



Fake or real news about COVID-19? Pretrained transformer model to detect potential misleading news

SreeJagadeesh Malla^a  and P. J. A. Alphonse^b

Department of Computer Applications, National Institute of Technology, Thuvakudi, Tiruchirappalli, Tamil Nadu 620015, India

Received 12 October 2021 / Accepted 18 December 2021 / Published online 13 January 2022
© The Author(s), under exclusive licence to EDP Sciences, Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract The World Health Organization declared the novel coronavirus disease 2019 a pandemic on March 11, 2020. Along with the coronavirus pandemic, a new crisis has emerged, characterized by widespread fear and panic caused by a lack of information or, in some cases, outright fake messages. In these circumstances, Twitter is one of the most eminent and trusted social media platforms. Fake tweets, on the other hand, are challenging to detect and differentiate. The primary goal of this paper is to educate society about the importance of accurate information and prevent the spread of fake information. This paper has investigated COVID-19 fake data from various social media platforms such as Twitter, Facebook, and Instagram. The objective of this paper is to categorize given tweets as either fake or real news. The authors have tested various deep learning models on the COVID-19 fake dataset. Finally, the CT-BERT and RoBERTa deep learning models outperformed other deep learning models like BERT, BERTweet, ALBERT, and DistilBERT. The proposed ensemble deep learning architecture outperformed CT-BERT and RoBERTa on the COVID-19 fake news dataset using the multiplicative fusion technique. The proposed model's performance in this technique was determined by the multiplicative product of the final predictive values of CT-BERT and RoBERTa. This technique overcomes the disadvantage of these CT-BERT and RoBERTa models' incorrect predictive nature. The proposed architecture outperforms both well-known ML and DL models, with 98.88% accuracy and a 98.93% F1-score.

1 Introduction

Reports of Wuhan Municipal Health Commission, China, have mentioned the coronavirus evolution on Dec 31st, 2019. It was initially named SARS-CoV-2. Later on Jan 12th, 2020, World Health Organization (WHO) renamed this disease like the 2019 novel coronavirus (2019-nCoV). On Jan 30th, 2020, a health emergency was declared by WHO. Upon subsequent discussions on this disease outbreak, it was renamed coronavirus disease 2019 (COVID-19) on Feb 11th, 2020. This COVID-19 pandemic has tremendously affected worldwide, and it faces an incredible threat to public health, food systems, psychology, and workplace safety.

According to the survey, COVID-19 is caused by the SARS-CoV-2 virus, which spreads from person to person, especially when they are in immediate contact. Furthermore, when people cough, sneeze, speak, sing, or breathe loudly, the virus can spread from an infected person to close contacted people. To deal with these critical pandemic situations, the government has promoted physical distancing by limiting close face-to-

face contact with others. Further to reduce the disease spread, the government has established cantonment zones where positive cases have considerably increased. Hence it is highly essential to alarm the social organizations and government organizations to avoid the spread of disease to other regions that are not affected. Social media has taken an active step in developing contact with and through various sectors of people across the globe. Especially in critical times, Twitter content individuals can interact with each other during the lockdown period, update their knowledge about the disease, and take the necessary steps to get rid of the disease outbreak. During the lockdown era, precautions like physical separation, wearing a mask, keeping rooms adequately aired, avoiding crowds, washing hands, and coughing into a tissue or bent elbow were adopted. This information was updated to the public consistently by Twitter posts.

The COVID-19 pandemic has had a negative impact on the world in a variety of areas, including public health, tourism, business, economics, politics, education, and people's lifestyle. In the last two years, researchers have paid more attention to COVID-19. Some researchers have concentrated on Natural Language Processing [1–3], which includes disease symp-

^a e-mail: malla.sree@gmail.com (corresponding author)

^b e-mail: alphonse@nitt.edu

Table 1 Different COVID-19 disease related tweets

Fake COVID-19 tweet

- Video of Muslims violating lockdown conditions in old city (Hyderabad)
- A photo shows a 19-year-old vaccine for canine coronavirus that could be used to prevent the new coronavirus causing COVID-19

Real COVID-19 tweet

- A common question: why are the cumulative outcome numbers smaller than the current outcome numbers? A: most
- States reported 1121 deaths a small rise from last Tuesday. Southern states reported 640 of those deaths

toms, medical reports of COVID patients, patient health conditions, information about pandemic preventions and precautions, and social media messages/tweets, among other things. Other researchers concentrate on image processing [4–6], which includes patient X-ray analysis to confirm whether the COVID-19 is positive or negative. During the COVID-19 outbreak, respiratory analysis research became popular [7–10]. Deep learning models were used to categorize the respiratory sounds of patients in this study, yielding better results. Mathematical researchers are more focused on COVID-19 statistical reports [11–14], such as the number of cases identified, the number of deaths, and the number of patients recovered, among other things.

Twitter posts contain both fake and real news (source: COVID-19 FakeNews dataset) as shown as Table 1. In a real sense, all real news may not be informative. For example, let us consider an accurate report containing some predictable content along with COVID-19 disease information. Only COVID-19 related content brings much hype to the tweet posted in public, and hence it is considered informative. In our proposed work, our objective is to highlight such informative content from the tweets and predict the severity of disease in a particular location based on geolocation, age, gender, and time. In detail, what sort of gender and where the outbreak of illness tends to be serious is identified within a particular period.

