A Novel Paradigm for Sentiment Analysis on COVID-19 Tweets with Transfer Learning Based Fine-Tuned BERT

Amit Pimpalkar[1,2,*], Jeberson Retna Raj[1]

[1]School of Computing, Sathyabama Institute of Science and Technology, Chennai, India

[2]Department of Computer Science and Engineering (AIML), Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Received 13 March 2023; received in revised form 04 July 2023; accepted 06 July 2023

DOI: https://doi.org/10.46604/aiti.2023.11743

Abstract

The rapid escalation in global COVID-19 cases has engendered profound emotions of fear, agitation, and despondency within society. It is evident from COVID-19-related tweets that spark panic and elevate stress among individuals. Analyzing the sentiment expressed in online comments aids various stakeholders in monitoring the situation. This research aims to improve the performance of pre-trained bidirectional encoder representations from transformers (BERT) by employing transfer learning (TL) and fine hyper-parameter tuning (FT). The model is applied to three distinct COVID-19-related datasets, and each of the datasets belongs to a different class. The evaluation of the model's performance involves six different machine learning (ML) classification models. This model is trained and evaluated using metrics such as accuracy, precision, recall, and F1-score. Heat maps are generated for each model to visualize the results. The performance of the model demonstrates accuracies of 83%, 97%, and 98% for Class-5, Class-3, and binary classifications, respectively.

Keywords: COVID-19, pre-trained, sentiment analysis, BERT, transfer learning

1. Introduction

The COVID-19 pandemic has become synonymous with the year 2020, and India stands among the countries most severely affected by the outbreak. In 2019, scientists identified the severe acute respiratory syndrome-associated coronavirus (SARS-CoV) as the cause of the pandemic, and it was later renamed COVID-19. The pandemic has led to numerous deaths and has caused both health and economic crises across the world. The resurgence of COVID-19, particularly with the Omicron variant, has caused widespread fear, agitation, and despondency [1].

Due to constant lockdowns and social isolation, people have increasingly turned to online activities, and social media platforms have become the most effective medium for communication. Twitter is a social networking site where people share news and relevant information. Meanwhile, it's also one of the most widely used microblogging platforms that provides invaluable data sets for public opinion surveys. As a result, the COVID-19 pandemic has catalyzed the increasing use of Twitter for discussions on various topics [2]. However, spreading false and misleading information on social media has become a significant concern. Therefore, it is crucial to verify the accuracy of information shared on social media and prevent the spread of false news by assessing people's sentiments.

To resolve the problem, researchers can use the information collected from analyzed tweets by Twitter for academic and research purposes. Consequently, nations must implement measures to protect themselves by revealing the truth and data about the pandemic. According to Yella [3], the later stages of COVID-19 now have a lower fatality rate but a more significant

* Corresponding author. E-mail address: amit.pimpalkar@gmail.com

English language proofreader: Chih-Wei Chang

contamination and spread rate than previously transmitted SARS, Omicron variant, and MERS COVID-19. Around 2.9 million new cases were recorded worldwide in the first week of January 2023, an increase of 10% week-on-week compared to the prior weeks. The same week witnessed almost 11,000 recorded fatalities, representing a 22% rise from the preceding week. 659 million confirmed illnesses and 6.6 million fatalities have been recorded globally as of 8 January 2023 [4]. When millions of people throughout the globe succumb to COVID-19, a new pandemic—the black fungus—emerges to threaten everyone's life—especially those who are still recuperating from COVID-19.

This study presents a framework for sourcing data to train the pre-trained model for sentiment analysis (SA). Against this background, research exploration work attempts to address the accompanying research questions (RQ) related to COVID-19.

RQ-1. What are the well-known catchphrases that appear in Indian tweets in English?

RQ-2. How did these tweets influence general well-being frameworks?

RQ-3. How do ML calculations serve to investigate individuals' sentiments?

RQ-4. How far can the deep learning (DL) 12-layer bidirectional encoder representations from transformers (BERT) model outperform other conventional ML models?

RQ-5. How is the performance comparison of the BERT model for COVID-19 SA on the bi-class and multi-class datasets (Class-3 and Class-5)?

The majority contributions of this research are as follows:

(1) This work adopts an integrated approach to understanding SA for COVID-19.
(2) A transfer learning (TL)-based pre-trained BERT model is employed to address SA challenges in English.
(3) The pre-trained BERT model is adapted to perform sentiment classification on three different datasets containing bi-class or multi-class data.
(4) Various encoding techniques, including one-hot and TF-IDF, along with TL variants, are explored to address scenarios.
(5) All fine-tuned network parameters are incorporated to efficiently capture gaps of varying lengths and weights in forming feature maps.
(6) A range of Standard ML methods, such as multinomial naive Bayes (MNB), random forest (RF), logistic regression (LR), extreme gradient boosting (XGB), stochastic gradient descent (SGD), and support vector classifier (SVC), are employed alongside the pre-trained BERT for COVID-19.
(7) This research conducts a comparative analysis of seven models to showcase the transformer design's superiority over prior state-of-the-art methods. With appropriate fine-tuning, the transformer design emerges as the new state-of-the-art model with superior performance.

The inspiration for this work, a brief demonstration of SA, RQ, and significant contributions of the research are provided in Section 1. The rest of the paper is organized as follows. Section 2 discusses related works done by the researchers, followed by the limitations. Section 3 gives a foundation for the work chosen, explaining the essentials of the fundamentals of BERT and TL on a deeper level. The study also highlights the proposed system's methodology, architecture, dataset, feature selection, hyper-parameter tuning, and evaluation process. The trial results depicted in Section 4 compare the outcomes of the innovative classifiers with those obtained using the proposed strategy for SA. At last, Section 5 concludes the paper and suggests directions for future research.

2. Related Work

Recently, related researchers reported on SA of COVID-19 Twitter data; they have developed many text-mining techniques to investigate the various elements of COVID-19, utilizing textual data from online social media sites. In this context, TL is becoming extremely important in the research field. Developments in DL have led to significant improvements in modeling, machine translation, and other natural language processing (NLP) tasks, including text categorization, language translation, and others. This is especially true for neural network designs like recurrent neural networks (RNN), long short-term memory (LSTM), and convolutional neural networks (CNN).

