Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

FUJECE

Derin Öğrenme ve Özellik Seçimi Yaklaşımları Kullanılarak Twitter Verilerinden COVID-19 Aşı Karşıtlığı Tespiti

Serdar ERTEM [1], Erdal ÖZBAY [2*]

[1,2]Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye.
[1]serdarertem01@gmail.com, [2]erdalozbay@firat.edu.tr

Geliş Tarihi: 27.02.2024
Kabul Tarihi: 29.03.2024

Düzeltme Tarihi: 19.03.2024

doi: https://doi.org/10.62520/fujece.1443753
Araştırma Makalesi

Alıntı: S. Ertem ve E. Özbay, "Derin öğrenme ve özellik seçimi yaklaşımları kullanılarak twitter verilerinden covid-19 aşı karşıtlığı tespiti", Fırat Üni. Deny. ve Hes. Müh. Derg., vol. 3, no 2, pp. 116-133, Haziran 2024.

Öz

COVID-19 pandemisi, dünya genelinde sağlık, ekonomi ve toplumsal yaşamı derinden etkileyen bir krize dönüşmüştür. Bu kriz sırasında aşı karşıtlığı, salgının kontrolü ve aşılama kampanyalarının etkinliği açısından önemli bir engel oluşturmaktadır. Bu çalışmada, derin öğrenme ve özellik seçimi yaklaşımlarının birleşimi kullanılarak COVID-19 aşı karşıtlığının twitter verilerinden tespiti amaçlanmıştır. Önerilen yöntem, derin öğrenme modeli ile özellik seçimi tekniklerinin entegrasyonunu içermekte ve metin verilerindeki önemli özellikleri belirleyerek aşı karşıtlığını tanımlamaktadır. Özellik çıkarımı için TF-IDF ve N-gram yöntemleri hibrit kullanılmış, ardından Ki-kare özellik seçimi gerçekleştirilmiştir. Veri seti twitter text verilerinden ve iki etiketten oluşmaktadır. Etiketlerin dengelenmesi için Sentetik Azınlık Aşırı Örnekleme Tekniği (SAAÖT) yöntemi uygulanmıştır. Sınıflandırma işlemi için derin öğrenme mimarilerinden Uzun Kısa-Süreli Bellek (UKSB) kullanılmıştır. Önerilen özellik çıkarımı, özellik seçimi ve UKSB yöntemlerinin birlikte kullanılmasıyla elde edilen deneysel sonuçlara göre %99.23 ile en yüksek doğruluk değerine ulaşılmıştır. Bu sonuçlar, COVID-19 aşı karşıtlığının metin verileri üzerinde etkili bir şekilde tespit edilmesi için önerilen yöntemlerin başarılı bir şekilde kullanılabileceğini göstermektedir. Çalışmanın sonuçları, sağlık politikalarının ve kamuoyu bilgilendirme stratejilerinin geliştirilmesine yönelik değerli bilgiler sunabilmektedir. Bu bakımdan, aşılama kampanyaları ve halk sağlığı müdahaleleri planlanırken, aşı karşıtlığı belirlemede yeni ve güçlü bir araç geliştirilmiştir.

Anahtar kelimeler: COVID-19 aşı-karşıtlığı, Derin öğrenme, Özellik seçimi, UKSB, Metin sınıflandırma

*Yazışılan yazar

İntihal Kontrol: Evet – Turnitin
Şikayet: fujece@firat.edu.tr

Firat Univ Jour. of Exp. and Comp. Eng., 3(2), 116-133, 2024

Firat University Journal of Experimental and Computational Engineering

Detection of COVID-19 Anti-Vaccination from Twitter Data Using Deep Learning and Feature Selection Approaches

Serdar ERTEM [1], Erdal ÖZBAY [2*]

[1,2]Computer Engineering Department, Faculty of Engineering, Firat University, 23119, Elazig, Türkiye.
[1]serdarertem01@gmail.com, [2]erdalozbay@firat.edu.tr

Received: 27.02.2024
Accepted: 29.03.2024
Revision:19.03.2024
doi: https://doi.org/10.62520/fujece.1443753
Research Article

Citation: S. Ertem and E. Özbay, "Detection of covid-19 anti-vaccination from twitter data using deep learning and feature selection approaches", Firat Univ. Jour.of Exper. and Comp. Eng., vol. 3, no 2, pp. 116-133, June 2024.

Abstract

The COVID-19 pandemic has evolved into a crisis significantly impacting health, the economy, and social life worldwide. During this crisis, anti-vaccination sentiment poses a considerable obstacle to controlling the epidemic and the effectiveness of vaccination campaigns. This study aimed to detect COVID-19 anti-vaccination sentiment from Twitter data using a combination of deep learning and feature selection approaches. The proposed method integrates a deep learning model with feature selection techniques to identify anti-vaccination sentiment by pinpointing important features in text data. Hybrid TF-IDF and N-gram methods were utilized for feature extraction, followed by Chi-square feature selection. The dataset comprises Twitter text data and two labels. The Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the labels. Long Short-Term Memory (LSTM), a deep learning architecture, was employed for the classification process. The experimental results, obtained by leveraging the proposed feature extraction, feature selection, and LSTM methods, achieved the highest accuracy value of 99.23%. These findings demonstrate the proposed methods' success in effectively detecting COVID-19 anti-vaccination sentiment in text data. The study's results can offer valuable insights for developing health policies and public information strategies, presenting a new and powerful tool for detecting anti-vaccine sentiment in planning vaccination campaigns and public health interventions.

Keywords: COVID-19 anti-vaccine, Deep learning, Feature selection, LSTM, Text classification

*Corresponging author

1. Introduction

The COVID-19 pandemic has demonstrated the profound effects of the global health crisis. In this process, the existence of an important problem such as society's anti-vaccination complicates the control of the epidemic and vaccination efforts. The causes and prevalence of anti-vaccine sentiment have been the subject of extensive research. In this context, social media platforms have become an important platform for people to express their thoughts, concerns, and attitudes [1].

In this study, a dataset consisting of Twitter text data was used to detect COVID-19 anti-vaccination. Twitter is a suitable source for this type of analysis because it has a large user base and is a platform where real-time opinions can be shared. Analysis of the tweets that make up our data set with deep learning and feature selection approaches was used as a tool to understand and define anti-vaccine sentiment [2].

Text classification holds significant importance in the field of natural language processing (NLP) due to the ever-increasing size of textual data. Additionally, the automatic labeling of data plays a crucial role. In NLP, transforming words into vectors—thereby converting them into numerical values comprehensible to computers—is a critical task. The techniques employed to convert words into numerical values are pivotal in text classification processes, as they considerably influence the accuracy of the system (model) to be developed. Hence, the vectors generated by these techniques should accurately represent the words, with the goal of producing vectors that enhance classification success. Improving the quality of vectors not only aids in representing words more accurately but also enriches the multifaceted relationships between words, contributing to better model performance [3].

