



# FakeBERT: Fake news detection in social media with a BERT-based deep learning approach

Rohit Kumar Kaliyar<sup>1</sup> · Anurag Goswami<sup>1</sup> · Pratik Narang<sup>2</sup> 

Received: 1 May 2020 / Revised: 24 August 2020 / Accepted: 11 November 2020 /

Published online: 7 January 2021

© Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

In the modern era of computing, the news ecosystem has transformed from old traditional print media to social media outlets. Social media platforms allow us to consume news much faster, with less restricted editing results in the spread of fake news at an incredible pace and scale. In recent researches, many useful methods for fake news detection employ sequential neural networks to encode news content and social context-level information where the text sequence was analyzed in a unidirectional way. Therefore, a bidirectional training approach is a priority for modelling the relevant information of fake news that is capable of improving the classification performance with the ability to capture semantic and long-distance dependencies in sentences. In this paper, we propose a BERT-based (Bidirectional Encoder Representations from Transformers) deep learning approach (FakeBERT) by combining different parallel blocks of the single-layer deep Convolutional Neural Network (CNN) having different kernel sizes and filters with the BERT. Such a combination is useful to handle ambiguity, which is the greatest challenge to natural language understanding. Classification results demonstrate that our proposed model (FakeBERT) outperforms the existing models with an accuracy of 98.90%.

**Keywords** Fake news · Neural network · Social media · Deep learning · BERT

## 1 Introduction

In the past few years, various social media platforms such as Twitter, Facebook, Instagram, etc. have become very popular since they facilitate the easy acquisition of information and

---

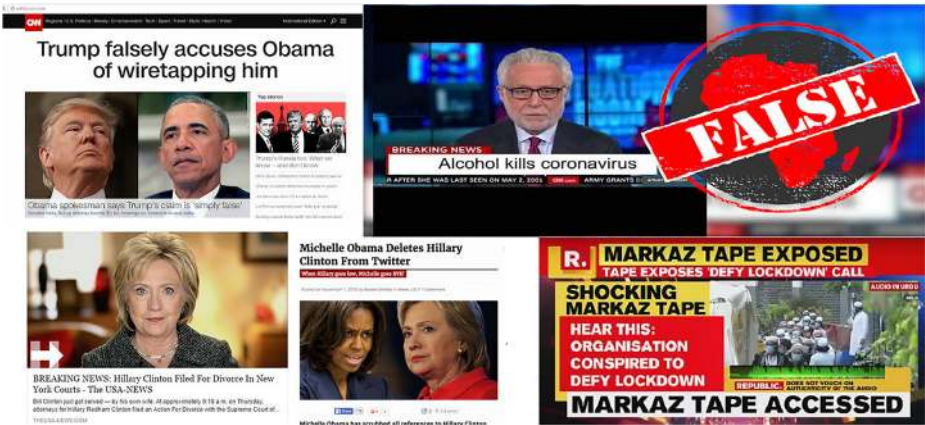
✉ Pratik Narang  
pratik.narang@pilani.bits-pilani.ac.in

Rohit Kumar Kaliyar  
rk5370@bennett.edu.in

Anurag Goswami  
anurag.goswami@bennett.edu.in

<sup>1</sup> Departement of Computer Science Engineering, Bennett University, Greater Noida, India

<sup>2</sup> Departement of CSIS, BITS Pilani, Pilani, Rajasthan, India



**Fig. 1** Examples of some fake news spread over social media (Source: Facebook<sup>®</sup>)

provide a quick platform for information sharing [10, 21]. The availability of unauthentic data on social media platforms has gained massive attention among researchers and become a hot-spot for sharing fake news [16, 46]. Fake news has been an important issue due to its tremendous negative impact [16, 46, 53], it has increased attention among researchers, journalists, politicians and the general public. In the context of writing style, fake news is written or published with the intent to mislead the people and to damage the image of an agency, entity, person, either for financial or political benefits [14, 35, 39, 53]. Few examples of fake news are shown in Fig. 1. These examples of fake news were in trending during the COVID-19 pandemic and 2016 U.S. General Presidential Election.

In the research context, related synonyms (keywords) often linked with fake news:

- Rumor: A rumour [4, 12, 16] is an unverified claim about any event, transmitting from individual to individual in the society. It might imply to an occurrence, article, and any social issue of open public concern. It might end up being a socially dangerous phenomenon in any human culture.
- Hoax: A hoax is a falsehood deliberately fabricated to masquerade as the truth [43]. Currently, it has been increasing at an alarming rate. Hoax is also known as with similar names like prank or jape.

### 1.1 Existing approaches for fake news detection

Detection of fake news is challenging as it is intentionally written to falsify information. The former theories [1] are valuable in guiding research on fake news detection using different classification models. Existing learnings for fake news detection can be generally categorized as (i) News Content-based learning and (ii) Social Context-based learning.

News content-based approaches [1, 14, 51, 53] deals with different writing style of published news articles. In these techniques, our main focus is to extract several features in fake news article related to both information as well as the writing style. Furthermore, fake news publishers regularly have malignant plans to spread mutilated and deluding, requiring specific composition styles to interest and convince a wide extent of consumers that are not present in true news stories. In these learnings, style-based methodologies [12, 35, 53] are helpful to capture the writing style of manipulators using linguistic features for identifying

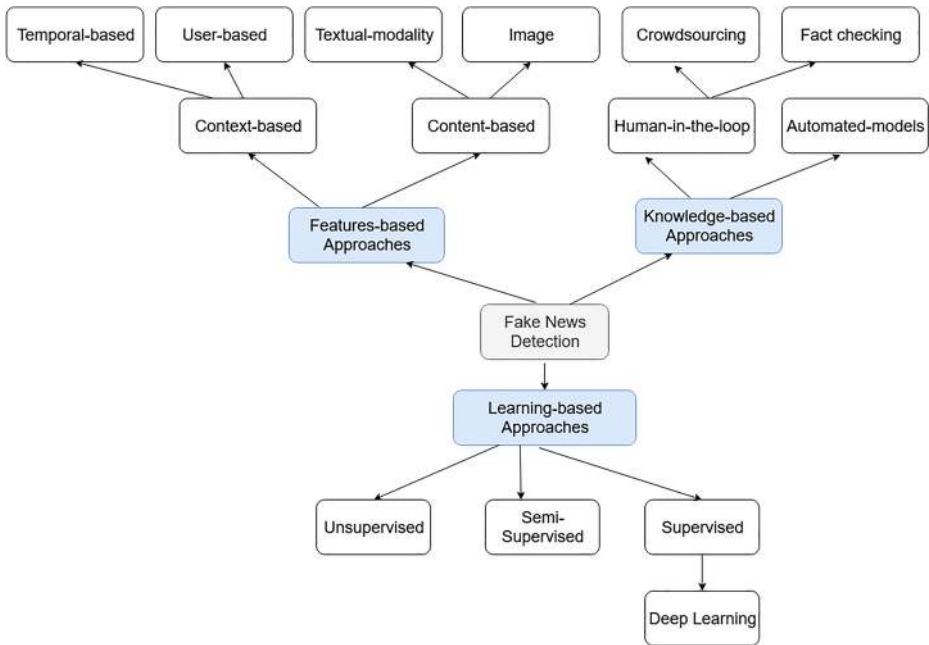


Fig. 2 Approaches for fake news detection

fake articles. Thus, it is difficult to detect fake news more accurately by using only news content-based features [14, 33, 46]. Thus, we also need to investigate the engagement of fake news articles with users.