The following are the highlights of this paper.:-

1. The ensemble transformer model with fusion vector multiplication technique was addressed.
2. The CT-BERT and RoBERTa transformers are utilised in a combination.
3. The FBEDL paradigm produces significant outcomes.
4. The dataset is based on the most recent COVID-19 labelled English fake tweets collection.
5. The model has a 98.93% F1-score and a 98.88% accuracy in identifying fake tweets.

Following on from the discussion of related work in Sect. 2, the Sect. 3 delves into methodology and data, following with a discussion of the experimental results in Sect. 4. In Sect. 4, we examine the results and look at the errors, and Sect. 5 bring the paper to a conclusion.

2 Related work

The authors Easwaramoorthy et al. [15] illustrated the transmission rate in both times by comparing and predicting the epidemic curves of the first and second waves of the COVID-19 pandemic. Kavitha et al. [16] have investigated the duration of the second and third waves in India and forecasts the outbreak's future trend using SIR and fractal models. Gowrisankar et al. [17] have explained multifractal formalism on COVID-19 data, with the assumption that country-specific infection rates exhibit power law growth.

Minaee et al. [18] present a detailed quantitative analysis of over 100 DL models proposed after over 16 popular text classification datasets. Kadhim [19] automatically classified a collection of documents into one or more known categories. Discussed weighing methods and comparison of different classification techniques. Aggarwal and Zhai [20] have presented a survey of a broad range of text classification algorithms and have talked about classification in the database, machine learning, data mining and information retrieval communities, as well as target marketing, medical diagnosis, news group filtering, and document organisation. Kowsari et al. [21] have discussed different text feature extractions, dimensionality reduction methods, existing algorithms and techniques, and evaluations methods along with real-world problems. De Beer and Matthee [22] have pointed various language approaches like Topic-Agnostic, machine learning and knowledge based.

Uysal and Gunal [23] have discussed the impact of preprocessing on text classification in terms of classification accuracy, text domain, dimension reduction and text language. Wenet al. [24] employ a clarity map by using two-channel convolutional network and morphological filtering. The fusion image is created by combining the clear parts of the source images. Castillo Ossa et al. [25] have developed a hybrid model that combines the population dynamics of the SIR model of differential equations with recurrent neural network extrapolations. Wiysobunriet al. [26] have presented an ensemble deep learning system based on the Max (Majority) voting scheme with VGG-19, DenseNet201, MobileNet-V2, ResNet34, and ResNet50 for the automatic detection of COVID-19 disease using chest X-ray images.

Table 2, The identification and classification of tweets related to disaster and current disease pandemic COVID-19 have been discussed. It has covered

Table 2 Summary for text classification based papers

S. no.	Year	Author and Paper	Important discussed topic	Model/technique
1	2021	Madichetty and Sridevi [27]	Detecting situational tweets in the aftermath of a disaster.	Feature-based approach and Fine-tuned RoBERTa model.
2	2021	Malla and Alphonse [28]	Detection of useful information tweets about COVID-19	Majority voting technique, RoBERTa, BERTweet, and CT-BERT
3	2020	Jagadeesh and Alphonse [1]	Identify and classify informative COVID-19 tweets	RoBERTa
4	2007	Danesh et al. [29]	An ensemble ML model for text classification.	Naive-Bayes, k-NN classifier and Racchio with fusion method
5	2021	Kranthi Kumar and Alphonse [30]	Impact of respiratory sounds on COVID-19 disease identification	CNN

Table 3 COVID-19 fake tweets detection papers summary

S. no.	Year	Author and Paper	Model	Accuracy	F1-score
1	2020	Gautam et al. [31]	XLNet + LDA	93.90	94.00
2	2021	Shushkevich and Cardiff [32]	Ensemble model	93.90	94.00
3	2020	Glazkova et al. [33]	CT-BERT+ hard voting	98.50	98.69
4	2021	Paka et al. [34]	BERTa + BiLSTM	95.40	95.30
5	2021	Li et al. [35]	BiLSTM	89.00	88.00
6	2017	Singhania et al. [36]	3HAN + features	96.30	96.77
7	2021	Ahmed et al. [37]	LSVM + TF-IDF	92.15	92.08

the summary of text mining and sentiment analysis-based papers based on AI techniques. In this table, We have represented the published year, author with the cited article, the main content in that paper, and model or approach used in that paper. Machine Learning models have explained Naive Bayes and k-NN classifier for text classification with the help of the top-most frequency word features and low-level lexical features. Transformer pre-trained deep learning models CT-BERT, BERTweet, RoBERTa, and other models outperformed traditional machine learning models and neural networks (CNN)

Table 3 has explained the summary for COVID-19 fake news detection-based papers. As shown in the table, the authors have discussed the automatic fake news detection AI models (on the different dataset) with performance metrics as F1-score and accuracy. Transformer model papers have achieved good results than other Artificial Intelligence models.

3 Framework methodology

During the COVID-19 epidemic, the FBEDL model detects fake COVID-19 tweets with an accuracy of 98.88% and an F1-score of 98.93%. Figure 1 depicts a high-level overview of the FBEDL model. The following subsections go over the FBEDL model in greater

depth: The FBEDL model's data collection and its preprocessing are described in Section A and B. Section C and D describes the pre-trained deep learning classifiers and section E had discussed fusion multiplication technique.

