Contents lists available at ScienceDirect

Applied Soft Computing

COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets

SreeJagadeesh Malla *, Alphonse P.J.A.

Department of Computer Applications, National Institute of Technology, Tiruchirappalli, 620015, India

ABSTRACT

On 11 March 2020, the (WHO) World Health Organization declared COVID-19 (CoronaVirus Disease 2019) as a pandemic. A further crisis has manifested mass fear and panic, driven by lack of information, or sometimes outright misinformation, alongside the coronavirus pandemic. Twitter is one of the prominent and trusted social media in this current outbreak. Over time, boundless COVID-19 headlines and vast awareness have been spreading, with tweets, updates, videos, and explosive posts. Few studies have been performed on the pandemic to detect and interrelate various disease types, including current coronavirus. However, it is pretty tricky to discriminate and detect a specific category. This work is motivated by the need to inform society about limiting irrelevant information and avoiding spreading negative emotions. In this context, the current work focuses on informative tweet detection in the pandemic to provide relevant information to the government, medical organizations, victims services, etc. This paper used a Majority Voting technique-based Ensemble Deep Learning (MVEDL) model. This MVEDL model is used to identify COVID-19 related (INFORMATIVE) tweets. The state-of-art deep learning models RoBERTa, BERTweet, and CT-BERT are used for best performance with the MVEDL model. The "COVID-19 English labeled tweets" dataset is used for training and testing the MVEDL model. The MVEDL model has shown 91.75 percent accuracy, 91.14 percent F1-score and outperforms the traditional machine learning and deep learning models. We also investigate how to use the MVEDL model for sentiment analysis on 226668 unlabeled COVID-19 tweets and their informative tweets. The application section discussed a comprehensive analysis of both actual and informative tweets. According to our knowledge, this is the first work on COVID-19 sentiment analysis using a deep learning ensemble model.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

SARS-CoV-2 is a new viral disease that first appeared in 2019. Later it was named COVID-19. In late December 2019, COVID-19 was detected first in Wuhan, China, and quickly spread worldwide. The World Health Organization (WHO), which is relentlessly trying to control the spread of the COVID-19 outbreak, declared the pandemic on 30 January 2020. COVID-19 is an infectious disease transmitted by contacts and small droplets when people cough, sneeze, or talk—by Quarantine, Limiting activities, and separating suspects from others who are not ill to be unable to spread the infection or contamination. Most countries have been locked down for strict quarantine implementation. Precautions such as clean, safe distance, wear the mask, do not touch the eyes, nose, or mouth, etc., are proposed. Due to the COVID-19 lockdown and other precautions during the COVID-19 pandemic, the global economic crisis began. The automotive industry, Tourism, Restaurants, Retail, Transportation, and Energy, etc., are the top-rated service sectors are affected by the COVID-19 recession. The vaccination process has been started in many countries to prevent people from becoming seriously ill with COVID-19.

In this pandemic situation, social media sites such as Instagram, Facebook, WhatsApp, Twitter, etc., help gather insightful messages allied to COVID-19 disease. During this pandemic condition, the situations correlate with specific social media messages. The content includes epidemic signs, communities affected by disease outbreaks and other medical services. Today, most NLP researchers focus on social media text classification. This paper has discussed the messages of the most popular social media site Twitter.

Twitter is one of the famous social media quotes most widely used for sharing short messages. These short messages are called tweets with a length of up to 280 characters. The Twitter API supports open access to Twitter in an advanced and exclusive way. Active Twitter user tweets multiple types of information

in a large amount of data at a tremendous pace in a health emergency, consisting of tweets related to both disease and non-disease. Informative tweets provide information about suspected, confirmed, recovered, and death cases and the location or travel history of the patients and contain symptoms of illness like cold, fever, headache, running nose, body pains, etc. The COVID-19 related tweets [1] are not following the "INFORMATIVE" annotation guidelines, are labeled with the "UNINFORMATIVE". Another type of tweets is "Fake" messages. These fake news are purposefully created messages to mislead the social media society, and uninformative messages are not fake, but not related to the COVID-19 pandemic as shown in Table 1. Precautions and prevention for various diseases are to be noticed by various social and government organizations or departments'. For this purpose they needs resources like social media for creating awareness and providing medical kits, medicines. The tweet data is classified with unique resources to help the suffering and suspected users to know their status in a health emergency . Tweets relate to various diseases should be classified for health events to enable the authorities to provide healthcare facilities to prevent the public from developing the disease, leading to the final phase of breathing. To diagnose these kinds of circumstances, we need a common automated framework to gather the above situational tweeting. The NLP researchers analyze the Twitter text by using traditional methods as well as advanced Artificial Intelligence techniques. During this decade, the use of AI techniques increases compared to conventional techniques with better performance and speed. Many of the deep learning strategies are capable of generating better results. Deep learning models' performance depends on the defined problem. Therefore, it has steadily increased the need for robust techniques to detect informative tweets in the disease-relevant corpus.

Disease prediction is one of the essential sub-parts of a health emergency. It is challenging to classify disease and non-disease-related tweets from a source in a health emergency. The following Table 1 provides sample informative and uninformative tweets on the corona pandemic. As stated, tweets relating to a low-grade disease are mentioned in most previous studies. This article explores the issue of the detection of tweets associated with health emergencies (COVID-19). For this purpose, we have proposed an ensemble pre-trained deep learning model with a majority voting technique (MVEDL) for solving the above problem. It is mainly based on the three latest state-of-art deep learning transformers, such as RoBERTa [2,3], CT-BERT [4], and BERTweet [5].

NLP and its use for social media analysis exponential growth have been experienced. The text classification machine learning models face some challenges due to gradient vanishing or exploding, and they are unable to learn long-term dependencies. Sometimes binary or numerical characteristics derived from word frequency are noisy. Moreover, a text classification problem consists of a large number of closely related classes. Human labelers expressively bias the training data, which may yield a wrong training of the model.

Deep learning models overcome the above issues, but they have their problems with text classification. Although most DL models have supervised models that require large amounts of domain labels and have achieved promising performance on challenging benchmarks, most of these models are not interpretable. A state-of-art deep learning model outperforms another model on one dataset but underperforms on other datasets.

ML models are fast and less accurate, where Deep learning models are slow and more accurate. Our MVEDL model is a combination of three state-of-art pre-trained deep learning models (ensemble model), in which the probability of predicting is more than the individual model.

This work is motivated by notifying society of the need to restrict the widespread use of social media worldwide, as the

Table 1

Types of COVID-19 Twitter tweets

COVID-19 Informative Tweet (Real Information and also related to COVID-19 pandemic)
● Over 200,000 in the US now infected. Over 5000 in the US now dead. Coronavirus infections due to increase drastically while the death toll is projected to reach 100k to 240k in the next weeks. My question: why hasna∈™t
COVID-19 Uninformative Tweet (Real Information ,but not related to COVID-19 pandemic)
● Stop and Shop Donates Meal For COVID-19 Healthcare Workers * Stop and Shop announced it will be providing 5000 meals daily for frontline healthcare workers battling novel .
COVID-19 Fake Tweet
● Obama Calls Trump's Coronavirus Response A Chaotic Disaster https://t.co/DeDqZEhAsB.
COVID-19 Real Tweet
● Schools are struggling to cope with a lack of #COVID19 tests - with new infections increasing since it became compulsory for pupils to return. But when should you get your child tested for the virus? Here's our explainer.

paper contributes to the dissemination of irrelevant information as a pandemic has spread to human society. The novelty of the work proposes a model for a technique based on deep learning set to assess the tweets' informative content. This paper has shown as a contribution to society how shocking it is that people share informative and useless data at first, but some share these worthless tweets more. This paper underlines with significant evidence the necessity of using "mechanisms for monitoring" to prevent the dissemination of negative psychology into people's minds in social media. The paper's work also covers the title, as it leads to a decrease in tweets' popularity and does not constitute a trustworthy source. The state-of-art deep learning pre-trained transformers RoBERTa, BERTweet, and CT-BERT, have been applied as an ensemble model on the datasets, and The model with the highest accuracy has been tested. The model used gives the COVID-19 English tweet dataset accuracy of over 91.75%.

The Highlights of this paper are given below:-

(1) This article presents an ensemble deep learning model for classifying Informative Tweet.
(2) The combination of pre-trained state-of-art transformers RoBERTa, BERTweet, and CT-BERT, are used in this paper.
(3) The MVEDL model obtains significant results from machine learning, deep learning, and the latest ensemble deep learning models.
(4) The latest COVID-19 labeled English dataset is trained for the model.
(5) The model has shown 91.75 percent accuracy and 91.14 percent F1-score in detecting the tweets linked to informative English tweets in the ongoing corona pandemic.
(6) The MVEDL model can be helpful in real-time applications like live tweets sentiment analysis/classification, depression status of the patient, COVID-19 outbreak statistics, etc.
(7) In this paper, the latest TextBlob algorithm is used for sentiment analysis.

The layout of the paper is structured like this. Section 1 outlined the summary of the work to be done based on the problem identified. Section 2 has highlighted the pitfalls involved in the existing works and concluded the discussion by giving a better solution to help discriminate the informative tweets very effectively. Section 3 is about a description of methodology. Section 4 is about the Experimental Results and Analysis followed on

Table 2
Summary for Artificial Intelligence based papers.

Paper	Year	Important discussed topic	Advantage/Limitations
[6]	2021	Identify damage assessment tweets during a disaster	The model used linear regression, SVR and random forest technology
[7]	2020	Identify multi-modal informative disaster tweets	Model based on BERT and DenseNet
[8]	2020	Dense classification with contextual representation	ELMo embedded classifier is applied.
[9]	2020	Detecting Informative Tweets	ensemble model of CNN, ANN, fine-tuned VGG-16 architecture
[10]	2020	Reduces dynamic routing computational complexity	Capsule networks have several advantages over CNN
[11]	2019	Model Combination of capsules encoded features and capsule networks	benefit of simplistic capsule networks compared to existing HMC methods.
[12]	2018	Sentiment analysis decision support systems	DSocial model was used for automating the processing of social network information
[13]	2017	Voting Algorithm for sentiment classification	Model Used SVM,Naive bayes, Bagging.
[14]	2009	Ensemble method for sentiment classification	Rule-based classification, supervised learning and machine learning

COVID-19 disease-related tweet classification. Section 5 is about Real-Time Application of the proposed model. Section 6 concludes the work done so far in meeting the objectives.