For real-time visualization and granular event classification in public opinion research, Zhuang et al. [5] introduced the latent Dirichlet allocation-autoregressive moving average model deep neural network (LDA-ARMA DNN). LDA was used to determine what the remarks were about. To perform multi-faceted SA and variation prediction, they used the ARMA on massive amounts of textual data connected to COVID-19. Malla and Alphonse [6] proposed a method majority voting-based ensemble deep learning (MVEDL) model for locating critical tweets during the COVID-19 outbreak using the COVID-19-Twitter BERT, BERTweet, and RoBERTa DL-based models.

The study attempted to establish the sentiment of people in eight nations [7]. They built a unification model for SA with five different deep-learning classifiers and then consolidated them to refine the result via a meta-learning method. Researchers have focused on mucormycosis related to COVID-19, and a time series analysis of tweets has shown that negative tweets are becoming less common with time [8]. Karthikeyan et al. [9] propose the use of a modified NN architecture for an AI system, incorporating a hybrid learning-based network classifier that relies on both CNN and support vector machine (SVM) [9]. Yan et al. [10] proposed an attention parallel dual-channel deep learning hybrid model (ADDHM) with BERT and a dual-channel architecture built of CNN and bidirectional long short-term memory (BiLSTM).

Furthermore, in each channel, an attentional mechanism was incorporated to separate the words that influence the most emotional inclination and strengthen the quality of the model for sentiment classification. Researchers Suganya and Kalpana [11] proposed a COVID-19 detection classification utilizing a pre-trained model Mask regional-convolutional neural network (R-CNN) to improve performance on both binary and multi-class classification. The authors utilized a shot learning strategy using a ResNet-50 baseline classifier and softmax as an activation function.

Table 1 Limitations of the existing research

Referenced research	Limitations of the research
[7]	The data's country of origin did not influence selected keywords for gathering and filtering tweets. At the outset of March 2020, collected news items and obtained data were found from a particular source.
[12]	The time lasted only 23 days, from 3 December 2021 to 26 December 2021.
[13]	System performance was minimal on the smaller datasets.
[14]	The model was constrained using sequential DL models with limited performance.
[15-16]	The feature selection and hyper-parameter tuning operations were not carried out.
[17]	The authors did not explore various perspectives, including geographical ones.
[18]	Insufficient tweets from a particular nation were evaluated.
[19]	The research focused on sentiments primarily related to medical services only.
[20]	Since most Weibo users are young individuals, the outcomes of the analysis may be skewed. The study only highlights the emergence of negative attitudes among young individuals following the pandemic.
[21]	The cross-validation mechanism was not utilized.
[22]	The sentiments presented in these tweets are solely based on the word "fear", and they were expressed by citizens of the United States. The tweets in question are relatively short in length.

There are many unnoticed problems and probable pitfalls in deciphering the work that has been circulating lately. Previous efforts focused on information analysis of tweets related to COVID-19 had a small corpus of tweets. SA using Twitter has several practical implications. The first issue is the data quality, as Twitter data is often noisy, with abbreviations, slang, and misspellings, which can affect the accuracy. The second issue is the limitation of Twitter's API, which only provides access to a small portion of tweets, limiting the sample's representativeness. Another issue is the potential bias in the training data, as the training data used to develop SA models may not represent the population or contain biases that affect the model's accuracy. Finally, there is the issue of ethical implications, which means using SA on social media raises privacy concerns involving analyzing personal data without consent. There is also the potential for misuse of SA, such as using SA to identify and target vulnerable individuals for political or commercial purposes. Table 1 below shows the limitations of existing research.

Considering the shortcomings of current research, the proposed solution advocates for a pre-trained model, BERT, which is based on the transformer architecture and includes both TL and optimization of hyper-parameters in the study. The study extracted the datasets from freely available repositories and explored them as in section 3.1. Furthermore, traditional ML and DL approaches have been used in many COVID-19-related tweets with bi-class and multi-class data. Nonetheless, this proposed technique beats various issues and elasticities with near-perfect accuracy. This research offers valuable insights into the potential of transformer-based models to enhance performance and advance the field of NLP.

3. Methodology

As shown in Fig. 1, the starting point of the suggested architecture is the cleaning and transformation of the three datasets utilized in this study. Both unlabeled and labeled datasets undergo a data refinement process, including steps like HTML parsing, POS tagging, stop word removal, tokenization, word sense disambiguation, and non-alphanumeric character removal. Then, the labeled corpus was combined with the unlabeled corpus for training, yielding a more robust labeled corpus. The research conducted two separate analyses, the first sorting tweets about COVID-19 in the labeled corpus into positive and negative categories and the second one using an improved labeled corpus. Training and testing sets were created from the tagged corpora; TLFT-BERT-based DL architecture was trained and verified using these datasets before being applied to SA and prediction from COVID-19 tweets and then calculated the suggested model's accuracy, precision, recall, and F1 score.

