ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING TECHNIQUES FOR EMOTION RECOGNITION USING ENGLISH TEXT

Abhishek A. Vichare*[1] and Satishkumar L. Varma[2]
[1]Department of Computer, SVKM's NMIMS Mukesh Patel School of Technology Management & Engineering, Mumbai, India
[2]Department of Information Technology, Pillai College of Engineering, Navi Mumbai, India
[1]Vichare1@gmail.com and [2]vsat2k@gmailcom
[1]ORC Id: 0000-0001-5855-3861 and [2]ORC Id: 0000-0002-5203-8655

ABSTRACT

Emotion gives thoughts and feelings of a person. Emotions are reflected from text, speech, gestures as well facial expressions. Recognizing and interpretation of emotion from such reflections plays an essential task in the interaction between human and machine conversation. Different views or opinions in terms of sentiments are derived from such emotions. Sentiment analysis provides negative, positive, or neutral terms; whereas emotional analysis provides deeper analysis of participant's emotions that tries to drill down into the various user behaviors. Emotion Recognition and Sentiment Analysis together can help to identify, process and illuminate human affects.

In this study, a diverse array of algorithms, including LSTM, BiLSTM, CNN, and RCNN, as well as ML algorithms such as RF, MNB, SVM, and Logistic Regression, were employed for the purpose of recognizing emotions from text. Models were trained and tested along with different word embedding methods and activation functions. Well balanced ISEAR dataset (International Survey On Emotion Antecedents and Reactions) is utilized in experimentation. Implementation of a fine-tuning and evaluation process for a BERT-based model using the Hugging Face Transformers library is also done for ISEAR dataset. Performance of BERT model (98.72%.) surpasses the accuracy achieved by other algorithms examined in this research and results reported in the existing literature.

Importance, applications and obstacles in the area of emotion recognition are also briefly discussed in this paper. Evolution of emotion is surveyed and discussed along with emotion models. The paper also delves into the available datasets for emotion analysis.

Keywords: Emotion recognition, BERT, ISEAR dataset, Machine learning, Deep learning algorithms

1. INTRODUCTION

Frequently, "emotions" and "sentiments" are considered interchangeable, yet "sentiment" conveys a broader concept. Sentiment analysis provides data points by whether they reflect a negative or positive feeling, or neutral, whereas emotional analysis provides deeper analysis of participant's emotions that tries to drill down into the psychology of different user behaviors. Sentiment analysis can be an effective method to extract facts about person emotions from the textual data. There are many studies done on sentiment analysis for textual information but there have not been many experiments done in the area of emotion recognition. Human beings have characteristics of expressing emotion in verbal (speech), nonverbal or written communication. Speech/verbal communication consist all aspects of tone, stress and words which convey information. Normally it happens over social media or telephone, radio or TV. Nonverbal communication consists of facial expressions, gestures or body language. Written communication holds textual information shared over blogs, email, messages or letters. Since last few years Emotion Recognition (SER) systems are becoming field of attraction for researchers [1,2].

Use of affective computing with combination of sentiment analysis can become beneficial for enormous applications. These two research field, affective computing and sentiment analysis lure research attention from educational field as well business community [3].

Emotion is natural to human beings and emotion analysis is an integral part of artificial intelligence. Because of the large data available on public platforms like Facebook, Youtube, Reddit, Twitter etc. emotion recognition has become hot research field of natural language processing owing to numerous applications [4, 5]. Facts explored from emotion recognition can provide more comprehensive feedback in treatment of unhealthy behavioral cases specially for mental illness like depressive disorder or emotional stress [6]. Wide application of NLP like voice based systems, chat/message bots consists speech as input data. Normally, Automatic Speech Recognition are used to firstconvert this input data to text followed by classification methods. It helps human computer interaction to get personalized and improved interaction feel by analyzing speaker emotional state. Automatic Speech Recognition can resolve variations in a speech from multiple users with the help of probabilistic acoustic [4].

K. Mannepalli et al. [7] also mentioned about emotion recognition as a multifaceted field, which has attracted scientist in recent years. It has many applications in health care, education and many more. Emotion recognition in conversation is required for emotion enabled dialogue systems. But it is difficult research problem due to several challenges. As per the Santosh Kumar Bharti et al. [8] existing methods doesn't provide accurate solution to recognize emotions from input text.

Emotion analysis can be extensively applied to interpret understanding the detail perspective of speech or input text. There are many application areas where emotion recognition can be used extensively such as decision making, hate speech detection, suggestion/recommendation systems, preparing business strategies, psychological assessment (depression or mental status examination, criminal psychology assessment, legal trials, personal or telephonic interviews).

EMOTION RECOGNITION CHALLENGES

Emotion recognition itself is a challenging assignment [1]. Age factor become crucial for identifying emotions. Emotion expression by elder people is different than emotion expressed by young people from speech. So it becomes difficult to identify emotions from speech of elders [9]. In emotion recognition labelling of emotions is difficult stage for other than the labelled datasets available on the web [10]. Mel frequency cepstral coefficients (MFCC) is frequently extracted feature in speech identification system. MFCC neglect relation between neighboring factors of the conversation, that create impact on emotion identification from speech [11]. Emotions differs from person to person, it depends on the internal and external surroundings. Emotions are also depending on utterance in person speech, which is important andnot an easy task.

2. RELATED LITERATURE

2.1. Evolution of Emotion

Roman statesman Cicero categorized emotion with classes fear, pain, lust, pleasure. Psychologist Robert Plutchik categorized emotion into eight basic categories of emotions [12].

Table 1: Models in emotions

Emotion Model, Year	Emotions labels	Approach	Model Structure
Tomkin model, 1962	Excitement, joy, surprise, distress, fear, shame, disgust and dis-smell	Categorical	–
Plutchik, 1980	Eight basic categories of emotions: joy, trust, fear, surprise, sadness, anger, anticipation, and disgust.	Dimensional	Wheel
Shaver et al., 1987	Anger, fear, joy, love, sadness, surprise	Categorical	Tree

Oatley et al.,1987	Anger, anxiety, disgust, happiness, sadness	Categorical	–
Ekman , 1992	Anger, disgust, fear, happiness, sadness, surprise	Categorical	–
Lovheim, 2012	Anger, contempt, distress, enjoyment, fear, interest, shame, surprise	Dimensional	Cube

Further, P. Ekman [13] mentioned about correlation between sentiments and facial gesture. In 2012, Lovheim proposed additional views to emotions like fear with terror, surprise with startle. Dr. Hugo Lövheim proposed the Lövheim Cube of Emotion, which combines the aspects of arousal, valence, and dominance to create a three-dimensional model of emotions.

