

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/359950590>

Dynamic Light Weight Recommendation System for Social Networking Analysis Using a Hybrid LSTM–SVM Classifier Algorithm

Article in *Optical Memory and Neural Networks* · March 2022

DOI: 10.3103/S1060992X2201009X

CITATIONS

14

READS

54

2 authors, including:



Thailambal Rajagopalan

Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University

22 PUBLICATIONS 104 CITATIONS

SEE PROFILE

Dynamic Light Weight Recommendation System for Social Networking Analysis Using a Hybrid LSTM-SVM Classifier Algorithm

N. S. Kiruthika^{a, *} and Dr. G. Thailambal^a

^a Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Pallavaram, Chennai, Tamil Nadu, 600117 India

*e-mail: sathishpoojaa5@gmail.com

Received December 9, 2021; revised January 8, 2022; accepted January 10, 2022

Abstract—Social media has become one of the significant platforms for information sharing. At that same time, the influence of fake news is a growing cause for all people using social networking platforms. The entire world faces a difficult situation with the COVID-19 pandemic. Simultaneously, the leaking of info in social media concerning COVID-19 also increases exponentially. These information cause serious worries, which affects people psychologically. Several social media analysis methods are developed over years. However, those had several difficulties due to short text social media comments, which causes significant data sparsity. To overcome such difficulties, this paper proposed a recommendation system for social networking to predict whether the information is fake or real using a hybrid LSTM-SVM classifier. Initially, the proposed model gathered real-time COVID-19 related commands from Twitter social media to form a dataset. The collected data is preprocessed by splitting, stop word removal, lemmatization, and spell correction. After preprocessing, the features from the data are extracted and converted to binary with the assist of a count vectorizer. The obtained features are further classified with a hybrid LSTM-SVM model. The predicted data is compared with the preprocessed data, consisting of real information. If the predicted data is equal to the preprocessed data, it will be real news or else fake news. The proposed model is implemented to attain better performance. Some of the performance metrics such as accuracy, sensitivity, specificity, and error are 90, 88, 97, and 0.1% respectively for the proposed model. The overall expected outcome of the recommendation system using hybrid LSTM-SVM is better than the existing techniques such as CNN-SVM, GRNN, LSTM, CNN, and SVM. The Hybrid LSTM-SVM model attained the best accuracy for predicting fake or real news.

Keywords: social networking analysis, fake news detection, count vectorizer, LSTM, SVM, hybrid LSTM-SVM

DOI: 10.3103/S1060992X2201009X

1. INTRODUCTION

Social media is one of the exclusive sources in people's daily lives, where people can express their thoughts, interact, and access news [1]. In social media networks, the news can spread fast, and the participating users in the social network display their sentiments and opinions based on the source stories. Social media provides lots of information in a low-cost and convenient way. Still, this leads to the worst by-products and the propagation of misinformation like fake news [2]. Fake news is wrong information that spread on social media that is obviously false. Political, economic parties, or psychological profit can mislead the fake news and it will widespread among a huge population. Recently various machine learning and data mining techniques have been used to identify fake news easily. It can help to know the real news [3]. Recent techniques depend on extracting texture attributes using a bag of words and n-gram. Subsequently, several machines or deep learning approaches, such as support vector machine, random forest, etc were also used for classifying real news from fake news. Factive or assertive verbs, subjectivity, writing style, and consistency are the advanced features used for Natural Language Processing. Presently, Scholars learn improved features to detect fake news on social media [4, 5].

Identifying fake news through online is still very difficult. Several existing content-based methods need a long text for processing. However, when compared to long text the sentences and words can be learned

better [6]. Social media, especially Twitter information, are generally short text, which causes major data sparsity issues. The fake news prediction model needs diverse user comments for each news scenario. In this way, fake news can be easily recognized [7]. But, many users share their comments in simple re-share the source scenario. It can also be used for effective prediction of fake news. Classifying the misinformation is expensive due to diffusion structure of retweets because of privacy concerns like social media. Many of them delete or hide the information [8]. Various researchers developed different prediction approaches [9, 10] to deal with these concerns. But, still several issues are prevailing when identifying the fake news in Twitter comments.

The source of misleading information has a harmful influence on social interactions [11]. In recent years, the worldwide has faced a situation is coronavirus pandemic. COVID-19 has become an important discussed topic on social platforms. Hence, the pandemic of the COVID-19 issue produces various spread of misinformation like fake news [12–14]. The COVID-19 pandemic information leaks extremely on social media. This information causes significant difficulties and influences society [15]. Fake news not only endangers people’s physical health but also creates negative impression on people’s mental health, causing excessive anxiety or terror due to disinformation. So, to overcome this issue, the recommendation system for social networking to predict whether the information is fake or real is designed by using a hybrid LSTM-SVM classifier. This paper focuses on detecting the fake news in Twitter social media regarding the COVID-19 pandemic. The proposed consists of three phases: preprocessing, feature selection, and classification to predict the COVID-19 pandemic-related information that is real or fake. The main focus of this model is to attain the best accuracy for predicting real or fake news. The main contribution of the paper is summarized as follows,

- A dynamic, lightweight recommendation system for social networking analysis is attained using a Hybrid LSTM-SVM classifier algorithm.
- The preprocessing technique is utilized to improve accuracy and reduce the process complexity of the real-time Twitter dataset to provide better performance.
- The count vectorizer converts the text features to binary features for effective feature selection.
- A Hybrid LSTM-SVM classification model is proposed to attain a better accuracy for prediction, which produces effective predictions than other existing models.
- The hybrid LSTM-SVM classification model is designed to predict fake news on Twitter to avoid the spreading of unwanted news related to the COVID-19 pandemic.

The upcoming section of the paper is structured as follows, and Section 2 describes some of the research articles related to the fake news detection model using various machine learning and deep learning techniques. Section 3 provides a detailed description of the proposed methodology and process involved in the proposed model. Section 4 explains the results gathered through executing the proposed fake news detection model. Section 5 finally concludes the entire research work.

2. LITERATURE REVIEW

Several researches have been performed by various scholars using recommendation techniques for fake news detection. Most existing techniques are designed based on Neuro-Fuzzy, CNN, and SVM. From them, a few are reviewed below.

In [16], the researcher had designed an Ontology and Context-Based Recommendation System (OCBRS) to consider the review’s context to determine the opinion. In this model, the Neuro-Fuzzy Classification method was used to extract the context of the review. The reviews were automatically classified in this method by using the fuzzy rule. Ontology facilitates the systematic and hierarchical methodology to cluster the context and act as a repository of context. This method approaches an improvement in the accuracy of recommendation systems.

In [17], the researchers designed an online recommendation system strategy utilizing deep learning specifically Convolutional Neural networks. The CNN architecture contained diverse class patterns that can choose regarding the preferences of users and designers. The color compatibility for textile items was the deep learning model’s recommended pattern. The pattern dataset, which contains 12000 photos, was used to test and train the model. Experiment analysis was performed with pattern datasets. The researcher of [18] presented an automatic prediction of COVID-19 regarding deliberations from social media. A natural language process approach depends on topic modeling to reveal multiple difficulties about COVID-19 from public sentiments. In addition, the model revealed the sentimental classification of COVID-19 comments using LSTM recurrent neural network.

In [19], the researcher had designed a recommendation approach utilizing the SVM machine learning technique for attribute extraction as well as combined ontology strategy. In this method, the verbal demand from the inactivated user was captured by a humanoid robot and translated into a full-text query. This technology analyzed the comprehensive queries to extract the demands of impaired users and convert them into search engine-friendly formats. The SVM was used to locate important data and eliminate use-less information. Combined ontology-based sentiment analysis was used to determine the item's positive polarity for recommendation to the user. Java was used to create the intelligent model and combined ontology.

In [20], the researcher had designed a social assembly recommendation utilizing a dynamic connection technique. A new framework was employed in this strategy to study the links between group users and group interests. A large group was broken into various interest subgroups, each of which had strongly linked people and shared similar interests. The system then used the connections between group users to compile a possible compact candidate set of media-user pairs for each interest subgroup, which further helped in developing a subgroup-based recommendation list. Then, as the final group recommendation results, a novel aggregation algorithm was devised to merge the recommended media lists of all interest subgroups.

