Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset

K. Nimmi [a,*], B. Janet [a], A. Kalai Selvan [b], N. Sivakumaran [c]

[a] *Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India*
[b] *Centre for Development of Advanced Computing (C-DAC), Thiruvananthapuram, India*
[c] *Department of Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli, India*

ARTICLE INFO

ABSTRACT

The COVID-19 precautions, lockdown, and quarantine implemented throughout the epidemic resulted in a worldwide economic disaster. People are facing unprecedented levels of intense threat, necessitating professional, systematic psychiatric intervention and assistance. New psychological services must be established as quickly as possible to support the mental healthcare needs of people in this pandemic condition. This study examines the contents of calls landed in the emergency response support system (ERSS) during the pandemic. Furthermore, a combined analysis of Twitter patterns connected to emergency services could be valuable in assisting people in this pandemic crisis and understanding and supporting people's emotions. The proposed Average Voting Ensemble Deep Learning model (AVEDL Model) is based on the Average Voting technique. The AVEDL Model is utilized to classify emotion based on COVID-19 associated emergency response support system calls (transcribed) along with tweets. Pre-trained transformer-based models BERT, DistilBERT, and RoBERTa are combined to build the AVEDL Model, which achieves the best results. The AVEDL Model is trained and tested for emotion detection using the COVID-19 labeled tweets and call content of the emergency response support system. This is the first deep learning ensemble model using COVID-19 emotion analysis to the best of our knowledge. The AVEDL Model outperforms standard deep learning and machine learning models by attaining an accuracy of 86.46 percent and Macro-average F1-score of 85.20 percent.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The coronavirus 2019 (COVID-19) pandemic had both direct and indirect impact on the healthcare systems around the world. SARS CoV 2 information caused immense fear in the society that was totally unprepared to tackle the outbreak. The rapid rise in the reported cases, the wide dissemination of information as well as the measures taken to combat the outbreak generated panic among the public, whose normal routined lives were turned topsy turvy. Majority of the countries have been placed under tight quarantine. Cleanliness, wearing a mask, a safe distance, not touching the nose, eyes, or mouth, and other precautions are suggested. The Federal Emergency Management Agency (FEMA) has identified social media as an important component in future disaster management [1]. During the pandemic the social media communications are linked to the COVID-19 illness, and the material includes epidemic symptoms, disease outbreak communities, and other medical services.

The adverse effects of the Covid-19 epidemic have resulted in a global health crisis. As has been mentioned earlier the lockdown implemented by the governments world wide altered practically every aspect of life. It has to be duly acknowledged that the pandemic affects everyone — financially, physically, and mentally. Hence it is essential to examine and comprehend the emotional responses throughout the crisis. Negative emotional responses like anger and fear may cause long-term socioeconomic damage. The Pandemic has created psychological impact on both health workers and common man. Assessing the mental state can aid in creating psychological programs as quickly as possible to manage mental health care of the general public and front line employees. It has been noted that there was a drastic increase in the number of calls landing in emergency response services during the pandemic. In India, the emergency response service is structured through the 112 (Indian emergency number) system, which reroutes calls to the police, medical department, or fire department based on need. For example, when an emergency call is received, calls requiring medical aid are directed to the appropriate control room for hospital assistance, based on the caller's location, to evaluate the patient and identify the best course of action.
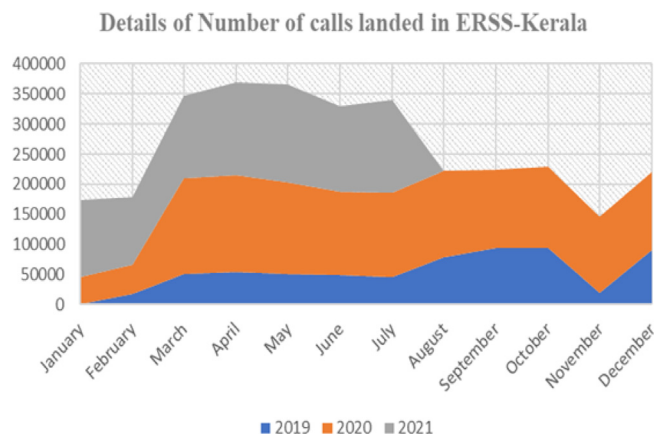
Fig. 1. Statistics of calls received by ERSS-Kerala from February 2019 to July 2021.

In March 2020, Kerala's ERSS received 1,59,237 calls, compared to 51,219 in March 2019. In April 2020, 1,62,108 calls landed, compared to 53,260 in April 2019. Due to the current pandemic situation, the number of calls made by users increased drastically. Fig. 1 depicts the specifics of calls received by ERSS-Kerala from February 2019 to July 2021. Fig. 1 shows that there has been a massive rise in the quantity of calls. Most of the calls during this period were inquiry and panic calls related to Covid-19. The worldwide economic crisis occurred as a result of the COVID-19 lockdown and other safeguards taken during the epidemic. People are facing unprecedented levels of intense threat, necessitating professional, systematic psychiatric intervention and assistance. New psychological services must be introduced as soon as possible to meet the mental healthcare needs of people in this COVID-19 pandemic situation.

The purpose of this study is to examine call content and classify calls based on emotion, a combined analysis of tweets along with emergency services call transcribed dataset could be valuable in assisting people in this pandemic crisis, as well as understanding and supporting people's emotions, and there by halting the propagation of negativity and spreading optimism.

Understanding the caller's feelings and talking to a person with emotional issues with empathy and a positive mindset might help to alleviate their pain, especially when talking to older individuals who are isolated due to the pandemic lockdown. During this pandemic many people have relinquished their life due to emotional breakdown. The suicide rate related to COVID-19 was 16.7 per 100,000 individuals, which is higher than that of the national norm of 10.2 per 100,000 individuals, [2]. This condition could have been changed if the government and peer groups provided sufficient moral and emotional support during crises. The motivation of the work: (1) Raise awareness in society that the extensive usage of social media during a pandemic should be curtailed to avoid panic and terror caused by the widespread propagation of negative psychology among social media users.

(2) The relevance of emotion recognition in ERSS contact centers is to assist callers experiencing emotional breakdowns, particularly during a pandemic where the entire world is experiencing extreme fear and tension.

Most NLP academics now focus on text classification in social media using both standard and advanced Artificial Intelligence methods. It is hard to assess people's emotions during a health emergency like covid since the crisis is unfamiliar, and individuals need time to acclimatize to the circumstance. As a result, there is a chance that more emotional breakdowns will occur throughout the globe.

The novelty lies in proposing an ensemble model called AVEDL Model for assessing the emotion exhibited in the ERSS call

transcribe dataset and tweets during Covid-19. Using calls from the Emergency Response Support System (transcribed), a labeled dataset based on emotion during Covid 19 was created. For labeling, the authors considered five possible emotion categories: Anticipation, sadness, anger, no specific emotion, and fear. This research has contributed to society by demonstrating the need for emotion analysis in ERSS call transcripts with facts demonstrating the fear and panic situation created due to the social media misuse.

Emotion detection on emergency call, can aid in monitoring and reacting progressively in the event of a crisis, hence aiding in the understanding of people's behavioral patterns. The difficulties encountered during crisis management are:

1. How to take advantage of technical breakthroughs and deal with emergencies or disasters in order to maximize resource usage.
2. What contexts should be considered when creating emotion based emergency response systems?

This study emphasizes the importance of understanding people's emotions to assist individuals in need, particularly during crises, since the lockdown and the rapidly spreading rate of the disease caused many emotional breakdowns that could have been avoided with the help of volunteers and NGOs who are willing to offer assistance. Research on the pandemic that has been conducted to detect relevant and irrelevant tweets along with emotion detection found discriminating and detecting a certain category difficult.

The pandemic's unpredictability keeps people on edge and makes them panic about their future leaving them terrified and overwhelmed with a sense of fear of the future. We have reached a point where even a minor fever can cause concern and tension at home. This circumstance can cause our emotions to run wild and exacerbate the problem. People are experiencing fear, anxiety, and tension due to the increasing number of exponential instances around the world. The global population's mental and physical health has been proven to be directly proportional to this pandemic.