There are several datasets available on the internet with different parameters. Below is the list of few emotion recognition datasets.

Table 2: Models in Emotions

No.	Dataset Name	Availability	Details
1	IEMOCAP	Publicly available	• Consists Multimodal data (audio, text and visualdata). • Consists both categorical and dimensional model ofemotion labels.
2	DailyDialog	Publicly available	• Consists only textual information. • DailyDialogue (Han, K. et al., 2014) consists onlycategorical model.
3	MELD	Publicly available	• Consists audio, text and visual data. • Contains categorical emotion labels.
4	SEMAINE	Publicly available	• Consists audio, text and visual data.
5	EmoContext	Publicly available	• EmoContext dataset contains textual information.
6	Emotionlines	Publicly available	• Consists of emotion labels for every utterance. • Textual data consists TV show script as well privateFacebook chat messages.
7	eNTERFACE'05	Publicly available	• Consists video dataset of 43 persons with basic emotions (happy, angry, disgust, sadness, surprise and fear). • Video files are categorized into facial and audiocategories (O. Martin et al., 2006).
8	RAVDESS	Publicly available	• Contains multimodal data of songs and emotional speeches. • Speech data consists calm, happy, sad, angry, and fearful, surprise as well disgust emotions.

9	Emotional Prosod ySpeech and Transcripts	Available on Fees	• This dataset is created by Linguistic Data Consortium. • Consists of audio files and related transcripts ofaudio files.
10	TIMIT Acoustic -Phonetic Continuous Speec hCorpus	Available on membership or payment of fees	• Consists speeches in American English withdifferent genders and dialects. • It is a joint venture of DARPA, MIT, SRI int. andTexas Instru.
11	ISEAR-International Survey on Emotion Antecedentsand Reactions	Publicly available	• Consist around seven thousand sentences • Consists anger, disgust, fear, sadness, shame, joy, and guilt.
12	SemEval 2019 Task 3: EmoContext	Publicly available	• Consists annotated Twitter conversations • Consists emotion happy, sad, and angry

2.2. Related work in Emotion Analysis from Speech

In 2004, multimodal emotion recognition was proposed which highlighted on body language, faces and audio [14] average 80% accuracy was achieved by using a Bayesian classifier. Mairesse et al. [15] analyzed manually collected short voice reviews with the help of openEAR/openSMILE toolkit. 72.9 % accuracy was achieved which calculated positive and negative sentiments. As a traditional classification techniques Ververidis et al. [16] and Mao, X et al. [17] used Bayesian Network model for classification. Hao, H. et al. [18], Neiberg, D. et al. [19] used SVM, Gaussian Mixture Model as well Multi-Classifier Fusion for classification respectively. HMM was used by Ntalampiras, S. et al. [20] as classifier.

New auto encoder based technique was proposed by J. Deng et al. [21] for recognizing emotions from speech. Positive and negative sentiments were categorized by applying ANN on FAU Aibo dataset. Since few years, deep learning techniques have become attractive topic in NLP applications. Kim Y. et al. [22] and Zheng W. L. et al. [23], presented Deep Belief Networks for conversation sentiment recognition and

performed significantly well over traditional method. Kim [24] used CNN with already trained word vectors to categorize data at sentence level.

DNN-Extreme Learning Machine (ELM) was presented by Han et al. [25]. To identify utterance level emotions Utterance level features were used. Accuracy level of system was not impressive. Zhang et al. [26] used character level CNNs for text categorization and demonstrated accuracy better than traditional models like bag-of-words, n-grams, word-based ConvNets and RNN.

In general speech emotion recognition consists features (e.g. spectral, pitch frequency, formant and energy related) extraction and classification to identify emotion [27, 28]. For favorite music classification Sawata et al. [29] used kernel discriminative locality with the help of EEG (electroencephalogram). 81.4% accuracy was achieved for music classification.

In 2017, Deb et al. [30] proposed conversation sentiment categorization with the help of vowel regions and non-vowel regions. Accuracy of 85.1%, 64.2% and 45.2% was achieved on EMODB, IEMOCAP and FAU AIBO dataset. Tzirakis et al. [31] introduced a SER method which consists auditory as well visual modalities to extract emotional information from different styles of speeches. M. Trotzek et al. [2] presented method to provide recognizing of stress using machine learning techniques on the basis of social media messages. This system consists CNN with word embedding.

Importance of emotion recognition in criminal behavior assessment is mentioned by J. Kaur et al. [32]. Business companies like FB, Twitter and YouTube are instructed by many governments for not taking steps against spread of hate speech on their websites. German government agencies have instructed social media to pay hefty fine up to fifty million euros per year if these companies failed to take precautionary measures against hateful messages [33].

Sung-Lin Y. et al. [34] presents an interaction aware attention network to identify emotions on the basis of utterance-base. Meaningful data is included at two levels, first at learning of utterance portrayal and at the last forecasting phase.

Xingmei Wang et al. [35] presents a method for self-supervised acoustic representation learning. The proposed method involves utilizing an acoustic-embedding memory unit and a modified space autoencoder to learn discriminative acoustic representations from unlabelled underwater acoustic data. By leveraging self-supervised learning, the authors aim to increase the performance of underwater target recognition systems by extracting informative and robust acoustic features from unannotated data.

Ayoub Ghriss et al. [36] introduces an approach for enhancing speech emotion recognition through sentiment-aware automatic speech recognition (ASR) pre-training. The authors proposed a two-stage architecture where an ASR model is first trained to recognize sentiment-related features from huge amount of labelled speech data. ASR model is tuned for speech emotion recognition using a smaller emotion-labeled dataset.

In 2023, Lixu Sun et al. [37] introduces a method to enhance speech recognition in low-resource languages using contrastive learning. The authors propose leveraging contrastive learning techniques to train deep neural networks that can effectively capture acoustic representations from limited amounts of labelled data.

2.3. MILESTONES IN STUDY OF EMOTION RECOGNITION FROM TEXTUAL DATA

Kao et al. [38] provides survey of emotion recognition techniques from text and explores potential areas for improvement in this field. This paper proposed case based reasoning technique for emotion recognition. The authors discussed various methods and challenges associated with detecting emotions from text. The paper aims to shed light on the need for advancements in text-based emotion detection and proposes potential avenues for enhancing the existing approaches. OCC model (twenty-two emotions) is used by author for research work.

