Signal Processing and Communications Applications Conference (SIU), 2013.
- [2]. C. Özdemir, “Yapay bağışıklık sistemi ile spam filtreleme”, Master’s Thesis, Fen Bilimleri Enstitüsü, 2013.
- [3]. M. E. Yüksel ve Ş. D. Odabaşı, “SMTP Protokolü ve Spam Mail Problemi”, Akad. Bilişim, 2010.
- [4]. E. E. Eryılmaz ve E. Kılıç, “İstenmeyen Epostaların Tespiti için Kullanılan Yöntemlerin İncelenmesi”, Dicle Üniversitesi Mühendis. Fakültesi Mühendis. Derg., c. 11, sy 3, ss. 977-987, 2020.
- [5]. C. Altunyaprak, “Bayes yöntemi kullanarak istenmeyen elektronik postaların filtrelenmesi”, PhD Thesis, Yüksek Lisans Tezi, Muğla Üniversitesi Fen Bilimleri Enstitüsü, 2006.
- [6]. Y. Gedik, “E-Posta Pazarlama: Teorik Bir Bakış”, Uluslar. Önetim Akad. Derg., c. 3, sy 2, ss. 476-490, 2020.
- [7]. K. Tekeli ve R. Aşlıyan, “Çok Katmanlı Algılayıcı, K-NN ve C4. 5 Metotlarıyla İstenmeyen E-postaların Tespiti”, Adnan Menderes Üniversitesi, 2016.
- [8]. Ü. Cahide ve İ. Şahin, “İstenmeyen Elektronik Postaların (SPAM) Filtrelenmesi için Bir Uzman Sistem Tasarımı ve Gerçekleştirilmesi”, Politek. Derg., c. 20, sy 2, ss. 267-274, 2017.
- [9]. N. Nazlı, "Analysis of machine learning-based spam filtering techniques", Master's Thesis, 2018.
- [10]. Y. Kaya ve C. Özdemir, “Spam filtrelemek için kaydırmalı ikili örüntüler tabanlı yeni bir yaklaşım”. XVIII. Akademik Bilişim Conference, 2016.
- [11]. O. Çitlak, İ. A. Doğru, ve M. Dörterler, “A spam detection system with short Link Analysis”. 10. Uluslararası Bilgi Güvenliği Ve Kriptoloji Konferansı, Ankara, Türkiye, 20 - 21 Ekim 2017.
- [12]. K. Zainal, N. F. Sulaiman, ve M. Z. Jali, "An analysis of various algorithms for text spam classification and clustering using RapidMiner and Weka", Int. J. Comput. Sci. Inf. Secur., c. 13, sy 3, s. 66, 2015.
- [13]. N. F. Rusland, N. Wahid, S. Kasim, ve H. Hafit, "Analysis of Naive Bayes algorithm for email spam filtering across multiple datasets", içinde IOP conference series: materials science and engineering, 2017, c. 226, sy 1, s. 012091.
- [14]. A. K. Sharma ve S. Sahni, "A comparative study of classification algorithms for spam email data analysis", Int. J. Comput. Sci. Eng., c. 3, sy 5, ss. 1890-1895, 2011.
- [15]. G. H. AL-Rawashdeh ve R. B. Mamat, "Comparison of four email classification algorithms using WEKA", Int. J. Comput. Sci. Inf. Secur. IJCSIS, c. 17, sy 2, ss. 42-54, 2019.
- [16]. H. C. Gündüz, “Spam 2.0, Tespit ve Engelleme Yöntemleri”, s. 6, 2007.
- [17]. E. Aydemir, Weka İle Yapay Zeka, 2. Baskı. Ankara: Seçkin, 2019.
- [18]. E. Koç, S. Çalışkan, S. A. Yazıcıoğlu, U. Demirci, ve Z. Kuş, “Yapay Sinir Ağları, Kelime Vektörleri ve Derin Öğrenme Uygulamaları”, 2018.
- [19]. G. AKSOY, “Ağırlıklı Bayes sınıflandırıcıda ağırlıkların optimizasyonu/Optimization of the weights of weighted Naive Bayesian classifier”, 2018.
- [20]. A. Suat, “KNN, Naive Bayes ve Karar Ağacı Makine Öğrenme Algoritmaları, Bu Algoritmaların Sosyal Bilimlerde Kullanım İmkânları”, 2020.
- [21]. Ş. Demirel ve S. G. Yakut, “Karar Ağacı Algoritmaları ve Çocuk İşçiliği Üzerine Bir

- Uygulama”, Sos. Bilim. Arařt. Derg., c. 8, sy 4, ss. 52-65, 2019.
- [22]. B. S. Kuzu ve S. G. Yakut, “Destek Vektör Makineleri Yardimiyla Imalat Sanayisinde Mali Başarisizlik Tahminlerinin Teknoloji Yoğunluđuna Göre İncelenmesi”, Osman. Korkut Ata Üniversitesi İktisadi Ve İdari Bilim. Fakültesi Derg., c. 4, sy 2, ss. 36-54, 2020.
- [23]. D. Akmaz ve M. S. Mamiş, “Bayes, Lazy, Trees, Rules Sınıfı Makine Öğrenme Algoritmaları”. 2nd International Mediterranean Science and Engineering Congress, 2017.
- [24]. Seckin, “MetaSezgisel Algoritmalar”, Bir Yazılımcının Günlüğü, 16 Mayıs 2017. <https://biryazilimciningunlugu.wordpress.com/2017/05/16/metasezgisel-algoritmalar/> (erişim 16 Nisan 2022).
- [25]. “Rules extraction system family”, Wikipedia. 27 Kasım 2019. Erişim: 16 Nisan 2022. [Çevrimiçi]. Erişim adresi: https://en.wikipedia.org/w/index.php?title=Rules_extraction_system_family&oldid=928196017



Journal of Soft Computing and Artificial Intelligence

Journal homepage: <https://dergipark.org.tr/en/pub/jscai>

International
Open Access 

Volume 03
Issue 01

June, 2022

Research Article

Applying Machine Learning Prediction Methods to COVID-19 Data

Adnan Kece¹, Yigit Alisan^{2*}, Faruk Serin³

¹Council of Forensic Medicine, Computing Forensic Department, İstanbul, Turkey

^{2*}Distance Education Application and Research Center Sinop University, 57000 Sinop, Turkey

³Mersin University, Faculty of Engineering, Department of Computer Engineering, Mersin, Turkey

ARTICLE INFO

Article history:

Received **April 25, 2022**

Revised **May 23, 2022**

Accepted **May 31, 2022**

Keywords:

Machine learning

SEIR

BFGS

LSTM

SVM

COVID-19

ABSTRACT

The Coronavirus (COVID-19) epidemic emerged in China and has caused many problems such as loss of life, and deterioration of the social and economic structure. Thus, understanding and predicting the course of the epidemic is very important. In this study, the SEIR model and machine learning methods LSTM and SVM were used to predict the values of Susceptible, Exposed, Infected, and Recovered for COVID-19. For this purpose, COVID-19 data from Egypt and South Korea provided by John Hopkins University were used. The results of the methods were compared by using MAPE. A total of 79% of MAPE were between 0 and 10. The comparisons show that although LSTM provided better results, the results of all three methods were successful in predicting the number of cases, the number of patients who died, and the peaks and dimensions of the epidemic.

1. Introduction

Epidemics are large-scale communicable diseases that greatly increase morbidity and mortality over a wide geographical area and cause significant economic, social, and political disruptions [1]. In December 2019, an epidemic of atypical pneumonia of a novel coronavirus (SARS-CoV-2) of zoonotic origin occurred in Wuhan, the capital of China's Hubei province [2]. On January 30, 2020, the World Health Organization (WHO) declared the outbreak a "Public Health Emergency of International Concern". On 12 February 2020, WHO named the disease caused by the novel coronavirus "Coronavirus Disease 2019" (COVID-19) [3]. On March 11, 2020, the epidemic was declared a pandemic by the World Health Organization [4]. COVID-19 pandemic is a global public health crisis that threatens people's health [5]. In the analysis of epidemics, there are

many epidemic modelling methods used to both predict the number of cases, the number of patients who died, and other parameters of the epidemic and to simulate the epidemic process. Among these models, the SIR model is one of the most preferred models. The SIR model has a compartment-type structure and was introduced to the literature by [6]. Compartment S represents Susceptible, compartment I Infected, and compartment R Recovered/Removed. Until today, many improvements have been made to the main model, such as adding new compartments and editing existing compartments, and new models have been created. In the literature, SIR [7], [8], SIRS [9], SEIR [10] – [12], and SIRD [13] are examples of use in this area. However, the SEIR model is more appropriate and comprehensive to describe information on the spread of the epidemic [14].

Time series are frequently preferred to produce helpful results for policymakers or managers in many

* Corresponding author

e-mail: alisan.yigit@gmail.com

DOI: 10.55195/jscai.1108528

different fields such as water resources [15], soil mapping [16], traffic [17][18], and tourist numbers [19]. In addition, the number of studies using machine learning methods in addition to traditional methods is increasing. Along with epidemic modelling methods, many machine learning methods are used in the analysis of epidemic diseases in epidemiology. Examples of usage in this field are ARIMA [20]– [22], LSTM [23], [24], SVM [25], [26] methods. Both epidemic modelling methods and machine learning methods can be used together in the analysis of epidemics. Hybrid studies are generally performed for comparison and confirmation of results. SEIR model and Regression techniques [27], SEIR model and DNN method [28], and SEIR model and LSTM method [29] are examples of hybrid studies in the literature.

This study consists of four stages and aims to predict some COVID-19 data. Firstly, a dataset was created using COVID-19 data from South Korea and Egypt due to their consistency in the number of cases and deaths. In the second section, SEIR model were optimized using various optimization methods. In the third section, epidemic parameters were predicted using Long Short-Term Memory (LSTM) and Support Vector Machine (SVM), which are machine learning methods frequently mentioned in the literature. In the last section, the data of the outputs obtained from the methods were evaluated using the MAPE metric.

1.1. Motivation and Background

The SEIR model interactively models four different conditions (Susceptible, Exposed, Infected, and Recovered/Removed) using historical data from an outbreak. In the compartmental models, firstly, the main parameters are given as input to the system. Then the output created by the first compartment is given as input to the next compartment. This interaction continues until the last compartment. The study, it is aimed to optimize the values of the parameters produced by the SEIR model in the transitions between the compartments by using machine learning methods and thus enhancing the overall success.

2. Material and Methods

The data set used in the study was obtained from the data updated daily by John Hopkins University via the GitHub platform. “Number of Confirmed Cases”, “Number of Patients Who Died”, “Number

of Recovered Patients”, “Date” and “Country Population” information were extracted from the data [30].

2.1. SEIR

Epidemic modelling methods simplify the mathematical modelling of infectious diseases. In these models, people are assigned to compartments according to their health status. People can advance between groups/compartments. The order of the labels usually shows the flow patterns between the compartments. Models attempt to estimate how far a disease has spread, the total number of infected, the duration of the epidemic, and various epidemiological parameters such as the reproduction coefficient. Such models can show how their response could affect the outcome of the epidemic.

The SEIR model simulates the time history of an epidemic. The model interactively models between four different conditions using historical data from an epidemic. These conditions are Susceptible (S), Exposed (E), Infected (I), and Recovered/Removed (R) [31]. The first compartment of the model, susceptible people, is equal to the entire population on the first day of the epidemic and has not yet encountered the disease. The second compartment consists of people who are exposed to the disease while being susceptible. In the third compartment, some people are infected. These consist of people with symptoms whose incubation period has ended after exposure to the disease. Finally, some are infected people who come out of the cycle because of death or recovery. The dynamic transfer of the population during the epidemic period is shown in Figure 1.



Figure 1. SEIR model

SEIR model parameters are calculated using the formulas in Equations. (1), (2), (3) and (4).

Equations

$$\frac{dS(t)}{dt} = -\beta I(t) \frac{S(t)}{N} \quad (1)$$

$$\frac{dE(t)}{dt} = \beta I(t) \frac{S(t)}{N} - \gamma E(t) \quad (2)$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t) \quad (3)$$

$$\frac{dR(t)}{dt} = \delta(t)I(t) \quad (4)$$

The formula for the number of susceptible in Eq. (1), the number of exposed in Eq. (2), the number of infected in Eq. (3), the number of removed in Eq. (4) at time t .

The terms in these equations are expressed as (S) parameter is target time histories of susceptible cases, (E) parameter is target time histories of exposed cases, (I) parameter is target time histories of infectious cases, (R) parameter is target time histories of removed/recovered cases, (β) infection rate, (γ) the inverse of the mean delay time and (δ) refers to the coefficients used depending on the mortality rate [31].

2.2. Curve Fitting Methods

In the predictions made on epidemic modelling methods, curve fitting methods can be used to keep the epidemic curves at a certain level or to optimize the curves. The curve fit is a data analysis method that attempts to construct a linear or nonlinear $f(x)$ model function. Some optimization algorithms such as L-BFGS-B are proposed to optimize the parameters of SIR and SEIR models and improve their prediction capabilities [32]. The curve fitting methods used in this study are explained as follows.

The L-BFGS-B algorithm defines constant and free variables at each step. The method only works by repeating the L-BFGS method on free variables to achieve high accuracy.

The Powell method is an algorithm proposed by Michael J. D. Powell to find the local minimum of a function. It is quite useful to use for minimizing functions of a large number of variables as it does not require storing any matrix with a conjugate gradient [33]. The conjugate gradient (CG) method commonly attributed to Magnus Hestenes and Eduard Stiefel is an algorithm used in mathematics for the numerical solution of certain systems of linear equations. COBYLA is a gradient-free optimization algorithm that can manage nonlinear inequality constraints. For this reason, Powell developed the COBYLA algorithm in 1994, which creates linear polynomial approximations to the objective and constraint functions by interpolating the vertices of the simplifications [34]. The SLSQP method is generally

used when the objective function and constraints are continuously differentiable. In mathematical problems using the SLSQP algorithm, each step solves two subproblems [35].

2.3. Support Vector Machine

The Support Vector Machine (SVM) was originally introduced by Vapnik et al and later expanded by another researcher [36], [37]. Vapnik stated that SVM method is based on VC (Vapnik–Chervonenkis) theory, and their experimental work demonstrated superior generalization performance over the neural network and statistical learning technique [38]. SVM is implemented in the form of classification and regression. Classification is performed with a linear or nonlinear separation surface in the input field of the dataset [39]. Regression, on the other hand, uses Vapnik's non-precision approach, creating a flexible channel with a minimum radius symmetrically around the predicted function, thus ignoring values outside the threshold as a result of error calculation [40]. SVM is basically divided into two according to whether the data set can be separated linearly or not. Therefore, in this section, it is discussed in two parts as linear and non-linear SVM.

