FMDB *Transactions on Sustainable Computer Letters*

Exploring Hybrid Multi-View Multimodal for Natural Language Emotion Recognition Using Multi-Source Information Learning Model

J. Angelin Jeba[1], S. Rubin Bose[2,*], Roja Boina[3]

[1]Department of Electronics and Communication Engineering, Anna University, Chennai, Tamil Nadu, India.
[2]Department of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, India.
[3]Department of Computer Science, University of New Haven, West Haven, North Carolina, United States of America.
jebaangelin@gmail.com[1], rubinbos@srmist.edu.in[2], rboin1@unh.newhaven.edu[3]

Abstract: In recent years, the popularity of predictive models employing machine learning and natural language processing for identifying emotions from various text sources, including news articles, microblogs, and social media posts, has grown significantly. However, deploying such models in real-world sentiment and emotion applications faces challenges, particularly concerning poor out-of-domain generalizability. This difficulty likely arises due to the complexities of transferring multiple emotion identification models, which are influenced by domain-specific characteristics like themes, communicative aims, and annotation techniques. As an online broadcasting platform, microblogging has emerged as a prominent forum for expressing ideas and opinions, prompting researchers from various fields to explore emotion recognition (ER) from microblogs. Automatic emotion recognition from microblogs presents a formidable challenge in machine learning, especially when seeking improved results across different types of content. Emoticons have become increasingly common in microblog materials as they aid in conveying content meaning. This research proposes a method for emotion recognition from microblog data incorporating text and emoticons. Emoticons are considered distinctive ways for users to communicate their feelings, and their meanings can be further enriched by incorporating appropriate emotional phrases. A Multi-Source Information Learning Model is employed to classify emotions, which considers the sequence of emoticons appearing in the microblog data. Experimental findings demonstrate that the suggested emotion recognition algorithm outperforms other methods when tested on Twitter data.

Keywords: Artificial Intelligence; Transfer Learning Algorithm; Convolutional Neural Network (CNN); Recurrent Neural Network (RNN); Receiver Operating Characteristic (ROC) Curve.

Received on: 09/10/2022, **Revised on:** 27/11/2022, **Accepted on:** 08/01/2023, **Published on:** 09/02/2023

Cited by: J. A. Jeba, S. R. Bose, R. Boina, "Exploring Hybrid Multi-View Multimodal for Natural Language Emotion Recognition Using Multi-Source Information Learning Model," *FMDB Transactions on Sustainable Computer Letters*., vol. 1, no. 1, pp. 12–24, 2023.

1. Introduction

Emotion recognition is a fascinating research and application area focusing on understanding and interpreting human emotions through various technologies, typically within computer science, artificial intelligence, and psychology. Emotion recognition aims to enable machines, such as computers or robots, to identify and understand human emotions based on different cues, such as facial expressions, speech patterns, physiological signals, and body language.

*Corresponding author.

Emotions are crucial in human communication and behaviour, influencing decision-making, social interactions, and overall well-being. By developing systems that recognise emotions, we can create more empathetic and responsive human-computer interactions, improve mental health support, and enhance the overall user experience in various applications and industries.

1.1. The key modalities used in emotion recognition are Facial Expression Analysis

Analyzing facial expressions is one of the most common methods for emotion recognition. It involves detecting and interpreting facial features, such as eye movements, smiles, frowns, and other muscle movements associated with different emotions.

Speech Analysis: Emotions are often reflected in speech's tone, pitch, and other acoustic features. Emotion recognition systems can analyze spoken language to detect emotions like happiness, sadness, anger, or fear.

Body Language and Gestures: Human emotions are expressed through body language, such as posture, hand movements, and gestures. Capturing and interpreting these cues can provide valuable insights into a person's emotional state.

Physiological Signals: Emotions can trigger changes in physiological responses, such as heart rate, skin conductance, and brain activity. Biometric sensors can measure these signals, allowing for emotion recognition in real-time.

Text Analysis: Emotion recognition from text involves analyzing written language to detect emotional content, sentiment, and intent. This is particularly useful in sentiment analysis for social media, customer feedback, and chatbot interactions.

1.2. Emotion recognition technology is applied in many areas, including Human-Computer Interaction (HCI)

Enhancing the interaction between humans and computers by making systems more responsive to users' emotional states. Healthcare: Assisting in diagnosing and treating mental health disorders, autism, and other conditions by monitoring patients' emotions and responses.

Marketing and Advertising: Understanding consumers' emotional reactions to products, advertisements, or services to improve marketing strategies.

Education: Creating emotionally aware learning systems that adapt to students' emotional needs and engagement levels. Virtual Reality and Gaming: Enhancing user experiences in virtual environments and games by adapting the content based on emotional feedback.

However, despite the significant progress in emotion recognition technology, challenges remain, particularly in handling cultural and individual variations in expressing emotions accurately. Ethical concerns about privacy and data usage must also be addressed as emotion recognition systems become more prevalent daily.

Emotion recognition is a rapidly evolving field with tremendous potential to revolutionize various aspects of human-computer interaction and improve our understanding of emotions and their impact on our lives. In natural language processing (NLP) and affective computing, emotion recognition from natural language has grown in importance. Accurately recognizing emotions from the text can help improve communication between humans and machines, enable personalized user experiences, and aid mental health diagnosis and treatment.

However, traditional emotion recognition methods rely on limited features and modalities, reducing performance and generalization ability. To solve this problem, we suggest a hybrid multi-view multimodal approach for natural language emotion recognition, which utilizes multiple sources of information to increase the recognition model's reliability and accuracy. Specifically, we combine text, audio, and visual modalities to create a multi-source learning model that can capture the different aspects of emotions expressed in natural language. Furthermore, we use a multi-view approach to capture different perspectives of the input data, which can help to overcome the limitations of single-view models.