The following are the paper's main contributions:

1. A dataset based on calls in the Emergency Response Support System (transcribed) was created and labeled based on emotion. No study has explicitly studied emotions linked to emergency calls during covid-19 pandemic situations, quarantine, and people's emotional breakdown during an emergency.
2. To extract the emotion during this pandemic, we used multiple state-of-the-art pre-trained models such as DistilBERT, RoBERTa and BERT and compared the results. The model is trained and tested using a combination of tweets and an Emergency Response Support System (transcribed) emotion-based labeled English dataset generated during COVID-19.
3. An ensemble deep learning model (Average voting based ensemble model), a combination of pre-trained transformer models BERT, DistilBERT and RoBERTa for categorizing emotion, is presented in this work . The proposed model outperforms machine learning utilizing the most recent ensemble deep learning models in terms of results. The suggested model has a 86.46% accuracy and Macro-average F1-score of 85.20% for classifying emotions.
4. The model proposed will be helpful to real-time applications such as emotion detection in emergency response support systems, call centers and so on.

The paper is organized as follows. Section 2 contains information on the related work. The framework methodology is described in Section 3. The findings are discussed in Section 4, which compares and contrasts the results obtained using

state-of-the-art classifiers and the results obtained using the proposed AVEDL Model to determine emotion. Section 5 illustrates the application of this work. Finally, Section 6 brings the conclusion outlining future research prospects.

2. Related works

Emotion detection is a subset of sentiment analysis that extracts fine-grained emotions from text, speech [3] or image. Emotions determine human behavior, hence developing a system to recognize emotions automatically is essential. 'Categorical' and 'Dimensional' models [4] can be found in all known emotion models. As in Ekman's model [5], categorical emotion models divide all emotions into a few basic categories: fear, happiness, anger, sadness, disgust, and so on. Dimensional emotion models like Plutchik's [6], classify emotions using various dimensions: arousal, dominance, valence and intensities: intense, moderate, basic of emotion. In the proposed work we are considering 'Categorical' emotional model. There are many emotion detection datasets available from emotion detection, but during the pandemic, tweets to detect emotions could help researchers better understand how individuals reacted to COVID-19. Analyzing Twitter data, and human emotions can be extremely valuable in stock market volatility, predicting crimes [7], and disaster management [8] and election polarity [9]. According to our knowledge, Covid-19 pandemic-related labeled emotion data is scarce though a few machine learning models that automatically classify tweets' emotions based on context are available. Due to the lack of available datasets, most COVID-19 sentiment research is limited to negative, neutral and positive sentiments. In [10] researchers manually labeled 10 000 tweets and classified them into ten different emotions with the help of three annotators. However, none of these studies classifies emotions in crises using an emergency dataset, representing a wide range of emotions with unbalanced data. To address these issues, we began manually annotating COVID-19 related ERSS dataset to 5 different emotion categories (Anticipation, sadness, Anger, No specific emotion and fear).

According to Acheampong et al. [11], regardless of the quantity of text data, detecting emotions from texts has been difficult, mainly owing to lack of speech modulation, facial expressions, and other indications that could help with context and relation extraction. There are various methods for detection of emotion; they are lexicon-based, keyword-based, machine-learning, Hybrid methods and Deep learning. Emotion detection using 'keyword-based' approaches is done by searching for a word match between terms in text data and emotion keywords [12,13], while 'lexicon-based' methods use an emotion lexicon to distinguish the proper emotion from text data. Prabhakar Kaila et al. [14] used the NRC sentiment lexicon to quantify emotions based on a random sample of COVID-19 tweets. For emotion detection, machine-learning uses two methods: supervised learning and unsupervised learning. 'Hybrid' methods integrate two or more of the above techniques and apply them to text emotion. Yang et al. [15] propose a hybrid approach for classification of emotion that comprises lexicon-keyword detection, emotion cue identification based on machine-learning-based emotion classification utilizing Max Entropy, Nave Bayesian and SVM, and Conditional Random Field. A vote-based approach is used to combine the outcomes of the preceding techniques. Aditya Vijayvergia and Krishan Kumar [16] propose a technique for emotion detection that combines many shallow models to outperform a single large model. Because the models work independently, they can be run in parallel, resulting in a faster execution time and smaller memory. In 00.98 ms per input, this model achieved 86.16 percent accuracy. Deep learning technique [17] includes Long Short Term Memory networks (LSTM) [18], and BiLSTM [19].

A research conducted in Belgium by Cauberghe et al. [20] during the COVID-19 quarantine period found that social media was positively associated with productive coping for teens with anxious feelings. During the COVID-19 pandemic, De Las Heras-Pedrosa et al. in Spain conducted research on social media [21] was able to capture the additional stress on people's mental health. Li et al. [22] in their research found that text data gives more details about people's natural emotional attitudes and expressions, with social media offering a platform for risk communication and emotional interchange to minimize social isolation. Aslam et al. [23]in their research discusses the use of latent-topics and sentiment techniques to COVID-19-related comments on social media. Venigalla et al. [24] during COVID-19, examined real-time messages on Twitter and were able to detect the nation's mood through their analysis. H. Jelodar et al. [25] for COVID-19 sentiment classification used LSTM Recurrent Neural Network and it outperformed other machine-learning methods. A Malte et al. [26] detected Cyber abuse words in languages Hindi and English texts were analyzed using the BERT model. Deepak et al. [27] addressed the issue of Global Filter-based Feature Selection Scheme in text corpus by implementing a soft voting technique, which resulted in a considerable improvement in the performance of classifiers. When compared to machine learning models, the pre-trained transformer based classification models ALBERT, BiLSTM, distilBERT, BERT, XLNet, and RoBERTa demonstrate greater accuracy. Realizing that ensemble models and pretrained transformer models outperform individual models, our study focuses on ensemble of pre-trained transformer model.

3. Framework methodology

Deep learning pertained models are used for emotion detection, the models include DistilBERT, RoBERTa, BERT. The proposed algorithm (AVEDL Model) which is depicted in detail in Fig. 1 is based on DistilBERT, RoBERTa and BERT. There are 4 emotions taken into consideration: Anticipation, Fear, Anger, Sadness and there is a "No specific emotion" label for denoting no specific emotion being identified. The subsections provide further details on the proposed model. Section 3.1 describes details of dataset gathering and Section 3.2 deals with pre-processing procedures. Section 3.3 to Section 3.5, explains the models used and finally in Section 3.6 the proposed AVEDL Model is discussed.

3.1. Data collection

The dataset used as an input to the proposed model is a combination of two dataset namely twitter dataset and ERSS call transcribed dataset. The first dataset is the Twitter dataset that we had gathered by utilizing the tweetid supplied by Gupta et al. [28], which included a labeled dataset for emotion analysis on over 63 million tweets during the pandemic covid 19. To collect data, we have used Twitter's standard search API [29]. The total number of emotions taken into account is 124 740. The first dataset is a twitter dataset that contains 18 900 samples of each emotion. The second dataset is the emergency response support system (ERSS) call recordings from India (Kerala) which were randomly picked and transcribed from January to March 2021 related to covid-19. The transcribed call containing texts connected to "corona", "covid-19", "quarantine", "covid" and "lockdown" are taken into consideration in this work. Each call was annotated based on the text content; Three different persons manually annotated the tweets in order to avoid biases. One of the three annotators has been a PhD scholar in data analysis since 2018. The second and third annotators were postgraduate students in the English department. The final label is determined based on
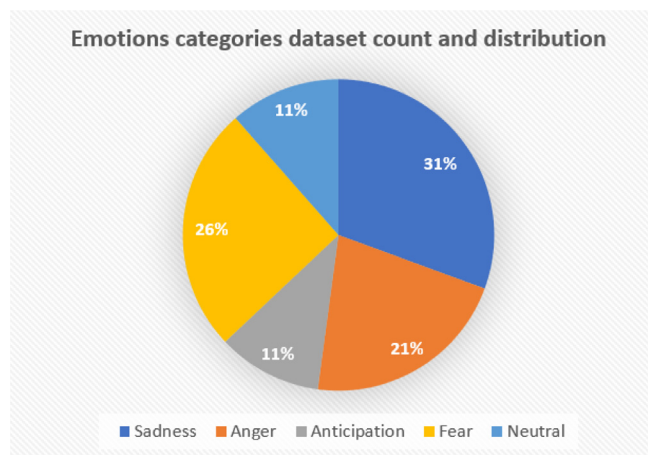
Fig. 2. Emotions categories dataset count and distribution.