Social context-based approaches [14, 17, 38, 51, 53] deals with the latent information between the user and news article. Social engagements (the semantic relationship between news articles and user) can be used as a significant feature for fake news detection. In these approaches, instance-based methodologies [51] deals with the behaviour of the user towards any social media post to induce the integrity of unique news stories. Furthermore, propagation-based methodologies [51] deals with the relations of significant social media posts to guide the learning of validity scores by propagating credibility values between users, posts, and news. Approaches related to fake news detection show in Fig. 2. In most of the existing and useful methods [14, 38, 51] consists of news content and context level features using unidirectional pre-trained word embedding models (such as GloVe, TF-IDF, word2Vec, etc.) There is a large scope to use bidirectional pre-trained word embedding models having powerful feature extraction capability.

### 1.2 Our contribution

In the existing approaches [1, 33, 40], for the detection of fake news, many useful methods have been presented using traditional machine learning models. The primary advantage of using deep learning model over existing classical feature-based approaches is that it does not require any handwritten features; instead, it identifies the best feature set on its own. The powerful learning ability of deep CNN is primarily due to the use of multiple feature extraction stages that can automatically learn representations from the dataset. In the existing

approaches [18, 19, 26], several inspiring ideas have been discussed to bring advancements in deep Convolutional Neural Networks (CNNs) like exploiting temporal and channel information, depth of architecture, and graph-based multi-path information processing. The idea of using a block of layers as a structural unit is also gaining popularity among researchers. In this paper, we propose a BERT-based deep learning approach (FakeBERT) by combining different parallel blocks of the single-layer CNNs with the Bidirectional Encoder Representations from Transformers (BERT). We utilize BERT as a sentence encoder, which can accurately get the context representation of a sentence. This work is in contrast to previous research works [9] where researchers looked at a text sequence in a unidirectional way (either left to right or right to left for pre-training). Many existing and useful methods had been [9, 24] presented with sequential neural networks to encode the relevant information. However, a deep neural network with bidirectional training approach can be an optimal and accurate solution for the detection of fake news. Our proposed method improves the performance of fake news detection with the powerful ability to capture semantic and long-distance dependencies in sentences.

To design our proposed architecture, we have added a classification layer on the top of the encoder output, multiplying the output vector by the embedding matrix, and finally calculated the probability of each vector with the Softmax function. Our model is a combination of three parallel blocks of 1D-convolutional neural networks with BERT having different kernel sizes and filters following by a max-pooling layer across each block. With this combination, the documents were processed using different CNN topologies by varying kernel size (different n-grams), filters, and several hidden layers or nodes. The design of FakeBERT consists of five convolution layers, five max-pooling layers followed by two densely connected layers and one embedding layer (BERT-layer) of input. In each layer, several filters have been applied to extract the information from the training dataset. Such a combination of BERT with one-dimensional deep convolutional neural network (1d-CNN) is useful to handle large-scale structure as well as unstructured text. It effectively addresses ambiguity, which is the greatest challenge to natural language understanding. Experiments were conducted to validate the performance of our proposed model. Several performance evaluation parameters (training accuracy, validation accuracy, False Positive Rate (FPR), and False Negative Rate (FNR)) have been taken into consideration to validate the classification results. Extensive experimentations demonstrate that our proposed model outperforms as compared to the existing benchmarks for classifying fake news. We illustrate the performance of our bidirectional pre-trained model (BERT) achieved an accuracy of 98.90%. Our proposed approach produces improved results by 4% comparing to the baseline approaches and is promising for the detection of fake news.

## 2 Related work

This section briefly summarizes the work in the field of fake news detection. Kumar et al [21] have explored a comprehensive survey of diverse aspects of fake news. Different categories of fake news, existing algorithms for counterfeit news detection, and future aspects have been explored in this research article. In one of the research, Shin et al [37] have investigated about fundamental theories across various disciplines to enhance the interdisciplinary study of fake news. In their study, authors have mainly investigated the problem of fake news from four perspectives: False knowledge it carries (what type of false message you get from the content), writing styles (different writing styles for creating fake news),

propagation patterns (when it is shared in a network, then which trends it follows), and the credibility of its creators and spreaders (the credibility score of a news creator and spreader). Bondielli et al [4] have presented a hybrid approach for detecting automated spammers by amalgamating community-based features with other feature categories, namely meta-content and interaction-based features. In another research, Ahmed et al [1] have focused on automatic detection of fake content using online fake reviews. Authors have also explored two different feature extraction methods for classifying fake news. They have examined six different machine learning models and shown improved accomplishments as compared to existing state-of-the-art benchmarks. In one of the researches, Allcott et al [2] have focused on a quantitative report to understand the impact of fake news on social media in the 2016 U.S. Presidential General Election and its effect upon U.S. voters. Authors have investigated the authentic and unauthentic URLs related to fake news from the BuzzFeed dataset. In one of the studies, Shu et al [38] have investigated a way for robotization process through hashtag recurrence. In this research article, authors have also presented a comprehensive review of detecting fake news on social media, false news classifications on psychology and social concepts, and existing algorithms from a data mining perspective. Ghosh et al [14] have investigated the impact of web-based social networking on political decisions. Quantity research [2, 53, 54] has been done in the context of detecting political-news-based articles. Authors have investigated the effect of various political gatherings related to the discussion of any fake news as agenda. Authors have also explored the Twitter-based data of six Venezuelan government officials with a specific end goal to investigate bot collaboration. Their discoveries recommend that political bots in Venezuela tend to imitate individuals from political gatherings or basic natives.

In one of the studies, Zhou et al [53] have investigated the ability of social media to aggregate the judgments of a large community of users. In their further investigation, they have explained machine learning approaches with the end goal to develop a better rumours detection. They have investigated the difficulties for the spread of rumours, rumours classification, and deception for the advancement of such frameworks. They have also investigated the utilization of such useful strategies towards creating fascinating structures that can help individuals in settling on choices towards evaluating the integrity of data gathered from various social media platforms. Vosoughi et al [46] have recognized salient features of rumours by investigating three aspects of information spread online: linguistic style, characteristics of people involved in propagating information, and network propagation subtleties. Authors have analyzed their proposed algorithm on 209 rumours representing 938,806 tweets collected from real-world events, including the 2013 Boston Marathon bombings, the 2014 Ferguson unrest, and the 2014 Ebola epidemic. They have expressed the effectiveness of their proposed framework with all existing methods. The primary objective of their study was to introduce a novel way of assessing style-similarity between different text contents. They have implemented numerous machine learning models and achieved an accuracy of 51% for fake news detection.

Chen et al [7] have proposed an unsupervised learning model combining recurrent neural networks and auto-encoders to distinguish rumours as anomalies from other credible micro-blogs based on users' behaviours. The experimental results show that their proposed model was able to achieve an accuracy of 92.49% with an F1 score of 89.16%. Further, Yang et al [49] have arrived with comparative resolutions for detecting false rumours. During the 2011 riots in England, authors have noticed and investigated that any improvement in the false rumours based stories could produce good results. In their investigation of the 2013 Boston Marathon bombings, they have found some exciting news stories, and most

of them were rumours and produced a significant impact on the share market. Shu et al [39] have explored the connection between fake and real facts available on social media platforms using an open tweet dataset. This dataset was created by gathering online tweets from Twitter that contains URLs from reality checking facts. In their investigation, they have found that URL's are the most widely recognized strategy to share news articles on various stages for the measurement of client articulation (for example, Twitter's limit is with 140 characters constraint). In their further investigation, they have used a Hoax-based dataset that gives a more accurate prediction for distinguishing fake news stories by conflicting them against known news sources from renowned inspection sites.