Our approach builds on recent advancements in deep learning and multimodal fusion techniques, which have demonstrated positive outcomes in various applications, including speech recognition, image classification, and natural language processing. In particular, we leverage a deep neural network architecture that can effectively learn and integrate information from multiple modalities while also being able to handle missing or noisy data.

In this paper, we test our suggested strategy's efficacy via an experiment on a dataset open to the general public. Our findings demonstrate the promise of our technique for real-world applications by proving that our multi-source learning model beats current state-of-the-art approaches for natural language emotion recognition. Overall, our work shows the significance of utilising many sources of information for improved natural language emotion recognition and adds to the expanding corpus of research on multimodal and multi-view learning.

2. Literature Review

Lee and Choi [1] introduced DCWS-RNNs, a multimodal model for understanding emotions during dialogue. In contrast to the cutting-edge approach, Dialogue RNN, this approach feels that a test utterance's circumstances should be broken down into four categories. Additionally, different contextual window sizes can represent the implicit weights of these various context-related aspects, enhancing accuracy performance.

Poria, et al. [2] suggested a novel technique for identifying human emotions based on deep and acoustic features. The proposed method is assessed on the widely used in the literature RADVES, EMO-DB, and IEMOCAP datasets. First, speech signals' acoustic features and spectrograms are extracted.

Alharbi, et al. [3] suggested the dimensional model of emotions as an alternative to the popular auditory perceptual representation space. The main advancement is the establishment of a perceptual boundary for the oversampling of minority emotion classes in this context. This limit, which uses arousal and valence criteria, yields two methods for achieving balance in the data: perceptual borderline oversampling and perceptual borderline SMOTE (Synthetic Minority Oversampling Technique). The results of utilising neural network classifiers for emotion recognition show that the recommended perceptual oversampling methods greatly surpassed the state-of-the-art in all scenarios and languages.

Vaswani, et al. [4] suggested the dimensional model of emotions as an alternative to the popular auditory perceptual representation space. The most important contribution is setting a perceptual boundary for oversampling minority emotion classes in this domain. Based on arousal and valence criteria, this limit produces two strategies for balancing the data: perceptual borderline oversampling and perceptual borderline SMOTE. The results of emotion recognition using neural network classifiers indicate that the proposed perceptual oversampling methods outperformed the state-of-the-art in all scenarios and languages.

Katsis et al. [5] proposed adding conversational semantic role labelling information and the commonsense knowledge feature from ATOMIC in order to improve semantics network design for emotion recognition in conversation. For fusion extraction, a knowledge-enhanced language representation layer based on self-attention has been developed.

Shen, et al. [6] introduced a device that controls the production of emotionally congruent dialogue. The control unit is a framework for selecting a technique to stop emotional sway. The author generates replies using the control unit and emotional channel, which represent the analysis and the emotion, respectively, as opposed to the existing approaches, which directly inject emotional words into the decoder.

Hazarika et al. [7] proposed the brand-new Topic-Extended Emotional Conversation Generation Model TECM-JD. The model embeds the selected emotion category as an additional input into the emotionally independent unit of the decoder in order to reduce the expression of the material affected by the inclusion of emotion. The Twitter LDA model employs a combined attention technique to obtain the input sequence content and the input sequence subject word content, ensuring that both the output topic and the input topic are categorised under the same heading. The results of the experiment indicate that the proposed model performs well and is superior to standard dialogue models in providing topic-relevant emotional content with more depth.

Hazarika, et al. [8] studied the problem of efficiently combining lexical modalities during training while providing a deployable acoustic model that does not require linguistic inputs. The advantage of this paper is that it tries to understand emotions with contextual information. The disadvantage of this paper is that it does not properly recognize big data like images.

Abdul-Mageed and Ungar [9] said that the Ekman model includes happiness or joy, sadness, disgust, anger, fear and surprise as basic emotions. Emotion Detection from textual sources can be done using concepts of Natural Language Processing. The highlight of this paper is that it detects emotions in textual form, which is increasingly used in microblogs and apps like Twitter. This paper explores emotion detection using NLP. This paper does not provide a detailed methodology for the proposed emotion detection system.

Coolican et al., [10] proposed a method of recognising emotion from facial images with pose, occlusion, and illumination. This has never been done and implemented in a hybrid model. It addresses facial emotion recognition, which is a trending topic. Another advantage is that it implements a hybrid model for improved accuracy. The disadvantage is that this model is not tested on other big data types.

Collins and D. G. Bobrow [11] pushed for a technique that draws on the correlation between the AffWild2 dataset's seven core emotions and twelve action components. The ResNet50-inspired approach utilises multi-task learning. This paper's strength is its emphasis on data-type-optimal learning. This study has the limitation that it cannot be used in expansive settings.

Kashii, et al. [12] published a method that compiles an affective ad dataset capable of evoking coherent emotions across users. It explores the efficacy of CNN features. The advantage of this paper is that it addresses affect recognition in adverting with practical applications. The methodology for compiling the dataset and ad selection is not detailed.

Herzig, et al. [13] analysed signal-based emotion detection, including different emotion models, emotion elicitation methods, classifiers, and the whole framework for emotion recognition. This paper emphasizes the importance of detecting emotions in text. The disadvantage of this paper is that there is no clear methodology.

Hareli and Rafaeli, [14] inferred emotion and sentiment modeling due to unstructured big data with different modalities. The advantage of this paper is that it tries to address the big data challenge and emotion or sentiment modelling. The disadvantage of this paper is that this is not tested on image data. This paper weighs in on the importance of consumer feelings.

Kumar et al. [15] proposed a model where emotions are detected by the speaker using speech recognition. Different emotions are conveyed at different speeds. Example: Rage and fear are always acted upon quickly. The advantage of this paper is that compared to traditional methods, these produce more precise findings. This paper helps analyse different types of emotions in different scenarios.