Emotion	Class	Count
No_specific_emotion	0	19 581
Anticipation	1	19 935
Fear	2	23 138
Anger	3	19 962
Sadness	4	21 099

the annotators' majority agreement on labels. With the help of these experts, a total of 30 240 call transcribed ERSS datasets were labeled. From 30 240, a total of 21 025 was used for ERSS data testing (details are provided in Section 5), and the remainder is mixed with the first dataset (tweets) to form the dataset used to train the model in this paper. The details of combined dataset is provided in the Table 1.

The weekly ratio of emotion types of ERSS is shown in Fig. 2. From Fig. 2, We can infer that the most common emotions detected in the dataset are Anticipation, sadness, neutral and fear. While those five emotions are represented in 89% of the dataset, there are 11% of dataset that are either neutral or can be classified in other emotion categories. The distribution is obtained by dividing the total number of identified dataset for an emotion by the total number of dataset in that time period. There are slight changes in emotions from week to week.

Reason for combining the dataset: The number of Corona-related calls received by ERSS has skyrocketed. The calls landing in the emergency response and support system during the pandemic were primarily due to the caller's misconceptions through social media worldwide. When we analyzed ERSS calls, we discovered that the influence of social media posts caused the majority of calls; in fact, some calls stated that the reason for calling was due to the knowledge they gained from social media. For instance, "My kid suffers from heavy fever and loss of taste; we doubt it is covid related, please send health care workers or doctors immediately to my location". Many calls similar to this landed in ERSS which are just panic calls. We decided to treat tweets as a source of information because we felt labeled tweets can be a rich source of data for processing since the emotion of a tweet and the emotion of a caller in the ERSS are similar or relatable and the content is related to pandemic. The maximum number of characters in a tweet is 280 and that of ERSS calls transcribed varies from 117 to 450 characters.

The data acquired from ERSS contain several interesting observations. Even the appearance of a headache causes panic and is regarded as a symptom of covid. Individuals have had emotional imbalances as a result of sudden lockdowns, job losses, and online classes. The stress caused due to these factors should be regarded as a serious social issue and measures should be taken to provide psychosocial and emotional support. Home quarantine, where people are requested to stay in a room in some cases (solitary confinement) and is allowed to leave that room only after 14 days, is associated with more significant emotional fluctuations, whilst the rest of the family is free to roam around. Providing emotional support, can assist these individuals in restoring mental peace. Mental well-being is just as essential as physical well-being.

3.2. Pre-processing

Tweets downloaded from the Twitter API include emoticons, emojis, hashtags, user mentions, standard web constructs like user mentions, Uniform Resource Locators (URLs), email addresses and other noise. As a result, they are noisy and may conceal valuable information. Emoticons are made up of short strings of characters (letters, symbols, or numerals) that depict various facial emotions and are used to express emotions on social media. These emoticons need to be replaced using the term that expresses the same emotion. To search and substitute emoticons, we utilized a customized Python program. In 1997, a standardized collection of tiny graphical symbols that portray a variety of objects, like sad faces, happy faces, were used as emojis. These have been very popular in social media and can reflect universal emotional expressions. The method for transforming emojis suggested here uses a brief sentence to make their meaning apparent. A python library module is used to search for and substitute emojis. Hashtags are phrases that begin with the hash symbol "#" and end with alphanumeric characters without spaces. The removal of hashtags is done.

As a language text processor, the Natural Language Toolkit library (NLTK) [30] is used. Pre-processing approach translates the dataset into sentences that are cleaner and more informative. Due to the absence of user mentions, hyperlinks, or hash symbols in the call transcribed ERSS dataset, these steps were omitted during pre-processing. Both the tweet dataset and ERSS dataset were combined after pre-processing.

The dataset is subjected to the following pre-processing step.

1. The dataset was changed to lowercase.
2. Removing the special characters, Punctuation symbols, numbers, unknown words.
3. Remove any unwanted spaces, tabs, or newlines.
4. Replace the emojis with words that are related.
5. Removing the hyperlinks (URLs) and user mentions listed in the tweets

3.3. BERT

The BERT encoder [31] is a multi-layer encoder. The BERT-base-uncased model was utilized, which consisted of a transformer block with twelve layers, each with 768 hidden layers and twelve head self-attention layers , and a total of 110 million parameters. At a time, a single sentence was fed into the model. The BERT tokenizer library was used to break the input phrases into tokens and map them to their indices, which are indicated as input ids. Every sentence has the (classification token)[CLS], a unique token that appears at the start of the first sentence and (separate segment token) [SEP] at the end of it. Each block goes via a Self Attention layer before being fed into a feed-forward neural network. It is then forwarded to the next encoder. Finally, each point will produce a hidden size vector (768 in BERT Base). This is an example of word embedding.

Table 2
System requirements.

Integrated Development Environment (IDE)	Jupyter notebook
Processor	Tesla K80
RAM	25GB
Model name	Intel(R) Xeon(R)
Programming space allocated	160 GB
Programming Language	Python (version 3.6.5)
Packages	Numpy, nltk, Pandas, and scikit-learn

3.4. DistilBERT

DistilBERT [32] is a compact, fast, inexpensive, and light transformer model that has been trained using a BERT base. It has 40% less parameters than BERT-base-uncased, and it runs 60% quicker while keeping over 95% of BERT's performance. DistilBERT employs a process known as distillation to simulate Google's BERT, which entails replacing a huge neural network with a smaller one. DistilBERT-uncased model is used in this experiment for emotion detection. Six transformer levels, 768 hidden layers, and twelve attention heads were included. The input texts were tokenized and converted into input ids before being padded and sent into the DistilBERT model for the multi classification job.

3.5. RoBERTa

Robustly Optimized BERT Pre-Training Approach [33] is a re-training of BERT with enhanced training technique, 1000 percent more data, and 1000 percent more compute power, introduced at Facebook. For pre-training, RoBERTa uses 160 GB of text. The Google BERT language masking approach is used to create RoBERTa. When compared to BERT, RoBERTa has a 2–20 percent higher accuracy. Both of these methods use a transformer to learn a representation of a language. The experiment employed the RoBERTa-base model, consisting of 768 hidden layers, twelve transformer layers, and twelve attention heads, with 125 million parameters. The input texts were encoded into tokens and assigned as input ids using the RoBERTa tokenizer. To eliminate variations per row, these ids were padded to a constant length. The features collected from these tokens were then used to classify sentence pairs.

The models was trained using various hyperparameter combinations, including batch size, epochs, learning rate and maximum length for the emotion dataset. The details of the parameter chosen is provided in Table 3. Various performance metrics are used to evaluate the results obtained for each combination like precision, F1-score, recall and accuracy to obtain the best results. For this experiment models were trained on the emotion dataset with batch sizes: 16, 24, and 32, maximum length: 50 and 125, learning rate: $5e-5$, $2e-5$, $1e-5$, $3e-5$ and $4e-5$.

3.6. Proposed Average Voting Ensemble Deep Learning model (AVEDL Model)

A Vote is a meta-algorithm that uses participant classifiers to accomplish the decision process [34]. The proposed AVEDL Model employs soft voting [35], in which each classifier provides a probability value that a single data point belongs to a specified target class. A weighted summing of the predictions is made based on the relevance of the classifier. Then voting is done, and the target label with the highest weighted probability sum will be the winner. Here all the models have equal weights. The AVEDL Model integrates three deep learning classifiers in this multiclass classification problem: DistilBERT, RoBERTa, and BERT. The AVEDL model, sums and averages the predicted probability

vectors for each model. The class with the highest average probability value is declared the winner and returned as the predicted class. Here the average probability of each category is utilized as the voting score. The soft voting ensemble model produces the best prediction during this procedure. The accuracy of the ensemble model is determined by the batch size of the individual models, the learning rate, and the optimizer. The block diagram of AVEDL Model is provided in Fig. 3. Combining different classifiers is a better way to change the model's accuracy. The algorithm 1 provides the details of emotion detection based on the ensemble model. The average voting based ensemble model offers an improvement in model performance. Various classifiers perform well on other data points in the combined ERSS call transcribe dataset, and some classifiers accurately identify it, while others may misclassify it. The model's correctness is extremely difficult to predict. However, the model's performance can be improved by incorporating the voting technique. Utilizing the voting process to integrate ensemble models improves the model significantly.

