IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3692~3702
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3692-3702

Text emotion recognition based on deep learning and attention with whale optimization algorithm

Tabarek Alwan Tuib[1,2], Mohammad-Reza Feizi-Derakhshi[1], Yaqdhan Mahmood Hussein[2,3], Fahad Taha Al-Dhief[4]
[1]Computerized Intelligence Systems Laboratory, Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
[2]College of Engineering Electronics and Communication Engineering, Al-Muthanna University, Samawah, Iraq
[3]College of Engineering, Al Ayen University, Nasiriyah, Iraq
[4]Center for Software Technology and Management, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia

Article Info

Article history:

Received Dec 27, 2024
Revised Apr 26, 2026
Accepted Jul 9, 2026

Keywords:

Bidirectional gated recurrent unit
Convolutional neural network
Gated recurrent unit
Text emotion recognition
Whale optimization algorithm

ABSTRACT

Text emotion recognition (TER) represents one of the crucial tasks in natural language processing (NLP), and it is highly important for many applications. To improve the performance of TER, advanced deep learning (DL) model GRU-BiGRU-CNN-Att is presented in this research. It is a combination of gated recurrent units (GRU), bidirectional gated recurrent units (BiGRU) with a convolutional neural network (CNN), and an attention mechanism. Using a GRU, BiGRU, and CNN as part of a deep feature extraction model, combined with an attention mechanism that gives important information different weights, the method that is proposed in the present work enhances the quality of the word vectors and leads to increasing sentiment analysis judgment accuracy. The whale optimization algorithm (WOA), which selects the most informative features for further enhancing the model, is used lastly for the optimization of the feature selection process. After such optimization, those chosen features are trained on a multi-layer perceptron (MLP), successfully combining machine learning (ML) and DL approaches for improving TER. A varied corpus of tweets, sentences, and dialogues had been used in the present work for the assessment of the performance of the suggested model. The proposed method achieves 83.76% accuracy in emotion recognition.

Corresponding Author:

Fahad Taha AL-Dhief
Center for Software Technology and Management, Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Bangi 43600, Selangor, Malaysia
Email: fahad@ukm.edu.my

1. INTRODUCTION

Text emotion recognition (TER) is an essential and rapidly developing activity in the sphere of natural language processing (NLP). Its significance can be seen in its wide range of applications in different applications, like social network analysis, where it can be used to gauge public sentiment from extensive online discourse; mental health assessment, delivering insights into emotional states based on textual analysis; and public opinion monitoring, where policymakers and businesses can be given the necessary feedback [1]–[3]. Regardless of its importance, the interpretation of emotions in a text is one of the challenging tasks, because of the challenging nature of the framework behind human language and expression. The recent fast and unceasing development of digital communication technology has led to the rise in interest in the TER subject. The TER will be able to analyze the emotional content of the document, giving it an advanced insight into its sentiment, intent, and expressive subtleties. The technology has been applied in many sectors such as education, communications, and business, among others [4].

Human emotions can be categorized [5] the emotions of humans into six namely joy, surprise, sadness, disgust, fear, and anger. There are many complicated forms of emotions such as optimism and love. The dimensional approach was inspired by the theory described in [6], where feelings (not to be confused with emotions) could be described in three dimensions: excitement-inhibition, tension-relaxation, and pleasantness-unpleasantness. A prominent example is Plutchik’s wheel of emotions [7]. Plutchik believed that eight fundamental emotions are present in everyone and modeled them in pairs and placed them on a wheel, each opposite to the other: anticipation and surprise, anger and fear, sadness and joy, and disgust and trust. The eight emotions are methodically organized depending on their physiological functions. An example of a conceptual analogy of the wheel of emotions by Plutchik is a cone that unfolds to create a 2D wheel. According to Plutchik's theory, those basic emotions may become stronger, weaker, or combined to form a broad spectrum of emotions. The separation between emotions, specifically the secondary, primary, and tertiary dyads indicated on the wheel of emotions, is relative to each other.

– Primary dyads: combine adjacent emotions (e.g., joy and trust).
– Secondary dyads: combine emotions with one intervening emotion (e.g. joy and fear, with trust in between).
– Tertiary dyads: combine emotions separated by two other emotions (e.g., joy and surprise, with trust and fear in between).