3. Methodology

3.1. Microblog Emoticon Management

Due to the development of technology and science, social media has been the most popular means of expressing one's emotions. People publish videos, texts, audio, images, and other media to convey their feelings. Among them, microblogs are the most well-liked. Microblogs contain millions of words, pictures, videos, audio files, hashtags, signs and symbols, all with different meanings. Twitter is one of the most popular microblogging services. Emoticons should receive special consideration in emotion recognition alongside texts because they enhance the meaning of information.

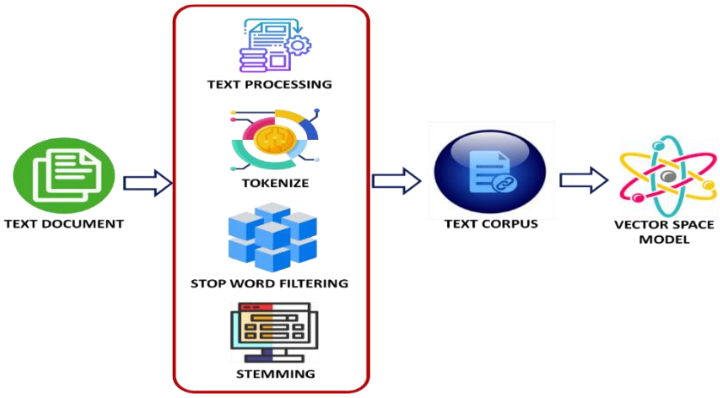


Figure 1: Block diagram of Emotion Recognition

The suggested method shown in the above Fig 1: Block diagram containing different modules pays special attention to the emoticon and its interaction with content. Text and emoticons are equally important for determining a person's actual emotion. However, as suggested by several research studies in the literature, emoticons should not be excluded from the pre-processing stage. With the help of emotive terms and other texts found in the microblog, the suggested model performed accurate emotion analysis.

The operational procedure of the suggested emotion recognition model is shown in Fig. 1 for a sample post that uses an emoticon. The proposed CNN plan consists of four sequential segments. In Task 1, a lookup database is used to translate emoticons into words that convey the same sentiments. In Task 2, the integer encoding, or IE, process is finished, which converts words into a string of integers. The next step in Task 3 is padding to produce a vector series of numbers that are all the same length. CNN is used in Task 4 to categorise certain emotions (Sad, Happy, Angry, or Love).

The full process is described in Algorithm 1, which also illustrates four unique steps. Data processing and CNN classification are two of this algorithm's main tasks. The first three phases of processing are included under data processing. The subsections that follow provide a quick explanation of the steps.

Algorithm 1: Proposed ER scheme [49]

```
Input: Microblog Data D of word size N
Output: Category of Emotion
//Task 1: Replacing emoticon(s) to corresponding meaning
        For t=1 to N, do
            If D(i) is emoticon(s), then
                    D(i) ← Emoticon, meaning (D(i))
            End If
        End For
//Task 2: Integer Encoding (IE) using Tokenizer
        For t=1 to N, do
            IE(i) = Tokenizer D(i)
        End For
//Task 3: Zero padding at first to make fixed L length
        For t=1 to N, do
            P(i) ← 0 // considering 0 for initial values
        End For
        For i=L-N+1 to N, do
            P(i) ← IE(i) // copy the rest values
        End For
//Task 4: Emotion classification using CNN
//Embedding integer to 2D vector
        For t=1 to L, do
            U(x,y) = Embedding P(i)
      End For
```

3.2. Data processing

The processing of microblog data is one of the most important steps in our proposed technique shown in Fig 2: The proposed ER architecture for a sample set of text and emoticon-based microblog data. Twitter data, including emoticons and text, is used with data from other social media networks. It is necessary to do some pre-processing processes to remove extraneous information and noisy input from the data. Cleaning includes changing the case and removing user names, hashtags, punctuation marks, and other elements.

The clean microblog data is then processed in three processes: text and emoticons. The emoticon. The meaning() function is used in Task 1's (emoticon conversion phase) search for emoticons in the microblog data and to replace them with the proper meaning. The function uses a lookup table to maintain the meaning of each emoticon word. The tokenization stage (Task 2) uses the Text_to_sequence() method to omit extraneous information and achieve IE. In Task 3, padding is accomplished using the pad_sequence() method to create an equal-length word vector sequence. Initial padding of zero is used in this instance. The CNN model is used in Task 4 to recognise certain emotions (such as happiness, love, sadness, and anger).

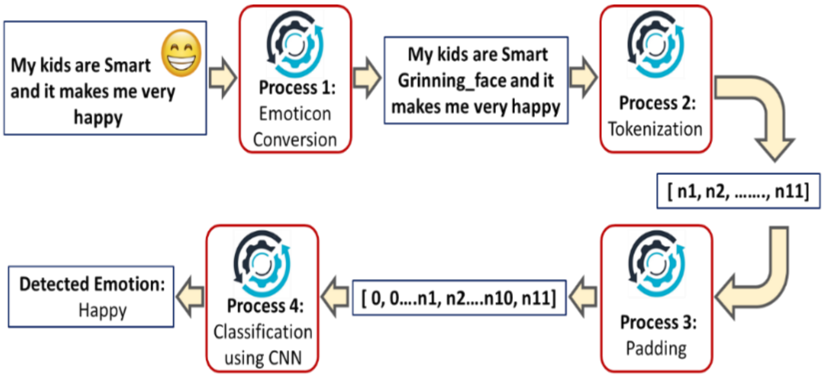


Figure 2: The proposed ER architecture for a sample set of text and emoticon-based microblog data [49]

3.3. Emotion Recognition using 2D CNN

CNN is a deep neural network often used to analyse 2D visual pictures. Numerous layers in CNN perform convolution and pooling operations. Convolutional layers from CNN offer an overview of an image's features. Sample feature maps are created by pooling layers down and summarising the presence of features on the feature map. The Conv2D layer architecture is generally utilised in regular CNN for pictures or image-like 2D inputs. The Dense Layer, the final layer used for classification, is entirely connected.

