Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
December 2023, vol. 28, no. 2, pp. 180–188
https://doi.org/10.2478/acss-2023-0018
https://content.sciendo.com

An Improved FakeBERT for Fake News Detection

Arshad Ali[1*], Maryam Gulzar[2]
[1] *FAST School of Computing, National University of Computer and Emerging Sciences, Lahore, Pakistan*,
[2] *Software Engineering Department, LUT University, Lappeenranta, Finland*

***Abstract* – In the present era of the internet and social media, the way of information dissemination has changed. However, due to rapid growth in the amount of news generated regularly and the unsupervised nature of social media, fake news turns out to be a big problem. Fake news can easily build a false positive or negative perception about a person, or an event. Fake news was also used as a tool by propagandists during the Coronavirus (COVID-19) pandemic. Thus, there is a need to use technology to tag fake news and prevent its dissemination. Previously, different algorithms were designed to detect fake news but without considering the semantic meaning and long sentence dependence. This research work proposes a new approach to the detection of fake news in the context of COVID-19. The suggested approach uses a combination of Bidirectional Encoder Representations from Transformers (BERT) for extracting context meaning from sentences, SVM for pattern identification to detect fake news in a better way from the COVID-19 dataset, and an evolutionary algorithm called Non-dominated Sorting Genetic Algorithm II (NSGA-II) to distribute text for Support Vector Machine (SVM) classification. The suggested approach improves accuracy by 5.2 % by removing a certain amount of ambiguity from sentences.**

***Keywords* – Covid-19, fake news, semantic analysis.**

I. Introduction

With the advent of the internet and the arrival of social media websites, people and organisations can easily and quickly share their ideas and beliefs to the masses, in a short amount of time. Different kinds of news are conveyed by different social media outlets to end consumers. The use of electronic media can quickly transmit fake news which can cause various problems such as biased views about the perception of different news. In recent years, some newspapers have used fake news to gain political and financial advantage. The political landscape and public opinion can be altered easily by manipulating the news as per their interest resulting in manipulation of results.

In 2020, the coronavirus spread from China to the whole world affecting the lives of people directly. The virus affected various people across the globe and many people feared encountering the virus. During the spread of the virus, many news outlets and social media spread news that could not be identified as fake or real news.

Different people claimed different kinds of cures against this virus which were later found to be false. Due to the unsupervised nature of the internet and especially social media, no check exists on whether news is false or true. Due to fake news, panic and fear were created forcing people to buy food in bulk. Australia prepared tissue papers, but a spread of fake news that Australia imported tissue papers resulted in a shortage of the same. Such propagation of fake news increased anxiety and fear among people which was later used for opinion alteration and urgency.

The research community is working on the detection of fake news using social media. Many people are affected, directly and indirectly, by the influence of social media in their daily lives. The relationship between the users and social media websites can be seen as influenced by different fake social trends. Detection of such engagements can help in detecting fake news. A. Koirala [1] prepared a fake news dataset called COVID-19. The behaviour of users on social media websites hints at their mental and cognitive thinking about such social media websites [2]. Different propagation methods are used to convince the user of fake news. Biased comments toward the news give these websites an unparalleled experience to spread fake news to different people. Many pre-trained word embedding techniques such as Global Vectors for word representation (GloVe), Term Frequency-Inverse Document Frequency (TF-IDF), and Word2Vec are used to extract features from the news websites to detect fake news [3].

Many Machines Learning (ML) algorithms-based techniques are used to detect fake news [3]. We can categorize fake news into two types: (i) news based and (ii) media based. The research community is using various techniques to extract relevant news from social media and news sites, and different extraction methods to extract different features from the information. Researchers are considering more advanced methods like BERT to detect fake news to obtain better results.

II. Research Problem

With the rise of fake news in different electronic media, many machine learning algorithms are implemented to detect fake news. Although a recently proposed FakeBERT [4] provides improved results, there is still room for improvement, which motivates us to try a hybrid model. This study tries to answer the following research question: "*Can preserving the context of the sentence give us better results in detecting fake news for the COVID-19 dataset*?"

* Corresponding author's e-mail: arshad.ali1@nu.edu.pk

©2023 Arshad Ali, Maryam Gulzar.
This is an open access article licensed under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0).

Many ML models relying on traditional methods are used to classify the news as fake or real. These methods work well with specific datasets and fail to provide similar results when the dataset gets changed. This problem is solvable by using a hybrid model that can not only preserve context in long sentences but also give good results on different datasets. The existing algorithms do not maintain context in the long run and their architecture cannot understand the context in any text. Maintaining a meaningful semantic meaning in a sentence gives an advantage to ML algorithms. To give contextual meaning to a sentence, different algorithms are developed. In recent years, many organisations are making use of evolutionary algorithms to solve their problems.

Suggested Solution: This work suggests a model that detects fake news, which is the combination of BERT [5], NSGA-II, and SVM. The BERT gives context to the text and SVM classifies it after training on a dataset. NSGA-II distributes datasets for SVM classification. This hybrid model tries to detect fake news with better results compared to other models, including one suggested by authors of [4].