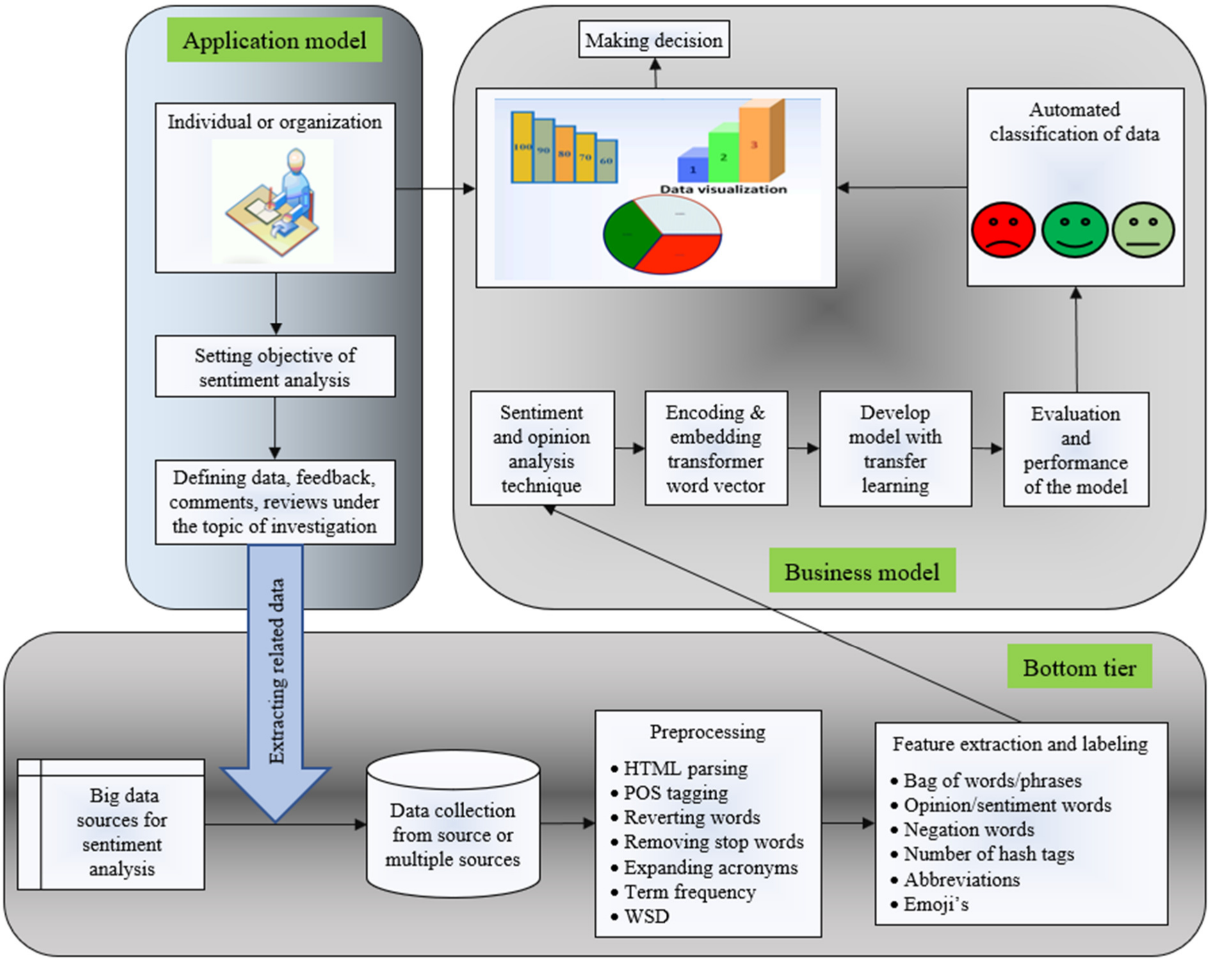


Fig. 1 Representative mechanism of the proposed model

3.1. Dataset

The experimental investigation employed three real-world datasets from the Kaggle repository to evaluate the classifier's classification performance. Tables 2, 3, and 4 summarize the datasets used in this investigation. An 80:20 train: test split was employed on the dataset to minimize overfitting and accurately assess the model. The testing technique provides data to fine-tune the model's hyper-parameters and verify model performance on 20% of the data after training it on 80% of the dataset. The Class-5 dataset has 41,157 distinct emotional tweets from Twitter users gathered around one month (March 2020 to April 2020) when there were uttermost cases of COVID-19. The dataset was categorized into five classes for the task, as shown in Table 2.

Table 2 The Class-5 dataset statistics

Class-5	Actual dataset	After processing	20% of processed data	80% of processed data
Extremely negative (0)	5,481	5,481	1,096	4,385
Negative (1)	9,917	9,890	1,978	7,912
Neutral (2)	7,713	7,565	1,513	6,052
Positive (3)	11,422	11,380	2,276	9,104
Extremely positive (4)	6,624	6,620	1,324	5,296
Total samples	41,157	40,936	8,187	32,749

Table 3 The Class-3 dataset statistics

Class-3	Actual dataset	After processing	20% of processed data	80% of processed data
Negative (0)	16,335	16,200	3,240	12,960
Neutral (1)	67,835	65,035	13,007	52,028
Positive (2)	6,280	6,175	1,235	4,940
Total samples	90,450	87,410	17,482	69,928

Table 4 The Class-2 dataset statistics

Class-2	Actual dataset	After processing	20% of processed data	80% of processed data
Negative (0)	40,322	8,986	1,798	7,188
Positive (1)	139,537	48,176	9,635	38,541
Total samples	179,859	57,162	11,433	45,729

The Class-3 dataset has 90,450 distinct emotional tweets from 70,000 Twitter users gathered within two months (February 2020 to March 2020) when there was a peak of COVID-19 [23]. For the classification job, three different classes, 6,175 positive sentiments, 16,200 negative sentiments, and 65,035 neutral sentiments, after pre-processing the dataset. The Class-2 dataset has 179,859 distinct emotional tweets from Twitter users gathered within three months (February 2020 to April 2020).

3.2. Data pre-processing

The information obtained through social media networks has frequently been raw, noisy, varied, and in a primitive state that makes study impossible; so, it is cleaned first before the research. There are several steps involved in pre-processing. Removing URLs from the text, all usernames, stop words, converting all letters to lowercase, and eliminating incorrect characteristics are all data pre-processing strategies. Several libraries were utilized in Python for text pre-processing tasks, including the natural language toolkit (NLTK), spacy, a regular expression library, and other general libraries.

3.3. BERT

Google introduced a novel language representation method known as BERT [24]. BERT enables the pre-training of deep bidirectional representations from the unlabelled text in each layer, concurrently modifying the context to the left and right of a given text. BERT includes tokens such as unknown [UNK], separation [SEP], classification [CLS], and padding [PAD], as illustrated in Fig. 2 [24]. The BERT model was trained using the masked language model (MLM) and next sentence prediction (NSP) techniques. The deep bidirectional model of the MLM is prepared by randomly selecting and masking input tokens and then making predictions for the masked tokens. The NSP method helps to set up the connection between sentences in the model. For each pair of phrases P1 and P2, the NSP examine 50% of the pre-trained sample to determine their correlation. Table 5 displays the characteristics of the BERT base and the BERT large.