In [21], geometric, deep learning was presented for an automatic fake news identification model. The fundamental algorithm simplified the classical CNN to graphical representation, which enables the fusion of varied data such as new propagation, social graph, user activity, profile, and content. The developed model used a Twitter dataset to train and test the information. In addition, fake news spread within a few hours of propagation the model predicted the fake news at an early stage. The model tested and trained the data separately at different times. The model utilized the content-based strategy to predict fake news. In [22], the researcher had presented an automatic fake news detection strategy for Facebook fake news prediction in a chrome environment. The developed model extrzseveral features from a Facebook account with a lot of news content attributes to investigate the characteristics of the account over deep learning. The implementation study used real-time data from Facebook.

The researcher in [23] had developed a fake news identification model based on Graph-aware Co-Attention Networks (GCAN). The developed model detected fake news issues in real-time events on social media networks. The dataset execution used the short-text and relevant sequence of retweet users without text comments from Twitter social media. GCAN classifier predict whether the source Twitter comment was fake or not.

According to the above-discussed literature, the existing strategies still contain several difficulties such as rapid propagation, access method and high cost. Various recommendation systems are designed based on Neuro-Fuzzy [16], CNN [17] and SVM [18] techniques. The performance of the existing model is less compared to the proposed model. Automatic fake news detection models on social media still have challenges because usually, social media comments are too short text, which causes major data sparsity issues [21]. Fake news detection model needs a diverse amount of user comments for each news scenario but, many users share their comments in simple and re-share the source scenario. This also causes difficulty for achieving effective prediction of fake news [22]. The fake news detection issues are a more realistic scenario on social media [23]. The proposed model designed a fake news detection model for Social Networking Analysis using a Hybrid LSTM-SVM Classifier to overcome these difficulties. Therefore, a hybrid LSTM-SVM model is proposed to recommend real or fake news effectively.

3. PROPOSED METHODOLOGY

A dynamic, lightweight recommendation system for social networking analysis using a hybrid LSTM-SVM classifier is presented to predict fake news. Social media has been developed as one of the major sources for sharing and gathering information from worldwide in the past decade. The continuous evolution of social media is progressing fast because its access methods are generally convenient ways for users. People continue to benefit from the easy and convenient accessibility of social media, but they also show to certain noisy and inaccurate information spread on social media, particularly fake news, a large number of misleading contents like COVID-19 related false news endure online social platforms like Twitter. Many natural language programming researchers have been developed various techniques for identifying the online COVID-19 fake news. For accurate prediction, the proposed model developed an effective fake news detection model using LSTM-SVM classifier and the architecture of the proposed model is illustrated in Fig. 1.

The proposed fake news detection architecture consists of three phases: preprocessing, count vectorizer, and classification for fake news prediction. Initially, the dataset is gathered from Twitter social media.

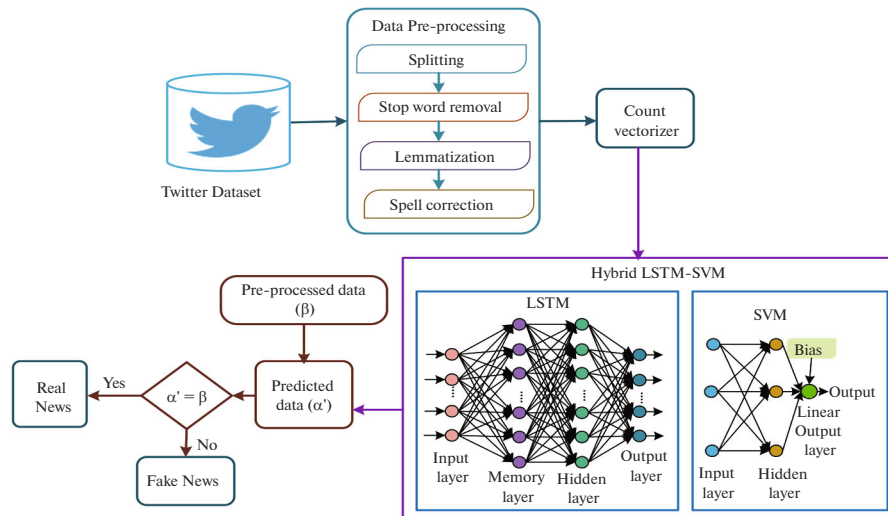


Fig. 1. Proposed Architecture of Social Networking Analysis using a Hybrid LSTM-SVM Classifier.

The preprocessing phase involves Splitting, Stop word Removal, Lemmatization, and Spell correction. A count vectorizer is utilized for extract the features from text to binary. The extracted features are classified with the hybrid LSTM-SVM model. The predicted data is compared with the preprocessed data, consisting of real information. If the predicted data is equal to the preprocessed data, it will be real news or fake news.

3.1. Data Gathering

Initially, real-time Twitter data are gathered to make a dataset [30]. The dataset contains 5000 manually acquired Twitter instructions in both fake and actual news, and it is completely based on COVID-19 pandemic-related information. These data are given into the preprocessing for converting the raw data into a specified format, i.e., the system readable format.

3.2. Data Pre-Processing

In this proposed model, preprocessing is performed in four ways: splitting, stop word removal, lemmatization, and spell correction. These methods preprocess the Twitter data as well as prepare the data for the upcoming step to extract effective features. This process eradicates the noise in the Twitter dataset by normalizing or eliminating the unwanted data.

— **Splitting:** The tweet data are initially in the form of sentences. Therefore, sentences are split into individual words for processing. String splitting is the technique of systematically dividing a text string into distinct components that can be processed.

— **Stop word Removal:** The output attained from the splitting process is again given into the stop word removal process, and it a commonly utilized in Natural Language Processing (NLP). For instance, stop words such as “the”, “a”, “an” that occur frequently are eliminated across all the documents in quantity and allow applications to focus on the important words instead.

— **Lemmatization:** The algorithmic process of determining the ‘lemma’ of a word based on its meaning is known as lemmatization. Lemmatization is a term that relates to doing things correctly using a vocabulary and morphological analysis of words, with the goal of removing inflectional endings solely and returning the base or dictionary form of a word, known as the lemma.

— **Spell correction:** Spell-checker analyzes each and every word with thousands of proper spell words. The majority of techniques use data from many sources of noisy and correct word mappings as training data for automatic spelling correction.

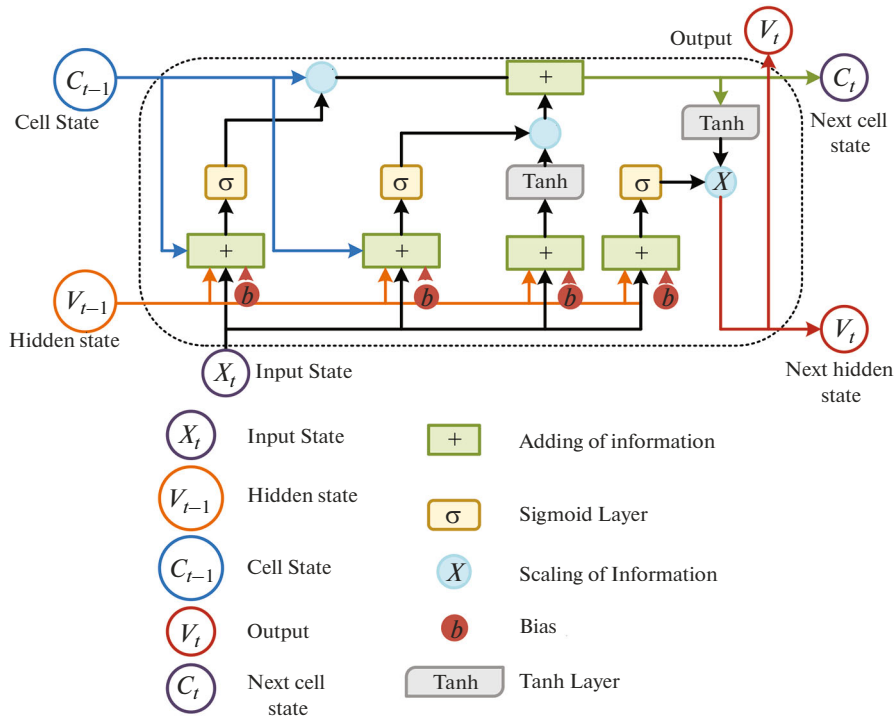


Fig. 2. Single cell Architecture of LSTM.