2. Related work

There are recently published survey papers on deep learning sentiment analysis and classification [15–17]. In their papers, [Minaee et al. 2020][15] Various text classification tasks were provided, including sentiment analysis, news categorization, question answers, and inference of natural languages. Also, they Provided a full review of over 120 profound learning models developed for the text classification in recent years, a summary of over 40 standard datasets used widely for deep learning text classification like RNN-based models, feed-forward networks, Siamese Neural Networks, CNN-based models, attention mechanism, Memory-augmented networks, Graph neural networks, Hybrid models, Transformers etc. Sreenivasulu and Sridevi (2018) [16] focuses on and categorizes various processes for event detection in various types of social media. Besides, various social media features and datasets will also be discussed. [17] explains that millions of related or unrelated tweets/messages to the disasters like earthquakes and floods are posted on social media during the disaster using the SVM classifier by using various statistical combinations. From the analysis, the earthquake keyword and frequency of hashtag features provide better results than the other combinations. Ravi and Ravi (2015) [18] are conducted a survey based on subcategories to be performed, machine learning, processing techniques for languages in nature, and the application of sentiment analysis, covering published literature during 2002–2015. The paper also contains open questions and a summary table with a hundred and sixty-one articles. Authors like Ozbayoglu et al. (2020) [19] offer state-of-the-art DL models for financial applications, such as LSTM, CNN (not very suitable for financial applications). Lella and Alphonse, 2021b [20] are aimed at diagnosing COVID-19 through Machine Learning and Deep Learning techniques, and the authors describe the first step to collect COVID-19 diagnostic reviews from patient respiratory sound data.

Madichetty has focused informative tweets identification of natural disasters, and Sridevi (2020c) [21] are examined the model using SVM (meta classifier) and KNN (base classifier) with the combination of CNN outperforms the other algorithms The Deep Learning models (CNN, LSTM, BLSTM, and BLSTM attention) are used to identify situational information during a disaster in Hindi language tweets, besides English language tweets. Deep learning model results outperform existing classic disaster-set approaches, such as Hagupit cyclone, Hyderabad bomb blast, Sandhy shooting, Harda rail accident, and Earthquake.

Lella and Alphonse, 2021a [22] focus on the diagnosis of COVID-19 disease using a CNN model with respiratory sound parameters. The model improves efficiency to classify COVID-19 sounds for detecting COVID-19 positive symptoms. (Kwon et al. 2021a) [23] Extract spatial information with the help of MLP and CNN for speech emotion recognition; The proposed system shows consistent improvements for IEMOCAP, RAVDESS, and EMODB datasets. (Kwon et al. 2021b) [24] A 1D CNN-dilated end-to-end MLT-SER system, learn local and global emotional functions can automatically from speech signals. Due to the nature of this model, it takes a while to learn and test but proves its effectiveness and solidity. It is suitable for real-time speech processing. Sajjad et al. (2020) [25]A new SER framework using an essential sequence segment measurement selection based on a redial function network (RBFN). Improve the system effectiveness by using key segments instead of the complete pronouncement to reduce the calculation complexity of the overall model and Standardize the CNN functionality before processing. Kwon et al. (2020) [26] This paper discussed the hierarchical blocks of long-term convolutionary (ConvLSTM) memory with sequence learning to extract the highest distinctive emotional characteristics. The center loss function increases the final classification results on IEMOCAP and RAVDESS datasets.

In Table 2, we explain the summary for text/sentiment classification (using Machine Learning and Deep Learning) based papers and tabulated the main strengths and weaknesses of these existing approaches. In this table, We represent the reference of paper, published year, main discussed content in that paper, and advantage/disadvantage of the model. Some papers have explained Machine Learning models with the help of the low-level lexical features,top-most frequency word features, syntactic features, SVM, Naive Bayes, Bagging and Random Forest, etc. Deep Learning models like ANN, CNN, Capsule networks, DenseNet, VGG-16, and BERT, etc. achieved better results than traditional machine learning models as shown in Table 2. (CatalandNangir, 2017) [13] used multiple classifiers in a voting mechanism for better accuracy than individual classifiers. But, later individual Deep learning techniques achieves better results. Therefore, our study is related to this hybrid approach category, but we have used a hybrid approach model (majority voting) than an individual deep learning model for better accuracy.

In Table 3, we explain the current world's pandemic issue COVID-19 related papers and Deep learning models with COVID-19 related datasets. In this table, sentiment classification, negative emotions during the pandemic, detecting fake news, segmentation of CT-images, classification of social media posts, public opinion on vaccination(non-COVID-19) and discussed raw tweets from Twitter for analysis are discussed. Chakraborty et al. (2020) [31] analyzed tweets between 1st Jan 2019 to 23rd March
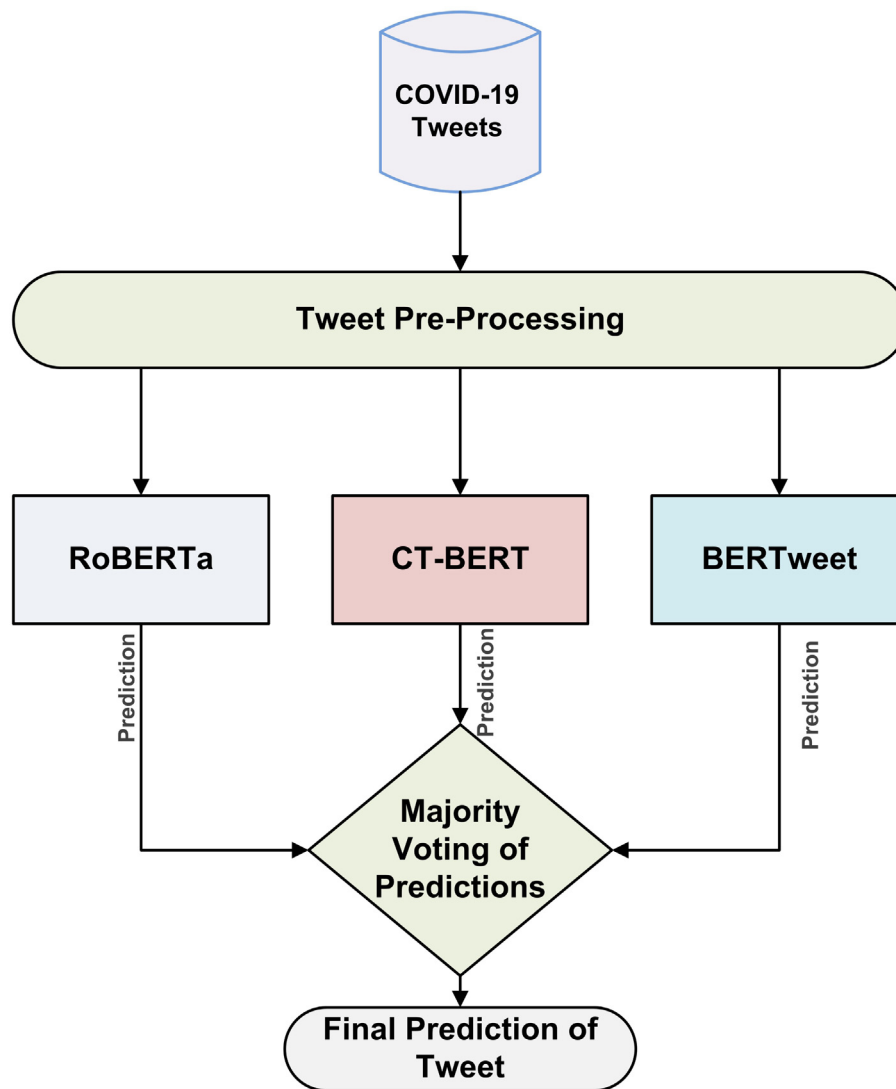
Fig. 1. Proposed (MVEDL) Ensemble Deep Learning Model Overview.

2020 for sentiment analysis, as shown in Table 3. These papers have utilized the advantage of the latest pre-trained transformer-based sequence classification models — Modified-LSTM, BiLSTM, distilBERT, CNN, BERT, ALBERT, RoBERTa, and XLNet etc. and gain state-of-the-art accuracy. Therefore, our study is related to the deep learning-based classification technique.

In Table 4,we explain the summary for COVID-19 informative tweets detection (on same Dataset) based papers. WNUT-2020 Task 2 conducted a worldwide open competition on informative tweets (COVID-19 related) detection task. In that task, we got an excellent accuracy score by using RoBERTa model on COVID-19 English tweets. (Kumar and Singh, 2020) [40] Nutcracker team has been placed 1st rank on the leaderboard, WNUT-2020 Task 2. In This table, we have discussed the models used by the papers concerning accuracy and F1-score.

3. Framework methodology

The MVEDL model detects the English COVID-19 "INFORMA-TIVE" tweets during the COVID-19 outbreak with the accuracy of 91.75% and F1-score of 91.14%. The overview of the MVEDL model is shown in Fig. 1. More details of the MVEDL model are described in the following subsections: Section 3.1 describes the tweet collection and pre-processing steps used in the MVEDL model. Section 3.2 describes the pre-trained deep learning classifiers used

in the Majority Voting-based Ensemble Deep Learning(MVEDL) model.

3.1. Tweets collection and data preprocessing

Tweets are collected from the organizers of the WNUT 2020 Shared Task2[46](Nguyen et al. 2020). The organizers collect a general Tweet corpus related to the COVID-19 pandemic based on a predefined list of 10 keywords, including: "covid-19", "coronavirus", "covid_19", "covid19", "covid_2019", "covid-2019", "covid2019", "CoronaVirusUpdate", "SARS-CoV-2" and "Coronavid19". . The obtained tweets are preprocessed using the following techniques.

3.2. Data preprocessing

Twitter data contains a lot of noise. Therefore, preprocessing on data(Twitter tweets) may help the pre-trained models in giving better performance. We perform the following data preprocessing steps, most of which have been inspired from [47]

(1) Unescape HTML tags
(2) Remove unnecessary spaces, tabs, and newlines

Table 3
Summary for COVID-19 related Text/Sentiment Classification based papers.