2.3.1. Linear SVM

In classification with SVM, it is aimed to separate the samples belonging to two classes, which are generally shown with class labels as $\{-1, +1\}$, with the help of a decision function obtained from the training data [36], [38] for the training of SVM in a linearly separable two-class classification problem, assuming that the training data consisting of k samples $\{x_i, y_i\}$, $i = 1, \dots, k$. The inequalities of the optimum hyperplane are calculated as Equations (5) and (6).

$$\min \frac{1}{2} \|w\|^2 \quad (5)$$

$$y_i(w x_i + b) - 1 \geq 0 \text{ and } y_i \in \{-1, 1\} \quad (6)$$

2.3.2. Non-Linear SVM

Data may not always be linearly separated in problem-solving. In such cases, the problem arising from the fact that some of the training data remain on the other side of the optimal hyperplane is solved by

defining a positive dummy variable (ξ). The objective function is specified in Eq. (7) and the constraints of the function are specified in equations (8) and (9) [36].

$$\frac{1}{2} \|w\|^2 + C(\sum_{i=1}^t \xi_i)^k \quad k > 1 \quad (7)$$

$$y_i(wx_i + b) \geq 1 - \xi_i \quad i = 1 \dots t \quad (8)$$

$$\xi_i \geq 0 \quad i = 1 \dots t \quad (9)$$

2.4. Long Short-term memory

Long short-term memory (LSTM) was first processed by Hochreiter & Schmidhuber in 1997 and developed later by the Flex Gers team [41]. The LSTM layer consists of a series of repetitively linked blocks known as memory blocks. These blocks can be thought of as a different versions of memory chips in a digital computer. Each contains one or more repetitively linked memory cells and three multiplicative units (input, output, and forget gates) [42].

Memory cells store their entries for a period. Input gates control the flow of new values into the cell. Forget gates decide which values should be stored in memory. The output gates, on the other hand, control which values in memory should be used to calculate the output activation of the LSTM unit. The LSTM model is shown in Figure 2.

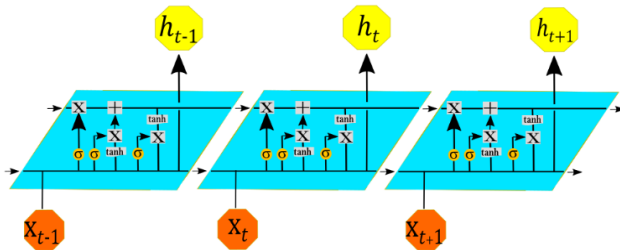


Figure 2. LSTM module

In Figure 2, each line carries a vector from the output of one node to the inputs of the others. Orange circles represent operations for points, such as vector addition, while blue boxes learn neural network layers. Merging lines indicate the merge, while a line stamp indicates that its content is copied, and the copies go to different locations.

In LSTM, the hidden state of memory cells is calculated by the formulas in Equations (10), (11), (12), (13) and (14).

Here i_t is the input gate, f_t is the forget gate, o_t is

the output gate, w is the weight on the corresponding gate, σ is the standard sigmoid value, x_t is the current timestamp input, c_t is the cell memory timestamp, and b is the deviation value [43].

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (10)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \quad (11)$$

$$c_t = f_t * c_{t-1} + i_t * g(W_{cx}x_t + W_{ch}h_{t-1} + W_{cc}c_{t-1} + b_c) \quad (12)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_{t-1} + b_o) \quad (13)$$

$$h_t = o_t * h(c_t) \quad (14)$$

2.5. Performance Measurement

Mean Absolute Percentage Error (MAPE) is one of the statistical performance measurement methods that shows how consistent the results obtained in applications are. The closer the result obtained in this method is to 0, the more successful it is. The MAPE formula in Eq. (15):

$$MAPE = \left(\frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \right) * 100 \quad (15)$$

3. Experimental Results

The initial parameters of the SEIR Model in this study were taken from studies published by WHO. R_0 , mean incubation time and infection time was taken as below to calculate pass rates between parameters [44].

Reproduction Number (R_0) = 3,9 (people)

Incubation Period = 2,9 (day)

Infection Period = 5,2 (day)

A total 70% of the data set was reserved for training and 30% for testing. "Number of Confirmed Cases", "Number of Patients Who Died", "Number of Recovered Patients" and "Country Population" information included in the data set were introduced to the model. R_0 and fatality rate were calculated on the SEIR model, along with the optimized number of cases and deaths. R_0 and mortality rates were used to find transition rates between epidemic parameters. Daily values of epidemic parameters were calculated using pass rates. Five different optimization algorithms were used for the optimization of the "Case" and "Death" numbers on the SEIR model. The

MAPE ratios related to the results are shown in Table 1.

Table 1. MAPE results of optimization algorithms

Method	Egypt		South Korea	
	Case	Fatality	Case	Fatality
L-BFGS-B	1.294	0.438	16.991	17.176
Powell	1.294	0.438	16.991	17.177
SLSQP	1.285	0.444	17.009	17.172
CG	1.294	0.438	16.991	17.177
COBYLA	4.822	3.868	18.187	17.698

As seen in Table 1, all methods produced acceptable results. The data of the L-BFGS-B method in calculating the transition ratios between parameters on the SEIR model are presented below. Calculated R0 and pass rates are shown in Table 2.

Table 2. Pass rates and R0 value of the parameters

Country/ Parameter	S>E	E>I	I>R	R0
Egypt	0.679	0.340	0.050	1.97
South Korea	0.651	0.344	0.016	1.89

The MAPE resulting from the prediction of the epidemic parameters of Egypt and South Korea are shown in Table 3 and Table 4. While the lowest MAPE value was obtained by the SVM to predict S and R, LSTM provided the lowest MAPE value for E and I.

Table 3. MAPE results of S, E, I and R for Egypt

Method/ Parameters	S	E	I	R
SEIR (L-BFGS-B)	2.570	6.100	5.270	6.550
SVM	0.002	12.500	5.880	0.430
LSTM	0.280	0.930	1.180	0.510

Table 4. MAPE results of S, E, I and R for South Korea

Method/ Parameters	S	E	I	R
SEIR (L-BFGS-B)	2.010	39.030	37.540	15.950
SVM	0.015	31.500	6.640	0.250
LSTM	0.110	0.860	0.450	1.540

In Figure 3, it is seen that the difference between the actual and predicted data widened after 150 days, although the predicted and actual data for the first 100 days were very close in predicting the number of susceptible people in Egypt with the SEIR. In Figure 3, it is also seen that the curves of the predicted and actual data of the people who were exposed, sick, and died in the epidemic in Egypt and South Korea are similar to each other.

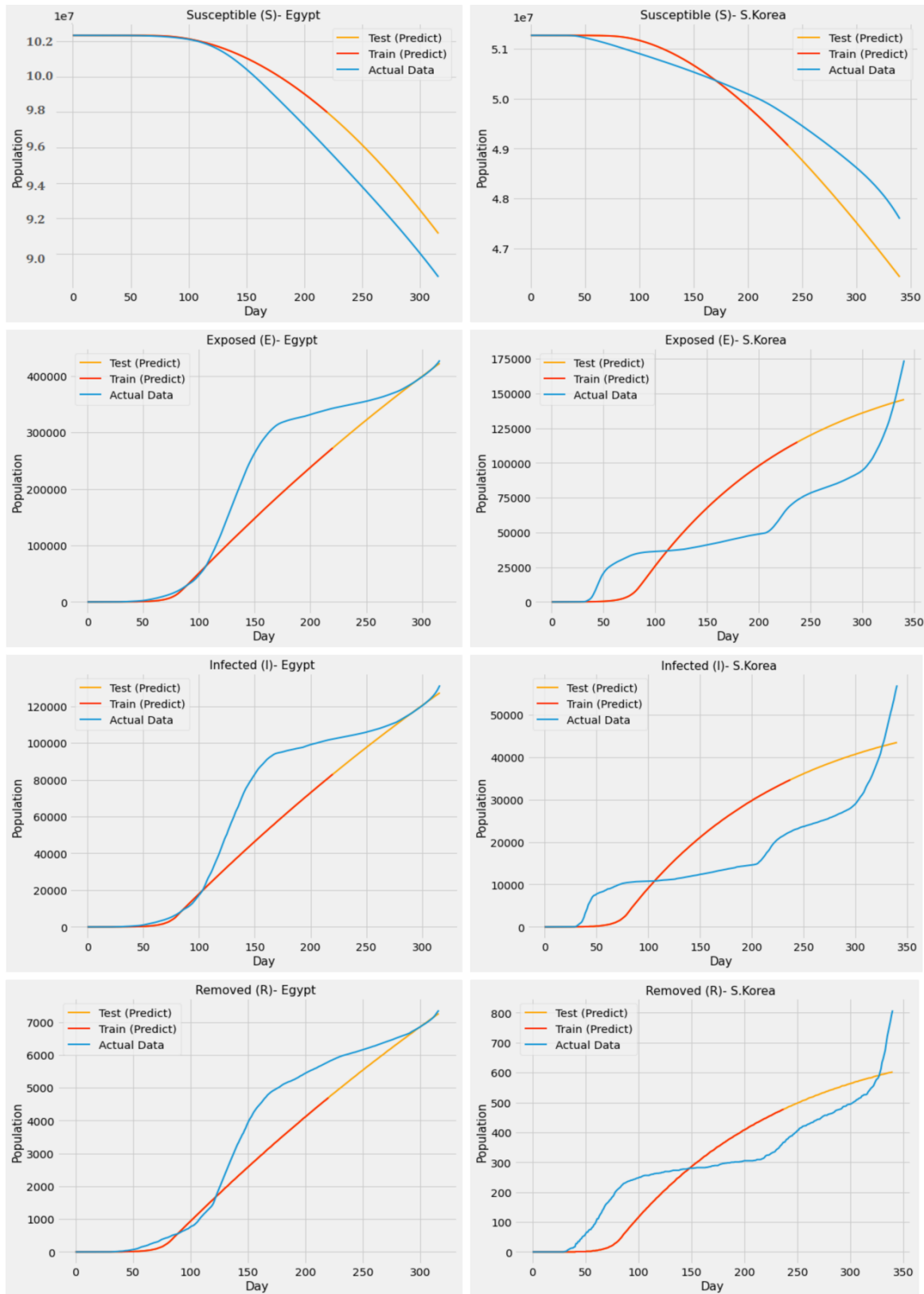


Figure 3. Prediction results of S, E, I and R in Egypt and South Korea using SEIR model

In the prediction of the COVID-19 epidemic parameters, LSTM and SVM (SVR) methods were used as machine learning methods. Susceptible, Exposed, Infected, and Removed persons in Egypt and South Korea were predicted by SVR and LSTM. A total 70% of the dataset was reserved for training

and 30% for testing. The prediction results of SVR and LSTM are shown in Figure 4 and Figure 5 respectively. As seen in the figures, the curves of the actual data and the predicted data are close to each other.

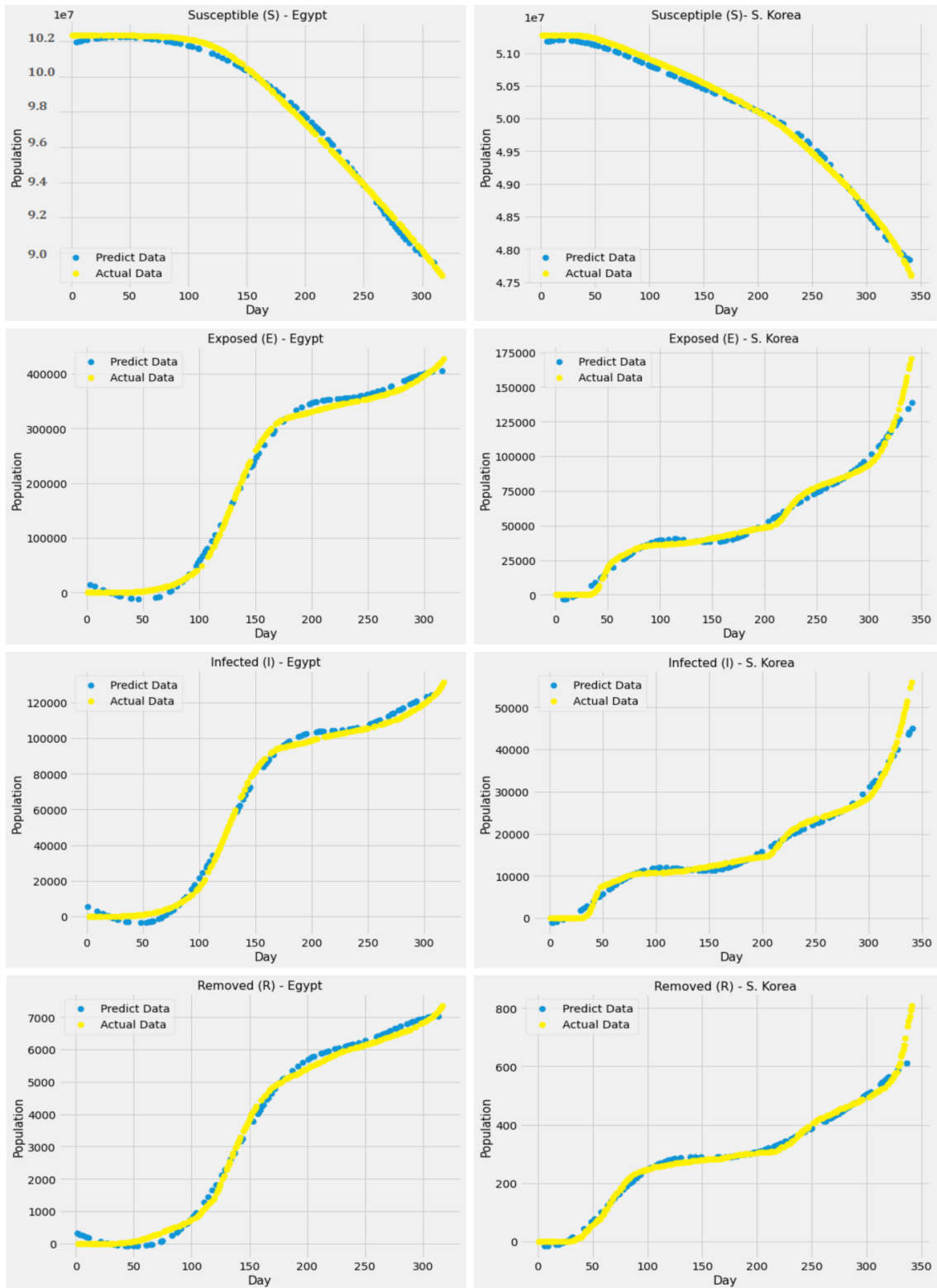


Figure 4. Prediction results of S, E, I and R parameters in Egypt and South Korea by SVM