3.1 Tweets collection and data preprocessing

In the COVID-19 pandemic (2020), organizers provided the COVID-19 fake news English dataset [38] with the id, tweet, label (“Fake” and “Real”) in the tsv format. Data is collected from the organizers of the Constraint@AAAI2021 workshop [39]. The organisers considered only textual English contents and captured a generic corpus linked to the coronavirus epidemic using a predetermined list of ten keywords including: COVID-19, cases, coronavirus, deaths, tests, new, people, number and total. The attained tweets are preprocessed using the methods described below.

3.2 Preprocessing of data

In Twitter information, there is a lot of noise. As a result, pre-trained models may benefit from data preparation. The following data preprocessing steps were inspired primarily by [40].

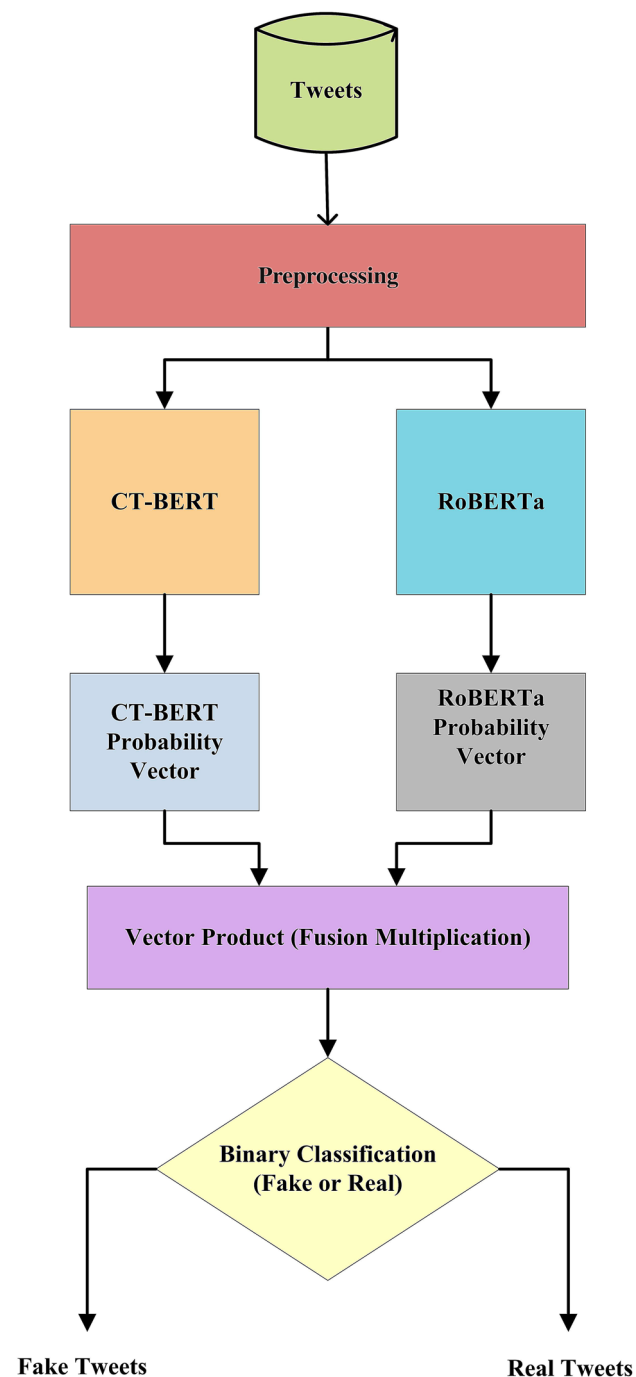


Fig. 1 Overview of the proposed (FBEDL) ensemble deep learning model

1. Remove all English stop words and non alphanumeric characters.
2. Remove tabs, newlines and unnecessary spaces.
3. All links in the tweets (shown as HTTPURL) are replaced with URL.

Because the user handles in the tweets had already been replaced by @USER, no further processing was required.

3.3 RoBERTa

RoBERTa [41] improves on BERT by deleting the next-sentence pretraining target and train with considerably learning rates and huge mini-batches, as well as modifying important hyper parameters. Google announced transformer method, which has improved the NLP (Natural Language Processing) systems using encoder representations. RoBERTa enhanced the efficiency than BERT, which increased the benefit of the masked language modelling objective. Furthermore, when compared to the base BERT model, RoBERTa is explored with higher magnitude data.

RoBERTa is a retraining of BERT with improved training methodology, 1000% more data, and compute power. So it outperforms both BERT and XLNet. But generally, the text is derived from all sources of text (not only tweets).

For the given COVID-19 fake dataset, the model has trained using various hyperparameter combinations (learning rate and batch size). The four metric parameters used to evaluate the results obtained for each combination are accuracy, recall, precision and F1-score. This model has been trained on the COVID-19 English fake dataset with batch sizes of 8, 16, and 32. However, the model performs well when the batch size is 8 and the learning rate is $1.12e-05$ as shown in Table 4. This results may vary from dataset to dataset. Finally, RoBERTa's performance measures are accuracy of 98.55, F1-score of 98.62, recall of 98.84, and precision of 98.40, all of which improves the proposed FBEDL model's performance.