Algorithm 1 Emotion detection based on COVID-19 based on AVEDL Model.

0: Emotion class, no specific emotion.
1: Emotion class Anticipation.
2: Emotion class Fear.
3: Emotion class Anger.
4: Emotion class Sadness.
BERT(): An array of predicted probability values for each class obtained from BERT model.
RoBERTa(): An array of predicted probability for each class obtained from RoBERTa model.
DistilBERT(): An array of probability for each class obtained from DistilBERT model.
Input : X_test, BERT(), RoBERTa(), DistilBERT()
Output:Final_prediction (0, 1, 2, 3, 4)
Ensemble classifiers: BERT(), RoBERTa(), DistilBERT()
Begin
for $i = 1$ to $size(X\_test)$ **do**
$Prediction = Average(BERT(i), RoBERTa(i), DistilBERT(i))$
$Final_Prediction = $ Maximum $(Prediction)$
end for
End

---

4. Experimental setup and analysis of results

The details related to the system requirements used to train the model are provided in the Table 2. The whole dataset is divided into the test, train and validation set using Scikit-learn library [36]. The splitting is done in different ratios: 85:15 and 75:25 and their results were analyzed. The Huggingface library [37] was used to carry out the implementation in Python. The "AdamW" optimizer is used. Fine-tuning the models for various epochs: 8, 16 and 32 epochs was done using and the optimum checkpoint was chosen depending on the validation set's performance. The choice of a classifier is crucial to the model's prediction. The multiple classifiers were fine-tuned using various hyper parameters before picking the best classifier. The following classifiers are used to carry out multiple tests.

1. The Pre-trained transformer model RoBERTa.
2. The BERT model.
3. The DistilBERT model.
4. The Average Voting Ensemble Deep Learning model (AVEDL Model).

On the dataset collected during COVID-19, the model is trained with several hyper parameters such as batch size, learning rate,
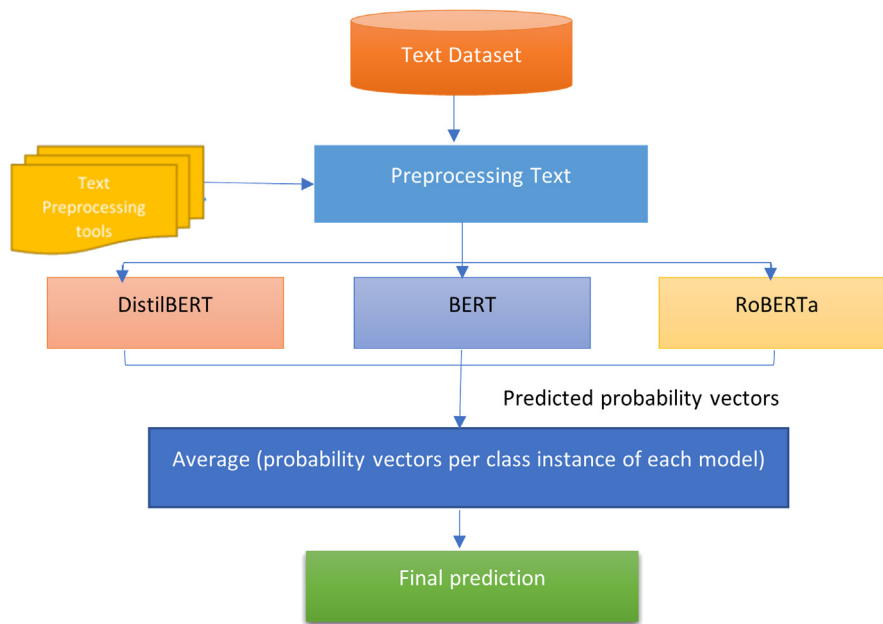
Fig. 3. Overview of Proposed AVEDL model.

Table 3
Hyper-parameters tuning.

Sl no	Model / Hyper-parameters	DistilBERT	BERT	RoBERTa
1	Pre-trained model	DistilBERT-base-uncased	Bert-base-uncased	Roberta-base
2	Learning rate	1e−5, 5e−5, 3e−5, 4e−5, 2e−5	1e−5, 5e−5, 3e−5, 4e−5, 2e−5	1e−5, 5e−5, 3e−5, 4e−5, 2e−5
3	Activation	Softmax	Softmax	Softmax
4	Batch size	8, 16, 32	8, 16, 32	8, 16, 32
5	Number of epochs trained	10,5	10,5	10
6	Maximum sequence length	125	125	125
7	Dropout	0.2	0.2	0.2
8	Hidden size	768	768	768
9	Optimizer	AdamW, Adam	AdamW, Adam	AdamW, Adam

Table 4
Wrongly classified sentence using BERT model.

Sl.no.	Text	Actual class	Predicted class
1	Sunil calling to inform that an migrant labor is having fever need immediate medical aid need ambulance shift hospital loction xxxx do something fast	Fear	Anger
2	During covid there is no geographic identification for containment zones but we wanted to serve people there so what all things are to be done we are not afraid of anything	Anger	No specific emotion
3	Know you would bounce back soon and give it back to the wuhan virus get well soon	No specific emotion	Anticipation
4	Resister complainant regarding auto radhika taking vegetables along persons without permission from police in lockdown time location xxxx	Sadness	No specific emotion

epochs, and optimizer. The findings are given in Tables 5, 8. Table 10 and are analyzed using several performance metric parameters such as recall, precision, accuracy and F1-score.

4.1. Performance evaluation metrics

The models performance such as Precision, Macro Precision, F1-score, Macro F1-score, Recall, Macro Recall and Accuracy are evaluated. A confusion matrix is represented using a table that shows how well a classifier performs on a set of test data. True positives (T_Pos) are classification in which the label "fear" was predicted as "fear" and belongs to class "fear". True negatives (T_Neg) are situations in which the label was predicted as not fear, and they belong to the class, not fear. False positives (F_Pos) are classification in which the label was predicted as "fear", but their actual label is "not fear" ("Type I error"). False negatives (F_Neg) are classification in which the label was predicted as the class "not fear", yet they belong to class "fear" ("Type II error"). Confusion matrix provides the details of (T_Pos), (T_Neg), (F_Pos), and (F_Neg).

Eqs. (1) to (7) are used to calculate precision, recall, F1-score, accuracy, F1 micro average score etc.

$$Precision(P) = \left(\frac{T\_Pos}{T\_Pos + F\_Pos} \right) \tag{1}$$

$$Recall(R) = \left(\frac{T\_Pos}{T\_Pos + F\_Neg} \right) \tag{2}$$

Table 5
Performance evaluation metrics of BERT.

Batch size	Learning rate	Macro precision	Macro recall	Macro F1-score	Accuracy	Time
32	1e−5	82.72%	82.70%	82.69%	84.11%	
	2e−5	83.71%	83.31%	83.41%	84.82%	
	3e−5	82.74%	82.01%	82.14%	83.77%	1229 s
	4e−5	82.79%	82.24%	82.20%	83.78%	
	5e−5	82.78%	82.39%	82.54%	83.91%	
16	**1e−5**	**83.92%**	**83.36%**	**83.55%**	**84.91%**	
	2e−5	82.60%	82.29%	82.43%	83.87%	
	3e−5	82.73%	82.22%	82.38%	83.79%	1345 s
	4e−5	82.66%	82.09%	82.27%	83.75%	
	5e−5	82.74%	82.01%	82.14%	83.77%	
8	1e−5	82.92%	82.58%	82.67%	84.16%	
	2e−5	81.89%	81.35%	81.51%	81.35%	
	3e−5	81.92%	81.59%	81.69%	81.59%	1642 s
	4e−5	81.35%	80.66%	80.87%	80.66%	
	5e−5	83.47%	83.34%	83.29%	84.69%	

Table 6
Detailed performance evaluation metrics of BERT showing best results.