Deep learning and machine learning approaches have been utilized to identify the negative sentiments associated with COVID-19 vaccines on social media. Typically, researchers categorize tweets concerning COVID-19 vaccines into two principal groups: those expressing negative sentiments about vaccines and those unrelated to such sentiments. Valid tweets are further classified into three subcategories: personal experiences, informational content, and advisory messages. In a particular study, four machine learning models were trained using the collected data: Support Vector Machine (SVM), Logistic Regression (LR), Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANN). The LSTM model achieved the highest accuracy, with a rate of 97.64%. It was observed that the SVM model presented the lowest accuracy rate, at 80%, for the research conducted. Additionally, the researchers evaluated the performance of various natural language processing models in detecting anti-vaccine sentiments during the COVID-19 pandemic. They discovered that the BERT (Bidirectional Encoder Representations from Transformers) model surpassed the other models in terms of performance on the test set [5].

In this study, meaningful features were obtained from English tweets by using Term Frequency-Inverse Document Frequency (TF-IDF) and N-gram methods together for feature extraction from text data to detect anti-vaccination. Additionally, feature selection was made using the Chi-square method to select the most prominent among these features. Synthetic Minority Oversampling Technique (SMOTE) method was applied to balance the data set. Finally, tweets were classified using the Long Short Term Memory (LSTM) deep learning algorithm [6].

The contributions of this study are as follows:

- Different feature extraction techniques were applied to obtain meaningful features from text data.
- Feature selection was applied to the features extracted from text data using the Chi-square method.
- Using these features with the LSTM deep learning model, signs of anti-vaccination were detected.
- Experimental results showed that the combination of deep learning and feature selection methods was effective in detecting COVID-19 anti-vaccination.

The results of the study show the effectiveness of the proposed methods. The results obtained show that the detection of anti-vaccination has been successfully carried out with a high accuracy rate. This study highlights the importance of data-driven approaches to identifying and managing anti-vaccine sentiment.

2. Related Works

In recent years, vaccination efforts related to the COVID-19 pandemic have become one of the most important agenda items worldwide. However, anti-vaccine sentiment has been a factor that has significantly impacted these efforts. In the literature, research on anti-vaccination has increased and many studies have been conducted on this subject. These studies have created an important basis for understanding the causes, effects and spread mechanisms of anti-vaccination, and the current status of anti-vaccination in the literature has been examined.

Qorib et al. showed in their study that the combination of TextBlob + TF-IDF + LinearSVC achieved the best performance in classifying public sentiment as positive, neutral, or negative. This combination resulted in 96.75% accuracy, 96.92% sensitivity, 92.80% sensitivity, and 94.70% F1 score. Additionally, combining two vectorization methods such as CountVectorizer and TF-IDF has been found to reduce model accuracy [7].

In a study where an AI-based framework was developed to detect COVID-19 vaccine misinformation, the BERT model showed the best performance. In the evaluation performed on the test set, a 98% F1 score was obtained. These findings were found to show that artificial intelligence models are effective in detecting misinformation about COVID-19 vaccines on social media platforms [8].

Aygün et al. focused on understanding people's views on vaccines and vaccine types during the period when vaccine development and community vaccination studies were carried out during the COVID-19 pandemic. Researchers collected 928,402 different vaccine-focused tweets to examine Twitter users' attitudes towards vaccination in the US, UK, Canada, Turkey, France, Germany, Spain and Italy. Data sets were prepared in two different languages (English and Turkish), and 4 different directions and 4 different BERT models were used to classify the tweets. Sentiment analysis was conducted on the 6 most widely used COVID-19 vaccines and the results are presented by country. While the success of the method varies between 84% and 88% in terms of F1 Score, the total accuracy value is determined as 87%. This study represents the first attempt to understand community views on the vaccination process [9].

Çelik and Kaplan evaluated the classification success using balanced data in a data set containing 4203 SMS and according to the study results, the highest classification successes were obtained using Logistic Regression with SMOTE (80.1%), Condensed Nearest Neighbor with XGBoost (62.1%) and Random Undersampling Technique with Logistic Regression (73.8%) [10].

Avvaru et al. implemented various versions of LSTM models (LSTM, stacked LSTM, Bi-LSTM, and CNN-LSTM) and BERT and XLNet transformer models on Twitter and Reddit [11]. Özbay provided an effective solution to the problem of aggression detection in Twitter data using the transformer-based CNN model with the Bert model proposed in his study [12].

In recent years, researchers have been applying artificial intelligence techniques to predict diabetes. In this context, a new SMOTE-based deep LSTM system was developed to deal with the class imbalance in the diabetes dataset and its prediction accuracy was measured. In a study where CNN, CNN-LSTM, ConvLSTM, and deep 1D-convolutional neural network (DCNN) techniques were examined and a SMOTE-based deep LSTM method was proposed, it was stated that the proposed model achieved the highest prediction accuracy of 99.64% measured based on the diabetes data set. These results show that, based on classification accuracy, this method outperforms other methods [13].

Bhatti et al. proposed a solution to classify emails into four classes: fraud, suspicious, harassment, and normal. They used the LSTM deep learning approach with stratified sampling to identify email classes. Additionally, sampling methods were used to balance the input dataset. The proposed model achieved a classification accuracy of over 90% by stratified sampling alone and an accuracy of over 95% by applying data balance techniques on the dataset [14].

In another study, a classification was conducted using machine learning methods to classify comments into toxic categories for social media use. The method used in the study is TF-IDF as feature extraction and SVM with Chi-square as feature selection. Additionally, various exploratory scenarios were also carried out to achieve the best performance by applying SVM kernels and preprocessing stages [15].

In their study, Hussein and Ozyurt compared the performances of LSTM, Bi-LSTM and GRU models and the Chi-square feature selection method for sentiment analysis. As a result of experiments conducted on two scaled data sets such as Yelp and US Airways, it has been observed that feature selection methods significantly increase classification accuracy. On the Yelp dataset, the Bi-LSTM model achieved 100% accuracy with Chi-square when using 500 features. In the US Airlines dataset, the GRU-LSTM model achieved 97.9% accuracy with Chi-square when 20 features were used [16].

In another study, CNN and LSTM models were specifically used to classify Spam and non-Spam text messages. The proposed models relied solely on text data and extracted the feature set automatically. 99.44% accuracy was achieved on a reference dataset consisting of 747 Spam and 4,827 non-Spam text messages [17].

In their study, Zhang & Rao presented a method to accurately and quickly perform text classification, which is commonly found in fields such as e-commerce and daily message analysis, with two methods: one-by-one comparison and one-to-one classification. The approach, called "n-BiLSTM," is used to transform natural language text sentences into similar features with N-gram techniques, and these features are then fed into a bidirectional LSTM. The results of the study are presented for n = 1, 2, 3 and when n = 2, they achieved the best result with 88.8% [18].