Using blog text data and a 1D convolutional operation, this work simulates a CNN operation on time series data using a CNN architecture. CNN employs the Conv1D architecture for temporal data. As time series, text data is processed using CNN's Conv1D architecture for ER. Conv1D operates using a gliding kernel in a single dimension. Emotion identification in text shows promise when constructed with two Conv1D layers, two max-pooling layers, a flattened layer, and two dense layers. We take a different approach by viewing emotion recognition as a multi-class problem using time-series data, which is often utilised for forecasting or single output prediction. Therefore, several nodes exist in the output layer.

The created method for emotion recognition from microblog data makes use of a CNN architecture, as shown in Figure 2. CNN has a stellar reputation for its ability to analyse text and extract key concepts. The layers of the proposed method's architecture are as follows: input, embedding, two convolutional, two max-pooling, flattening, two dense, and output.

The suggested model has an output dimension of 128, and an input dimension of the word size. The output spaces of the first and second convolutional layers are 64 and 32 dimensions large, respectively. The kernel size is 3, and the relu activation is used, in both convolutional layers. Each time the pooling window is advanced by one step, the maximum pooling window size is increased by 2, and so on for both maximum pooling layers. The input is then smoothed by the layer that follows. The activation function relu is used by 16 in the first dense layer. Since there are four alternative class labels, the final layer of the proposed CNN architecture has four nodes and is dense. The softmax activation function is used to determine the probabilities associated with each class.

The detailed CNN architecture is illustrated in Figure 3 to perform the emotion recognition. It consists of Input, Convolutional, Pooling Layer, fully connected, and Output layers [16].

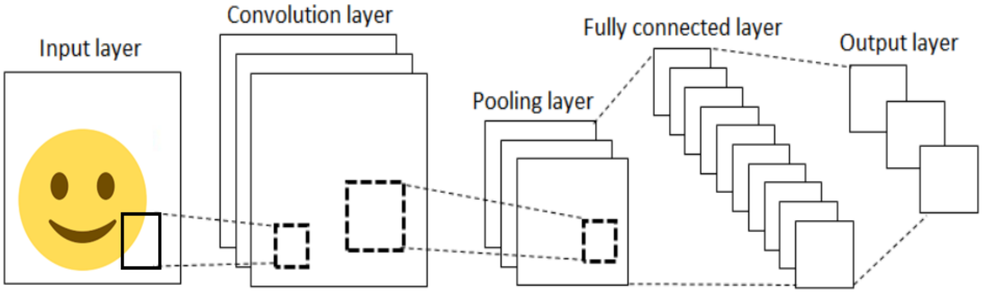


Figure 3: CNN Architecture used for Emotion Recognition

A Convolutional Neural Network (CNN) is a specialized deep learning neural network designed to process and analyze visual data, such as images and videos. The human visual system inspires it and is well-suited for tasks like image recognition, object detection, and image classification. CNNs have achieved remarkable success in various computer vision applications [17].

The key components of a CNN include Convolutional layers: These layers apply convolutional operations to the input data using a set of learnable filters (also known as kernels). Convolution helps to extract different features from the input images. As the network learns during training, these filters can detect patterns like edges, corners, textures, and other visual attributes [18].

Activation functions: After each convolutional operation, an activation function (e.g., ReLU - Rectified Linear Unit) is applied element-wise to introduce non-linearity in the network. This allows CNNs to model more complex relationships between input and output [19].

Pooling layers: Pooling layers downsample the spatial dimensions of the feature maps produced by the convolutional layers. They help reduce the network's computational complexity and make it more robust to spatial variations and translations in the input data [20].

Fully connected layers: Towards the end of the network, one or more fully connected layers combine the learned features and make predictions. These layers often lead to the final classification or regression output [21].

The typical CNN architecture consists of multiple stacked convolutional layers, pooling layers, and then fully connected layers. The deeper the network, the more complex features it can learn and the more parameters it has, requiring more data and computational resources for training [22].

Training a CNN involves feeding it a large dataset of labelled images (input-output pairs) and adjusting the network weights using optimization techniques (e.g., gradient descent) to minimize the difference between predicted outputs and true labels. Dataset preparation [23].

The Twitter dataset used in this study is made up of English tweets that were collected using the Twitter API. The Python library tweepy is used to gather the tweets [24]. This study extracts only English tweets using the Twitter API's language filtering feature by specifying the optional language parameter in the Twitter Search URL to 'en'. There are 16,011 tweets in the dataset overall, each with a label for one of the four emotion classes [25-29]. These classifications are represented numerically, with 1, 2, 3, and 4 designating the associated feelings of sadness, joy, infatuation, and rage. 75% of the data (12,008 tweets) are used as the test set to train the proposed CNN architecture, while the remaining 25% (4,003 tweets). A sample of tweets and their accompanying emotion class labels (ECs) are shown in Table 1.