Class	Precision	Recall	F1-score
0	78.45%	79.56%	79.00%
1	85.14%	75.78%	80.19%
2	95.53%	94.93%	95.23%
3	75.30%	83.09%	79.00%
4	85.21%	83.45%	84.32%

$$F1 - Score = 2 * \left(\frac{P \ * \ R}{P \ +_C R} \right) \qquad (3)$$

$$Macro\ F1 - Score = \frac{1}{|C|} \sum_{i=1} F1 - Score_i \qquad (4)$$

$$Macro\ Recall = \frac{1}{|C|} \sum_{i=1}^{C} R_i \qquad (5)$$

$$Macro\ Precision = \frac{1}{|C|} \sum_{i=1}^{C} P_i \qquad (6)$$

$$Accuracy \ = \left(\frac{T\_Pos + T\_Neg}{T\_Pos + \ F\_Pos + T\_Neg + F\_Neg} \right) \qquad (7)$$

where, the letter C stands for the number of label categories.

4.2. Results analysis

The BERT, DistilBERT, RoBERTa models results may vary from dataset to dataset. In Table 5, the performance metrics of each instance of Bert model is provided. From Table 5 it can be understood that the batch size: 16 and 1e−5 is the learning rate, which outperforms other tested parameters. The best results were obtained using the learning rate of 1e−5, maximum length of 125, epoch of 10 and batch size of 16. The performance obtained were Macro precision: 83.92%, Macro recall: 83.36%, Macro F1-Score: 83.55% and accuracy: 84.91%. The Table 6 contains the details of the best performing Bert model prediction for each class. According to the performance metrics in the Table 6, class 2 was predicted with the highest F1-score of 95.23%, precision of 95.53%, recall of 94.93% and accuracy of 95.23%.

Performance evaluation metrics of DistilBERT utilizing varieties of hyper parameter combinations is provided in Table 8. When a batch size of 16 along with learning rate of 2e−5 is used, best results were obtained. The best performance obtained is, Macro recall: 83.39%, Macro precision : 83.84%, Macro F1-Score: 83.52% and accuracy: 84.86%. We evaluate DistilBERT's performance metrics as specified in Table 9. We can see that class 2 was predicted with the highest F1-score 95.13%. Class 0 had the
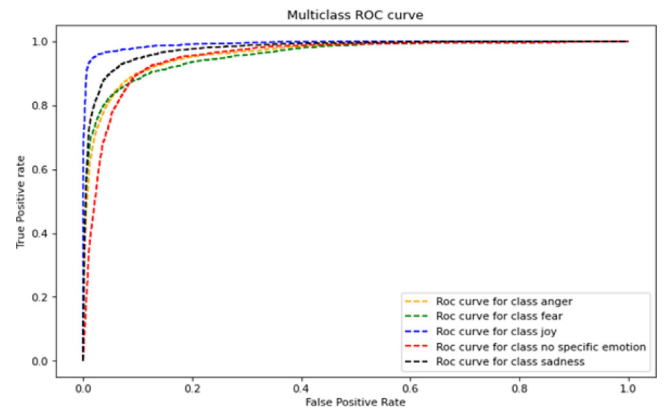

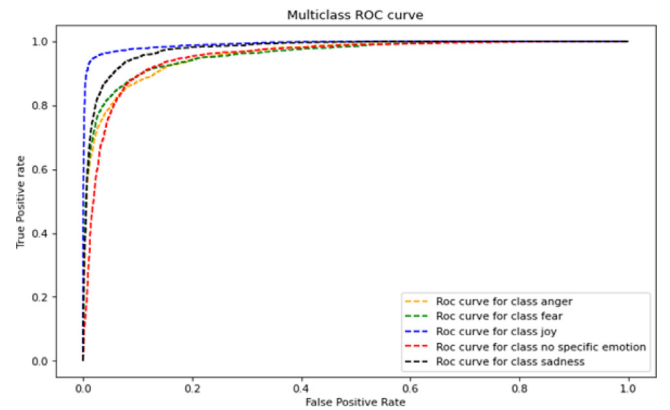**Fig. 4.** BERT ROC curves for emotion classification.

Fig. 5. DistilBERT ROC curves for emotion classification.

lowest F1-score : 77.94%, while class 3 had F1-score of 77.31%. Class 4 and class 1 have F1-scores of 83.93

The performance evaluation metrics of RoBERTa is provided in table Table 10. The best results were obtained when a batch size of 8 along with learning rate of 2e−5 and epoch of 10 is used. The best performance obtained is, Macro recall: 84.19%, Macro precision : 84.87%, Macro F1-Score: 84.47% and accuracy: 85.69%. The performance metrics of best performing RoBERTa model is provided in Table 12, when we analyze the details we can see that class 2 was predicted with the highest F1-score 95.13%, precision is 94.90% and recall is 95.35%.

The receiver operating characteristic (ROC) curves for emotion classifications, using annotators' labels as the ground truth (actual

Table 7

Misclassified sentence using DistilBERT model.

Sl.no.	Text	Actual class	Predicted class
1	There is always a mismatch in their numbers negative reports are sent later focus is on what is posted	Anger	Sadness
2	Please look department of posts at kolkata offices specially head post offices and sub post offices too no precaution about corona	Sadness	Anger
3	What ever may be the religion of yours if you are covid positive or suspected and you are hiding from people	No specific emotion	Fear
4	The may have been covid positive but the cause of death covid or due to some other reason people are spreading rumors	Fear	Sadness
5	caa corona applied appeared	Anticipation	No specific emotion

Table 8

Performance evaluation metrics of DistilBERT.

Batch size	Learning rate	Macro precision	Macro recall	Macro F1-score	Accuracy	Time
	1e−5	82.60%	82.16%	82.32%	83.78%	
	2e−5	83.13%	82.95%	82.96%	84.29%	
32	3e−5	79.61%	79.58%	79.55%	81.30%	619 s
	4e−5	80.05%	79.52%	79.70%	81.37%	
	5e−5	80.22%	79.70%	79.90%	81.56%	
	1e−5	83.62%	83.35%	83.41%	**84.86%**	
	2e−5	**83.84%**	**83.39%**	**83.52%**	**84.86%**	
16	3e−5	79.61%	79.58%	79.55%	81.30%	691 s
	4e−5	81.86%	81.66%	81.64%	83.31%	
	5e−5	82.27%	81.92%	81.90%	83.36%	
	1e−5	82.08%	82.66%	81.81%	83.34%	
	2e−5	83.30%	82.83%	82.92%	84.24%	
8	3e−5	81.94%	81.88%	81.86%	83.28%	572 s
	4e−5	81.12%	80.70%	80.88%	82.44%	
	5e−5	80.13%	79.59%	79.77%	81.36%	

Table 9

Detailed performance evaluation metrics of DistilBERT showing best result.

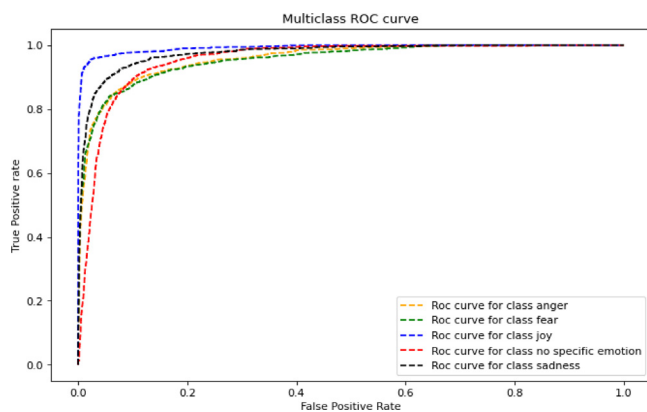| Class	Precision	Recall	F1-score
0	82.78%	73.63%	77.94%
1	79.67%	80.98%	80.32%
2	95.59%	94.66%	95.13%
3	77.31%	82.38%	79.76%
4	82.78%	85.11%	83.93%

Fig. 6. ROC curves for emotion classifications using RoBERTa.

class). The better the classifier, the closer the curve replicates the left vertical and top horizontal axis. The dashed line depicts the emotions' classification by the classifier. The ROC curve of best performing parameters of BERT, DistilBERT and RoBERTa models are provided in Figs. 4–6.