III. Domain Knowledge

A. Fake News

Fake news is the kind of news that does not communicate the correct information to the end user. The aim of fake news is to create a false negative (or even positive) perception or provoke users so that a specific bias about a person, event, or policy can be created. Individuals, organisations, and groups use electronic media and social media websites to convey their message to society. Sometimes false or misleading information is generated to influence the readers and prompt a change in their opinion. Fake news impacts the credibility of real news as well. Fake news can easily proliferate, particularly in times of political turbulence and instability. During the COVID-19 pandemic, fake news was used as a tool to create panic, fear, and bias. Some examples of fake news are as follows:

- Many fake coronavirus cures propagate on social media [6];
- Use of fake vaccination news to put people in fear and anxiety [7];
- Use of many media outlets fake news [8] by Donald Trump during his election campaign and his tenure as predicant of the United States of America.

B. Pre-processing

In pre-processing, data received from different sources are changed to a different format for any model to process. The received raw data are transformed into meaningful data which can be easily processed by any algorithm. This pre-processing is used to check the quality of the data. First, it is checked whether the data are correct or not. Then, it is checked whether the records are correctly labelled or not. Then, data are converted into meaningful simple words to reduce the ambiguity of words. Inconsistency can be removed from the sentences after verification of data accuracy and availability. A few works use pre-defined datasets to get data for experiments while some use different sources to collect data. It is important to ensure that data are available and complete. Data sources use the same standard to label records. Data mismatch can cause different types of problems, which can be difficult to understand for algorithms. This work applies various pre-processing techniques such as tokenization, stop word removal, lemmatization, and padding.

Tokenization: The purpose of tokenization [9] is to split the records into smaller parts such as paragraphs, sentences, and words. Tokenization is performed on sentences to break them into words.

Stop Word Removal: Natural language consists of multiple parts of different vocabulary. Stop words [10] are those words that consistently occur in natural languages and mainly include pronouns, prepositions, and conjunctions. They help complete a sentence by connecting words. However, stop words are not considered important words that can give any meaningful context to the document. Therefore, these are not used in Natural Language Processing (NLP) applications. Removing stop words from a dataset can not only reduce the size of the document but also help in the fast execution of any model.

Lemmatization: Lemmatization [11] is a process of morphological computation of words. The group of the same root or lemma is made together which can be analysed easily later on. To create a link to its lemma, a dictionary is needed.

Padding: Padding is used to make every record the same by adding empty 0 numbers at the end to make every sentence equal in size before feeding as input to any model. For example, the BERT layer ensures that the tokens passed to the model are of the same size.

C. Semantic Analysis

Semantic analysis [12] is considered an important field of NLP that helps in understanding the meaning of natural language. For humans, it is very easy to understand natural language but the same becomes a challenging task for machines. Analysing and understanding the complexity of natural language is not easy for a machine. Semantic analysis considers context, structuring, and grammar to understand the meaning of natural language. Text disambiguation and relationship extraction are major processes to understand the context of a sentence.

Text disambiguation is used to understand the meaning of the word in the context of usage in the sentence. Every word varies in a sentence in the sense of meaning in natural language. It is dependent on the usage of the word in the sentence and context. For example, the word “Bat” can be an animal or “Cricket Bat”. Likewise, the word “crane” may be an animal or a machine. The ability of a machine to understand ambiguous words is called text disambiguation.

Relationship extraction is used to know different words in the sentence and identify the relationship between them. For example, Paris in France can be interpreted in two different ways, i.e., (i) as a question and (ii) as information.

Semantic classifiers and semantic extractors are the main techniques used in semantic analysis, while sentiment analysis is used to see the dependency of every word in the sentence, i.e., to detect sentiment inside a text. Sentiments are categorized

into three types: (i) positive, (ii) negative, and (iii) neutral. On social media, people share their opinions about different products. If a client of a certain company writes on a social media platform like Facebook that "This company is a joke." Then, it is possible that a joke can be considered a positive word by a machine without analysing the sentiment. Various ML approaches are used to find sentiment in the text. For example, BERT is one of the most widely used machine learning approaches.

Semantic extraction models are used to extract relationships from sentences. Keyword extraction is used to extract the relationship of the sentence by using BERT. Keyword extraction is used to extract words that are relevant to the text and expressions on a very small base level. Different keywords are analysed and classified to detect which words are positive or negative. It helps in giving insight into the intent of the text.

D. BERT

BERT [13], developed by Google and pre-trained on a large English dataset, is a transformer-based language representation model. Firstly, it splits sentences into tokens and then encodes these words or tokens into numeric forms as classifiers only understand numbers. BERT consists of two steps, i.e., pre-training and fine-tuning. In pre-training, a large amount of unlabelled text is given as input to the BERT model for training where it learns the contextual relations between words and sentences in the language. BERT has two sub-models: masked language modelling and next-sentence prediction. The former takes in a sentence with some masked words, and the latter needs to predict the masked words. During fine-tuning, the last output layer of BERT is replaced by a new fully connected layer trained for the specific task. The fake news is labelled as 0, while the real news is labelled as 1. BERT relies on a self-attention mechanism because different words in the same context or the same words in different contexts can change the meaning of a sentence. The context of a word is the words preceding or following the word. BERT is bi-directional and processes sentences from both left-to-right and right-to-left directions. This enables BERT to capture the word context dependency in a better way. BERT can convert words into word encodings and embeddings. This work uses BERT to encode the text. BERT encoder converts every 3116 sentences into a vector of length 100, in which every single value contains a 768-sized vector.

