ORIGINAL RESEARCH

Fake News Detection Using Hybrid Deep Learning Method

Arun Kumar Yadav[1] · Suraj Kumar[1] · Dipesh Kumar[1] · Lalit Kumar[1] · Kapil Kumar[1] · Sandeep Kumar Maurya[1] · Mohit Kumar[1] · Divakar Yadav[2]

Abstract

The growth of online social networks platforms in recent years has resulted in the widespread dissemination of social news such as commercial adverts, political news, celebrity gossip, and more. This proliferation has resulted in distribution of fake news, that is many times disguised as, genuine real news. This paper evaluates several machine learning and deep learning approaches to automatically identify fake news from headlines and news descriptions. As there is no single large, standard dataset available for fake news detection, this study derives a large dataset by integrating two publicly available datasets—*Fake and real news* and *allData*. There are 64,934 labeled news articles in the dataset. For classification, deep learning models such as Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Convolutional Neural Network-LSTM (CNN-LSTM), and CNN-BiLSTM were employed. On the consolidated dataset, the Word2Vec word-embedding technique yielded higher performance when combined with CNN-BiLSTM model, with accuracy, precision, recall, F1 measure, and AUC-ROC values of 0.975, 0.984, 0.970, 0.977, and 0.992, respectively. The proposed methodology outperforms recent state-of-the-art methods in recognizing fake news.

Keywords Fake news · Machine learning · Deep learning · GloVe · FastText · Word2Vec · CNN · BiLSTM

Introduction

The internet has become an inextricable element of our daily lives. In regard to the consumption of news, the more traditional information media such as newspapers and television are seeing a continued decline in their popularity. Certainly, the emergence of various social media platforms has had a substantial influence on this transformation. The currency of social networking websites such as Twitter [1] and Facebook [2] has grown exponentially. In November 2017, Facebook, for example, announced that it had 2.07 billion active users. On a daily basis, 1.37 billion of these people log on to Facebook. Since January 2018, Twitter has had 330 million users. Since the establishment of such relevant forums, these numbers have continued to rise as shown in Fig. 1.

An increasingly large number of people rely on various social media platforms, not only to communicate with friends and family, but also to keep up with current events and news from around the world. As a result, the success of news relies heavily on social media. According to [3], a British study has shown an increase in social media activity and, more importantly, its importance in media consumption. Furthermore, the news consumption also varies by age group of the users as shown in Fig. 2. According to Zubiaga et al. [4], social media has evolved into an essential publishing tool for journalists [5, 6], as well as a primary source of information for people [7]. Journalists can use social media to report on popular perception about current events and even to obtain new information, while citizens can follow the development of news and events through official channels of news outlets and their official social media accounts, along with content from their own circle, such as family and friends, on various social media platforms. Furthermore, due to their inherent capacity to propagate information far quicker than traditional media, social networks have shown to be quite valuable, particularly during several crises [8].
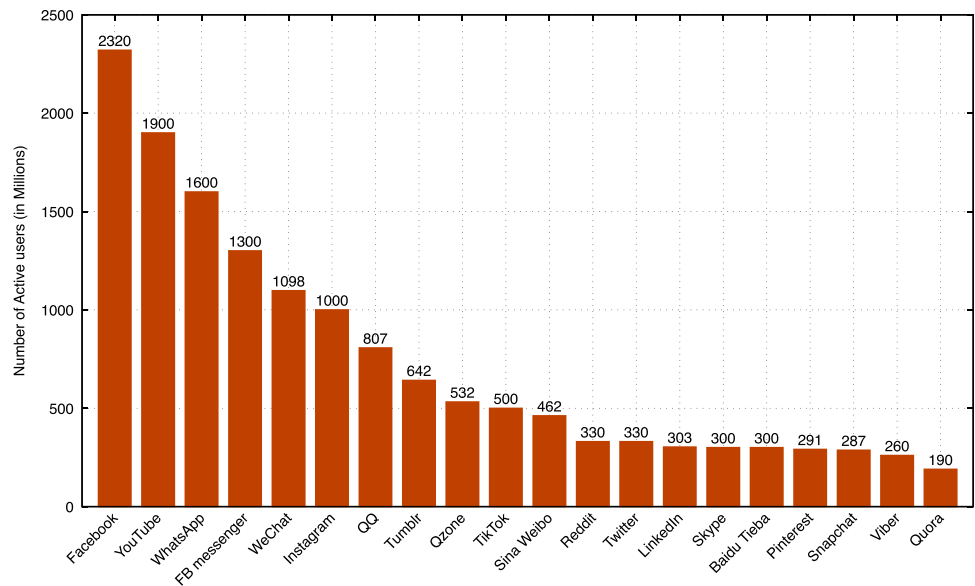
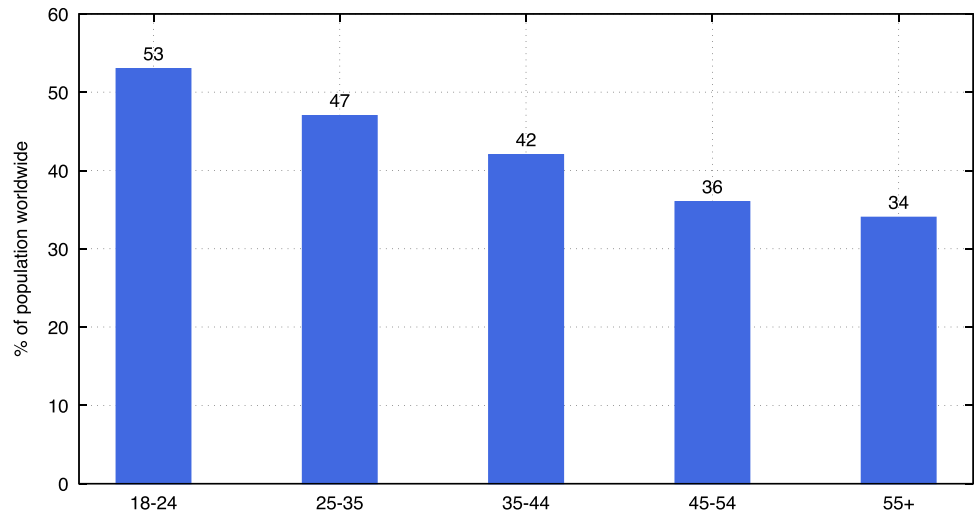✉ Divakar Yadav
  dsy99@rediffmail.com

1 Dept of Computer Science & Engineering, NIT Hamirpur, Hamirpur, HP, India

2 School of Computer & Information Sciences, IGNOU, New Delhi, India

Fig. 1 Monthly active users of widely used social networking websites and platforms [9]

Fig. 2 Use of social media platforms as a news outlet by age group [9]

However, social media's beneficial influence comes at a price: a lack of control and trustworthiness—posting content, news, and reviews makes social media a fecund platform for propagating unconfirmed and/or misleading material [4]. People frequently publish or share other people's postings without confirming the source, legitimacy, or reliability of such content. Most of the times, an intriguing headline is enough to have an article shared thousands of times, even if the substance is possibly incorrect or inaccurate.

Fake news may be yellow journalism that uses the media and the internet to spread purposeful or deliberate misinformation and falsehood. Fake news, however, is not a digital-era invention. Even before the Internet, journalists were tasked with verifying and assessing the accuracy of news and sources, therefore limiting public opinion on false information. The authors in [10, 11], for example, state that in 2016, a bogus rumor about Hillary Clinton being involved in child trafficking prompted a man to shoot a pizzeria that was believed to be one of the trafficking hot spots. Today's social media platform promotes the spread of unconfirmed and incorrect information to a wide number of users, having a significant impact on global perception and comprehension of events [4]. The 2016 US presidential campaign was probably one of the most prominent examples of how misleading news may impact belief of the general public. The authors of [12] examined the article and discovered several intriguing conclusions, such as voters were exposed to a strong pro-Trump viewpoint during the election campaign. Furthermore, republican voters were more likely to trust in both true and fake news headlines, according to polls [12]. However, it may be safe to conclude that fake news, and misinformation in general, is becoming a huge problem on the internet and therefore, society. As a direct outcome, both the social media platforms and the academic community, along with various

independent journalist groups, are quite active in finding and validating potentially fraudulent statements. Facebook, for example, has released a series of recommendations that would assist people avoid becoming victims of fake news articles.