Paper	Year	Important discussed topic	Advantages/Limitations
[27]	2021	COVID-19 dataset from Feb 2020 to March 2020 was used for classification of sentiments	BiLSTM,CNN, distilBERT,BERT,XLNET and ALBERT ware used.
[28]	2021	Negative emotions of COVID-19 pandemic ware discussed	The keywords can be used to remove content related to COVID-19 from some relevant tweets.
[29]	2021	Detecting fake news from COVID-19	Modified-LSTM and Modified GRU used for improve accuracy.
[30]	2021	An automatic lung segmentation of CT-images of COVID-19 patients	A new fully connected (FC) layer of paralleling quantum-installed self-controlled network (PQIS-Net) gives better results.
[31]	2020	Sentiment analysis on latest COVID-19 dataset with 226668 tweets	Implementing a fuzzy rule base for Gaussian membership for analysis.
[32]	2020	Deep sentiment classification on COVID-19 comments	LSTM Recurrent Neural Network achieved higher accuracy than other machine-learning algorithms for COVID-19 — sentiment classification .
[33]	2020	The classification of the positions of critical patients	Consider different classifiers of Bayesian, linear and support vector machine (SVM).
[34]	2020	Classify data incredible or non-trustworthy.	Ensemble learning model(SVM and Random Forest) had better performance than individual models.
[35]	2020	comparative analysis of quantum backpropagation multilayer perceptron (QBMLP) and continuous variable quantum neural networks	Promising results on convoluted and sporadic data.
[36]	2020	Analysis on the largest English Twitter depression dataset(COVID 19)	Pre-trained transformer classification models BERT, RoBERTa and XLNet ware used.
[37]	2020	Auto-assign sentences for COVID-19 press briefings corpus	CNN + BERT (combined) outperforms CNN combined with other embeddings (Word2Vec, Glove, ELMo)
[38]	2019	Automatically sense the public opinion on vaccination from tweets	bag-of-words(n-grams as tokens), and for classification(SVM) is used.
[39]	2021	Automatically sense the public opinion on vaccination from tweets	bag-of-words (n-grams as tokens), and for classification(SVM) is used.

Table 4
Summary for COVID-19 Informative Tweets Detection (on same Dataset) based papers.

Paper	Year	Author	Model	Accuracy	F1-Score
[40]	2020	(Kumar and Singh, 2020)	CT-BERT +RoBERTa	91.50	90.96
[41]	2020	(Møller et al. 2020)	CT-BERT	91.40	90.96
[42]	2020	(Maveli, 2020)	RoBERTa + XLNet +BERTweet	90.40	90.11
[43]	2020	(Bao et al. 2020)	RoBERTa +MLP	90.30	90.05
[44]	2020	(Nguyen, 2020)	Majority Voting	90.15	90.08
[3]	2020	(Jagadeesh and Alphonse, 2020)	RoBERTa	89.35	89.14
[45]	2020	(Babu and Eswari, 2020)	CT-BERT+RoBERTa+ SVM(TFIDF)	89.35	88.87

Table 5
RoBERTa : Results obtained on the Test Dataset.

Model	Accuracy	F1-score	Recall	Precision
RoBERTa	**90.30**	**90.77**	**91.20**	**90.34**

(3) Replacing the mentioned hyperlinks in the tweets (depicted as HTTPURL), with URL. A simple explanation for this could be that "URL" is a more commonly used expression of hyperlinks than HTTPURL.

(4) Using the python emoji 2 library to demojise the emojis i.e. replace them with a short textual description.

The user handles were already replaced by @USER in the tweets, hence no processing was required.

3.3. RoBERTa

It is stated in Google's autonomous method published in 2018 that their robustness and optimized approach has improved the processing of natural language systems with bidirectional encoder representations by transformers (BERT). Roberta [2] applied to mask strategy on BERT's language to predict intentionally hidden text sections. For this purpose, RoBERTa has modified key hyperparameters and has trained the data in larger mini-batches and learning rates. In comparison with the original BERT,

RoBERTa increased the utility of the masked language modeling objective by offering better work efficiency. Moreover, RoBERTa is also studied with higher magnitude data compared to the original BERT.

The model was trained with different combinations of hyperparameters (batch size and learning rate) for the given COVID-19 dataset. The results obtained for each combination are evaluated using four metric measurements, namely Accuracy, F1-Score, Recall, and Precision. This RoBERTa model has trained with batch sizes of 8,16, and 32 on the COVID-19 English dataset. Table 6 clearly shows that batch size equals 16 and the learning rate equals 4e-5, which performs well compared to other combinations of batch size and learning rate on the COVID-19 dataset. Results of this RoBERTa model might change from dataset to dataset. Finally, the metrics (Accuracy(90.30%), F1-Score(90.77%), Recall(91.20%), and Precision(90.34%)) are mentioned in Table 5 are improving the performance of RoBERTa as well as the proposed MVEDL model.

(1) 'roberta-base' is used
(2) 4e-5 tends to work well with this RoBERTa transformer model.
(3) We have used a batch size of 16.
(4) The maximum sequence length of a training tweet has been fixed to 143.

Table 6
RoBERTa experimental results on COVID-19 English Tweets Dataset.

Batch size	Learning rate	TP	FP	FN	TN	Accuracy	F1-Score	Recall	Precision
	1e−5	890	54	143	913	90.14	90.26	**94.41**	86.45
	2e−5	860	84	123	933	89.64	90.26	91.74	88.35
8	3e−5	864	80	121	935	89.95	90.29	92.11	88.54
	4e−5	838	106	111	945	89.14	89.70	89.91	89.48
	5e−5	832	112	114	942	88.7	89.28	89.37	89.20
	1e−5	864	80	133	923	89.35	89.65	92.02	87.40
	2e−5	867	77	126	930	89.85	90.15	92.35	88.06
16	3e−5	858	86	135	921	88.94	89.28	91.45	87.21
	4e−5	852	92	102	954	**90.30**	**90.77**	91.20	**90.34**
	5e−5	849	95	119	937	89.30	89.75	90.79	88.73
	1e−5	873	71	138	918	89.55	89.77	92.82	86.93
	2e−5	866	78	117	939	90.25	90.59	92.33	88.92
32	3e−5	864	80	126	930	89.7	90.02	92.07	88.06
	4e−5	868	76	141	915	89.14	89.39	92.33	86.64
	5e−5	850	94	119	937	89.35	89.79	90.88	88.73

Table 7
RoBERTa model wrong predicted tweets.

Tweets	True label	Predicted label
@USER @USER Absolutely! They've been blaming NHS structure, public, NHS staff 4 changing PPE 2 often & now claim those poor NHS staff who've died looking after Covid patients probably got it outside work! They're completely devoid of any decency or respect for the dead or living! #ToryScum	INFORMATIVE	UNINFORMATIVE
I'll pull a Kurt Kloss and ask peeps here if their companies have a written COVID policy on what happens when a coworker tests positive post the mandated work from home period. Curious to know if anyone's read published solutions to the new workplace normal.	INFORMATIVE	UNINFORMATIVE
@USER In authoritarian countries, officials tend to sanitize ugly truths to please their big boss. Signs of that when sec. duque said that confirmed covid 19 in a Pinoy without travel history doesnt mean local transmission. Putok sa buho yung virus?	INFORMATIVE	UNINFORMATIVE
I have a sore throat, dry cough, headache & im feeling weak.Yesterday on @USER permanet secretary for ministry of health said if anyone suspects they have #COVID2019 they must stay at home & call a number there's a team deployed to assist from home. I need the contact	INFORMATIVE	UNINFORMATIVE

(5) The hidden dropout was equivalent to 0.05 to avoid over-fitting.

(6) The hidden size has been set to 768 for 'roberta-base.'

(7) For the query optimizer, an 'adam' has been used.

(8) Trains up to 10 epochs.

RoBERTa wrongly predicted tweets:

Some highly misclassified tweets are mentioned in Table 7. The reason might be the assign the labels of samples from English experts to English experts. The above reason has created a noise in the label data, which is used for model training.

3.4. CT-BERT

In an attempt to read and analyze the Twitter content relevant to covid-19, a Covid-Twitter-BERT (CT-BERT) [47] model is implemented. The model depends on the bert-large model (English, non-cased, entire word mask). The bert-large is trained on raw text data extract from a free book corpus (0.8b words) and Wikipedia (3.5b words). To improve the performance in the subdomain, numerous transformer-based models are trained on specialized corpora. These models are a proxy for traditional language models and are often qualified for downstream work.

As same as RoBERTa, This model is also trained with different combinations of hyperparameters (batch size and learning rate) for the given COVID-19 dataset. The results obtained for each combination are evaluated using four metric measurements, namely Accuracy, F1-Score, Recall, and Precision, and tabulated in Table 9. CT-BERT model has trained with batch sizes of 8 and 16 on The COVID-19 English dataset. In this paper, this model has not been trained due to the CT-BERT word embedding size (1.47 GB). Table 9 clearly shows that batch size is equals 8 and learning rate equals 1e-5, which performs well compared

Table 8
CT-BERT : Results obtained on the Test Dataset.

Model	Accuracy	F1-score	Recall	Precision
CT-BERT	**91.10**	**91.57**	**91.57**	**91.57**

to other combinations of batch size and learning rate on the COVID-19 dataset. Results of this CT-BERT model might change from dataset to dataset. Finally the metrics (Accuracy(91.10%), F1-Score(91.57%), Recall(91.57%), and Precision(91.57%)) are mentioned in Table 8 are improving the performance of CT-BERT as well as the proposed ensemble model.

(1) 'bert-large' is used.

(2) Trained on a collection of 22.5M corona-related tweets.

(3) The data consisted of 40.7M sentences and 633M tokens.

(4) set batch size to 8.

(5) "learning rate" has been set to 1e-5.

CT-BERT wrongly predicted tweets:

Some highly misclassified tweets are mentioned in Table 10. The reason might be the assign the labels of samples from English experts to English experts. The above reason has created a noise in the label data, which is used for model training.

3.5. BERTweet

BERTweet [48] used the same bert-base architectural design that is trained to have a masked language modeling purpose. BERTweet pre-training procedure is based on RoBERTa [2], which optimized the BERT pre-training approach for more robust performance. The pre-train BERTweet corpus consists of 850 m English tweets (16b word tokens 80GB), 845m streamed tweets

Table 9
CT-BERT experimental results on COVID-19 English Tweets Dataset.