3.3. Count Vectorizer

Count Vectorizer is a machine learning model for extracting the features that are the main concept of Natural Language Processing. Afterwards, when the data has been imported into the data frames, it is processed to ensure that only the required fields are taken into account. After the preprocessing is completed, the Count Vectorizer generates a dictionary from the training dataset. Individual characters are extracted to gain features in the method. The count vectorizer converts the preprocessed word vectors into binary vectors from the bag of words based on words frequency. The count vectorizer approach functions on the basis of Sklearn Library. The features are created into representative vectors [24]. According to the detailed representation of tweets, the features are required to examine an assessment of high-frequency words before executing a count vectorizer. Finally, it can convert the text vectors into binary vectors for effective classification.

3.4. Classification Process

In this mode, the hybrid LSTM-SVM classification model is used for predicting the fake news on Twitter to avoid spreading the fake news concerning the COVID-19 pandemic. The binary vectors from the count vectorizer are given into the hybrid LSTM-SVM classifier. Hence, a detailed description of the proposed hybrid classification approach is given.

3.4.1. LSTM. Long Short Term Memory (LSTM) networks are a specialization of Recurrent Neural Network (RNN), and it can be used to detect long-term dependency in a single sentence. This kind of neural network was employed in the various fake news detecting methods. This shows a better classification performance than several existing [25]. LSTM units can transmit a key characteristic from the beginning of the input sequence over a considerable distance, capturing potential long-distance dependencies. LSTM is a state-of-the-art learning technique for semantic composition that allows computing a document's expression from its words' description at many abstraction levels. The representation of all words is denoted, such as real-valued vector, continuous and a low dimensional, which is also termed as embedding of words whereas, the word embedding matrix is used to stack each word vector.

LSTM is a kind of recurrent neural network (RNN) and it is developed to solve stability and speed difficulties [26]. The functioning of LSTM is almost similar RNN, a new idea is presented with interaction

per module or one cell. The LSTM network contain memory blocks which is termed as cells and this cell state is considered as an significant state which enables the forward data flow and remains unchanged. With the help of sigmoid activation functional gates the data can be removed or added. These gates with different specific weights so, it is connected like series of matrix operations. While implementing the gates long term dependency difficulties can be occurred this can be solved by using these memorizing LST. The steps involved in the process of LSTM is given as below.

Step 1: The forget gate. The initial step is responsible for recognize the data that is no need for this process. The sigmoid function is responsible for this process and consider the current input X_t at time t and output as V_{t-1} at time $t - 1$. Based on the previous output the sigmoid function make a decision for eliminate the unwanted portion in the output [27]. This process is termed as forget gate (f_t). Moreover, (f_t) lies in-between the limits 0 and 1, which is almost matching to every value in the cell C_{t-1} .

$$f_t = \sigma(w_f [V_{t-1}, X_t] + b_f). \quad (1)$$

In Eq. (1), w_f is the weight, b_f is the bias of forgot gate and σ is the sigmoid function.

Step 2: Learn gate. In this step include two states such as, one ignore state and one storing state form the X_t in cell state. The current input X_t is and V_{t-1} are combined together thus required data that is learned from the hidden state V_{t-1} can be applied to the X_t . In addition, this process have two layers such as, sigmoid layer and tanh layer. The sigmoid layer is responsible for decides whether new data is updated or removed utilizing 0 and 1. The tanh layer is responsible for passes the values among -1 to 1 , this can update the weights. Values are selected depending on their level of importance.

$$m_t = \sigma(w_m [V_{t-1}, X_t] + b_m), \quad (2)$$

$$N_t = \tanh(w_n [V_{t-1}, X_t] + b_n). \quad (3)$$

Step 3: The remember gate. The cell state information that is not forget in this stage. The hidden state (V_{t-1}) and X_t are combined together in remember gate that can update the cell state information C_{t-1} . It combine the previous hidden state output (V_{t-1}) and current input (X_t) to produce the remember gate.

$$C_t = C_{t-1}f_t + N_t m_t. \quad (4)$$

Step 4: The output gate. In this stage combine the significant data from previous cell state (C_{t-1}) and previous hidden state V_{t-1} to make a hidden state (V_{t-1}) next cell as well as produce the output of the current input.

$$Q_t = \sigma(w_q [V_{t-1}, X_t] + b_q), \quad (5)$$

$$Q_t = Q_t \tanh(C_t), \quad (6)$$

where b_q and w_q are the bias and weight respectively for the output gate. The output from the LSTM is given to the input of the SVM classifier.

3.4.2. SVM. A support vector machine (SVM) is one of the supervised machine learning algorithm, it is used for interchange with a SV network. SVM is the most efficient classification technique and it outperforms other classification methods. It perform features in document classification task, making it extremely difficult to discover relationships in sentences, which is frequently the most important aspect in determining overall sentiment polarity in documents. The SVM classifier makes use of a diverse set of attributes that ranging across multiple levels of language description. Furthermore, the SVM technique's goal is to discriminate categories each new data that enters within it, as well as to enhance the margin between the two labels. The SVM algorithm can find a hyperplane that splits the dataset into two groups [28]. The other way to elaborate this, the SVM are "the data points nearest to the hyperplane" and whether eliminate the changed position of the separating hyperplane, hence in a dataset the support vectors are considered as critical factors. Therefore, the hyperplane can be termed as "the further from the hyperplane our data points lie" and "a line that linearly separates and classifies a set of data", in the data points the maximum possibilities are accurately classified the data. The SVM have more benefits as it predict the very accurately and the dataset are more concise and smaller also it execute better. Furthermore, it can be utilized to classify or determine the number in a very flexible way. Also, support vector machines is more memory efficient and it contain the ability to handle large dimensional spaces.

The binary classification is performed utilizing Support Vector Machine (SVM) machine learning strategy. The major ideas is to discover a superior hyperplane which divide data properly into two labels.

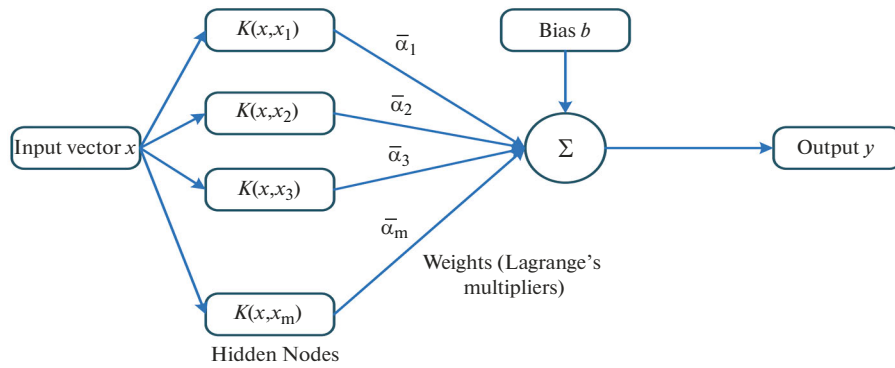


Fig. 3. Architecture of the SVM classifier.

Furthermore, combining multiple binary SVMs to form a multi-label classification model. Figure 3 illustrates the basic architecture of the SVM classifier.

Let i be the training instances $\{x_i, y_i\}, i = 1, \dots, I$ each case consist of an input x_i as well as a label $y_i \in \{-1, 1\}$. Every hyper plane has a bias (b) and a weight vector (w) that may be calculated using the equation below (7) [29].

$$wx + b = 0. \tag{7}$$

The following equation can be used to define the hyper plane function that classifies training and testing data (8),

$$f(x) = \text{sgn}(wx + b). \tag{8}$$

Prior function can be given as equation when dealing with kernel function (9),

$$f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i k(x_i, x) + b \right), \tag{9}$$

where b is a bias, x_i represents the input of training case, N is the number of training instances, y_i denotes the corresponding label and the kernel function $K(x_i, x)$ is used to map the input vectors into an extended features space. The coefficients α_i are determined by two constraints stated using Eqs. (10) and (11).

$$0 \leq \alpha_i, i = 1, \dots, N, \tag{10}$$

$$\sum_{i=1}^N \alpha_i y_i = 0. \tag{11}$$

3.4.3. Hybrid LSTM-SVM. In this proposed model combines two techniques such as, LSTM and SVM. Figure 4 illustrates the hybrid LSTM-SVM architecture.