Alfarizi et al.'s aim in their study is to compare the classification method with the LSTM model by adding the word-weighted TF-IDF and LinearSVC model to increase the accuracy in determining sentiment. The dataset used is 18000 divided into 16000 training data along with 2000 test data with 6 emotion classes namely sadness, anger, fear, love, joy and surprise. While the classification accuracy of emotions using the LSTM method was 97.50%, an accuracy value of 89% was obtained using the LinearSVC method [19].

3. Material and Method

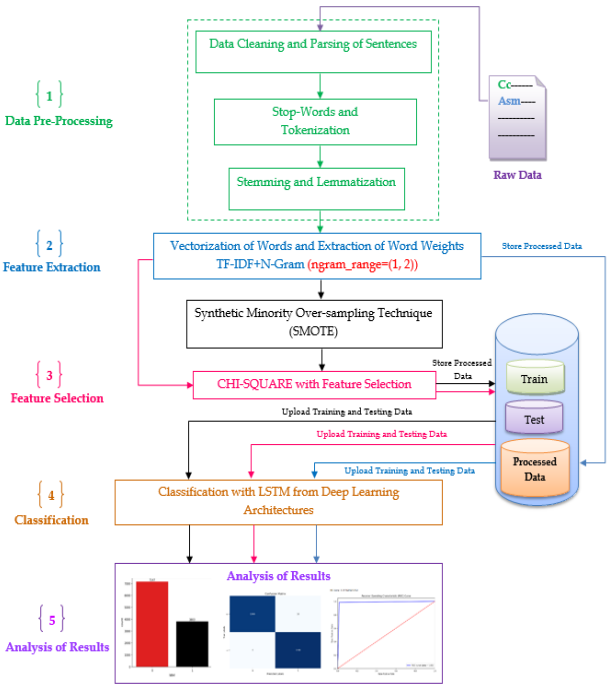

Figure 1. Flow diagram of the proposed methodology

Recently, anti-vaccination against COVID-19 has become a growing source of concern. The strategy of the model and method proposed in this study, which aims to analyze the trends and prevalence of anti-vaccine sentiment by using text data spread on social media platforms such as Twitter, is shown in Figure 1.

3.1. Dataset Construction

Twitter is one of the most popular social media platforms with 353 million active users and more than 500 million tweets are shared every day. Additionally, the Twitter API is allowed to pull public tweets, including tweet text, user information, retweets, and mentions, in JSON format, and a Python library called Twarc is used for this purpose. Specific keywords are used to collect relevant tweets about COVID-19 vaccines, and only English-language tweets are taken into account. Additionally, replies to tweets, retweets, and quoted tweets are ignored. In total, a total of 15,465,687 vaccine-related tweets were collected from December 1, 2020, to July 31, 2021 [8]. In accordance with Twitter's terms of service, the dataset is anonymized, contains only tweet IDs, and is publicly shared with users via the GitHub platform.

A function called get_tweet_by_id is defined to retrieve the content of a particular tweet and an associated reply using the Twitter API. By obtaining the tweet ID and response from GitHub, this function uses Twitter's oEmbed API to extract the tweet content. The content of this tweet was later removed. Then, tweet contents were retrieved from a specific tweet list by using the get_tweet_by_id function in a loop and these contents were added to a DataFrame. Any missing data is then dropped from this data frame. The number of missing or None values was checked and the cleaned data was saved in a CSV file.

A data set was created by creating a code block and extracting English tweet contents and their responses from the tweet ID list. The raw data set consists of 10950 rows and 2 columns after the specified cleaning process. The columns are named "text" and "label". While 7147 of the lines in the text column are in the label '0', 3803 of them are in the label '1'. As seen in Figure 2, in the "label" column of the data set used in the study, those who are not anti-vaccine are labeled as '0' and those who are anti-vaccine are labeled as '1'.

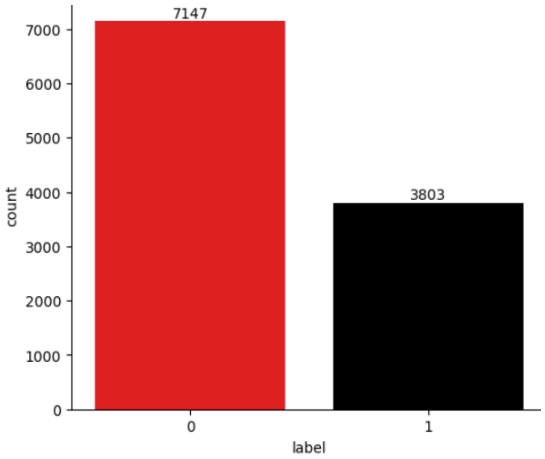

Figure 2. Distribution of the dataset via the label column

3.2. Data Pre-processing

Data preprocessing is an important stage that includes cleaning and editing steps on text data [20]. These steps ensure that text data becomes analyzable, consistent, and meaningful. More detailed explanations of the steps used in the data preprocessing phase are as follows:

3.2.1. Data Cleaning and Parsing of Sentences

Text cleaning involves removing noise (e.g., special characters, numbers, etc.) from the data. Text normalization involves bringing texts into a certain format, such as case conversion.

Conversion to Lowercase: All characters in the text were converted to lowercase, eliminating the distinction between uppercase and lowercase letters. In this way, "COVID" and "covid" are treated as the same word.

Removing Hashtags: Hashtags (#) encountered in these texts taken from the Twitter platform have been removed from the text. Hashtags often contain additional information on the topic and can be distracting in analysis.

Removing URLs: Web addresses (URLs) contained within the text are unnecessary to understand the content of the text and have therefore been removed. Removing URLs helps clean up the text and make it analyzable.

Removing Special Characters: Special characters, markings, or punctuation marks in the text are removed. This step increases the homogeneity of the text and simplifies the analysis process.

Removing Usernames: Removed usernames (@username) in the text. Usernames are unnecessary to understand the content of the text and can be misleading in analysis.

Remove Double Spaces: Double spaces within the text have been replaced with single spaces. This step preserves the order of the text and ensures consistency in subsequent processing.

Removing Numbers and Numbers: Numbers in the text have been removed. Since text analysis generally focuses on textual content, it is desired to reduce the effect of numbers in the text.

Removing Single Letters: Single letters (\b\w\b) that do not make sense on their own have been removed. For example, single letters such as "a" or "I" generally do not change the meaning of the text and can therefore be removed.

Removing Emojis: Emojis in the text were removed by converting them to ASCII characters. This step is implemented considering that emojis may cause undesirable effects in the analysis.

3.2.2. Stop-words and Tokenization

This step involves removing words from the text that do not have meaning or are not necessary for analysis. A stop_words set was used to extract English stop words. Tokenization is used to split the text into pieces or "tokens". For example, breaking sentences into words. The text is separated into tokens by a simple space-splitting process.

3.2.3. Stemming and Lemmatization

This step involves finding the root of the word. For example, the word "keeping" is reduced to its root as "keep". Lemmatization was performed using WordNetLemmatizer. By completing these steps, text data becomes ready for analysis and more robust results are obtained. Therefore, raw data turns into processed data.