Batch size	Learning rate	TP	FP	FN	TN	Accuracy	F1-Score	Recall	Precision
	1e−5	855	89	89	967	**91.10**	**91.57**	91.57	91.57
	2e−5	846	98	86	970	90.80	91.33	90.82	**91.85**
8	3e−5	853	91	93	963	90.8	91.27	91.36	91.19
	4e−5	863	81	110	946	90.45	90.83	92.11	89.58
	5e−5	875	69	125	931	90.3	90.56	93.10	88.16
	1e−5	860	84	111	945	90.25	90.64	91.83	89.48
	2e−5	874	70	126	930	90.2	90.46	93.00	88.06
16	3e−5	880	64	119	937	90.85	90.73	**93.60**	88.73
	4e−5	854	90	122	944	89.45	89.90	91.29	88.55
	5e−5	875	69	120	936	90.55	90.82	93.13	88.63

Table 10
CT-BERT model wrong predicted tweets.

Tweets	True label	Predicted label
@USER @USER 2 weeks later: Artist Torey Lanez is the first celebrity to pass away from CoVid19. ?? He really gonna be roasting with that fever of 105.	UNINFORMATIVE	INFORMATIVE
@USER I don't listen to any news since COVID19 killed the first 50 patients. I read news, but I don't have to listen to the ??????? .	UNINFORMATIVE	INFORMATIVE
In light of the confirmation of a #COVID19 case in Uganda, President Museveni will today,at 4pm, address the country on what further steps to take in a bid to curb the possible spread of the pandemic. #MonitorUpdates #CoronavirusPandemic HTTPURL	UNINFORMATIVE	INFORMATIVE
I did notice from a recent Folha article that 9 out of the first 10 Coronavirus deaths in Brazil were all people who had died in private hospitals. This could be for several reasons but it's definitely something to watch out for.	UNINFORMATIVE	INFORMATIVE

Table 11
BERTweet : Results obtained on the Test Dataset.

Model	Accuracy	F1-score	Recall	Precision
BERTweet	**89.8**	**90.15**	**91.92**	**88.44**

from 2012 to 2019, and 5 m covid-19 pandemic-related tweets. BERTweet is seen outperforming the preceding state-of-art models such as roberta-base and xml-r-base rivals on text classification tagging, named object identification, and part-of-speech downstream tweet NLP tasks.

As same as RoBERTa, this model is also trained with different combinations of hyperparameters (batch size and learning rate) given the COVID-19 dataset. The results obtained for each combination are evaluated using four metric measurements, namely Accuracy, F1-Score, Recall and Precision, and also tabulated in Table 12. BERTweet model has trained with batch sizes of 8,16, and 32 on the COVID-19 English dataset. Table 12 clearly shows that with batch size 8 and learning rate 2e-5, this model performs well compared to other combinations on the COVID-19 dataset. Results of this BERTweet model might change from dataset to dataset. Finally, the metrics are (Accuracy(89.80%), F1-Score(90.15%), Recall(91.92%), and Precision(88.44%)) listed in the table Table 11 improve the performance of both BERTweet and the proposed ensemble model. Some highly misclassified tweets are mentioned in Table 13.

(1) 'bertweet-base' is used
(2) 2e-5 tends to work well with this BERTweet transformer model.
(3) Batch size of 8 used.
(4) The maximum sequence length of a training tweet is fixed to 143.
(5) To avoid an over-fitting, set hidden dropout is equated to 0.05.
(6) The hidden size for 'bertweet-base' is fixed to 768.
(7) An 'adam' optimizer is used.
(8) Trained up to 10 epochs.

3.6. Majority voting

A vote is a meta-algorithm that performs the decision process by applying participant classifiers [13,49]. There are several combination rules for the Vote algorithm, such as majority voting, minimum probability, maximum probability, multiplication of probabilities, and an average of probabilities. In this paper, Majority voting combines deep learners using a majority vote or predicted probabilities for the sample classification. For the majority vote, a sample class label is used to identify most of the class trying to label predicted by each classifier. The proposed model has used three deep learning classifiers and a binary classification problem. For example, If the prediction results in the majority voting rule for a sample is like (i.e., $Dl\_Classifier1 \rightarrow Class2$, $Dl\_Classifier2 \rightarrow Class2$, $Dl\_Classifier3 \rightarrow Class1$), then the data sample would be classified as Class2. To avoid a tie, the number of classifiers should be odd (> 1) .

The voting mechanism uses to increase the accuracy of the model, to combine different classifiers (classifiers mentioned above). Various classifiers in COVID-19 English datasets have performed well at other data points. For instance, consider the tweets related to COVID-19 tweets as data points. Tweets are correctly classified by some classifiers, and specific other classifiers may misclassify the same tweets. The correctness of the model is hard to predict. However, by combining various classifiers using the voting system, it is better to alter the accuracy of the model. The performance metrics of each ensemble algorithm depend on the batch size and learning rate used within it.

It aims to combine three deep learning models, namely RoBERTA, CT-BERT, and BERTweet (the above-mentioned models), following the voting mechanism to attain better accuracy of the proposed model. Each of the models that are involved in the ensemble process has given desired performance at various data points pertaining to the COVID-19 dataset. For instance, consider the tweets related to the coronavirus as data points. In this process, if any two of these three models mentioned above give the best prediction that is declared as the correct prediction of the tweet. The correctness of the model is hard to predict. However, by integrating the voting system, the exactness of the

Table 12
BERTweet experimental results on COVID-19 English Tweets Dataset.

Batch size	Learning rate	TP	FP	FN	TN	Accuracy	F1-Score	Recall	Precision
	1e−5	854	90	127	929	89.14	89.71	91.16	87.97
	2e−5	862	82	122	934	**89.80**	90.15	91.92	88.44
8	3e−5	861	83	132	924	89.25	89.57	91.75	87.50
	4e−5	856	88	155	901	87.85	88.11	91.10	85.32
	5e−5	837	107	144	912	87.45	87.90	89.49	86.36
	1e−5	853	91	131	925	88.90	89.06	91.04	87.59
	2e−5	859	85	119	937	**89.80**	89.28	91.68	88.73
16	3e−5	839	105	118	938	88.85	**90.18**	89.93	**88.82**
	4e−5	847	97	131	925	88.6	89.37	90.50	87.59
	5e−5	869	75	136	920	89.45	89.02	**92.46**	87.12
	1e−5	867	77	150	906	88.64	88.86	92.16	85.79
	2e−5	864	80	151	905	88.44	88.68	91.87	85.70
32	3e−5	846	98	133	923	88.44	88.87	90.40	87.40
	4e−5	863	81	135	921	89.20	89.50	91.91	87.21
	5e−5	864	80	144	912	88.80	89.06	91.93	86.36

Table 13
BERTweet model wrong predicted tweets.

Tweets	True label	Predicted label
@USER @USER Absolutely! They've been blaming NHS structure, public, NHS staff 4 changing PPE 2 often & now claim those poor NHS staff who've died looking after Covid patients probably got it outside work! They're completely devoid of any decency or respect for the dead or living! #ToryScum'	INFORMATIVE	UNINFORMATIVE
.@USER McConnell is closely following the confirmed case of coronavirus in #Kentucky. Federal funding is on the way to bolster efforts throughout the Commonwealth to keep families safe.	INFORMATIVE	UNINFORMATIVE
#TedCruz to self-quarantine after interacting with a CPAC attendee who is positive for coronavirus But I thought its only a hoax? You mean its not a hoax when you are concerned, @USER Ben Carson says you can go to work. Trump does too. Just sayin'	UNINFORMATIVE	INFORMATIVE
This is Jains Kences Retreat, Virugambakkam. (Yes, the same apartments that got a lot of attention on social media because of the Covid Positive guy right OUTSIDE our community) . We have been cautions and indoors…HTTPURL	INFORMATIVE	UNINFORMATIVE

model can be increased. The accuracy of this ensemble depends on the above mentioned individual models' batch size and the learning rate. The above algorithm decides the best learning rate for each ensemble model. Thus, the combined ensemble models using the voting mechanism bring significant improvement to the model. A majority voting is used to predict the tweet's class label, as shown in algorithm 1.

$$Final\_Tweet\_prediction = Majority(RoBERTA(k),$$
$$CTBERT(k), BERTweet(k))$$

The majority voting counts the votes of all the models and selects the class with the most votes as a prediction. Formally, the final prediction is given by:

$$M_v = \sum_{i=1}^{N} P_i(c)$$

where M_v denotes the majority voting value of the model, N is the total no of classifiers, $P_i(c)$ is ith classifier prediction value, which is 0 or 1.

4. Experimental results and analysis

All the experiments in our work have been carried out in Google CoLab interface with chrome browser and the following configuration: GPU processor name: GP100, GPU variant : GP100-893-A1 with 100% performance, Tesla P100-PCIE-16GB graphic card , 16 GB RAM allocated, 160 GB programming space allocated by the NVIDIA Corporation.

This section addresses datasets, explanations of model parameters, and performance assessments. In addition, the proposed solution is compared with current methods. The implementation

---

Algorithm 1: Proposed Method based on Majority based Ensemble Technique

0: INFORMATIVE tweet related to the COVID-19.
1: UNINFORMATIVE tweet non-related to the COVID-19.
ROBERTA(): An array of RoBERTa model predicted values(values 0 or 1)
CTBERT(): An array of CTBERT model predicted values(values 0 or 1)
BERTweet(): An array of BERTweet model predicted values(values 0 or 1)
S: Test dataset
Input: S, ROBERTA(), CTBERT(), BERTweet()
Steps:
for $k = 1$ to size of (S) **do**
  Final_Tweet_prediction = Majority(RoBERTA(k), CTBERT(k), BERTweet(k))
end for

Output: Final_Tweet_prediction (0, 1)

---

was performed in python language by using the Huggingface library [50]. To fine-tune our baseline models, we employ "ktrain" package [51]. We use "AdamW" optimizer [52] with a list batch size of 8, 16, 32 and learning rates in the set 1e−5, 2e−5, 3e−5, 4e−5, 5e−5. We fine-tune the models for 25 epochs and select the best checkpoint based on the performance of the model on the validation set. Datasets for this experiment have been obtained from WNUT-2020 at Task 2. DATASET contains around 10000 tweets, whereas the 7000 tweets were used for training, 1000 tweets were for validation and 2000 tweets were used for testing.