Table 1: Some tweets and their corresponding categories [49]

Microblog data (emoticons with texts)	Emotion Category
Good morning 😁	2
Today is not my day 😭	1
😭 I can't handle this	1
He looks ginger lol 😡	3
😡 I don't need the vaccine	3
Coming home to this period 😍	4

3.4. Experimental setup

The text tokenization utility class from Keras, a potent open-source Python library, is used to transform the data words into numerical entities [30] for this multi-class classification problem in emotion recognition. Except under specific conditions, handling "Out of Vocabulary" (OOV) words is complete. This challenge employs the activation functions softmax and relu. Optimizer and loss functions are categorical-cross entropy and rmsprop, respectively [31-33]. As well as the data processing, the suggested CNN model is implemented using Python [34]. The examination is administered in a web-based data science environment, such as "www.kaggle.com." The suggested CNN training technique utilises batch sizes of 32, 64, and 128. An Intel(R) Core (TM) i7-7700 processor operating at 3.60 GHz and 16 GB of RAM are utilised in the experimental setup on a Windows 10 PC [35-37].

4. Results and Discussions

The performance of the emotion recognition model is analyzed using the parameters such as Precision, Recall and F1-score, as expressed in equations 1, 2, and 3 [38]. The performance of each category of emotion is individually analyzed. Table 2 illustrates the results obtained for emotion recognition [39-41]. The precision values obtained are 0.96, 0.92, 0.77, 0.90, 0.93, and 0.71 for the emotion categories 0, 1, 2, 3, 4 and 5, respectively. Similarly, the Recall and F1-Score values are obtained for the various categories of Emotion recognition.

$$\text{Precision} = TP / (TP + FP) \ldots\ldots. (1)$$

Recall = TP / (TP + FN)…….. (2)
F1 Score = 2 * (Precision * Recall) / (Precision + Recall…….. (3)

Table 2: Precision analysis for each emotion category based on numerical order

Emotion Category	Precision	Recall	F1-Score
0	0.96	0.91	0.93
1	0.92	0.91	0.92
2	0.77	0.81	0.79
3	0.90	0.97	0.93
4	0.93	0.81	0.86
5	0.71	0.83	0.77

The proposed CNN model has a substantial advantage because it uses real-world Twitter data and emoticons to recognise emotions [42]. The proposed CNN method only considers text data and assesses how emoticons affect emotion identification. Figure 4 shows the model accuracy graph of various training epochs. Figure 5 illustrates the loss function graph for various training epochs. The accuracies of the test set and training set and loss of the test set and training set are illustrated.

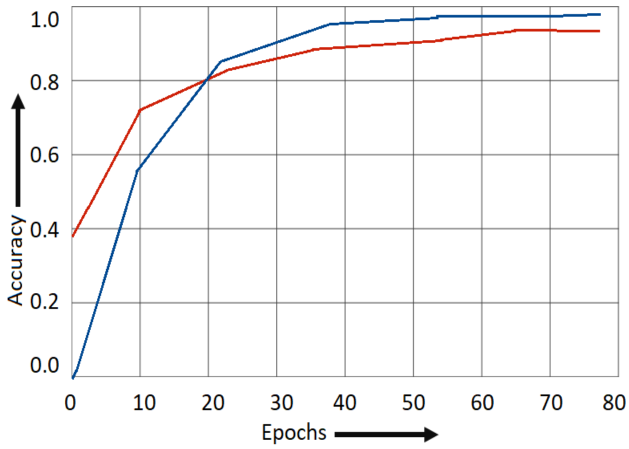


Figure 4: Model Accuracy graph

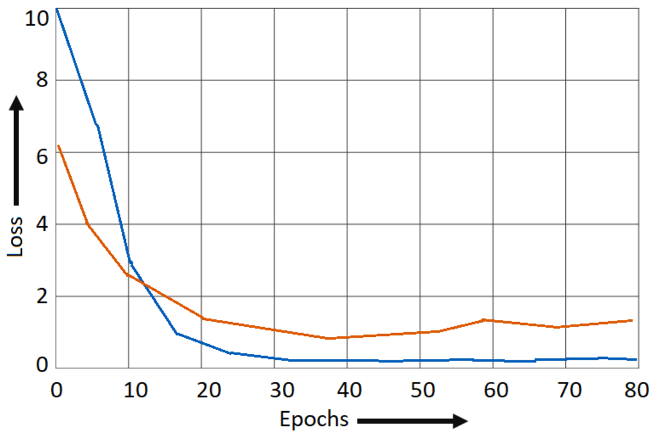


Figure 5: Loss function graph

Notably, the proposed CNN architecture beats the text-only scenario in terms of test accuracy while maintaining a consistent training set accuracy. The suggested CNN approach achieves a test accuracy of 39.9% for the text-only data in just ten epochs with a batch size of 128. In comparison, the approach obtains a test accuracy of 88.0% in under ten epochs when mixing emoticons and text with a batch size of 32. Higher test set accuracy is preferable in every machine learning system because it shows how well the system generalises. The test set's improved precision shows that the suggested CNN model's capacity to reliably identify emotions has improved due to including emoticons alongside text.

