International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I3P112

Volume 72 Issue 3, 116-126, March 2024

Original Article

Fake Account Detection in Twitter using Long Short-Term Memory and Convolutional Neural Network

Louzar Oumaima[1], Ramdi Mariam[2], Baida Ouafae[3], Lyhyaoui Abdelouahid[4]

[1,2,3,4]*Department of LTI Lab, ENSA of Tangier, Abdelmalek Essaâdi University, Tangier, Morocco.*

[1]*Corresponding Author : oumaima.louzar@etu.uae.ac.ma*

Received: 28 November 2023 Revised: 15 February 2024 Accepted: 23 February 2024 Published: 17 March 2024

Abstract *- With the growing influence of social media platforms, the identification and prevention of fake accounts has become a crucial challenge for maintaining the integrity of online interactions. The proliferation of Online Social Network (OSN) platforms has given rise to a significant increase in the number of fake accounts, leading to numerous detrimental effects on online communities. Many strategies have been suggested by various communities to deal with false accounts in OSN. Therefore, this paper proposes an innovative approach for detecting fake accounts on Twitter based on the content of tweets. It incorporated Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). It conducted this research in several processes, including data collection, data preprocessing, data reduction by applied Correlation-based Feature Selection (CFS) and Principal Component Analysis (PCA), and data classification. The suggested method, the LSTM-CNN approach, is to cluster more than 2,000,000 accounts from the MIB dataset, and the experimental results show that the approach has the highest accuracy of 98.95% compared with other research.*

Keywords *- Fake account, Twitter, Online Social Network, LSTM, CNN.*

1. Introduction

In recent years, Online Social Networks (OSNs) like Twitter, Facebook, and LinkedIn have experienced a remarkable surge in popularity. These platforms serve as communication tools for individuals to create profiles, connect with others, share content and information, coordinate activities, conduct online businesses, and interact in various ways, such as messaging, commenting, and liking posts. These platforms facilitate social interaction and networking in the virtual world. The growth of social media has been substantial over the past two decades. A significant number of people have joined various social networking platforms, numerous events have been organized, and social networking has a culture of creating false profiles and news.

Moreover, fraudulent accounts employ their profiles for various objectives, including disseminating misinformation that can impact entire markets, economies, or cultures. Finding information about dishonesty is a persistent issue. People set up accounts on various social networking sites to exchange social media content. To spread fake news without disclosing their identities, users frequently create accounts with inaccurate or anonymous information. Additionally, users frequently open accounts in other people's names (identity theft) or hack into other people's accounts. False accounts frequently interact and follow the posts of influencer individuals on the network. Through an impersonation policy, even social media sites, for example, Facebook, Twitter, and WhatsApp, erase or block these phoney accounts [1]. The establishment of fake accounts contributes to the propagation of false information and hoaxes in society. As a result, identifying false accounts is crucial to stopping the spread of fake content on social networks [2]. However, this comes with a fair share of issues, chief among them the swift dissemination of false information. Traditional false information is regarded as deliberate deception. Writing and publishing fake news is typically done to deceive, harm, or benefit politically or financially from an organization, entity, or individual. The fundamental processes in the propagation of false news include creation, publication, and dissemination. Traditional false news mostly preys on consumers by taking advantage of their weaknesses. Intentionally overstated content, crafted with striking language or emotional appeal, accompanied by compelling visuals that evoke strong user sentiment, and employing clickbait to attract clicks to the provided links are frequently linked to the success of fake news transmission [3]. Because of the way news is shown on user's feeds and home pages, users frequently only interact with specific types of content.Additionally, users frequently create groups of like-minded individuals, polarizing their views. Consumers are inherently susceptible to bogus news for two important reasons. According to the Theory of Perception or relationism, naive realism is the inclination of a user to accept the news that they perceive.

At the same time, confirmation bias is the preference of a user to seek information that aligns with their existing beliefs [4]. Either an economic model of fake news or an epidemiological model exists. According to the economic model, the notion of false news is comparable to a two-player strategic game in which publishers and consumers act as the central players. While customers want to make the most of getting correct information and news that supports their preconceived notions, publishers aim to optimize their profits by reaching more consumers and enhancing their reputation for authenticity [5].

The impact of fake profiles on online social networks (OSN) can be significant and multifaceted, affecting users, businesses, and the platform itself. Here are some key impacts:

- Fraud and Scams: Fake profiles are frequently involved in fraudulent activities and scams. They may engage in phishing, identity theft, romance scams, or financial fraud schemes by tricking unsuspecting users into providing personal information or money.
- Trust and Credibility: Fake profiles undermine trust and credibility within the online community. When users encounter fake accounts, they may become skeptical of the authenticity of other profiles and content on the platform, leading to a decrease in trust in the platform as a whole.
- Cyberbullying and Harassment: Fake profiles can be used for cyberbullying, harassment, and identity theft. Individuals can create fake accounts to harass or defame others anonymously, resulting in emotional distress, reputational damage, and even legal consequences for victims.
- Platform Integrity and User Experience: The presence of fake profiles undermines the integrity of the platform and degrades the overall user experience. This can lead to spammy content, irrelevant interactions, and lower user engagement, which can drive genuine users away from the platform.
- Regulatory and Legal Issues: Social media platforms are subject to regulatory scrutiny and legal liabilities related to fake profiles and their associated activities. Governments and regulatory agencies can impose fines or regulations on platforms that fail to adequately address the problem of fake profiles and the distribution of harmful content.