The proposed hybrid LSTM-SVM classifier contain four layers such as, input layer, hidden layer, FC layer and softmax layer. 80% of feature extracted data from count vectorizer is initially given to the LSTM input layer and which passes the input into hidden layer. The gated cell or gated unit refers to the LSTM's hidden layer. It has four layers, each of which interacts with the others to produce the cell's output as well as the cell's state. After that, both these things are passed onto the next hidden layer. RNN only have single neural network layer of tanh, but LSTM contain one tanh layer and three logistic sigmoid gate. These gates determine which information are required for the following cell and which information are eliminated. The finding ranges from 0 to 1, with "0" denoting "reject information" as well as "1" denoting "include information". After completed hidden layer process, the output is fed to the FC layer, it is used to discover the exact complete structures of the features detected by the lower layers in the network. These are generally sit at the upper of the network hierarchy, at that time, the input has been minimized to a dense illustration of features. Then each node in the FC layer learns its own set of weights on all of the nodes in the layer. Then the output from the FC layer is given to the softmax layer.

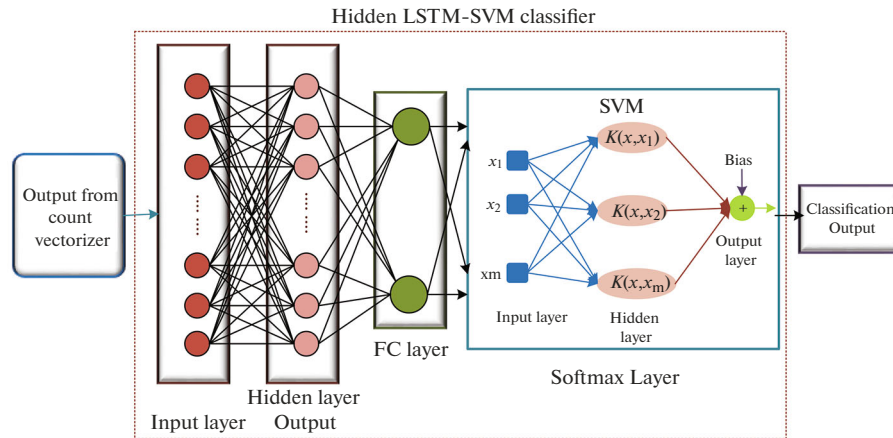


Fig. 4. Architecture of the Hybrid LSTM-SVM.

Generally the softmax layer is the final layer of the LSTM network. The softmax layer is a softmax function that generally used for multi-class classification and its formula is almost close to sigmoid function which is utilized for logistic regression. The softmax function can be utilized in a classifier only the classes are mutually exclusive. Thus, the proposed mode modify the softmax layer of the LSTM network and it use SVM classifier in the softmax layer for accurate prediction. The SVM perform the classification by finding the hyperplanes that differentiate the classes. Here, both techniques involved in classification and training phases. The considered document is given as input into the LSTM network, the statistical model of the LSTM neural network is constructed, and a feature vector is computed for each vector in the training set using the weights gained from the penultimate network layer. The training phase completed using SVM classifier. In the classification phases, the same stages are used for each vector that needs to be classed, and an embedding vector is obtained using the previously trained LSTM network. After completed training process, the trained model is tested with reaming 20% of feature extracted data. Lastly, the SVM classifier uses the embedding vector in combine with other classification features to provide predicted classes.

3.5. Compare Predicted Data with Preprocessed Data

After complete the classification process, the predicted data is compared with preprocessed data, which contain real information about COVID-19 pandemic and which is gathered from [31] like that dataset, these kinds dataset contain real information regarding to coronavirus such as, its symptoms, medicines, guidelines and so on which is provided from World Health Organization (WHO). If the outcome from the proposed LSTM-SVM classifier is similar to the preprocessed data, then the tweet information is considered as real news otherwise it is considered as fake news. In this way, the proposed Hybrid prediction model achieved a best accuracy for predicting fake news in Twitter about COVID-19 pandemic. Algorithm 1 illustrates the pseudo code of overall proposed model.

Algorithm 1: Pseudo Code for Social Networking Analysis using a Hybrid LSTM-SVM Classifier

Input dataset = X,

pre-processed data = Q,

Pre-processing

A = Splitting (X) # splits combined words are separated for individual processing

A1 = Stop word removal (A) # stop words such as "the", "an", "a" are removed

A2 = Lemmatization (A1) # Grouping of words together with different derivatives

A3 = Spell correction (A2) # Spell correction of wrong words

feature extraction

B = Count Vectorizer (A3) # convert text vectors into binary vector

```

# classification
C = LSTM-SVM (B) # classification utilizing Hybrid LSTM-SVM
If (Q = C) # compare predicted data with pre-processed data
{
  real news
}
Else
{
  fake news
}

```

Output = detect either real or fake news in the tweet information

4. RESULT AND DISCUSSION

The proposed Dynamic Light Weight Recommendation System for Social Networking Analysis using a Hybrid LSTM-SVM Classification model is executed in Matlab software to estimates its performance. The text classification is testing with some system configurations such as, Intel(R) Core(TM) i5-10300H Processor, CPU @ 2.50GHz, 16.0 GB Memory (RAM) and System type of 64-bit operating system. The real time dataset is collected from Twitter social media [30] and the dataset includes 5000 manually gathered Twitter commands in both fake and real news. The commands consist of uppercases, special characters and keywords. The preprocessing stage consist splitting, stop word removal, lemmatization and spell correction for extract the sentences. The preprocessing findings are given to the count vectorizer which extract the features as well as, convert text to binary vectors. The outcome from the count vectorizer is divided into two sets for training and testing the classifier. Initially, feature extracted 4000 commands are fed into the proposed hybrid LSTM-SVM classifier for train the model. Once completed training, the remaining 1500 feature extracted commands are then given into the trained model for testing the classifier. After that, the predicted classes are compared with preprocessed data. If the predicted class is similar to the preprocessed data, the news is considered as real otherwise considered as fake commands from Twitter social media.

Table 1 illustrates the extraction of input in the preprocessing phase, the data are initially given to the Splitting stage and this can split each and every word. After completed splitting process, the data are processed into stop word removal, in this step is to remove the stop words such as the, an, a, if, are etc. Lemmatization aim is to take away inflectional suffixes and prefixes to bring out the word's dictionary form. Finally the data are spell corrected and produced the preprocessed findings. Then these findings are given into the Count vectorizer for feature selection. After that these features are given into the classification process. The proposed model attained better performance compared to others.

4.1. Experimental Analysis

The experimental study on the proposed social media networking analyzing model was carried out using several performance metrics. Several metrics considered for this analysis are accuracy, error, sensitivity, specificity, precision, false positive rate (FPR), fl_score and kappa. These performance metrics are estimated among the proposed model (LSTM-SVM) and existing fake news detection model in Twitter social media. Several existing techniques based on fake news detection models that are considered for comparison analysis such as, Convolutional Neural Network and Support Vector Machine (CNN-SVM), bidirectional Gated Recurrent Neural Network (GRNN), Long Short Term Memory (LSTM) and Support Vector Machine (SVM). Table 2 shows the parameter values determined for the proposed and existing techniques for fake news detection strategies.