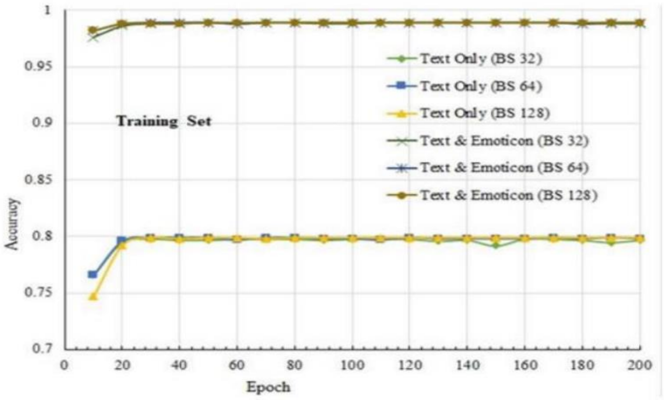
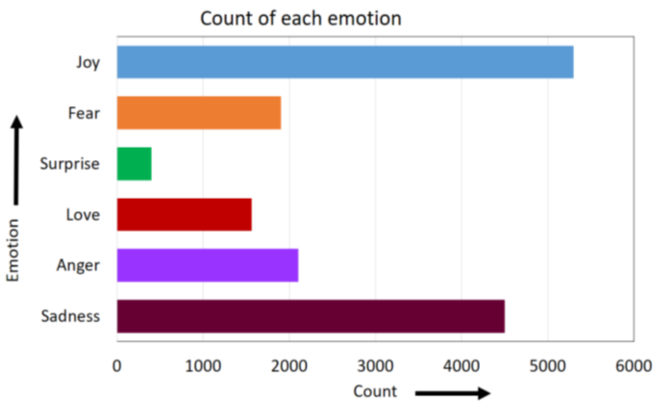
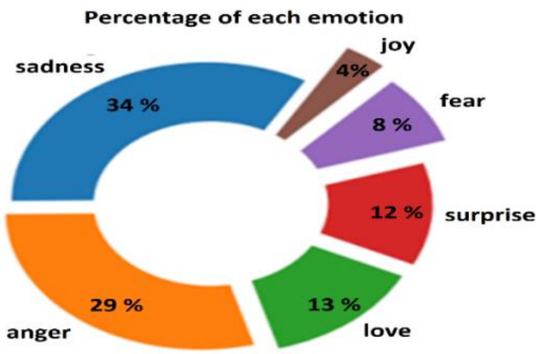


Figure 6: Training set accuracy plot (BS – Batch size) [49]

Figure 6 shows the training accuracy plot for different batch sizes. The suggested strategy in Figure 6 achieved more accuracy for emoticon and text data than text-only accuracy. The CNN approach is recommended to perform best in the "Angry" category when both text and emoticon data are combined, correctly categorizing 890 cases. In comparison, the text-only instance correctly categorizes 429 out of 1000 cases in the "Sad" category, which is where it performs best. Other performance evaluation criteria can be established using the provided confusion matrices for the two scenarios of the suggested CNN strategy. Figure. 7 (a) and (b) illustrates the count of each emotion category and the percentage division of each emotion category.

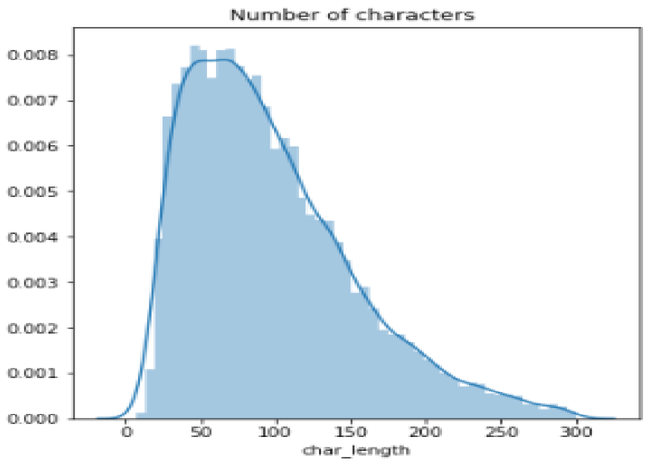


(a)

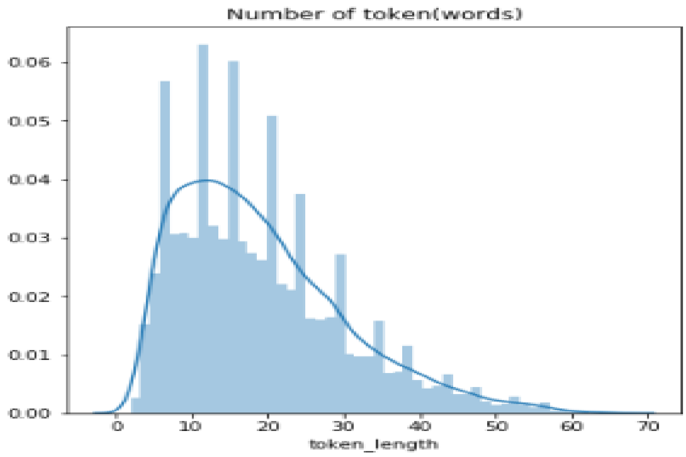


(b)

Figure 7: (a) Count of each emotion (b) Percentage division of each emotion



(a)



(b)

Figure 8: (a) Characters input graph (b) tokenization input graph

Figure 8 (a) and (b) demonstrate the characters and tokenization input graph. The techniques and the corresponding dataset sizes are contrasted with the proposed CNN strategy for emotion categorization in Twitter data. 16012 samples total, with 12009 tweets utilised for training and 4003 for testing, make up the suggested approach. Khan [43], achieved 72.3% accuracy with Naive Bayes training using just 116 out of 268 tweets. For a 72.6% accuracy rate, Khan, [44] used 4600 tweets and CNN. In order to get 61.3% accuracy, Gadhavi, et al. [45] used 19,678 tweets with CNN and BiLSTM, but they left out the train-test split ratio. For 82.1% accuracy, Kumar et al., [46] used 2313 tweets with LSTM. Nevertheless, Pattana-Anake and Joseph [47] employed the Bi-GRU architecture on the Sina Weibo NLPCC2013 and NLPCC2014 datasets, each with 15,000 and 10,000 sentences, and achieved accuracy rates of 85.76% and 86.35%, respectively. Even though there were only 5400 emotive sentences in each sample, the proposed model outperformed all earlier ones with an accuracy rate of 88.0% on the test set [48]. The suggested method accurately classifies emotions in microblog data while considering both sentences and emoticons using CNN Conv1D architecture, which is suitable for time series data. Overall, this strategy outperforms and significantly contributes to earlier approaches.