E. NSGA-II

The evolutionary algorithms (EAs) [14] are considered a natural phenomenon in which populations containing different solutions are evolved into the next generation. This next generation is dependent on the fitness of the problem. An evolutionary algorithm is used to solve different problems related to optimization. Evolutionary algorithms use nature-inspired processes to solve complex problems. EAs are preferred over traditional methods as they provide near-optimal, if not optimal, solutions very quickly.

NSGA-II or non-dominated sorting genetic algorithm II [15] is considered one of the best evolutionary algorithms for solving multiple objective optimization problems having three main characteristics: (i) non-dominated sorting, (ii) crossover operator, and (iii) mutation for creating different genes in the population. It consists of five steps. (i) Population initialization initializes the population based on the problem range and constraint and genes are usually randomly created from the input data. (ii) Non-dominated sort sorts the population based on non-domination, grouped by fronts. (iii) Selection removes the worst-performing individuals from the population. Since the sorting algorithm groups individuals by fronts, sometimes the fronts need to be split. The individuals that reside in the non-dense areas are selected using crowd distancing. (iv) Genetic operators use the crossover to create a different set of offspring and parents. Usually, a one-point crossover is used to divide the gene into different parts and combine it with other parts of the same size. (v) Recombination and selection create a new generation of the population by combining the individuals after selection, mutation, and crossover. The resulting individuals are called offspring.

F. SVM

SVM is a supervised ML approach that is applied to binary classification problems. After feeding labelled data to the SVM classifier, it can classify data into different classes. It works well and very fast with a limited number of samples due to which it is mostly considered in text classification problems. In SVM, the data are plotted onto n-dimensional space where n is the number of features. We then classify the different classes by finding the hyperplane. SVM is used for the classification of discrete labels. The main disadvantage of SVM is that it can take a long time to train and test a large dataset.

IV. Literature Review

This section mainly reviews studies focusing on detection of fake news [16]–[24], [4]. However, a few example studies pertaining to sentiment classification using BERT and/or SVM [5], [25], [26], text summarization via multi-objective optimization [27], [28] are also covered considering that either BERT/SVM is used or evolutionary/multi-objective optimization approaches are used in such studies.

Lee et al. [5] used BERT to find stock market sentiment classification. This work considered all the news from the U.S. stock market from a news website. BERT is used to find if the sentiment in any chosen news is positive or negative for the stocks. Authors claimed that the BERT model achieved good results which can make a relationship between stock market news and stock market prices. Lui et al. [25] used a combination of the Bayesian network with BERT to classify a Chinese dataset into two categories. The data categorized by the Bayesian network were fed to the BERT classifier and achieved an accuracy of 94 %. The suggested approach was compared with traditional methods and a gain of 4 % was observed. It was found that simple BERT and Bayesian systems separately performed worse than the proposed model. Hassonaha et al. [26] introduced a hybrid system that uses SVM to classify emotions collected by considering different accounts on Twitter into three different categories. Relief and multi-verse optimizer algorithms were used for feature

extraction from the dataset. The suggested method provided better results in terms of feature extraction and accuracy.

Fors-Isalguez et al. [27] proposed a method to generate text summarization based on the query by using word embedding and evolutionary algorithms. Pre-processing of data was done by cleaning and removing unrelated stop words from the text. A corpus, namely, TAC-2009 was used to train and test the data in the beginning. Later MOED was used to evolve the dataset to make a summary of the text provided. Three different experiments were conducted to see the effect of the proposed model. For semantic meaning, the simple word2vec was used to give meaning to the context of sentences. It was claimed that the text summarization created by the suggested model was better in terms of semantic meaning and context. Alqaisi et al. [28] proposed a text summarization model with the combination of an evolutionary algorithm named NSGA-II and a clustering-based algorithm. A multi-language dataset, namely, TUC-2011 was used, however, with a focus on Arabic data. Text pre-processing techniques, i.e., removing stop words, skewing, and tokenization of the words were performed. Each sentence was represented using a bag of words. Clustering was used to identify topics that were present in the text. To summarize the text, NSGA-II was used to add diversity and relevancy. In the end, sentences were combined to get a summary of the topic. It was observed that the F1-score improved by 8.46 % as compared to existing algorithms.

Authors of [22] explored different linguistic features to detect fake news in the Portuguese language using different machine learning techniques. The suggested machine learning model, namely, SVM achieved an accuracy of 49 % on the chosen dataset. This work used different automatic methods to get fake content from different social media websites. To detect different linguistic features in fake news, authors used news websites to analyse the credibility of news. Different domains, such as science, business, and sports, were considered to analyse the news data.

Lui et al. [17] used BERT as the encoder to represent news. Due to training on large-scale data, BERT increases the knowledge of the language by capturing meaningful meanings to train the model. Based on BERT, authors fine-tuned a fake news dataset, and the effect of the short news representation was given attention to some extent. Data were divided into multiple labels to distinguish the articles from the news.