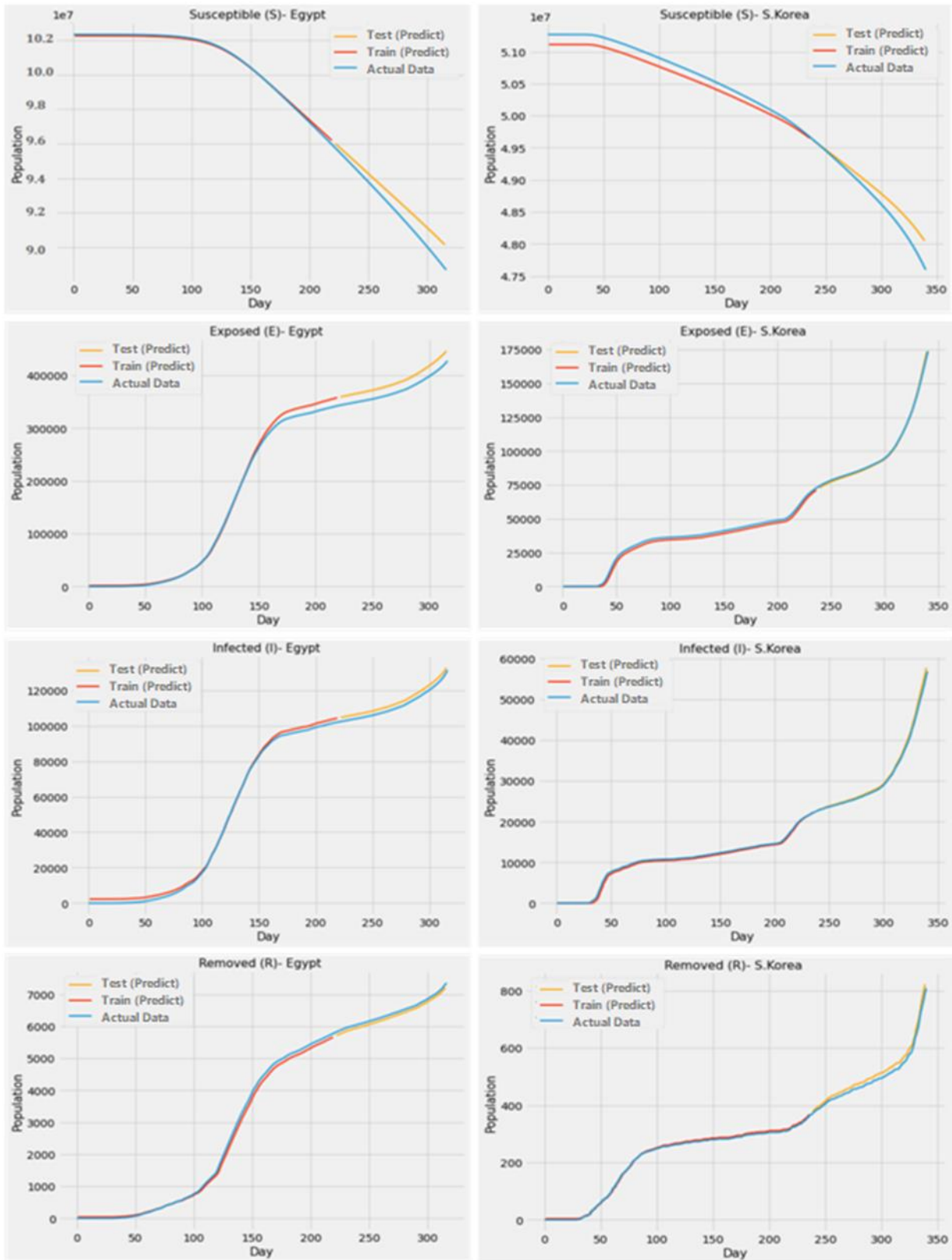


Figure 5. Prediction results of S, E, I and R parameters in Egypt and South Korea by LSTM

4. Discussion and Conclusion

Studies in the literature show that instead of estimating time series with a single method, estimating by combining multiple methods that can model different functional relationships in the data set gives more effective results. In addition, the success of the previously used machine learning methods compared to traditional methods also supports the results of the study [24], [26]. In the study, which was carried out by bringing an additional perspective to the existing studies, machine learning methods gave very successful results in estimating the parameters of epidemic diseases. Societies are cultural, social, economic, etc. It has different structures due to many factors. For this reason, the real-life consequences of epidemics differ according to the structure of society. However, compartment models used in the field of epidemiology accept these characteristics of societies as the same. It is aimed to provide memory features to the SEIR model, which has a memoryless structure, by using the LSTM method. Thus, instead of approaching every data in the same way, training the data according to the patterns it has will increase the accuracy and confidence rates of the results to be obtained.

4.1. Limitations and future research

The official numbers announced by each country may not reflect the truth, as countries may impose bans such as entry and import bans to countries with a high number of infected individuals to protect their own citizens in pandemic-level epidemics. In addition to the inherent limitations of the time series, this situation may cause erroneous data, and thus the results may be erroneous. To overcome such limitations in the study, the countries used were carefully researched to choose their data. Models derived from the SIR main model, such as SEIR, generally make some assumptions. Some of these assumptions are that during the spread of the disease, there is no birth, or death other than a disease, external migration, or immigration, and the society has a homogeneous structure. In future studies, using the obtained data by eliminating the possibilities such as misdiagnosis and different causes of death will increase the success rate of the model. In addition, it will show significant success in studies where all the

features that shape the basic structure of the society can be added to the model, considering the super-spreaders in the spread of the disease. It is thought that more reliable results will be produced because of adding a compartment that can provide memory properties to such models, providing the data of similar events in the past in the same population to the model, and training the model using these.

Information on the fund or financing

The authors have not received financial support for the research, authorship, or publication of this work.

Contributions of the Authors

Adnan Kece: Methodology, Data curation, Writing Original draft preparation, Visualization, Software.

Yigit Alisan: Reviewing, Editing, Visualization.

Faruk Serin: Conceptualization, Original draft preparation.

References

- [1] N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, and N. Wolfe, "Pandemics: Risks, Impacts, and Mitigation," in *Disease Control Priorities: Improving Health and Reducing Poverty*, 3rd ed., D. T. Jamison, H. Gelband, S. Horton, P. Jha, R. Laxminarayan, C. N. Mock, and R. Nugent, Eds. Washington (DC): The International Bank for Reconstruction and Development / The World Bank, 2017.
- [2] Q. Li et al., "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia," *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199–1207, Oct. 2020, doi: 10.1056/NEJMoa2001316.
- [3] <https://www.who.int/director-general/speeches/detail/who-director-general-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020>. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. (Access date: 10 July 2021)
- [4] CSSEGISandData · GitHub. (n.d.). Retrieved June 6, 2021, from <https://github.com/CSSEGISandData> (Access date: 10 July 2021)
- [5] N. Bernardini et al., "How lockdown measures, during COVID-19 pandemic, matter on psoriatic patient's perception: study on 600 patients on biologic therapy," *Journal of Infection and Public Health*, 2021, doi: 10.1016/j.jiph.2021.03.010.
- [6] W.O. Kermack, A.G. McKendrick, "A contribution to the mathematical theory of epidemics", *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 115 (772) (1927),

- pp. 700-721
- [7] S. Dil, N. Dil, and Z. H. Maken, "COVID-19 Trends and Forecast in the Eastern Mediterranean Region With a Particular Focus on Pakistan," *Cureus*, vol. 12, no. 6, May 2021, doi: 10.7759/cureus.8582.
- [8] L. Roques, E. Klein, J. Papax, A. Sar, and S. Soubeyrand, "Using early data to estimate the actual infection fatality ratio from COVID-19 in France (Running title: Infection fatality ratio from COVID-19)," *Biology*, vol. 9, no. 5, p. 97, Jun. 2020, doi: 10.3390/biology9050097.
- [9] C.-H. Li, C.-C. Tsai, and S.-Y. Yang, "Analysis of epidemic spreading of an SIRS model in complex heterogeneous networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 4, pp. 1042–1054, Jun. 2014, doi: 10.1016/j.cnsns.2013.08.033.
- [10] C. Reno et al., "Forecasting COVID-19-Associated Hospitalizations under Different Levels of Social Distancing in Lombardy and Emilia-Romagna, Northern Italy: Results from an Extended SEIR Compartmental Model," *Journal of Clinical Medicine*, vol. 9, no. 5, p. 1492, May 2020, doi: 10.3390/jcm9051492.
- [11] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," arXiv:2002.06563 [q-bio], Jun. 2020.
- [12] P. Teles, "A time-dependent SEIR model to analyse the evolution of the SARS-CoV-2 epidemic outbreak in Portugal," arXiv:2004.04735 [q-bio], Jul. 2020.
- [13] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the COVID-19 outbreak," *PLOS ONE*, vol. 15, no. 3, p. e0230405, May 2020, doi: 10.1371/journal.pone.0230405.
- [14] D. Zhao, J. Sun, Y. Tan, J. Wu, and Y. Dou, "An extended SEIR model considering homepage effect for the information propagation of online social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 512, pp. 1019–1031, Jan. 2018, doi: 10.1016/j.physa.2018.08.006.
- [15] D. Hussain, T. Hussain, A. A. Khan, S. A. A. Naqvi, and A. Jamil, "A deep learning approach for hydrological time-series prediction: A case study of Gilgit river basin," *Earth Sci Inform*, vol. 13, no. 3, pp. 915–927, Sep. 2020, doi: 10.1007/s12145-020-00477-2.
- [16] A. Gasmı, C. Gomez, P. Lagacherie, H. Zouari, A. Laamrani, and A. Chehbouni, "Mean spectral reflectance from bare soil pixels along a Landsat-TM time series to increase both the prediction accuracy of soil clay content and mapping coverage," *Geoderma*, vol. 388, p. 114864, Apr. 2021, doi: 10.1016/j.geoderma.2020.114864.
- [17] F. Serin, Y. Alisan, and A. Kece, "Hybrid time series forecasting methods for travel time prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 579, p. 126134, Oct. 2021, doi: 10.1016/j.physa.2021.126134.
- [18] F. Serin, Y. Alisan, and M. Erturkler, "Predicting Bus Travel Time Using Machine Learning Methods with Three-Layer Architecture," *Measurement*, p. 111403, May 2022, doi: 10.1016/j.measurement.2022.111403.
- [19] X. Chen and D. Cong, "Application of Improved Algorithm Based on Four-Dimensional ResNet in Rural Tourism Passenger Flow Prediction," *Journal of Sensors*, vol. 2022, pp. 1–8, Apr. 2022, doi: 10.1155/2022/9675647.
- [20] A. S. Ahmar and E. B. Del Val, "SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain," *The Science of the Total Environment*, vol. 729, p. 138883, 2020, doi: 10.1016/j.scitotenv.2020.138883.
- [21] T. Chakraborty and I. Ghosh, "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," *Chaos, Solitons & Fractals*, vol. 135, p. 109850, May 2020, doi: 10.1016/j.chaos.2020.109850.
- [22] N. Chintalapudi, G. Battineni, and F. Amenta, "COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach," *Journal of Microbiology, Immunology and Infection*, vol. 53, no. 3, pp. 396–403, May 2020, doi: 10.1016/j.jmii.2020.04.004.
- [23] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons & Fractals*, vol. 135, p. 109864, May 2020, doi: 10.1016/j.chaos.2020.109864.
- [24] P. Wang, X. Zheng, G. Ai, D. Liu, and B. Zhu, "Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran," *Chaos, Solitons & Fractals*, vol. 140, p. 110214, Dec. 2020, doi: 10.1016/j.chaos.2020.110214.
- [25] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons & Fractals*, vol. 138, p. 109942, May 2020, doi: 10.1016/j.chaos.2020.109942.
- [26] V. Singh et al., "Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 23, no. 8, pp. 1583–

- 1597, Feb. 2020, doi: 10.1080/09720529.2020.1784535.
- [27] R. Gupta, G. Pandey, P. Chaudhary, and S. K. Pal, "SEIR and Regression Model based COVID-19 outbreak predictions in India," *Public and Global Health*, Jun. 2020.
- [28] S. Feng, Z. Feng, C. Ling, C. Chang, and Z. Feng, "Prediction of the COVID-19 Epidemic Trends Based on SEIR and AI Models," *Epidemiology*, Dec. 2020.
- [29] Z. Yang et al., "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, pp. 165–174, Sep. 2020, doi: 10.21037/jtd.2020.02.64.
- [30] "CSSEGISandData · GitHub." <https://github.com/CSSEGISandData> (accessed Jun. 06, 2021).
- [31] A. Godio, F. Pace, and A. Vergnano, "SEIR Modeling of the Italian Epidemic of SARS-CoV-2," *MATHEMATICS & COMPUTER SCIENCE*, Jul. 2020.
- [32] I. Rahimi, A. Gandomi, and F. Chen, *Analysis and Prediction of COVID-19 using SIR, SEIR, and Machine Learning Models: Australia, Italy, and UK Cases*. 2020.
- [33] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 1, pp. 241–254, Feb. 1977, doi: 10.1007/BF01593790.
- [34] M. J. D. Powell, "A View of Algorithms for Optimization without Derivatives," p. 12.
- [35] M. Gupta and B. Gupta, "An Ensemble Model for Breast Cancer Prediction Using Sequential Least Squares Programming Method (SLSQP)," in *2018 Eleventh International Conference on Contemporary Computing (IC3)*, Mar. 2018, pp. 1–3. doi: 10.1109/IC3.2018.8530572.
- [36] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning* 1995 20:3, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [37] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, May 2000, doi: 10.1093/bioinformatics/16.10.906.
- [38] V. N. Vapnik, "The Nature of Statistical Learning Theory," *The Nature of Statistical Learning Theory*, 1995, doi: 10.1007/978-1-4757-2440-0.
- [39] O. L. Mangasarian and D. R. Musicant, "Active Support Vector Machine Classification," p. 7, 2000.
- [40] M. Awad and R. Khanna, "Support Vector Regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Eds. Berkeley, CA: Apress, 2015, pp. 67–80.
- [41] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning Precise Timing with LSTM Recurrent Networks," p. 29, 2002.
- [42] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *ICASSP 2013 - 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 6645–6649. doi: 10.1109/ICASSP.2013.6638947.
- [43] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, May 2016, pp. 324–328. doi: 10.1109/YAC.2016.7804912.
- [44] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the COVID-19 epidemic?," *Infectious Disease Modelling*, vol. 5, pp. 271–281, Jun. 2020, doi: 10.1016/j.idm.2020.03.001.