5. Conclusion

The study's conclusion indicates that social media is the dominant platform for expressing emotions, often utilizing emoticons to enhance politeness in texts. The researchers have made a significant breakthrough in emotion recognition from microblog data using a CNN-based model incorporating text and emoticons. Unlike the prevailing methods solely relying on text, the proposed approach with emoticons outperforms earlier algorithms, emphasizing the significant impact of emoticons on people's emotional behaviour. In summary, this study successfully developed an emotion identification model and evaluated the effectiveness of emoticons for emotion recognition in microblog data. The work also opens up new avenues for further research, leading to substantial growth in English microblog emotion recognition and possibly extending its applicability to other languages. Future investigations should explore the inclusion of additional emotions like surprise and disgust to enhance the realism of the recognition system. Moreover, utilizing larger datasets may improve the accuracy of the CNN approach, as suggested.

Acknowledgment: I am deeply grateful to my coauthors for their expertise and dedication, greatly enriching this work. Special thanks to my friends for their unwavering support and encouragement throughout the research process.

Data Availability Statement: The data for this study can be made available upon request to the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding

Conflicts of Interest Statement: The authors have no conflicts of interest to declare. This work represents a new contribution by the authors, and all citations and references are appropriately included based on the information utilized.

Ethics and Consent Statement: This research adheres to ethical guidelines, obtaining informed consent from all participants. Confidentiality measures were implemented to safeguard participant privacy.

References

1. Y.-J. Lee and H.-J. Choi, "Comparative study of emotion annotation approaches in Korean dialogue," in 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea (South), pp. 354-357, 2021.
2. S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "MELD: A multimodal multi-party dataset for emotion recognition in conversations," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 527–536, 2019.
3. B. Alharbi, H. Alamro, M. Alshehri, Z. Khayyat, M. Kalkatawi, I. I. Jaber, and X. Zhang, ASAD: A Twitter-based benchmark Arabic sentiment analysis dataset, 2021 arXiv: https://doi.org/10.48550/arXiv.2011.00578.
4. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, Dec 2017, arXiv: https://doi.org/10.48550/arXiv.1706.03762
5. C. D. Katsis, Y. Goletsis, G. Rigas, and D. I. Fotiadis, "A wearable system for the affective monitoring of car racing drivers during simulated conditions," Transp. Res. Part C Emerg. Technol., vol. 19, no. 3, pp. 541–551, 2011.
6. Z. Shen, J. Wang, Z. Pan, Y. Li, and J. Wang, Cross attention-guided dense network for images fusion, Aug. 2021, arXiv: https://doi.org/10.48550/arXiv.2109.11393.
7. D. Hazarika, S. Poria, R. Mihalcea, E. Cambria, and R. Zimmermann, "ICON: Interactive conversational memory network for multimodal emotion detection," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.
8. D. Hazarika, S. Poria, A. Zadeh, E. Cambria, L.-P. Morency, and R. Zimmermann, "Conversational memory network for emotion recognition in dyadic dialogue videos," Proc. Conf., vol. 2018, pp. 2122–2132, 2018.
9. M. Abdul-Mageed and L. Ungar, "EmoNet: Fine-grained emotion detection with gated recurrent neural networks," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers, 2017.
10. H. Coolican et al., "Reviews: Response to David booth's review of research methods and statistics in Psychology (4th edition), developmental psychology in action" The Sage handbook of social psychology: Concise student edition, psychology and the life sciences (4th edition)," Psychol. Learn. Teach., vol. 7, no. 2, pp. 46–58, 2008.
11. A. Collins and D. G. Bobrow, "Representation and Understanding: Studies in Cognitive Science," Audiovisual emotion recognition using big data towards, G, Mobile, 1975.
12. A. Kashii, K. Takashio, and H. Tokuda, "Ex-amp robot: Expressive robotic avatar with multimodal emotion detection to enhance communication of users with motor disabilities," in 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2017.
13. J. Herzig, D. Konopnicki, T. Sandbank, and M. Shmueli-Scheuer, Emotion detection and expression integration in dialog systems, U.S. Patent: US20190286705. 2019.
14. S. Hareli and A. Rafaeli, "Emotion cycles: On the social influence of emotion in organizations," Res. Organ. Behav., vol. 28, pp. 35–59, 2008.
15. A. Kumar, O. Irsoy, P. Ondruska, M. Iyyer, J. Bradbury, I. Gulrajani, V. Zhong, R. Paulus, and R. Socher, Ask me anything: Dynamic memory networks for natural language processing, in Proc. Int. Conf. March 2015.
16. R. Bose and S. Kumar, "Hand gesture recognition using faster R-CNN inception V2 model," in Proceedings of the Advances in Robotics 2019, 2019.