3.3. Feature Extraction

Text classification is a widely used technique in the field of natural language processing and allows the assignment of a text to a specific category or class. The feature extraction step is very important to carry out this process effectively. Because extracting meaningful features from text data directly affects text classification performance. Frequently used feature extraction techniques include TF-IDF and N-gram models. These methods are word embedding or word vector methods used to represent texts, in other words, to digitize texts.

3.3.1. TF-IDF

TF-IDF is a numerical statistical method that reflects the importance of a term in a document relative to the entire collection of documents. It consists of two main components.

Term Frequency (TF): Measures how often a term occurs in a document. It calculates the number of times a term appears in the document by dividing it by the total number of terms in the document. This indicates the uniqueness of a term in the document. Example: If the term "covid" appears 5 times in a document containing 100 words, the term frequency of the term "covid" will be 5/100 = 0.05.

Inverse Document Frequency (IDF): Measures the rarity of a term within the entire document collection. It is calculated by taking the logarithm of the ratio of the number of documents in which a term occurs to the total number of documents. This determines the importance of a term in the entire corpus. Example: If there are 1,000,000 documents in the collection and the term "covid" appears in 1000 documents, the inverse document frequency of the term "covid" will be log(1,000,000/1,000) = 3.

After TF and IDF are calculated, the TF-IDF value is obtained. This gives higher weights to terms that occur frequently in the document but are rare in the entire collection.

3.3.2. N-gram

N-grams are sequences of n consecutive elements (words in text processing). It is used to capture the local structure of a text by taking into account the relationship between neighboring words.

Uni-Gram (1-Gram): Single words in the text. For example: “covid”, “vaccine”, “detected”.

Bi-Gram (2-Gram): Two consecutive word pairs. For example: “covid vaccine”, “vaccine detected”.

Tri-Gram (3-Gram): Three consecutive words. For example: “covid vaccine detected”.

N-grams are used to better capture word orders and contexts in text data. It is valuable in tasks such as language modeling, machine translation, and text generation. These techniques are widely used in natural language processing tasks to extract meaningful features from text data and improve the performance of machine learning models. It can be useful to try different N-gram ranges, especially when you want to examine language structure and expression patterns. However, higher N-gram ranges may result in greater dimensionality and thus a larger feature space, which may increase the training time of the model. Therefore, a balanced approach is important when choosing the N-gram range.

In this study, pre-processed text data is converted into a high-dimensional digital representation (vectorization) by applying the TF-IDF method. Here, each dimension represents a unique term or combination of terms. In addition, the range of n-grams used in the study was determined as (1, 2). Here, in addition to uni-grams, consecutive words, i.e. bi-grams, are also used. This means that features containing each term and pairs of consecutive words will be created. These transformed feature vectors can be fed into deep-learning models for tasks such as text classification, clustering, or information retrieval.

3.4. SMOTE

SMOTE is a technique used to address the issue of class imbalance when working with unbalanced datasets. It is used to increase performance, especially in classification problems, when the number of examples of the rare class is low [21]. SMOTE synthetically increases the minority class samples by creating random points among existing data points. This increases the representation of the minority class while helping the model generalize better against that class. Using SMOTE can reduce the tendency of the model to misclassify the minority class due to class imbalance, and ultimately help achieve a more balanced classification model. In the data set used in the study, the number of tweets labeled ‘0’, that is, non-anti-vaccine, is 7147, while the number of tweets labeled ‘1’, that is, anti-vaccination, is 3803. SMOTE was used in the proposed model to

eliminate class imbalance. The number of samples labeled '0' and '1' was equalized, that is, the number of samples labeled '0' became 7147, and the number of samples labeled '1' became 7147. The total number of samples used in the proposed model was 14294.

3.5. Feature Selection

Feature selection, which has an important place among data analysis methods, especially in recent years, helps researchers extract meaningful information from complex data sets. In this study, the use of the Chi-square method for feature selection in detecting COVID-19 anti-vaccination was examined. Chi-square is considered a powerful tool to identify relationships and connections between variables in the data set. This study addresses the effectiveness and applicability of Chi-square in the feature selection process.

3.5.1. Chi-Square Test

Feature selection is a critical step that directly affects the performance of the text classification model. One of the important reasons why feature selection affects text classification is the determination of meaningful features. Texts often contain large and complex bodies of data. Feature selection can help the model identify meaningful features, reducing the impact of redundant or noisy features. This allows the model to work on more focused and effective features [22].

The use of Chi-square as a feature selection method played an important role in the study carried out for COVID-19 anti-vaccine detection. Initially, a hybrid of TF-IDF and N-gram was used. 10000 features were determined with the max_features parameter. In feature selection, the ngram_range parameter was selected as (1, 2). Chi-square method was used to improve the performance of the model and eliminate unnecessary features. According to Equation 1, Chi-square values are calculated to determine the relationship between each feature and the target, and the desired number of features with the highest Chi-square scores are selected.

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

here, x^2 represents the Chi-square statistic, O_i represents observed frequencies, E_i represents expected frequencies. Σ shows the sum of terms calculated and summed for all categories. In Equation 1, O_i refers to the observed frequency, that is, the actual frequency in the data, and E_i refers to the expected frequency. Expected frequencies are calculated to be equally distributed across categories if the two variables are independent. The chi-square statistic represents the sum of the squares of the difference between the observed and expected frequencies divided by the expected frequencies. This statistic determines how large the relationship between variables is. A higher Chi-square statistic value indicates a stronger relationship between variables. The Chi-square test is a method used for categorical feature selection and evaluates the significant difference between observed and expected values. This test is used to measure the strength of the relationship between variables and examines the difference between observed and expected frequencies to determine statistical significance [23].

The SelectKBest class was created to select the top 6000 features in the dataset using the chi2 statistic. This helps preserve important features while reducing the training time of the model and reduces the risk of overfitting. This selection process uses a statistical method to find the most significant relationships between features, allowing the model to generalize better. As a result of the analysis made with the Chi-square test, the most decisive $k = 6000$ feature with the highest accuracy was selected. This choice enabled the model to perform better and accurately detect anti-vaccination. These results highlight the effectiveness of Chi-square in the feature selection process and its importance for COVID-19 anti-vaccine detection.

3.6. LSTM Architecture

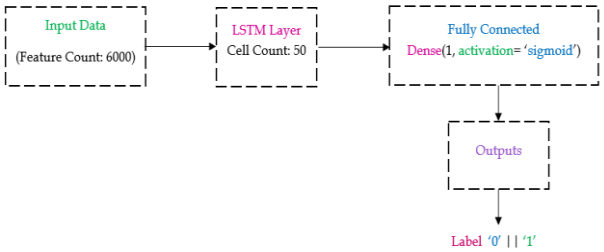

Figure 3. The structure of an LSTM model

In this study, LSTM deep learning architecture was used to classify pro-vaccine or anti-vaccine tweets about COVID-19 [24]. 10000 features were selected using TF-IDF and N-gram in training the model. SMOTE was used to balance the dataset. Then, 6000 features were selected using the Chi-square technique. The resulting balanced data set was classified with the LSTM model. The model evaluated its ability to detect anti-vaccine content by analyzing the textual content of tweets. The results demonstrated the ability of the LSTM model to identify anti-vaccine tweets with high accuracy and efficiency. The structure of an LSTM model is shown in Figure 3. The stages of creating the LSTM model used in the study are as follows:

Creating LSTM Layer: LSTM is used specifically for processing sequential data such as time series data. The LSTM layer is a type of RNN layer but with more advanced memory mechanisms than traditional RNNs. Different classification performances were evaluated as 20, 50, 70, and 100 cells in the LSTM layer. In the proposed model, an LSTM layer consisting of 50 cells is defined.