Table 14
Details of COVID-19 English Tweets Dataset.

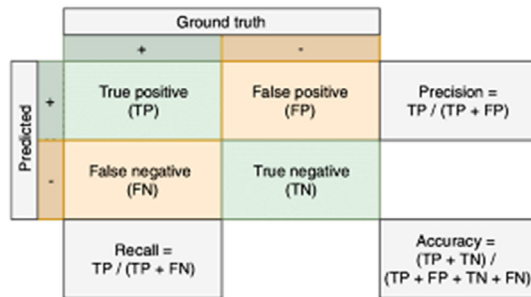| COVID-19 dataset	UNINFORMATIVE	INFORMATIVE
Training Tweets	3697	3303
Validation Tweets	528	472
Test Tweets	1056	944

Fig. 2. Sample confusion matrix.

4.1. Dataset

In the COVID-19 pandemic (2020), WNUT-2020 at Task 2 organizers provided the COVID-19 English tweets [46](Nguyen et al. 2020) dataset with the tweet-ids, tweet text, and label ("INFORMATIVE" and "UNINFORMATIVE") in the tsv format. The above dataset size is 10,000, which was collected from Twitter API. This dataset has been used for train the proposed model for predicting informative tweets in a testing phase. 70%(7000) of tweets were used for training, 20%(2000) of tweets were used for testing, and 10%(1000) of tweets were used for validation. The complete details of the COVID-19 dataset as shown in Table 14.

4.2. Experiment setup

The model's outcome relies on the use of a classifier. Therefore, different experiments are performed with the help of the following classifiers.

(1) RoBERTa deep learning classifier.
(2) CT-BERT deep learning classifier.
(3) BERTweet deep learning classifier.
(4) Majority Voting-based Ensemble.

4.3. Performance measures

In the following parameters, the performance of the models such as Precision, F1-score, Accuracy, and Recall are evaluated. They are explained by means of the matrix of confusion.

4.3.1. Confusion matrix

Also known as the error matrix is the confusion matrix. The results of the test algorithm can be summarized. The row represents the predictive values of the tweets in the confusion matrix, and the column represents the actual tweet value.

The first line, the first column, is true positive in Fig. 2. (TP). The number of tweets associated with information is properly predicted. In the first row, the second column is false positive (FP) in the matrix. The number of tweets associated with information is incorrectly predicted. In the first column, the second row in the confusion matrix denotes false negative (FN). Uninformative tweets are incorrectly predicted by It. In the second column, second row in the confusion matrix represents true negative (TN). It specifies the correctly foreseen number of tweets related to uninformative.

4.4. Performance analysis

This section can be divided into four subsections. The first subsection describes the comparison of the machine learning model's performance. The second subsection describes the comparison of the deep learning model's performance. The third subsection describes the comparison of the ensemble deep learning model's performance, and the last subsection describes a comparison of the proposed model (Majority Voting-based Ensemble Deep Learning) performance with the existing methodologies.

4.4.1. Performance metrics evaluation among state-of-art machine learning models

The organizers of the WNUT-2020 Shared Task 2 did not provide any baseline method for the English COVID-19 dataset. In this subsection, SVM, Decision Tree, BoW, and Random Forest have been considered for predicting informative tweets. The Random Forest classifier has achieved an accuracy of 82.06 percentage and precision of 81.92 percentage respectively, SVM classifier has performed better in F1-score with 82.71 percentage and Recall with 84.19 percentage respectively. The remaining machine learning algorithms have performed better than the BOW method, as concluded with Fig. 3.

(1) Random Forest classifier has performed a better accuracy compare with Decision Tree, BoW, and SVM (see Fig. 3(a)).
(2) SVM classifier has achieved a better F1-score compare with Decision Tree, BoW, and Random Forest (see Fig. 3(b)).
(3) Random Forest classifier has gained a better precision compare with Decision Tree, BoW, and SVM (see Fig. 3(c)).
(4) SVM classifier has achieved a better Recall compare with Decision Tree, BoW, and Random Forest (see Fig. 3(d)).

4.4.2. Performance metrics evaluation among state-of-art deep learning models

This subsection has considered state-of-art deep learning models CNN, BERT, DistliBERT, BERTweet, RoBERTa, and CT-BERT. The MAX_LENGTH of the tweet is set to 143 for better train the models, and training the corpus is related to English. Here testing tweets are in the English language. We have used batch sizes of 8, 16, and 32 for train the models and learning rate of values 1e−5, 2e−5, 3e−5, 4e−5, 5e−5.

CT-BERT has achieved better than the RoBERTa, BERT, BERTweet, DistilBERT, and CNN models as shown in Fig. 4. CT-BERT model has performed better because it has attained better true positive and false negative values than other competitors in the race. Accuracy and F1-score of the CT-BERT are the best values as shown in the 4(a), 4(b) than the other models due to the high precision value. As shown in 4(d) the recall value of BERTweet has occupied the first place and, RoBERTa and CT-BERT have engaged second and third place, respectively. The precision value of the CT-BERT is tiny high than the DistilBERT but better than the other models as shown in 4(c).

Almost CT-BERT has outperformed well than the other deep learning models because it has pre-trained on a large corpus of Twitter messages on the topic of COVID-19 (see Fig. 6).

4.4.3. Performance metrics evaluation among ensemble deep learning models

This subsection has considered recent ensemble deep learning models as shown in Fig. 5 (RoBERTa + MLP, CT-BERT + RoBERTa, RoBERTa + XLNet + BERTweet, CT-BERT + RoBERTa + TFIDF(SVM)) in terms of performance metrics. The comparisons are shown in Fig. 5(a) for accuracy, 5(b) for F1-score, 5(c) for precision, and 5(d) for recall. Almost in above all cases, the combination of CT-BERT and RoBERTa has attained the best performance than other
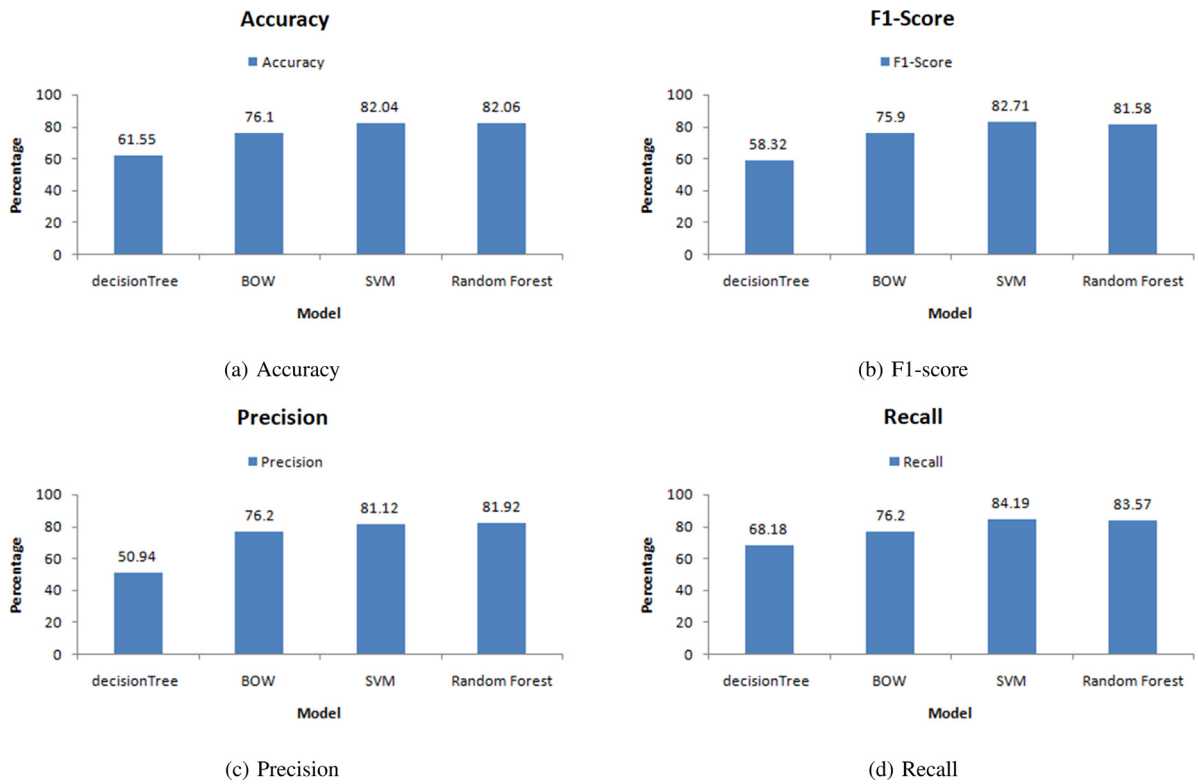
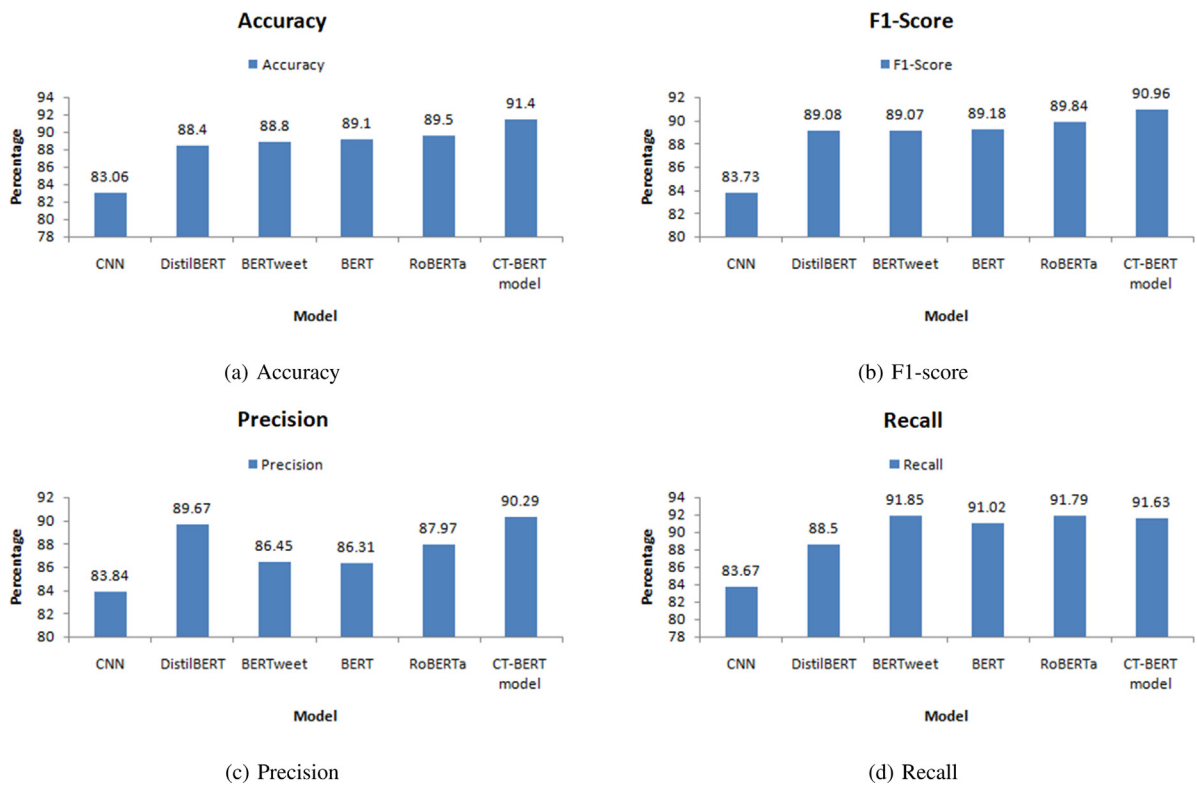**Fig. 3.** Machine Learning models evaluation metrics.