3.4 CT-BERT

CT-BERT (COVID-Twitter-BERT) [40], a recent transformer based model, which has trained on a massive corpus of Twitter tweets on the issue of current on going COVID-19 outbreak. This model shows a better improvement of 05–10% when compared to its basic model, BERT-LARGE. The most substantial improvements have been made to the target domain. CT-BERT as well as other pretrained transformer models are trained on a specific pretraining target domain and can be used for a variety of NLP tasks, such as mining and analysis. CT-BERT was designed with COVID-19 content in mind.

Covid Twitter-BERT includes domain (COVID-19) as well as specific information, and it can better handle noisy texts like tweets. CT-BERT performs similarly well on other classification problems on COVID-19-related data sources, particularly on text derived from social media platforms.

For the given COVID-19 fake dataset, this model has trained using various hyperparameter combinations (batch size and learning rate). As indicated in Table 5, the best results were obtained when the batch size was equal to 8 and the learning rate was equal to $1.02e-06$. The CT-BERT model's results may vary from dataset to dataset. Finally, CT-performance BERT's metrics

Table 4 RoBERTa results have obtained using the test data set

Epocs	bs	lr	Loss	TN	FN	FP	TP	Accuracy	F1-score	Recall	Precision
25	8	1.12e−05	0.0780	1002	18	13	1107	98.55	98.62	98.84	98.40

are accuracy of 98.22, F1-score of 98.32, recall of 99.02, and precision of 97.62, all of which improve the performance of the proposed FBEDL model.

3.5 Fusion vector multiplication

To overcome the disadvantages of CT-BERT and RoBERTa models, an ensemble model is introduced. For concatenation of output for internal models, fusion techniques are more popular. These techniques include max, min, mean, avg, sum, difference, and product probability values.

The probability vector of a tweet is calculated using the fine-tuned RoBERTa model and the CT-BERT model. The multiplicative fusion technique [42] performs element-wise multiplication to combine both (array of the last layer) probability vectors into a single vector [27]. The predicted tweet label is based on the generated vector.

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ a_n & b_n \end{bmatrix} \tag{1}$$

$$B = \begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ c_n & d_n \end{bmatrix} \tag{2}$$

where $a_i + b_i = 1$ and $c_i + d_i = 1$

where a_i and b_i are the probabilities of fake and real news of i th tweet from RoBERTa model respectively and c_i and d_i are the probabilities of fake and real news of i th tweet from CT-BERT model respectively. Consider A is the probability vector (last layer) of the RoBERTa and B is the probability vector (last layer) of the CT-BERT. The fusion vector multiplication of A, B is $FVM(A, B) = AB$ (single vector)

$$FVM(A, B) = \begin{bmatrix} a_1 * c_1 & b_1 * d_1 \\ a_2 * c_2 & b_2 * d_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ a_n * c_n & b_n * d_n \end{bmatrix} \tag{3}$$

$$FBEDL_Test(tweet_i) = \begin{cases} \text{Fake,} & \text{if } a_i * c_i > b_i * d_i \\ \text{Real,} & \text{if } a_i * c_i < b_i * d_i \\ \text{Neutral,} & \text{otherwise} \end{cases} \tag{4}$$

Algorithm 1 Proposed Method based on Fusion vector multiplication Technique

- 1: 0: COVID-19 Fake tweet.
- 2: 1: COVID-19 Real News/tweet
- 3: A: An array of ROBERTA model probability vector (last layer).
- 4: B: An array of CT-BERT model probability vector (last layer).
- 5: S: Size of Test dataset
- 6: **Input:** S, A,B
- 7: **Steps:**
- 8: for $k = 1$ to size of (S)
- 9: Final_prediction_value = FVM(A_k, B_k)
- 10: end
- 11: **Output:** Final_prediction_value (0, 1) (ie: Fake, Real)

where a_i, c_i are the first column i th elements of A and B respectively.

where b_i, d_i are the second column i th elements of A and B respectively.

From Eq. (4), The following possible observations are:

1. if $a_i * c_i > b_i * d_i$ then the proposed model predicts the tweet as “Fake”.
2. if $a_i * c_i < b_i * d_i$ then the proposed model predicts the tweet as “Real”.
3. Our proposed model is trained and tested by Fake news COVID-19 dataset (ie Fake, Real). In this case Neutral case is very rare to occur.

4 Results and analysis

All of our experiments in this paper have been completed using the Google Colaboratory (CoLab) interface and the Chrome browser. This section covers data sets, model parameter explanations, and performance evaluations. Furthermore, the proposed solution is evaluated in comparison to existing methods. The Huggingface package [43] has used in the implementation through Python. The “ktrain” package [44] has been used to fine-tune our baseline models.