Figure 5 displays the accuracy and losses of the training process. The accuracy of the proposed method is 100%, since it was more rapidly detect the error and eliminate it more effectively. The error of the proposed method is 0% because the hybrid LSTM-SVM classification model attained accurate prediction. Figure 6 illustrates the confusion matrix obtained in this proposed model. Here, the X and Y labels represents the predicted class and true class. The proposed model predict two classes such as fake and real therefore 2×2 dimensional confusion matrix used. The comparison analysis based on accuracy (%) among the proposed and existing strategies depend on fake news detection on social media is shown in Fig. 7. The graphical representation is based on the several machine learning and deep learning techniques

Table 1. Input and Output attained in pre-processing

Input	Splitting	Stop_word	Lementize	Spell_correction
“My folks ordered pizza and now they doing the most to make sure its corona free, bitch why did yall order pizza then”	“My” “folks” “ordered” “pizza” “and” “now” “they” “doing” “the” “most” “to” “make” “sure” “its” “corona” “free” “,” “bitch” “why” “did” “yall” “order” “then”	“Folks” “ordered” “pizza” “make” “sure” “corona” “free” “,” “bitch” “why” “yall” “order”	“Folk” “order” “pizza” “make” “sure” “corona” “free” “bitch” “why” “yall”	“Folk order pizza make sure corona free bitch why tall order pizza”
“The problem of pov-erty has now covered the cover of religion. The issue has changed. There is relief from corona. All is well”	“The” “prob-lem” “of” “poverty” “has” “now” “cov-ered” “the” “cover” “religion” “.” issue” “changed” “There” “is” “relief” “from” “corona” “All” “well”	“Problem” “pov-erty” “covered” “cover” “religion” “.” “issue” “changed” “relief” “corona” “well”	“Problem” “pov-erty” “cover” “reli-gion” “issue” “change” “relief” “corona” “well”	“Problem poverty cover cover religion issue change relief corona well”
“Neighbours, if u can hear me coughing, it’s because my water went down the wrong hole and NOT because I have corona”	“Neighbours” “,” “if” “u” “can” “hear” “me” “coughing” “it’s” “because” “my” “water” “went” “down” “the” “wrong” “hole” “and” “NOT” “I” “have” “corona”	“Neighbours” “,” “u” “hear” “coughing” “s” “water” “went” “down” “wrong” “hole” “corona”	“Neighbour” “u” “hear” “cough” “s” “water” “go” “down” “wrong” “hole” “corona”	“Neighbor hear cough water down wrong hole corona”
“Corona its like a strange unreal horror movie everybody out there stay strong stay home and take care much love for all of you”	“Corona” “its” “like” “a” “strange” “unreal” “horror” “movie” “every-body” “out” “there” “stay” “strong” “home” “and” “take” “care” “much” “love” “for” “all” “of” “you”	“Corona” “like” “strange” “unreal” “horror” “movie” “everybody” “stay” “strong” “home” “take” “care” “love”	“Corona” “like” “strange” “unreal” “horror” “movie” “everybody” “stay” “strong” “home” “take” “care” “love”	“Corona like strange unreal horror movie everybody stay strong stay home take care love”
“Those white butter-fly clearly gave us a clear signal that something bad is going to happen.... Corona Virus”.	“Those” “white” “butterfly” “clearly” “gave” “us” “a” “clear” “signal” “that” “something” “bad” “is” “going” “to” “happen” “.” “Corona” “Virus”	“White” “butterfly” “clearly” “gave” “clear” “signal” “something” “bad” “going” “happen” “...” “Corona” “Virus” “.”	“White” “butterfly” “clearly” “give” “clear” “signal” “something” “bad” “go” “happen” “corona” “virus”	“White butterfly clearly give clear sig-nal something bad happen corona virus”

Table 2. Parameters Estimated for proposed and existing techniques in fake news detection

Parameters	LSTM-SVM (proposed)	CNN-SVM	GRNN	LSTM	CNN	SVM
Accuracy, %	90	87	83	71	70	65
Sensitivity, %	88	80	79	78	68	60
Specificity, %	97	92	82	68	68	60
Error	0.1	0.3	0.17	0.28	0.30	0.35
Precision, %	98	90	73	67	60	55
F1_Score, %	90	80	73	71	62	52
FPR	0.03	0.1	0.13	0.28	0.30	0.35
Kappa, %	80	79	62	55	5	48

and value of accuracy in percentage on both X and Y labels respectively. The accuracy is found for the proposed method is 90% which is greater compared to existing methods such as CNN-SVM is 87%, GRNN is 83%, LSTM is 71%, CRNN is 70% and SVM is 65%. This show that the performance of the proposed model is better compared to others.

The comparison analysis based on sensitivity (%) between the proposed and existing approaches depend on fake news detection on social media is shown in Fig. 8. The graphical illustration is based on the several machine learning and deep learning techniques and value of sensitivity in percentage on both X and Y labels respectively. The sensitivity is found for the proposed method is 88% which is greater compared to existing methods such as CNN-SVM is 80%, GRNN is 79%, LSTM is 78%, CRNN is 68% and SVM is 60%. This can show the performance of the proposed model works better compared to others. The comparison analysis based on specificity (%) between the proposed and existing strategies depend on fake news detection on social media is shown in Fig. 9. The graphical illustration is based on the several machine learning and deep learning techniques and value of specificity in percentage on both X and Y labels respectively. The specificity is found for the proposed method is 97% which is greater compared to

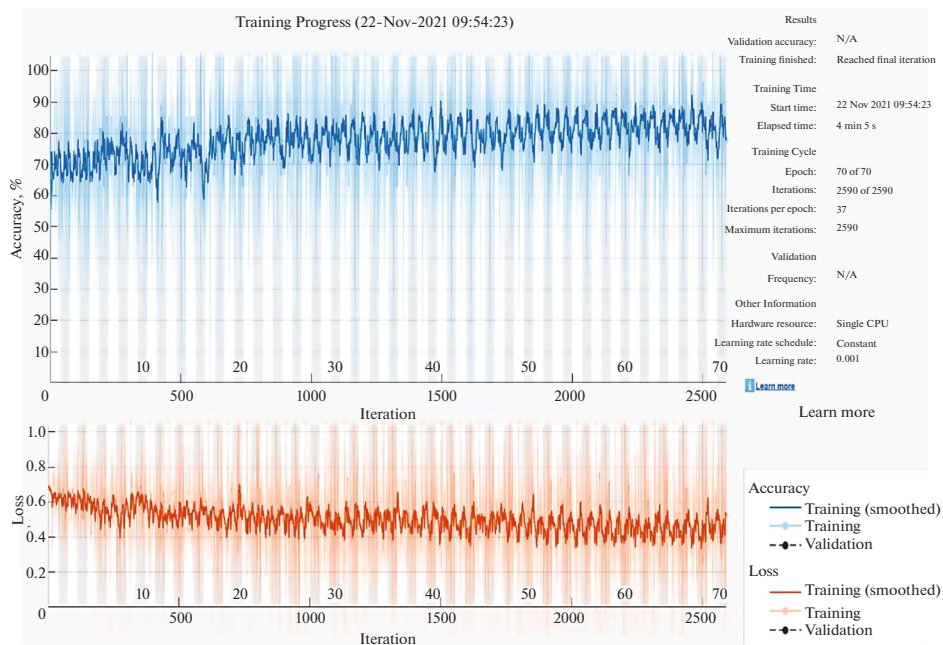


Fig. 5. Training phase.

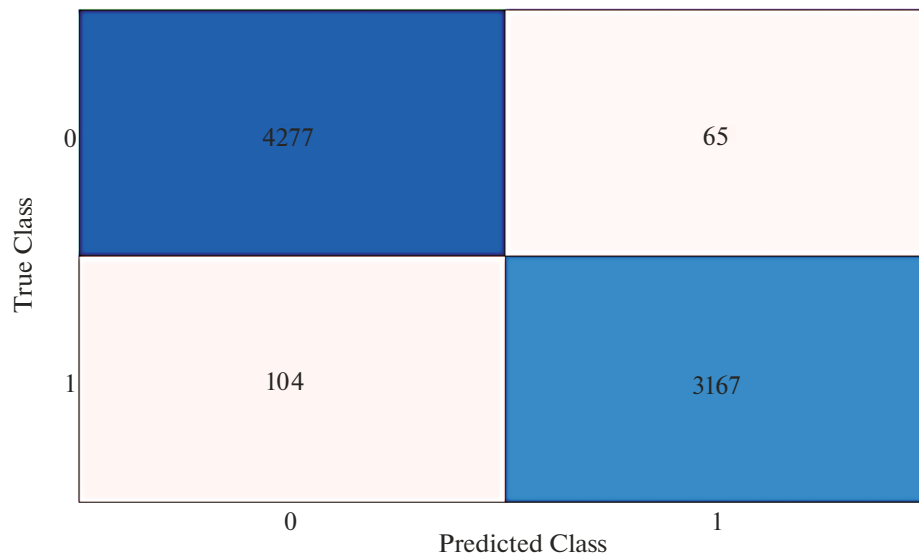


Fig. 6. Confusion matrix obtained for the proposed model.