Among the social networks, Twitter is one of the most popular social network microblogging platforms. Toward the task of fake account detection, it proposed a novel method using Long Short-Term Memory (LSTM) with a Convolutional Neural Network (CNN). The objective of this study is to properly prepare the dataset using feature reduction techniques and then train the model by combining both CNN and LSTM algorithms.The subsequent sections of the paper are structured as follows: Section 2 provides a summary of the current literature.

Section 3 outlines the objectives of the study. Section 4 details the methodology employed to attain these objectives, including data preprocessing, reduction, and classification techniques. Section 5 introduces the evaluation metrics utilized. Section 6 describes the experimental results and discussion with the conclusion presented in Section 7 & 8 respectively.

2. Related Work

In recent years, there have been great strides in fake account detection on OSNs (Online Social Networks) and on various datasets using various machine learning and deep learning algorithms. Some of the works have been described below:

Santosh Kumar Uppada et al. [6] proposed a Social Engagement-based News Authenticity Detection (SENAD) model that evaluates the genuineness of news articles shared on Twitter by examining the credibility and objectivity of the users involved. The novel concept of authenticity score is incorporated into the suggested SENAD model, which also takes into consideration user-centric social engagement metrics like account age, following-followers ratio bias, etc. The approach substantially improves the ability to identify fake news and accounts, as evidenced by its classification accuracy of 93.7%. Maria Grazia Vigliotti et al. [7] examined the evolution of communication in a social network graph over time to investigate the identification of anomalous behavior.

They presented a novel method that enabled them to deduce a subset of nodes within the social network that may share similar qualitative characteristics, operating on the assumption that certain nodes in the network could be qualitatively labeled. Nasira Perveen et al. [8] suggested a novel method for distinguishing between real tweets and spam tweets on Twitter, a well-known social networking website. The suggested method, which was informed by Twitter's spam detection policies and studies of spam practices, mixes sentimental and POS-based data with content/user-based information. Using the Twitter API, a dataset of the top 29 trending topics from 2012 was compiled. The research uses five conventional classifiers—BayesNet, Naive Bayes, Random Forest, Support Vector Machine (SVM), and J48 schemes—to assess the performance of the suggested features in spam identification. Naive Bayes, J48, and Random Forest classifiers achieved 93% precision and 95% F-measure in spam detection. Nadav Voloch et al. [9] presented a novel method for user privacy protection that is divided into three main stages and addresses three of its key components: information flow, role-based access control, and trust. The proposed method classifies a user's straight connections to roles and takes into account the user's sub-network. It uses publicly available data to assess the quality of network connections, including overall friend count, user account age, and length of friendship. Suneet Joshi et al. [10] developed a relationship identification (RIF) framework based on

graphics, linguistics, and social theory to identify harmful end-users on social media. This concept combines profile and graphical data with the grammatical, temporal, and contextual ethics of user-generated material. To identify user behaviors and similarity indices across social media, the RIF framework extracts feature vectors. Maximum precision, recall, F1-Score, and accuracy were 82.49%, 87.76%, 86.19% and 98.54%, respectively, with the RIF framework.

Vishal Sharma et al. [11], in their work, a problem of cross-platform anomalies are taken into consideration when an individual exhibits various behavior with various users across the various OSNs. Cognitive tokens create an Intelligent Sensing Model for Anomaly Detection (ISMA) by intentionally generating misleading data to lure anomalous users, which is presented as a solution to this issue. Add the results % of the studies below.

Khalid Binsaeed et al., in their research [12], a novel method proposed for identifying spam in Twitter microblogging that makes use of machine learning (ML) methods and domain popularity services. The suggested strategy consists of two key phases: 1) Tweets are regularly collected and undergo filtration by choosing the ones that show up more often than a predetermined threshold in the given time frame like tweets that are widely observed. Subsequently, the linked URL domain undergoes verification against the top one million globally viewed websites, according to Alexa, to conduct an inspection of the common tweets.

A tweet is marked as potentially spam if it frequently appears on Twitter but does not rank in the top a million websites accessed worldwide. 2) The second stage begins by using machine learning (ML) algorithms to flag tweets to extract attributes that aid in the real-time detection and prevention of spam clusters. The proposed strategy's effectiveness has been assessed using the three most often used classification models (random forest, Naive Bayes, and J48). Results for all classifiers demonstrated the usefulness of the suggested strategy utilizing several performance metrics such as accuracy, precision, F1-score, and sensitivity) and test phase.

Prabhu Kavin et al. [13] provided an artificial intelligence technique for Twitter social networks' spam identification in their study. To create the model, they used a random forest approach, a neural artificial network, and a vector support machine. According to the results and in comparison to the RF and ANN algorithms, the proposed support vector machine algorithm demonstrates superior precision, recall, and F-measure. Sowmya P and Madhumita Chatterjee [14] proposed a method of detecting Fake and Clone profiles on Twitter. A collection of rules that can distinguish between false and real profiles is used to detect Fake profiles, and there are two ways to identify Clone profiles, one employing the C4.5 decision tree technique and the other utilizing two types of similarity measures, similarity of attributes and similarity of network relationships.From these articles, it observed that there is a huge variation in the application of machine learning algorithms but a minority who use deep learning algorithms know that they performed good results in different fields. There is also diversity in the methods used and the components on which the authors relied to detect anomalies. Some relied on tweets and their content, others on user information, while some focused on the similarity of attributes and relationships in the network.

3. Objective of the work

The objective of the work is to define a new approach to detecting fake accounts that are not authentic on Twitter or that are created to deceive users by developing a more efficient, precise, and robust method. This will include improving the ability to detect newly created fake accounts, identify suspicious behavior, and differentiate between genuine and fake profiles. By developing such an approach, it sought to strengthen the security and reliability of the Twitter platform, protect users against misinformation and fraud, and maintain trust in the platform as a source of information and communication by line. This approach will move towards deep learning by combining the two algorithms CNN and LSTM, evaluating the algorithm on a large Twitter database, and comparing the results obtained by those of the article [15], which was based on the Bidirectional Gated Recurrent Unit (BiGRU) model to detect fake accounts on the same database that it used in this approach.