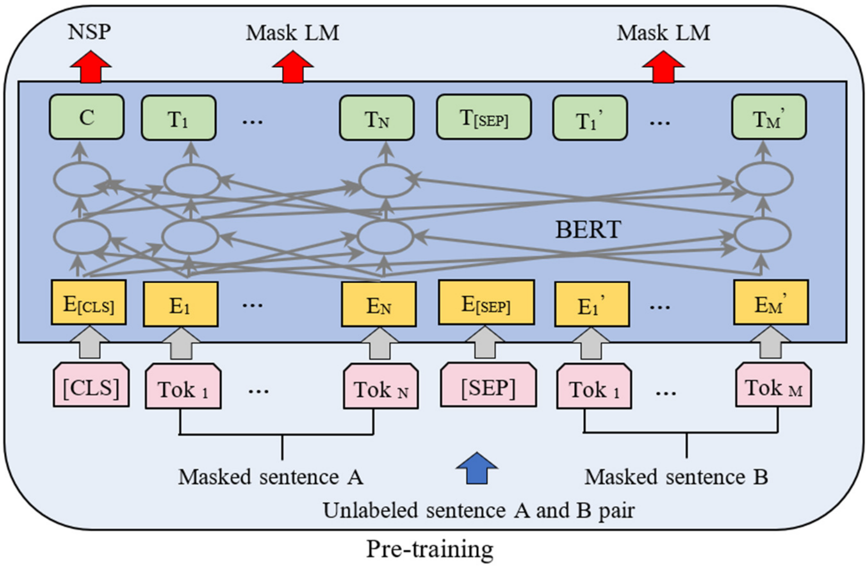


Fig. 2 BERT Representation [24]

Table 5 BERT model detail view

Model/Dimensions	Number of layers (L)	Hidden layers size (H)	Self-attention heads (A)	Total parameters
Bert base	12	768	12	110M
Bert large	24	1024	16	340M

3.4. Transfer learning in NLP

Today, the most powerful language models rely significantly on transformers, and they are often regarded as the best for all fundamental NLP and Natural Language Understanding (NLU) tasks. Since training neural networks from scratch with enormous datasets is computationally expensive and takes a long time, and access to high-end computing resources, such as clusters of CPUs and GPUs, is required. To avoid these challenges, the transfer function approach is presented as an easy technique. The idea behind the TL is that rather than starting from a blank and using random weights, it is possible to solve future problems with the help of the consequences gained from one training. The pre-trained models can be utilized directly or to extract characteristics other than those learned from the model during prior training.

In the context of TL, typically, two stages are involved. Firstly, a model is pre-trained on one dataset to serve as a feature extractor. Subsequently, the pre-trained model is used to apply the acquired knowledge by fine-tuning it on another dataset. In this specific case, the BERT model in its original form has been pre-trained on the combination of two massive collections, BookCorpus and English Wikipedia, and it is used for solving a task for classifying COVID-19 reviews. This approach leads to a reduction in training time and an improvement in neural network performance.

3.5. Experimental hyper-parameters

Upon completing the training set, researchers reconfigured crucial hyper-parameters to obtain a refined model with comparable assessment metrics. ML algorithms may effectively train the data but fail to summarize inconspicuous data, which means overfitting. Meanwhile, regularization, optimizer, learning rate, and batch standardization are generally employed to avoid overfitting.

The researchers trained a model with input layers on top of the embedding layer via a pre-trained BERT base uncased. They conducted experiments to evaluate the model's performance using different optimizers and learning rates while keeping other parameters, such as Conv1D kernel size, number of filters, and batch size constant. The evaluation results were documented, and it was observed that the "Adam" optimizer with a learning rate of 1e-5 and decay learning rate of 1e-7 had the most influence on the model's performance. Subsequently, the researchers selected the aforementioned hyper-parameters and fine-tuned the number of dense units and epoch size while keeping other parameters constant. The experimental results were thoroughly reviewed and implemented, leading to the establishment of model correlation hyper-boundaries. The model correlation hyper-boundaries are set as follows, as shown in Table 6.

Table 6 Hyper-parameter used in the research

Hyper-parameter	Values
Max review length	128
Learning rate	1e-5
Decay	1e-7
Batch size	32
Number of epochs	10
Number of Conv1D filters	64
Conv1D kernel size	3
Number of dense units	32
Threshold	0.9
Optimizer	Adam
Activation function	Softmax
Batch normalizations	Yes
Loss function	Binary cross entropy, categorical cross entropy
Output function	Sigmoid

The experiments were conducted using Python 3.8, with TensorFlow 2.10 and Keras 2.9 libraries on Windows 10. To train state-of-the-art DL models, a Kaggle Kernel that provides Nvidia P100 GPU, 73.1 GB of Disk, 15.9 GB of GPU memory, 13 GB of RAM, 1.32 GHz Memory Clock, with a reasonably high-end performance of 9.3 TFLOPS.

4. Result and Discussion

To strengthen the experimental hypothesis, traditional ML classifier techniques and experiments from past examinations explore different Twitter datasets for SA as baseline classifiers including MNB, RF, LR, XGB, SGD, and SVC, to establish bi-class and multi-class classification. Regarding identifying the optimal model evaluation technique, the study investigated two feature vectors - TF-IDF and one-hot encoding. Hyper-parameter tuning and TL techniques were employed to pre-trained BERT encoding, utilizing the same feature vectors. The researchers obtained remarkable outcomes that exceeded those of conventional ML classifiers. They then further illuminated their findings by comparing them with the existing state-of-the-art BERT model on the COVID-19 dataset. The following section provides detailed statistics and a summary of the discoveries made by the research methodology.