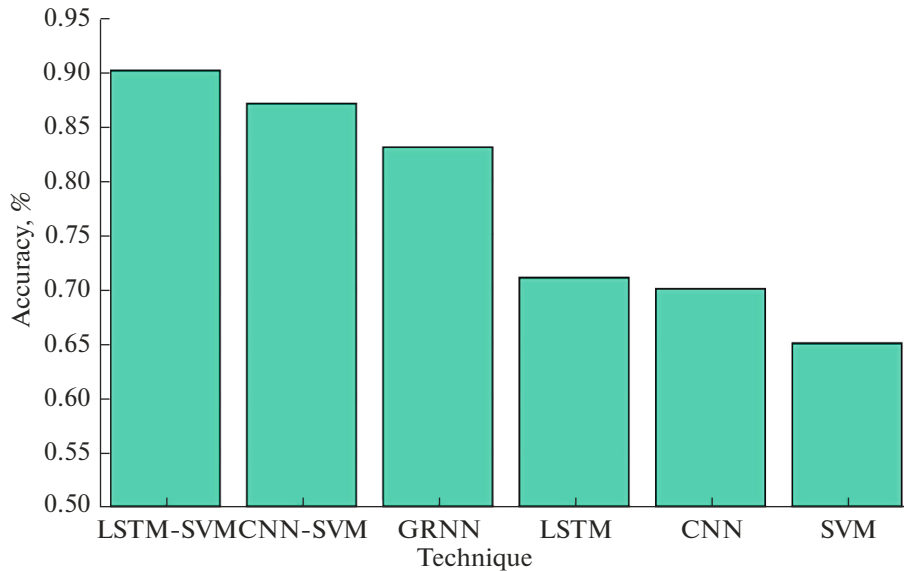


Fig. 7. Comparison of Accuracy among Proposed and Existing Approaches.

existing methods such as CNN-SVM is 92%, GRNN is 82%, LSTM is 68%, CRNN is 68% and SVM is 60%. This can show the performance of the proposed model is better compared to others.

The comparison analysis based on error between the proposed and existing techniques based on fake news detection on social media is shown in Fig. 10. The graphical illustration is based on the several machine learning and deep learning techniques and value of error in percentage on both X and Y labels respectively. The error is found for the proposed method is 0.1 which is less compared to existing methods such as CNN-SVM is 0.3, GRNN is 0.17, LSTM is 0.28, CRNN is 0.30 and SVM is 0.35. This can show the performance of the proposed model works better compared to others. Figure 11 displays the comparison analysis based on precision (%) among the proposed and existing strategies depend on fake news detection on social media. The graphical illustration is based on the several machine learning and deep learning techniques and value of precision in percentage on both X and Y labels respectively. The precision is found for the proposed method is 98% which is greater compared to existing methods such as CNN-SVM is

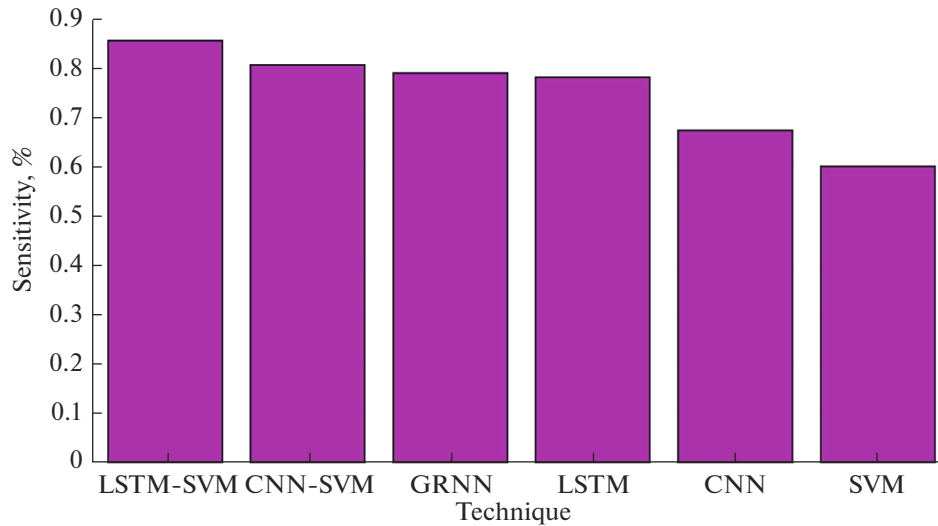


Fig. 8. Comparison of Sensitivity among proposed and existing approaches.

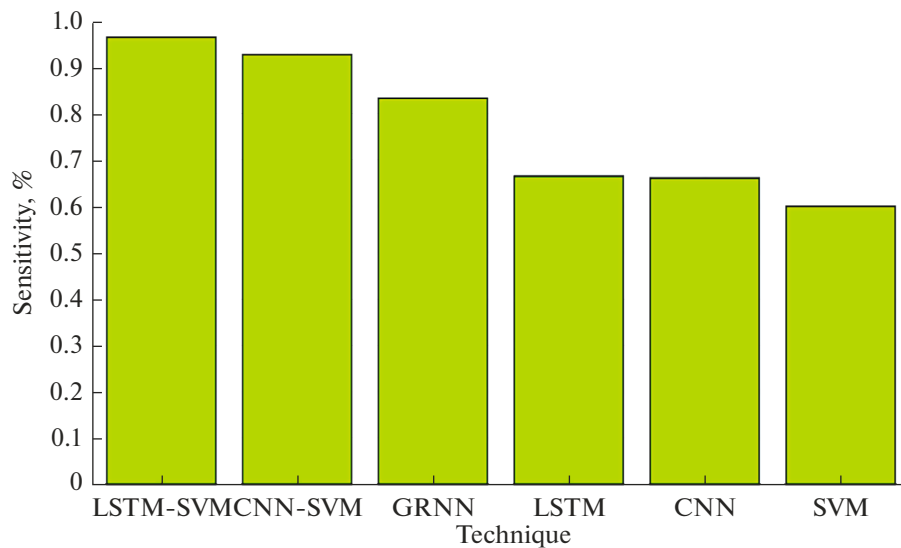


Fig. 9. Comparison of Specificity among Proposed and Existing Approaches.

90%, GRNN is 73%, LSTM is 71%, CRNN is 62% and SVM is 52%. This shows that the performance of the proposed model is better compared to others.

Figure 12 illustrates the comparison analysis based on F1_score (%) among the proposed and existing strategies depend on fake news detection on social media. The graphical illustration is based on the several machine learning and deep learning techniques and value of F1_score in percentage on both X and Y labels respectively. The F1_score is found for the proposed method is 90% which is greater compared to existing methods such as CNN-SVM is 80%, GRNN is 73%, LSTM is 71%, CRNN is 62% and SVM is 52%. This can show the performance of the proposed model works better compared to others. Figure 13 illustrates the comparison analysis based on FPR (%) among the proposed and existing strategies depend on fake news detection on social media. The graphical illustration is based on the several machine learning and deep learning techniques and value of FPR in percentage on both X and Y labels respectively. The FPR is found for the proposed method is 0.03 which is greater compared to existing methods such as CNN-SVM

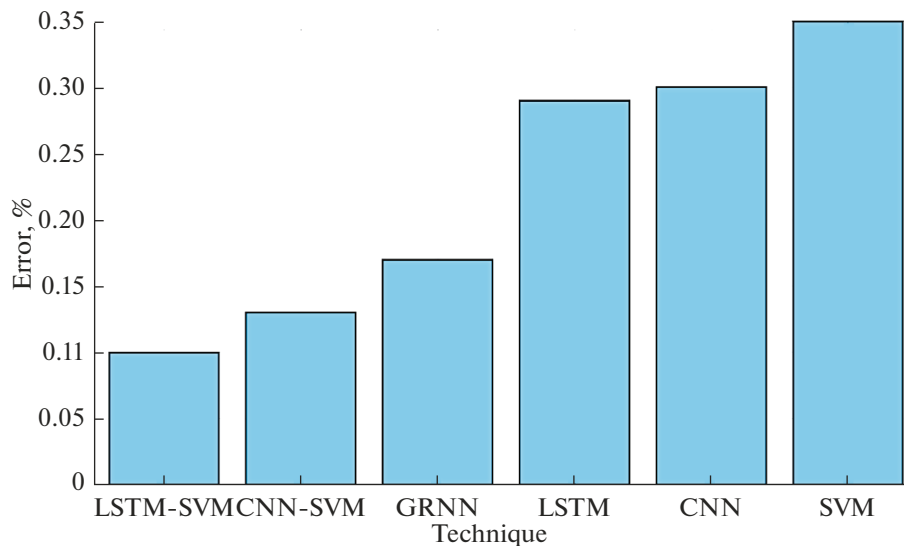


Fig. 10. Comparison of Error among Proposed and Existing Approaches.