Paper by Ricardo Calix et al. [39] explores the process of identifying emotional information from text data and utilizing it to generate realistic facial expressions in a three-dimensional (3-D) environment. The author presents a methodology that involves emotion recognition techniques applied to text inputs and their integration into a facial expression rendering system. The paper endeavors to connect textual emotion analysis and the visual representation of emotions in 3-D facial expressions, providing insights into the potential of text-based emotion recognition for realistic facial animation and virtual human interactions.

Sophia Y. et al. [40] presents a rule-based system designed for detecting the causes of emotions in text.The author proposes a set of linguistic rules that can identify triggers or factors contributing to specific emotions. This paper provides emotion cause corpus in Chinese language. Correlation between emotion and cause was calculated by linguistic analysis and cause detection was done by using linguistic rules. five primary emotions are considered in this paper.

Sunghwan K. et al. [41] focuses on the assessment of unsupervised emotion models for text recognition. The author assesses the performance and effectiveness of different unsupervised emotion

models in capturing and recognizing emotions from textual data. SemEval-2007, ISEAR and fairy tales were used as dataset for experimentation.

Chaffar et al. [42] explores the utilization of a heterogeneous dataset for emotion recognition fromtext. The authors investigate the benefits of combining data from distinct sources (news, fairy tales and blogs) to improve emotion analysis. The paper proposes a methodology for integrating and analysing diverse datasets to enhance the understanding of emotions expressed in text.

Erdenebileg Batbaatar et al. [43] proposes technique for emotion identification from text using a Semantic-Emotion Neural Network. The proposed model incorporates semantic information and utilizes a neural network architecture to explore the relation in words and emotions. By leveraging the semantic context of the text, the model tries to improve performance and accuracy of emotion recognition.

Asif Iqbal Middya et al. [44] reviews the existing literature on multimodal emotion recognition and highlights the challenges associated with effectively integrating audio and visual information. It then presents the proposed model-level fusion approach, which involves training separate DL models, such as CNNs and RNNs, on audio and visual data respectively. The paper explores different fusion strategies and evaluates their performance on datasets. The results presents that the proposed model-level fusion approach secured enhanced accuracy compared to using individual modalities.

Saurabh et al. [55] suggested strategies for text-based emotion identification involve the integration of transformer models (such as BERT and RoBERTa) and BiLSTM. For BERT and RoBERTA accuracy reported by authors was 71 % on ISEAR dataset. Authors achieved 72 %, 73 % for their BERT+PsyLing model and RoBERTa+PsyLing model, respectively, on ISEAR dataset.

Acheampong et al. [59] conducted comparative assessment on BERT, RoBERTA, DistilBERT, and XLNet for the task of emotion identification emplying ISEAR dataset. RoBERTa provided the maximum accuracy of 74.31% among other algorithms. Authors claimed that results achived have outperformed other methods and results achieved so far. Park et al. [57] achieved F1 score of 75.2% on the ISEAR dataset. Authors finetunes a RoBERTa-Large model in this work.

Üveges and O. Ring [60] described method of fine-tuning the Hungarian BERT to classify emotions and sentiments from data. This study obtained 0.7029 F1 for ISEAR using the BERT- base model. Authors highlighted importance of ISEAR dataset and BERT model for emotion recognition.

3. METHODOLOGIES

3.1. Dataset Details

The ISEAR dataset, the International Survey on Emotion Antecedents and Reactions, is a collection of labelled textual data that captures individuals' self-reported emotional experiences [57]. It was developed to support research in the field of psychology and emotion identification. The ISEAR dataset serves as a commonly utilized benchmark dataset and balanced dataset [55] [60].

Emotions: The ISEAR dataset focuses on seven fundamental emotions: anger, fear, joy, sadness, disgust, shame, and guilt.

Collection Method: The dataset was created through a survey-based approach. Participants were presented with a list of situational triggers, and they were asked to recall and describe an event from their personal experiences that caused them to experience a particular emotion. The descriptions provided by the participants form the textual data in the dataset.

Language and Format: The ISEAR dataset consists of sentences or short text snippets written in the English language. Each entry represents an individual's description of an emotional experience.

Labelling: Each entry in the ISEAR dataset is labelled with one of the emotion mentioned above. The labelling was done based on the self-reported emotional state indicated by the participant's description. Each entry is related with a single emotion label.

Size: The ISEAR dataset contains a moderate number of entries. It consists of approximately 7,000 labelled instances, with an equal number of instances of each emotion.

Use in Research: The ISEAR dataset has found extensive application in research studies focused on emotion identification, sentiment analysis, and NLP. ISEAR is used with emotion identification algorithms, and gaining insights into the relationship between language and emotions.

Figure 1 presents equal allocation of percentage of each emotion present in dataset.