After evaluating the training model and measuring its performance using several metrics, as discussed earlier, the experiment tested the model on both bi-class and multi-class (Class-3 and Class-5) datasets. For assessment reasons, a confusion matrix and heat maps also facilitated data exploration and visualization of several characteristics. Because COVID-19 datasets have an unequal class distribution, accuracy is the most logical performance statistic to employ here. The display of word clouds can yield a rapid overview of the dominant lexicon within a given text or may identify the top themes based on the relative frequency of individual words. This presentation method often renders the most frequent terms in larger font sizes while relegating less commonly used terms to a more minor, secondary position. Fig. 3 shows the most prominent keyword used for Class-5 and Class-3 datasets.

(a) Class-5 dataset (b) Class-3 dataset

Fig. 3 Word cloud for Class-5 and Class-3 datasets

Table 7 reveals a comparative analysis of accuracy which scores on many benchmarking ML algorithms and the proposed method of the BERT model with TL. Meanwhile, Fig. 4 presents a fascinating contrast of models being assessed for various accuracy classifications.

Table 7 Comparison of Classes-2, 3, and 5 test results between baseline and proposed methods

Model/Evaluation	Precision			Recall			F1-score		
	Class-2	Class-3	Class-5	Class-2	Class-3	Class-5	Class-2	Class-3	Class-5
MNB	0.91	0.60	0.49	0.89	0.69	0.51	0.90	0.63	0.50
RF	0.96	0.75	0.54	0.98	0.80	0.60	0.97	0.77	0.55
LR	0.96	0.72	0.62	0.97	0.81	0.63	0.96	0.76	0.63
XGB	0.92	0.63	0.58	0.96	0.80	0.60	0.94	0.68	0.59
SGD	0.95	0.70	0.60	0.97	0.79	0.57	0.96	0.74	0.58
SVC	0.95	0.67	0.59	0.98	0.83	0.64	0.97	0.72	0.60
Proposed BERT	**0.98**	**0.97**	**0.83**	**0.98**	**0.97**	**0.84**	**0.98**	**0.97**	**0.83**

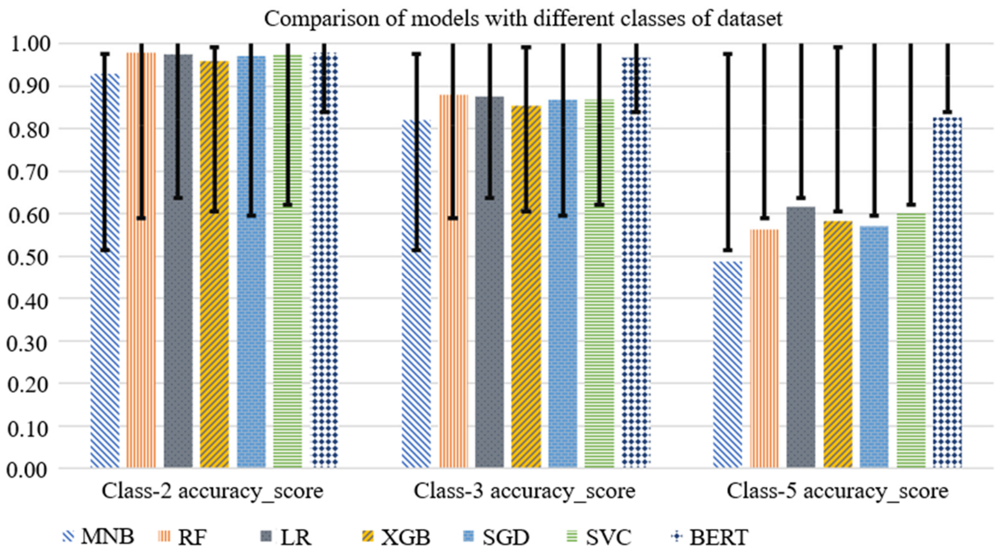
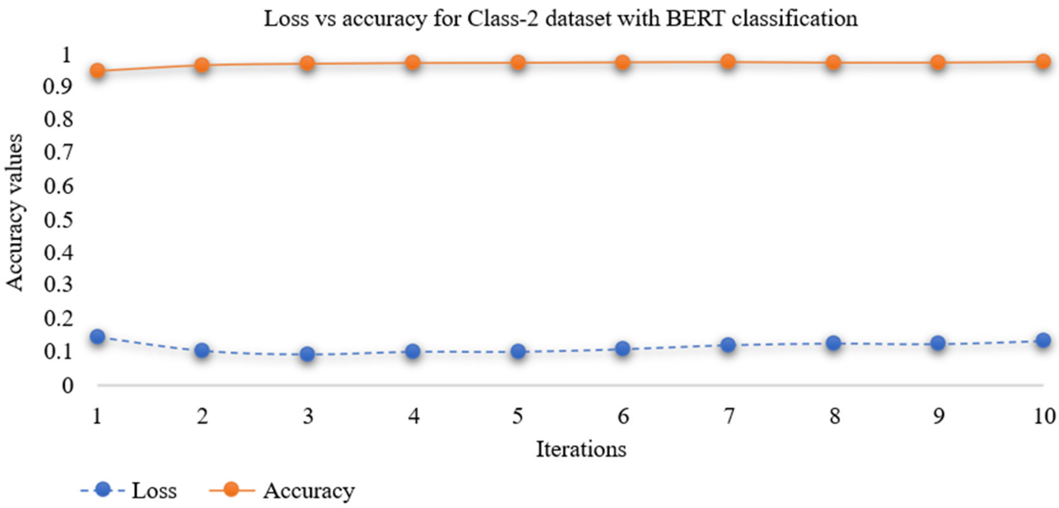