Such dyads explain the complex interrelations between emotions, and this gives an organized approach to learning the systematic interconnections between them on the wheel. This model provides a subtle explanation of the interaction and transformation of emotions in the human emotional spectrum. Although emotions are usually conveyed through facial expression, gestures, speech, and text, the latter does not possess the expressive elements inherent in spoken language and non-verbal communication [8]. This fact complicates text-based emotion recognition and makes it especially difficult. Text, by definition, does not have any facial or tonal indicators (tasteless), which makes the situation of identifying the emotional states more difficult. Besides that, the complexity and ambiguity of the inherent textual communication contribute to the augmentation of these issues, because the words might have a wide variety of meanings and morphological forms, which will lead to complicating the process of recognizing emotions in written text. The TER relies on the use of NLP. NLP is a dynamic field of study that is urged by the increasing amount of online textual materials. The preparation of the text is the first stage of the TER, it involves word breaking, part of speech tagging, meaning as well as the structure of the sentences. The preprocessing will be followed by the feature extraction stage, whereby the word vectors are quantified, formatted for computational processing, and fed into the classification systems aiming at identifying the emotional states [4], [9]. In the case of feature extraction, some common algorithms are applied, such as one-hot encoding where a binary representation of the types of texts is obtained [4], [9], and the approach of term frequency-inverse document frequency (TF-IDF), where the word relevance is established based on its frequency to end of having a convenient and simplistic textual representation [10].

Designing reliable classification algorithms and efficiently selecting and weighting textual data present challenges that extend beyond just the textual feature extraction. The previous models in the area of TER had faced a wide range of limitations. Usually, they lacked a comprehensive taxonomy of emotions and have been built on lexicons with insufficient kinds of vocabulary. Many disregarded semantic contexts, failed to extract contextual information effectively from sentences, and performed poorly in detecting specific emotions. Additionally, these models had trouble pulling out important meaning from the text, were slow to process information, missed connections between different features, had too little data, and often made mistakes in classifying emotions. Handling of commonly appearing emojis, poor extraction of semantic information, and failure to adapt to different sentence structures were also other major concerns. Although some of these gaps have been addressed in previous studies, current methodologies often face challenges in collecting complex emotional patterns as well as improving feature representation effectively and limit performance. In this paper, a more advanced deep learning (DL) architecture, GRU-BiGRU-CNN-Att is proposed to have a significant effect on the accuracy and stability of TER. The combination of gated recurrent units (GRU), bidirectional gated recurrent units (BiGRU), a convolutional neural network (CNN), and an attention mechanism (AM).

This research offers several key contributions. First, propose a hybrid DL architecture, GRU-BiGRU-CNN-Att, which integrates GRU and BiGRU with a CNN to extract deep features from text effectively. An attention mechanism is incorporated, allowing the model to assign varying weights to crucial textual information, thereby enhancing word vector quality and sentiment analysis accuracy. Second, the use of whale optimization algorithm (WOA) for feature optimization marks a first in this context, identifying key features to improve model predictions. Third, the strengths of the conventional machine learning (ML) and

DL methods are then merged through training optimized features with (MLP) to achieve better results in terms of emotional recognition in text. Fourth, the suggested model works well when it is applied to text-based emotion recognition, though there is no vocal tone and face expressions. The model is applicable to use on different platforms because three different datasets (simple sentences, dialogues, and Tweets) enhance robustness and generalizability over other previous research studies that used a single dataset only. This adaptability is useful in implementation like the analysis of customer reviews and enhancing social media user security.

The rest of this research paper is structured as follows: section 2 is the review of the related literature about emotion detection. Section 3 describes the proposed method and the computational methods to overcome the previous limitations. Section 4 discusses and presents the results of the experiment in all of the datasets. Lastly, section 5 also provides the conclusion of the research and recommends future research directions.