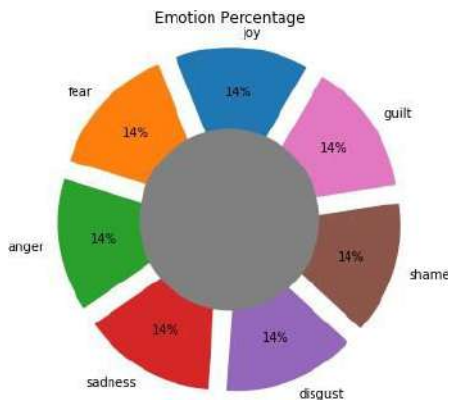

Fig. 1 Emotions in Dataset

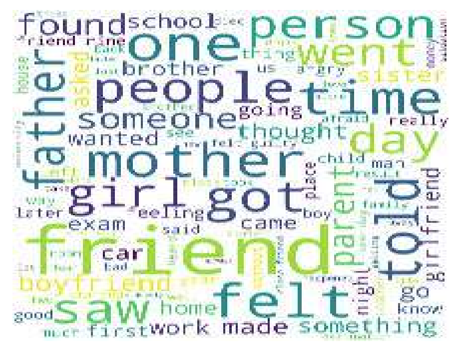

Fig. 2 Word Cloud

3.2 Machine Learning Approach

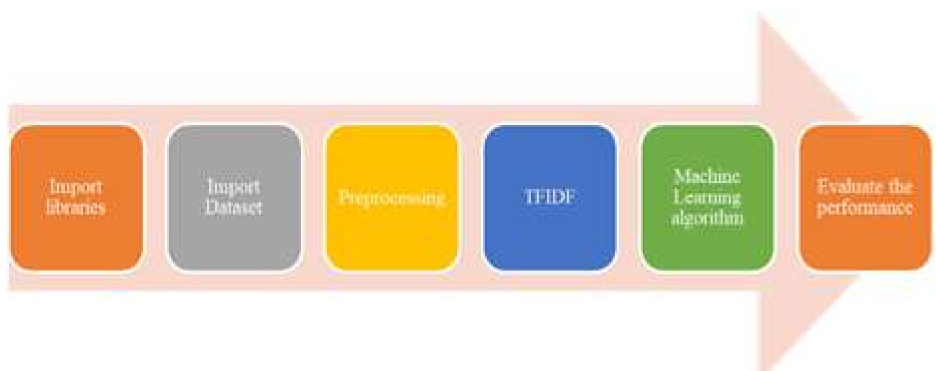

Fig. 3 Steps In Ml Approach

The initial experiment involved the training of algorithms such as random forest, multinomial NB, SVM, and logistic regression. Data splitting was carried out, allocating 80% for training and 20% for testing. Pre-processing steps were initiated, encompassing the removal of HTML markup, URLs, hashtags, punctuation, non-ASCII digits, and whitespace. Tokenization using NLTK and stemming with NLTK were executed as part of the pre-processing steps. Subsequently, vectorization of words was executed using TF-IDF.

3.3 BERT Implementation Details

The BertForSequenceClassification model is fetched from the 'bert-base-uncased' pre-trained model. The num_labels parameter is set to the number of emotion classes in the task-specific dataset. The final layer consists a softmax activation, appropriate for multi-class categorization.

Fine-tuning a BERT model involves accepting a pre-trained BERT and further training it using specific dataset. In this script, the fine-tuning process is implemented in the training loop. BERT model is loaded from the 'bert-base-uncased' pre-trained model, which is crafted for sequence classification tasks, and the num_labels parameter is set to the number of emotion labels in the task-specific dataset. The key idea behind fine- tuning is that the pre-trained BERT model, which has learned rich contextual representations from big corpus, is adapted to a particular task by changing parameters based on a smaller, job specific dataset. The process helps the model capture domain-specific patterns and better results.

BERT does not use traditional word embeddings like FastText or GloVe. BERT relies on a transformer architecture [55], which evaluate the whole context of a word in a sentence rather than relying on fixed-size word embeddings. Other methods like FastText or GloVe, assign a fixed-size vectorto each word. These embeddings do not employ the meaning of words in a sentence. In contrast, BERT uses a transformer model that fetch the sequence of words bidirectionally. This enables BERT to capture context- dependent representations for each word, considering its surroundings in the sentence. The resulting embeddings are dynamic and context-aware, making BERT well-suited for various NLP tasks, encompassing sentiment analysis, NER, and question answering [53] [56].

Mathematical formulations for the key components of the BERT-based sequence classification model in the code are as follows.

1. Tokenization and Embedding Layer:

- Tokenization: encoded_dict=tokenizer. encode_plus (text,…) where text is the input text.
- Input IDs: input_ids=encoded_dict[′input_ids′]
- Attention Masks: attention_mask=encoded_dict[′attention_mask′]

2. Encoder Layer:

- Model Instantiation: model=BertForSequenceClassification.from_pretrained (′bert−base−uncased′, num_labels=len(emotion_label_mapping))
- Forward Pass: outputs=model (input_ids, attention_mask=attention_mask, labels=labels) where labels are the ground truth emotion labels.
- Loss Calculation: loss=outputs. loss
- Backward Pass and Optimization:loss. backward ()

Torch.nn.utils.clip_grad_norm_(model. parameters (),1.0) optimizer.step() optimizer=Adam W (model.parameters() ,lr=2e-5, eps=1e-8)

Optimizer Type: AdamW is used as the optimizer. AdamW stands for "Adam with Weight Decay," and it is a popular choice for training neural networks.

Model Parameters: model.parameters(). Learning Rate (lr): The learning rate is set to 2×10−5.

Epsilon (eps): eps is a small value added to the denominator of the update rule.In this case, it is set to 1×10^{-8}.

- Learning Rate Scheduler: scheduler=get_linear_schedule_with_warmup(optimizer,num_warmup_steps=0, num_training_steps= total_steps)

4 RESULTS AND DISCUSSION

4.1 Machine Learning Approach Results

Table 3: Isear Dataset Accuracy With Ml

Algorithm	Accuracy Achieved	Accuracy from Literature
Naïve Bayes	55.91%	43.24 % [47]
RandomForest classifier	55.11%	49 % [45], 37.26 % [47], 47.0 % [48], 56 % [49]

SVM	57.19%	14.48 % [47], 54.3 % [48], 58 % [49], 55 % [45]
Logistic regression	58.06%	57.16 [52]

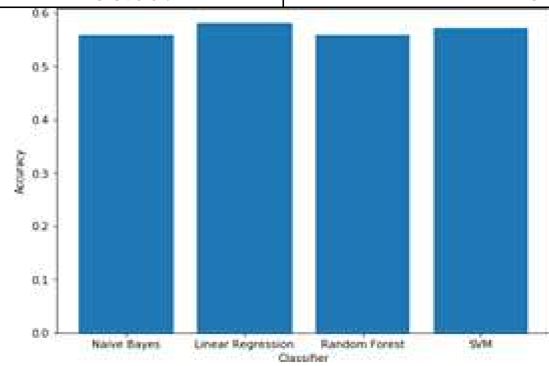

Fig. 4 Ml Model Accuracy

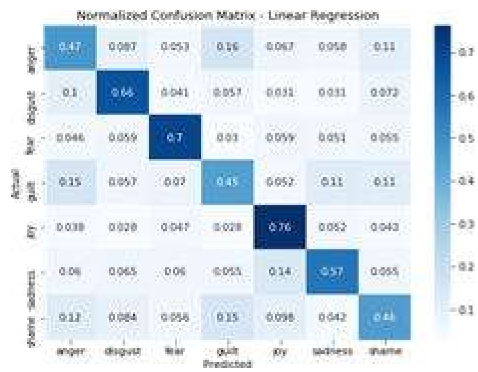

Fig. 5 Lr Confusion Matrix

Logistic regression provides 58.06 % accuracy which is comparatively higher than the NB, random forest classifier and logistic regression.