Fig. 4. Deep Learning models evaluation metrics.

ensemble deep learning models as shown in Fig. 5. It has shown that the ensemble deep learning models are best compared with the deep learning models. It is well understood that CT-BERT + RoBERTa model has achieved an accuracy of 91.5% and an F1 score

of 90.96% compared with other ensemble models as mentioned. From the above results, The MVEDL model has been included the deep learning classifiers CT-BERT and RoBERTa.
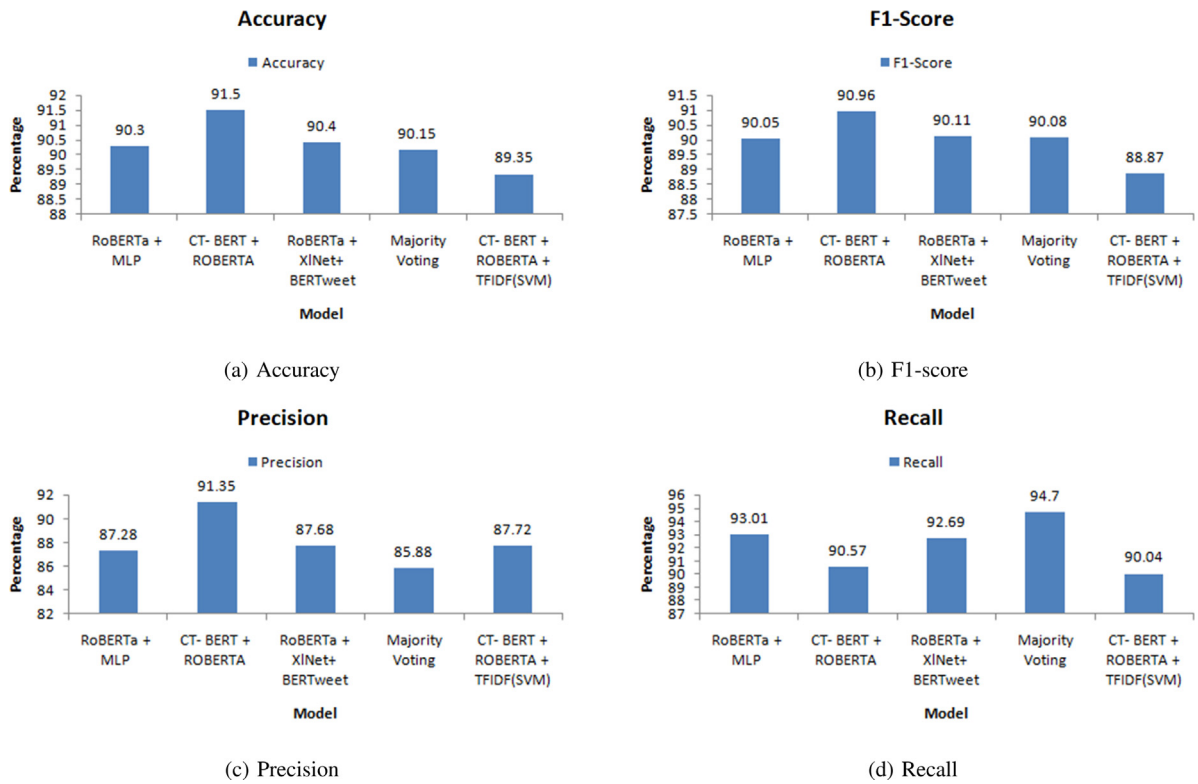
(a) Accuracy

(b) F1-score

(c) Precision

(d) Recall

Fig. 5. Ensemble Deep Learning models evaluation metrics.
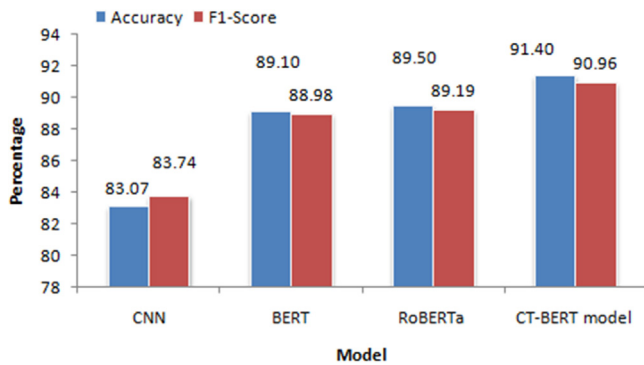

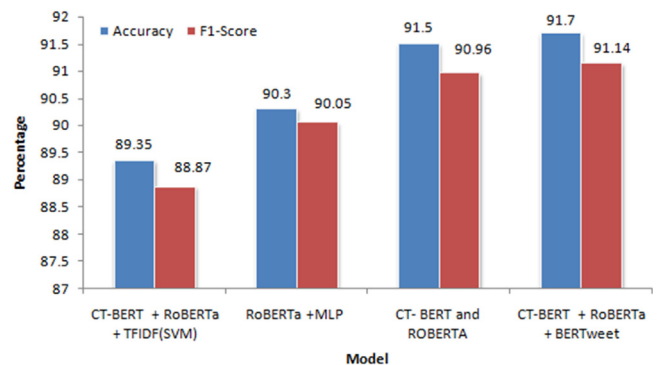**Fig. 6.** Deep Learning models performance metrics.

Fig. 7. Ensemble Deep Learning models performance metrics.

Table 15
Performance comparison of the proposed model with state-of-art models.

Model	F1-score	Accuracy
Random Forest Classifier	81.58	82.06
SVM Classifier	82.71	82.04
CNN	83.73	83.06
BERT	88.97	89.10
DistilBERT	89.08	88.40
RoBERTa [3]	89.18	89.50
RoBERTa + MLP [43]	90.05	90.30
CT-BERT model [41]	90.96	91.40
CT-BERT + RoBERTa [40]	90.96	91.50
Proposed Model(MVEDL)	**91.14**	**91.75**

Table 16
Proposed model results obtained on the test Dataset.

Model	Accuracy	F1-score	Recall	Precision
Proposed Model(MVEDL)	**91.75**	**91.14**	**93.58**	**89.94**

4.5. Proposed model comparison with ensemble deep learning techniques

In this subsection, the proposed model (MVEDL) is compared against state-of-art machine learning models, deep learning models (CNN, BERT, RoBERTa, DistilBert, BERTweet, and CT-BERT) and ensemble models (see Fig. 7) in terms of accuracy and F1-score, and shown in Table 15. Fig. 8 shows that the deep learning models are exceptional when compared with the other machine learning models. The MVEDL has performed with 91.75% accuracy, 91.14% F1-score, 93.58% recall, and 89.94% precision, as shown in Table 16.

Referring to Table 15, it is well understood that our MVEDL model has achieved an accuracy of 91.75% and an F1 score of 91.14% compared with existing models as mentioned. This gives a clear indication that the model has succeeded in differentiating the informative tweets related to COVID-19 disease outbreaks from combined tweets.
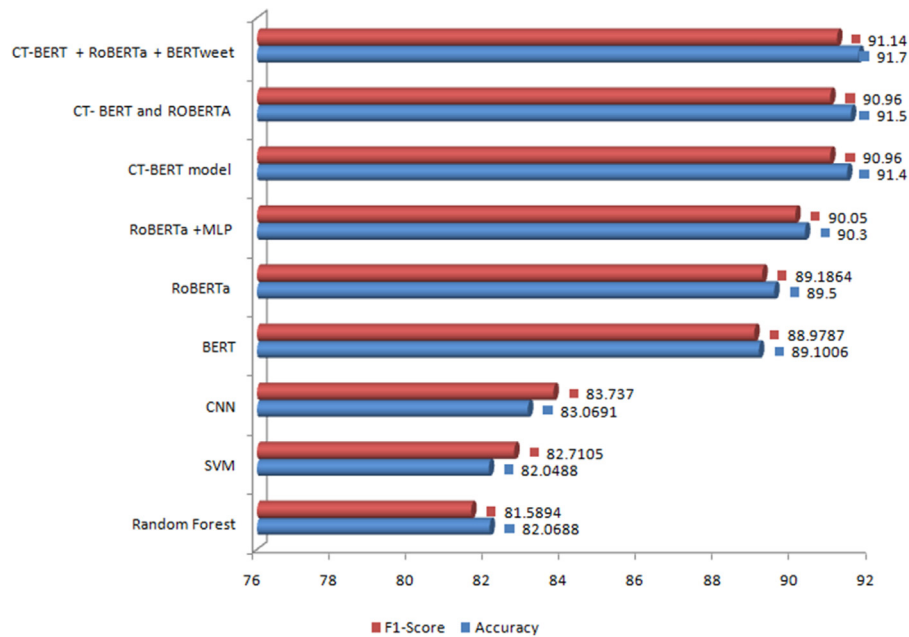
Fig. 8. Proposed model Results comparing with state-of-art Models.