In one of the researches, Monteiro et al [25] have collected a fake news dataset in the Portuguese language and investigated their results based on different linguistic features. Authors have achieved the highest accuracy of 49% using machine learning techniques. One of the researches, Karimi et al [20] have analyzed 360 satirical news articles including civics, science, business, and delicate news. They have also proposed an SVM-based model. In their investigation, their five highlights are Absurdity, Humor, Grammar, Negative effect, and punctuation. Their proposed framework achieved an accuracy of 38.81%. One of the researches, Perez-Rosas et al [29] have explained the automatic identification of fake content in online news articles. They have presented a comprehensive analysis for the identification of linguistic features in the false news content. In one of the studies, Castillo et al [5] have investigated feature-based methods to assess the credibility of tweets on Twitter. Roy et al [34] have explored the neural embedding approach using the deep recurrent model. They have used weighted n-gram bag of word model using statistical features and other external features with the help of featurizing engineering. Subsequently, they have combined all features and classifying fake news with the accuracy of 43.82%. One of the researches, Wang et al [47] have presented a novel dataset for fake news detection. They have proposed a hybrid architecture to solve fake news problem. They have created a model using two main components; one is a Convolutional Neural Network for meta-data representation learning, followed by a Long Short-Term Memory neural network (LSTM). Although being complicated with many parameters to be optimized, their proposed model performs poorly on the test set, with only 27.4% inaccuracy. One of the researches, Peters et al [30] took a different perspective on detecting fake news by looking at its linguistic characteristics. Despite substantial dependence on lexical resources, the performance on political-set was even slower than [47], with an accuracy of 22.0% only.

In many existing studies [13, 23, 28, 42], authors have explored the problem of fake news employing a real-world fake news dataset: Fake-News. In one of the studies, Ahmed et al [1] have utilised TF-IDF (Term Frequency-Inverse Document Frequency) as a feature extraction method with different machine learning models. Extensive experiments have performed with LR (Linear-regression model) and obtained an accuracy of 89.00%. Subsequently, they have shown an accuracy of 92% using their LSVM (Linear Support Vector Machine). Liu et al [23] have investigated the methods for recognizing false tweets. In their investigation, authors have utilized a corpus of more than 8 million tweets gathered from the supporters of the presidential candidates in the general election in the U.S. In their investigation, they have employed deep CNNs for fake news detection. In their approach, they have utilised the concept of subjectivity analysis and obtained an accuracy of 92.10%. O'Brien et al [28] have applied deep learning strategies for classifying fake news. In their study, they have achieved an accuracy of 93.50% using the black-box method. Ghanem et al [13] have adopted different word embeddings, including n-gram features to detect the stances in fake articles. They have obtained an accuracy of 48.80%. Ruchansky et al [35] have employed a deep

hybrid model for classifying fake news. They have utilized news-user relationships as an essential factor and achieved an accuracy of 89.20%. In one of the studies, Singh et al [42] have investigated with LIWC (Linguistic Analysis and Word Count) features using traditional machine learning methods for classifying fake news. They have explored the problem of fake news with SVM (support vector machine) as a classifier obtained an accuracy of 87.00%. In one of the studies, Jwa et al [18] have explored the approach towards automatic fake news detection. They have used Bidirectional Encoder Representations from Transformers model (BERT) model to detect fake news by analyzing the relationship between the headline and the body text of the news story. Their results improve the 0.14 F-score over existing state-of-the-art models. Weiss et al [48] have investigated the origins of the term “fake news” and the factors contributing to its current prevalence. This lack of consensus may have future implications for students in particular and higher education. Crestani et al [8] have proposed a novel model that can classify a user as a potential fact checker or a potential fake news spreader. Their model was based on a Convolutional Neural Network (CNN) and combined word embeddings with features that represent users’ personality traits and linguistic patterns.

### 3 Methodology

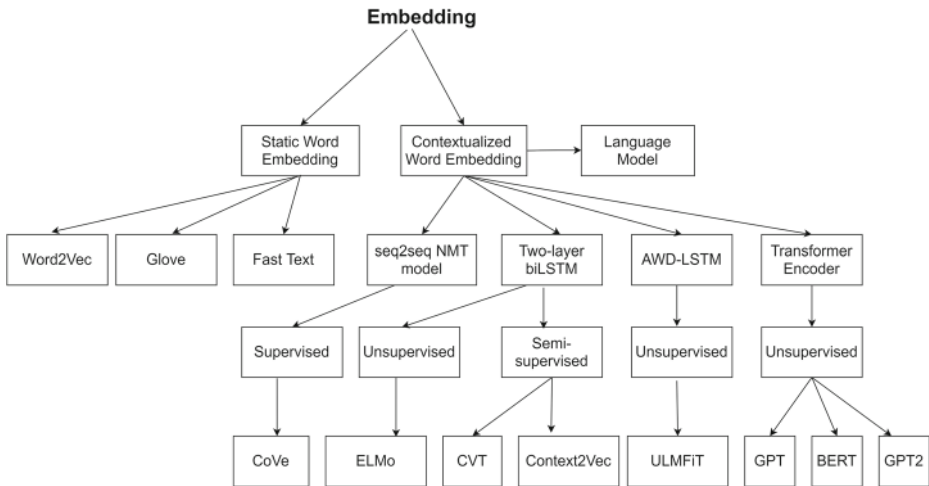
In this section, an overview of word embedding, GloVe word embedding, BERT model, fine-tuning of BERT, and the selection of hyperparameters discussed. Our proposed model (FakeBERT) and other deep learning architectures also investigated in this section.

#### 3.1 Word embedding

Word embeddings [30] are widely used in both machine learning as well as deep learning models. These models perform well in cases such as reduced training time and improved overall classification performance of the model. Pre-trained representations can also either be static or contextual (refer Fig. 3 for more details). Contextual models generate a representation of each word that is based on the other words in the sentence. Word2Vec and GloVe [50] are currently among the most widely used word embedding models that can convert words into meaningful vectors. For using pre-trained embedding models for training, we displace the parameters of the processing layer with input embedding vectors. Primarily, we maintain the index and then fix this layer, restricting it from being updated throughout the method of gradient descent [30, 31]. Our experiment shows that embedding-based input vectors perform a valuable role in text classification tasks.

#### 3.2 GloVe

The GloVe is a weighted least square model [3] that train the model using co-occurrence counts of the words in the input vectors. It effectively leverages the benefits of the statistical information by training on the non-zero elements in a word-to-word co-occurrence matrix. The GloVe is an unsupervised training model that is useful to find the co-relation between two words with their distance in a vector space [31]. These generated vectors are known as word embedding vectors. We have used word embedding as semantic features in addition to n-grams because they represent the semantic distances between the words in the context. The smallest package of embedding is 822Mb, called “glove.6B.zip”. GloVe model is trained on a dataset having one billion words with a dictionary of 400 thousand words.



**Fig. 3** An Overview of existing word-embedding models

There exist different embedding vector sizes, having 50, 100, 200 and 300 dimensions for processing. In this paper, we have taken the 100-dimensional version.

### 3.3 BERT

BERT [11] is a advanced pre-trained word embedding model based on transformer encoded architecture [44]. We utilize BERT as a sentence encoder, which can accurately get the context representation of a sentence [30]. BERT removes the unidirectional constraint using a mask language model (MLM) [44]. It randomly masks some of the tokens from the input and predicts the original vocabulary id of the masked word based only. MLM has increased the capability of BERT to outperforms as compared to previous embedding methods. It is a deeply bidirectional system that is capable of handling the unlabelled text by jointly conditioning on both left and right context in all layers. In this research, we have extracted embeddings for a sentence or a set of words or pooling the sequence of hidden-states for the whole input sequence. A deep bidirectional model is more powerful than a shallow left-to-right and right-to-left model. In the existing research [11], two types of BERT models have been investigated for context-specific tasks, are:

- BERT Base (refer Table 1 for more information about parameters setting): Smaller in size, computationally affordable and not applicable to complex text mining operations.
- BERT Large (refer Table 2 for more information about parameters setting): Larger in size, computationally expensive and crunches large text data to deliver the best results.