Table 1. Comparison of the different models cited in the related work

Reference	Model
[6]	Evaluate the genuineness of news articles shared on Twitter by examining the credibility and objectivity of the users involved.
[7]	Classify 'behavior' to be normal or abnormal based on the nodes of the network.
[8]	Distinguishing between real tweets and spam tweets based on the content of users
[9]	Protection of user privacy based on information flow, role-based access control, and trust.
[10]	Identify harmful end-users on social media based on graphics, linguistics, and social theory.
[11]	Detection of anomalies by generating misleading data to lure anomalous users.
[12]	Identify spam in Twitter microblogging based on the content of the tweet.
[13]	Identify Twitter social networks' spam based on the network content.
[14]	Detecting Fake and Clone profiles on Twitter based on the similarity of attributes and similarity of network relationships.

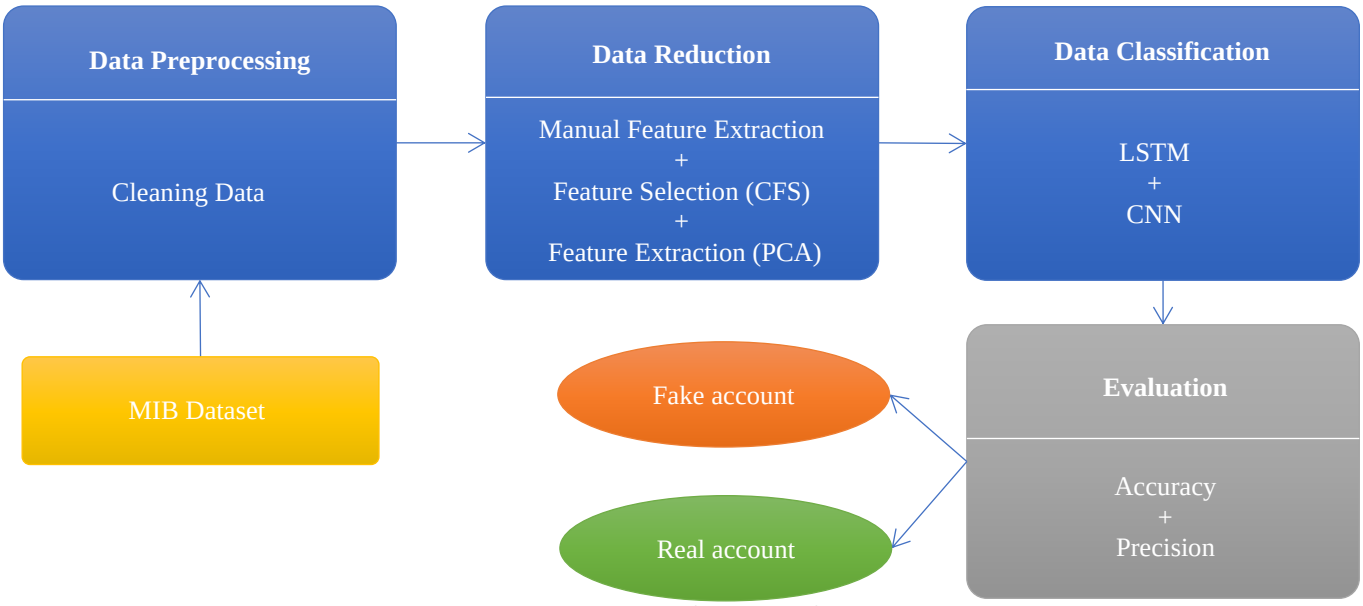

Fig. 1 Proposed approach for detecting fake account

4. Methodology

This section provides a comprehensive explanation of the proposed approach. The model architecture is illustrated in Figure 1, comprising three key stages: data preprocessing, data reduction, and data classification. The approach started with dataset preprocessing. Subsequently, in the second phase, it incorporated various reduction techniques. Within the reduction phase, data underwent filtration and reduction using specific mechanisms, preparing it for the subsequent classification phase. During classification, the filtered data was subjected to classification algorithms, culminating in the presentation of the ultimate outcomes.

4.1. Data Description

It made use of a well-balanced dataset generously provided by Cresci et al. [16] from Twitter. This labeled dataset comprises individual accounts/users and their associated tweets, as described in Table 2. To maintain cohesiveness, tweets from the same user were aggregated into single documents, considering that users may have authored multiple tweets. Users without any tweets were excluded from consideration.

The dataset is composed as follows [17]:
genuine accounts:
Accounts collected and verified by humans
social spambots #1:
followers of a political candidate in Italy
social spambots #2:
spammers that promote premium mobile apps
social spambots #3:
spammers who advertise things on Amazon.com
traditional spambots #1:
Authors employed a spammer training set
traditional spambots #2:
scam URL spammers
traditional spambots #3:
spamming job offers through automated accounts
traditional spambots #4:
additional automatic accounts, unwanted job postings
fake followers.
simple accounts that artificially boost another account's number of followers.

4.2. Data Preprocessing

The dataset contains 3,474 accounts/users and more than two million tweets. In the preprocessing phase, it started by merging the different databases into a single dataset that contains 40 attributes with several types. Then, it identified and handled the missing values, removed punctuation marks, stop words (a, an, the, of, …), mentions, and hashtags in tweets. In the article [15], the authors focused on the semantic and syntactic analysis of the tweets; on the other hand, the approach took into consideration the content of the tweets and the other attributes that are related to the characteristics of the profile.