Fake news online has become a considerable concern in recent years (Fig. 3) because of the proliferation of online media platforms. People are frequently unable to devote sufficient time and expertise to examining clues and evaluating news trustworthiness, as seen in Fig. 4. Consequently, given the large number of online content consumers and suppliers, automatic identification of fake news is considered the only viable option to tackle this problem. Thus, fake news detection has become an active research topic in academic circles and industry.

This research study proposes a novel deep learning-based strategy for automatic detection of fake news after evaluating several machine learning and deep learning approaches. Experiments using 16 distinct configurations of machine learning models were conducted and evaluated. In terms of deep learning models, model configurations with Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and CNN-BiLSTM models along with 3 different word-embedding techniques (a total of 12 distinct configurations) were utilized for experiments and thereafter evaluated against state-of-the-art approaches.

The major contributions of proposed work are listed as follows:

Fig. 3 Number of articles produced on fake news and rumors [13]
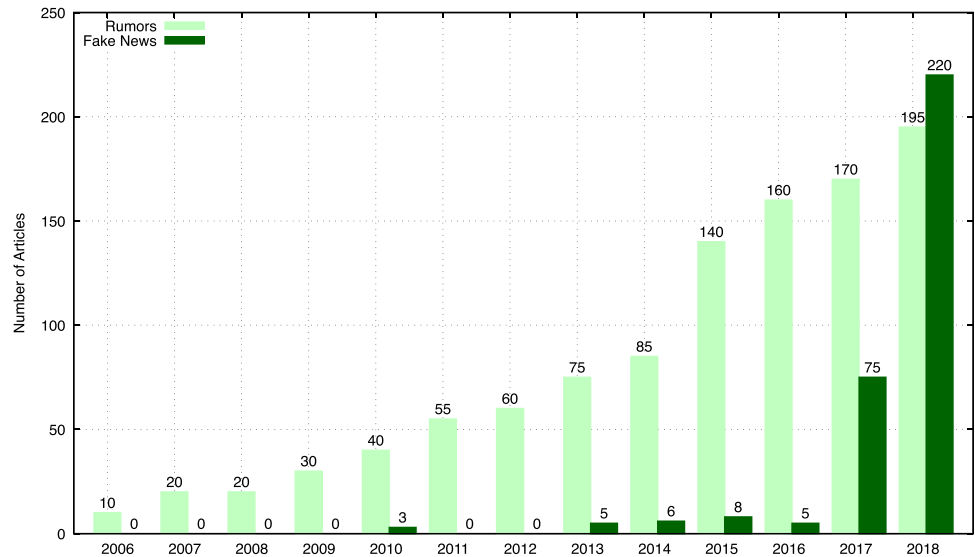
Fig. 4 Awareness of people about the accuracy of news [9]
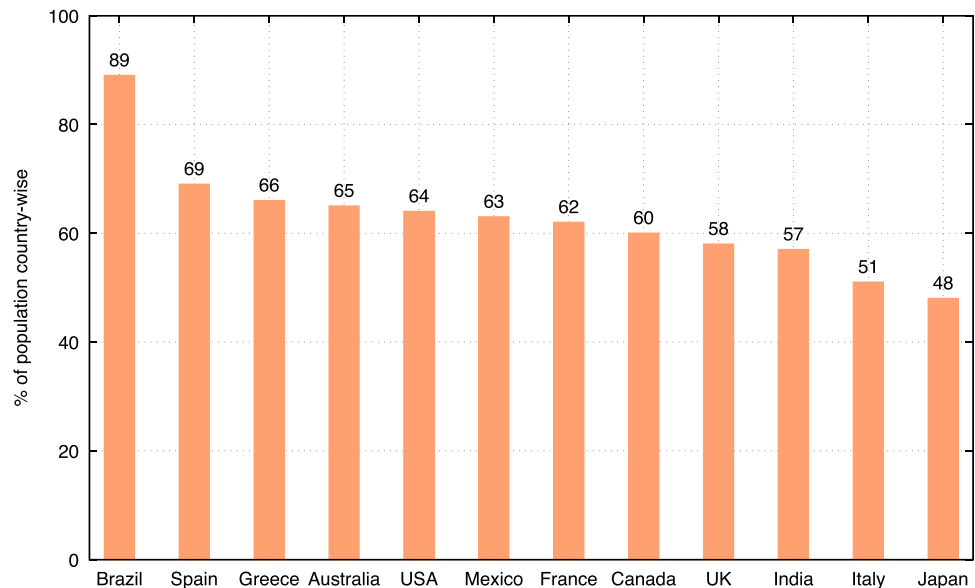
- A novel deep learning-based approach is presented to automatically detect fake news. The proposed approach significantly outperforms the existing state-of-the-art approaches for fake news detection.
- Evaluation of 16 configurations of machine learning techniques and 12 configurations of deep learning techniques is presented.
- The evaluation of the proposed, along with the competing approaches, is done using a consolidated dataset consisting of 64,934 news records. Two publicly available datasets—*Fake and real news* and *allData*—were integrated to create the final dataset for experiments.

The rest of the paper is organized as follows. Section "Literature Survey" of the paper briefly reviews the recent additions to the literature, with an emphasis on data processing, machine learning, and deep learning approaches. The proposed pre-processing approach and model are described in Sect. "Methodology". Section "Results and Analysis" summarizes the findings and analysis based on the various methods. Section "Comparison with State of the Art" compares the proposed approach's outcomes to current state-of-the-art methodologies. Conclusions and future scope of the study are presented in Sect. "Conclusion and Future Work".

Literature Survey

Fake news is not a new phenomenon. Presently, however, it propagates mainly through social media platforms, similar to other related issues like cyberbullying [14] and hate speech [15, 16], with a concentration on breaking events such as criminal gossip, satirical news, natural calamities, and the like. The emergence of fake news has been related to the availability of digital mass media technology. The unverified information may be the part of fake news, and it plays an important role to influence the society connected through social media platforms such as Facebook and Twitter. With the increasing dependency on social media, fake news transforms from textual to visual forms and increase its impact on society. It contains emotionally triggering language and targets the users of the low emotional intelligence. The emotional content-based fake news easily targets the users and it achieves its objective [17, 18]. According to the study in the article [19–21], higher emotional intelligence people can differentiate the fake and real news. A research is carried out to identify the emotional intelligence in the article [22]. The authors proposed a study to identify role of emotional intelligence in identifying fake news articles from Facebook. In this study, they concluded that people with high emotional intelligence fall less for fake news than people with less emotional intelligence. A psychological characteristics

of text of social networking platform concluded that people use manipulative language that focus on dual meaning and away from original information [23]. Also, they concluded that a fake news contains less motion verbs as compared to real news. Again, in the paper [24–26], the authors discuss the role of sentiments in increasing the fake news on social media and other platform. They concluded that the sender feels more tension, and they contain negative emotions as compared to real news. Since anybody can contribute news pieces to digital media platforms, online news items feature well-researched snippets as well as opinion-based arguments that are just plain false information. As there is no curator and/or moderator regarding the reliability standards for the material on these sites, false news may propagate easily. To make problems worse, it is difficult to tell the difference between factual news and half-true or false news. To address the issue of fake news, automatic identification has been attempted in the recent years using deep learning and machine learning techniques as identified in the following brief literature review.