17. A. A. Alarood, E. Alsolami, M. A. Al-Khasawneh, N. Ababneh, and W. Elmedany, “IES: Hyper-chaotic plain image encryption scheme using improved shuffled confusion-diffusion,” Ain Shams Eng. J., vol. 13, no. 3, p. 101583, 2022.
18. A. J. Joseph, F. J. Joseph, O. M. Stanislaus, and D. Das, “Classification methodologies in healthcare,” in Evolving Predictive Analytics in Healthcare: New AI techniques for real-time interventions, pp. 55–73, 2022.
19. A. Khan et al., “Optimal policy learning for disease prevention using reinforcement learning,” Scientific Programming, vol. 2020, pp. 1–13, 2020.
20. D. D. Pandya, A. Jadeja, S. Degadwala, and D. Vyas, “Ensemble Learning based Enzyme Family Classification using n-gram Feature,” in 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1386–1392, 2022.
21. F. J. J. Joseph and Thai-Nichi International College, Thai-Nichi Institute of Technology, Bangkok, Thailand, “Time series forecast of Covid 19 Pandemic Using Auto Recurrent Linear Regression,” J. Eng. Res., 2022.
22. Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabduallah, B. I., Siddiqui, T., Albahlal, B. M., et al. (2023). DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. Diagnostics, 13(6), 1093.
23. H. Gupta, D. Patel, A. Makade, K. Gupta, O. P. Vyas, and A. Puliafito, “Risk prediction in the life insurance industry using federated learning approach,” in 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON), pp. 948–953, 2022.
24. J. Joseph, “IoT Based Aquarium Water Quality Monitoring and Predictive Analytics Using Parameter Optimized Stack LSTM,” in 2022 International Conference on Information Technology (InCIT), IEEE, 2022.
25. J. Mahale, S. Degadwala, and D. Vyas, “Crop prediction system based on soil and weather characteristics,” in 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2022.
26. S. Khan, “How Data Mining Can Help Curb City Crime,” International Journal of Control Theory and Applications (IJCTA), vol. 9, no. 23, pp. 483–488, 2016.
27. M. A. Al-Khasawneh, S. M. Shamsuddin, S. Hasan, and A. A. Bakar, “MapReduce a Comprehensive Review,” in 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), pp.1-6, 2018.
28. M. A. Saleh, S. Hajar Othman, A. Al-Dhaqm, and M. A. Al-Khasawneh, “Common investigation process model for internet of things forensics,” in 2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE), pp. 84-89, 2021.
29. M. Alajmi and S. Khan, Mobile Community Networks Information Investigation for Additional Significance. 2013.
30. M. Alajmi and S. Khan, Part of Ajax And Openajax In Cutting Edge Rich Application Advancement For E-Learning. 2015.
31. M. I. Uddin, S. A. Ali Shah, M. A. Al-Khasawneh, A. A. Alarood, and E. Alsolami, “Optimal policy learning for COVID-19 prevention using reinforcement learning,” J. Inf. Sci., vol. 48, no. 3, pp. 336–348, 2022.
32. M. W. Sabir et al., “Segmentation of liver tumor in CT scan using ResU-Net,” Appl. Sci. (Basel), vol. 12, no. 17, p. 8650, 2022.
33. N. Mast et al., “Channel Contention-Based Routing protocol for wireless ad hoc networks,” Complexity, vol. 2021, pp. 1–10, 2021.
34. N. Srisook, O. Tuntoolavest, P. Danphitsanuparn, V. Pattana-Anake, and J. Joseph, “Convolutional Neural Network Based Nutrient Deficiency Classification in Leaves of Elaeis guineensis Jacq,” International Journal of Computer Information Systems and Industrial Management Applications, vol. 14, pp. 19–27, 2022.
35. P. Chopra, N. Junath, S. K. Singh, S. Khan, R. Sugumar, and M. Bhowmick, “Cyclic GAN model to classify breast cancer data for pathological healthcare task,” Biomed Res. Int., vol. 2022, p. 6336700, 2022.
36. R. Rani et al., “Towards green computing oriented security: A Lightweight Postquantum signature for IoE,” Sensors (Basel), vol. 21, no. 5, p. 1883, 2021.
37. S. Bhoumik, S. Chatterjee, A. Sarkar, A. Kumar, and F. J. John Joseph, “Covid 19 prediction from X ray images using fully connected convolutional neural network,” in CSBio ’20: Proceedings of the Eleventh International Conference on Computational Systems-Biology and Bioinformatics, pp. 106–107, 2020.
38. S. Dave, S. Degadwala, and D. Vyas, “DDoS detection at fog layer in internet of things,” in 2022 International Conference on Edge Computing and Applications (ICECAA), pp. 610–617, 2022.
39. S. Dhanush, S. C. Mohanraj, V. S. Sruthi, F. J. Cloudin, and J. Joseph, “CODEDJ-Private Permissioned Blockchain Based Digital Wallet with Enhanced Security,” in IEEE International Conference on Bio-Neuro Informatics Models and Algorithms, IEEE, 2022.
40. S. Khan and A. Alfaifi, “Modeling of Coronavirus Behavior to Predict it’s Spread,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 5, 2020.
41. S. Khan, “Artificial Intelligence Virtual Assistants (Chatbots) are Innovative Investigators,” International Journal of Computer Science Network Security, vol. 20, no. 2, pp. 93–98, 2020.
42. S. Khan, “Data visualization to explore the countries dataset for pattern creation,” Int. J. Onl. Eng., vol. 17, no. 13, pp. 4–19, 2021.

43. S. Khan, “Development of Arabic evaluations in information retrieval,” Int. J. Adv. Appl. Sci., vol. 6, no. 12, pp. 92–98, 2019.
44. S. Khan, “Study Factors for Student Performance Applying Data Mining Regression Model Approach,” International Journal of Computer Science Network Security, vol. 21, no. 2, pp. 188–192, 2021.
45. V. B. Gadhavi, S. Degadwala, and D. Vyas, “Transfer learning approach for recognizing natural disasters video,” in 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 793–798, 2022.
46. V. Kumar et al., “Grouping and sponsoring centric green coverage model for Internet of Things,” Sensors (Basel), vol. 21, no. 12, p. 3948, 2021.
47. V. Pattana-Anake and F. J. J. Joseph, “Hyper parameter optimization of stack LSTM based regression for PM 2.5 data in Bangkok,” in 2022 7th International Conference on Business and Industrial Research (ICBIR), 2022.
48. Z. Ullah et al., “Certificateless proxy reencryption scheme (CPRES) based on hyperelliptic curve for access control in content-centric network (CCN),” Mobile Information Systems, vol. 2020, pp. 1–13, 2020.
49. M. A. Habib, M. A. H. Akhand, and M. A. S. Kamal, “Emotion recognition from microblog managing emoticon with text and classifying using 1D CNN,” J. Comput. Sci., vol. 18, no. 12, pp. 1170–1178, 2022.