Creating the Fully Connected (Dense) Layer: Following the LSTM layer, one or more fully connected layers are usually added. This layer receives the outputs of the LSTM layer and is used to produce the desired outputs. In the proposed model, a fully connected layer with a single neuron is added. This neuron has a sigmoid activation function that is often used in binary classification problems. This ensures that the output falls between 0 and 1 and can be interpreted with probability values.

Compiling the Model: The compilation step determines the model's loss function, optimization algorithm, and optional metrics. Binary cross-entropy (binary_crossentropy) is used as the loss function of the model. This is a commonly used loss function for binary classification problems. "Adam" was chosen as the optimization algorithm. Adam is an effective optimization algorithm that accelerates training using the adaptive momentum method. Additionally, the accuracy metric is also monitored during training.

Training the Model: Training the model is performed over a certain number of iterations (epochs) on the data. During training, the epoch value was selected as 5. batch_size, the number of samples to be used in each training step is determined as 32. The study focuses on the classification of COVID-19 anti-vaccine tweets using the LSTM deep learning architecture, making a valuable contribution to solving an important problem in this field. Details and analysis of the results obtained with the feature selection and balancing techniques used will be presented. The training hyperparameters of the LSTM model are given in Table 1.

Table 1. Hyperparameters of the LSTM model

Model	Batch_size	Epochs	LSTM Cell
LSTM	32	5	50

4. Experimental Results and Discussion

4.1. Classification Performance Metrics

The study presents a review of commonly used metrics to evaluate the performance of classification models. Metrics used to evaluate the accuracy of the model in classification problems include accuracy, precision, sensitivity, and F1 score. The use of graphical methods such as the ROC curve is also examined. Complexity matrix-based metrics detail the impact of false positives and false negatives on model performance. Classification performance is evaluated through the metrics included in the complexity matrix shown in Table 2.

Table 2. Two-label confusion matrix

		Predicted Labels	
		Positive	Negative
True Labels	Positive	TP	FP
	Negative	FN	TN

Confusion matrix is a widely used tool to evaluate the performance of classification algorithms [25]. The key terms found in the confusion matrix are:

True Positive (TP): True positive. The number of true positive instances that the model correctly predicted as positive.

True Negative (TN): True negative. The number of true negative instances that the model correctly predicted as negative.

False Positive (FP): False positive. The number of instances in which the model incorrectly predicted a negative instance as positive.

False Negative (FN): False negative. The number of instances in which the model incorrectly predicted a positive instance as negative.

The following metrics were determined to evaluate classification performance using the confusion matrix and are listed below with their descriptions:

Accuracy: It is the ratio of correctly predicted samples to the total number of samples. It is a measure of overall model performance, but may not be sufficient on its own for unbalanced classes or datasets. It is calculated as given in Equation 2.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2}$$

Precision: The ratio of samples predicted to be positive to samples that are actually positive. It is important in situations aiming to reduce false positives (FP). It is calculated as in Equation 3.

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

Recall: The ratio of true positive samples to correctly predicted positive samples. It is important in situations aiming to reduce false negatives (FN). It is calculated as in Equation 4.

$$Recall = \frac{TP}{(TP + FN)} \tag{4}$$

F1 score: It is the harmonic mean of precision and sensitivity. It can be used as a substitute for accuracy for unbalanced classes or datasets. It is calculated as in Equation 5.

$$F1\ score = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{5}$$

ROC Curve: Receiver Operating Characteristic Curve and Area Under the Curve (AUC) It is a graphical tool used to evaluate the performance of classification algorithms. AUC refers to the area under the ROC curve, and the closer it is to 1, the better the performance of the model.

4.2. Performance Evaluation

First of all, after the Twitter data set was pre-processed with data cleaning and parsing of sentences, stop-words and tokenization, and stemming and lemmatization, feature extraction was performed with TF-IDF and N-gram methods and then classified using the LSTM deep learning algorithm. The results obtained show that the highest accuracy rate is 97.35%. The confusion matrix obtained at this stage is shown in Figure 4 and the ROC Curve is shown in Figure 5.

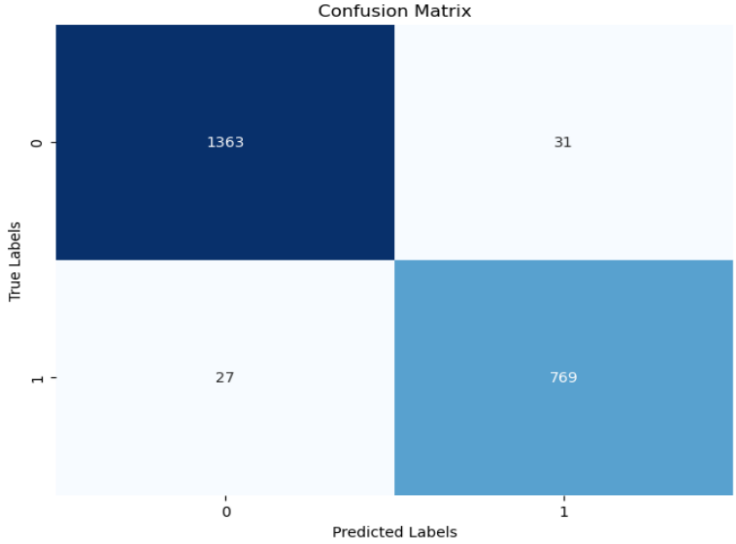

Figure 4. Confusion matrix of LSTM with TF-IDF and N-gram

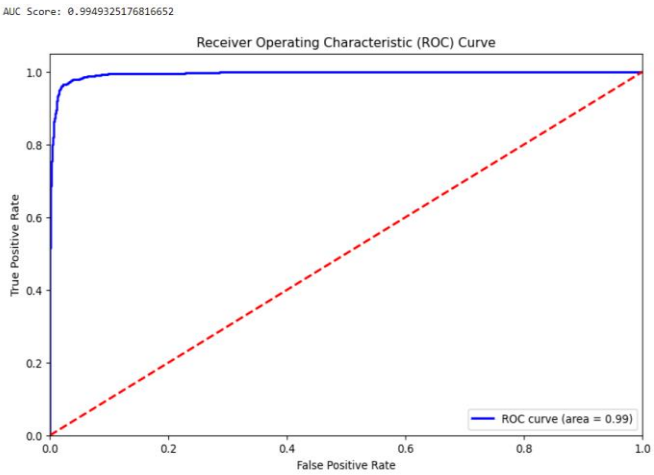

Figure 5. ROC curve of LSTM with TF-IDF and N-gram

In the second stage, feature extraction was made with TF-IDF and N-gram without using SMOTE, and the highest accuracy rate was determined as 98.36% according to the results obtained by performing feature selection with Chi-square. The confusion matrix obtained at the second stage is shown in Figure 6 and the ROC Curve is shown in Figure 7.