4.1 Fake news COVID-19 dataset

In the COVID-19 outbreak (2020), Constraint@AAAI 2021 workshop organizers provided the COVID-19 fake news English dataset [38] with the id, tweet, label (“Fake” and “Real”) in the form of tsv. The above dataset, which contains fake news collected from tweets,

Table 5 CT-BERT results from the test dataset

Epocs	bs	lr	Loss	TN	FN	FP	TP	Accuracy	Recall	Precision	F1-score
25	8	1.02e−06	0.0313	993	27	11	1109	98.22	99.02	97.62	98.32

Table 6 COVID-19 fake english data set details

Fake news (COVID-19) dataSet	Fake	Real
Training data	3060	3360
Validation data	1020	1120
Test data	1020	1120

instagram posts, facebook posts, press releases, or any other popular media content, has a size of 10,700 records. Using the Twitter API, real news was gathered from potential real tweets. Official accounts such as the Indian Council of Medical Research (ICMR), the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), Covid India Seva, and others may have real tweets. They give valuable COVID-19 information such as vaccine progress, dates, hotspots, government policies, and so on.

The dataset is divided into three sections: 60% for train, 20% for validation, and 20% for testing. Table 6 illustrates the distribution of all data splits by class. The dataset with 52.34% of the samples containing legitimate news and 47.66% including fraudulent news.

4.2 Experiment setup

The outcome of the model is dependent on the use of a classifier. As a result, the following classifiers are used to conduct various tests.

1. CT-BERT transformer model.
2. RoBERTa transformer model.
3. Fusion vector multiplication technique.

4.3 Performance measures

The model performance is evaluated using the following parameters: Precision, F1-score, Accuracy, and Recall. These metrics have been depended on the confusion matrix.

4.3.1 Confusion matrix

The performance of a classification model has been evaluated by an $N \times N$ matrix, where N indicates number of target classes. For binary classification N is equals to 2, so a 2×2 matrix containing four values, as shown below.

True Positive (TP): the expected and actual values are identical. The model actual result was positive and anticipated a positive value. True Negative (TN): the expect value comparable to the real value. The model

actual value is negative and the anticipated a negative value. False Positive (FP): The expected value has incorrectly predicted. Although the actual number is negative, the model projected that it would be positive.

False Negative (FN): the expected value is incorrectly predicted. Although the actual number was positive, the model predicted that it would be negative.

4.4 Performance analysis

There are three subsections in this section. The performance of the ML (machine learning) models are compared in the first subsection. In the second subsection, the performance of the deep learning models are compared. The proposed model's performance is compared to existing approaches in the third subsection.

4.4.1 Performance metrics in machine learning models

The Constraint@AAAI2021 workshop organisers have provided baseline results for the English COVID-19 fake dataset. Logistic Regression, Decision Tree, Gradient Boost and SVM have been considered for baseline results for predicting fake news tweets. The Support Vector Machines (SVM) classifier has achieved an accuracy of 93.32%, F1-score of 93.32%, precision of 93.33%, and recall of 93.32%. As a result, the SVM classifier outperformed all metrics values as shown as Table 7.

4.4.2 Deep Learning models performance metrics evaluation

The transformer pretrained deep learning models like DistliBERT, ALBERT, BERT, BERTweet, RoBERTa and CT-BERT have been considered in this subsection. The MAX_LENGTH(tweet) has been fixed to 143 in order to train the model's better with the English language corpus. The tweets that are being tested are in English. For training the models and learning the rate of values $1e-4$, $1e-5$, $1e-6$, $1e-7$, $1e-8$ and tested with batch sizes of 8, 16, and 32.

CT-BERT and RoBERTa have occupied first two places as shown in the Table 8 than the BERTweet, BERT, DistilBERT, and ALBERT models as exhibit in Fig. 2a–d. They outperformed the other competitors in the race, according to the experiment results, because they had higher TP (true positive) and FN (false negative) values. CT-BERT performed well because it has pre-trained on a large corpus of COVID-19-related Twitter messages.

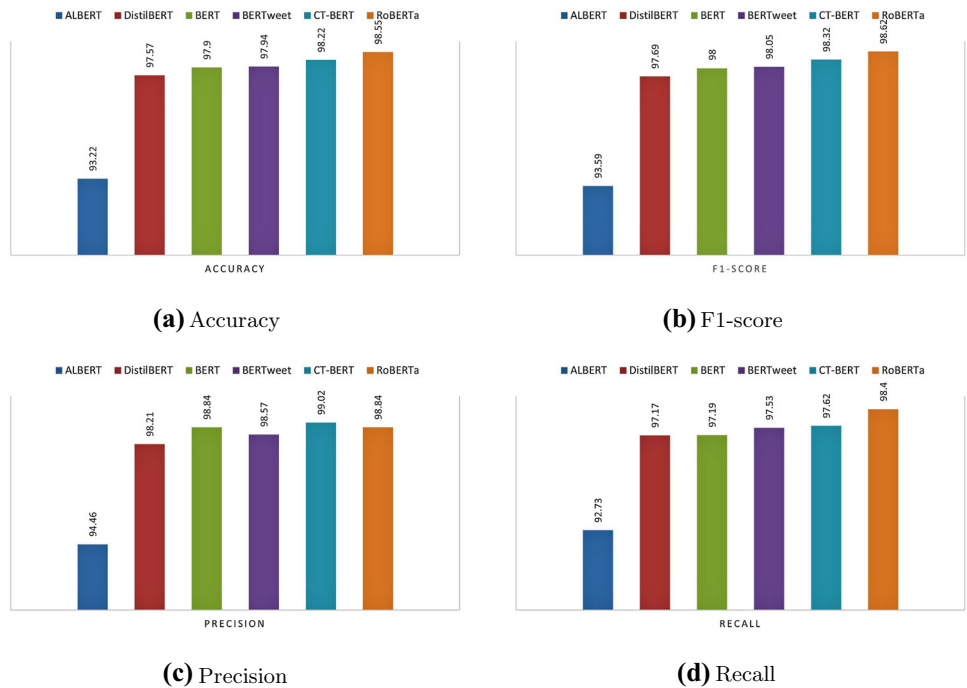
Table 7 Machine learning models: results from the test data set