Table 2. The full dataset

Name of group	Accounts	Tweets	Year
genuine accounts	3 474	8 377 522	2011
social spambots #1	991	1 610 176	2012
social spambots #2	3 457	428 542	2014
social spambots #3	464	1 418 626	2011
traditional spambots #1	1 000	145 094	2009
traditional spambots #2	100	74 957	2014
traditional spambots #3	433	5 794 931	2013
traditional spambots #4	1 128	133 311	2009
fake followers	3 351	196 027	2012

Table 3. All features and their types

No.	Feature Name	Type
01	user id	integer
02	tweet	string
03	name	string
04	Screen name	string
05	statuses count	integer
06	followers count	integer
07	friends count	integer
08	favourites count	integer
09	listed count	integer
10	url	string
11	lang	string
12	time zone	string
13	location	string
14	default profile	float
15	default profile image	float
16	geo enabled	float
17	profile image url	string
18	profile banner url	string
19	profile use _background image	float
20	profile background image url https	string
21	profile text color	string
22	profile image url https	string
23	profile sidebar border color	string
24	profile background tile	float
25	profile sidebar fill color	string
26	profile background image url	string
27	profile background color	string
28	profile link color	string
29	utc offset	float
30	is translator	float
31	follow request sent	float
32	protected	float
33	verified	float
34	notifications	float
35	description	string
36	contributors enabled	float
37	following	float
38	created at	string
39	updated	string
40	target	integer

The Table 3 presents the different features and their types.

4.3. Data Reduction

Since the dataset comprises numerous attributes, it focused on evaluating the most impactful ones. Attributes lacking significance were omitted from the model. The approach holds significance for the application of various machine learning algorithms to the dataset.

Achieving this was based on the methods of feature selection and feature extraction after manual extraction.

4.3.1. Manual Feature Extraction

For manual feature extraction, it is essential to identify and describe the characteristics relevant to a specific situation, along with implementing a method to extract these features. A thorough understanding of the context or domain is often beneficial for making informed decisions about which characteristics could be valuable.After visualization of the data and its meaning, there are a few attributes that are useless for the current study and cannot add any value to the research, such as profile picture links and backgrounds:

- profile image url
- profile banner url
- profile uses the background image
- profile background image url https
- profile image url https
- profile background image url

4.3.2. Feature Selection

The process of choosing a subset from the initial feature set based on the significance of the features is known as feature selection. There are two key phases to feature selection. First, choose appropriate features from the original dataset's class (base). Next, find and eliminate any superfluous features [18].The advantages of feature selection [19] will help mitigate overfitting in the model, minimize storage needs, improve dataset accuracy, lower computational costs, make predictive models easier to understand, use fewer computing resources, and alleviate the curse of dimensionality. Wrappers, filters, and embedding techniques are the three different types of feature selection methods. Wrapper approaches use statistical reassembling or cross-validation to assess a subset of characteristics' prediction accuracy. A smaller collection of features that are selected iteratively are used to train the model. Due to the extensive training and cross-validation, the approach exhibits slower performance. Filter methods preprocess the data. To compute and predict the target feature, these methods take into account the interrelationships among features. Various statistical tests are then conducted on the features to identify high-ranking ones. The filter method and wrapper method are combined to create the hybrid embedded method. In the current case, it applied Correlation-based Feature Selection (CFS) because the database is large in terms of features, and it is a technique among the Filter methods. CFS measures the correlation between features and the target variable. Features with high correlation are selected. It aims to find the most informative and least redundant features to improve the performance of predictive models and reduce dimensionality. CFS first calculates the correlation between each characteristic and the target variable. The Pearson correlation coefficient for continuous target variables or other metrics like the chi-squared statistic for categorical target variables can both be used to calculate this correlation. Then, using each feature's association with the target variable as a guide, CFS creates subsets of features one at a time. Starting with a blank subset, characteristics that have the highest individual correlation

with the target are incrementally added. CFS analyzes the pairwise correlation between the characteristics in each subset to reduce repetition among the chosen features. High levels of correlation between features are punished, and only one of them is kept in the subset. This phase is essential to ensuring that the features chosen offer unique information and are not redundant. For each collection of features, the CFS calculates a merit score. The merit score penalizes feature redundancy and combines the individual feature-target correlations. Maximizing this merit score is the aim. The final set of features to be utilized for modeling is chosen by CFS from the subset of features with the greatest merit score. To ensure that the selected feature subset leads to improved model performance on the data, it chose to apply feature extraction in the next step.

4.3.3. Feature Extraction

The process of extracting a new set of features from the collection of features generated during the feature selection step is known as feature extraction. Examples of feature extraction techniques are Principal Component Analysis (PCA), Latent Semantic Indexing (LSI), clustering techniques, etc. By comparing the advantages of these methods, PCA is more effective. One of PCA's main advantages is that it reduces the data's dimensionality.

When the initial data has a lot of variables and is difficult to visualize or understand, this can be useful. Additionally, PCA can be used to extract new features or components from the initial data that may be more insightful or intelligible than the original components. When the original features are linked or noisy, this is especially beneficial. For these advantages, it decided to use PCA in this step. PCA can be used to find patterns and variations in a dataset. The method looks for elements that help achieve the variance maximization objective. The PCA algorithm's identified components are the major components with the highest variances [20]. To eliminate redundant data from the dataset, PCA projects the dataset from high dimension to low dimension. Using the ideas of covariance, eigenvector, and eigenvalue, PCA transforms the data.