### 3.4 Fine-tuning of BERT

Fine-tuning of BERT [11] is a process that allows it to model many downstream tasks, irrespective of the text form (single text or text pairs). A limited exploration is available to enhance the computing power of BERT to improve the performance on target tasks. BERT model uses a self-attention mechanism to unify the word vectors as inputs that include bidirectional cross attention between two sentences. Mainly, there exist a few fine-tuning



**Table 1** Parameters for BERT-Base

Parameter Name	Value of Parameter
Number of Layers	12
Hidden Size	768
Attention Heads	12
Number of Parameters	110M

strategies that we need to consider: 1) The first factor is the pre-processing of long text since the maximum sequence length of BERT is 512. In our research, we have taken the sequence length of 512. 2) The second factor is layer selection. The official BERT-base model consists of an embedding layer, a 12-layer encoder, and a pooling layer. 3) The third factor is the over-fitting problem. BERT can be fine-tuned with different learning parameters for different context-specific tasks [44] (refer Table 2 for more information).

### 3.5 Deep learning models for fake news detection

Deep learning models are well-known for achieving state-of-the-art results in a wide range of artificial intelligence applications [31]. This section provides an overview of the deep learning models used in our research with their architectures to achieve the end goal. Experiments have been conducted using deep learning-based models (CNN and LSTM [15]) and our proposed model (FakeBERT) with different pre-trained word embeddings.

a) Convolutional Neural Network (CNN): In Fig. 4, the computational graph of our designed Convolutional Neural Network (CNN) model is shown. This CNN model (Fig. 4) truncates, zero-pads, and tokenizes the fake news article separately and passes each into an embedding layer. In this architecture (refer Table 3 and Fig. 4), first convolution layer holds 128 filters with `kernels_size=5`, which decreases the input embedding vector from 1000 to 996 after convolution process. In the network, after each convolution layer, a max-pooling layer is also present to reduce the input vector dimension. Subsequently, a max-pooling layer with `filter_size=5`; that further minimises the embedding vector to 1/5th of 996, i.e. 199. The second convolution layer holds 128 filters with `kernels_size=5`, which decreases the input embedding vector from 199 to 195. Subsequently, this is the max-pooling layer with filter size 5; that further reduces the input vector to 1/5th of 199, i.e. 39. After three convolution layers, a flatten layer is added to convert 2-D input to 1-D. Subsequently, there are two hidden layers having 128 neurons in each one. The outputs of the CNNs are passed through a dense layer with dropout and then passed through a softmax layer to yield a stance classification. Number of trainable parameters are also shown in Table 3.

**Table 2** Parameters for BERT-Large

Parameter Name	Value of Parameter
Number of Layers	24
Hidden Size	1024
Attention Heads	16
Number of Parameters	340M

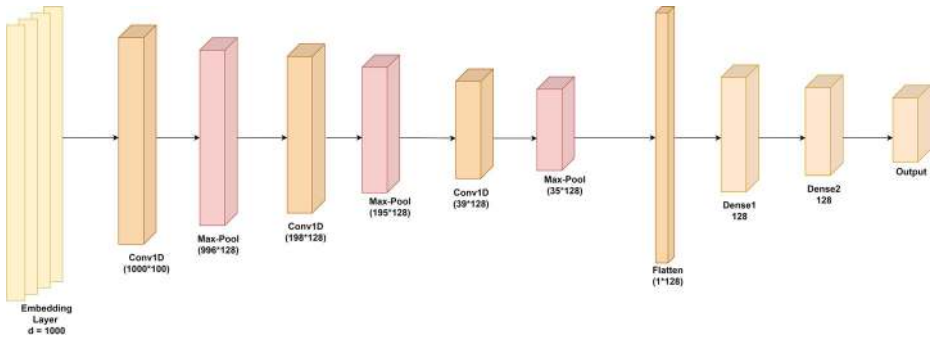


Fig. 4 CNN model

Table 3 CNN layered architecture

Layer	Input size	Output size	Param number
Embedding	1000	1000 × 100	25187700
Conv1D	1000 × 100	996 × 128	64128
Maxpool	996 × 128	199 × 128	0
Conv1D	199 × 128	195 × 128	82048
Maxpool	195 × 128	39 × 128	0
Conv1D	39 × 128	35 × 128	82048
Maxpool	35 × 128	1 × 128	0
Flatten	1 × 128	128	0
Dense	128	128	16512
Dense	128	2	258

Table 4 LSTM layered architecture

Layer	Input size	Output size	Param number
Embedding	1000 × 100	1000 × 100	25187700
Dropout	1000 × 100	1000 × 100	0
Conv1D	1000 × 100	1000 × 32	16032
Maxpool	1000 × 32	500 × 32	0
Conv1D	500 × 32	500 × 64	6208
Maxpool	500 × 64	250 × 64	0
LSTM	250 × 64	100	66000
Batch-Normalization	100	100	400
Dense	100	256	25856
Dense	256	128	32896
Dense	128	64	8256
Dense	64	2	130

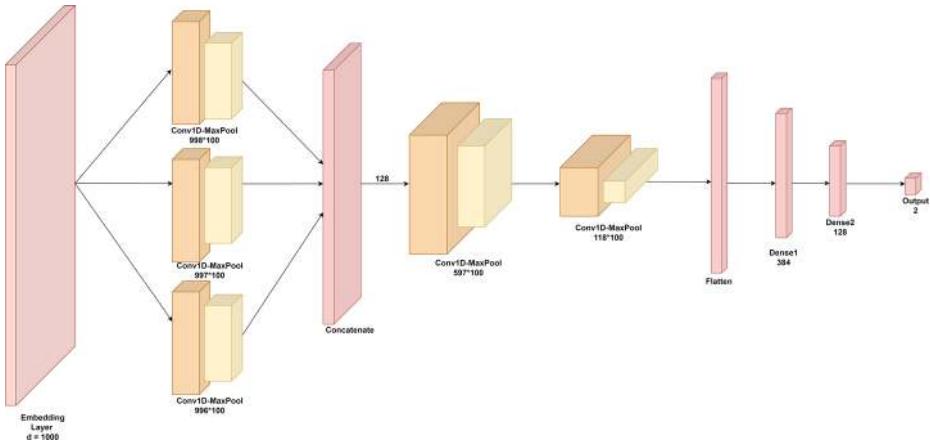


Fig. 5 FakeBERT model

b) Long Short Term Memory Network (LSTM): In this paper, we have implemented the LSTM model having four dense layers with a batch normalization process for the classification of fake news. The selection of optimal hyperparameters is also made for accurate results. From Table 4, we can observe the layered architecture of the LSTM model.

### 3.6 Proposed model: FakeBERT

In this paper, the most fundamental advantage of selecting a deep convolutional neural network is the automatic feature extraction. In our proposed model, we pass the input in the form of a tensor in which local elements correlates with one another. More concrete results can be achieved with a deep architecture which develops hierarchical representations of learning. From Fig. 5, we can perceive the computational graph of our proposed approach (FakeBERT). In many existing and useful studies [6, 52], the problem of fake news has examined utilising a unidirectional pre-trained word embedding model followed by a 1D-convolutional-pooling layer network [52]. Our suggested model obtains the advantages of automated feature engineering approach [36]. In our model, inputs are the vectors generated after word-embedding from BERT. We give the equal dimensional input vectors to all three convolutional layers present in parallel blocks [26] followed by a pooling layer in each block. In our proposed model, the decision of chosen number of convolutional layers, kernels\_sizes, no. of filters, and optimal hyperparameters etc.[19, 26] to make our model more accurate as follows:

**Convolutional layer** The convolutional layer consists of a set of filters and kernels [52] for better semantic representations of words having a different length. The significant actions performed are matrix multiplications (non-linear operation) passes through an activation function to produce the final output. In our proposed model, we have used three parallel blocks of 1D-CNN having one layer in each block and two straight forward layers after the concatenation process with different kernel sizes and filters.