4.2 DEEP LEARNING APPROACH RESULTS

In second experiment LSTM, BiLSTM, RCNN as well CNN is applied on ISEAR dataset. BiLSTM uses two LSTM for training sequential input. First LSTM is applied on input without any change. Second LSTM is applied on reverse representation of input. Word embedding techniques used are FastText and GloVe. wiki-news-300d-1M.vec pre-trained word vector is used from fastText library. Glove.6B.100d.word2vec.txt pre-trained word vector is used from GloVe.

Table 4 Isear Dataset Accuracy with Dl

Algorithm	Word embedding	Activation function	Accuracy
BiLSTM	FastText	Relu	36.32 %
BiLSTM	GloVe	Softmax	58.79 %
BiLSTM	FastText	Sigmoid	59.53 %
BiLSTM	FastText	Softmax	61.61 %
LSTM	GloVe	Softmax	58.12 %
LSTM	FastText	Softmax	62.20 %
RCNN	FastText	Softmax	58.93 %
CNN	FastText	Softmax	63.75 %

Results from above table shows that CNN with FastText word embedding along with Softmax activation function gives better accuracy compared to other combinations of word embedding methods and activation functions.

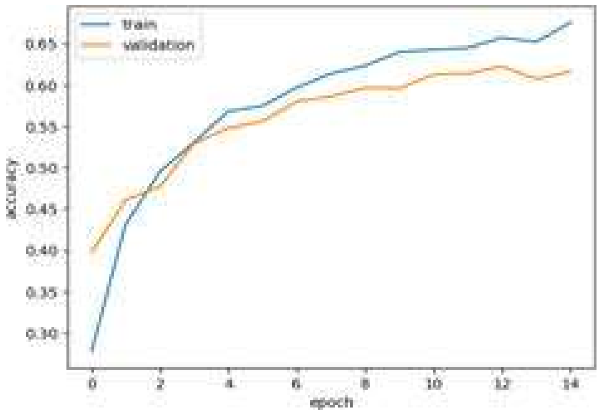

Fig. 6 Bilstm Model Accuracy

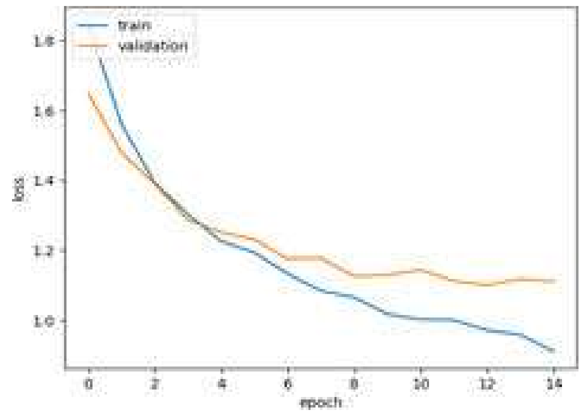

Fig. 7 Bilstm Model Loss

Fig. 8 Bilstm Confusion Matrix

Figure 6, 7 and 8 provides result for BiLSTM applied along with fastText word embedding and softmax activation function on ISEAR dataset. Accuracy achieved here is 61.61%. Figure 9, 10 and 11 provides result for LSTM applied along with fastText word embedding and softmax activation function. Accuracy achieved is 62.20 %.

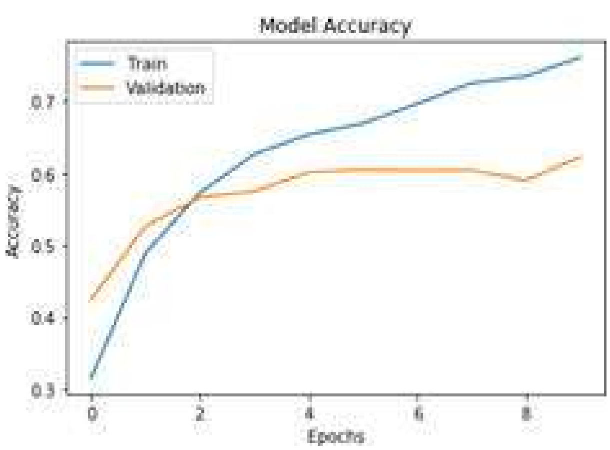

Fig. 9 Lstm Model Accuracy

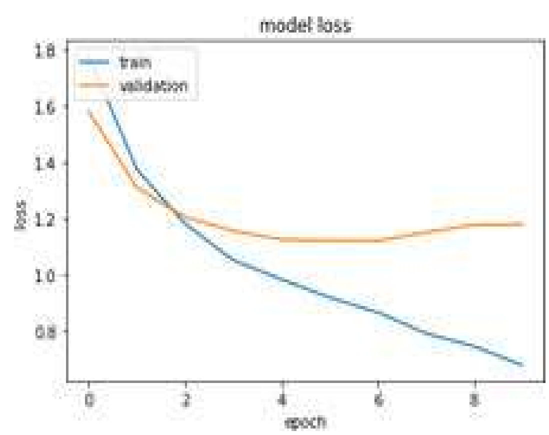

Fig. 10 Lstm Model Loss

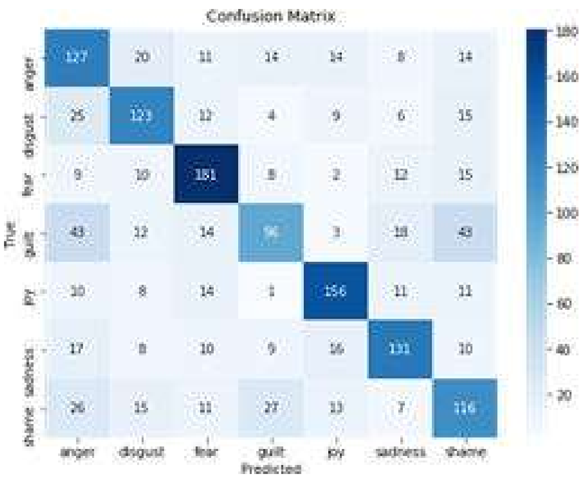

Fig. 11 LSTM confusion matrix

Figure 12, 13 and 14 provides result model accuracy, loss and confusion matrix for CNN applied along with fastText word embedding and softmax activation function. CNN provide accuracy 63.75% for ISEAR dataset.