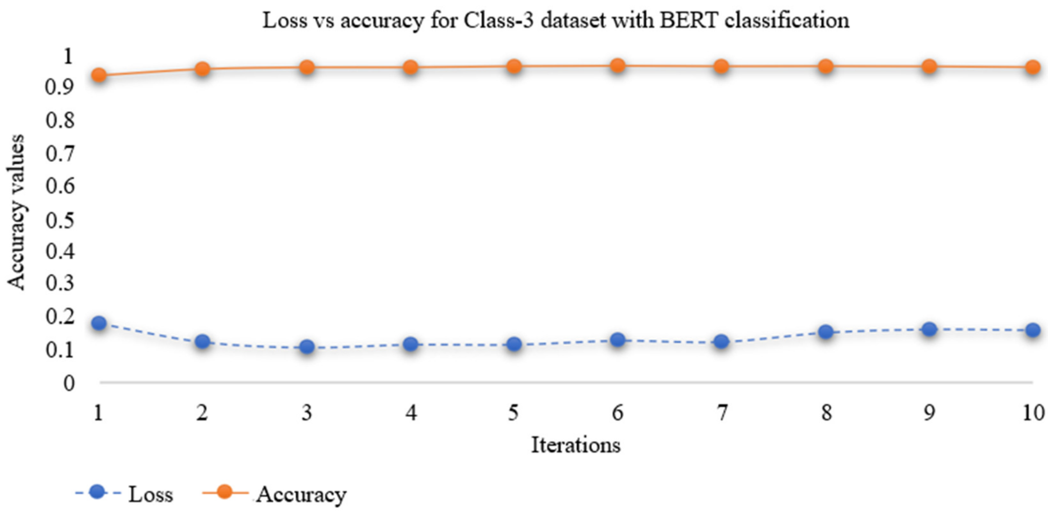
Fig. 4 The comparison of models under evaluations with different classes on accuracy

Fig. 5 depicts the suggested approach's training and testing loss ratio, as determined by computing the 80% of trained samples and the randomly selected 20% testing inputs. According to the experimental results, the suggested model's accuracy, F1-score, precision, and recall were high. Even for the Class-5 classification, the average improvement is around 30% using the proposed method. The suggested model's inference time to assess a single sample was determined to be as low as 0.516 s/step, which is just around half a second.

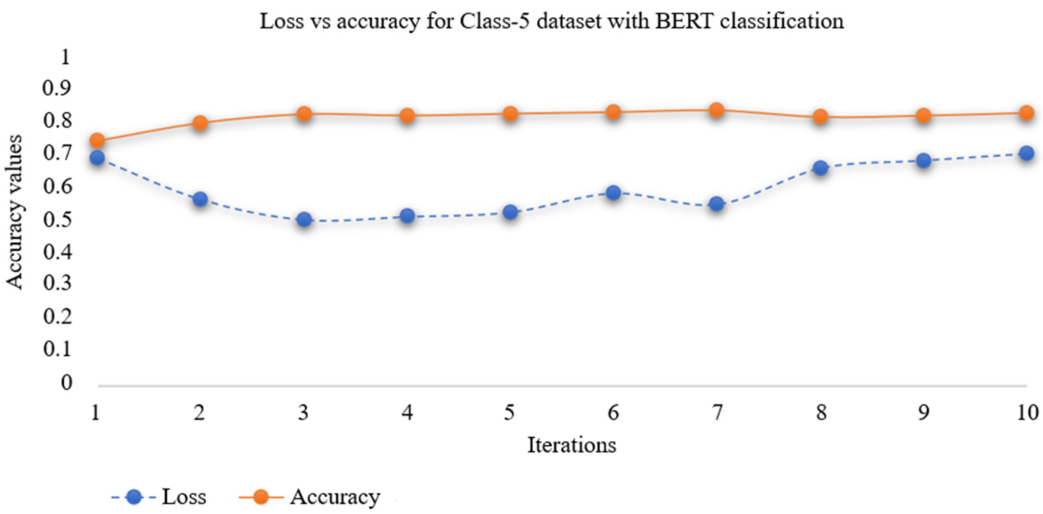
In the illuminating Fig. 6, the predicted emotions were scrutinized using heat maps, which disclosed the instances of a particular emotion vis-à-vis other opinions in the identical sentiment group. These heat maps, manifesting the expressions of two sentiments, furnish insights about the attitudes associated with positive, negative, and neutral related classes. The experimental results compared to previous work in text sentiment classification algorithms utilizing COVID-19 datasets to verify the model's effectiveness. The experimental results were validated using classical ML and the TLFT-BERT model. Table 8 compares the empirical model's accuracy to previous models, and the findings demonstrate the superiority of the proposed technique. This modeling approach can be utilized to improve COVID-19 management in many situations.