In the paper [27], authors discuss the machine learning techniques to detect false news in Russian language. They used TV shows and fact-checking websites as dataset and Support Vector Machine (SVM) model as the classifier. They reported accuracy of 0.600. In the paper [28], authors proposed Vector Space model on TV show transcripts. They reported the 0.630 accuracy on TV show dataset. Again, in the paper [29], authors proposed Credibility Propagation Network of Tweets (CPNT) model on *Sina Weibo* dataset (real-time self-made dataset). This study reported their performance using different popular metrics such as accuracy (0.840), precision (0.786), and recall (0.933). They commented that proposed approach outperform and obtained significant improvements over the state-of-the-art approaches. In the paper [30], the authors proposed FakeFlow model by combining topic and effective information extracted from text. They evaluated the model's performance through several experiments on real-world datasets. They achieved the F1-score of 0.960 on *TruthShades*, 0.88 on *PoliticalNews*, and 0.85 on *FakeNewsNet* datasets. They reported that the results of the proposed model are superior when compared against state-of-the-art methods. In the paper [31], the authors discuss the fake news detection using BERT (Bidirectional Encoder Representations from Transformers) model with 0.860 accuracy. They evaluated the results on multiple datasets and concluded that proposed approach outperforms on state-of-the-art methods on all evaluation matrices. In the paper [32], the authors proposed the linguistic feature-driven model for fake news detection. They reported accuracy of 0.860 on the *fakenewsdata1* dataset. Also, they compared their proposed model with machine learning and LSTM-based deep learning models. In the paper [33], the authors discuss the detailed analysis of fake news detection

on *LIAR* dataset and concluded that Naive Bayes perform better as compared to other machine learning methods on mentioned dataset. In the paper [34], the authors proposed that fake news was detected using Bidirectional LSTM and Recurrent Neural Network (RNN) models. They evaluated the proposed model on two publicly available, unstructured datasets. They reported that proposed model shows an accuracy of 0.980. In the paper [35], the authors proposed multiple languages and platforms-based fake news detection using k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) methods. They used three datasets (*TwitterBR*, *Fake-BrCorpus*, and *btvlifestyle*) and reported 0.910 accuracy. In the paper [36], the authors proposed automatic filtering of fake news detection in Portuguese language. The proposed approach used SVM model to evaluate the accuracy on *Fake. Br* (self-made) dataset. In the paper [37], the authors discuss the approaches to detect malicious profiles on Twitter using Random Forest method. They reported that their proposed model got accuracy of 0.990. In the paper [38], the authors discuss the fake news, its effect on society and individual, challenges and opportunities to work toward its detection. They achieved 0.990 accuracy in their experiments. Again in the paper [39], the authors discuss the detailed study on fake news detection and its implications along with the future directions.

In recent years, deep learning is also playing an important role in fake news detection. A hybrid deep model is proposed in the paper [40] for automatic detection of fake news. In this paper, the authors used CSI (Capture, Score, and Integrate) model along with two real-world datasets from Twitter and Weibo. They commented that their proposed model reported 0.890 accuracy. An SRSR (Seriously Rapid Source Review) method is developed to detect fake and real information in journalism [41]. In this paper, the authors apply CNN-based model on SRSR system on *fake and real news* dataset. They reported 0.92 precision and 0.92 recall. In the paper [42], the authors proposed the WeFEND (WEakly-supervised FakE News Detection) model, a reinforcement learning-based method. In this approach, the authors used their proposed model on *WeChat* dataset and reported 0.824 accuracy. A deep CNN-based algorithm for fake news detection is proposed in the paper [43]. The proposed method is evaluated on *fake news* dataset and reported 0.980 accuracy. In the paper [44], the authors proposed deep learning-based model for online fake news detection from multiple domains. They evaluated the proposed model on recently released datasets—*FakeNewsAMT* and *Celebrity*—and reported 0.830 accuracy on *FakeNewsAMT* and 0.790 accuracy on *Celebrity*. They commented that proposed system outperforms the current handcrafted feature engineering-based state-of-the-art system by significant margins of 3.08% and 9.3%, respectively.

A CNN-RNN-based hybrid deep learning model is proposed in the paper [45] for fake news detection. In this paper, the authors evaluated their proposed model on two new datasets (*ISO* and *FA-KES*) and reported 0.600 accuracy which is significantly better than other non-hybrid baseline models. In the paper [46], a domain-specific pre-trained model is proposed. The proposed model, self-ensemble SCIBERT (Scientific Bidirectional Encoder Representations from Transformers)-based model, is evaluated with the *Fake-Health* dataset. The reported accuracy is 0.690 and precision is 0.720. Again, in the paper [34], the authors present BiLSTM-based model for fake news detection and reported 0.960 accuracy. In the paper [30], the authors used the Longformer model on *FakeNewsNet* dataset and reported an accuracy of 0.970. In the paper [43], the authors proposed deep learning-based methods with *MediaEval*, *GossipCop*, and *PolitiFact* datasets. They reported 0.920 accuracy.

Several crucial study findings were uncovered after looking at past studies on automatic fake news detection. First, it is noticed that machine learning approaches were utilized in the majority of research to detect false news, which necessitated a great deal of feature engineering. Because of the numerous privacy regulations on social media sites, collecting user information is not always viable. This research investigates whether pre-trained models, rather than hand-made characteristics, may be utilized for representation of words. Second, there are not many large datasets to detect phoney news. When it comes to learning algorithms, it is common knowledge that larger datasets tend to generalize well and yield more dependable conclusions, especially with deep learning methods. As a result, the current study examines whether larger datasets may be created by intelligently integrating publicly accessible, comparatively smaller datasets and then used in experiments to acquire more accurate and dependable results? Third, the majority of the research reported their findings in terms of one or two assessment measures such as accuracy, recall, and precision. When it comes to spotting false news, accuracy may not be the only statistic to consider; recall, precision, and F-score are also essential considerations. As a result, improved performance, as measured via most of the popular metrics is desired, as we do not want to mistakenly label any legitimate news as false news. This research explores whether reporting results using all accessible metrics can offer a more fair and just evaluation of the fake news detection system. Table 1 shows the recent work on machine learning and deep learning methods for detecting fake news.

To address the above identified research points, we employed a variety of machine learning and deep learning methodologies to tackle the first point. In addition, to address the second issue, we consolidated two publicly accessible datasets to have enough input data to train and validate the proposed system. With a large number of fake

Table 1 Recent work on fake news detection using machine learning and deep learning methods

S. No.	Year	Model	Dataset	Accuracy
1	2015 [27]	SVM	TV shows, fact-checking websites	0.650
2	2016 [29]	CPCV model	Sina Weibo	0.840
3	2017 [40]	CSI	Social media datasets Twitter	0.892
4	2018 [47]	MMFD	LIAR	0.38
5	2019 [42]	LSTM	WeChat	0.824
6	2021 [34]	BiLSTM	Real_or_fake	0.91
7	2021 [48]	Hybrid CNN-RNN	FA-KESdataset	0.60
8	2021 [31]	Naive Bayes	Fake or real news dataset	0.86

and real news samples, we anticipate the proposed system's performance to be unaffected by data limits. Finally, to handle the third point, we analyzed and reported the experimental results using a variety of assessment metrics such as accuracy, precision, recall, F1-score, and AUC-ROC (area under the ROC curve) to evaluate the system as objectively as possible.

Methodology

Fake news detection may be posed as a binary classification problem. In this section, fake news detection is done on the basis of determining if the information in the article is accurate or not. In this regard, the current research makes a contribution by providing a comprehensive comparative analysis of several different models based on unique approaches that demonstrate the utility of deep learning models in detecting fake news. The analysis concludes that a CNN- and BiLSTM-based hybrid approach improves the performance of the fake news detection model in comparison with the recent state-of-the-art approaches. The overall experimental framework used in this work is shown in Fig. 6.