Table 17
Comparative table representing positive, negative and neutral tweets labeled by TEXTBLOB method for Total tweets and Informative tweets.

Tweet type	TOTAL TWEETS		INFORMATIVE TWEETS	
	TEXTBLOB	Percentage	TEXTBLOB	Percentage
Total tweets in the dataset	226 668	100	19877	100
Total tweets with sentiment	226 668	100	19877	100
No. of positive tweets	98 504	43.46	10975	55.21
No. of negative tweets	53 566	23.63	4398	22.13
No. of neutral tweets	74 598	32.91	4504	22.66

5. Real-time application

This pandemic has claimed millions of lives and leads the world to a complete health crisis and financial recession. For the victims, government agencies, and NGOs, it would have been easier at this stage to gather structured social media information. This requires strict tweet quality control to ensure that valuable content on these most used blogs is shared.

The aim of this paper will lie in the initiation of fact-checking implied on social sites before its wide sharing, detecting false & uninformative news, and prevented from being disseminated within netizens.

For this, we considered real-time tweet classification using the Sentiment Analysis Process(SAP) for extracting informative tweets with the help of the proposed MVEDL model to make sense of the uncountable tweets posted per second in social media. Sentiment analysis is one of the best methods for expressed desires by transforming data into a structured format from unstructured texts.

This SAP aims to extract most frequency words [31] and classify the sentiments into positive, negative, and neutral tweets on informative tweets. The Natural Language Toolkit library (NLTK) is used as an appropriate language text processor. The data flow diagram of Real-Time informative tweet classification using the MVEDL model for sentiment analysis is shown in Fig. 9. Towards this start, a dataset named Tot_DATA_SET 226,668 [53] distinct tweets from December 2019 to May 2020 is taken into account and used for the classification of informative tweets. For data preprocessing, used the "re" python module cleanses symbols

such as @, RT, #, URL, numeric values and removes duplicate rows and punctuation marks.

The proposed Majority Voting-based Ensemble Deep Learning (MVEDL) model has classified Tot_DATA_SET dataset into 19877 tweets as "INFORMATIVE" and remaining as "UNINFORMATIVE". The informative tweets are named as Info_DATA_SET that is used for SAP. In the SAP process, TextBlob algorithm is used to calculate the total numbers of positive, negative, and neutral tweets. Table 17 shows the number of positive, negative, and neutral tweets labeled by TextBlob. From the Tot_DATA_SET dataset, we have observed that the number of positive tweets is nearly 20 percent more than the number of negative tweets and about 10 percent more than the number of neutral tweets, as shown in Fig. 10(a). But from Info_DATA_SET, the number of positive tweets are nearly 33 percent more than both the number of negative tweets and neutral tweets as shown in Fig. 10(b).

The Sentiment Analysis Process(SAP) helps (see Fig. 10) to understand over time positive, negative, and neutral trends. As shown in Fig. 10(a), people were more likely to express non-positive feelings than positive feelings and express more neutral feelings than negative feelings from sentiment analysis on total raw tweets. On the other side as shown in Fig. 10(b), people were more likely to express positive feelings than non-positive feelings and tell an almost identical number of negative and neutral feelings from sentiment analysis on informative tweets.

As shown in Table 18, from total raw tweets, the top 20 frequency words of negative, positive, and neutral sentiments are listed. We have observed some non-related COVID-19 words are placed in the top 20, for example, "trump", "work" etc. As shown in the Table 19, from informative tweets, the top 20 frequency
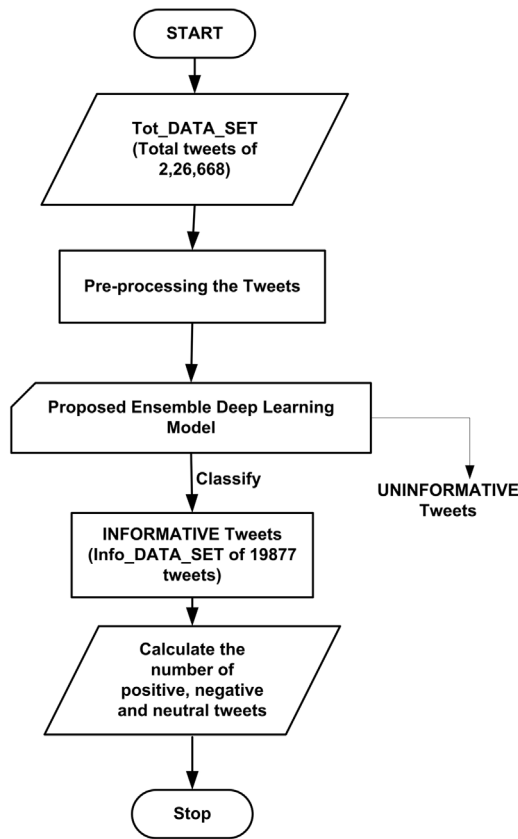
Fig. 9. Various steps involved in the COVID-19 informative tweets sentiment analysis.

Table 18
Analysis of Total COVID-19 English tweets from Jan 2020 to March 2020.

Negative		Positive		Neutral	
Words	Count	Words	Count	Words	Count
covid19	18605	covid19	40593	covid19	27976
covid	17236	covid	26675	coronavirus	20074
coronavirus	11362	coronavirus	22514	covid	15004
corona	5437	new	11848	corona	7407
virus	3518	people	11401	pandemic	3612
deaths	3002	cases	8300	people	3539
trump	2997	corona	8202	virus	2784
pandemic	2499	deaths	7292	deaths	2681
sick	2101	corona	6895	trump	2496
home	2016	virus	5190	cases	2255
death	1922	pandemic	5125	lockdown	1848
bad	1910	positive	3999	health	1819
new	1898	trump	3970	death	1785
cases	1894	great	3598	news	1763
patients	1796	health	3512	world	1672
ill	1755	death	3442	home	1510
work	1606	patients	3103	patients	1493
health	1525	news	3018	hydroxy chloroquine	1389
china	1519	safe	2707	china	1357
dead	1515	lockdown	2635	vaccine	1313

Table 19
Analysis of Informative COVID-19 English tweets from Jan 2020 to March 2020.

Positive		Negative		Neutral	
Words	Count	Words	Count	Words	Count
covid19	5916	covid19	2068	covid19	2095
cases	5666	covid	1519	coronavirus	1481
new	4469	deaths	1002	deaths	1118
deaths	3447	coronavirus	966	cases	1092
coronavirus	3372	cases	852	covid	952
positive	2293	people	708	died	489
covid	2134	died	576	death	434
confirmed	1779	hospital	547	people	351
total	1656	dead	511	hospital	329
people	1109	100000	486	total	311
tested	1102	new	410	home	309
reported	955	tested	384	100000	256
died	926	virus	340	ewing	220
death	816	home	321	corona	208
100000	680	americans	312	dies	200
reports	664	sick	309	symptoms	188
county	656	death	301	trump	187
recovered	640	symptoms	298	reported	160
health	613	negative	290	2020	157
hospital	609	positive	269	county	152

words of negative, positive, and neutral sentiments are listed. We have observed some related COVID-19 words in the top 20 compare with the above Table 18, for example, "confirmed", "tested" and "recovered" etc.,

6. Conclusion

The main aim of this paper is to show a novel NLP application for the detection of meaningful latent issues and an emotional classification on COVID-19 tweets. We believe that the results of the paper will help people to understand the concerns and needs of COVID-19 tweet analysis. Our results may also contribute to enhancing practical public health services strategies and COVID-19 interventions.

In this paper, we have proposed Majority Voting-based Ensemble Deep Learning (MVEDL) model for detecting informative tweets during the COVID-19 pandemic. The approach of majority voting is designed to boost the Entrenching our model. To improve model performance, we experiment with different combinations of machine learning and deep learning models, but finally, we achieved state-of-art performance using COVID-Twitter BERT, BERTweet, and RoBERTa deep learning models. The proposed model has shown 91.75 percent accuracy and 91.14 percent F1-score and outperforms the traditional machine learning and deep learning models.

A dataset of 226,668 COVID-19 tweets is collected from December 2019 to May 2020 to identify informative tweets as a proposed model application. We have applied the sentiment analysis technique on both datasets and have listed the most frequent words(up to 40).

The limitations of our MVEDL model are CT-BERT, BERTweet, and RoBERTa are pre-trained models with large memory (1.47GB,850MB, and 657MB, respectively) for corpus training. The time complexity of the models also very high compared to machine learning models. In this paper, these models have taken 3207.518 s (batch size=8), 1184.481 s (batch size=8), and 1064.962 s (batch size=16) respectively. The process of voting has taken 62.592 s Our model has run on parallel python interfaces, so the total time complexity is 3270.11 s (CT-BERT time complexity + voting time complexity). We are planning to apply data compression techniques to improve the model performance.

This paper currently describes only COVID-19 pandemic English tweets. The MVEDL model performance can increase with a large training dataset. This model may be capable of predicting tweets related to similar types of diseases in the future. We can train other combinations with new transformer-based models on a large COVID-19 dataset for better results in the future.
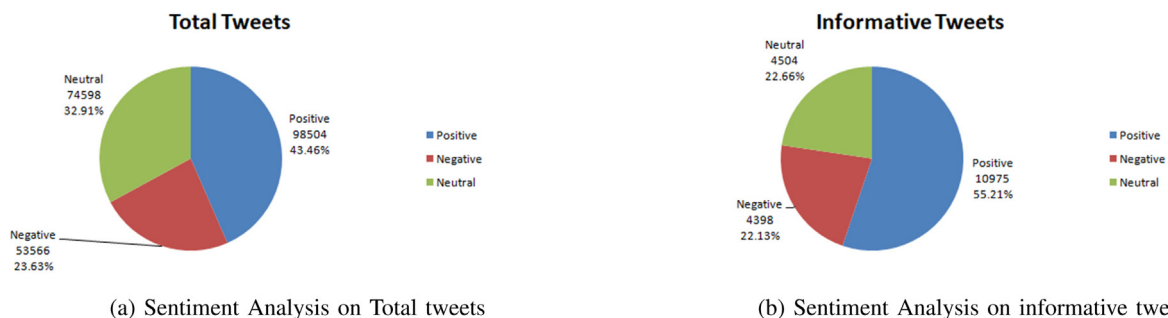
(a) Sentiment Analysis on Total tweets

(b) Sentiment Analysis on informative tweets

Fig. 10. Comparison of the Sentiment Analysis on Total and informative COVID-19 tweets.