(a) Loss and accuracy comparison for Class-2 classifiers

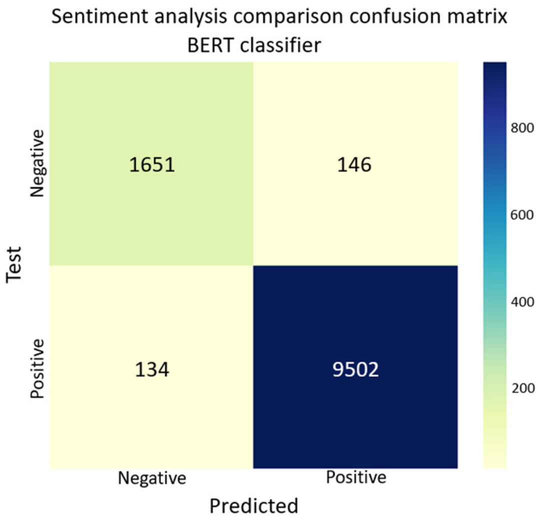


(b) Loss and accuracy comparison for Class-3 classifiers

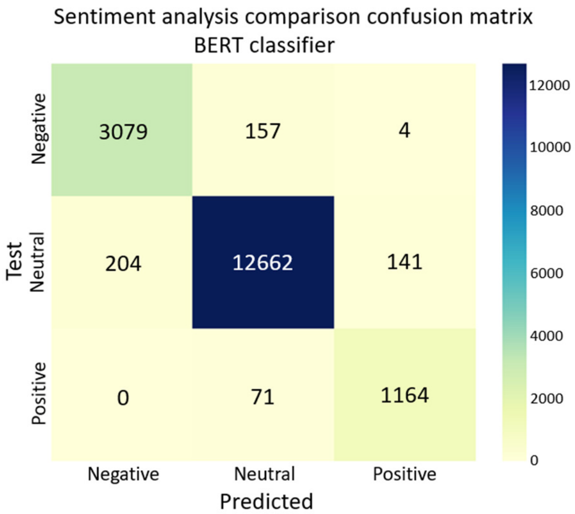


(c) Loss and accuracy comparison for Class-5 classifiers

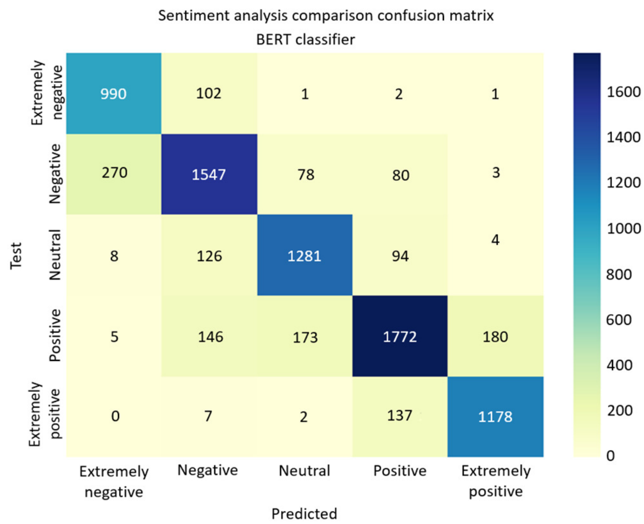
Fig. 5 Loss vs. accuracy for specific classes of datasets with BERT classification



(a) Confusion matrix for Class-2 classifiers



(b) Confusion matrix for Class-3 classifier



(c) Confusion matrix for Class-5 classifier

Fig. 6 Heat map of the proposed models under evaluation

Table 8 Comparison of transfer learning-based approaches with state-of-the-art approaches in the literature

Author	Model	Dataset language	Best results
Malla and Alphonse [6]	Majority voting-based ensemble	English	91.75
Basiri et al. [7]	CNN, BiGRU	English	85.80
Naseem et al. [23]	BERT, ALBERT, DistilBERT	English	94.80
Garcia and Berton [25]	Multilingual universal sentence encoder	English & Portuguese	84.00
Chintalapudi et al. [26]	LSTM, BERT	English	89.00
Satu et al. [27]	LSTM, TClustVID	English	97.80
Kabakus [28]	CNN, LSTM	Turkish	97.89
Jalil et al. [29]	DistilBERT	English	96.66
Pimpalkar and Raj [30]	MBiLSTM	English	93.55
Proposed model	**Transfer learning-based BERT**	**English**	**98.00**

The following solutions can be drawn to respond to the RQ relating to COVID-19 indicated in Section 1.

RQ-1: The well-known top twelve catchphrases in Indian tweets were China, coronavirus, COVID, death toll, food, outbreak, pandemic, people, price, sanitizer, store, and supermarket in alphabetically sorted order in English.

RQ. 2: COVID-19 is comparable to other stress in that it causes cognitive, psychological, and emotional stress and it is sufficient to affect people's well-being. COVID-19 has influenced people's lifestyles and contributed to stress, fear of infection, and concern for working individuals' family lives.

RQ. 3: This research concludes that personal feelings or opinions can aid in matching and comprehending the sentiments of others by communicating feelings and providing feedback to others. SA may extract individuals' views from the language used in social media postings, conversations, reviews, and more. ML algorithms can help to predict whether sentiment is positive or negative.

RQ-4 & RQ-5: Table 7 demonstrates that the 12-layer BERT model outperforms other conventional ML models, and even for Class-5 classification. The average improvement is over 30% utilizing the proposed method, and the proposed model's inference time for a single sample was determined to be as low as 0.516s/step, or around half a second.

5. Conclusion and Future Work

The manuscript presents six ML and a DL technique to categorize COVID-19-related tweets. The pre-trained BERT base model is enhanced with TL as TLFT-BERT for class-2 and multi-class classification, specifically designed to tackle SA challenges in English, and uses three COVID-19 Twitter datasets to develop a robust SA model.

(1) The model achieves high accuracy, with 98% for binary classification and 97% for Class-3 classification. It is worth mentioning that the suggested model requires less training time.

(2) By incorporating all fine-tuned network parameters, the TLFT-BERT model captures varying lengths and weights of features for effective feature map summation and understanding of complex linguistic patterns.

(3) The research demonstrates the pivotal role of transformer models in SA when appropriately fine-tuned. The results highlight the superiority of the transformer-based approach, outperforming traditional ML methods.

(4) These findings have broader implications for social scientists and governments seeking insights into global sentiments surrounding COVID-19 tweets. The findings open new avenues for research and applications in SA and other NLP tasks, driving advancements in the field.

(5) The research has a limitation since all the utilized datasets were scribbled in English. It would be intriguing to compare and differentiate native Indian languages.

A more extensive dataset may need training in the Large BERT architecture before being utilized. Furthermore, researchers may use Twitter streaming API to obtain real-time tweets to do SA and research various social networks. This work does not have the highlights to go to multilingual tweets, which could be considered a likely future work toward this path. It is still a research subject on how to embed a long document for enhancing classification performance, particularly when utilizing pre-trained contextualized word embedding like ELMo, XLNet, RoBERTa, and others.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] S. Hosgurmath, V. Petli, and V. K. Jalihal, "An Omicron Variant Tweeter Sentiment Analysis Using NLP Technique," Global Transitions Proceedings, vol. 3, no. 1, pp. 215-219, June 2022.

[2] C. Guo, S. Lin, Z. Huang, and Y. Yao, "Analysis of Sentiment Changes in Online Messages of Depression Patients before and during the COVID-19 Epidemic Based on BERT+BiLSTM," Health Information Science and Systems, vol. 10, no. 1, article no. 15, December 2022.

[3] H. M. Yella, "How Do SARS and MERS Compare with COVID-19?" https://www.medicalnewstoday.com/articles/how-do-sars-and-mers-compare-with-covid-19, September 20, 2022.