The covariance matrix of the standardized data is calculated through PCA. The diagonal elements of the covariance matrix represent the variance of individual features, and it describes the correlations between all pairs of features. Then, it decomposes the covariance matrix using its eigenvalues. Either eigenvalues or eigenvectors result from this decomposition. Each eigenvalue is the percentage of variation explained by the corresponding eigenvector, and each eigenvector indicates a major component. The eigenvectors are ordered in descending order by the corresponding eigenvalues. Predefined criteria (for example, keeping components that account for 95% of the variance) or the explained variance ratio is frequently used to decide the number of major components to keep.

Table 4. Resultant features

01	tweet
02	followers count
03	friends count
04	favourites count
05	listed count
06	url
07	lang
08	default profile
09	default profile image
10	geo enabled
11	follow request sent
12	statuses count
13	verified
14	protected
15	target

The original data is converted into a new, lower-dimensional space using the principal components that have been chosen. The data points are transformed by being projected onto the new coordinate system established by the primary components. The dataset processing has yielded a total of 15 resultant features, which are presented in the accompanying table.

4.4. Data Classification

Now, after preparing the dataset, the result is transmitted as input to the model as presented in the proposed architecture. The objective of the research is to create a predictive system for classifying fake and real profiles in the Twitter dataset based on 60% of tweets and 40% of profile information. It is crucial to choose the best classification strategy for the model after the features, training sets, and test sets have been determined. In the field of analytics, it would be overstated to say that every data set has a flawless classification strategy. The objective will be achieved by employing LSTM and CNN in the construction of the system. The reason for this choice will be detailed in the following.

4.4.1. Long Short-Term Memory

Variable-length sequence processing has been the domain of the Recurrent Neural Network (RNN). However, dealing with lengthy sequences can lead to challenges such as vanishing gradients and information loss due to the abundance of historical data., as the usual RNN is equal to multi-layer feed-forward neural networks. The Long Short-Term Memory network (LSTM) represents a modified version of the RNN architecture, possessing the ability to grasp prolonged dependencies and effectively address the challenge of vanishing gradient descent [21].

Among the reasons for choosing LSTM in the study are the following:

- The ability of LSTM to handle variable-length sequences is crucial in applications where input sequence length

fluctuates. Because sentences in natural language processing tasks might vary in length, like tweets in the instance, this flexibility is especially helpful.

- When data becomes available over time, and the model needs to update its parameters, LSTMs can be applied to online learning scenarios. They are, therefore, appropriate for uses requiring ongoing education.

LSTM was specifically designed to overcome the challenge of long-term dependencies. One of LSTM's distinctive features is its capacity for retaining information. The core concept of the LSTM model centers around the cell state, which remains relatively unchanged as it traverses through the sequence, undergoing only minor linear interactions. Based on RNN, Long Short-Term Memory (LSTM) is an enhanced network architecture. By incorporating a memory cell and three control gates, LSTM efficiently preserves historical data in lengthy sequences, mitigating the defeat of archival data and disappearance brought that may arise from training deep RNN layers.

In Figure 2, the LSTM structure is displayed. The structure includes an additional memory cell to store historical data. Three gates (an input gate, a forget gate, and an output gate) control the update, deletion, and output of historical data, respectively. The input is employed to ascertain the impact of input vectors on the state of the memory cell and the outputs affected by the output gate. Ultimately, the memory cell's ability to recall or forget its prior knowledge is determined by the forget gate. There are H LSTM hidden layer nodes when the input sequences are X and x_t is a d-dimensional word embedding vector where X is presented as follows:

$$X = [x_1, x_2, \ldots, x_t].$$

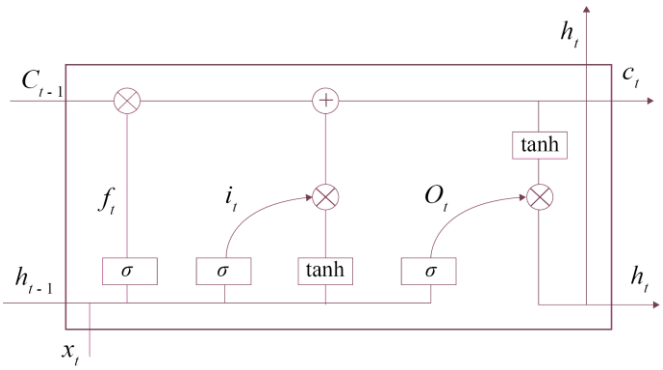

Fig. 2 LSTM structure

The following is an update of the three gates (forget f_t, output o_t, and input i_t) at time step t [22].

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o \quad (2)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f \quad (3)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$h_t = o_t \otimes tanh(c_t) \quad (5)$$

The input gate, input vector, and forget gate determine the memory cell's state at time step t, which is c_t. Its dimension matches the number of nodes in the hidden layer, denoted as H. This is the LSTM's final output. σ is the sigmoid function. The LSTM model was presented in the previous section. Regarding the extraction of features and the semantic analysis of extended text sequences, the LSTM model's unique structure gives it a significant edge over the conventional RNN model; in text classification tasks based on LSTM, two simple techniques can be employed to extract local features from the LSTM outputs: Maximum Pooling and Average Pooling.