Dataset

In the machine and deep learning community, it is common knowledge that data is the most important aspect of any task. However, there is no single standard dataset that is large enough to be utilized for effectively and reliably detecting fake news, especially with deep learning methods. Therefore, the dataset employed in this study was derived from two different sources: *fake and real news dataset* [49] and *allData* [41]. The final dataset was created by merging the datasets mentioned above using the panda's library's `con-cat` function.

The *fake and real news dataset* [49] contains news headlines and news descriptions which are labeled as "fake" or "real". This dataset contains 44,919 labeled news items, out of which 21,417 are labeled as real news and 23,502 as fake news. *allData* [41] contains 20,015 labeled news items, with

11,941 as fake news and 8074 as real news. The final consolidated dataset used for this study contains 64,934 labeled news items. It contains three columns: News Headline, News Description, and Label (0-fake, 1-real). Approximately 54.6% of news items in the consolidated dataset are labeled as fake news, while 45.4% are labeled as real news. Out of the complete dataset, 70% (45454 news) of the samples were used for training, 15% (9740 news) were used for validation, and 15% (9740 news) were used for testing. The distribution of all three datasets, including the consolidated dataset, is shown in Fig. 5.

Preprocessing

Preprocessing is the first step that is accomplished to prepare the input data for feeding into the model. This stage also incorporates the cleaning of the input data, the removal of irregularities in the raw data, data normalization, and other similar issues so that the model can learn more effectively in the later stages of the pipeline.

Many portions in the news descriptions were found to be ineffective in relation to detecting fake news. To reduce the number of distinct terms, the entire data were transformed to lowercase. Special characters, hashtags, and other elements were eliminated. Stop words like "is", "the", and "are" were removed since they are not useful for detecting fake news. Following the initial pre-processing, several feature extraction and word-embedding techniques were utilized to extract and represent the input data in a way that the models can fully exploit. The following textual features extraction techniques are used in this work:

1. BOW (bag of words): In this technique, words from sentences are tokenized and placed into bags or groups marking the token and its count.
2. TF-IDF (term frequency–inverse document frequency): In this method, statements of tokenized words are converted into sparse matrices. TF-IDF is calculated using the term frequency (TF) and inverse document frequency (IDF).

Fig. 5 Dataset distribution of the datasets used in this work
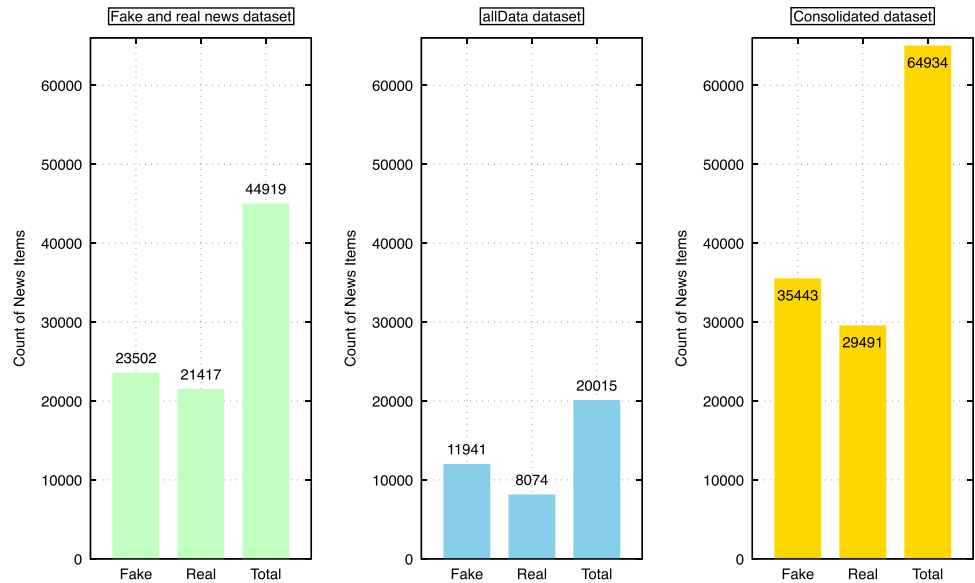
Fig. 6 Experimental framework used in this work. NB stands for Naïve Bayes
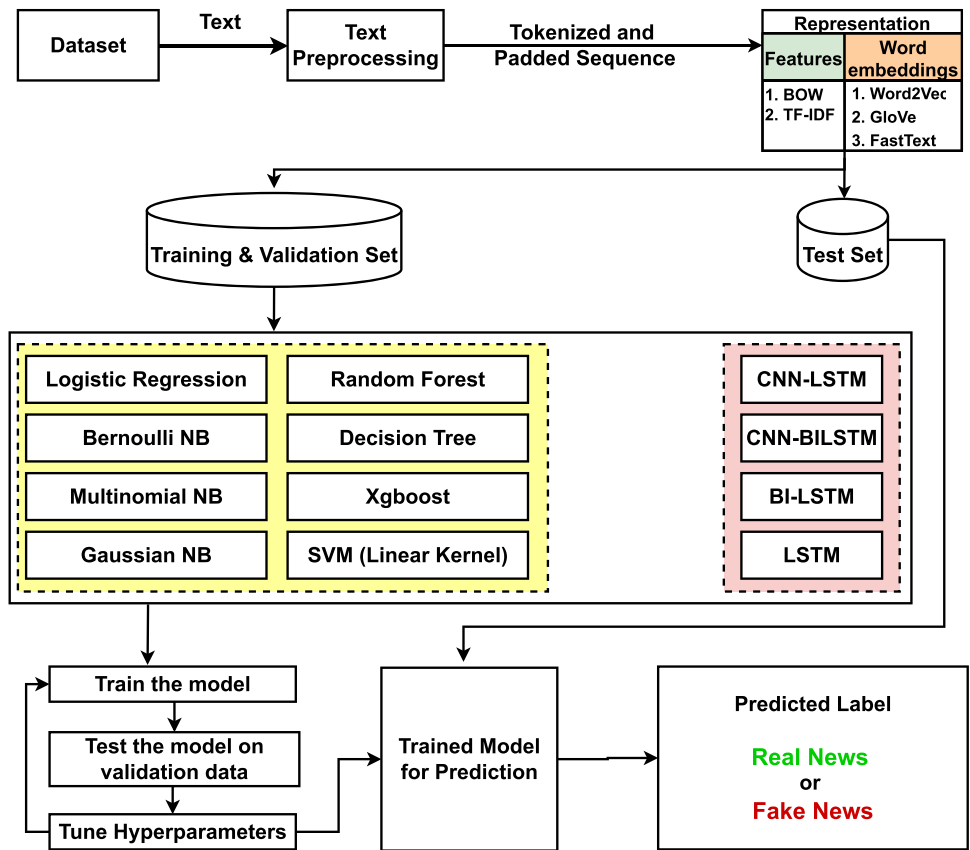
In conjunction with the deep learning methods, the following word-embedding techniques are used in this work:

1. Word2Vec: It represents each distinct word with a particular vector. Those vectors are chosen carefully such that a simple mathematical function indicates a level of semantic similarity between the words represented by vectors.

2. GloVe (global vectors): It is a word vector representation method where training is performed on aggregated global word–word occurrence from the corpus.

3. FastText: It is an extension of the Word2Vec model. It represents each word as an n-gram of characters. It helps to capture the meaning of shorter words and allows the embeddings to understand suffixes and prefixes.

Model, Training, and Evaluation

The outperforming model in this study exploits the CNN's ability to extract spatial information as well as the ability of Long Short-Term Memory (LSTM) to learn long-term relationships. To analyze the input vectors and extract the feature points present at the text level, a 1D convolutional layer is used first. The feature maps of the convolutional layer are used as input for the BiLSTM layer that learns the long-term interdependence of local characteristics in news descriptions to categorize them as real or fake.