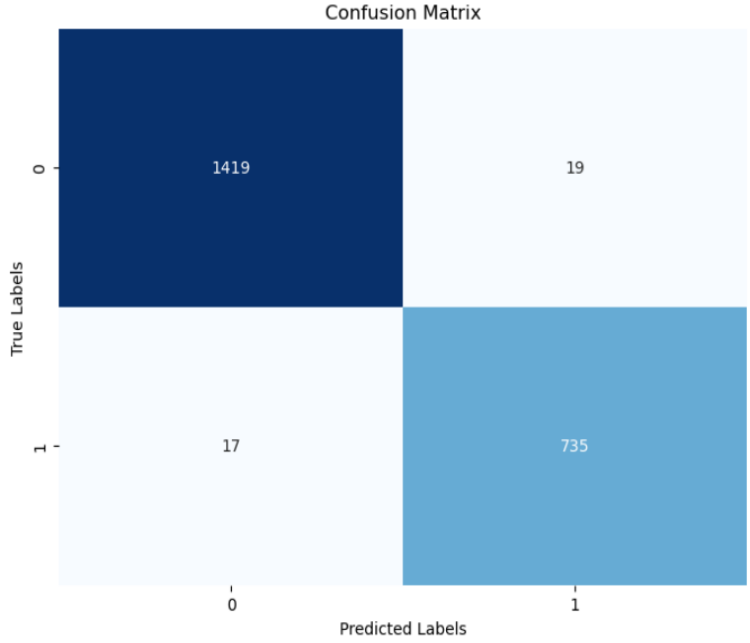

Figure 6. Confusion matrix of LSTM with TF-IDF, N-gram, and Chi-square feature selection

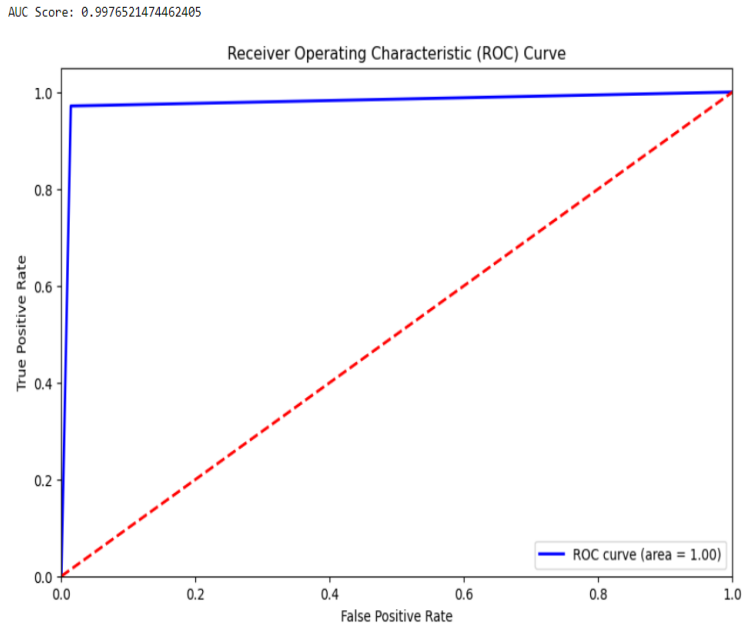

Figure 7. ROC curve of LSTM with TF-IDF, N-gram, and Chi-square feature selection

In the third stage, the data set was balanced using SMOTE, feature extraction was performed with TF-IDF and N-gram, and feature selection was performed with Chi-square, and according to the results obtained, the highest accuracy rate was determined as 99.23%. The confusion matrix obtained at the third stage is shown in Figure 8 and the ROC Curve is shown in Figure 9.

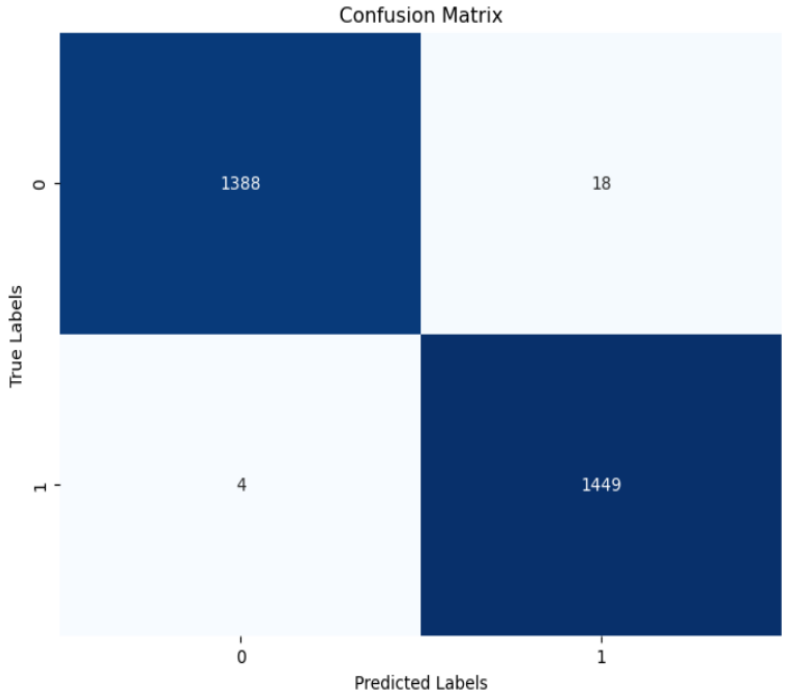

Figure 8. Confusion matrix of LSTM with TF-IDF, N-gram, Chi-square feature selection, and SMOTE