Jwa et al. [18] proposed exBAKE that extends the pre-trained BERT model to understand the meaning of different articles. exBAKE built a relationship between headlines and new articles to predict the authenticity of the news. For experiments, the Fake News Channel-I (FNC-1) dataset was used, and the proposed model outperformed the basic BERT. The F1-score rose to 0.14 after using this model. Jain et al. [19] used SVM to detect fake news. The suggested approach classified data into two categories, either true or false. This work collected data from different websites, then extracted features from the dataset, and then applied different techniques like SVM and naïve bayes. SVM achieved the highest accuracy of 93.50 %.

Zhang et al. [20] proposed a BERT-based Domain Adaptation Neural Network (BDANN) model that can detect fake news from different social media websites such as Twitter and weibo. BDANN consisted of three parts: feature extraction, a classifier to identify the domain, and a classifier to detect fake news. Feature extractors extracted different features from images and text and removed the dependency of events by adding them to the domain classifier. The suggested approach detected fake news with an accuracy of 86 %.

Kaliyar et al. [21] applied CNN to detect fake news using GloVe as a pertained model which was trained to work on test data, and experiments were performed on datasets. This work calmed an accuracy of 98.36 %. GloVe, an embedding vector technique, was used to classify news as fake or real.

Umer et al. [22] used a hybrid model comprising CNN and LSTM to detect fake news. The data from a fake news website were fed to principal component analysis and chi-square algorithms, which provided contextual meaning to the sentences. According to the authors, 97.8 % accuracy was achieved. It was observed that Principal Component Analysis (PCA) performed better than chi-square. Pre-processing was done on the dataset to remove any non-alphabet characters. CNN was used to keep the contextual meaning in contact. Then, data were passed from CNN to Long Short-Term Memory (LSTM) to obtain the final output.

Nasir et al. [23] used a combination of CNN and RNN. CNN was used to extract features from the dataset and RNN was used to learn long-term dependencies. A CNN layer of Conv1D was used for processing the input vectors and extracting the local features that resided at the next level. CNN and RNN layers were interconnected with each other. The CNN layer forwarded the learned features to RNN layers to classify the news article data in a better way. The RNN layer was able to differentiate whether the news was fake or real.

Ahmad et al. [24] proposed an evolutionary method to detect fake news in COVID-19 news. This study employed three algorithms: genetic, particle swarm optimization, and salp swarm, on the COVID-19 dataset [1]. An accuracy of 76 % was achieved by the model.

Kaliyar et al. [4] provides a Fake BERT system that used BERT with Deep Learning (DL) techniques to identify fake news. BERT generated a vector after word embedding and then passed it to three input layers in parallel blocks. There was a pooling layer in each block. The authors used different numbers of layers, kernel sizes, filters, and optimal hyper-parameters to tune the model to obtain accurate results. Almost 98.90 % of accuracy was achieved by Fake BERT. The work compared the suggested model with other simpler algorithms and achieved better results.

Table I provides a summary of existing works in terms of the used algorithm, dataset, and accuracy results.

TABLE I
SUMMARY OF RELATED STUDIES

Sr. #	Name	Architecture	Dataset	Comments
1	FakeBERT [4]	CNN+BERT	2016 US elections (Kaggle)	98.90 % of accuracy
2	BDANN [20]	BERT	Twitter and Weibo	86 % of accuracy
3	BERT-based stock market Sentiment Analysis [5]	BERT	StockTwits	88 % of accuracy
4	Bayesian network with BERT [25]	Bayesian+BERT	Chinese dataset	94 % of accuracy
5	FNDNet [21]	Different layers (Word embedding/CNN)	Kaggle fake news dataset	98.36 % of accuracy
6	Smart System to detect fake news [19]	SVM/Naïve Bayes/NLP	Using an Authenticated news site	93.60 % of accuracy
7	exBAKE [18]	Predefined BERT with Headline and comments	FNC-1 dataset	F1-score improved by 0.14 %
8	Text summarization using Evolutionary algorithm [27]	NSGA-II and Clustering	TUC-2011	F1-score improved by 8.46 %
9	Text summarization using Evolutionary algorithm and word embeddings [28]	Word embedding/MOED	TAC-2009	Accuracy improved by 3.4 %
10	Evolutionary Fake News Detection Method [24]	Particle swarm optimization/Binary genetic +KNN	COVID-19 news data [1]	75 % of accuracy

Research gap: To the best of our knowledge and analysis, the existing works except [4] do not consider semantic analysis as an important process while detecting fake news. Traditional machine learning algorithms work well with specific datasets; however, their performance degrades significantly if the dataset changes. Machine learning models heavily rely on dataset records to improve results in terms of accuracy. In the case of long sentences, it is very tough for machine learning models to perform well, and accuracy decreases drastically. Thus, there is a need for a hybrid model that can solve the problem not only by preserving the context in log sentences but also by providing good results with different datasets.

The present work suggests a hybrid approach for the prediction of fake news in the COVID-19 dataset. The suggested hybrid model is a combination of BERT for sentence encoding, NSGA-II to distribute dataset records according to the fitness function, and SVM to classify data (i.e., predict) which is trained and tested whenever the fitness function of NSGA-II is called.

V. PROPOSED HYBRID MODEL

The section presents the proposed hybrid model which improves the prediction of fake news in the COVID-19 dataset.