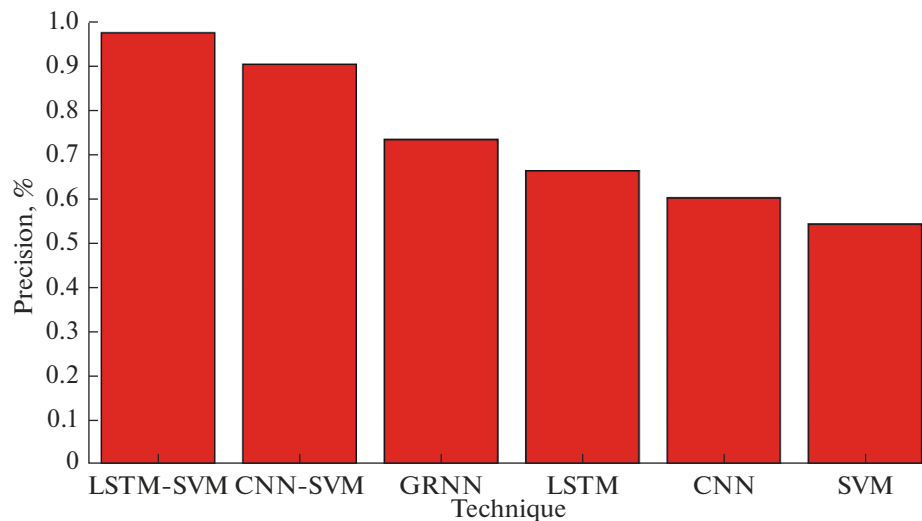


Fig. 11. Comparison of Precision among Proposed and Existing Approaches.

is 0.1, GRNN is 0.13, LSTM is 0.28, CRNN is 0.30 and SVM is 0.35. This shows better performance of the proposed model on comparison with existing.

Figure 14 illustrates the comparison analysis based on kappa (%) among the proposed and existing strategies depend on fake news detection on social media. The graphical illustration is based on the several machine learning and deep learning techniques and value of accuracy in percentage on both X and Y labels respectively. The accuracy is found for the proposed method is 80% which is greater compared to existing methods such as CNN-SVM is 79%, GRNN is 62%, LSTM is 55%, CRNN is 5% and SVM is 48%. This shows better performance of the proposed model on compared with others.

5. CONCLUSIONS

A fake news detection model utilizing a hybrid LSTM-SVM classifier is designed to improve the performance of the fake news prediction. The real-time data are collected from Twitter social media. These data are given into preprocessing, and it performed splitting, stop word removal, lemmatization and spell

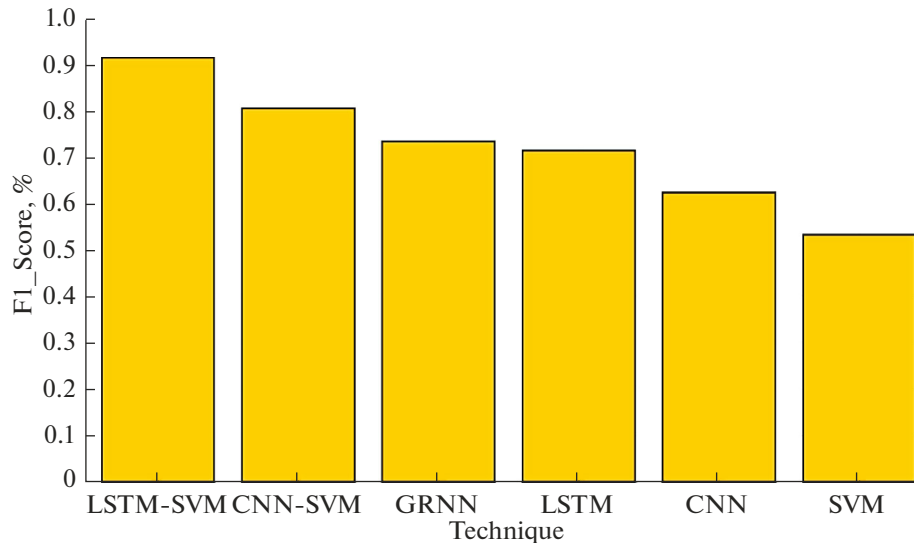


Fig. 12. Comparison of F1_Score among Proposed and Existing Approaches.

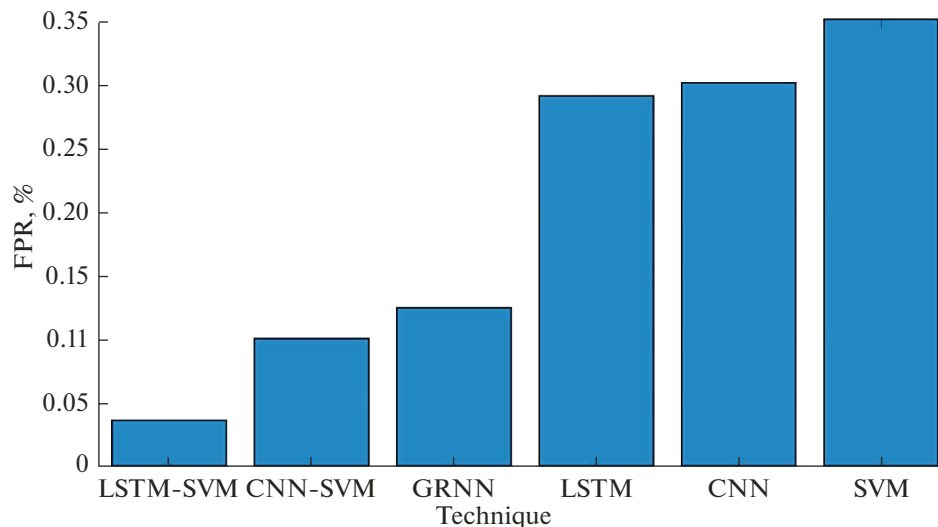


Fig. 13. Comparison of FPR among Proposed and Existing Approaches.

correction. A count vectorizer is utilized for text feature selection and binary conversion. After extracting features, 80% of data are initially given for training the hybrid LSTM-SVM classification model. Then the trained model is tested with remaining 20% of extracted features, and the model produces the classes effectively. The predicted data is compared with the preprocessed data, which consist of real information. If the predicted data is equal to the preprocessed data, then it will be real news or else fake news. The proposed model is implemented to estimate the performance metrics such as accuracy is 90%, sensitivity is 88%, specificity is 97%, the error is 0.1, precision is 98%, F1_Score is 90%, FPR is 0.03, and kappa is 80%. The overall expected outcome of the recommendation system using hybrid LSTM-SVM is better compared to the existing techniques such as CNN-SVM, GRNN, LSTM and SVM. The proposed social networking analysis model delivers effective fake news detection that can be utilized to detect the Twitter comments related to the COVID-19 pandemic, either real or fake, and this can be used for the user to know the reality of the news for aware about the fake news.