**Max-pooling layer** Max-pooling layer effectively down-samples [27, 36] the output obtained from the convolutional layer and reduce the number of computation operations

needed in the system. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. In our proposed model, we have used five max-pooling layers (three using parallel blocks of 1D-CNN and two with straight forward convolutional layers).

**Flatten layer** In between the convolutional layer and the fully connected layer, there is a Flatten layer. Flattening transforms a two-dimensional matrix of features into a vector that can be fed into a fully connected neural network classifier.

**Dense layer** A dense layer is just a regular layer of neurons in a neural network. Each neuron receives input from all the neurons in the previous layer, thus densely connected. The layer has a weight matrix  $W$ , a bias vector  $b$ , and the activations of previous layer  $a$ . In many existing and useful methods [36, 45], authors have mostly used one or two dense layers in their proposed networks to prevent over-fitting. In our proposed model, we have also taken two dense layers with a diverse number of filters.

**Dropout** Dropout is a regularization technique [36, 45] where randomly selected neurons are ignored during training. Its main contribution to the activation of downstream neurons is temporarily removed on the forward pass and any weight updates are not applied to the neuron on the backward pass. We have applied dropout to dense layers in the network. Dropout works by randomly setting the outgoing edges of hidden units to 0 at each update of the training phase. We have used the value of dropout is 0.2 in our investigations.

**Activation Function** ReLu refers to the Rectifier Unit, the most commonly deployed activation function [22, 41] for the outputs of the CNN neurons. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. ReLU is computed after the convolution and is a non-linear activation function like tanh or sigmoid. The equation of ReLU can be written as:

$$\sigma = \max(0, z) \quad (1)$$

here  $z = \text{input}$

**Loss Function ( $L$ )** The cross-entropy compares the model's prediction with the label which is the true probability distribution. The cross-entropy goes down as the prediction gets more and more accurate. It becomes zero if the prediction is perfect. As such, the cross-entropy can be a loss function to train a classification model. So predicting a probability of .014 when the actual observation label is 1 would be bad and result in a high loss value. In binary classification, where the number of classes ( $M$ ) equals 2, cross-entropy can be calculated as:

$$L = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2)$$

If  $M > 2$  (i.e. multi-class classification), we calculate a separate loss for each class label per observation and sum the result.

$$- \sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (3)$$

**Table 5** FakeBERT layered architecture

Layer	Input size	Output size	Param number
Embedding	1000	1000 × 100	25187700
Conv1D	1000 × 100	998 × 128	38528
Conv1D	1000 × 100	997 × 128	51328
Conv1D	1000 × 100	996 × 128	64128
Maxpool	998 × 128	199 × 128	0
Maxpool	997 × 128	199 × 128	0
Maxpool	996 × 128	199 × 128	0
Concatenate	199 × 128, 199 × 128, 199 × 128	597 × 128	0
Conv1D	597 × 128	593 × 128	82048
Maxpool	593 × 128	118 × 128	0
Conv1D	118 × 128	114 × 128	82048
Maxpool	114 × 128	3 × 128	0
Flatten	3 × 128	384	0
Dense	384	128	49280
Dense	128	2	258

Here  $y$  - binary indicator (0 or 1) if class label  $c$  is the correct classification for observation  $o$ ,  $p$  - predicted probability observation  $o$  is of class  $c$

We can observe the computational graph and layered architecture of our proposed FakeBERT model using Table 5 and Fig. 5. In this design, the input is scattered into three parallel blocks of 1D-CNN having 128 filters and one convolutional layer across each block. First convolution layer consists of 128 filters and kernel\_size=3, which reduces input embedding vector from 1000 to 998, second layer has 128 filters and kernel\_size=4, which reduces input vector from 1000 to 997, and third layer has 128 filters and kernel\_size=5, which decreases input vector from 1000 to 996. After a particular convolution layer, a max-pooling layer is also present to decrease the dimension. Subsequently, a max-pooling layer with kernel\_size=5 further reduces the vector to 1/5th of 996, i.e. 199. After concatenation of three above conv-layers, a convolution layer is applied having kernel\_size=5 including 128 filters.

**Table 6** Optimal hyperparameters with CNN

Hyperparameter	Value
Number of convolution layers	3
Number of max pooling layers	3
Number of dense layers	2
Number of Flatten layers	1
Loss function	Categorical-crossentropy
Activation function	Relu
Learning rate	0.001
Optimizer	Ada-delta
Number of epochs	10
Batch size	128

**Table 7** Optimal hyperparameters with LSTM

Hyperparameter	Value
Number of convolution layers	2
Number of max pooling layers	2
Number of dense layers	4
Dropout rate	.2
Optimizer	Adam
Activation function	Relu
Loss function	Binary-crossentropy
Number of epochs	10
Batch size	64

**Table 8** Optimal hyperparameters with FakeBERT

Hyperparameter	Value
Number of convolution layers	5
Number of max pooling layers	5
Number of dense layers	2
Number of Flatten layers	1
Dropout rate	.2
Optimizer	Adadelata
Activation function	Relu
Loss function	Categorical-crossentropy
Number of epochs	10
Batch size	128

**Table 9** Attributes in the fake news dataset

Attribute	Number of Instances
ID (unique value to the news article)	20800
title (main heading related to particular news)	20242
author (name of the creator of that news)	18843
text (complete news article)	20761
label (information about that the article as fake or real)	20800

**Table 10** Fake news dataset with the class labels

Class label	Number of Instances
True	10540
False	10260

Subsequently, there are two hidden layers having 384 and 128 nodes respectively. The number of trainable parameters across each layer is also presented (for more details refer column Param number) in Table 5. This model is not both computationally complex for training at any real-world fake news dataset. The work was carried using the NVIDIA DGX-1 V100 machine. The machine is equipped with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM and 1000 TFLOPS speed.

## 4 Experiments

Experiments have been conducted using deep learning models (CNN and LSTM) and our proposed model (FakeBERT) using pre-trained word embedding techniques (BERT and GloVe). Performances are recorded of different classification models and analyzed with the benchmark results.

### 4.1 Dataset description

In this paper, we have done extensive experiments using the real-world fake news dataset.<sup>1</sup> It (refer Table 9) consists of two files (i) train.csv, and (ii) test.csv: A testing dataset without the label. It is a collection of the fake and real news of propagated during the time of the U.S. General Presidential Election-2016. In Table 10, we can see the instances with the class labels in the respective fake news dataset.

#### 4.1.1 Hyperparameter setting

The selection of optimal hyperparameters is one of the main methods of any deep learning solution. Existing deep learning models explicitly define optimal hyperparameters that examine several factors such as memory and cost. Optimal selection of best numbers depends on the balanced or imbalanced dataset. For selecting optimal numbers, there are two fundamental approaches: automatic and manual selection. Both the methods are equally valid, but for manual selection, deep knowledge of the model is needed. For automatic selection, the high computational cost is required. From Tables 6, 7 and 8, we can observe the values of hyperparameters used in our investigations (Tables 9 and 10).

### 4.2 Evaluation parameters

To evaluate the performance of FakeBERT, we have considered the accuracy, cross-entropy loss, FPR (False Positive Rate), FNR (False Negative Rate), and confusion matrix (refer Table 11 for more details) as evaluation matrices.