Model	Accuracy	F1-score	Precision	Recall
DecisionTree	85.37	85.39	85.47	85.37
Logistic Regression	91.96	91.96	92.01	91.96
Support Vector Machine	93.32	93.32	93.33	93.32
Gradient Boost	86.96	86.96	87.24	86.96

Table 8 Deep learning models: results from the test data set

Model	TN	FN	FP	TP	Accuracy	F1-score	Precision	Recall
ALBERT	937	83	62	1058	93.22	93.59	94.46	92.73
DistilBERT	988	32	20	1100	97.57	97.69	98.21	97.17
BERT	988	32	13	1107	97.90	98.00	98.84	97.19
BERTweet-COVID-19	992	28	16	1104	97.94	98.05	98.57	97.53
CT-BERT	993	27	11	1109	98.22	98.32	99.02	97.62
RoBERTa	1002	18	13	1107	98.55	98.62	98.84	98.40

Fig. 2 Deep learning models performance in terms of evaluation metrics



4.4.3 Ensemble deep learning models performance metrics

This segment examined modern ensemble deep learning models. The ensemble model with BiLSTM + SVM + Linear regression + Navaiy Baiyes + combination of LR+NB has obtained F1-score as 94% and accuracy as 93.90%. The combination of XLNet and LDA technique has given F1-score as 96.70% and accuracy as 96.60%. The ensemble model using CT-BERT and hard voting technique has given better performance than other ensemble models.

4.5 Performance comparison: proposed model versus ensemble deep learning techniques

The proposed model (FBEDL) is evaluated in terms of accuracy and F1-score to the machine learning models, deep learning models, and ensemble models. In comparison to existing models, our FBEDL model attained an F1 score of 98.93% and an accuracy of 98.88%, as shown in Tables 9 and 10 as well as Fig. 3. This indicates that the model was successful in distinguishing fake tweets/News about the COVID-19 disease outbreak.

Table 9 Performance comparison: proposed model versus existing models

Model	F1-score	Accuracy
Decision Tree [38]	85.39	85.37
Gradient Boost [38]	86.96	86.96
Logistic Regression [38]	91.96	91.96
Support Vector Machine [38] (Baseline)	93.32	93.32
XLNet + LDA [31]	96.70	96.60
Ensemble [32]	94.00	93.90
CT-BERT + hard voting [33]	98.69	98.50
Proposed model (FBEDL)	98.93	98.88

Table 10 FBEDL model results from the test dataset

F1-score	Accuracy	Recall	precision
98.93	98.88	98.75	99.11

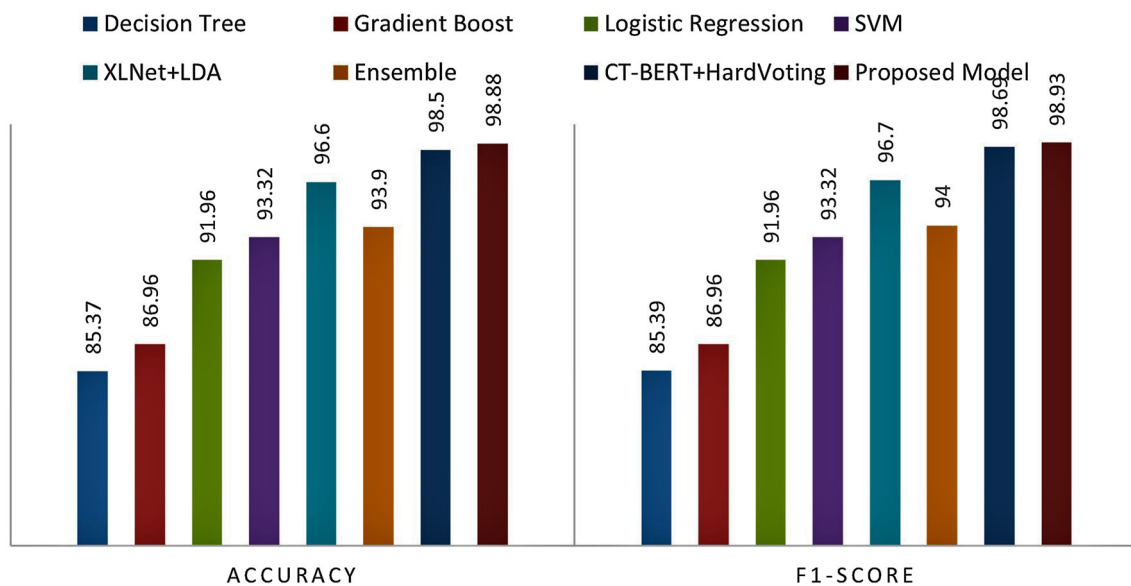
5 Conclusion

The principal goal of this work is to demonstrate how to use a novel NLP application to detect real or fake COVID-19 tweets. The conclusions of the paper assist individuals in avoiding hysteria about COVID-19 tweets. Our findings may also aid in the improvement of COVID-19 therapies and public health measures.