Journal of Soft Computing and Artificial Intelligence

Journal homepage: <https://dergipark.org.tr/en/pub/jscai>

International
Open Access 

Volume 03
Issue 01

June, 2022

Research Article

Investigation of Albedo Factor Parameters in Some Selected Sn Compounds

Ahmet TURŞUCU,^{1,2} 

¹Department of Electric-Electronic Engineering, Sırnak University, 73000, Sırnak, Turkey

²Technology and Research Centre, Sırnak University, 73000, Sırnak, Turkey

ARTICLE INFO

Article history:

Received **May 19, 2022**

Revised **June 15, 2022**

Accepted **June 20, 2022**

Keywords:

Gamma-ray Scattering

Albedo Factors

Albedo Number

Albedo Energy

Albedo Dose

ABSTRACT

In recent years, studies on the elements used in producing electronic device components and the interaction of their different compounds with radiation have been emphasized. In developing this situation, giving importance to space studies and other searches in energy production have been very effective. In the light of these developments, the interaction of tin, which is widely used in producing electronic device components and different industrial areas, with radiation has been investigated. For this purpose, the variation of albedo factor values in some compounds of the tin element was analysed and presented. Am-241 radioactive sources were used as incident radiation in determining the albedo factor values mentioned in the study. The albedo factor values obtained by examining the Compton and coherent scattering peaks were used to determine the albedo factor values.

1. Introduction

Tin is the oldest metal known to human-being. It is increasingly used in many fields such as coatings, compounds, alloys, and advanced technology from ancient times. Today, tin is an essential metal, although it is used in industry in small quantities. This is because it is used in small amounts in many areas. The tin element and some of its compounds, essential among the coating materials, are widely used in the transportation and chemical industry. Tin ore, which has a wide usage area and provides ease of use in many tools, has been the focus of attention of many researchers, and its different properties have been investigated. Tin is used in can making, coating, various alloys, solder, and chemicals. It is also used in engine bearings, bodywork, radiators, oil, and air filters in the automotive industry. It has a wide use area in the

aircraft and ship industry and the electrical and electronics industry. It is consumed in a wide¹ area, from the production of paint, perfume, soap, and polyurethane to toothpaste production in the chemical industry. Besides, it is also used in printing, kitchen equipment, and the glass industry.

We can reach some crucial atomic information by directing X-ray photons on materials. The obtained data is evaluated whether the directed X-rays make elastic or inelastic collisions. [1] These scattering processes help understand the atomic features of each material. In this way, we gain important information about how materials behave when radiation exposure. In calculating albedo factor values, which are the subject of this study, x-rays irradiated the material are scattered from the material by inelastic collision. If we want to give an example of inelastic scattering,

* Corresponding author

e-mail: ahmettursucu@sirnak.edu.tr

DOI: 10.55195/jscai.1118684

soft photon emitting, and target recoil events, all inelastic scattering events come to mind first. Photoelectric, Compton scattering, and pair production events occur because of inelastic scattering. Also, scattering from bound electrons (Rayleigh scattering), nuclear Thomson scattering, Delbrück scattering, and nuclear resonance scattering can be given as the phenomena observed because of elastic scattering.

The photon reflection capability of the target sample is called an albedo factor and is related to the atomic number, thickness, and incident x-ray photon energy [2]. Each target's derived albedo factor values are used in manufacturing radiation shielding material [3-8]. For this purpose, different materials used or recommended in the industry have been researched by various researchers and brought to the literature. Experimental studies were carried out using other materials and radiation sources.

This article uses experimental methods to explain the elastic and inelastic scattering intensity ratios, that is, the albedo factors of the selected samples. Defining the albedo factor parameter, which indicates the photon scattering ability of a material consisting of a combination of different elements or a pure element, will help select the elements to be used in material production. The albedo factor, which is only one of the screening parameters, consists of 3 different components: albedo number, albedo dose, and albedo energy.

2. Material and Methods

2.1. Experimental Arrangement

The experimental procedure of this study was carried out using a semiconductor detector (HPGe) and a circular radioactive source, respectively. These said components are in the scattering geometry and are shown in Figure 1. In addition, the experimental setup was covered with the sample chamber. The sample chamber provided radiation shielding for the researchers,

and at the same time, the chamber prevented unexpected backscatter counts. The structural equipment of the sample chamber is shown in the Figure. 2. The sample chamber is lined with a lead layer and conical shape. The primary use of the sample chamber is also related to this structure. Thus, this shape of the lead-lined and conical-shaped sample chamber provides a monochromatic gamma-ray. Paired monochromatic gamma rays were directed into the sample chamber for scattering to occur so that the scattering photons were detected through an HPGe semiconductor X-ray detector. To align the photons coming from the radioactive source and scattered from the sample and to protect the experimental environment from unwanted harmful x-rays, the experimental setup was kept in the sample room. Gamma rays emitted from the source were directed at the target with a scattering angle of 168 degrees. It is aimed to minimize unwanted gamma-ray scattering with a scattering angle of 168 degrees.

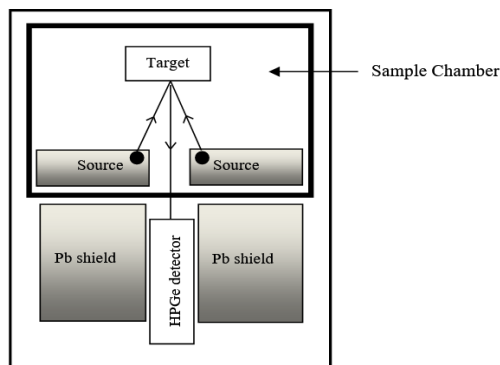


Figure 1 Experimental setup (radius of the collimator is 0.53 cm)

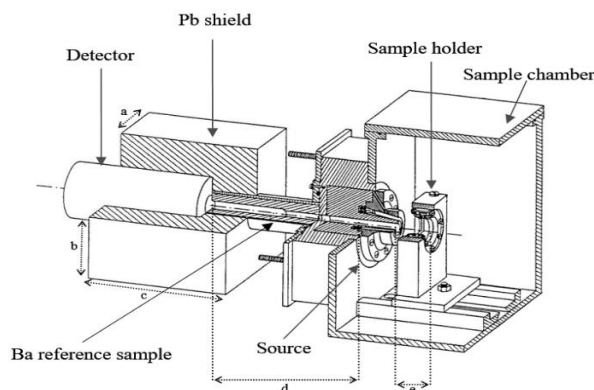


Figure 2 Sample chamber (a=6.5 cm, b=6.3 cm, c=13.5 cm, d=11 cm, e=5 cm)

2.2. Calculation of Albedo Factors

The experimental setup has an HPGe semiconductor detector, and the detector was linked to the multichannel analyzer, Genie-2000, and controlled by a computer. With the mentioned Genie-2000 program, characteristic photo peaks are revealed, and analysis is provided with the data obtained. The determined photopic areas were calculated by using the Origin 7.5 program. The calculated photo-peak areas consist of backscattered and incident constituents and are represented by N_{bs} and N_i , respectively [7]. At this moment, these related photopeak constituents were used in the calculation of the albedo number by the following equation,

$$A_N = \left[\frac{N_{bs}/\varepsilon(E_{bs})}{(N_i/\varepsilon(E_i))(1/d\Omega)(1/2)} \right] \quad (1)$$

Where $\varepsilon(E_{bs})$ and $\varepsilon(E_i)$ terms represent the backscattered and the incident photo-peak efficiencies of the HPGe detector, respectively. It is also seen that the term solid angle ($d\Omega$) is used together with the terms mentioned in the formula. The solid angle term ($d\Omega$) expressed here is the solid angle seen by the detector collimator opening from the center of the target. A factor of 1/2 added to the denominator means that half of the gamma photons emitted by the radioactive source hit the target.

Another constituent of the albedo factor parameter is the albedo energy and calculated by using the following equation,

$$A_E = \left[\frac{E_{bs}}{E_i} \right] A_N \quad (2)$$

Equation (2) has the photon energies of the incident and the backscattered gamma rays, represented by E_i and E_{bs} , respectively. Finally, the calculation of albedo energy has been used to derive the last term of the albedo parameter called the albedo dose. The albedo dose is

directly proportional to the albedo energy, as shown below,

$$A_D = \left[\frac{\sigma_a(E_{bs})}{\sigma_a(E_i)} \right] \quad (3)$$

Equation (3) has the gamma photon absorption coefficient of the air layer. So, the backscattered and incident gamma photon absorption coefficient terms were expressed separately. Also, the air layer is composed of different elements in different percentages, which are taken from Table 1. These coefficient terms have been calculated using related percentages from the XCOM photon cross-section database [8].

Table 1 Table of gaseous composition of dry air

Constituent	Chemical symbol	Mole percent
Nitrogen	N ₂	78.084
Oxygen	O ₂	20.947
Argon	Ar	0.934
Carbon dioxide	CO ₂	0.0350
Neon	Ne	0.001818
Helium	He	0.000524
Methane	CH ₄	0.00017
Krypton	Kr	0.000114
Hydrogen	H ₂	0.000053
Nitrous oxide	N ₂ O	0.000031
Xenon	Xe	0.0000087
Ozone	O ₃	0.0000001

3. Results

This study investigated the radiation absorption and shielding properties of some tin compounds. The fact that tin and its compounds are frequently used in industry and technology has led to this. The tin element has a wide range of uses. The element Sn, which requires a high degree of polishing, is used to protect metal materials prone to corrosion, such as coating steel cans. Materials such as soft solder, tin, bronze, and phosphor bronze are important tin alloys. In addition, the niobium-tin alloy is widely used for superconducting magnets. It is obtained using molten glass floating on the molten tin to create a flat surface on most window glass. In addition, molten tin salts are sprayed onto the glass surface to form a

conductive structure. The most essential tin salt used in these processes is tin (II) chloride, used as a reducing agent and mordant in dyeing calico and silk. Tin (IV) oxide is frequently used in ceramic and gas sensor production.

Zinc stannate (Zn_2SnO_4) is a fire retardant used in plastics. Some tin compounds have been used as antifouling paints for ships and boats to prevent mussels. However, these compounds are deadly to marine life, especially oysters, even at low levels. Its use is now banned in most countries. Tin is a bidirectional metal because it can react with strong acids and bases. Pure water does not affect it. In aqueous solutions, oxygen accelerates the corrosion process. In the absence of oxygen, the high potential of tin (0.75 V) causes a hydrogen layer to form on its surface, which will reduce even the effect of acids. Normally, its surface and thickness are covered with a thin oxide layer that increases the temperature. When a tin-plated steel in an airless environment is brought into contact with acidic solutions, a tin-iron couple is formed at negative potential. During this event, the tin plating is used as the anode, and the molecule itself, not the iron, undergoes aggregation. Otherwise, molten iron; In canning, which is its main area of use, the process is a cornerstone because it impairs taste and vision. Tin; does not react directly with hydrogen, nitrogen, carbon dioxide and gas ammonia. When moist, sulfur dioxide affects tin. It reacts readily with chlorine, bromine, and iodine and slowly with phosphorus at room temperature. Halogen acids react violently with tin, especially when hot and concentrated. Hot sulfuric acid melts the tin. Reacts gradually with nitric acid; however, it becomes hydrated tin oxide when the acid is concentrated. It reacts rapidly with sulfuric, chlorosulfuric and pyrosulfuric acids. The tin dissolving effect of phosphoric acid is much less than other mineral acids. In the atmospheric environment, it reacts slowly with organic acids such as lactic acid, citric acid, tartaric acid, and oxalic acid. The effect on diluted solutions such as ammonium

hydroxide and sodium carbonate are weak; However, strong alkalis such as sodium and potassium hydroxide dissolve tin when cold and diluted to form tin compounds with (+4) valence. Oxidizing salts and their solutions, such as potassium peroxy sulphate, ferric chloride, sulfate, and stannous chloride, dissolve tin.

X-ray room armouring is one of the foremost safety procedures for X-ray rooms. According to the medical study regulation prepared by the Ministry of Health, it is forbidden to carry out radioactive treatment and diagnostic studies without radiation shielding. Taking radiation shielding measures is very, very important for the protection of human health. In addition, it includes significant measures for both human health and the stable operation of other devices. To prevent the radiation emitted by X-rays and minimize the harmful effects, it is necessary to cover the rooms where the X-ray films are taken. In medical treatment environments, especially in chemotherapy units where cancer patients undergo chemotherapy, the rays that may cause radioactive damage should be minimized without damaging the room design, and the beam source should be well protected. In such environments, shielding can be made with copper, reducing the radiation intensity. High radiation level, maximum ideal level is reached with shielding. How to determine the ideal armour thickness to reduce the radiation level by making some calculations. The calculation is based on the radiation dose value and MMED value that may occur in case of overwork. Products used for armouring can be lead, concrete, copper, barite aggregate and heavy concrete. The materials used in the armour vary according to the radioactive source and radiation type. In treatment environments with low radiation intensity, the shielding process can be made of light aluminium. In high-risk environments, the material with the highest atomic number among the materials to protect the rooms is lead.

In this study, albedo factor analyses were performed for some tin compounds, one of the most widely used materials for shielding environments where radiation-based treatment methods are used. Albedo factor is only one of the radiations shielding parameters and consists of 3 different sub-terms. The sub-terms mentioned are called albedo number, albedo energy and albedo dose.