<sup>1</sup> The dataset can be downloaded from: <https://www.kaggle.com>

**Table 11** Representation of confusion matrix

	Predicted negative	Predicted positive
Actual negative	True negative (TN)	False positive (FP)
Actual positive	False negative (FN)	True positive (TP)

## 5 Results and discussion

We have investigated and analyzed the results with several classifiers having different types of learning paradigms (different optimal hyper-parameters and architectures). Classification results demonstrate that the capability of automatic feature extraction with deep learning models plays an essential role in the accurate detection of fake news. Our proposed model (FakeBERT) produced more accurate results as compared to existing benchmarks with an accuracy of 98.90%.

### 5.1 Classification results using machine learning models

Firstly, several experiments conducted for estimating the performance of elected machine learning classifiers. (Multinomial Naive Bayes (MNB), Random Forest (RF), Decision Tree (DT), K-nearest neighbor (KNN)) using real-world fake news dataset. In our investigation, we have found that using MNB; we have achieved an accuracy of 89.97% with GloVe. Respective confusion matrix is shown in Table 12. Confusion matrices with others machine learning classifiers are shown in Tables 13, 14 and 15. The decision-tree algorithm also provides an accuracy of 73.65%. The confusion matrix using the MNB classifier predicts more labels accurately closer to actual labels with the testing dataset (for more details refer to Table 12). Dealing with the balanced dataset, MNB provided more accurate results. Machine Learning-based classification results are tabulated in Table 21 and Fig. 6. In this research, we have investigated the performance of different machine learning models with uni-directional pre-training model. In our investigation, we found that accuracy is not up to the mark with real-world fake news dataset. Further, a bidirectional training model which is a more powerful feature extractor [44] was on priority for investigation. Motivated this fact, we introduced BERT, a bidirectional transformer encoder-based pre-trained word embedding model. BERT is a more powerful feature extractor than GloVe and provides effective results for NLP-tasks. Experiments have been conducted using the BERT-based machine learning approach and achieved improved classification results. Deep Learning is a subset of Machine Learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts. One of the deep learning's main advantages over other machine learning is its capacity to execute feature engineering on its own. A deep learning algorithm will scan the data to search for features that correlate and combine them to enable faster learning.

**Table 12** Confusion matrix for MNB with GloVe

	Predicted negative	Predicted positive
Actual negative	853 (TN)	111 (FP)
Actual positive	73 (FN)	898 (TP)



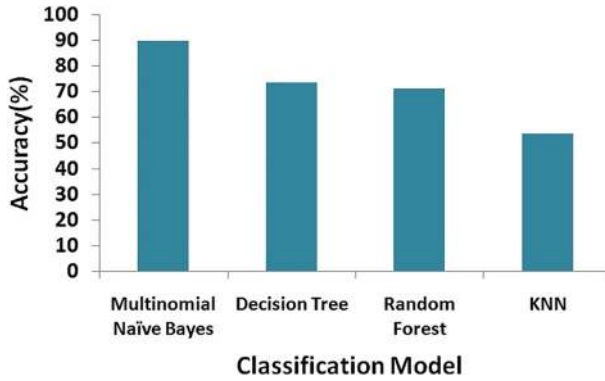


Fig. 6 Classification results with GloVe

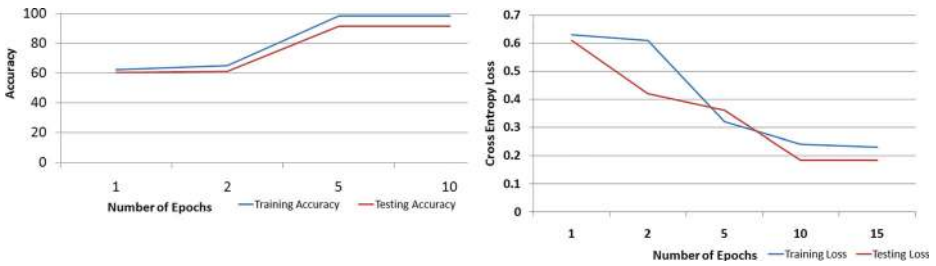


Fig. 7 Accuracy and cross entropy loss using CNN

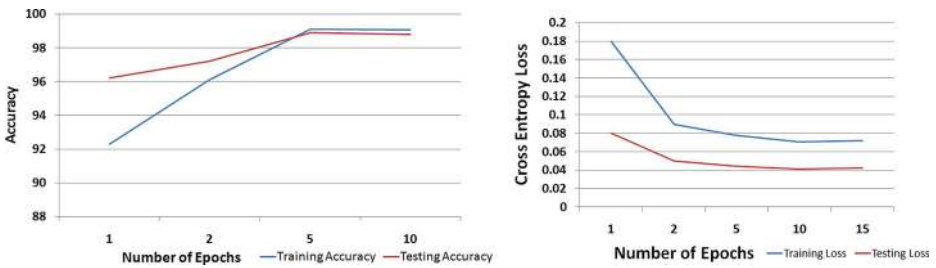


Fig. 8 Accuracy and cross entropy loss using FakeBERT

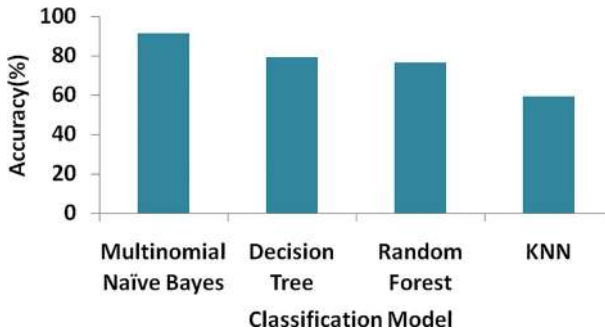


Fig. 9 Classification results with BERT

**Table 13** Confusion matrix for KNN with GloVe

	Predicted negative	Predicted positive
Actual negative	282 (TN)	762 (FP)
Actual positive	200 (FN)	836 (TP)

**Table 14** Confusion matrix for DT with GloVe

	Predicted negative	Predicted positive
Actual negative	631 (TN)	413 (FP)
Actual positive	135 (FN)	901 (TP)

**Table 15** Confusion matrix for RF with GloVe

	Predicted negative	Predicted positive
Actual negative	683 (TN)	361 (FP)
Actual positive	234 (FN)	802 (TP)

**Table 16** Confusion matrix for LSTM with GloVe

	Predicted negative	Predicted positive
Actual negative	1030 (TN)	8 (FP)
Actual positive	47 (FN)	995 (TP)

**Table 17** Confusion matrix for CNN with BERT

	Predicted negative	Predicted positive
Actual negative	1004 (TN)	63 (FP)
Actual positive	90 (FN)	942 (TP)

**Table 18** Confusion matrix for LSTM with BERT

	Predicted negative	Predicted positive
Actual negative	1032 (TN)	7 (FP)
Actual positive	44 (FN)	998 (TP)

**Table 19** Confusion matrix for FakeBERT with BERT

	Predicted negative	Predicted positive
Actual negative	1045 (TN)	6 (FP)
Actual positive	17 (FN)	1012 (TP)

**Table 20** Our proposed model vs existing benchmarks with real-world fake news dataset

Authors	Accuracy(%)
Ghanem et al [13]	48.80
Singh et al [42]	87.00
Ahmed et al [1] using LR-unigram model	89.00
Ruchansky et al [35]	89.20
Ahmed et al [1] using LSVM model	92.00
Liu et al [23]	92.10
O'Brien et al [28]	93.50
Our Proposed model (FakeBERT)	<b>98.90</b>