Implementation Decisions

The models are implemented using Google Colab Pro platform. For intensive computing, Google Colab is a cloud environment for Jupyter notebooks that incorporates GPUs (Graphical Processing Units) and TPUs (Tensor Processing Units). The experiment code was written in Python.

Mapping Text to Vectors

The text is tokenized and then converted to vectors. The tokenizer is used to convert a pre-processed training corpus into integer sequences. The length of vector (sequence) is set to 2000 and pre-padding is applied to it. The CNN is given pre-trained word embeddings as input to accomplish local feature extraction. An embedding matrix is created using the FastText, GloVe, and Word2Vec word embeddings. With the current dataset, Word2Vec outperforms the other two embeddings. The dimension of embeddings was set to 200 in this study.

Proposed Model Details

The proposed model is implemented using Keras API. The Sequential model is composed of several layers. The embedding layer is the first layer of the network. This is the input layer where the model is fed the training data. The pre-trained word embeddings are used by giving the prepared embedding matrix as starting weights. To reduce the effect of overfitting, the next layer is a Dropout layer with a rate of 0.3. The third layer is a 1D CNN layer (Conv1D) that has 64 filters of size 5×5 to extract local features, along with ReLU as the activation function. In the next layer, vectors are pooled (MaxPooling1D) with a window size of 4. The BiLSTM layer that follows receives the pooled feature maps. This information is utilized to train the BiLSTM that outputs long-term dependent features of input while keeping memory. The dimension of the output is set to 128. Next, another Dropout layer is added with a rate of 0.3. The final layer of model is a Dense layer. Here the vectors are classified as real or fake. Sigmoid is used as a activation function in this layer. Binary cross-entropy is used as the loss function and Adaptive Moment Estimation (Adam) is used as the optimizer. Training of the model is performed with a batch size of 64. The proposed model is depicted in Fig. 7 and Algorithm 1.

Fig. 7 Architecture of the proposed hybrid deep learning model
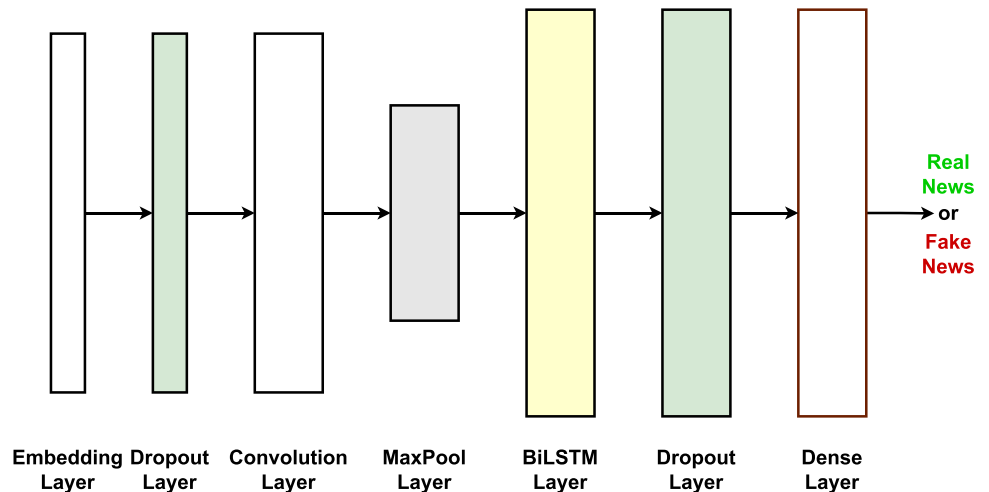
```
1: for each input sample in the training set, do
2:     Pre-process the input text
3:     Extract features using one of BOW, TF-IDF, Word2Vec, GloVe
4:     Pass the features to the Convolution layer for further features extraction from existing features
5:     Send the features from the CNN layer to the BiLSTM layer for modeling dependencies between words in a
       sentence
6:     Forward the outputs from BiLSTM layer to the Dense layer for final classification into fake or real news
7: end for
```

Algorithm 1 Steps in proposed method

Results and Analysis

This section describes the evaluation of machine learning and deep learning methods along with various feature engineering methods. Section 4.1 describes evaluation of the machine learning methods and Sect. "Deep Learning Methods" describes the evaluation of deep learning methods.

Machine Learning Methods

To achieve as accurate results as possible, eight different machine learning models were tried for fake news detection experiments. Among the various classification algorithms used, SVM with linear kernel was found to be the best classification algorithm with an accuracy of 92.3%. The details of these models along with their accuracy, precision, recall, F1-score, and AUC-ROC are depicted in Tables 2 and 3. The graphs for accuracy and F1-score of machine learning methods are shown in Fig. 8.

As per the results obtained, the Multinomial Naive Bayes algorithm with the TF-IDF count vectorizer feature performed the worst among all the classifiers. XgBoost performed the best in terms of precision (0.950) with both BOW and TF-IDF features. SVM with linear kernel and TF-IDF count vectorizer feature provides the best accuracy (0.923), precision (0.949), F1-score (0.917), and ROC-AUC with 0.921. Logistic Regression performs better than other experimented methods in terms of recall (0.892) while SVM performed better in terms of most other performance metrics. The recent advancements in deep learning motivated the authors to apply deep learning-based approaches for fake news detection, as discussed in Sect. Deep Learning Methods.

Deep Learning Methods

Different deep learning-based models vanilla and hybrid models were trained with the abovementioned three word-embedding techniques. The details of these models along with their accuracy, precision, recall, F1-score, and AUC-ROC are given in Tables 4, 5, and 6. The graphs for accuracy and F1-score of deep learning methods with different word embeddings are shown in Fig. 9. Among all these deep learning- and word-embedding-based models, the hybrid model **CNN-BiLSTM** with Word2Vec word-embedding outperforms all other models. It provides accuracy (0.975), precision (0.984), recall (0.970), F1-score (0.977), and AUC-ROC (0.992). It can be seen that, in general, using BiLSTM tends to improve performance over LSTM. This is expected as BiLSTM models the sequence in two directions. Furthermore, when combined with CNN, BiLSTM further improves the results. This indicates that the features identified by CNN layer contribute in the final classification, even after word embeddings have already been used for word representation. Finally, it was expected that FastText embedding would provide the best results among the word embeddings as it can, to some extent, handle out of vocabulary issues. However, this was not observed and the reason for that remains to be explored.

Comparison with State of the Art

In this section, we compare the current study with other research already done on fake news detection. Since there is no standard dataset available, a completely fair comparison is not possible. Therefore, we attempt to compare the methodology and results of other researchers with the present work as far as practicable.
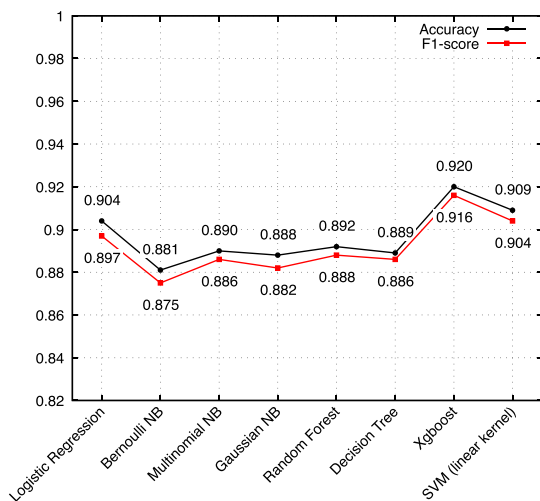
Machine Learning Methods

In this study, different machine learning models were utilized among which SVM proved to be the best that yielded an accuracy of 0.923. The dataset is created by merging 2 different publicly available datasets *Fake and real news* dataset and *allData* containing a total of 64,934 data points. The comparison of the current study with state-of-the-art approaches, with respect to the machine learning methods, is shown in Table 7. The graphical comparison of accuracy and F1-score is illustrated in Fig. 10.