4. Discussion

The calculation of each parameter according to the gamma-ray counts obtained from the experiment was determined using the relevant formulas. The albedo factor values calculated using the formulas are given in Figures 1-3. If we interpret the numbers, the albedo parameter values decrease with the increasing value of the atomic numbers of the target samples. The increase in photoelectric cross-section and decrease in Compton cross-section with a rising atomic number are the main reasons for this situation.

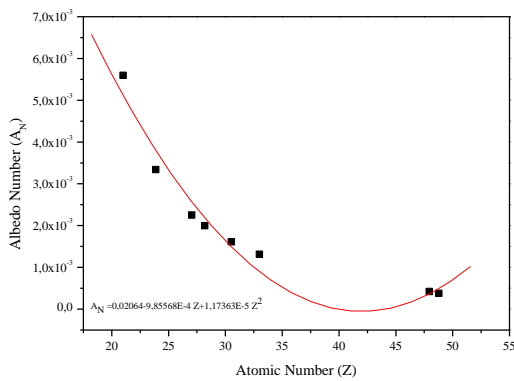


Figure 3 The albedo number distribution of target samples is a function of the atomic number

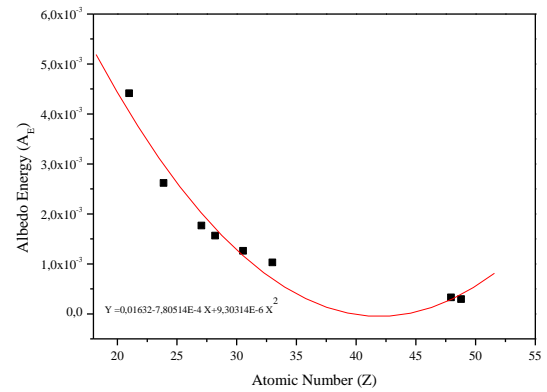


Figure 4 The albedo energy distribution of target samples is a function of the atomic number

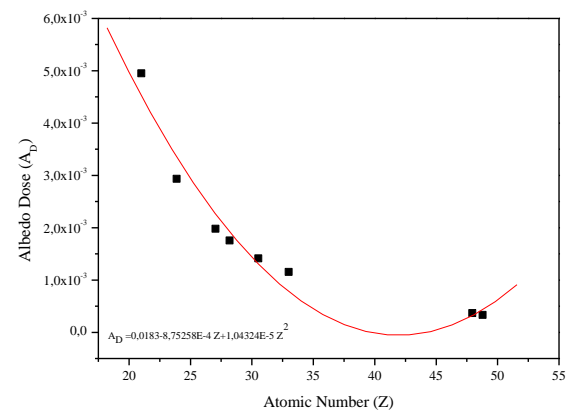


Figure 5 The albedo dose distribution of target samples is a function of the atomic number

References

- [1] Du Mond, Jesse W. M., 1929. "Compton modified line structure and its relation to the electron theory of solid bodies." *Physic. Rev.*, 33 (5). Pp
- [2] Uzunoglu, Z., Yilmaz, D., Şahin, Y., 2017 "Determination of the saturation thickness and albedo factors for mercury(II) oxide and lead(II) oxide." *Instrum. SCi. Technol.*, 45(1): 111-121.
- [3] Ljubenov, V., et al., 2010. "Total reflection coefficients of low-energy photons presented as universal functions." *Nucl. Technol. Radiat.* 25(2), 100-106
- [4] Sabharwal, A. D., Singh., B., Sadhu, B., S. 2009. "Investigation of multiple backscattering and albedos of 1.12 MeV gamma in elements and alloys." *Nucl. Instrum. Meth. B.* 267, 151-156.
- [5] Gilfrich, J., V., Brown, D., B., Burkhalter, P., G., 1975. "Integral reflection coefficient of X-ray Spectrometer Crystal." *Appl. Spectrosc.* 29(4), 322-326.
- [6] Kadotani, H., Shimizu A., 1998. "Gamma ray albedo data generated by the invariant embedding method." *J. Nucl. SCi. Technol.* 35:8, 584-594.
- [7] Kiran, K. U., et al., 2016. "Albedo factors of 123, 320, 511, 662 and 1115 keV gamma photons in carbon, aluminium, iron and copper." *Eur. Phys. J. Plus.*, 131(4).
- [8] Sabharwal, A. D., Sadhu, B., S., Singh., B., 2011. "Albedo factors of 279, 320, 511 and 662 keV backscattered gamma photons." *Radiat. Eff. Defect. S.*, 166(6): 451-458.
- [9] Gerward, L., et al., 2001. "X-ray absorption in matter. reengineering XCOM." *Radiat. Phys. Chem.*, 60(1-2): 23-24.
- [10] Weyrich, W., 1975. "The electron momentum distribution in solidpotassium fluoride, studied by Compton scattering." *Berichte der Bunsen-Gesellschaft für physikalische Chemie(früher Zeitschrift für Elektrochemie)* 79(11), 1085-1095.
- [11] Yilmaz, D., Uzunoğlu, Z., Demir, C., 2017. "Albedo factors of some elements in the atomic number range $26 \leq Z \leq 79$ for 59.54 keV." *Appl. Radiat. Isotopes* 122: 68-71.



Journal of Soft Computing and Artificial Intelligence

Journal homepage: <https://dergipark.org.tr/en/pub/jscai>

International
Open Access 

Volume 03
Issue 01

June, 2022

Research Article

Hybrid experimental investigation of MR damper controlled tuned mass damper used for structures under earthquakes

Huseyin Aggumus^{1*} , Rahmi Guclu² 

¹Department of Mechanical and Metal Technologies, Sirtak University, Sirtak, 73000, TURKEY

²Department of Mechanical Engineering, Yildiz Technical University, Besiktas 34349, Istanbul, TURKEY

ARTICLE INFO

Article history:

Received **May 27, 2022**

Revised **June 13, 2022**

Accepted **June 19, 2022**

Keywords:

Structural vibration control, MR damper, Real-time hybrid simulation (RTHS), Semi-active tuned mass damper (STMD)

ABSTRACT

This paper aims to investigate the performances of a semi-active tuned mass damper (STMD) used to reduce the vibrations of buildings under different seismic excitation by the real-time hybrid simulation (RTHS) method. In the STMD, the MR damper is used as a control element with a variable damping feature. The RTHS method is an alternative to experimentally studying the STMD system. MR damper is critically significant for the system and is experimentally installed. At the same time, the other parts are designed in numerical simulation and tested simultaneously. MR damper is a control element whose damping value can change according to the amount of voltage transmitted. Therefore, the groundhook control method determines the MR damper voltage variations. The results show that the control method applied to MR damper-controlled STMD effectively suppresses structural vibrations.

1. Introduction

Passive control applications have been implemented to improve the response of a structural system under the influence of earthquake excitations. In order to increase the performance of passive control applications, semi-active control studies [1]–[4] and active control studies [5]–[7] are performed. Tuned mass dampers (TMD) are commonly used in control applications in structural systems. TMDs applied to structural systems are used as passive control elements [8], [9]. Also, the performance of suppressing structural vibrations can be improved by applying active and semi-active control methods. Active control applications are successful in reducing structural vibrations.

Nevertheless, the cost and reliability of the equipment for this application are disadvantageous. The performance of TMDs with passive control is

limited. Their performance decreases under variable conditions. Semi-active control methods applied to TMDs are safer than active control and perform better than passive control. In this study, STMD reduces the vibrations of a building with MDOF under seismic excitation.

Many simulation studies are related to STMDs using MR damper in the literature. Causal sub-optimal control [10], Clipped optimal control [11], LQR control [12], Sky-groundhook control with optimal fuzzy control [13], MIMO fuzzy logic control [14], Type 1 and 2 fuzzy logic control [15], Bang-Bang control [16]. Besides, many experimental studies have been made; Groundhook control [17], LQR control [18], and model-based feed-forward control [19]. Numerical simulations may not ultimately reflect reality. Experimental studies are of great importance in scientific research.

* Corresponding author

e-mail: haggumus@sirtak.edu.tr

DOI: 10.55195/jscai.1122514

However, the experimental setup and operation of structural systems are also significant problems.

In recent years, hybrid simulation methods have been applied for structural systems, including simulation and experimental studies. These methods combine the advantages of numerical simulation and experimental setups. The objective here is to experimentally construct system elements that are difficult to model mathematically and to use the measurement data obtained in this way in simulation in real-time. The responses of structures under harmonic excitation were investigated by the RTHS method [2]. However, the response of a building under earthquake reactions is an issue that needs to be investigated.

In this study, the performance analysis of the STMD, which is used to reduce the vibrations of MDOF building under seismic excitations by the RTHS method, is investigated. MR damper is a control element whose damping value can change according to the amount of voltage transmitted. For this reason, the groundhook control method is used for the voltage changes in the MR damper because it is simple and easy to implement.

2. Equation motion of system model

In this study, the semi-active control application of the 10-story building model, where only lateral vibrations are taken into account, is shown in Figure 1.

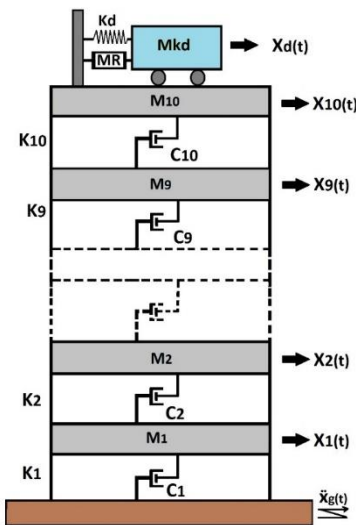


Figure 1. The building model with the STMD

The general equation of motion of the model is as follows.

$$M_s \ddot{x}(t) + C_s \dot{x}(t) + K_s x(t) = -T_s f(t) - M_s Z \ddot{x}_g \quad (1)$$

The mass, damping, and stiffness matrices of the system are M_s, C_s ve $K_s \in R^{11 \times 11}$ and the acceleration, velocity, and displacement vectors are $\ddot{x}(t), \dot{x}(t), x(t) \in R^{11 \times 1}$ respectively. The damping force of the MR damper is $f(t)$, and the earthquake ground acceleration is \ddot{x}_g . In Equation 2 and Equation 3, the displacement and seismic vectors of the system are given.

$$x = [x_1 \ x_2 \ x_3 \ \dots \ x_9 \ x_{10} \ x_d]^T \quad (2)$$

$$Z = [1 \ 1 \ 1 \ \dots \ 1 \ 1 \ 1]^T \quad (3)$$

The vector showing the location of the controller is shown below.

$$T_s = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ -1]^T \quad (4)$$

3. Semi-Active Control Application

The groundhook configuration shown in Figure 2 cannot be practiced. Because the damper cannot be fixed to a stationary inertia frame, this semi-active control policy aims to mimic the ideal structural configuration of a passive damper between structure and ground [13] [20]. In Figure 2, the relative velocity is defined by subtracting the velocity of the structure from the velocity of the STMD, that is, $(dx_{10} - dx_d)$. Also, x_{10} is defined as displacement.

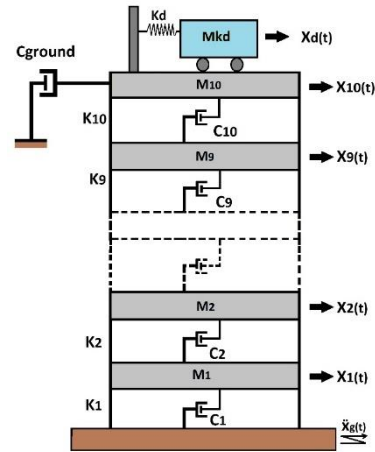


Figure 2. The groundhook control configuration

The groundhook control algorithm alternates between two different options for voltage determination, maximum and minimum. These are accomplished with a simple algorithm shown below.

$$V = \begin{cases} V_{\max} & \text{if } x_{10}(dx_{10} - dx_d) \geq 0 \\ V_{\min} & \text{else } x_{10}(dx_{10} - dx_d) < 0 \end{cases} \quad (5)$$

Here, V_{\min} and V_{\max} represent the minimum and maximum voltages, respectively. As shown in Equation (5), the voltage transmitted to the MR damper can be determined by the command based on groundhook control. It has been effectively applied to real-time systems due to its simplicity and ease of application [13]. Therefore, the groundhook control is preferred for the hybrid system in this study.

4. Experimental System and Performance Analysis

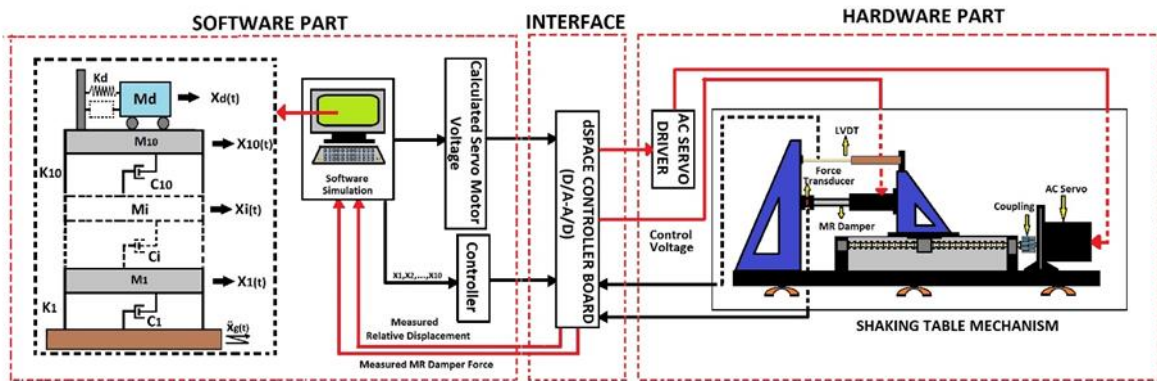


Figure 3. Application scheme of the RTHS method [21]

Structural control studies with the RTHS method are examined using the shaking table in Yıldız Technical University Vibration Research and Control Laboratory. In the experimental part, where force and relative displacement are measured, one force sensor and one linear variable position sensor are used. One RD-8041-1 type MR damper is used, which is the experimental part of the STMD control application. A computer communicates with the dSpace interface and is used to run the data processing, control, and shaking table.

4.2 Hysteresis properties of MR damper and system parameters

The characteristic features of the MR damper connected to the experiment setup are obtained with a sinusoidal input of 1 Hz and 5 mm amplitude. The determined force-displacement and force-velocity curves are shown in Figure 4.