## 5.2 Classification results using deep learning models

To improve the classification results and to consider the issues in machine learning implementations, more experiments have been conducted with the deep learning models (CNN, LSTM, and FakeBERT) and recorded the performances with real-world fake news dataset. We have designed a deep convolutional network with BERT as a word embedding model. Our deep learning-based approach has built on the top of BERT. In our deep investigation, we have found that using the GloVe-based deep approach with Long Short Term Memory (LSTM) and convolutional neural network (CNN, we found the improved classification results with an accuracy of 92.70% and 97.55% respectively with 10 epochs. The respective confusion matrix show with the help of Table 16. Experiments have been conducted using CNN, LSTM, and our proposed BERT approach. Respective confusion matrices are shown with the help of Tables 17 and 18. Using BERT, we achieved a validation accuracy

**Table 21** Classification results with BERT and GloVe

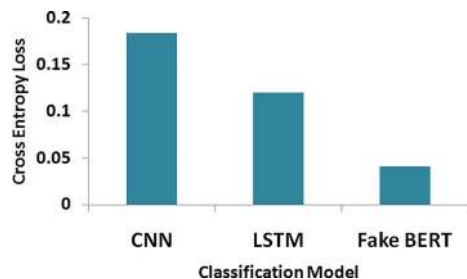
Word embedding model	Classification model	Accuracy (%)
TF-IDF (using unigrams and bigrams)	Neural Network	94.31
BOW (Bag of words)	Neural Network	89.23
Word2Vec	Neural Network	75.67
GloVe	MNB	89.97
GloVe	DT	73.65
GloVe	RF	71.34
GloVe	KNN	53.75
BERT	MNB	91.20
BERT	DT	79.25
BERT	RF	76.40
BERT	KNN	59.10
GloVe	CNN	91.50
GloVe	LSTM	97.25
BERT	CNN	92.70
BERT	LSTM	97.55
BERT	<b>Our Proposed model (FakeBERT)</b>	<b>98.90</b>

**Table 22** False Positive Rate (FPR) and False Negative Rate (FNR)

Word Embedding Model	Classification Model	FPR	FNR
TF-IDF (using unigrams and bigrams)	Neural Network	0.04684	0.0742
BOW (Bag of words)	Neural Network	0.1040	0.0862
Word2Vec	Neural Network	0.1320	0.3416
GloVe	MNB	0.1151	0.0752
GloVe	DT	0.3956	0.1303
GloVe	RF	0.3458	0.2259
GloVe	KNN	0.7299	0.1931
BERT	MNB	0.0985	0.0789
BERT	DT	0.1660	0.2429
BERT	RF	0.1245	0.3318
BERT	KNN	0.4037	0.4110
GloVe	CNN	0.0989	0.0776
GloVe	LSTM	0.0080	0.0482
BERT	CNN	0.0590	0.0872
BERT	LSTM	0.0077	0.0451
BERT	<b>FakeBERT</b>	<b>0.0160</b>	<b>0.0059</b>

of 92.70% with CNN and 97.55% with LSTM respectively with 10 epochs. We have found in our investigation that our BERT approach provided state-of-the-art results in fake news classification.

To validate the performance of our BERT-based deep learning model (FakeBERT), several experiments have been conducted with optimized hyperparameters. In our investigation, we have found that our model achieved more accurate results with an accuracy of 98.90%. A respective confusion matrix shows with the help of Table 19. In our approach, the selection of hyperparameters shows in Table 8. From Fig. 7, we can examine the accuracy and cross-entropy loss of our implemented CNN model with a real-world fake news dataset. As seen from Fig. 8, the training loss decays more quickly with BERT-based model as compared to the previous word embedding model (like GloVe, word2Vec etc.) From Fig. 5 and Table 5, we can observe the architecture of our implemented BERT-based model (FakeBERT). From Table 21, we can see the accuracy of the implemented FakeBERT model with 98.90% using the test set. As investigated above, the pre-trained embedding-based models consistently outperform with a significant margin of improvement. The training loss of BERT approach decays comparatively fast and without any inconsistencies. It shows clearly

**Fig. 10** Cross entropy loss with CNN, LSTM, and FakeBERT

from Fig. 9 that cross-entropy loss is reducing fastly using FakeBERT model. We achieved more accurate results with our proposed model as compared to other implemented models with minimal losses of data. To validate the performance of our recommended model; we have considered two more evaluations parameters (FPR and FNR). Results are tabulated in Table 22. In these results, it is clear that with our proposed model (FakeBERT), both FPR and FNR are minimum with the value of 1.60% and 0.59% respectively. It shows the performance of our proposed model with real-world fake news dataset. With other classification models, the values of FPR and FNR are high.

It perceived that using bidirectional pre-trained word embedding (BERT), leads to faster training of model and lower cross-entropy loss. Consistently in classification tasks, precision and recall improve when we use pre-trained word embedding (trained on a sufficiently large corpus). From Table 21, we can observe the results using both machine learning as well as deep learning models. It demonstrates clearly that our proposed model (FakeBERT) performs state-of-the-art results as compared to existing benchmarks results using different classification models. From Table 20, we can comprehend the comparative analysis of the proposed method with the existing benchmarks using the Kaggle real-world fake news dataset. It is a precise observation that the highest classification accuracy is reported with an accuracy of 93.50%. Table 21 demonstrates clearly that our proposed model gives comparatively more accurate results and better performances (testing accuracy, FPR, FNR, Cross-entropy loss). Cross-Entropy loss is also very less using BERT as training model (more details refer to Fig. 10). Using our BERT-based in-depth convolutional approach (FakeBERT), we were capable of achieving an accuracy of 98.90% as compared to 98.36% with GloVe (Table 22).

## 6 Conclusion and future scope

In this research, we have demonstrated the performance of our proposed model (FakeBERT- a BERT-based deep convolutional approach) for fake news detection. Our model is a combination of BERT and three parallel blocks of 1d-CNN having different kernel-sized convolutional layers with different filters for better learning. Our model is built on the top of a bidirectional transformer encoder-based pre-trained word embedding model (BERT). Classification results demonstrate that FakeBERT provides more accurate results with an accuracy of 98.90%. The accuracy of FakeBERT is better than the current state-of-the-art models with real-world fake news dataset: Fake-News. This dataset consists of thousands fake and real news articles during the 2016 U.S. General Preseantiaial Election. We evaluated our models with different parameters (Accuracy, FPR, FNR, and Cross-entropy loss).

In future work, we will design a hybrid approach (combining content, context, and temporal level information from news articles) applying for both the binary as well as multi-class real-world fake news dataset. This hybrid approach can be valuable to detect the instances of fake news for multi-label datasets which propagate in a graph. We will further study the problem of fake news from the viewpoint of different echo-chambers that exists in social media data, which can consider as a group of personalities having the same opinion for any social concern. The prime motivation to introduce echo-chambers is that every user is co-related in a graph like structure (not in isolation) to any social media platform like a community.