Table 2 Performance of different machine learning methods on the consolidated dataset using bag of words count vectorizer feature

S. No.	Model	Accuracy	Recall	Precision	F1-score	AUC-ROC
1	Logistic Regression	0.904	0.892	0.902	0.897	0.889
2	Bernoulli NB	0.881	0.858	0.891	0.875	0.880
3	Multinomial NB	0.890	0.882	0.891	0.886	0.889
4	Gaussian NB	0.888	0.862	0.904	0.882	0.887
5	Random Forest	0.892	0.886	0.891	0.888	0.889
6	Decision Tree	0.889	0.886	0.887	0.886	0.887
7	Xgboost	0.920	0.889	0.950	0.916	0.921
8	SVM (linear kernel)	0.909	0.886	0.924	0.904	0.908

Table 3 Performance of different machine learning methods on the consolidated dataset using TF-IDF Feature

S. No.	Model	Accuracy	Recall	Precision	F1-score	AUC-ROC
1	Logistic Regression	0.913	0.887	0.931	0.908	0.856
2	Bernoulli NB	0.877	0.860	0.883	0.871	0.876
3	Multinomial NB	0.856	0.848	0.855	0.851	0.856
4	Gaussian NB	0.888	0.862	0.904	0.882	0.887
5	Random Forest	0.873	0.873	0.868	0.870	0.856
6	Decision Tree	0.889	0.886	0.887	0.886	0.856
7	Xgboost	0.920	0.889	0.950	0.916	0.922
8	SVM (linear kernel)	0.923	0.888	0.949	0.917	0.921

(a) Bag of Words feature
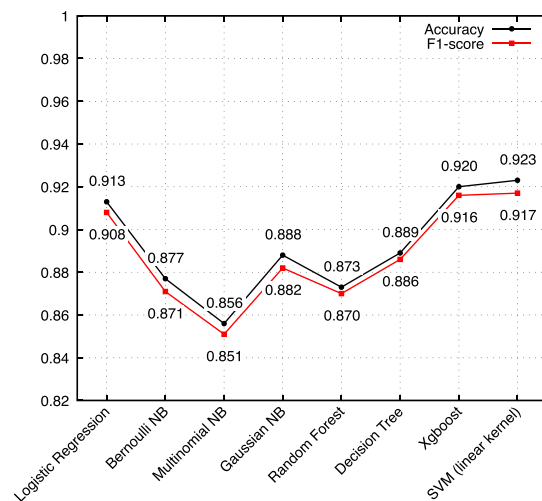
(b) TF-IDF feature

Fig. 8 Graphs showing accuracy and F1-score of Machine Learning methods with different feature extraction techniques

Deep Learning Methods

In this study, various deep learning-based models and hybrid models were trained among which the CNN-BiL-STM model with Word2Vec word-embedding outperformed all other models. It yielded an accuracy of 0.975, precision of 0.984, recall of 0.970, and F1-score of 0.977 on the consolidated dataset containing 64,934 sample points. The comparison of the current study with state-of-the-art approaches, with respect to the deep learning methods, is shown in Table 8. The comparison of accuracy and F1-score is depicted in Fig. 11. It is important to note, however, that a fair comparison of the proposed approach is not possible with state-of-the-art methods.

Table 4 Performance of deep learning models on the consolidated dataset with Word2Vec embeddingc

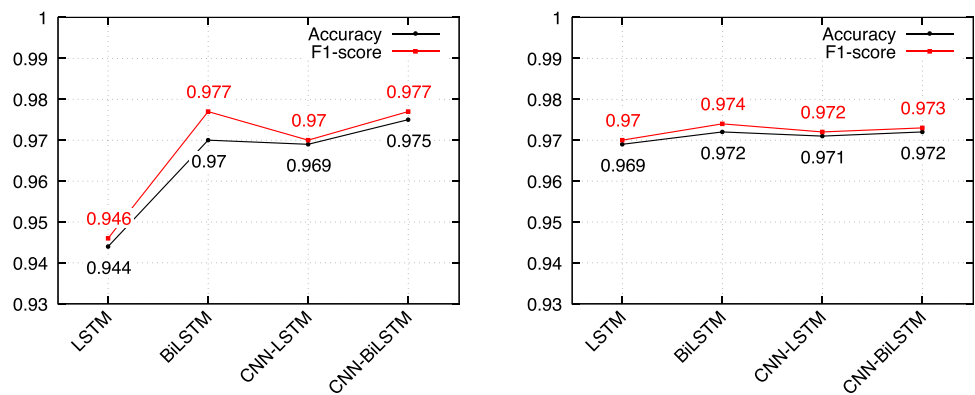| S.No.	Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
1	LSTM	0.944	0.944	0.948	0.946	0.987
2	BiLSTM	0.970	0.974	0.968	0.977	0.993
3	CNN-LSTM	0.969	0.966	0.974	0.970	0.991
4	CNN-BiLSTM	0.975	0.984	0.970	0.977	0.992

Table 5 Performance of deep learning models on the consolidated dataset with GloVe embedding

S.No.	Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
1	LSTM	0.969	0.974	0.967	0.970	0.994
2	BiLSTM	0.972	0.973	0.974	0.974	0.995
3	CNN-LSTM	0.971	0.972	0.972	0.972	0.993
4	CNN-BiLSTM	0.972	0.981	0.965	0.973	0.994

**Table 6** Performance of deep learning models on the consolidated dataset with FastText embedding

S.No.	Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
1	LSTM	0.969	0.976	0.965	0.971	0.988
2	BiLSTM	0.968	0.965	0.965	0.965	0.993
3	CNN-LSTM	0.966	0.967	0.969	0.968	0.992
4	CNN-BiLSTM	0.972	0.980	0.966	0.973	0.993

Fig. 9 Graphs showing accuracy and F1-score of deep learning methods with different word embeddings


(a) Word2Vec Embedding

(b) GloVe Embedding
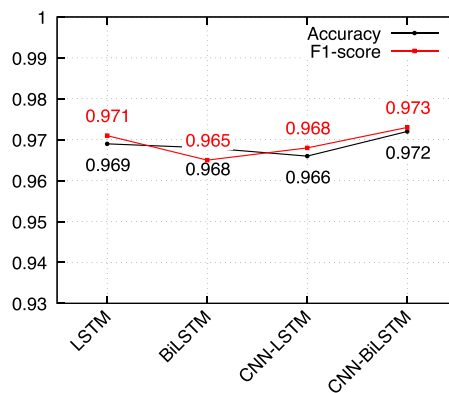
(c) Fasttext Embedding