The parameters of a ten-story building are used by scaling in STMD control application with the RTHS method [22], [23]. Parameters of the model in Figure 1; $m_{1...10} = 72$ t, $k_{1...10} = 13 \times 10^7$ N/m, $c_{1...10} = 1.24 \times 10^6$ Ns/m [2]. The mass ratio for the tuned

4.1 Introduction of Experimental Setup

In this study, the implementation of the RTHS method takes place in two steps. Firstly, the relative speed data read from the building model in the computer simulation is transmitted to the experimental setup. Secondly, the MR damper force generated in the experiment set up by the effect of control methods is transmitted to computer simulation. The schematic view of the RTHS method is shown in Figure 3

mass damper is 0.03, and its parameters are determined as $m_d = 21.6$ t ve $k_d = 81 \times 10^4$ [8]. The system is excited by the accelerations of the El-Centro and Kocaeli earthquakes. Excitations are scaled due to the stroke limitation of the MR damper (El-Centro (0.04), Kocaeli (0.075)). The maximum and minimum voltages transmitted to the MR damper are $V_{\max} = 10$ volts and $V_{\min} = 0$.

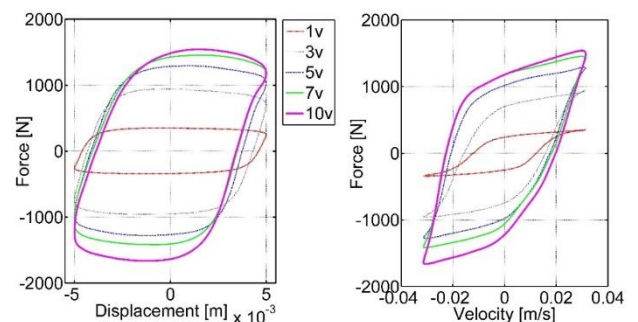


Figure 4. MR damper characteristics

4.3 Time responses

El-Centro and Kocaeli Earthquake excitations effect of the structural system of the maximum displacement and displacement RMS responses of all floors are researched. Figure 5 shows the

displacement and displacement RMS values of all floors. The groundhook control method (STMDg) has successfully reduced the vibrations of the structural system in all excitation situations.

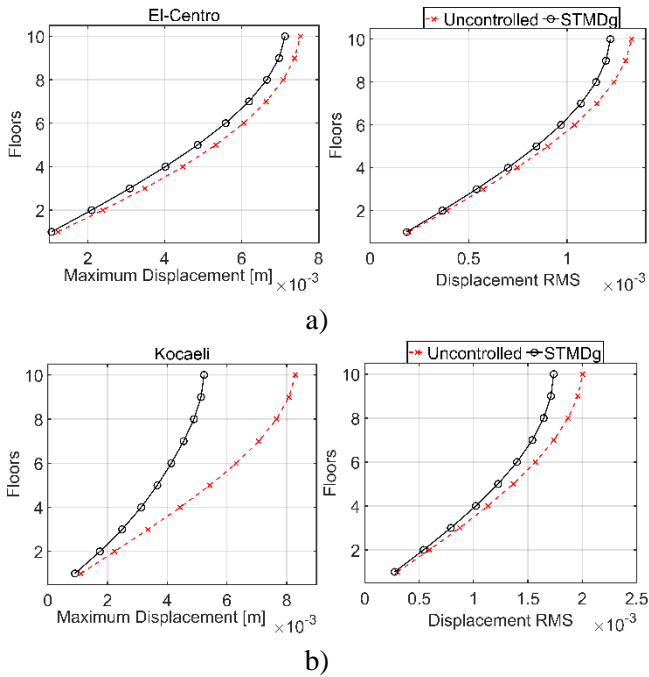


Figure 5. Maximum displacement and displacement RMS responses of all floors of the STMD a) El-Centro earthquake b) Kocaeli earthquake

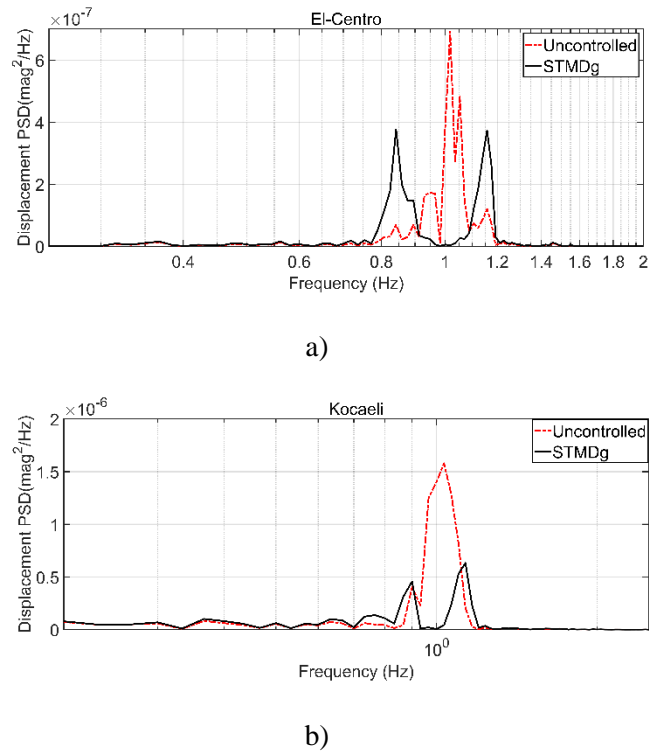


Figure 6. 10. Floors PSD responses of the STMD a) El-Centro earthquake b) Kocaeli earthquake

4.4 Frequency analysis and MR damper data

For the frequency-domain analysis, power spectrum density (PSD) responses of the frequency variations of the system are researched. The PSD curves of the displacements of the 10th floors for all excitation cases are shown in Figure 6. It is observed that the semi-active controllers successfully suppress the resonance peaks.

Table 1 shows the MR damper data read from the experimental setup and used in the computer simulation. It is seen that the maximum forces are almost equal in both earthquake excitations, and more force is produced in the Kocaeli earthquake in the force RMS values

Table 1. Experimental data of the MR damper

Earthquakes	Forces [N]		Voltages [V]	
	Maximum	RMS	Maximum	RMS
El-Centro	1959.64	490.8016	10	5.80
Kocaeli	1985.10	624.3378	10	5.94

4.5 Structural Vibration Performance Evaluations

Performance indices for the structural system are given as follows [24].

$$J_1 = \max \left\{ \frac{\max_{t,i} \left\{ \frac{|d_i(t)|}{h_i} \right\}}{\delta_{\max}} \right\}, J_2 = \max \left\{ \frac{\max_{t,i} \{ |\ddot{x}_{ai}(t)| \}}{\ddot{x}_a^{\max}} \right\} \quad (6)$$

$$\delta_{\max} = \frac{|d_i(t)|}{h_i} \quad (7)$$

Here δ_{\max} , \ddot{x}_a^{\max} , d_i and h_i the maximum inter-story drift ratio, absolute acceleration, the distance between floors, and relative displacement between floors, respectively. The performance indices in Table 2 compare the cases of no control application, and the cases of STMD are compared. Among the performance indices, J_1 is calculated with the maximum displacement between floors, while J_2 is calculated with the maximum acceleration value. Performance indices are expected to be smaller than one to improve system responses. It is seen that STMDg improves system responses in all performance indices in both seismic excitations affecting the structural system. In Kocaeli excitation, the performance index of j_2 in all control cases is similar to the uncontrolled state. However, this is acceptable due to improvements in displacement performance indices.

Table 2. Performance Indices

Performance Indices	El-Centro	Kocaeli
	STMDg	STMDg
J ₁	0,9126	0,8012
J ₂	0,9932	1,0046

5. Conclusion

This paper investigates the structural vibration suppression performances of the STMDs with the RTHS method. The software part of the RTHS method consists of a building model computer and control algorithm, and the experimental part consists of an MR damper and sensors. The semi-active control element in STMD is the MR damper. The groundhook control algorithm determines the voltage transmitted to the MR damper. The system is excited with the El-Centro and Kocaeli earthquakes, and its displacement and PSDs are investigated. In all time responses, it has been found that the STMDg has improved system responses. PSD curves are analyzed for the frequency domain of system responses. Displacement PSD analysis shows that the STMDg is suppressing resonance peaks. The results show that the performances of STMDs used in structural vibration control can be effectively executed with RTHS, an alternative method to experimental setups.

References

- [1] H. Aggumus and S. Cetin, "Experimental investigation of semiactive robust control for structures with magnetorheological dampers," *J. Low Freq. Noise Vib. Act. Control*, vol. 37, no. 2, pp. 216–234, Jun. 2018, doi: 10.1177/0263092317711985.
- [2] H. Aggumus and R. Guclu, "Robust H_∞ Control of STMDs Used in Structural Systems by Hardware in the Loop Simulation Method," *Actuators*, vol. 9, no. 3, p. 55, Jul. 2020, doi: 10.3390/act9030055.
- [3] M. Paksoy and H. Aggümüş, "MR Sönümleyicili Yarı Aktif Ayarlı Kütle Sönümleyicisinin Uyarlamalı Kontrolü," *Avrupa Bilim Ve Teknol. Derg.*, no. 33, Art. no. 33, Jan. 2022, doi: 10.31590/ejosat.1020498.
- [4] A. Turan, H. Aggümüş, Mr Damperli Yarı Aktif Yapısal Sistem İçin Optimal Pid Kontrolcü Tasarımı, Mühendislik Alanında Uluslararası Araştırmalar II, Eğitim Yayınevi, ss 101-110, 2021.
- [5] R. Guclu and H. Yazici, "Seismic-vibration mitigation of a nonlinear structural system with an ATMD through a fuzzy PID controller," *Nonlinear Dyn.*, vol. 58, no. 3, pp. 553–564, Nov. 2009, doi: 10.1007/s11071-009-9500-5.
- [6] R. Guclu and H. Yazici, "Vibration control of a structure with ATMD against earthquake using fuzzy logic controllers," *J. Sound Vib.*, vol. 318, no. 1–2, pp. 36–49, Nov. 2008, doi: 10.1016/j.jsv.2008.03.058.
- [7] R. Guclu, "Fuzzy logic control of vibrations of analytical multi-degree-of-freedom structural systems," *Turk. J. Eng. Environ. Sci.*, vol. 27, no. 3, pp. 157–168, 2003.
- [8] A. Y. T. Leung and H. Zhang, "Particle swarm optimization of tuned mass dampers," *Eng. Struct.*, vol. 31, no. 3, pp. 715–728, Mar. 2009, doi: 10.1016/j.engstruct.2008.11.017.
- [9] P. Xiang and A. Nishitani, "Optimum design for more effective tuned mass damper system and its application to base-isolated buildings: OPTIMUM DESIGN FOR MORE EFFECTIVE TMD," *Struct. Control Health Monit.*, vol. 21, no. 1, pp. 98–114, Jan. 2014, doi: 10.1002/stc.1556.
- [10] U. Aldemir, "Optimal control of structures with semiactive-tuned mass dampers," *J. Sound Vib.*, vol. 266, no. 4, pp. 847–874, Sep. 2003, doi: 10.1016/S0022-460X(03)00191-3.
- [11] P. Y. Lin, L. L. Chung, and C. H. Loh, "Semiactive Control of Building Structures with Semiactive Tuned Mass Damper," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 20, no. 1, pp. 35–51, Jan. 2005, doi: 10.1111/j.1467-8667.2005.00375.x.
- [12] K. T. Tse, K. C. S. Kwok, P. A. Hitchcock, B. Samali, and M. F. Huang, "Vibration control of a wind-excited benchmark tall building with complex lateral-torsional modes of vibration," *Adv. Struct. Eng.*, vol. 10, no. 3, pp. 283–304, 2007.
- [13] H.-S. Kim and J.-W. Kang, "Semi-active fuzzy control of a wind-excited tall building using multi-objective genetic algorithm," *Eng. Struct.*, vol. 41, pp. 242–257, Aug. 2012, doi: 10.1016/j.engstruct.2012.03.038.
- [14] H.-S. Kim, "Seismic response control of adjacent buildings coupled by semi-active shared TMD," *Int. J. Steel Struct.*, vol. 16, no. 2, pp. 647–656, Jun. 2016, doi: 10.1007/s13296-016-6030-0.
- [15] A. Bathaei, S. M. Zahrai, and M. Ramezani, "Semi-active seismic control of an 11-DOF building model with TMD+MR damper using type-1 and -2 fuzzy algorithms," *J. Vib. Control*, vol. 24, no. 13, pp. 2938–2953, Jul. 2018, doi: 10.1177/1077546317696369.
- [16] Aly, Aly Mousaad, "Control of wind-induced motion in high-rise buildings with hybrid TM/MR dampers," *Wind Struct.*, vol. 21, no. 5, pp. 565–595, Nov. 2015, doi: 10.12989/WAS.2015.21.5.565.

- [17] M. Setareh, J. K. Ritchey, T. M. Murray, J.-H. Koo, and M. Ahmadian, "Semiactive Tuned Mass Damper for Floor Vibration Control," *J. Struct. Eng.*, vol. 133, no. 2, pp. 242–250, Feb. 2007, doi: 10.1061/(ASCE)0733-9445(2007)133:2(242).
- [18] P.-Y. Lin, T.-K. Lin, and J.-S. Hwang, "A semi-active mass damping system for low- and mid-rise buildings," *Earthq. Struct.*, vol. 4, no. 1, pp. 63–84, Jan. 2013, doi: 10.12989/EAS.2013.4.1.063.
- [19] F. Weber, H. Distl, S. Fischer, and C. Braun, "MR Damper Controlled Vibration Absorber for Enhanced Mitigation of Harmonic Vibrations," *Actuators*, vol. 5, no. 4, p. 27, Dec. 2016, doi: 10.3390/act5040027.
- [20] J.-H. Koo, M. Ahmadian, M. Setareh, and T. Murray, "In Search of Suitable Control Methods for Semi-Active Tuned Vibration Absorbers:," *J. Vib. Control*, Feb. 2004, doi: 10.1177/1077546304032020.
- [21] H. Aggümüş, Simülasyon Çevriminde Donanım Yöntemiyle Yarı Aktif Ayarlı Kütle Sönümleyicilerinin Performans Analizi. Doktora Tezi, İstanbul: YTÜ Fen Bilimleri Enstitüsü, 2020.
- [22] M. N. S. Hadi and Y. Arfiadi, "Optimum Design of Absorber for MDOF Structures," *J. Struct. Eng.*, vol. 124, no. 11, pp. 1272–1280, Nov. 1998, doi: 10.1061/(ASCE)0733-9445(1998)124:11(1272).
- [23] M. P. Singh, E. E. Matheu, and L. E. Suarez, "Active and semi-active control of structures under seismic excitation," *Earthq. Eng. Struct. Dyn.*, vol. 26, no. 2, pp. 193–213, 1997.
- [24] Y. Ohtori, R. E. Christenson, B. F. Spencer Jr, and S. J. Dyke, "Benchmark control problems for seismically excited nonlinear buildings," *J. Eng. Mech.*, vol. 130, no. 4, pp. 366–385, 2004.
- [25] S. Cetin, E. Zergeroglu, S. Sivrioglu, and I. Yuksek, "A new semiactive nonlinear adaptive controller for structures using MR damper: design and experimental validation," *Nonlinear Dyn.*, vol. 66, no. 4, pp. 731–743, 2011.