Declaration of competing interest

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

References

[1] Dat Quoc Nguyen, Thanh Vu, Afshin Rahimi, Mai Hoang Dao, Linh The Nguyen, Long Doan, WNUT-2020 task 2: Identification of informative COVID-19 english tweets, in: Proceedings of the Sixth Workshop on Noisy User-Generated Text (W-NUT 2020), Association for Computational Linguistics, Online, 2020, pp. 314–318, http://dx.doi.org/10.18653/v1/2020.wnut-1.41, https://www.aclweb.org/anthology/2020.wnut-1.41.

[2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019, arXiv preprint arXiv:1907.11692.

[3] 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), 2020, pp. 450–454.

[4] Khiem Vinh Tran, Hao Phu Phan, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, UIT-Hse at WNUT-2020 task 2: Exploiting CT-BERT for identifying COVID-19 information on the Twitter social network, 2020, arXiv preprint arXiv:2009.02935.

[5] Dat Quoc Nguyen, Thanh Vu, Anh Tuan Nguyen, Bertweet: A pre-trained language model for english tweets, 2020, arXiv preprint arXiv:2005.10200.

[6] Sreenivasulu Madichetty, M. Sridevi, A novel method for identifying the damage assessment tweets during disaster, Future Gener. Comput. Syst. 116 (2021) 440–454.

[7] Sreenivasulu Madichetty, Sridevi Muthukumarasamy, P. Jayadev, Multimodal classification of Twitter data during disasters for humanitarian response, J. Ambient Intell. Humaniz. Comput. (2020) 1–15.

[8] Sreenivasulu Madichetty, M. Sridevi, Improved classification of crisis-related data on Twitter using contextual representations, Procedia Comput. Sci. 167 (2020) 962–968.

[9] Sreenivasulu Madichetty, M. Sridevi, Classifying informative and non-informative tweets from the twitter by adapting image features during disaster, Multimedia Tools Appl. 79 (39) (2020) 28901–28923.

[10] Jaeyoung Kim, Sion Jang, Eunjeong Park, Sungchul Choi, Text classification using capsules, Neurocomputing 376 (2020) 214–221.

[11] Rami Aly, Steffen Remus, Chris Biemann, Hierarchical multi-label classification of text with capsule networks, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2019, pp. 323–330.

[12] Vicente García-Díaz, Jordán Pascual Espada, Rubén González Crespo, B Cristina Pelayo G-Bustelo, Juan Manuel Cueva Lovelle, An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis, Appl. Soft Comput. 67 (2018) 822–833.

[13] Cagatay Catal, Mehmet Nangir, A sentiment classification model based on multiple classifiers, Appl. Soft Comput. 50 (2017) 135–141.

[14] Rudy Prabowo, Mike Thelwall, Sentiment analysis: A combined approach, J. Informetr. 3 (2) (2009) 143–157.

[15] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, Jianfeng Gao, Deep learning based text classification: A comprehensive review, 2020, arXiv preprint arXiv:2004.03705.

[16] Madichetty Sreenivasulu, M. Sridevi, A survey on event detection methods on various social media, in: Recent Findings in Intelligent Computing Techniques, Springer, 2018, pp. 87–93.

[17] Madichetty Sreenivasulu, M. Sridevi, Comparative study of statistical features to detect the target event during disaster, Big Data Min. Analyt. 3 (2) (2020) 121–130.

[18] Kumar Ravi, Vadlamani Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowl.-Based Syst. 89 (2015) 14–46.

[19] Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, Omer Berat Sezer, Deep learning for financial applications: A survey, Appl. Soft Comput. (2020) 106384.

[20] Kranthi Kumar Lella, P.J.A. Alphonse, A literature review on COVID-19 disease diagnosis from respiratory sound data, AIMS Bioeng. 8 (2) (2021) 140–153.

[21] Sreenivasulu Madichetty, M. Sridevi, A stacked convolutional neural network for detecting the resource tweets during a disaster, Multimedia Tools Appl. (2020) 1–23.

[22] Kranthi Kumar Lella, P.J.A. Alphonse, Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice, AIMS Publ. Health 8 (2) (2021) 240–264.

[23] Soonil Kwon, et al., Att-net: Enhanced emotion recognition system using lightweight self-attention module, Appl. Soft Comput. 102 (2021) 107101.

[24] Soonil Kwon, et al., MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach, Expert Syst. Appl. 167 (2021) 114177.

[25] Muhammad Sajjad, Soonil Kwon, et al., Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM, IEEE Access 8 (2020) 79861–79875.

[26] Soonil Kwon, et al., CLSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network, Mathematics 8 (12) (2020) 2133.

[27] Usman Naseem, Imran Razzak, Matloob Khushi, Peter W. Eklund, Jinman Kim, Covidsenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis, IEEE Trans. Comput. Soc. Syst. (2021).

[28] Klaifer Garcia, Lilian Berton, Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA, Appl. Soft Comput. 101 (2021) 107057.

[29] Diaa Salama Abdelminaam, Fatma Helmy Ismail, Mohamed Taha, Ahmed Taha, Essam H. Houssein, Ayman Nabil, Coaid-DEEP: An optimized intelligent framework for automated detecting COVID-19 misleading information on Twitter, IEEE Access 9 (2021) 27840–27867.

[30] Debanjan Konar, Bijaya K. Panigrahi, Siddhartha Bhattacharyya, Nilanjan Dey, Richard Jiang, Auto-diagnosis of COVID-19 using lung CT images with semi-supervised shallow learning network, IEEE Access 9 (2021) 28716–28728.

[31] Koyel Chakraborty, Surbhi Bhatia, Siddhartha Bhattacharyya, Jan Platos, Rajib Bag, Aboul Ella Hassanien, Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media, Appl. Soft Comput. 97 (2020) 106754.

[32] H. Jelodar, Y. Wang, R. Orji, S. Huang, Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach, IEEE J. Biomed. Health Inf. 24 (10) (2020) 2733–2742, http://dx.doi.org/10.1109/JBHI.2020.3001216.

[33] Lorenzo Carnevale, Antonio Celesti, Giacomo Fiumara, Antonino Galletta, Massimo Villari, Investigating classification supervised learning approaches for the identification of critical patients' posts in a healthcare social network, Appl. Soft Comput. 90 (2020) 106155.

[34] Mabrook S. Al-Rakhami, Atif M. Al-Amri, LieS kill, facts save: Detecting COVID-19 misinformation in Twitter, IEEE Access 8 (2020) 155961–155970.

[35] Pranav Kairon, Siddhartha Bhattacharyya, COVID-19 outbreak prediction using quantum neural networks, in: Intelligence Enabled Research, Springer, 2020, pp. 113–123.

[36] Yipeng Zhang, Hanjia Lyu, Yubao Liu, Xiyang Zhang, Yu Wang, Jiebo Luo, Monitoring depression trend on Twitter during the COVID-19 pandemic, 2020, arXiv preprint arXiv:2007.00228.

[37] Kakia Chatsiou, Text classification of COVID-19 press briefings using BERT and convolutional neural networks, 2020, arXiv preprint arXiv:2010.10267.

[38] Eleonora D'Andrea, Pietro Ducange, Alessio Bechini, Alessandro Renda, Francesco Marcelloni, Monitoring the public opinion about the vaccination topic from tweets analysis, Expert Syst. Appl. 116 (2019) 209–226.

[39] Lucy Lu Wang, Kyle Lo, Text mining approaches for dealing with the rapidly expanding literature on COVID-19, Brief. Bioinform. 22 (2) (2021) 781–799.

[40] Priyanshu Kumar, Aadarsh Singh, NutCracker at WNUT-2020 task 2: robustly identifying informative covid-19 tweets using ensembling and adversarial training, in: Proceedings of the Sixth Workshop on Noisy User-Generated Text (W-NUT 2020), 2020, pp. 404–408.

[41] Anders Giovanni Møller, Rob Van Der Goot, Barbara Plank, NLP North at WNUT-2020 Task 2: Pre-training versus ensembling for detection of informative COVID-19 english tweets, in: Proceedings of the Sixth Workshop on Noisy User-Generated Text (W-NUT 2020), 2020, pp. 331–336.

[42] Nickil Maveli, Edinburghnlp at WNUT-2020 task 2: Leveraging transformers with generalized augmentation for identifying informativeness in COVID-19 tweets, 2020, arXiv preprint arXiv:2009.06375.

[43] Linh Doan Bao, Viet Anh Nguyen, Quang Pham Huu, SunBear at WNUT-2020 task 2: Improving BERT-based noisy text classification with knowledge of the data domain, in: Proceedings of the Sixth Workshop on Noisy User-Generated Text (W-NUT 2020), 2020, pp. 485–490.

[44] Anh Tuan Nguyen, TATL At W-NUT 2020 task 2: A transformer-based baseline system for identification of informative COVID-19 english tweets, 2020, arXiv preprint arXiv:2008.12854.

[45] Yandrapati Prakash Babu, Rajagopal Eswari, Cia_Nitt at WNUT-2020 task 2: Classification of COVID-19 tweets using pre-trained language models, 2020, arXiv preprint arXiv:2009.05782.

[46] Dat Quoc Nguyen, Thanh Vu, Afshin Rahimi, Mai Hoang Dao, Linh The Nguyen, Long Doan, WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets, in: Proceedings of the 6th Workshop on Noisy User-Generated Text, 2020.

[47] Martin Müller, Marcel Salathé, Per E. Kummervold, COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter, 2020, arXiv preprint arXiv:2005.07503.

[48] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[49] Sreenivasulu Madichetty, M. Sridevi, Identification of medical resource tweets using majority voting-based ensemble during disaster, Soc. Netw. Anal. Min. 10 (1) (2020) 1–18.

[50] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al., Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[51] Arun S. Maiya, Ktrain: A low-code library for augmented machine learning, 2020, arXiv:2004.10703.

[52] Ilya Loshchilov, Frank Hutter, Decoupled weight decay regularization, 2017, arXiv preprint arXiv:1711.05101.

[53] Avishek Garain, English language tweets dataset for COVID-19, IEEE Dataport, 2020, https://doi.org/10.21227/2zg4-yx02.