Figure 1 presents the methodology of this work. After dataset collection, the cleaning and pre-processing of the dataset are performed. A pre-trained BERT can convert text into word embedding with contextual meaning. Next, NSGA-II distributes the dataset to run on the classifier SVM in its fitness function.

It keeps running until the stopping condition is met. In this case, when the 100th generation of a child becomes part of the population, the algorithm stops.

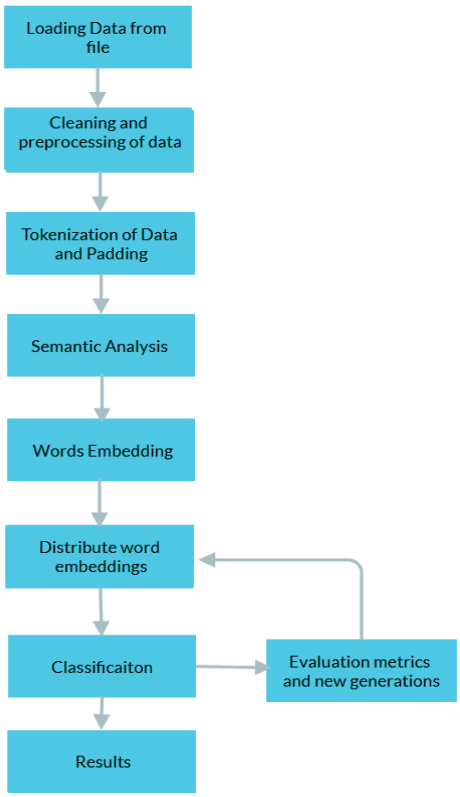


Fig. 1. The main flow of the proposed hybrid model.

A. Dataset Description

The COVID-19 dataset [1] considered for this study contains five columns and only two columns are used for data analysis and model training, namely, text and label. The text column contains articles from the COVID-19 news collected from many authentic sites. The label column is used to tell if the news is fake or not. The label column contains two values. One is 0 and the other is 1. There are a total of 3116 values in the dataset.

The most used words are China, said, corona, and virus as shown in Fig. 2. The class distribution of real and fake news is shown in Fig. 3. The number of real news samples is higher than fake news samples.

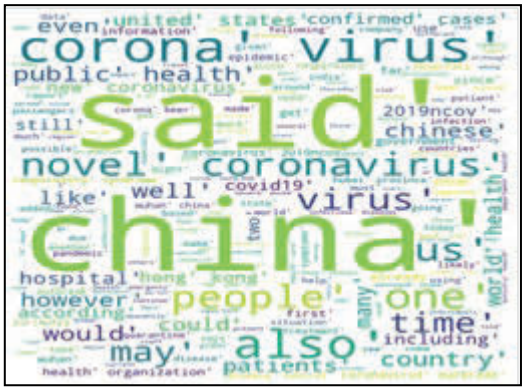


Fig. 2. Common words used in COVID-19 dataset.

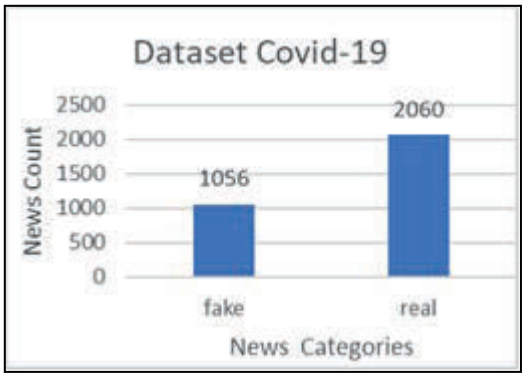


Fig. 3. COVID-19 dataset categories.

B. Pre-processing of Data

The data were first pre-processed. Figure 4 shows the various stages of the pre-processing phase. Sentences usually contain a lot of redundant information such as stop words that need to be removed.



Fig. 4. Basic flow of pre-processing applied onto the COVID-19 dataset.

First, the records were converted into tokens or words. All unimportant words and punctuation were removed. Words that were not in English grammar were also removed. This was followed by the removal of the stop words. Then lemmatization was applied to the text to normalize all inflected and derivative forms of words to one form. As a result, words like 'cat' and 'cats' were normalized to 'cat'. In a way, this reduced the number of unique words in the dataset. Finally, padding was added to make sure that every sentence was of the same length. The outcome of the pre-processing phase was reduced data. This small size of the records or data helped speed up the training process of the classification models. Table II demonstrates an example text from the COVID-19 dataset [1] along with its tokenization, stop word removal, lemmatization and padding. The new article was reduced to tokens and fed to a pre-trained model and data were converted into tokens to generate encodings at the word level.

C. Encoding Sentiments with BERT

A pre-trained BERT model was used for encoding the data or tokens. BERT is pre-trained over a large corpus of Wikipedia articles. Each sentence is converted into a word embedding of 100 which is then passed to NSGA-II to optimize and find the best population to be used for training and testing.

D. NSGA-II Shuffling and SVM Classification

NSGA-II algorithm is used to distribute word embedding in its fitness function; SVM is used to classify the data for every individual. It is run for 100 generations. A population of 10 is used with random assignment of word embedding. One point crossover is used to make new generations of parents and offspring with a probability of 50 %. For mutation, a probability of 40% is used. A random generation is going to tell that if the number is less than 40 then it can do mutation. NSGA-II distributes records of the dataset according to fitness function and tries to make such records that give better results.