Journal of Soft Computing and Artificial Intelligence

Journal homepage: <https://dergipark.org.tr/en/pub/jscai>

International
Open Access 

Volume 03
Issue 01

June, 2022

Research Article

Time Series Cleaning Methods for Hospital Emergency Admissions

Yigit Alisan¹, Olcay Tosun^{2*}

¹Distance Education Application and Research Center Sinop University, 57000 Sinop, Turkey

²Quality Management Department, Sinop University, 57000 Sinop, Turkey

ARTICLE INFO

Article history:

Received **June 6, 2022**

Revised **June 24, 2022**

Accepted **June 27, 2022**

Keywords:

LSTM

Data cleaning

Emergency services

Time series

ABSTRACT

Due to the nature of hospital emergency services, density cannot be easily estimated. It is one of the important issues that should be planned for emergency service managers to have sufficient resources continuously in services that develop suddenly, and emergency interventions are made for human life. Effective and efficient management and planning of limited resources are important not only for hospital administrators but also for people who will receive service from emergency services. In this situation, estimating the number of people who will request service in the emergency service with the least error is of great importance in terms of resource management and the operations carried out in the emergency services. The density of patients coming to the emergency department may vary according to the season, special dates, and even time zones during the day. The aim of the study is to show that more successful results will be obtained because of processing the time series by considering the country and area-specific features instead of the traditional approach. In this paper, the patient admission dataset of the public hospital emergency service in Turkey was used. Data cleaning and arranging operations were carried out by considering the official and religious special days of Turkey and the time periods during the day. The data set is first handled holistically, and its performances are measured by making predictions with the LSTM (Long Short Term Memory) model. Then, to examine the effect of time zones, performance values were calculated separately by dividing each day into 3 equal time zones. Finally, to investigate the effect of triage areas on the total density, the model performance was measured by dividing the data forming each time zone into 3 different triage areas in 3 equal time periods. Three stages were applied both on the raw data set and on the data created by extracting the official, religious holidays, and weekend data specific to Turkey. According to the MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error) results, more successful results are obtained thanks to the cleaning and editing processes. Thanks to the study, it is thought that the data sets used for demand forecasting studies in the health sector will produce results closer to reality by determining and standardizing the purification criteria in this way.

1. Introduction

Emergency services, unlike polyclinics, provides basic care support to patients and accepts patients in 24 hours a day, 7 days a week [1]. In general, patients may come to the emergency services due to a sudden onset of illness, an undiagnosed illness or sudden injury. As in every field, resources are limited in the field of health. In this case, hospital managers must

perform effective and efficient resource management and planning processes to meet the needs. Personnel needs in emergency services are generally met by a doctor and nurses on duty, regardless of their specialty, in 8-hour shifts. This brings additional workload for duty personnel such as doctors and nurses. Patients who apply to the emergency departments of hospitals operating in Turkey are first

* Corresponding author

e-mail: olcaytosun@gmail.com

DOI: 10.55195/jscai.1126611

determined by the diagnostic desk to which triage area they will follow [1]. In some countries in Europe, triage categorization is made over 5 colors. Red, orange, yellow, green, and blue colors indicate urgent, very urgent, urgent, standard, and non-emergency situations, respectively [2]. Three colors are used for triage categorization in Turkey. Category 1 (red), category 2 (yellow), category 3 (green) means immediate intervention, intervention within one hour at the latest, and intervention within three hours at the latest, respectively. In this way, it is aimed to use human and logistic resources correctly and effectively and to intervene quickly to patients. Today, the increase in people's expectations from the health system leads to a lack of capacity and resources, which in turn reduces their satisfaction with the service provided. In order to minimize such negative situations and to perform resource management in the most effective way, it has become important to dynamically adjust resources according to short-term forecasts. There are some factors having effects on the time series. These factors, time series components, are named as trend (T), regular fluctuations (M), irregular fluctuations. Trend shows the main tendency of the time series in the long run [3]. Regular fluctuations are fluctuations having both cyclical and periodicity properties [4]. Clinics may be out of service in case of special days (official-religious holidays), disasters, or epidemics that vary according to countries. The number of applications to the emergency department may increase in such cases. It causes a decrease in the performance of estimator models due to the effect of such irregular fluctuations on the time series.

In this study, it is aimed to reveal the effects of the above-mentioned conditions on time series forecasting performances. First, LSTM (Long Short Term Memory) model performance was measured on the raw data. Then, raw data was edited by extracting daily data for the weekend and holidays in 2015. Performance of LSTM was measured again on edited data. In addition, using the edited data, performance measurement was made with the LSTM model by dividing a day into three 8-hour slices and 3 different triage situations. In this study, the contribution of modifications made in the raw datasets used for forecasting of the hospital emergency service units was aimed to be revealed by comparing the model performance obtained because of the statistical metrics.

This study is divided in 6 sections. Basic information about the research subject is given in the introduction section. In the literature review section, studies on traditional time series models and the

LSTM model based on deep learning are introduced. In the material and method section, it is detailed the modifications performed on the data set used in the study, the LSTM model used and the performance criteria. In the findings section, the performance on modified data sets of LSTM model is given. In the fifth part, evaluation and conclusion part, the performance of the LSTM on the data preparation method are evaluated. In the last section, future studies, the limitations of this study and recommendation for future studies are shared.

2. Literature Review

Time series are in demand in a very wide area in daily life, from stock market to transportation, from telecommunication to energy sector. Time series are frequently used in decision-making mechanisms as they make it easier for us to make predictions for the future on management issues such as investment, planning and optimization. In the transportation sector, models have been developed using time series for calculating the transportation times of vehicles, especially in metropolitans that have transportation problems due to traffic density, for planning urban transportation vehicles [5] [6]. There are lots of studies on changes and fluctuations on time series data sets, especially in areas such as seasonal electricity consumption [4], [7], natural gas consumption [8], [9], [10] economy [11] and food [12]. Box-Jenkins (ARIMA- Autoregressive Integrated Moving Average) models (AR- Autoregression, MA- Moving average, ARMA- Autoregressive moving average) [13] have been used in many fields such as furniture [14], finance [15], energy [16], food [17] for discrete and linear time series datasets. In the healthcare field, in addition to the emergency department density estimation [18] [19], covid-19 [20], the number of calls to the 112-emergency call center [21], the average cost per prescription [22], the need for medical supplies [23], serum set consumption [24], Electrocardiogram (ECG) signal analyzes [25] and hospital disaster preparedness [26] have also been used for estimation purposes. LSTM (Long Short Time Memory) with recent success in deep learning approaches [27] [28] [29] has been used in many fields such as [30] financial [31], energy [32], health [33] [34] [35] [36]. In addition, the LSTM shows high performance in areas such as handwriting recognition [37] [38],

translation [39]. It produces more successful results than traditional/statistical models in predictions made using time series [40].

3. Material and Method

3.1 Dataset

The data set used in the study consists of information of patients who applied to emergency services of public hospital operating in Turkey between 01.01.2015 and 31.12.2015 in yearly, monthly, daily, hourly, minutely, and secondly. Triage sections which the patients they were referred are also included in dataset (Table 1).

Table 1 Emergency service data set sample

Examination	Triage Epicrisis Entry Date	Triage Information
Green field examination	1/1/2015 0:17	GREEN
Green field examination	1/2/2015 19:49	GREEN
Green field examination	1/2/2015 21:35	GREEN
Green field examination	1/2/2015 22:16	GREEN

First, the raw data set was grouped to show the total number of patients who applied in one day (Data Set-1). Secondly, weekends, public and religious holidays were excluded and grouped in the same way (Data Set-2). Finally, for both data sets, a day was divided into three equal time periods (shifts) and grouped daily for each time period (Data Set-3 and Data Set-4). Finally, for both data sets, a day was divided into three equal time periods (shifts) and grouped daily for each time period (Data Set-3 and Data Set-4). While creating Data Set-2, all weekend holidays of 2015 were excluded from the dataset. At the same time, only the days covering the Feast of Sacrifice were extracted from the data set. Only public holidays are excluded. The aim here is that official holidays are limited to one day and other religious holidays do not have a dominant effect on the data. At the same time, it was aimed to keep sufficient data to reach realistic results.

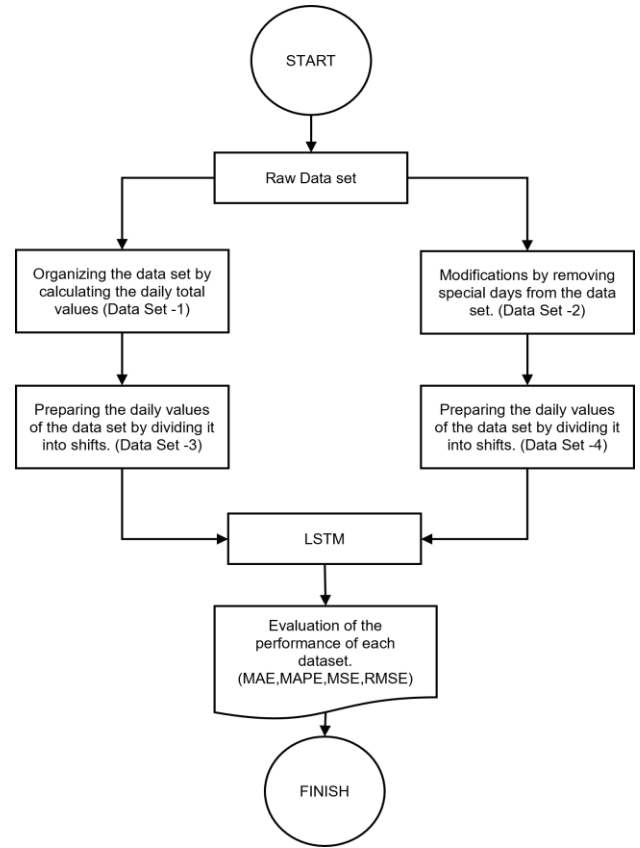


Figure 1 Data preparation and make regression process steps

3.2 LSTM (Long Short-Term Memory) Model

The LSTM was introduced to the literature by Sepp Hochreiter and Jurgen Schmidhuber in 1997 [41]. Later, the model took its current form [42] with some innovations and regulations such as adding the forget gate [43] and the activation function [44]. Since LSTM can learn long-term trends with its memory that can hold historical data, sequential or time series performs well [22]. Equations of input, output, forget gates and memory cells using in basic LSTM architecture are shown in Equations 1, 2, 3, 4, 5 and 6, respectively [45].

$$i_t = \sigma(W_{ix}x_t + W_{hh}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{ch}h_{t-1} + W_{cc}c_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_{t-1} + b_o) \quad (4)$$

$$h_t = o_t \odot h(c_t) \quad (5)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

i, f, o, c, w and $\sigma(x)$ represents input gate, forget gate, output gate, cell activation vector, weight matrix and sigmoid function respectively (Eq. 6). The \odot symbol represents the scalar product of two vectors or metrics.

3.3 Performance Measurements

Some statistics are used on the estimation results produced by the models to measure the model performance and to determine which of the generated models produces better results. In the performance evaluation of the estimation results of the model, MAPE (Mean Absolute Percentage Error), and RMSE (Root Mean Squared Error) equations are used. As the MAPE (Eq. 7), and RMSE (Eq. 8) results get closer to 0, it indicates that the model performs well, and on the contrary, the model performs poorly. Choosing which evaluation metric to use may vary depending on the problem. For regression problems, RMSE and MAPE are the most used performance evaluation tools [46]– [48].

For example, Results can be classified according to MAPE value. Results and its classifications are given below as:

- it means “very good” if the result less than 10%,
- it means “good” if the result is between 10% and 20%,
- it means “moderate” if the result is between 20% and 50%,
- it means “false” or “incorrect” if the result is more than 50% [49].

$$MAPE = \left(\frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \right) * 100 \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \tag{8}$$

In the above equations Y , \hat{Y} and n indicate the actual value, predicted value, and total number of values, respectively.

4. Experimental Results

Same LSTM (Long Short-Term Memory) model trained and tested on each dataset created. Python was used as programming language. In this study, very simple LSTM model was used. LSTM model architecture consists of 3 layers in total: LSTM with 10 units, drop out and dense layers. Sigmoid function was used as an activation function. Some training data parameters are as follows: 10 batch size, 100 epochs, Adam optimizer and binary cross entropy as lost function. If the training loss of the model did not change during the 10 epochs, the model training process was stopped. Dataset was split into 70% as training and 30% as a testing.

Data Set-1 and Data Set-3 consist of 365 daily data,

for Data Set-2 and Data Set-4 consist of 257 daily data after subtracting holidays. LSTM performance results on 4 data set are shown in Table 2.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 10)	640
dropout (Dropout)	(None, 10)	0
dense (Dense)	(None, 1)	11

Figure 1 LSTM architecture

Table 2 Datasets total values

Datasets	MAPE	RMSE
Data Set-1	0.207	22.884
Data Set-2	0.157	15.661
Data Set-3	00.00-08.00	0.529
	08.00-16.00	0.33
	16.00-00.00	0.163
Data Set-4	00.00-08.00	0.513
	08.00-16.00	0.28
	16.00-00.00	0.147

Considering the results shown in Table 3, if the raw data set (Data Set-1) is used without any modification, the MAPE (Mean Absolute Percentage Error) value is 20% and the RMSE (Root Mean Squared Error) value is 22.88. In this case, performance can be described as “medium”. It is seen that the MAPE and RMSE results of the Data Set-2, which was modified by subtracting the determined days, are 15% and 15.66, respectively. According to the results, Data Set-2 performance can be considered in the “good” category. There is a difference of about 5% between the two dataset performances.