In this study, a fusion technique-based ensemble deep learning model is used to detect fraudulent tweets in the ongoing COVID-19 epidemic. The use of fusion vector multiplication is designed to help our model become more entrenched. We tried various deep learning model combinations to improve model performance, but COVID-Twitter BERT and RoBERTa deep learning models have achieved state-of-art performance. With 98.88% accuracy and a 98.93% F1-score, the proposed model outperforms traditional machine learning and deep learning models.

One of the disadvantages of our proposed model is that RoBERTa and CT-BERT are pre-trained models with a lot of memory for corpus training (657MB and 1.47GB, respectively). When compared to machine learning models, the models' time complexity is likewise relatively high. To boost model performance, we plan to apply data compression techniques

This research focuses on COVID-19 pandemic English fake tweets for the time being. Our method may be able to predict fake tweets about diseases that are similar in the future. We can improve our results in the future by training other combinations on a sizeable COVID-19 dataset using alternative transformer-based models.

**Fig. 3** Performance: proposed model versus state-of-art models

Declarations

Conflict of interest The authors declare no competing interests.

References

- M.S. Jagadeesh, P.J.A. Alphonse, NIT_COVID-19 at WNUT-2020 task 2: deep learning model RoBERTa for identify informative COVID-19 English tweets. In: Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), pp. 450–454. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.wnut-1.66>
- Y. Prakash Babu, R. Eswari, CIA_NITT at WNUT-2020 task 2: classification of COVID-19 tweets using pre-trained language models. In: Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), pp. 471–474. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.wnut-1.70>
- M. Jamshidi, A. Lalbakhsh, J. Talla, Z. Peroutka, F. Hadjilooei, P. Lalbakhsh, M. Jamshidi, L. La Spada, M. Mirzozafari, M. Dehghani et al., Artificial intelligence and Covid-19: deep learning approaches for diagnosis and treatment. *IEEE Access* **8**, 109581–109595 (2020)
- S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-Covid: predicting Covid-19 from chest X-ray images using deep transfer learning. *Med. Image Anal.* **65**, 101794 (2020)
- R. Maroldi, P. Rondi, G.M. Agazzi, M. Ravanelli, A. Borghesi, D. Farina, Which role for chest X-ray score in predicting the outcome in Covid-19 pneumonia? *Eur. Radiol.* **31**(6), 4016–4022 (2021)
- M.A. Al-antari, C.-H. Hua, J. Bang, S. Lee, Fast deep learning computer-aided diagnosis of Covid-19 based on digital chest X-ray images. *Appl. Intell.* **51**(5), 2890–2907 (2021)
- K.K. Lella, P. Alphonse, A literature review on Covid-19 disease diagnosis from respiratory sound data. *AIMS Bioeng.* **8**(2), 140–153 (2021)
- K.K. Lella, A. Pja, Automatic Covid-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice. *AIMS Public Health* **8**(2), 240 (2021)
- L. Gattinoni, D. Chiumello, P. Caironi, M. Busana, F. Romitti, L. Brazzi, L. Camporota, *COVID-19 Pneumonia: Different Respiratory Treatments for Different Phenotypes?* (Springer, Berlin, 2020)
- J.J. Marini, L. Gattinoni, Management of Covid-19 respiratory distress. *JAMA* **323**(22), 2329–2330 (2020)
- M. Mandal, S. Jana, S.K. Nandi, A. Khatua, S. Adak, T. Kar, A model based study on the dynamics of Covid-19: prediction and control. *Chaos Solitons Fract.* **136**, 109889 (2020)
- A. Kumar, K.R. Nayar, S.F. Koya, Covid-19: challenges and its consequences for rural health care in India. *Public Health Pract.* **1**, 100009 (2020)
- J.D. Norrie, Remdesivir for Covid-19: challenges of underpowered studies. *Lancet* **395**(10236), 1525–1527 (2020)
- A. Ghosh, S. Nundy, T.K. Mallick, How India is dealing with Covid-19 pandemic. *Sens. Int.* **1**, 100021 (2020)
- D. Easwaramoorthy, A. Gowrisankar, A. Manimaran, S. Nandhini, L. Rondoni, S. Banerjee, An exploration of fractal-based prognostic model and comparative analysis for second wave of COVID-19 diffusion. *Nonlinear Dyn.* **106**(2), 1375–1395 (2021)
- C. Kavitha, A. Gowrisankar, S. Banerjee, The second and third waves in India: when will the pandemic be culminated? *Eur. Phys. J. Plus* **136**(5), 1–12 (2021)
- A. Gowrisankar, L. Rondoni, S. Banerjee, Can India develop herd immunity against Covid-19? *Eur. Phys. J. Plus* **135**(6), 1–9 (2020)
- S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, Deep learning-based text classification: a comprehensive review. *ACM Comput. Surv. (CSUR)* **54**(3), 1–40 (2021)
- A.I. Kadhim, Survey on supervised machine learning techniques for automatic text classification. *Artif. Intell. Rev.* **52**(1), 273–292 (2019)
- C.C. Aggarwal, C. Zhai, A survey of text classification algorithms. in *Mining Text Data*, ed. by C. Aggarwal, C. Zhai (Springer, Boston, 2012)
- K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, Text classification algorithms: a survey. *Information* **10**(4), 150 (2019)
- D. De Beer, M. Matthee, Approaches to identify fake news: a systematic literature review. In: *International Conference on Integrated Science*, pp. 13–22. Springer (2020)
- A.K. Uysal, S. Gunal, The impact of preprocessing on text classification. *Inf. Process. Manag.* **50**(1), 104–112 (2014)
- Y. Wen, X. Yang, T. Celik, O. Sushkova, M.K. Albertini, Multifocus image fusion using convolutional neural network. *Multimed. Tools Appl.* **79**(45), 34531–34543 (2020)
- L.F. Castillo Ossa, P. Chamoso, J. Arango-López, F. Pinto-Santos, G.A. Isaza, C. Santa-Cruz-González, A. Ceballos-Marquez, G. Hernández, J.M. Corchado, A hybrid model for Covid-19 monitoring and prediction. *Electronics* **10**(7), 799 (2021)
- B.N. Wiysobunri, H.S. Erden, B.U. Toreyin, An ensemble deep learning system for the automatic detection of COVID-19 in X-ray images (2020)
- S. Madichetty, M. Sridevi, A neural-based approach for detecting the situational information from Twitter during disaster. *IEEE Trans. Comput. Soc. Syst.* (2021)
- S. Malla, P. Alphonse, COVID-19 outbreak: an ensemble pre-trained deep learning model for detecting informative tweets. *Appl. Soft Comput.* **107**, 107495 (2021). <https://doi.org/10.1016/j.asoc.2021.107495>
- A. Danesh, B. Moshiri, O. Fatemi, Improve text classification accuracy based on classifier fusion methods. In: *2007 10th International Conference on Information Fusion*, pp. 1–6. IEEE (2007)
- L. Kranthi Kumar, P.J.A. Alphonse, Automatic diagnosis of Covid-19 disease using deep convolutional neural network with multi-feature channel from respiratory