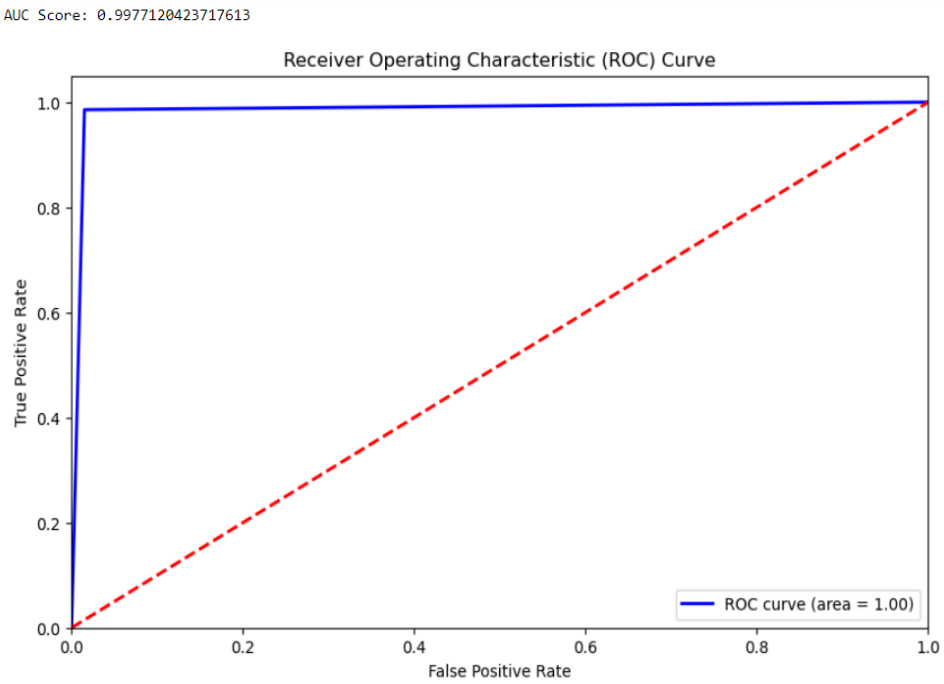

Figure 9. ROC curve of LSTM with TF-IDF, N-gram, Chi-square feature selection, and SMOTE

Table 3. Evaluation of the proposed method through performance metrics

Methods	Feature	Model	Accuracy	Precision	Recall	F1	ROC
TF-IDF + N-gram (1,2)	10000	LSTM	97.35%	97.8%	98.1%	97.9%	99.5%
TF-IDF + N-gram (1,2) + Chi-square	6000	LSTM	98.36%	98.7%	98.8%	98.8%	99.8%
TF-IDF + N-gram (1,2) + Chi-square + SMOTE (Proposed method)	6000	LSTM	99.23%	98.7%	99.7%	99.2%	99.8%

The results in Table 3 summarize a study evaluating the impact of feature selection and extraction on the performance of deep learning-based classification models. Initially, data obtained using a large feature set was processed with complex deep learning architectures such as LSTM. 10000 features derived using TF-IDF + N-Gram (1,2) initially yielded 97.35% accuracy. However, when the Chi-square method was applied to reduce the number of features and reduce the complexity of the model, 98.36% accuracy was achieved. At this point, remarkably, the feature selection and reduction process improved the overall performance of the model. 6000 features determined by chi-square appear to create a less complex model, which increases the accuracy rate. Moreover, 99.23% accuracy, a significant increase in the performance of the model, was observed when the SMOTE method was used to balance class labels in unbalanced datasets. This suggests that SMOTE, in particular, can help underrepresented classes to be better learned by the model, thereby increasing the precision and recall rate of the model. It has been shown that the proposed model achieves higher success in terms of other performance metrics, including Precision, Recall, F1 score, and ROC [26].

These results emphasize that feature selection and balancing methods can significantly affect the performance of deep learning models and that using appropriate methods can increase the reliability of the model. This work can be seen as an important step in the development of deep learning-based classification models, as choosing and balancing the right features can enable the model to generalize better to real-world data.

Table 4 compares the performance of existing SOTA methods and models in the literature for identifying anti-vaccine tweets. The latest methods used to identify anti-vaccine tweets demonstrate varying levels of success. Among these methods, the BERT method stands out, with Hayawi et al. [8] achieving an impressive 98% accuracy measured by the F1 score. Similarly, Kariyapperuma et al. [4] achieved a high accuracy of 97.6% using LSTM. However, the innovative approach developed in this study achieved a result that exceeds the success of existing methods with 99.23% accuracy and 99.21% F1 score. An impressive 99.23% accuracy result was achieved with this model, which is a combination of TF-IDF + N-Gram (1,2) + Chi-square + SMOTE and the deep learning-based LSTM model. In addition to establishing a new reference point in the field of identifying anti-vaccine sentiments, this model also highlights the effectiveness of advanced natural language processing techniques and deep learning architectures to deal with complex social media data.

Table 4. Evaluation of the proposed method through performance metrics

Study	Year	Method	Accuracy / F1
Qorib et al. [7]	2023	TextBlob + TF-IDF + LinearSVC	96.8%
Hayawi et al. [8]	2022	BERT	98% (F1)
Aygün et al. [9]	2021	mBERT-base	86%
Kariyapperuma et al. [4]	2022	LSTM	97.6%
To et al. [5]	2021	BERT	91.6%
Quintana et al. [27]	2022	LSTM	77%
Proposed method	2024	TF-IDF + N-gram + Chi-square + SMOTE	99.23%

Building on observer work by Quintana et al., this study of vaccine discussions on Twitter clearly identifies and visualizes language patterns between antivaxes (anti-vaccine campaigners and vaccine deniers) and other groups. Additionally, using features of Antivaxes' tweets, text classifiers have been developed to identify users using anti-vaccine language. This contributes to improving the health of the epistemic environment and supporting public health initiatives by creating an early warning mechanism [27].

5. Conclusions

In this study, Twitter text data was examined using a deep learning approach to determine COVID-19 anti-vaccination. In the first stage, TF-IDF and N-gram methods were used together to extract meaningful features from the texts. A chi-square statistical test was applied to select the most decisive features among the obtained features. In this way, it is aimed to increase the performance of the model and avoid unnecessary information. SMOTE was used to ensure that the classification model required a balanced training dataset. This technique allowed us to combat the oversampling problem while eliminating data imbalance by increasing the samples

of the minority class with synthetic samples. LSTM deep learning model was used to process text data more effectively and make decisions. This deep learning architecture enabled us to achieve a high accuracy rate by better capturing the hidden relationships between texts. The effectiveness of the model was evaluated through performance metrics. It has been shown that the proposed model can be used successfully in detecting COVID-19 anti-vaccination by achieving a high accuracy rate of 99.23%. Therefore, it shows that deep learning-based approaches can be used effectively to detect COVID-19 anti-vaccination. Using a larger data set and collecting data from different social media platforms in future studies may help increase the generalization ability of the model. Additionally, a comparative analysis of different feature extraction methods and deep learning architectures can be performed to improve the performance of the model. This methodology could potentially be valuable in important applications such as detecting and preventing the spread of anti-vaccine views on social media platforms.

6. Acknowledgement

This study was funded by Fırat University (FUBAP) with the scientific research project number MF.23.37.

7. Author Contribution Statement

In this study, Author 1 contributed to the development of the method, obtaining experimental results, and preparation of the paper; Author 2 contributed to the creation of the idea, design and organization.