SVM is trained and tested whenever the fitness function of NSGA-II is called. In every generation, every gene fitness function is calculated.

E. Evaluation Metrics

Four evaluation metrics, namely, accuracy, F1-score, precision, and recall were used to evaluate the proposed approach. Accuracy is the ratio of all the correct predictions to the overall predictions of the model. Precision is obtained by dividing the total positives by the sum of false and true positives, while recall is obtained by dividing the total positives by the sum of false negatives and true positives. F1-score indicates the accuracy of the model on the dataset.

TABLE II
AN EXAMPLE OF A RECORD FROM THE DATASET APPLYING PRE-PROCESSING

Sentence	Tokenization	Removing stop words	Lemmatization	Padding
Facebook has shuttered a popular group for Michiganders who oppose their governor's extreme lockdown measures	facebook', 'has', 'shuttered', 'a', 'popular', 'group', 'for', 'michiganders', 'who', 'oppose', 'their', 'governor', 's', 'extreme', 'lockdown'…	facebook', 'shuttered', 'popular', 'group', 'michiganders', 'oppose', 'governor', 'extreme',	facebook', 'shutter', 'popular', 'group', 'michigand', 'oppos', 'governor', 'extrem', 'lockdown'…	'facebook', 'shutter', 'popular', 'group', 'michigand', 'oppos', 'governor', 'extrem', 'lockdown',",",",

VI. Results and Discussion

This section presents the experimental setup and discusses the obtained results.

A. Experimental Setup

Different experiments were conducted on the COVID-19 dataset to compare the performance of different machine learning, deep learning, and hybrid techniques. In the first phase, ML and DL models were compared.

The pre-processing phase remains the same during all experiments. In the first phase, the data are encoded using TD-IDF. The dataset is split into training and testing ratio of 90:10. Six different algorithms are trained and tested on the dataset to check the accuracy. A pre-defined BERT model is used. It uses 12 hidden layers with an attention head with parameters of $H = 768$ and $A = 12$. For CNN, the Conv1D model is used.

In the second phase, hybrid models are used. First, BERT and CNN are combined. This model is used as a baseline to compare the result with the proposed hybrid model. Bert-base-uncased is used as a pre-defined BERT model to encode the dataset. The Bert base consists of 12 layers with a hidden size of 768. The size of the attention head is 12, and 110M parameters are used to train the model. For CNN, Conv2d was used. The max size of the sentence input to BERT is 100, where each word is embedded by a vector of 768 dimensions. CNN model is used to classify the output from the BERT-based encoder. In another experiment, SVM is used for classification instead of CNN.

In both experiments, the dataset is split into the same training-testing ratio. Finally, another baseline model as provided in [24] is used that combines particle swarm optimization with K-nearest neighbours. Lastly, in the fourth experiment, BERT, NSGA-II, and SVM combination are used. NSGA-II is run for 100 generations with a population of 10. 40 % mutation and 50 % one-point crossover are applied to create offspring. For classification, SVM is used to classify the word embedding received from the BERT layer.

B. Results

It was found that BERT performed worse than CNN as evident from Table III. The performance of CNN was 12.5 % better than BERT in terms of accuracy. This motivates us to use BERT with other machine learning algorithms to get better results as compared to other algorithms. In traditional machine learning techniques, CNN performed much better as compared to other models. CNN performed almost 4.1 % better than SVM. CNN performed better for small text; however, its performance was not comparable with other hybrid models due to some long sentences in the COVID-19 dataset [1].

Table IV clearly shows that after running the experiments, the results of the proposed hybrid model are 5.2 % better than the PSO + K-Nearest Neighbours model [24] in terms of accuracy.

All three hybrid models performed better than traditional models in terms of accuracy. 2.5 % of accuracy was improved from a simple BERT+SVM model with the usage of an evolutionary algorithm to create a population with the help of gene mutation and crossover.

This clearly shows that traditional methods of machine learning are more focused on characteristic words and categories. They ignore the structure and context of the sentence due to which important information is lost. The BERT model considers the semantic and contextual information of the sentence. The combination of BERT and other machine learning models with NSGA-II not only considers the characteristics of words important but also gives semantic meaning to the sentences in the records; thus, the accuracy improves further.

It is evident from Table IV that F1-score of the proposed model, i.e., BERT+NSGA-II+SVM is better as compared to other models. The precision score is less than KSO + K-Nearest Neighbours algorithm. BERT+NSGA2+SVM performs better with respect to recall. Among machine learning models, CNN performs better; but its performance is lower as compared to hybrid. It was only able to perform better in terms of F1-score than a hybrid model. CNN in terms of accuracy performs worse than hybrid models.