2. LITERATURE REVIEW

TER has utilized different methods with the aim of improving the emotion recognition process of textual information [4]. One of the key goals of such studies is to find models that best identify emotions. A keyword-based approach was explored in [11] using the ISEAR dataset [12] and focused on the phrasal verbs. They created a specific database following preprocessing the data and finding many phrasal verbs linked to emotions that had not been previously indexed. They achieved a noteworthy 65% accuracy rate by using this method to classify phrasal verbs and synonymous terms by emotion. Nevertheless, their approach was constrained by a dearth of emotion-specific keywords and a disregard for the semantic depth of word meanings. A learning-based method was developed in [13] using the ISEAR dataset [12] to create a learning-based method. After comparing several classifiers, they concluded that logistic regression performed better than extreme gradient boosting (XGBoost), k-nearest neighbor (KNN), and support vector machines (SVMs). The study also suggested that DL methods may be used to achieve further refinements.

The Emo2Vec approach, which encodes emotional meanings into vector forms, was first presented in [14]. They used both larger and smaller datasets, such as WASSA and ISEAR, for training Emo2Vec inside a multitask learning framework. Their results outperformed more conventional models such as CNN as well as deep moji embeddings, and their applications ranged from stress detection to sarcasm classification. Emo2Vec produced competitive results when paired with GloVe and logistic regression.

Emotion detection from textual conversations was investigated on emotion detection from textual conversations using an attention-based learning model [15]. Their architecture used tokenization, bidirectional long short-term memory (Bi-LSTM) units, self-attention, and dense classification layers, achieving a micro-F1 score of 75.82%. A two-step approach was proposed a two-step approach for emotion detection in Twitter data by combining a rule-based method with a supervised naive Bayes classifier. The rule-based and machine-learning approaches achieved accuracies of approximately 85% and 88%, respectively [16]. A supervised learning framework was developed a supervised learning framework for automatic emotion classification in textual streams. The approach combines an offline training stage with an online classification stage for tracking emotions in real-time Twitter data [17]. Emotion analysis techniques have also been applied emotion analysis techniques with the aim of identifying hate speech on social media, with a focus on the analysis of unstructured data of posts that were intended to spread hatred [18].

Machine-learning and deep-learning approaches were reviewed machine-learning and deep learning approaches for emotion analysis in textual data and discussed the main challenges associated with text-based emotion detection [19]. Text-based emotion detection approaches were surveyed text-based emotion detection and highlighted main approaches adopted in design of text-based emotion-detection systems [20]. A process for creating an emotion lexicon was described a process for creating an emotion lexicon enriched with word-level emotional intensity to improve emotion analysis in texts [21]. Twitter data were used Twitter data to identify emotions and sentiments and employed the resulting emotion and sentiment information to generate generalized and personalized recommendations based on users' Twitter activity [22].

3. METHOD

This section describes a hybrid architecture for improving TER, illustrated in Figure 1, which combines DL and ML to leverage the strengths of both. The pipeline starts with data collection and pre-processing to clean and prepare raw text data for training. Then several DL models will be trained and tested, and the three best models are chosen according to the performance metrics, including accuracy and F1-score. These best DL models are then combined (summed) to create a single latent feature vector, which captures rich, multidimensional emotional information from text. This deep latent vector is then passed over to the most performing ML classifier, pairing strong DL feature extraction with the stability and interpretability of conventional classification methods. Lastly, the performance of the hybrid system is compared to the performance of individual DL and ML models for the purpose of determining which configuration achieves the highest overall accuracy, addressing the weaknesses of the standalone methods as well as delivering a more robust TER system.