Table 7 Comparison of machine learning methods with state of the art

S. No.	Model/year	Dataset	Accuracy	Precision	Recall	F1-score
1	SVM (2015) [27]	TV shows, fact-checking websites	0.650	0.640	0.650	0.630
2	CPCV model (2016) [29]	Sina Weibo	0.840	0.918	0.933	0.853
3	LSVM (2017) [50]	Fake and real news	0.920	–	–	–
4	LFD model (2020) [32]	fakenewsdata1	0.860	–	–	–
5	Naïve Bayes (2020) [51]	REAL and FAKE news	0.920	0.960	0.970	0.930
6	Randomized DT (2020) [52]	Persian Twitter data	0.910	0.920	0.900	0.902
7	KNN model (2021) [53]	Web scraping	0.920	0.930	0.900	0.910
8	Naïve Bayes (2021) [31]	LIAR Dataset	0.600	0.590	0.600	0.590
9	Naïve Bayes (2021) [31]	Fake and real news	0.860	0.860	0.860	0.860
10	Logistic Regression (2021) [54]	Web scraping	0.850	0.840	0.870	0.850
11	This study (SVM + TF-IDF)	Consolidated dataset	0.923	0.949	0.888	0.917
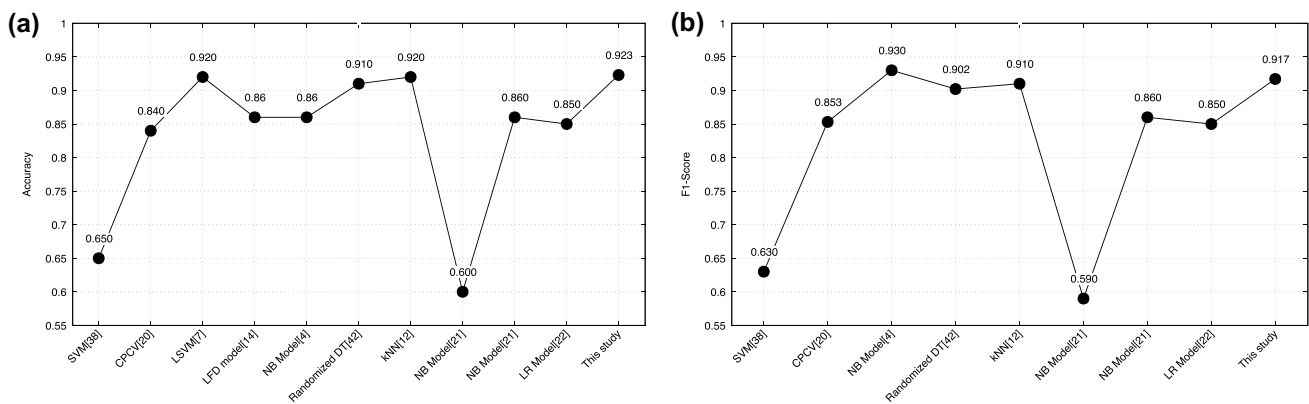
Fig. 10 Comparison of (**a**) accuracy & (**b**) F1-score of machine learning methods with state of the art

This is because there is no large, standard dataset available for the task. Hence, researchers tend to use the dataset that is available to them or their own scraped datasets. In Table 8, studies that have employed a reasonably large dataset are shown. Among those, the proposed method provides the best performance.

Conclusion and Future Work

In this study, several machine learning and deep learning models were assessed for the task of fake news detection and a deep learning-based hybrid model is proposed. The current study used word embeddings like Word2Vec, Glove, and FastText along with models such as LSTM, BiLSTM, CNN-LSTM, and CNN-BiLSTM to arrive at the proposed model. In the experiments, it is observed that all deep learning models outperform machine learning models. The BiLSTM model outperforms unidirectional models. In terms of accuracy (0.975), precision (0.984), recall (0.970), F1-score (0.977), and AUC-ROC (0.992), the CNN-BiLSTM (proposed) model with Word2Vec embedding surpassed all other models. Overall, the employment of deep neural networks in this domain appears to be promising.

In the future, further experiments may be carried out using different, recent word representation techniques to improve the performance. Along with word embeddings, recent high performing classifiers such as transformer architecture-based models may be utilized to improve the recognition of fake news from real news. In relation to products, the proposed methodology may be utilized to create an effective fake news recognition application for various social media platforms.

Table 8 Comparison of proposed deep learning method with state of the art

S. No.	Model/year	Dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC score
1	CSI model (2017) [40]	Twitter	0.893	–	–	–	–
2	CSI model (2017) [40]	Weibo	0.953	–	–	–	–
3	CNN-based model (2018) [41]	Scraped from websites	–	0.922	0.9271	0.921	–
4	CNN(2019) [42]	Wechat	0.824	0.88	0.751	0.810	0.873
5	CNN(2019) [55]	Scraped from websites	0.940	–	–	–	0.980
6	MLP(ELMo) (2020) [44]	FakeNewsAMT	0.833	–	–	–	–
7	MLP(ELMo) (2020) [44]	Celebrity	0.790	–	–	–	–
8	RNN (2020) [56]	LIAR	0.965	–	–	–	–
9	CNN (2020) [56]	FakeNews datset	0.968	–	–	–	–
10	Hybrid CNN-RNN(2021) [48]	FA-KES dataset	0.600	0.590	0.600	0.590	–
11	SciBERT(2021) [57]	FakeHealth dataset	0.690	0.720	0.690	–	–
12	Longformer(2021) [30]	FakeNewsNet	0.970	0.970	0.970	0.970	–
13	CNN-BiLSTM + Word2Vec	Consolidated dataset	0.975	0.984	0.970	0.977	0.992
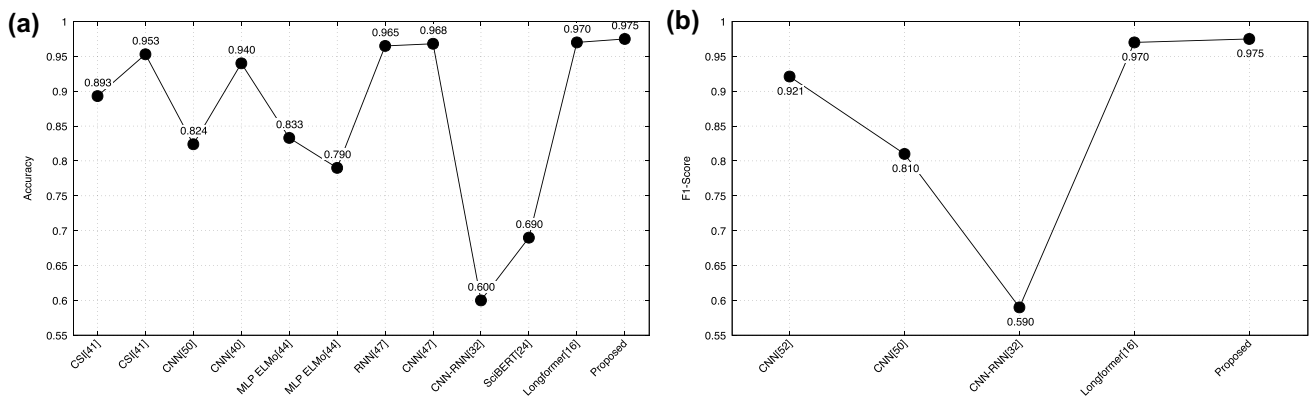
Fig. 11 Comparison of (**a**) accuracy & (**b**) F1-score of proposed deep learning model with state of the art

Declarations

Conflict of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Twitter. Accessed: 2021-12-30. https://twitter.com.
2. Facebook. Accessed: 2021-12-30. https://www.facebook.com.
3. Newman N, Dutton GBWH. Social media in the changing ecology of news: the fourth and fifth estates in Britain. Int J Internet Sci. 2012;7(1):6–22.
4. Zubiaga A, Aker KA. Detection and resolution of rumours in social media. ACM Comput Surv. 2018;51(2):32:1-32:36.
5. Diakopoulos N, De Choudhury MNM. Finding and assessing social media information sources in the context of journalism. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM. 2012; p. 2451-2460.
6. Tolmie P, Procter DR. Supporting the use of user generated content in journalistic practice. In: Proceedings of the CHI conference on human factors in computing systems, ACM. 2017;3632-3644.
7. Hermida A. Twittering the news: the emergence of ambient journalism. J Pract. 2010;4(3):297–308.
8. Vieweg S. Microblogged contributions to the emergency arena: discovery, interpretation and implications. In: Proceedings of the ACM conference on computer supported cooperative work, ACM. 2010;241-250.
9. Newman FN. Reuter Institute Digital News Report 2018. (2018).
10. Wang WY. "Liar, liar pants on fire": a new benchmark dataset for fake news detection. In: Proceedings of the fifty fifth annual meeting of the association for computational linguistics. 2017;2: Short Papers:422-426.
11. Kang CAG. In Washington Pizzeria Attack. Fake News Brought Real Guns. 2016;5:102437.
12. Allcott HMG. Social media and fake news in the 2016 election. Technical Report, National Bureau of Economic Research. 2017.