8. Ethics Committee Approval and Conflict of Interest

"There is no need for an ethics committee approval in the prepared article"
"There is no conflict of interest with any person/institution in the prepared article"

9. List of Abbreviations

COVID–19	Coronavirus Disease–2019
TF-IDF	Term Frequency-Inverse Document Frequency
SVM	Support Vector Machine
LR	Logistic Regression
LSTM	Long Short Term Memory
ANN	Artificial Neural Networks
BERT	Bidirectional Encoder Representations from Transformers
SMOTE	Synthetic Minority Oversampling Technique
AI	Artificial Intelligence
SMS	Short Message Service
CNN	Convolutional Neural Network
API	Application Programming Interface
CSV	Comma Separated Values
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve

10. References

[1] C. H. van Werkhoven, A. W. Valk, B. Smagge, H. E. de Melker, M. J. Knol, S. J. Hahné and B. de GierEarly, "COVID-19 vaccine effectiveness of XBB. 1.5 vaccine against hospitalisation and admission to intensive care, the Netherlands", Eurosurveillance, 29(1), 2300703, 9 October to 5 December 2023.

[2] P. Xu, D. A. Broniatowski and M. Dredze, "Twitter social mobility data reveal demographic variations in social distancing practices during the COVID-19 pandemic", Scientific reports, vol. 14, no 1, pp. 1165, 2024.

[3] M. Umer, Z. Imtiaz, M. Ahmad, M. Nappi, C. Medaglia, G. S. Choi and A. Mehmood, "Impact of convolutional neural network and FastText embedding on text classification", Multimedia Tools and Applications, vol. 82, no 4, pp. 5569-5585, 2023.
[4] K. R. S. N. Kariyapperuma, K. Banujan, P. M. A. K. Wijeratna and B. T. G. S. Kumara, "Classification of covid19 vaccine-related tweets using deep learning", In 2022 International Conference on Data Analytics for Business and Industry (ICDABI), IEEE, pp. 1-5, October, 2022.
[5] Q. G. To, K. G. To, V. A. N. Huynh, N. T. Nguyen, D. T. Ngo, S. J. Alley and C. Vandelanotte, "Applying machine learning to identify anti-vaccination tweets during the covid-19 pandemic,", International journal of environmental research and public health, vol. 18, no 8, pp. 4069, 2021.
[6] A. Mallik and S. Kumar, "Word2Vec and LSTM based deep learning technique for context-free fake news detection", Multimedia Tools and Applications, vol. 83, no 1, pp. 919-940, 2024.
[7] M. Qorib, T. Oladunni, M. Denis, E. Ososanya and P. Cotae, "Covid-19 vaccine hesitancy: text mining, sentiment analysis and machine learning on covid-19 vaccination twitter dataset", Expert Systems with Applications, vol. 212, pp. 118715, 2023.
[8] K. Hayawi, S. Shahriar, M. A. Serhani, I. Taleb and S. S. Mathew, "ANTi-Vax: a novel Twitter dataset for covid-19 vaccine misinformation detection", Public health, vol. 203, pp. 23-30, 2022.
[9] I. Aygün, B. Kaya and M. Kaya, "Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning", IEEE Journal of Biomedical and Health Informatics, vol. 26, no 5, pp. 2360-2369, 2021.
[10] Ö. Çelik and G. Kaplan, "Yeniden Örnekleme Teknikleri Kullanarak SMS Verisi Üzerinde Metin Sınıflandırma Çalışması", Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi, vol. 36, no 3, pp. 433-442, 2020.
[11] A. Avvaru, S. Vobilisetty and R. Mamidi, "Detecting sarcasm in conversation context using transformer-based models", In Proceedings of the second workshop on figurative language processing, pp. 98-103, July, 2020.
[12] E. Özbay, "Transformatör-tabanlı evrişimli sinir ağı modeli kullanarak twıtter verisinde saldırganlık tespiti", Konya Journal of Engineering Sciences, vol. 10, no 4, pp. 986-1001, 2022.
[13] S. A. Alex, N. Z. Jhanjhi, M. Humayun, A. O. Ibrahim and A. W. Abulfaraj, "Deep lstm model for diabetes prediction with class balancing by smote", Electronics, vol. 11, no 17, pp. 2737, 2022.
[14] P. Bhatti, Z. Jalil and A. Majeed, "Email Classification using LSTM: A Deep Learning Technique", In 2021 International Conference on Cyber Warfare and Security (ICCWS), IEEE, pp. 100-105, November, 2021.
[15] N. Azzahra, D. Murdiansyah and K. Lhaksmana, "Toxic comment classification on social media using support vector machine and chi square feature selection", International Journal on Information and Communication Technology (IJoICT), vol. 7, no 1, pp. 64-76, 2021.
[16] M. Hussein and F. Özyurt, "A new technique for sentiment analysis system based on deep learning using Chi-Square feature selection methods", Balkan Journal of Electrical and Computer Engineering, vol. 9, no 4, pp. 320-326, 2021.
[17] P. K. Roy, J. P. Singh and S. Banerjee, "Deep learning to filter sms spam. future generation computer systems", vol. 102, pp. 524-533, 2020.
[18] Y. Zhang and Z. Rao, "n-bilstm: bilstm with n-gram features for text classification", In 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), IEEE, pp. 1056-1059, June, 2020.
[19] M. I. Alfarizi, L. Syafaah and M. Lestandy, "Emotional text classification using tf-idf (term frequency-inverse document frequency) and lstm (long short-term memory)", JUITA: Jurnal Informatika, vol. 10, no 2, pp. 225-232, 2022.
[20] F. A. Özbay and B. Alataş, "Çevrimiçi sosyal medyada sahte haber tespiti", Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, vol. 11, no 1, pp. 91-103, 2020.
[21] A. Ciran and E. Özbay, "Optimization-based feature selection in deep learning methods for monkeypox skin lesion detection", In 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, pp. 1-6, October, 2023.
[22] İ. Sel, C. Yeroğlu and D. Hanbay, "Feature selection by using heuristic methods for text classification", In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) IEEE, pp. 1-6, September, 2019.

[23] X. Jin, A. Xu, R. Bie and P. Guo, "Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles", In Data Mining for Biomedical Applications: PAKDD 2006 Workshop, BioDM 2006, Singapore, Springer Berlin Heidelberg, pp. 106-115, April 9, 2006.
[24] M. Yildirim, "Detection of COVID-19 fake news in online social networks with the developed CNN-LSTM based hybrid model". Review of Computer Engineering Studies, vol. 9, no. 2, pp. 41-48, 2022.
[25] Y. Eroglu, M. Yildirim and A. Cinar, "Diagnosis of periventricular leukomalacia in children with artificial intelligence-based models developed using brain magnetic resonance images", Signal, Image and Video Processing, vol. 17, no. 8, pp. 4543-4550, 2023.
[26] F. B. Demir, M. Baygin, I. Tuncer, P. D. Barua, S. Dogan, T. Tuncer and U. R. Acharya, "MNPDenseNet: automated monkeypox detection using multiple nested patch division and pretrained densenet201,", Multimedia Tools and Applications, pp. 1-23, 2024 .
[27] I. O. Quintana, M. Cheong, M. Alfano, R. Reimann and C. Klein, "Automated clustering of covid-19 anti-vaccine discourse on twitter,", arXiv preprint arXiv:2203.01549, 2022.