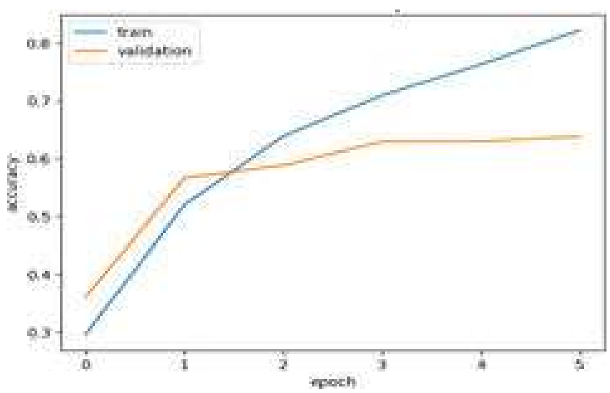

Fig. 12 Cnn Model Accuracy

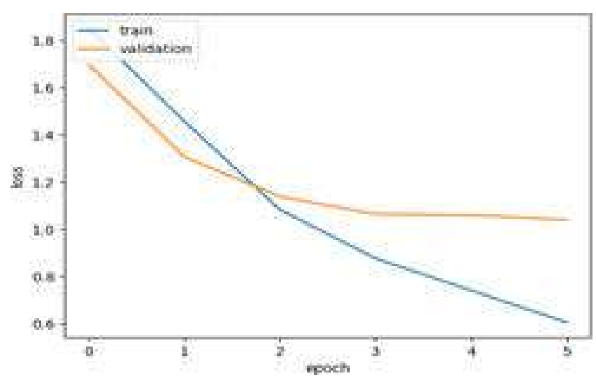

Fig. 13 Cnn Model Loss

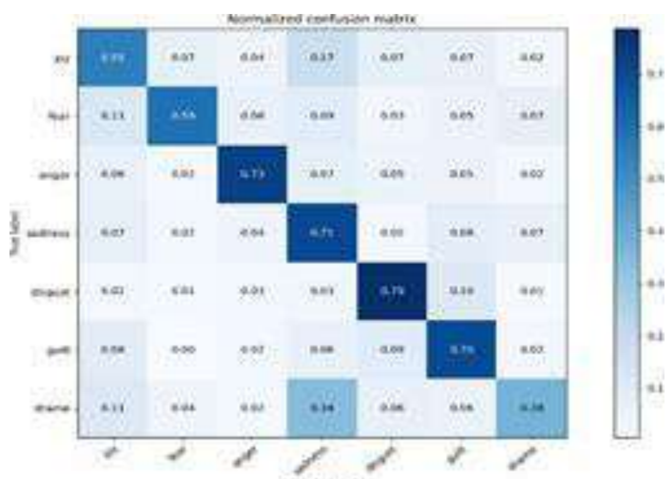

Fig. 14 Confusion matrix

4.3 B E R T RESULTS

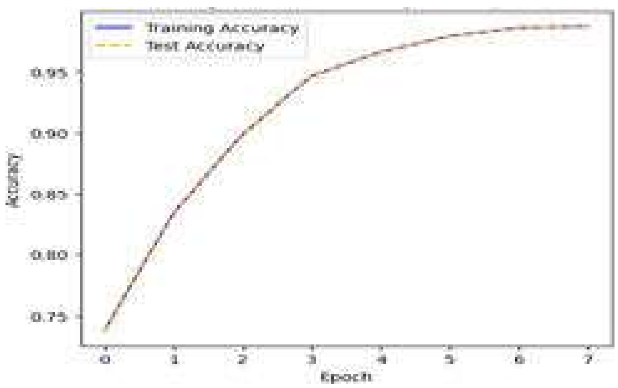

Fig. 15 Bert Model Accuracy

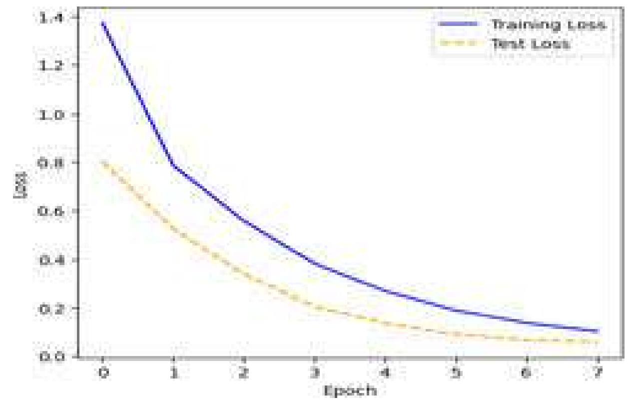

Fig. 16 Bert Model Loss

Figures 15 and 16 illustrate the model's performance concerning training and testing loss as well as accuracy. The figure 17presents the confusion matrix obtained from assessing the BERT model on ISEAR dataset.

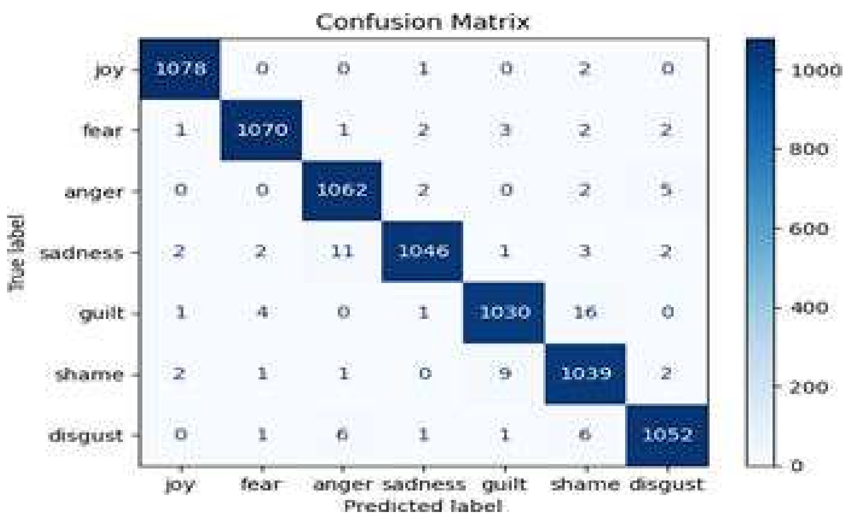

Fig. 17 Bert Confusion Matrix

Table 5: ISEAR dataset accuracy with BERT

Algorithm	Accuracy achieved	Accuracy from literature
BERT	98.72 %	72 % [54], 71 % [55], 70.09 % [58], 72.64% [59], 70.29 [60]
RoBERTa	--	75.2 % [57], 74. 31 % [58], 71 % [55]

5 CONCLUSION

Recognizing emotions from text presents a challenging research problem that has garnered surging interest in recent times, emerging as a notable field of study. This paper consists a comparative analysis of performance assessment of ML and DL algorithms for emotion recognition.