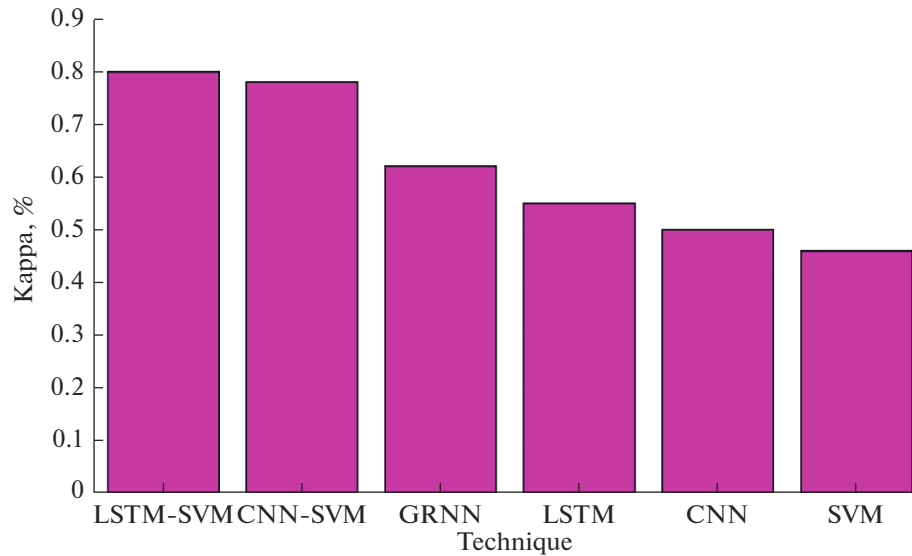


Fig. 14. Comparison of Kappa among Proposed and Existing Approaches.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., and Akbar, M., Fake news detection using deep learning models: A novel approach, *Trans. Emerging Telecommun. Technol.*, 2020, vol. 31, no. 2, p. e3767.
2. Arulmurugan, R., Sabarmathi, K.R., and Anandakumar, H.J.C.C., Classification of sentence level sentiment analysis using cloud machine learning techniques, *Cluster Comput.*, 2019, 22, no. 1, pp. 1199–1209.
3. Dong, X., Victor, U., and Qian, L., Two-path deep semisupervised learning for timely fake news detection, *IEEE Trans. Comput. Soc. Syst.*, 2020, vol. 7, no. 6, pp. 1386–1398.
4. Choudhary, M., Chouhan, S.S., Pilli, E.S., and Vipparthi, S.K., BerConvoNet: A deep learning framework for fake news classification, *Appl. Soft Comput.*, 2021, vol. 110, p. 107614.
5. Sabeeh, V., Zohdy, M., Mollah, A., and Al Bashaireh, R., Fake news detection on social media using deep learning and semantic knowledge sources, *Int. J. Comput. Sci. Inf. Secur. (IJCSIS)*, 2020, vol. 18, no. 2.
6. Agarwal, A., Mittal, M., Pathak, A., and Goyal, L.M., Fake news detection using a blend of neural networks: an application of deep learning, *SN Comput. Sci.*, 2020, vol. 1, no. 3, pp. 1–9.
7. Nasir, J.A., Khan, O.S., and Varlamis, I., Fake news detection: A hybrid CNN-RNN based deep learning approach, *Int. J. Inform. Manage. Data Insights*, 2021, vol. 1, no. 1, p. 100007.
8. Kaliyar, R.K., Goswami, A., and Narang, P., DeepFakeE: improving fake news detection using tensor decomposition-based deep neural network, *J. Supercomputing*, 2021, vol. 77, no. 2, pp. 1015–1037.
9. Jadhav, S.S. and Thepade, S.D., Fake news identification and classification using DSSM and improved recurrent neural network classifier, *Appl. Artif. Intell.*, 2019, vol. 33, no. 12, pp. 1058–1068.
10. Ahmad, I., Yousaf, M., Yousaf, S., and Ahmad, M.O., Fake news detection using machine learning ensemble methods, *Complexity*, 2020.
11. Akinyemi, B., Adewusi, O., and Oyebade, A., An Improved Classification Model for Fake News Detection in Social Media, *Int. J. Inform. Technol. Comput. Sci. (IJITCS)*, 2020, vol. 12, no. 1, pp. 34–43.
12. Abdullah, A., Awan, M., Shehzad, M., and Ashraf, M., Fake news classification bimodal using convolutional neural network and long short-term memory, *Int. J. Emerg. Technol. Learn.*, 2020, vol. 11, pp. 209–212.
13. Hakak, S., Alazab, M., Khan, S., Gadekallu, T.R., Maddikunta, P.K.R., and Khan, W.Z., An ensemble machine learning approach through effective feature extraction to classify fake news, *Future Gener. Comput. Syst.*, 2021, vol. 117, pp. 47–58.
14. Abdi, A., Shamsuddin, S.M., Hasan, S., and Piran, J., Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion, *Inf. Process. Manage.*, 2019, vol. 56, no. 4, pp. 1245–1259.
15. Khan, J.Y., Khondaker, M.T.I., Afroz, S., Uddin, G., and Iqbal, A., A benchmark study of machine learning models for online fake news detection, *Mach. Learn. Appl.*, 2021, vol. 4, p. 100032.

16. Sulthana, A.R. and Ramasamy, S., Ontology and context based recommendation system using neuro-fuzzy classification, *Comput. Electr. Eng.*, 2019, vol. 74, pp. 498–510.
17. Turkut, Ü., Tuncer, A., Savran, H., and Yilmaz, S., June. An online recommendation system using deep learning for textile products, in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, IEEE, 2020, pp. 1–4.
18. Jelodar, H., Wang, Y., Orji, R., and Huang, S., Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach, *IEEE J. Biomed. Health Inf.*, 2020, vol. 24, no. 10, pp. 2733–2742.
19. Ali, F., Kwak, D., Khan, P., Ei-Sappagh, S.H.A., Islam, S.R., Park, D., and Kwak, K.S., Merged ontology and SVM-based information extraction and recommendation system for social robots, *IEEE Access*, 2017, vol. 5, pp. 12364–12379.
20. Qin, D., Zhou, X., Chen, L., Huang, G., and Zhang, Y., Dynamic connection-based social group recommendation, *IEEE Trans. Knowl. Data Eng.*, 2018, vol. 32, no. 3, pp. 453–467.
21. Monti, F., Frasca, F., Eynard, D., Mannion, D., and Bronstein, M.M., Fake news detection on social media using geometric deep learning, *arXiv preprint arXiv:1902.06673*, 2019.
22. Sahoo, S.R. and Gupta, B.B., Multiple features based approach for automatic fake news detection on social networks using deep learning, *Appl. Soft Comput.*, 2021, vol. 100, p. 106983.
23. Lu, Y.J. and Li, C.T., GCAN: Graph-aware co-attention networks for explainable fake news detection on social media, *arXiv preprint arXiv:2004.11648*, 2020.
24. Hossny, A.H., Mitchell, L., Lothian, N., and Osborne, G., Feature selection methods for event detection in Twitter: a text mining approach, *Soc. Network Anal. Min.*, 2020, vol. 10, no. 1, pp. 1–15.
25. Kulkarni, A. and Shivananda, A., Converting text to features, in *Natural Language Processing Recipes*, Berkeley, CA: Apress, 2021, pp. 63–106.
26. Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., and On, B.W., Fake news stance detection using deep learning architecture (CNN-LSTM), *IEEE Access*, 2020, vol. 8, pp. 156695–156706.
27. Lv, S., Zhang, H., He, H., and Chen, B., Microblog rumor detection based on comment sentiment and CNN-LSTM, in *Artificial Intelligence in China*, Singapore: Springer, 2020, pp. 148–156.
28. Ghanem, B., Rosso, P., and Rangel, F., An emotional analysis of false information in social media and news articles, *ACM Trans. Internet Technol. (TOIT)*, 2020, vol. 20, no. 2, pp. 1–18.
29. Tharwat, A., Parameter investigation of support vector machine classifier with kernel functions, *Knowl. Inf. Syst.*, 2019, vol. 61, no. 3, pp. 1269–1302.
30. KAGGLE. <https://www.kaggle.com/c/sentiment-analysis-of-covid-19-related-tweets/data>.
31. KAGGLE. <https://www.kaggle.com/iamhungundji/covid19-symptoms-checker>.