TABLE III
Results after Running Different Traditional Models on COVID-19 Dataset

Models	Precision	Recall	F1-score	Accuracy
Naive Bayes	0.57	0.52	0.55	60 %
SVM	0.65	0.71	0.68	69 %
Random Forest	0.63	0.7	0.66	68 %
SVC	0.61	0.56	0.59	63 %
Gradient Boosting	0.72	0.68	0.7	66 %
K-Neighbours	0.69	0.65	0.7	64 %
BERT	0.72	0.63	0.67	63 %
CNN	0.73	0.82	0.77	72 %

TABLE IV
Results of Different Hybrid Models on COVID-19 Dataset

Models	Precision	Recall	F1-score	Accuracy
PSO + K-Nearest Neighbours [28]	0.88	0.76	0.79	76 %
BERT + CNN [4]	0.73	0.77	0.75	77 %
Proposed BERT+ SVM	0.77	0.77	0.77	78 %
Proposed BERT + NSGA-II + SVM	0.76	0.9	0.83	80 %

VII. Conclusion

The research suggested a model for the COVID-19 fake news dataset and demonstrated its performance. The suggested model is a combination of BERT, NSGA-II, and SVM. BERT helps in keeping the contextual and semantic meaning of the sentences. NSGA-II changes the word embedding orders received from the BERT encoder and optimizes word embedding by mutation and crossover.

NSGA-II fitness function is to maximize accuracy and F1-score. SVM is used to classify word embedding in the fitness function of NSGA-II. The result of the proposed hybrid model

is much better than that of simple models and hybrid models. The accuracy and F1-score of the BERT+NSGA-II+SVM model were 5.2 % better than the particle swarm optimization model [24]. The suggested hybrid model performed much better than other models in terms of accuracy and F1-score. Moreover, it was observed that hybrid models performed better than traditional machine learning algorithms.

In future, different hybrid models can be implemented to achieve the goal of improving different evaluation metrics of different datasets. They can be modified so that they are not only used to classify binary classes but multiple classes as well.

References

[1] A. Koirala, "COVID-19 fake news dataset", Mendeley Data, V1, Feb. 2021. https://doi.org/10.17632/zwfdmp5syg.1

[2] S. Singhania, N. Fernandez, and S. Rao, "3HAN: A deep neural network for fake news detection," in *Proc. of the International conference on neural information processing*, Guangzhou, China, Oct. 2017, pp. 572–581. https://doi.org/10.1007/978-3-319-70096-0_59

[3] V.V. Hirlekar and A. Kumar, "Natural language processing based online fake news detection challenges – a detailed review," in *Proc. ICCES*, Coimbatore, India, Jun. 2020, pp. 748–754. https://doi.org/10.1109/ICCES48766.2020.9137915

[4] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications,* vol. 80, pp. 11765–11788, Jan. 2021. https://doi.org/10.1007/s11042-020-10183-2

[5] C. Lee, Z. Gao, and C. Tsai, "BERT-based stock market sentiment analysis," in *IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, Taoyuan, Taiwan, Sep. 2020, pp. 1–2. https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258102

[6] R. Heilweil, "Coronavirus scammers are flooding social media with fake cures and tests," Apr. 2020. [Online]. Available: https://www.vox.com/recode/2020/4/17/21221692/digital-black-market-covid-19-coronavirus-instagram-twitter-ebay (Accessed: December 20, 2021).

[7] S. Reilly, J. Palamdino, J. Lambert, and M. Stiles, "Fake vaccine cards are everywhere. It's a public health nightmare," Grid News, 2022. [Online]. Available: https://www.grid.news/story/science/2022/01/25/fake-vaccine-cards-are-everywhere-its-a-public-health-nightmare/ (Accessed: March 20, 2022).

[8] Woodward and Alex, "'Fake news': A guide to Trump's favorite phrase – and the dangers it obscures," Independent Digital News and Media, 2020. [Online]. Available: https://www.independent.co.uk/news/world/americas/us-election/trump-fake-news-counter-history-b732873.html (Accessed: December 20, 2021).

[9] F. Barcala, J. Vilares, M. Alonso, J. Grana, and M. Vilares, "Tokenization and proper noun recognition for information retrieval," in *Proc.of 13th International Workshop on Database and Expert Systems Applications*, Aix-en-Provence, France, 2002, pp. 246–250.

[10] D. J. Ladani and N. P. Desai, "Stopword identification and removal techniques on TC and IR applications: A survey," in *Proc. of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2020, pp. 466–472. https://doi.org/10.1109/ICACCS48705.2020.9074166

[11] P. Han, S. Shen, D. Wang, and Y. Liu, "The influence of word normalization in English document clustering," in *Proc. of IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, Zhangjiajie, China, May 2012, pp. 116–120. https://doi.org/10.1109/CSAE.2012.6272740

[12] S. A. Salloum, R. Khan, and K. Shaalan, "A survey of semantic analysis approaches." in *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, *Advances in Intelligent Systems and Computing*, vol. 1153, A.E. Hassanien, A. Azar, T. Gaber, D. Oliva, and F. Tolba, Eds. Springer, Cham., Mar. 2020. https://doi.org/10.1007/978-3-030-44289-7_6

[13] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *Computation and Language,* pp. 1–16, 2018. [Online]. Available: https://aclanthology.org/N19-1423.pdf

[14] P. A. Vikhar, "Evolutionary algorithms: A critical review and its future prospects," in *Proc. of International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, Jalgaon, India, Dec. 2016, pp. 261–265. https://doi.org/10.1109/ICGTSPICC.2016.7955308

[15] Y. Yusoff, M. Salihin, N. Azlan, and M. Zain, "Overview of NSGA-II for optimizing machining process parameters," *Procedia Engineering,* vol. 15, pp. 3978–3983, 2011. https://doi.org/10.1016/j.proeng.2011.08.745