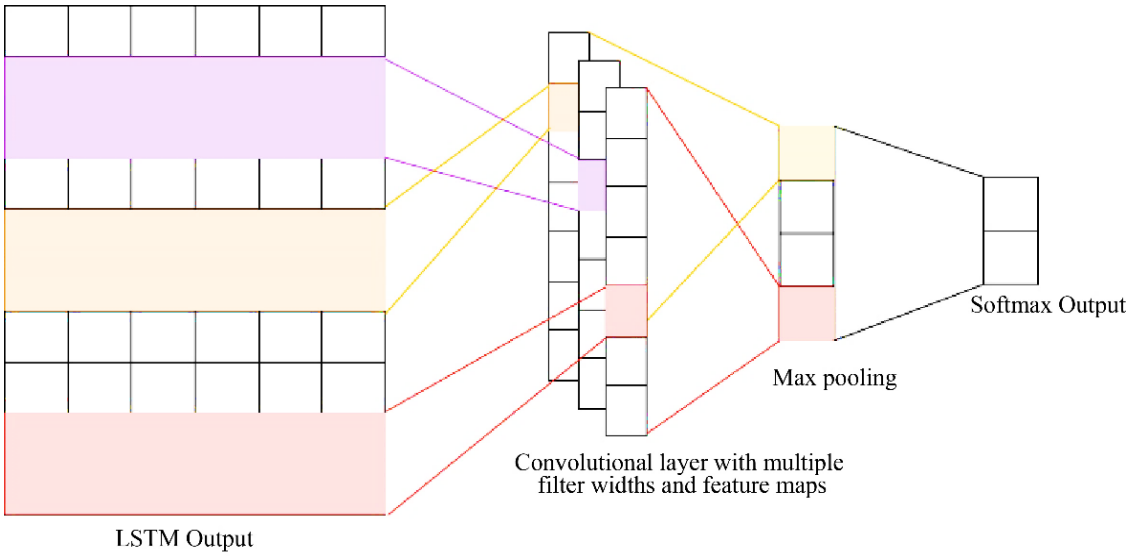

Fig. 3 CNN Structure

In this study, the proposed LSTM-CNN model enhances the architecture by integrating a CNN model on top of the multi-layer LSTM model. This additional layer aims to extract features from input text sequences further, enhancing classification accuracy.

4.4.2. Convolution Neural Network

A specific class of neural networks known as Convolutional Neural Networks (CNNs) has shown to be incredibly successful in a variety of computer vision-related applications. Tasks like object detection, image classification, and image recognition are especially well-suited for them. Here are several reasons why CNN was chosen for the study:

- CNNs use convolutional layers to implement parameter sharing. In comparison to fully connected networks, this helps lower the total number of parameters in the network. The network can learn translational invariance, the ability to identify patterns regardless of where they are in the input space by sharing parameters.
- CNNs achieve translation invariance because convolutional layers employ shared weights. This implies that patterns are resilient to spatial transformations since the network can identify them regardless of where they are in the input space.
- CNN architectures are made to handle high-dimensional input data, like the MIB dataset, with effectiveness and computational efficiency.

Convolutional layers in Natural Language Processing have demonstrated unexpectedly impressive results, even when applied to textual data. Typically, when training a model with sequential textual data, widely used Recurrent Neural Networks (RNN) models exhibit satisfactory performance but require a substantial amount of time. However, incorporating a Convolutional layer after the RNN layer can substantially decrease the training duration of the model. The input vector for the CNN is the output vector generated by the multi-layer LSTM. Fundamentally, the convolutional layer aims to uncover relationships among distinct sentences or paragraphs within a document by utilizing filters. In the current case, this layer allows higher-level feature extraction [23]. In this research, the CNN model is structured as in Figure 3. Following the LSTM model's extraction of each feature sequence, the output is $H = [h_1, h_2, \dots, h_t]$, h_t represents the m-dimensional feature vector of the t word in the text sequence. The number of nodes in the LSTM hidden layer is represented by the vector's length. The number of LSTM expansion steps is the same as the text sequence length, T. The input matrix of CNN is $H \in R^{m \times T}$.

$F \in R^{j \times k}$ Where j is the number of words in the window and k is the dimension of the word embedding vector. The convolutional filter $F = [F_0, \dots, F_{m-1}]$ generates the value at time step t as follows.

$$O_{F_t} = ReLU[\left(\sum_{i=0}^{m-1} h_{t+i}{}^T F_i\right) + b] \quad (6)$$

Table 5. Proposed Confusion Matrix

Predicted class / True class	**Fake account**	**Real account**
Fake account	a	b
Real account	c	d

F and b are the parameters of the single filter where b is a bias, and $ReLU$ is the activation function. It integrated two essential components, an LSTM and a Convolutional Network, to distinguish between genuine and fake accounts using the content of tweets and the nature of these tweets. Consequently, the central concept behind this study is to harness the strengths of both LSTM and CNN to enhance predictive accuracy.

5. Results and Discussion

The dataset is divided into three segments: 70% for training, 10% for validation, and 20% for testing purposes. The entire project is created using Python version 3.6. Additionally, it employed the NLTK library for data preprocessing tasks while enhancing data processing with the utilization of the numpy and pandas libraries. Moreover, it utilized the Python libraries Keras, TensorFlow, and scikit-learn to construct, compute, and assess the effectiveness of the proposed methodology.

6. Evaluation Metric

It gauged the system's performance by employing the following metrics: accuracy and precision. The assessment took into account the confusion matrix presented in Table 5, where a represents the number of correctly classified as fake accounts, b means the number of false accounts misclassified as real accounts, c refers to the number of real accounts misclassified as fake accounts, d indicates the number of correctly classified real account.

Accuracy (A) is expressed as;

$$A = \frac{a+d}{a+b+c+d} \quad (7)$$

It represents the proportion of instances correctly classified for both classes about the total number of instances in the dataset. Precision (P) expressed by:

$$P = \frac{a}{a+c} \quad (8)$$

It denotes the ratio of instances correctly classified to the total number of instances in the dataset.