The research employs the ISEAR dataset, comprising 7666 sentences categorized into seven classes representing seven fundamental emotions. In this work, we employed a diverse set of ML algorithms, including RF, MNB, SVM and Logistic Regression, to analyze the ISEAR dataset. For a comprehensive comparative analysis, we explored various methods, incorporating LSTM, BiLSTM, CNN and RCNN algorithms. Models were trained and tested along with distinct embedding techniques and activation functions.

BERT model stands out with the maximum accuracy among all the models, reaching 98.72 %. The BERT model's performance exceeds the accuracy achieved by other algorithms used in this study and surpasses results reported in the literature. Fine-tuning and evaluation process for a BERT-based model using the Hugging Face Transformers library is done in this work. To our knowledge, no such accuracy has been previously reported.

Only one dataset is used in this research work. In future, other datasets mentioned in table 2 can be used for more experiments with similar approaches.

Author contributions Research is conducted by the corresponding author Abhishek A. Vichare and supervision was done by Satishkumar L. Varma.

REFERENCES

[1] Poria S, Hazarika D, Majumder N, Naik G, Mihalcea R, Cambria E. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019:527-536.

[2] Trotzek M, Koitka S, Friedrich CM. Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences. IEEE Trans Knowl Data Eng. 2018;588-601.

[3] Han J, Zhang Z, Cummins N, Schuller B. Adversarial Training in Affective Computing and Sentiment Analysis: Recent Advances and Perspectives [Review Article]. IEEE Comput Intell Mag. 2019;14(2):68-81.

[4] Tripathi S, Kumar A, Ramesh A, Singh C, Yenigalla P. Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions. ArXiv. 2019. https://arxiv.org/abs/1906.05681.

[5] Kratzwald B, Ilic S, Kraus M, Feuerriegel S, Prendinger H. Deep learning for affective computing: Text-based emotion recognition in decision support. arXiv. 2018. https://arxiv.org/abs/1803.06397.

[6] Provoost S, Ruwaard J, Van Breda W, Riper H, Greenshaw AJ. Validating Automated Sentiment Analysis of Online Cognitive Behavioral Therapy Patient Texts: An Exploratory Study. Front Psychol. 2019;10:1-12.

[7] Mannepalli K, Sastry PN, Suman M. Emotion recognition in speech signals using optimization-based multi-SVNN classifier. J King Saud Univ - Comput Inf Sci. 2018;384-397.

[8] Bharti SK, Varadhaganapathy S, Gupta RK, Shukla PK, Bouye M, Hingaa SK, Mahmoud A. Text-Based Emotion Recognition Using Deep Learning Approach. Comput Intell Neurosci. 2022.

[9] Wang K, An N, Li BN, Zhang Y, Li L. Speech Emotion Recognition Using Fourier Parameters. IEEE Trans Affect Comput. 2015;6(1):69-75.

[10] Tashev IJ, Wang Z, Godin K. Speech emotion recognition based on Gaussian Mixture Models and Deep Neural Networks. In: Proceedings Information Theory and Applications Workshop (ITA). San Diego, CA; 2017:1-4.

[11] Sun Y, Wen G, Wang J. Weighted spectral features based on local Hu moments for speech emotion recognition. Biomed Signal Process Control. 2015;18:80-90.

[12] Plutchik R. Emotion: a psychoevolutionary synthesis. New York: Harper and Row; 1980.

[13] Ekman P. Facial expression and emotion. Amer Psychol. 1993;48(4):384.

[14] Caridakis G, Malatesta L, Kessous L, Amir N, Raouzaiou A, Karpouzis K. Modeling naturalistic affective states via facial and vocal expressions recognition. In: Proc. 8th Int. Conf. Multimodal Interfaces; 2004:146-154.

[15] Mairesse F, Polifroni J, Di Fabbrizio G. Can prosody inform sentiment analysis? Experiments on short spoken reviews. In: Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP); 2012:5093-5096.

[16] Ververidis D, Kotropoulos C. Fast and accurate sequential floating forward feature selection with the bayes classifier applied to speech emotion recognition. Signal Processing. 2008;88(12):2956-2970.

[17] Mao X, Chen L, Fu L. Multi-level Speech Emotion Recognition Based on HMM and ANN. In: WRI World Congress on Computer Science and Information Engineering; 2009:225-229.

[18] Hao H, Xu MX, Wu W. GMM Supervector Based SVM with Spectral Features for Speech Emotion Recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP; 2007:413-416.

[19] Neiberg D, Laskowski K, Elenius K. Emotion Recognition in Spontaneous Speech Using GMMs. INTERSPEECH; 2006.

[20] Ntalampiras S, Fakotakis N. Modeling the Temporal Evolution of Acoustic Parameters for Speech Emotion Recognition. IEEE Trans Affect Comput. 2012;3.99:116-125.

[21] Deng J, Zhang Z, Marchi E, Schuller B. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In: Proc Humaine Assoc Conf Affect Comput Intell Interact (ACII); 2013:511-516.

[22] Kim Y, Lee H, Provost EM. Deep learning for robust feature generation in audio-visual emotion recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing; 2013:3687-3691.

[23] Zheng WL, Zhu J, Peng Y. EEG-based emotion classification using deep belief networks. IEEE International Conference on Multimedia & Expo; 2014:1-6.

[24] Kim Y. Convolutional Neural Networks for Sentence Classification. In: Proceedings of the 2014 Conference on EMNLP; 2014:1746–1751.

[25] Han K, Yu D, Tashev I. Speech emotion recognition using deep neural network and extreme learning machine. INTERSPEECH; 2014.

[26] Zhang X, Zhao J, LeCun Y. Character-level Convolutional Networks for Text Classification. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH; 2015:3057–3061.

[27] Lee J, Tashev I. High-level feature representation using recurrent neural network for speech emotion recognition. INTERSPEECH; 2015.