13. Bondielli AMF. A survey on fake news and rumour detection techniques. Inform Sci. 2019;497:38–55.

14. Bharti S, Yadav AK, Kumar M, Yadav D. Cyberbullying detection from tweets using deep learning. Kybernetes. 2021.

15. Guellil I, Adeel A, Azouaou F, Chennoufi S, Maafi H, Hamitouche T. Detecting hate speech against politicians in Arabic community on social media. Int J Web Inform Syst. 2020.

16. Akuma S, Lubem T, Adom IT. Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets. Int J Inform Technol. 2022;p. 1–7.

17. Nielsen RK, Graves L. " News you don't believe": Audience perspectives on fake news. 2017.

18. Kumar P, Vardhan M. PWEBSA: Twitter sentiment analysis by combining Plutchik wheel of emotion and word embedding. Int J Inform Technol. 2022;1–9.

19. Salovey P, Mayer JD. Emotional intelligence. Imagination Cognit Personality. 1990;9(3):185–211.

20. Mayer JD, Salovey P. What is emotional intelligence? Emotional development and emotional intelligence: Implications for educators. New York: Basic Book; 1997. p. 3–31.

21. Hamid Y, Elyassami S, Gulzar Y, Balasaraswathi VR, Habuza T, Wani S. An improvised CNN model for fake image detection. Int J Inform Technol. 2022;1–11.

22. Preston S, Anderson A, Robertson DJ, Shephard MP, Huhe N. Detecting fake news on Facebook: the role of emotional intelligence. Plos one. 2021;16(3): e0246757.

23. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: Predicting deception from linguistic styles. Personal Soc Psychol Bull. 2003;29(5):665–75.

24. Markowitz DM, Hancock JT. Linguistic obfuscation in fraudulent science. J Language Soc Psychol. 2016;35(4):435–45.

25. Pennebaker JW, King LA. Linguistic styles: language use as an individual difference. J Personality Soc Psychol. 1999;77(6):1296.

26. Khanday AMUD, Khan QR, Rabani ST. Identifying propaganda from online social networks during COVID-19 using machine learning techniques. Int J Inform Technol. 2021;13:115–22.

27. Pisarevskaya D. Rhetorical structure theory as a feature for deception detection in news reports in the Russian language. In: Komp'juternaja Lingvistika i Intellektual'nye Tehnologii; 2017. p. 191–200.

28. Chen Y, Rubin VL, Conroy N. Towards news verification: deception detection methods for news discourse. 2015.

29. Jin Z, Cao J, Zhang Y, Luo J. News verification by exploiting conflicting social viewpoints in microblogs. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 30; 2016.

30. Ghanem B, Ponzetto SP, Rosso P, Rangel F. Fakeflow: Fake news detection by modeling the flow of affective information. arXiv preprint arXiv:2101.09810. 2021.

31. Khan JY, Khondaker MTI, Afroz S, Uddin G, Iqbal A. A benchmark study of machine learning models for online fake news detection. Mach Learn Appl. 2021;4: 100032.

32. Choudhary A, Arora A. Linguistic feature based learning model for fake news detection and classification. Expert Syst Appl. 2021;169: 114171.

33. Agarwal V, Sultana HP, Malhotra S, Sarkar A. Analysis of classifiers for fake news detection. Proc Comput Sci. 2019;165:377–83.

34. Bahad P, Saxena P, Kamal R. Fake news detection using bi-directional LSTM-recurrent neural network. Proc Comput Sci. 2019;165:74–82.

35. Faustini PHA, Covoes TF. Fake news detection in multiple platforms and languages. Expert Syst Appl. 2020;158: 113503.

36. Silva RM, Santos RL, Almeida TA, Pardo TA. Towards automatically filtering fake news in Portuguese. Expert Syst Appl. 2020;146: 113199.

37. Sahoo SR, Gupta BB. Hybrid approach for detection of malicious profiles in twitter. Comput Electr Eng. 2019;76:65–81.

38. Meel P, Vishwakarma DK. Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst Appl. 2020;153: 112986.

39. Wu Y, Ngai EW, Wu P, Wu C. Fake online reviews: literature review, synthesis, and directions for future research. Decision Supp Syst. 2020;132: 113280.

40. Ruchansky N, Seo S, Liu Y. Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management; 2017. 797–806.

41. Yang Y, Zheng L, Zhang J, Cui Q, Li Z, Yu PS. TI-CNN: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749. 2018.

42. Wang Y, Yang W, Ma F, Xu J, Zhong B, Deng Q, et al. Weak supervision for fake news detection via reinforcement learning. In: Proceedings of the AAAI conference on artificial intelligence. 34; 2020: 516–523.

43. MadhuBala M, Yadav AK, Sucharitha G, Kumar PP. A Deep Learning Based Algorithm Design for Fake News Detection Framework. Ann Roman Soc Cell Biol. 2021;25(6):4182–4192.

44. Saikh T, De A, Ekbal A, Bhattacharyya P. A deep learning approach for automatic detection of fake news. arXiv preprint arXiv:2005.04938. 2020.

45. Nasir JA, Khan OS, Varlamis I. Fake news detection: a hybrid CNN-RNN based deep learning approach. Int J Inform Manag Data Insights. 2021;1(1): 100007.

46. Kumari S, Reddy HK, Kulkarni CS, Gowthami V. Debunking Health Fake News with Domain Specific Pre-Trained Model. Global Trans Proc. 2021;.

47. Karimi H, Roy P, Saba-Sadiya S, Tang J. Multi-source multi-class fake news detection. In: Proceedings of the 27th international conference on computational linguistics; 2018:1546–1557.

48. Nasir JA, Khan OS, Varlamis I. Fake news detection: A hybrid CNN-RNN based deep learning approach. Int J Inform Manag Data Insights. 2021;1(1): 100007.

49. Ahmed H, Traore I, Saad S. Detecting opinion spams and fake news using text classification. Secur Privacy. 2017;12(1): e9.

50. Ahmed H, Traore I, Saad S. Detection of online fake news using n-gram analysis and machine learning techniques. In: International conference on intelligent, secure, and dependable systems in distributed and cloud environments. Springer; 2017: 127–138.

51. Adiba FI, Islam T, Kaiser MS, Mahmud M, Rahman MA. Effect of corpora on classification of fake news using naive Bayes classifier. Int J Automation Artificial Intell Mach Learn. 2020;1(1):80–92.

52. Sadr MM, Mousavi Chelak A, Ziaei S, Tanha J. A predictive model based on machine learning methods to recognize fake persian news on twitter. Int J Nonlinear Anal Appl. 2020;11:119–28.

53. Chaturvedi B, Dubey S. An Efficient Approach based on Machine Learning and Performance Improvement for Fake News Detection;.

54. Kulkarni P, Karwande S, Keskar R, Kale P, Iyer S. Fake News Detection using Machine Learning. In: ITM Web of Conferences. vol. 40. EDP Sciences; 2021. p. 03003.

55. Rodríguez ÁI, Iglesias LL. Fake news detection using Deep Learning. arXiv preprint arXiv:1910.03496. 2019.

56. Singh L. Fake News Detection: a comparison between available Deep Learning techniques in vector space. In: 2020 IEEE

4th Conference on Information & Communication Technology (CICT). IEEE; 2020. p. 1–4.

57. Kumari S, Reddy HK, Kulkarni CS, Gowthami V. Debunking health fake news with domain specific pre-trained model. Global Trans Proc. 2021;2(2):267–72.