7. Discussion

The approach's evaluation consists of three components. Initially, it assessed the model's performance before and after preprocessing in the first part. In the second segment, it compared the model with the outcomes of an alternative approach, which utilizes different algorithms but relies on the same MIB database.

Finally, it concluded by comparing the results with those reported in articles that share the common objective of detecting fake accounts on Twitter. As outlined earlier in the preprocessing phase, the dataset undergoes a three-step process. Initially, there is a manual reduction of features, wherein those deemed irrelevant to the model based on the significance of their data are excluded.

Subsequently, it implemented Correlation-based Feature Selection (CFS) to gauge the correlation between features and the target variable. This technique seeks to identify the most informative and least redundant features, enhancing predictive model performance and minimizing dimensionality. To further refine the feature extraction process, Principal Component Analysis (PCA) is employed, aiming to detect elements that contribute to the maximization of variance. It examined the outputs of the feature set both before and after the preprocessing step, specifically comparing the unreduced features.

Upon comprehensive analysis, it becomes evident that employing Correlation-based Feature Selection (CFS) during the feature selection phase and Principal Component Analysis (PCA) during the feature extraction phase proves more effective for substantial dimensionality reduction, leading to a significant enhancement in accuracy.

Table 6 presents the results of the proposed model before and after the preprocessing phase. In the classification phase of the approach, a combination of LSTM and CNN was used. It is effective in capturing both temporal dependencies and spatial patterns within the data. LSTM (Long Short-Term Memory) is adept at handling sequential information, making it suitable for tasks where the order of data is crucial. Meanwhile, CNN (Convolutional Neural Network) excels in capturing spatial relationships through its convolutional layers.

On the other hand, in the approach [15], they employed BiGRU, a bidirectional gated recurrent unit. BiGRU extends the capabilities of traditional GRU models by processing input data in both forward and backward directions. This bidirectional processing allows the model to capture contextual information from both the past and future, enhancing its understanding of the overall sequence.

Table 7 and Figure 4 present a comparison between the two approaches on the same MIB dataset. In comparison to the BiGRU model's outcomes, as reported in the paper [15], utilizing the same MIB dataset, the application of the LSTM-CNN model with CFS and PCA algorithms during the preprocessing phase yielded a notable increase in accuracy, reaching 98.95%. The BiGRU model, under the same conditions, demonstrated an improved accuracy of 98.87%.

This study's findings and the subsequent comparison underscore the pivotal role of the data preprocessing phase, showcasing its significant impact on results. The strategic combination of algorithms, such as LSTM and CNN, proved to be a judicious choice, leading to commendable outcomes. False profiles on Twitter possess the potential to manipulate concepts such as influence and popularity, thereby influencing the economy, political systems, and society at large. These deceptive accounts pose a threat to social media networks.

Table 6. Proposed model results before and after the preprocessing phase

Model	Accuracy	Precision
CNN + LSTM without preprocessing	82.18%	87.93%
CNN + LSTM with preprocessing	98.95%	99.15%

Table 7. Comparison of performance based on the same dataset

Performance	Classifiers	Accuracy	Precision
Proposal approach	LSTM	98.47%	98.90%
	CNN + LSTM	98.95%	99.15%
Comparative approach [15]	BiGRU	98.87%	99.23%

Table 8. Comparison with recent research

Research work	Classifier	Accuracy (%)
Proposed approach	LSTM-CNN	98.95
[24]	RF	98.6
[25]	LR	96.2
[26]	CNN	95.7

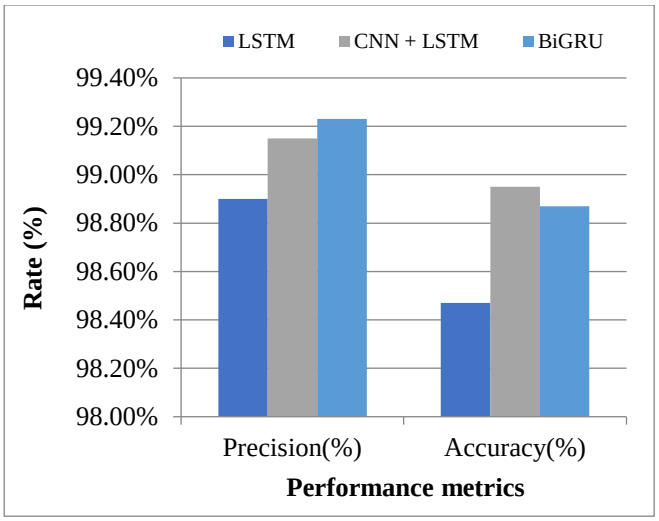

Fig. 4 Comparison of performance based on the same dataset

In the literature, various algorithms are employed to identify and flag fraudulent profiles, ensuring users are shielded from potential harm or misinformation caused by malicious actors. As illustrated in the table provided below, in the classification phase of the article [24], four classifier algorithms were employed, namely J48, Random Forest, KNN, and Naive Bayes. The outcomes reveal that the Random Forest algorithm, coupled with Correlation data reduction, attains the highest accuracy at 98.6%.Conversely, the Naive Bayes algorithm, in conjunction with Correlation data reduction, registers the lowest accuracy, amounting to 82.1%. Conversely, the study employed the Naive Bayes, Decision Tree, and Logistic Regression algorithms to instruct the system in identifying fraudulent Twitter accounts based on readily available information. Following a comparative analysis of all classifiers, the Logistic Regression algorithm demonstrated the highest accuracy, reaching 96.2%.