[28] Jin Q, Li C, Chen S, Wu H. Speech emotion recognition with acoustic and lexical features. In: IEEE International Conference on Acoustics, Speech and Signal Processing; 2015:4749-4753.

[29] Sawata R, Ogawa T, Haseyama M. Novel audio feature projection using KDLPCCA-based correlation with EEG features for favorite music classification. IEEE Trans Affect Comput. 2016;10(3):430-444.

[30] Deb S, Dandapat S. Emotion Classification using Segmentation of Vowel-Like and Non-Vowel-Like Regions. IEEE Trans Affect Comput. 2017;1-1.

[31] Tzirakis P, Trigeorgis G, Nicolaou MA, Schuller B, Zafeiriou S. End-to-end multi-modal emotion recognition using deep neural networks. IEEE J Sel Top Signal Process. 2017.

[32] Kaur J, Juglan K, Sharma V. Role of Acoustic Cues in Conveying Emotion in Speech. J Forensic Sci Crim Investig. 2018.

[33] Biere S, Bhulai S. Hate Speech Detection Using Natural Language Processing Techniques. Master's dissertation, Vrije Universiteit Amsterdam; 2018.

[34] Yeh S, Lin Y, Lee C. An Interaction-aware Attention Network for Speech Emotion Recognition in Spoken Dialogs. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019:6685-6689.

[35] Wang X, Meng J, Liu Y, Zhan G, Tian Z. Self-supervised acoustic representation learning via acoustic-embedding memory unit modified space autoencoder for underwater target recognition. J Acoust Soc Am. 2022;152(5):2905-2915.

[36] Ghriss A, Yang B, Rozgic V, Shriberg E, Wang C. Sentiment-Aware Automatic Speech Recognition Pre-Training for Enhanced Speech Emotion Recognition. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2022:7347-7351.

[37] Sun L, Nurmemet Y, Jiang L. A Method Improves Speech Recognition with Contrastive Learning in Low-Resource Languages. Appl Sci. 2023;13:4836.

[38] Kao E, Chieh Chun, Liu Y, Yang T-H, Hsieh C-T, Soo V. Towards Text-based Emotion Detection: A Survey and Possible Improvements. In: Proceedings of International Conference on Information Management and Engineering, ICIME 2009; 2009:70-74.

[39] Calix RA, Mallepudi SA, Chen B, Knapp GM. Emotion Recognition in Text for 3-D Facial Expression Rendering. IEEE Trans Multimedia. 2010;12(6):544-551.

[40] Lee SYM, Chen Y, Huang C-R. A Text-driven Rule-based System for Emotion Cause Detection. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, CA. Association for Computational Linguistics; 2010:45–53.

[41] Kim S, Valitutti A, Calvo R. Evaluation of unsupervised emotion models to textual affect recognition. In: Proceedings of the Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics; 2010:62–70.

[42] Chaffar S, Inkpen D. Using a Heterogeneous Dataset for Emotion Analysis in Text. 2011:62-67.

[43] Erdenebileg B, Li M, Ryu KH. Semantic-Emotion Neural Network for Emotion Recognition from Text. IEEE Access. 2019;111866-111878.

[44] Middya AI, Nag B, Roy S. Deep learning based multimodal emotion recognition using model-level fusion of audio–visual modalities. Knowledge-Based Systems. 2022;244:108580.

[45] Polignano M, Gemmis M, Basile P, Semeraro G. A comparison of word-embeddings in emotion detection from text using BiLSTM, CNN and self-attention. In: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization. Laranca, Cyprus; 2019:63–68.

[46] Yohanes D, Putra JS, Filbert K, Suryaningrum KM, Saputri HA. Emotion Detection in Textual Data using Deep Learning. Procedia Computer Science. 2023;227:464-473.

[47] Sailunaz K, Alhajj R. Emotion and sentiment analysis from Twitter text. Journal of Computational Science. 2019;36:101003.

[48] Kratzwald B, Ilić S, Kraus M, Feuerriegel S, Prendinger H. Deep learning for affective computing: Text-based emotion recognition in decision support. Decision Support Systems. 2018;115:24-35.

[49] Wassermair P. Emotion Classification of Text (Thesis). Institute of Computational Perception; 2023.

[50] Nazarenko D, Afanasieva I, Golian N, Golian V. Investigation of the Deep Learning Approaches to Classify Emotions in Texts. Presented at COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, Kharkiv, Ukraine, April 22–23, 2021.

[51] Li Y, Kazameini A, Mehta Y, Cambria E. Multitask Learning for Emotion and Personality Detection. IEEE Transactions on Affective Computing. 2021;1(1):1.

[52] Dobša J, Šebalj D, Bužić D. Classification of Emotions Based on Text and Qualitative Variables. In: 44th International Convention on Information, Communication and Electronic Technology (MIPRO). Opatija, Croatia; 2021:383-388. doi:10.23919/MIPRO52101.2021.9596747.

[53] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of NAACL-HLT 2019. Association for Computational Linguistics; 2019:4171–4186.

[54] Mahmud Chowdhury MS, Pal B. BERT-Based Emotion Classification Approach with Analysis of COVID-19 Pandemic Tweets. In: Applied Informatics for Industry 4.0; 2023.

[55] Zanwar S, Wiechmann D, Qiao Y, Kerz E. Improving the Generalizability of Text-Based Emotion Detection by Leveraging Transformers with Psycholinguistic Features. In: Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS). Abu Dhabi, UAE: Association for Computational Linguistics; 2022:1–13.

[56] Yin W, Shang L. Efficient Nearest Neighbor Emotion Classification with BERT-whitening. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics; 2022:4738–4745.

[57] Park S, Kim J, Ye S, Jeon J, Park H, Oh A. Dimensional emotion detection from categorical emotion. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:4367–4380.

[58] Adoma AF, Henry N-M, Chen W. Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-Based Emotion Recognition. In: 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). Chengdu, China; 2020:117-121. doi:10.1109/ICCWAMTIP51612.2020.9317379.

[59] Acheampong F, Nunoo-Mensah H, Chen W, Niyongabo Rubungo A. Recognizing Emotions from Texts Using a BERT-based Approach. 2020. doi:10.1109/ICCWAMTIP51612.2020.9317523.

[60] Üveges I, Ring O. HunEmBERT: A Fine-Tuned BERT-Model for Classifying Sentiment and Emotion in Political Communication. IEEE Access. 2023;11:60267-60278. doi:10.1109/ACCESS.2023.3285536.