The AVEDL Model performance metrics is shown in Table 14. With reference to Table 14, it is clear that our AVEDL model obtained achieved recall: 85.07%, precision : 85.49%, accuracy: 86.46%, F1-score: 85.20% when compared to existing models. This indicates that the model was successful in distinguishing specific emotions from a group of emotions. The Table 13 provides the details of AVEDL model prediction based on each class. An analysis of performance metrics on this table show that class 2 was predicted with the highest F1-score 95.94%. When we compare the F1 score of the AVEDL Model to the top performing performance metrics of the other models presented in Table 12, Table 9, and Table 6, we can observe that the AVEDL Model has a higher F1 Score in all classes.

The testing accuracy of DistilBERT ranges from 81.30% to 84.86% and the F1-score varies from 79.55% to 83.52%. The testing accuracy of BERT ranges from 80.66% to 84.91% and the F1-score varies from 80.87% to 83.55%. The testing accuracy of RoBERT ranges from 81.09% to 85.69% and the F1-score varies from 81.21% to 84.47%.

The Table 4 refers to the BERT model's misclassified texts. The model predicted sentence 4, which belongs to the class; " No specific emotion" was predicted, but "Sadness" was the actual label, indicating that the sentence was incorrectly labeled. From the analysis we can infer that there are cases of wrong labeling. The author assumes that the presence of implicit sentiments [38] is the cause of incorrect prediction. There is no specific sentiment term in this sentence, but it implies subjective sentiment, hence there is a labeling conflict. The details of the DistilBERT model's incorrectly classified text are presented in Table 7 where the actual and predicted classes are displayed to find out the ground truth. The classification Table 11 provide the details of classification made by RoBERTa model. Sentence 1 is a metaphor sentence, which describes one thing while contrasting it with another. A metaphor contrasts two things. The intensity of these kinds of metaphor sentences is commonly predicted wrongly. Corona is

Table 10
Performance evaluation metrics of RoBERTa.

Batch size	Learning rate	Macro precision	Macro recall	Macro F1-score	Accuracy	Time
32	1e−5	83.05%	82.54%	82.74%	84.20%	
	2e−5	83.58%	83.18%	83.33%	84.69%	
	3e−5	82.60%	82.29%	82.43%	83.87%	1256 s
	4e−5	81.34%	81.02%	81.13%	82.59%	
	5e−5	82.05%	81.98%	81.98%	83.49%	
16	1e−5	83.06%	82.96%	82.96%	82.96%	
	2e−5	82.74%	82.35%	82.47%	82.35%	
	3e−5	81.94%	81.65%	81.73%	81.65%	1437 s
	4e−5	83.74%	83.26%	83.44%	84.83%	
	5e−5	81.59%	81.09%	81.21%	81.09%	
8	1e−5	83.22%	82.88%	82.98%	82.88%	
	2e−5	**84.87%**	**84.19%**	**84.47%**	**85.69%**	
	3e−5	83.15%	82.67%	82.88%	84.33%	1428 s
	4e−5	82.84%	82.38%	82.57%	84.06%	
	5e−5	81.64%	81.41%	81.45%	83.00%	

Table 11
The RoBERTa model incorrectly predicted sentences.

Sl.no.	Text	Actual class	Predicted class
1	Before covid also more people died of diabetes cancer road accidents and malaria Covid is like fire	Sadness	Fear
2	Crowd infront of hotel xxxx covid will spread fast if no action taken take some action	Anger	Fear
3	Sujatha sheeba wife of subramanian staying lonely need adequate food not having enough food since last week due to lockdown location xxxx nearby am much worried because of lockdown	Anger	Sadness
4	He may have been covid positive but the cause of death covid or due to some other reason please try to understand	Fear	Sadness

Table 12
Detailed Performance evaluation metrics of RoBERTa showing best result.

Class	Precision	Recall	F1-score
0	82.98%	75.83%	79.25%
1	81.44%	78.04%	79.70%
2	94.90%	95.35%	95.13%
3	76.54%	83.26%	79.76%
4	83.94%	84.68%	84.31%

Table 13
Detailed performance evaluation metrics of AVEDL Model showing best result.

Class	Precision	Recall	F1-score
0	82.79%	78.30%	80.49%
1	85.71%	79.09%	82.27%
2	96.54%	95.34%	95.94%
3	77.23%	85.10%	80.98%
4	85.17%	87.53%	86.33%

Table 14
Performance evaluation metrics of Proposed AVEDL model.

Model	Recall	Precision	Accuracy	F1-score
Proposed AVEDL model	**85.07%**	**85.49%**	**86.46%**	**85.20%**

Table 15
Comparing AVEDL model results with state-of-art models.

Model	Macro Recall	Macro Precision	Accuracy	F1-score
SVM	58.32%	58.33%	58.41%	57.11%
SVM(tf-idf)	69.30%	63.72%	66.22%	65.34%
Random Forest(tf-idf)	57.65%	42.02%	42.35%	47.49%
BERT	83.92%	83.36%	84.91%	83.55%
RoBERTa	84.87%	84.19%	85.69%	84.47%
DistilBERT	83.62%	83.35%	84.86%	83.41%
Proposed AVEDL Model	85.07%	85.49%	86.46%	85.20%

4.3. State-of-the-art approaches compared to the proposed AVEDL model

This section summarizes the State-of-the-art approaches for identifying emotion and compares the findings to the proposed method based on the merged dataset. The machine learning models we looked at for emotion detection include Nave Bayesian combined with term frequency-inverse document frequency (tfidf) [39], Support vector machines (SVM), SVM combined with tfidf, and Random Forest combined with tfidf. The Macro F1-score of the Nave bayesian classifier was 61.97%, while the accuracy was 60.44%. The SVM classifier performed better with an accuracy of 66.22% and Macro F1-score of 65.34%. The performance of the other machine learning algorithms is lower, as concluded in Table 15. Deep learning models BERT, RoBERTa, and DistilBERT were used to evaluate performance measures with the proposed model. The maximum length is set to 125, and alternative batch sizes (eight, sixteen, thirty two) and various learning rate were evaluated as part of the hyper-parameter tuning process in order to maximize performance. After fine tuning the RoBERTa model, we discovered that it outperforms the DistilBERT and BERT models. There is a slight variation in the performance of DistilBERT and BERT models. As indicated in Figs. 7 and 8, RoBERTa has the

compared to fire in this instance, so the author surmises that this is the reason for the incorrect classification..

The main cause for incorrect text classification of all the model is that the text labels were assigned by English experts who may disagree. The noise in the labeling used for model training has resulted from the above-mentioned reason. Another significant reason is that we did not consider speech in this experiment, therefore the true emotion of a person could not be captured correctly. However, in an emergency, the tone in which a person talks will aid in effective emotion capture.

Macro Precision and Macro Recall Comparision with existing models
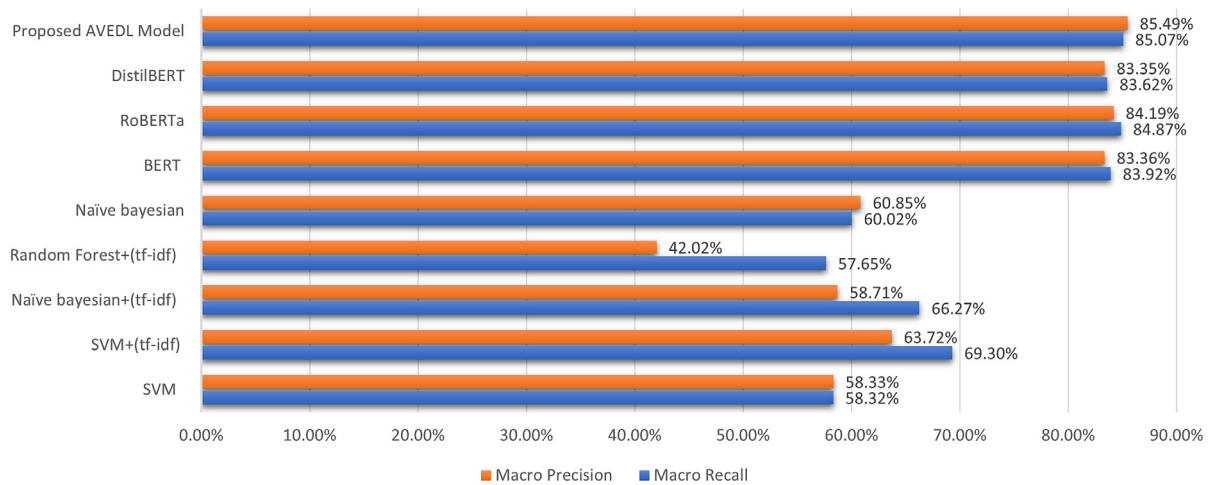

Fig. 7. Comparing the results of AVEDL Model with existing models based on Macro Precision and Macro Recall.