Furthermore, within the realm of studies on this subject, this paper introduces a novel approach. The authors propose a neural network-based ensemble technique that integrates deep learning methods with traditional feature-based methods for detecting spam at the tweet level. The experimentation involved the utilization of various word embeddings through CNN. The proposed method demonstrates an accuracy rate of 95.7%. Based on these comparisons, it can be inferred that the model contributes significantly to recent research. This contribution is realized through a preprocessing phase employing two feature extraction algorithms, namely Correlation-based Feature Selection (CFS) and Principal Component Analysis (PCA) in the MIB dataset. Additionally, the classification phase is executed using a combination of two robust algorithms: Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN).

8. Conclusion

This study delved into the detection of fake accounts on the Twitter social network by integrating both LSTM and CNN algorithms.

The approach blends deep learning techniques with conventional feature-based methods. The research comprises various stages, encompassing data collection, preprocessing, reduction through Correlation-based Feature Selection (CFS) and Principal Component Analysis (PCA), and classification. Extensive analysis of Twitter data revealed noteworthy characteristics of fake accounts. The fusion of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) techniques, considering tweet content and profile characteristics, resulted in a highly effective algorithm with an impressive accuracy of 98.95%. Future expansions of this research might explore additional features or algorithms, offering promising avenues for further study.

References

[1] Prasanta Kumar Sahoo, and K. Lavanya, "Identification of Malicious Accounts in Facebook," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 1, pp. 2917-2921, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] Priyanka Kondeti et al., "Fake Account Detection Using Machine Learning," *Evolutionary Computing and Mobile Sustainable Networks*, *Springer*, vol. 53, pp. 791-802, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Xinyi Zhou, and Reza Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1-40, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Kai Shu, Suhang Wang, and Huan Liu, "Exploiting Tri-Relationship for Fake News Detection," *arXiv*, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Vahit Çalişir, *Disinformation, Post-Truth, and Naive Realism in COVID-19: Melting the Truth*, Handbook of Research on Representing Health and Medicine in Modern Media, IGI Global, pp. 200-215, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Santosh Kumar Uppada et al., "Novel Approaches to Fake News and Fake Account Detection in OSNs: User Social Engagement and Visual Content Centric Model," *Social Network Analysis and Mining*, vol. 12, no. 52, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Maria Grazia Vigliotti, and Chris Hankin, "Discovery of Anomalous Behaviour in Temporal Networks," *Social Networks*, vol. 41, pp. 18-25, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[8] Nasira Perveen et al., "Sentiment Based Twitter Spam Detection," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 7, pp. 568-573, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Nadav Voloch, Nurit Gal-Oz, and Ehud Gudes, "A Trust-based Privacy Providing Model for Online Social Networks," *Online Social Networks and Media*, vol. 24, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Suneet Joshi, and Deepak Singh Tomar, "Deep Neural Network-Based Relationship Identification Framework to Discriminate Fake Profile over Social Media," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, pp. 599-611, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Vishal Sharma, Ilsun You, and Ravinder Kumar, "ISMA: Intelligent Sensing Model for Anomalies Detection in Cross Platform OSNs with a Case Study on IoT," *IEEE Access*, vol. 5, pp. 3284 - 3301, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[12] Khalid Binsaeed, Gianluca Stringhini, and Ahmed E. Youssef, "Detecting Spam in Twitter Microblogging Services: A Novel Machine Learning Approach Based on Domain Popularity," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 11-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] B. Prabhu Kavin et al., "Machine Learning-Based Secure Data Acquisition for Fake Accounts Detection in Future Mobile Communication Networks," *Security Threats and Challenges in Future Mobile Communication Systems*, vol. 2022, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] P. Sowmya, and Madhumita Chatterjee, "Detection of Fake and Clone Accounts in Twitter Using Classification and Distance Measure Algorithms," *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, pp. 67-70, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Faouzia Benabbou, Hanane Boukhouima, and Nawal Sael, "Fake Accounts Detection System Based on Bidirectional Gated Recurrent Unit Neural Network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 3129-3127, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Stefano Cresci et al., "The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race," *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 963-972, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] MIB Datasets, Consiglio Nazionale Delle Ricerche (CNR). [Online]. Available: https://mib.projects.iit.cnr.it/dataset.html.

[18] Mark A. Hall, "*Correlation-Based Feature Selection for Machine Learning*," Higher Degree Theses, University of Waikato, pp. 1-199, 1999. [Google Scholar] [Publisher Link]

[19] Santiago Egea et al., "Intelligent IoT Traffic Classification Using Novel Search Strategy for Fast-Based-Correlation Feature Selection in Industrial Environment," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1616-1624, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[20] Atiqur Rehman et al., "Performance Analysis of PCA, Sparse PCA, Kernel PCA and Incremental PCA Algorithms for Heart Failure Prediction," *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Istanbul, Turkey, pp. 1-5, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Weicong Kong et al., "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Jiarui Zhang et al., "LSTM-CNN Hybrid Model for Text Classification," *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, pp. 1675-1680, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Tara N. Sainath et al., "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, pp. 4580-4584, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[24] Ahmad Homsi et al., "Detecting Twitter Fake Accounts Using Machine Learning and Data Reduction Techniques," *Proceedings of the 10th International Conference on Data Science, Technology and Applications*, pp. 88-95, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Kusum Kumari Bharti, and Shivanjali Pandey, "Fake Account Detection in Twitter Using Logistic Regression with Particle Swarm Optimization," *Application of Soft Computing*, vol. 25, pp. 11333-11345, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Sreekanth Madisetty, and Maunendra Sankar Desarkar, "A Neural Network-Based Ensemble Approach for Spam Detection in Twitter," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 4, pp. 973-984, 2018. [CrossRef] [Google Scholar] [Publisher Link]