[16] R. A. Monteiro, R. L. S. Santos, T. A. S. Pardo, T. A. de Almeida, E.E.S. Ruiz, and O. A. Vale, "Contributions to the study of fake news in portuguese: new corpus and automatic detection results," in *Proc. of International Conference on Computational Processing of the Portuguese Language*, Canela, Brazil, 2018, pp. 324–334. https://doi.org/10.1007/978-3-319-99722-3_33

[17] C. Lui *et al*., "A two-stage model based on BERT for short fake news detection," in *Proc. of International Conference on Knowledge Science, Engineering and Management*, Athens, Greece, Aug. 2019, pp. 172–183. https://doi.org/10.1007/978-3-030-29563-9_17

[18] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (BERT)," *Applied Sciences.,* vol. 9, no. 19, Sep. 2019, Art. no. 4062. https://doi.org/10.3390/app9194062

[19] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, "A smart system for fake news detection using machine learning," in *Proc. of International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India, Sep. 2019, pp. 1–6. https://doi.org/10.1109/ICICT46931.2019.8977659

[20] T. Zhang *et al*. , "BDANN: BERT-based domain adaptation neural network for multi-modal fake news detection," in *Proc. of International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, Jul. 2020, pp. 1–8. https://doi.org/10.1109/IJCNN48605.2020.9206973

[21] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – A deep convolutional neural network for fake news detection," *Cognitive Systems Research,* vol. 61, pp. 32-44, Jun. 2020. https://doi.org/10.1016/j.cogsys.2019.12.005

[22] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access,* vol. 8, pp. 156695–156706, Aug. 2020. https://doi.org/10.1109/ACCESS.2020.3019735

[23] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: a hybrid CNN-RNN based deep learning approach," *International Journal of Information Management Data Insights,* vol. 1, no. 1, Apr. 2021, Art. no. 100007. https://doi.org/10.1016/j.jjimei.2020.100007

[24] B. Al-Ahmad, M. A. Al-Zoubi, R. A. Kurma, and I. Aljarah, "An evolutionary fake news detection method for COVID-19 pandemic information," *Asymmetry,* vol. 13, no. 6, Jun. 2021, Art. no. 1091. https://doi.org/10.3390/sym13061091

[25] S. Liu, H. Tao, and S. Feng, "Text classification research based on BERT model and Bayesian network," in *Proc. of Chinese Automation Congress (CAC)*, Hangzhou, China, Nov. 2019, pp. 5842–5846. https://doi.org/10.1109/CAC48633.2019.8996183

[26] M. A.Hassonaha, R. Al-Sayyeda, and A. Rodan, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter," *Knowledge-Based Systems,* vol. 192, Mar. 2020, Art. no. 105353. https://doi.org/10.1016/j.knosys.2019.105353

[27] Y. Fors-Isalguez and J. Hemosillo-Valadez, "Query-oriented text summarization based on multiobjective evolutionary algorithms and word embeddings," *Journal of Intelligent & Fuzzy Systems,* vol. 34, no. 5, pp. 3235–3244, May 2018. https://doi.org/10.3233/JIFS-169506

[28] R. Alqaisi, W. Ghanem, and A. Qaroush, "Extractive multi-document arabic text summarization using evolutionary multi-objective optimization with K-medoid clustering," *IEEE Access,* vol. 8, pp. 228206–228224, Dec. 2020. https://doi.org/10.1109/ACCESS.2020.3046494

Arshad Ali received a BSc degree in Mathematics and Physics from Punjab University, Lahore, in 1997. In 2003, he earned an MSc degree in Computer Science from Punjab University, Lahore. He received a Master degree (MPhil) in Information Technology (Informatique) with a qualification of specialist in mobile networks in 2009. He received the PhD degree in Information Technology (telecommunications and electronics) jointly from the Institute of Telecom SudParis and UPMC (Paris VI). Between 2005 and 2008, he served as an Audit Officer at the Office of the Auditor General of Pakistan. He was a Post-doctoral Researcher at Orange Labs, Paris. Currently, he is working as an Assistant Professor at the Computer Science Department of the University of Lahore, Pakistan. His research interests are in the areas of mobile ad-hoc networks, network coding, natural language processing, machine learning, deep learning, and image processing in healthcare.
E-mail: Arshad.ali1@nu.edu.pk
ORCID iD: https://orcid.org/0000-0003-0562-7403

Maryam Gulzar received her BS degree in Computer Science from the CS&IT Department of the University of Lahore (UOL), Pakistan in 2016. She completed her MS (SPM) study at the School of Computing at the National University of Computer and Emerging Sciences (NUCES), Lahore campus, Pakistan in 2018. Currently, she is undertaking PhD as a Junior Researcher at the Department of Software Engineering of LUT University, Finland. Earlier, she was a Lecturer at the Department of Software Engineering of the University of Lahore, Pakistan. Her research interests include digital transmission, information systems, and machine learning. She won a research grant from the Higher Education Commission of Pakistan for four years to complete her PhD studies
E-mail: Maryam.Gulzar@lut.fi
ORCID iD: https://orcid.org/0000-0003-1632-7588