Accuracy and Macro F1-score comparison with existing models
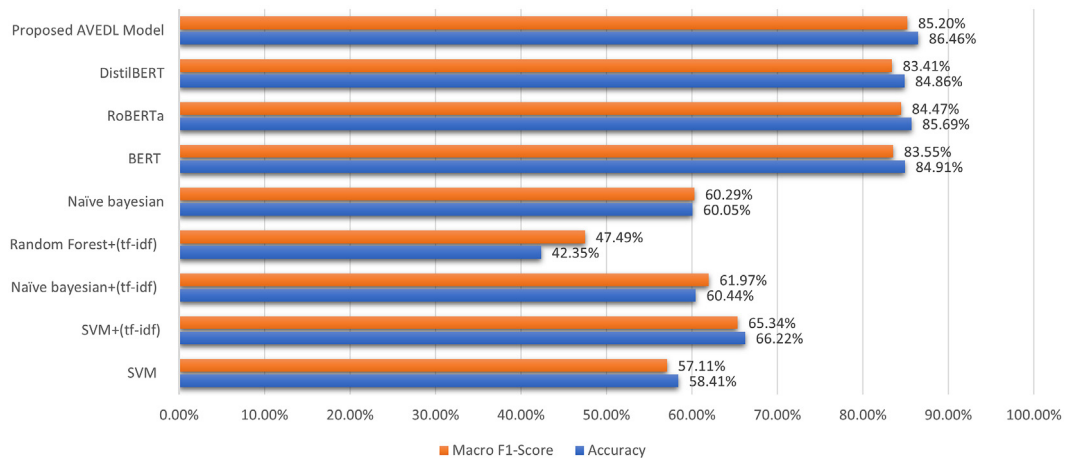

Fig. 8. Comparing the results of AVEDL Model with existing models based on Accuracy and Macro F1-score.

highest F1 score and accuracy, whereas BERT is in second and DistilBERT is in the third position. However, when compared to existing approaches and RoBERTa, the proposed AVEDL Model performs the best. The accuracy obtained by AVEDL Model is 86.46% and macro F1-score obtained is 85.20% which is better than all the models taken into consideration Based on the Table 15, it can be concluded that the proposed model outperforms all other state-of-the-art deep learning and machine learning models.

4.4. The proposed method's limitations

The AVEDL model is tested using the ERSS dataset relevant to pandemics yields significant results, but the model performs poorly while using the Normal ERSS dataset, which is not related to pandemics. This is something that has to be looked into further in the future. DistilBERT, RoBERTa, and BERT 16 GB are pre-trained models with huge memory for corpus training, which are the model's main limitation. In comparison to machine learning models, the models' time complexity is likewise relatively high. In this study, these models took 1256 s to execute with a batch size of 8, 1428 s with a batch size of 32, and 1437 s with a batch size of 16, respectively. The ensemble model procedure

took 1075 s. The models DistilBERT, RoBERTa, and BERT were run in parallel, resulting in total time complexity of 2512 s (voting time complexity+BERT time complexity). To increase the model's performance and to reduce the time complexity we intend to use more data and data compression techniques.

5. Real time application

The AVEDL model can be utilized in ERSS to aid in the process of understanding a caller's emotion and providing emotional support. The Fig. 9 shows a detailed flow diagram of how the AVEDL model can be utilized in ERSS. When a call is received in ERSS, it is transcribed and the text is pre-processed before the AVEDL Model is used to detect the emotion. We evaluated the AVEDL model's performance on a portion of the ERSS dataset that was not included in the combined tweet and ERSS dataset. The model performance of the ERSS dataset is provided in the Table 16. The proposed model was providing an accuracy of 84.33% and a Macro F1-score of 82.88%.

By understanding the caller's emotions, call takers will be able to assist people in coping with the situation and will also be able to help and support them throughout the difficult period. There are situations were the call takers can make feedback calls to
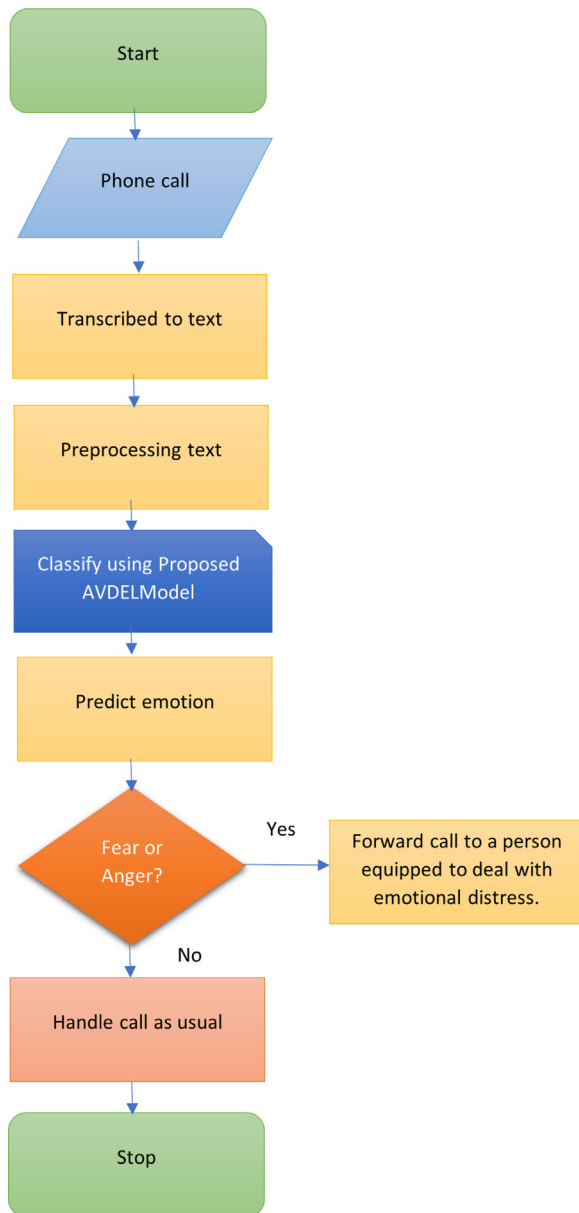
Fig. 9. Flow chart of emotion detection on ERSS-Kerala data.

Table 16
Performance evaluation metrics of AVEDL model on ERSS dataset.

Macro Precision	Macro Recall	Macro F1-score	Accuracy
83.15%	82.67%	82.88%	84.33%

callers in addition to the services that the emergency response support system can provide.

6. Conclusion and future scope

Digital health technology integrated with the power of Artificial Intelligence [40] is treated as an effective and excellent tool that can help the general population with focused health information quickly and accurately. This intervention may, for example, be sent to individuals as automated tailored messaging and education based on the content and attitudes of their social media posts, thereby providing emotional support. Personalized and educational interventions with unambiguous messages and recommendations on the nation's recovery from the pandemic,

vaccine safety and its availability, along with information related to prevention of the disease generates hope and positivity among people and assures emotional stability across the country. Detecting emotions amid a health-related crisis will always aid in improving people's mental health, allowing them to handle pandemic-like scenarios with confidence.