[4] World Health Organization, “Weekly Epidemiological Update on COVID-19 - 11 January 2023,” https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---11-january-2023, January 19, 2023.
[5] M. Zhuang, Y. Li, X. Tan, L. Xing, and X. Lu, “Analysis of Public Opinion Evolution of COVID-19 Based on LDA-ARMA Hybrid Model,” Complex & Intelligent Systems, vol. 7, no. 6, pp. 3165-3178, December 2021.
[6] S. Malla and P. J. A. Alphonse, “COVID-19 Outbreak: An Ensemble Pre-Trained Deep Learning Model for Detecting Informative Tweets,” Applied Soft Computing, vol. 107, article no. 107495, August 2021.
[7] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharrya, “A Novel Fusion-Based Deep Learning Model for Sentiment Analysis of COVID-19 Tweets,” Knowledge-Based Systems, vol. 228, article no. 107242, September 2021.
[8] M. Singh, H. K. Dhillon, P. Ichhpujani, S. Iyengar, and R. Kaur, “Twitter Sentiment Analysis for COVID-19 Associated Mucormycosis,” Indian Journal of Ophthalmology, vol. 70, no. 5, pp. 1773-1779, May 2022.
[9] S. Karthikeyan, G. Ramkumar, S. Aravindkumar, M. Tamilselvi, S. Ramesh and A. Ranjith, “A Novel Deep Learning-Based Black Fungus Disease Identification Using Modified Hybrid Learning Methodology,” Contrast Media & Molecular Imaging, vol. 2022, article no. 4352730, 2022.
[10] C. Yan, J. Liu, W. Liu, and X. Liu, “Research on Public Opinion Sentiment Classification Based on Attention Parallel Dual-Channel Deep Learning Hybrid Model,” Engineering Applications of Artificial Intelligence, vol. 116, article no. 105448, November 2022.
[11] D. Suganya and R. Kalpana, “Prognosticating Various Acute Covid Lung Disorders from COVID-19 Patient Using Chest CT Images,” Engineering Applications of Artificial Intelligence, vol. 119, article no. 105820, March 2023.
[12] M. Mahyoob, J. Algaraady, M. Alrahiali, and A. Alblwi, “Sentiment Analysis of Public Tweets Towards the Emergence of SARS-CoV-2 Omicron Variant: A Social Media Analytics Framework,” Engineering, Technology & Applied Science Research, vol. 12, no. 3, June 2022.
[13] B. N. Ramya, S. M. Shetty, A. M. Amaresh, and R. Rakshitha, “Smart Simon Bot with Public Sentiment Analysis for Novel COVID-19 Tweets Stratification,” SN Computer Science, vol. 2, no. 3, article no. 227, May 2021.
[14] G. M. Demirci, S. R. Keskin, and G. Dogan, “Sentiment Analysis in Turkish with Deep Learning,” International Conference on Big Data, pp. 2215-2221, December 2019.
[15] S. S. Aljameel, D. A. Alabbad, N. A. Alzahrani, S. M. Alqarni, F. A. Alamoudi, L. M. Babili, et al., “A Sentiment Analysis Approach to Predict an Individual’s Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia,” International Journal of Environmental Research and Public Health, vol. 18, no. 1, article no. 218, January 2021.
[16] A. Onan, “Sentiment Analysis on Product Reviews Based on Weighted Word Embeddings and Deep Neural Networks,” Concurrency and Computation: Practice and Experience, vol. 33, no. 23, article no. e5909, December 2020.
[17] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, “COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification,” Information, vol. 11, no. 6, article no. 314, June 2020.
[18] G. Barkur, Vibha, and G. B. Kamath, “Sentiment Analysis of Nationwide Lockdown Due to COVID 19 Outbreak: Evidence from India,” Asian Journal of Psychiatry, vol. 51, article no. 102089, June 2020.
[19] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah, “Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study,” Journal of Medical Internet Research, vol. 22, no. 4, article no. e19016, April 2020.
[20] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, “The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users,” International Journal of Environmental Research and Public Health, vol. 17, no. 6, article no. 2032, March 2020.
[21] M. A. Tocoglu, O. Ozturkmenoglu, and A. Alpkocak, “Emotion Analysis from Turkish Tweets Using Deep Neural Networks,” IEEE Access, vol. 7, pp. 183061-183069, 2019.
[22] S. Darad and S. Krishnan, “Sentimental Analysis of COVID-19 Twitter Data Using Deep Learning and Machine Learning Models,” Ingenius, Revista de Ciencia y Tecnología, no. 29, pp. 108-117, January-June 2023. (In Español)
[23] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, “COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis,” IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 1003-1015, August 2021.
[24] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” https://doi.org/10.48550/arXiv.1810.04805, May 24, 2019.
[25] K. Garcia and L. Berton, “Topic Detection and Sentiment Analysis in Twitter Content Related to COVID-19 from Brazil and the USA,” Applied Soft Computing, vol. 101, article no. 107057, March 2021.

[26] N. Chintalapudi, G. Battineni, and F. Amenta, "Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models," Infectious Disease Reports, vol. 13, no. 2, pp. 329-339, June 2021.

[27] M. S. Satu, M. I. Khan, M. Mahmud, S. Uddin, M. A. Summers, J. M. W. Quinn, et al., "TClustVID: A Novel Machine Learning Classification Model to Investigate Topics and Sentiment in COVID-19 Tweets," Knowledge-Based Systems, vol. 226, article no. 107126, August 2021.

[28] A. T. Kabakus, "A Novel COVID-19 Sentiment Analysis in Turkish Based on the Combination of Convolutional Neural Network and Bidirectional Long-Short Term Memory on Twitter," Concurrency and Computation: Practice and Experience, vol. 34, no. 22, article no. e6883, October 2022.

[29] Z. Jalil, A. Abbasi, A. R. Javed, M. B. Khan, M. H. A. Hasanat, K. M. Malik, et al., "COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques," Frontiers in Public Health, vol. 9, article no. 812735, January 2022.

[30] A. Pimpalkar and J. R. Raj, "MBiLSTMGloVe: Embedding GloVe Knowledge into the Corpus Using Multi-Layer BiLSTM Deep Learning Model for Social Media Sentiment Analysis," Expert Systems with Applications, vol. 203, article no. 117581, October 2022.

Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (https://creativecommons.org/licenses/by-nc/4.0/).