- sound data: cough, voice, and breath. *Alex. Eng. J.* (2021). <https://doi.org/10.1016/j.aej.2021.06.024>
31. A. Gautam, S. Masud, et al., Fake news detection system using xlnet model with topic distributions: constraint @ aaai2021 shared task. arXiv preprint [arXiv:2101.11425](https://arxiv.org/abs/2101.11425) (2021)
 32. E. Shushkevich, J. Cardiff, Tudublin team at constraint@aaai2021—covid19 fake news detection. arXiv preprint [arXiv:2101.05701](https://arxiv.org/abs/2101.05701) (2021)
 33. A. Glazkova, M. Glazkov, T. Trifonov, g2tmn at constraint@aaai2021: exploiting ct-bert and ensembling learning for Covid-19 fake news detection. arXiv preprint [arXiv:2012.11967](https://arxiv.org/abs/2012.11967) (2020)
 34. W.S. Paka, R. Bansal, A. Kaushik, S. Sengupta, T. Chakraborty, Cross-sean: a cross-stitch semi-supervised neural attention model for Covid-19 fake news detection. *Appl. Soft Comput.* **107**, 107393 (2021)
 35. X. Li, P. Lu, L. Hu, X. Wang, L. Lu, A novel self-learning semi-supervised deep learning network to detect fake news on social media. *Multimed. Tools Appl.* (2021). <https://doi.org/10.1007/s11042-021-11065-x>
 36. S. Singhania, N. Fernandez, S. Rao, 3han: a deep neural network for fake news detection. In: *International Conference on Neural Information Processing*, pp. 572–581. Springer (2017)
 37. H. Ahmed, T. Traore, S. Saad, 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*, pp. 127–138. Springer (2017)
 38. P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M.S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: COVID-19 fake news dataset. in *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation* (Springer, 2021), pp. 21–29
 39. P. Patwa, M. Bhardwaj, V. Guptha, G. Kumari, S. Sharma, S. Pykl, A. Das, A. Ekbal, M.S. Akhtar, T. Chakraborty, Overview of constraint 2021 shared tasks: detecting English Covid-19 fake news and Hindi hostile posts. in *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation* (Springer, 2021), pp. 42–53
 40. M. Müller, M. Salathé, P.E. Kummervold, Covid-twitter-bert: a natural language processing model to analyse Covid-19 content on twitter. arXiv preprint [arXiv:2005.07503](https://arxiv.org/abs/2005.07503) (2020)
 41. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: a robustly optimized bert pretraining approach. arXiv preprint [arXiv:1907.11692](https://arxiv.org/abs/1907.11692) (2019)
 42. A. Wu Y. Han, Multi-modal circulant fusion for video-to-language and backward. In: *IJCAI*, vol. 3, p. 8 (2018)
 43. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
 44. A.S. Maiya, ktrain: a low-code library for augmented machine learning. arXiv preprint [arXiv:2004.10703](https://arxiv.org/abs/2004.10703) [cs.LG] (2020)