The proposed study is helpful for pandemic conditions in recognizing people's emotions during a pandemic. The model is also designed to use a combination of ERSS dataset and tweets that aid in understanding fear tweets and allows authorities to stem the spread of fear during a pandemic. The methodology helps detect emotions in other public health initiatives that impact people's health. As the COVID-19 progressed, public health requirements became more stringent, imposing tight regulations to prevent the spread of overlooked people's mental and emotional health. The data inclusion is limited to emotions related to COVID-19 texts connected to "corona", "covid-19", "quarantine" and "covid", "lockdown" and size of the dataset are some of the limitations of this study. We feel that the existing number of dataset is sufficient to make valid inferences about the importance of emotion detection in social media and emergency calls. We accept that our work is limited by an unbalanced dataset that represents various emotions.

In the future, context-based emotion detection will be done using the ERSS dataset (emergency), which will help detect and monitor societal emotional levels in all situations. In this research, we combined tweets with the ERSS dataset, which is connected to Covid-19. However, in the future, we intend to use the ERSS dataset alone for emotion identification, which will be highly useful in pandemic-related scenarios. Although this paper focuses on covid-related ERRSS and tweets, employing a large training dataset can improve the performance of the average voting ensemble deep learning model. We can employ various combinations of transformer-based models in the future to get better performance by training the models on a larger COVID-19 dataset. We did not even consider slang as well as speech while computing emotions thus developing a multi-model that includes slang terms and speech relevant to COVID-19 concerns would be advantageous. Our findings will be valuable to disaster management decision-makers and public health policymakers because they demonstrate the significance of paying attention to emotional concerns arising from severe public health restrictions.

Credit authorship contribution statement

K. Nimmi: Conceptualization, Methodology, Software, Validation, Writing - Original draft preparation, Data Curation. **B. Janet:** Supervision, Reviewing. **A. Kalai Selvan:** Data Curation. **N. Sivakumaran:** Project Administration, Funding acquisition.

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] S.E. Vieweg, Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications (Doctoral dissertation), University of Colorado at Boulder, 2012.

[2] M.N. Sripad, M. Pantoji, G.S. Gowda, S. Ganjekar, V.S.K. Reddi, S.B. Math, Suicide in the context of COVID-19 diagnosis in India: Insights and implications from online print media reports, Psychiatry Res. 298 (2021) 113799.

[3] M. Magdin, T. Sulka, J. Tomanová, M. Vozár, Voice analysis using PRAAT software and classification of user emotional state, Int. J. Interact. Multimedia Artif. Intell. - IJIMAI 5 (6) (2019) 33–42.

[4] R.A. Calvo, S. Mac Kim, Emotions in text: dimensional and categorical models, Comput. Intell. 29 (3) (2013) 527–543.

[5] P. Ekman, An argument for basic emotions, Cogn. Emot. 6 (3–4) (1992) 169–200.

[6] R. Plutchik, Emotion. A psychoevolutionary synthesis, 1980.

[7] M.S. Gerber, Predicting crime using Twitter and kernel density estimation, Decis. Support Syst. 61 (2014) 115–125.

[8] M.Y. Kabir, S. Madria, A deep learning approach for tweet classification and rescue scheduling for effective disaster management, in: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2019, November, pp. 269–278.

[9] P. Grover, A.K. Kar, Y.K. Dwivedi, M. Janssen, Polarization and acculturation in US Election 2016 outcomes–can twitter analytics predict changes in voting preferences, Technol. Forecast. Soc. Change 145 (2019) 438–460.

[10] M.Y. Kabir, S. Madria, EMOCOV: Machine Learning for Emotion Detection, Analysis and Visualization using COVID-19 Tweets, Vol. 23, Online Social Networks and Media, 2021, 100135.

[11] F.A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, Eng. Rep. 2 (7) (2020) e12189.

[12] F. Mozafari, H. Tahayori, Emotion detection by using similarity techniques, in: 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), IEEE, 2019, pp. 1–5.

[13] A. Joshi, V. Tripathi, R. Soni, P. Bhattacharyya, M.J. Carman, Emogram: an open-source time sequence-based emotion tracker and its innovative applications, in: Workshops At the Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[14] D. Prabhakar Kaila, D.A. Prasad, Informational flow on Twitter–corona virus outbreak–topic modelling approach, Int. J. Adv. Res. Eng. Technol. (IJARET) 11 (3) (2020).

[15] H. Yang, A. Willis, A. De Roeck, B. Nuseibeh, A hybrid model for automatic emotion recognition in suicide notes, Biomed. Inform. Insights 5 (2012) BII–S8948.

[16] A. Vijayvergia, K. Kumar, Selective shallow models strength integration for emotion detection using GloVe and LSTM, Multimedia Tools Appl. 80 (18) (2021) 28349–28363.

[17] A. Chatterjee, U. Gupta, M.K. Chinnakotla, R. Srikanth, M. Galley, P. Agrawal, Understanding emotions in text using deep learning and big data, Comput. Hum. Behav. 93 (2019) 309–317.

[18] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[19] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681.

[20] . Cauberghe, . Verolien, Ini Van Wesenbeeck, Steffi De Jans, Liselot Hudders, Koen Ponnet, How adolescents use social media to cope with feelings of loneliness and anxiety during COVID-19 lockdown, Cyberpsychol. Behav. Soc. Netw. 24 (4) (2021) 250–257.

[21] C. de Las Heras-Pedrosa, P. Sánchez-Núñez, J.I. Peláez, Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems, Int. J. Environ. Res. Public Health 17 (15) (2020) 5542.

[22] Q. Li, C. Wei, J. Dang, L. Cao, L. Liu, Tracking and analyzing public emotion evolutions during COVID-19: a case study from the event-driven perspective on microblogs, Int. J. Environ. Res. Public Health 17 (18) (2020) 6888.

[23] Faheem Aslam, Tahir Mumtaz Awan, Jabir Hussain Syed, Aisha Kashif, Mahwish Parveen, Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak, Humanit. Soc. Sci. Commun. 7 (1) (2020) 1–9.

[24] A.S.M. Venigalla, S. Chimalakonda, D. Vagavolu, Mood of India during Covid-19-an interactive web portal based on emotion analysis of Twitter data, in: Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing, 2020, pp. 65–68.

[25] 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.

[26] A. Malte, P. Ratadiya, Multilingual cyber abuse detection using advanced transformer architecture, in: TENCON 2019-2019 IEEE Region 10 Conference (TENCON), IEEE, 2019, pp. 784–789.

[27] Deepak Agnihotri, Kesari Verma, Priyanka Tripathi, Bikesh Kumar Singh, Soft voting technique to improve the performance of global filter based feature selection in text corpus, Applied Intelligence 49 (4) (2019) 1597–1619.

[28] R.K. Gupta, A. Vishwanath, Y. Yang, Global reactions to covid-19 on twitter: A labelled dataset with latent topic, sentiment and emotion attributes, 2021.

[29] Twitter standard search API, 2020, https://developer.twitter.com/en/docs/tweets/search/overview/standard.

[30] J. Perkins, Python Text Processing with NLTK 2.0 Cookbook, PACKT publishing, 2010.

[31] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 4171–4186.

[32] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, A distilled version of BERT: smaller, faster, cheaper and lighter, 2019, arXiv preprint arXiv:1910.01108.

[33] S. Casola, A. Lavelli, FBK@ SMM4H2020: RoBERTa for detecting medications on Twitter, in: Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task, 2020, December, pp. 101–103.

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

[35] Saloni Kumari, Deepika Kumar, Mamta Mittal, An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier, International Journal of Cognitive Computing in Engineering 2 (2021) 40–46.

[36] J. Bergstra, D. Yamins, D.D. Cox, Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms, in: Proceedings of the 12th Python in science conference, Vol. 13, 2013, June, p. 20.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[38] J. Liao, S. Wang, D. Li, Identification of fact-implied implicit sentiment based on multi-level semantic fused representation, Knowl.-Based Syst. 165 (2019) 197–207.

[39] A. Guo, T. Yang, Research and improvement of feature words weight based on TF-IDF algorithm, in: IEEE Information Technology, Networking, Electronic and Automation Control Conference, Chongqing, China, 2016, pp. 415–419.

[40] A. Shaban-Nejad, M. Michalowski, D.L. Buckeridge, Health intelligence: how artificial intelligence transforms population and personalized health, NPJ Digit. Med. 1 (1) (2018) 1–2.