## References

1. Ahmed H, Traore I, Saad S (2017) 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, Cham, pp 127–138
2. Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. *J Econ Perspect* 31(2):211–36
3. Asparouhov T, Muthén B (2010) Weighted least squares estimation with missing data. *Mplus Technical Appendix* 2010: 1–10
4. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. *Inform Sci* 497:38–55
5. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on world wide web, pp 675–684
6. Cerisara C, Kral P, Lenc L (2018) On the effects of using word2vec representations in neural networks for dialogue act recognition. *Comput Speech Lang* 47:175–193
7. Chen W, Zhang Y, Yeo CK, Lau CT, Sung Lee B (2018) Unsupervised rumor detection based on users' behaviors using neural networks. *Pattern Recogn Lett* 105:226–233
8. Crestani F, Rosso P (2020) The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In: Natural language processing and information systems: 25th international conference on applications of natural language to information systems, NLDB 2020, Saarbrücken, Germany, vol 181. Springer Nature. Proceedings
9. De S, Sohan FY, Mukherjee A (2018) Attending sentences to detect satirical fake news. In: Proceedings of the 27th international conference on computational linguistics, pp 3371–3380
10. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Eugene Stanley H, Quattrociocchi W (2016) The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113(3):554–559
11. Devlin J, Chang M-W, Lee K, Kristina T (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (1)
12. Fazil M, Abulaish M (2018) A hybrid approach for detecting automated spammers in twitter. *IEEE Trans Inf Forensics Secur* 13(11):2707–2719
13. Ghanem B, Rosso P, Rangel F (2018) Stance detection in fake news a combined feature representation. In: Proceedings of the first workshop on fact extraction and VERification (FEVER), pp 66–71
14. Ghosh S, Shah C (2018) Towards automatic fake news classification. *Proc Assoc Inf Sci Technol* 55(1):805–807
15. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2016) LSTM: A search space odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232
16. Gorrell G, Kochkina E, Liakata M, Aker A, Zubiaga A, Bontcheva K, Derczynski L (2019) SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 845–854
17. Gupta M, Zhao P, Han J (2012) Evaluating event credibility on twitter. In: Proceedings of the 2012 SIAM international conference on data mining, pp 153–164. Society for industrial and applied mathematics
18. Jwa H, Oh D, Park K, Kang JM, Lim H (2019) exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT). *Appl Sci* 9(19):4062
19. Kaliyar RK, Goswami A, Narang P, Sinha S (2020) FNDNetA deep convolutional neural network for fake news detection. *Cognitive Systems Research* 61:32–44
20. Karimi H, Roy P, Saba-Sadiya S, Tang J (2018) Multi-source multi-class fake news detection. In: Proceedings of the 27th international conference on computational linguistics, pp 1546–1557
21. Kumar S, Shah N (2018) False information on web and social media: a survey. [arXiv:arXiv-1804](https://arxiv.org/abs/1804.08044)
22. Li Y, Yuan Y (2017) Convergence analysis of two-layer neural networks with relu activation
23. Liu Y, Yi-Fang BW (2018) Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-second AAAI conference on artificial intelligence
24. Malik S, Sentovich EM, Brayton RK, Sangiovanni-Vincentelli A (1991) Retiming and resynthesis: Optimizing sequential networks with combinational techniques. *IEEE Trans Comput-Aided Design Integr Circuits Syst* 10(1):74–84
25. Monteiro RA, Santos RLS, Pardo TAS, de Almeida TA, Ruiz EES, Vale OA (2018) Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: International conference on computational processing of the portuguese language. Springer, Cham, pp 324–334

26. Munandar D, Arisal A, Riswantini D, Rozie AF (2018) Text classification for sentiment prediction of social media dataset using multichannel convolution neural network. In: 2018 International conference on computer, control, informatics and its applications (IC3INA). IEEE, pp 104–109
27. Nagi J, Ducatelle F, Di Caro GA, Cireşan D, Meier U, Giusti A, Nagi F, Schmidhuber J, Gambardella LM (2011) Max-pooling convolutional neural networks for vision-based hand gesture recognition, IEEE
28. O'Brien N, Latessa S, Evangelopoulos G, Boix X (2018) The language of fake news: Opening the black-box of deep learning based detectors
29. Pérez-Rosas Verónica, Kleinberg B, Lefevre A, Mihalcea R (2018) Automatic detection of fake news. In: Proceedings of the 27th international conference on computational linguistics, pp 3391–3401
30. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proceedings of NAACL-HLT, pp 2227–2237
31. Qi Y, Sachan D, Felix M, Padmanabhan S, Neubig G (2018) When and why are pre-trained word embeddings useful for neural machine translation? In: Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: human language technologies, vol 2 (short papers), pp 529–535
32. Rashkin H, Choi E, Jang JY, Volkova S, Choi Y (2017) Truth of varying shades: Analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 conference on empirical methods in natural language processing, pp 2931–2937
33. Reema A, Kar AK, Vigneswara Ilavarasan P (2018) Detection of spammers in twitter marketing: a hybrid approach using social media analytics and bio inspired computing. *Information Systems Frontiers* 20(3):515–530
34. Roy A, Basak K, Ekbal A, Bhattacharyya P (2018) A deep ensemble framework for fake news detection and classification. arXiv:arXiv-1811
35. Ruchansky N, Seo S, Liu Y (2017) Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management. ACM, pp 797–806
36. Seide F, Li G, Chen X, Yu D (2011) Feature engineering in context-dependent deep neural networks for conversational speech transcription, IEEE
37. Shin J, Jian L, Driscoll K, Bar F (2018) The diffusion of misinformation on social media: Temporal pattern, message, and source. *Comput Hum Behav* 8:278–287
38. Shu K, Cui L, Wang S, Lee D, Liu H (2019) defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, pp 395–405
39. Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2020) FakeNewsNet: A data repository with news content, social context, and spatio temporal information for studying fake news on social media. *Big Data* 8(3):171–188
40. Shu K, Wang S, Liu H (2019) Beyond news contents: The role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp 312–320. ACM
41. Sibi P, Allwyn Jones S, Siddarth P (2013) Analysis of different activation functions using back propagation neural networks. *J Theor Appl Inf Technol* 47(3):1264–1268
42. Singh DSKR, Vivek RD, Ghosh I (2017) Automated fake news detection using linguistic analysis and machine learning. In: International conference on social computing, behavioral-cultural modeling, & prediction and behavior representation in modeling and simulation (SBP-BRIMS), pp 1–3
43. Tacchini E, Ballarin G, Vedova ML, Moret S, Hoax LucadeAlfaro. (2017) Some like it Della Automated fake news detection in social networks. In: 2nd workshop on data science for social good, SoGood 2017, pp 1–15. CEUR-WS
44. Tenney I, Das D, Pavlick E (2019) BERT rediscovers the classical NLP pipeline. In: Proceedings of the 57th annual meeting of the association for computational linguistics
45. Vasudevan V, Zoph B, Shlens J, Le QV (2019) Neural architecture search for convolutional neural networks. U.S Patent 10,521,729 issued December 31
46. Vosoughi S, 'Neo Mohsenvand M, Roy D (2017) Rumor gauge: Predicting the veracity of rumors on Twitter. *ACM Trans Knowl Discov Data (TKDD)* 11(4):1–36
47. Wang WY (2017) Liar, liar pants on fire: A new benchmark dataset for fake news detection. In: Proceedings of the 55th annual meeting of the association for computational linguistics (vol 2: short Papers), pp 422–426
48. Weiss AP, Alwan A, Garcia EP, Garcia J (2020) Surveying fake news: Assessing university faculty's fragmented definition of fake news and its impact on teaching critical thinking. *Int J Educ Integr* 16(1):1–30
49. Yang F, Liu Y, Xiaohui Y, Yang M (2012) Automatic detection of rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD workshop on mining data semantics, pp 1–7

50. Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing. *IEEE Comput Intell Mag* 13(3):55–75
51. Zhang X, Zhao J, LeCun Y (2015) Character-level convolutional networks for text classification. In: *Advances in neural information processing systems*, pp 649–657
52. Zhong B, Xing X, Love P, Wang Xu, Luo H (2019) Convolutional neural network: Deep learning-based classification of building quality problems. *Adv Eng Inform* 40:46–57
53. Zhou X, Zafarani R (2018) Fake news: a survey of research, detection methods, and opportunities. *arXiv:arXiv-1812*
54. Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: A survey. *ACM Comput Surv (CSUR)* 51(2):1–36

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.