Table 3 Performance results according to triage of Data Set-1 and Data Set-2

Triage Group	Data Set-1		Data Set-2	
	MAPE	RMSE	MAPE	RMSE
Red	-	0.419	-	0.393
Yellow	0.658	6.783	0.756	6.911
Green	0.223	21.739	0.159	15.177
Sum	0.207	22.884	0.157	15.661

Data Set-3 and Data Set-4 were obtained by separating Data Set-1 and Data Set-2 values according to 8-hour time periods and triage sections,

respectively. Performance results for Data Set-3 and Dataset-4 are shown in Table 4 and Table 5, respectively.

Table 4 Performance results according to triage of Data Set-3

		Red	Yellow	Green	Sum
00.00-08.00	MAPE	-	-	0.527	0.529
08.00-16.00	RMSE	0.193	2.034	4.023	5.031
16.00-00.00	MAPE	-	-	0.552	0.330
00.00-08.00	RMSE	0.350	4.335	14.249	13.960
08.00-16.00	MAPE	-	0.919	0.185	0.163
16.00-00.00	RMSE	0.151	3.804	11.098	11.164

MAPE value in Table-4 and Table5 couldn't be calculated for the "Yellow Triage" between 00.00-08.00 and 08.00-16.00 and for all time periods "Red Triage". The reason for this is that in the MAPE equation (Equation 8) if actual value is 0, division by zero cannot be performed. Model performance on Dataset-2 is seen as 2% more performant in general. When the "Green Triage" field MAPE and RMSE values are examined, it is seen that model performs much better.

Table 5 Performance results according to triage of Data Set-4

		Red	Yellow	Green	Sum
00.00-08.00	MAPE	-	-	0.539	0.513
08.00-16.00	RMSE	0.201	1.889	4.050	4.864
16.00-00.00	MAPE	-	-	0.392	0.280
00.00-08.00	RMSE	0.300	3.999	9.038	10.531
08.00-16.00	MAPE	-	0.929	0.150	0.147
16.00-00.00	RMSE	0.177	3.816	8.663	9.280

5. Conclusions

Time series are frequently used in the field of economics and data cleaning forms [50]–[53] prepared by considering many factors such as trends and seasonal effects are used and constantly improved. However, in the health sector, where many disciplines work intertwined and coordinated, there are some studies [19] that focus on seasonal effects in the field of emergency service demand forecasting, but there are not enough studies on triage or time zones in general. In the study, it has been shown that cleaning methods on time series data sets contribute positively to the performance of the model, as well as the selection of the methods or methods to be applied

on the time series. As stated in the study, emergency service admission criteria may vary according to country. Researchers can work on creating cleaning and purification forms by country in future studies. In addition, future works can be focused on to help determine general and special criteria in emergency patient admission procedures.

References

- [1] "Başbakanlık Mevzuatı Geliştirme ve Yayın Genel Müdürlüğü." <https://www.resmigazete.gov.tr/eskiler/2009/10/20091016-16.htm> (accessed May 14, 2021).
- [2] T. R. M. Azeredo, H. M. Guedes, R. A. Rebelo de Almeida, T. C. M. Chianca, and J. C. A. Martins, "Efficacy of the Manchester Triage System: a systematic review," *International Emergency Nursing*, vol. 23, no. 2, pp. 47–52, Apr. 2015, doi: 10.1016/j.ienj.2014.06.001.
- [3] Ö. Serper and N. Gürsakal, *Araştırma Yöntemleri Üzerine*, Filiz Kitabevi. İstanbul, 1983.
- [4] K. Gürtan, *İstatistik ve araştırma metotları*. İstanbul Üniversitesi, 1971.
- [5] F. Serin, Y. Alisan, and A. Kece, "Hybrid time series forecasting methods for travel time prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 579, p. 126134, Oct. 2021, doi: 10.1016/j.physa.2021.126134.
- [6] F. Serin, Y. Alisan, and M. Erturkler, "Predicting Bus Travel Time Using Machine Learning Methods with Three-Layer Architecture," *Measurement*, p. 111403, May 2022, doi: 10.1016/j.measurement.2022.111403.
- [7] E. Özkan, E. Güler, and Z. Aladağ, "Elektrik Enerjisi Tüketim Verileri İçin Uygun Tahmin Yöntemi Seçimi," *Endüstri Mühendisliği*, vol. 31, no. 2, pp. 198–214, 2020.
- [8] G. Altınay, "Aylık elektrik talebinin mevsimsel model ile orta dönem öngörüsü," 2010.
- [9] O. Çoban and C. C. Özcan, "Sektörel Açidan Enerjinin Artan Önemi: Konya İli İçin Bir Doğalgaz Talep Tahmini Denemesi," *Sosyal Ekonomik Araştırmalar Dergisi*, vol. 11, no. 22, pp. 85–106, 2011.
- [10] Yucesan, M., Pekel, E., Celik, E., Gul, M., & Serin, F. (2021). Forecasting daily natural gas consumption with regression, time series and machine learning based methods. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 1-16.
- [11] O. Kaynar, S. Taştan, and F. Demirkoparan, "Yapay sinir ağları ile doğalgaz tüketim tahmini," *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 25, 2011.
- [12] M. Akdağ and V. Yiğit, "Box-Jenkins Ve Yapay Sinir Ağı Modelleri İle Enflasyon Tahmini," *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 30, no. 2, 2016.
- [13] A. Bekin, "Türkiye'de bazı temel gıda fiyatları için

- yapay sinir ağları ve zaman serisi tahmin modellerinin karşılaştırmalı analizi,” Master’s Thesis, 2015.
- [14] B. Ataseven, “Yapay sinir ağları ile öngörü modellemesi,” *Öneri Dergisi*, vol. 10, no. 39, pp. 101–115, 2013.
- [15] M. Yucesan, M. Gul, and E. Celik, “Performance comparison between ARIMAX, ANN and ARIMAX-ANN hybridization in sales forecasting for furniture industry,” *Drvna industrija: Znanstveni časopis za pitanja drvne tehnologije*, vol. 69, no. 4, pp. 357–370, 2018.
- [16] G. Zhang, X. Zhang, and H. Feng, “Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion,” *Expert Systems*, vol. 33, no. 5, pp. 501–516, 2016.
- [17] C. Yuan, S. Liu, and Z. Fang, “Comparison of China’s primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model,” *Energy*, vol. 100, pp. 384–390, 2016.
- [18] R. Tripathi et al., “Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models,” *Advances in Agriculture*, vol. 2014, 2014.
- [19] G. Sariyer, “Acil Servislerde Talebin Zaman Serileri Modelleri ile Tahmin Edilmesi,” *International Journal of Engineering Research and Development*, vol. 10, no. 1, Art. no. 1, Jan. 2017, doi: 10.29137/umagd.419661.
- [20] F. Serin, A. Keçe and Y. Alişan, "Applying Machine Learning Prediction Methods to COVID-19 Data", *Journal of Soft Computing and Artificial Intelligence*, vol. 3, no. 1, pp. 11-21, doi:10.55195/jscai.1108528
- [21] V. Demir, M. Zontul, and İ. Yelmen, “Drug Sales Prediction with ACF and PACF Supported ARIMA Method,” in *2020 5th International Conference on Computer Science and Engineering (UBMK)*, Sep. 2020, pp. 243–247. doi: 10.1109/UBMK50275.2020.9219448.
- [22] E. Aydemir, M. Karaatli, G. Yilmaz, and S. Aksoy, “112 Acil çağrı merkezine gelen çağrı sayılarını belirleyebilmek için bir yapay sinir ağları tahminleme modeli geliştirilmesi,” *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 20, no. 5, pp. 145–149, 2014.
- [23] A. Sevda, “Türkiye’de Reçete Başına Ortalama Maliyet Serisinin Zaman Serisi Modelleriyle Öngörüsü ve Öngörü Performanslarının Karşılaştırılması,” *SGD-Sosyal Güvenlik Dergisi*, vol. 4, no. 2, pp. 176–192.
- [24] A. G. Özüdoğru and A. Görener, “Sağlık sektöründe talep tahmini üzerine bir uygulama,” 2015.
- [25] Kaya, Y., Kuncan, F. & Tekin, R. A New Approach for Congestive Heart Failure and Arrhythmia Classification Using Angle Transformation with LSTM. *Arab J Sci Eng* (2022). <https://doi.org/10.1007/s13369-022-06617-8>
- [26] V. Yiğit, “Hastanelerde tıbbi malzeme talep tahmini: Serum seti tüketimi üzerinde örnek bir uygulama,” *Manas Sosyal Araştırmalar Dergisi*, vol. 5, no. 4, pp. 207–222, 2016.
- [27] Ortiz-Barrios, M., Gul, M., Yucesan, M., Alfaro-Sarmiento, I., Navarro-Jiménez, E., & Jiménez-Delgado, G. (2022). A fuzzy hybrid decision-making framework for increasing the hospital disaster preparedness: The colombian case. *International journal of disaster risk reduction*, 72, 102831.
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. Accessed: Apr. 27, 2020. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [29] I. J. Goodfellow et al., “Generative Adversarial Networks,” arXiv:1406.2661 [cs, stat], Jun. 2014, Accessed: May 13, 2021. [Online]. Available: <http://arxiv.org/abs/1406.2661>
- [30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” arXiv:1512.00567 [cs], Dec. 2015, Accessed: May 13, 2021. [Online]. Available: <http://arxiv.org/abs/1512.00567>
- [31] J. Cao, Z. Li, and J. Li, “Financial time series forecasting model based on CEEMDAN and LSTM,” *Physica A: Statistical Mechanics and its Applications*, vol. 519, pp. 127–139, Apr. 2019, doi: 10.1016/j.physa.2018.11.061.
- [32] A. Sagheer and M. Kotb, “Time series forecasting of petroleum production using deep LSTM recurrent networks,” *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019, doi: 10.1016/j.neucom.2018.09.082.
- [33] S. Kaushik et al., “AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures,” *Front. Big Data*, vol. 3, 2020, doi: 10.3389/fdata.2020.00004.
- [34] V. K. R. Chimmula and L. Zhang, “Time series forecasting of COVID-19 transmission in Canada using LSTM networks,” *Chaos, Solitons & Fractals*, vol. 135, p. 109864, Jun. 2020, doi: 10.1016/j.chaos.2020.109864.
- [35] F. Kadri, F. Harrou, S. Chaabane, and C. Tahon, “Time series modelling and forecasting of emergency department overcrowding,” *Journal of medical systems*, vol. 38, no. 9, p. 107, 2014.
- [36] S. Khadanga, K. Aggarwal, S. Joty, and J. Srivastava, “Using Clinical Notes with Time Series Data for ICU Management,” arXiv:1909.09702 [cs, stat], Jan. 2020, Accessed: May 13, 2021. [Online]. Available: <http://arxiv.org/abs/1909.09702>
- [37] Z. C. Lipton, D. C. Kale, and R. C. Wetzel, “Phenotyping of Clinical Time Series with LSTM Recurrent Neural Networks,” arXiv:1510.07641 [cs], Mar. 2017, Accessed: May 13, 2021. [Online]. Available: <http://arxiv.org/abs/1510.07641>
- [38] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke and J. Schmidhuber “A Novel Connectionist

- System for Unconstrained Handwriting Recognition,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 855-868, May 2009, doi: 10.1109/TPAMI.2008.137.
- [39] V. Märgner and H. E. Abed, “ICDAR 2009 Arabic Handwriting Recognition Competition,” in 2009 10th International Conference on Document Analysis and Recognition, Jul. 2009, pp. 1383–1387. doi: 10.1109/ICDAR.2009.256.
- [40] C. Metz, “An Infusion of AI Makes Google Translate More Powerful Than Ever,” Wired. Accessed: May 13, 2021. [Online]. Available: <https://www.wired.com/2016/09/google-claims-ai-breakthrough-machine-translation/>
- [41] S. Siami-Namini and A. S. Namin, “Forecasting Economics and Financial Time Series: ARIMA vs. LSTM,” arXiv:1803.06386 [cs, q-fin, stat], Mar. 2018, Accessed: May 13, 2021. [Online]. Available: <http://arxiv.org/abs/1803.06386>
- [42] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [43] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” IEEE Trans. Neural Netw. Learning Syst., vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [44] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: continual prediction with LSTM,” pp. 850–855, Jan. 1999, doi: 10.1049/cp:19991218.
- [45] F. A. Gers, D. Eck, and J. Schmidhuber, “Applying LSTM to time series predictable through time-window approaches,” in Neural Nets WIRN Vietri-01, Springer, 2002, pp. 193–200.
- [46] N. Ahmadi, T. G. Constandinou, and C.-S. Bouganis, “Decoding hand kinematics from local field potentials using long short-term memory (LSTM) network,” in 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), 2019, pp. 415–419.
- [47] Ç. Elmas, “Yapay sinir ağları,” Seçkin Yayıncılık:2003.
- [48] E. Ulucan and İ. KIZILIRMAK, “Konaklama işletmelerinde talep tahmin yöntemleri: Yapay sinir ağları ile ilgili bir araştırma,” Seyahat ve Otel İşletmeciliği Dergisi, vol. 15, no. 1, pp. 89–101, 2018.
- [49] S. Şahin And D. Kocadağ, “Sağlık Sektöründe Talep Tahmini Üzerine Literatür Araştırması,” Düzce Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, vol. 10, no. 1, pp. 99–113.
- [50] C. D. Lewis, Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. London; Boston: Butterworth Scientific, 1982.
- [51] H. Kutbay, “Kayıt dışı istihdamın vergi gelirleri üzerindeki etkisi,” Sosyal Bilimler Araştırma Dergisi, vol. 7, no. 2, pp. 172–189, 2018.
- [52] F. Y. Contuk and B. Güngör, “Finansal Piyasaların Gelişmesinin Ekonomik Büyüme Üzerine Etkileri: 1998-2014 Türkiye Örneği,” Finans Politik ve Ekonomik Yorumlar, no. 611, pp. 15–27, 2016.
- [53] U. SİVRİ and B. SEVEN, “Ortalama Tüketim Eğilimi Durağan Mıdır? Türkiye Ekonomisi için Bir Zaman Serisi Analizi,” Anadolu İktisat ve İşletme Dergisi, vol. 1, no. 1, pp. 50–65, 2017.