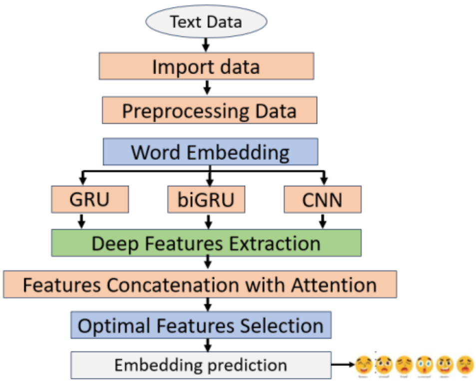

Figure 1. Suggested model for detecting emotions from text

3.1. Dataset description

The dataset to be utilized in the present study was compiled on three sources WASSA [23], ISEAR [12], and Emotion-stimulus [24], where each one offered text and emotion labels in different formats (dialogues, tweets, and sentences). ISEAR database was developed following a global effort of psychologists in the 1990s and contains reports of approximately 3000 participants in 37 nations with seven fundamental emotions (anger, joy, sadness, guilt, fear, disgust, and shame) [12]. The data was originally categorized into 173 emotions and later consolidated into seven main types: sadness, fear, joy, anger, surprise, disgust, and shame. In addition, the dataset includes Emotion-stimulus statements, where the "cause" subset contains 820 sentences annotated with both an emotion label as well as its triggering cause, while the "no cause" subset includes 1594 sentences labeled only with emotions [24], as summarized in Table 1. This study combined three datasets (text sentences, dialogues, and tweets) into a single corpus of over 14,500 English text samples. Each sample was labeled with one of six emotions—disgust, joy, fear, sadness, surprise, and anger—based on syntactic and semantic polarity [5], [23]. As shown in Figure 2, these emotions are formally defined, and the text may include extra punctuation and emoticons. The combined corpus exhibits an uneven distribution across the emotion classes, which should be considered because class imbalance can influence classification performance [25]. The final dataset contains only the sentences and their emotion labels, and it has been split to training and testing sets using a ratio of 80:20.

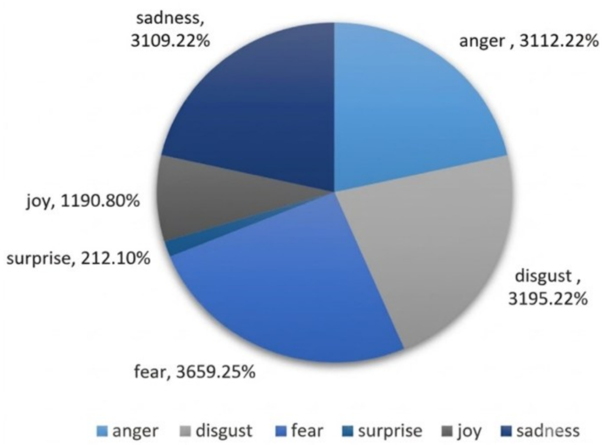

Figure 2. Six types of emotions in this work

Table 1. Description of each dataset individually

Dataset	Granularity	No. of emotions	Size	Description
ISEAR	Sentences	7 emotions	7666	Studied in 37 countries
WASSA	Tweets	4 emotions	4334	Tweets
Emotion-stimulus	Dialogs	7 emotions	2500	—

3.2. Data pre-processing

One of the essential processes in data mining is preprocessing, which is used to eliminate unnecessary texts and symbols in raw data to make them cleaner, easier to interpret, and more efficient for analysis. Due to the fact that the data is collected in various sources, it is to be integrated and formatted prior to further processing to achieve accurate findings. In this study, preprocessing was tailored to the dataset and included stop-word removal, tokenization, stemming, emoji-to-text conversion, and lemmatization using NLTK tools. As part of preprocessing, data cleaning was applied to eliminate noisy text (e.g., superfluous symbols and words) that does not contribute to model learning [9]. After cleaning, the resulting sentences were tokenized and used as model inputs, then split to training and testing sets and evaluated with multiple ML and DL algorithms to improve performance, especially since the datasets were drawn from three repositories (WASSA, ISEAR, and Emotion-stimulus).

3.3. Feature extraction

In order to speed up processing, feature extraction reduces dimensionality by dividing large amounts of raw data into smaller categories [10]. This process converts the preprocessed textual data into numerical representations that can be efficiently processed by the machine learning and deep learning models. In this study, TF-IDF and Word2vec are used to represent term importance and semantic relationships, respectively, as explained in the following subsections.

3.3.1. Term frequency-inverse document frequency

By combining term frequency with inverse document frequency, the TF-IDF method finds the terms that are utilized most often in a document as well as those that are utilized infrequently throughout the corpus. This method facilitates the process of choosing a unique term vector to act as a feature set for the training of the model. In this work, text is converted to feature vectors with the use of TF-IDF vectorizer, which allows them to be used as input for model estimators. A dictionary-like vocabulary is created, with each token (word) being mapped to a feature index according to how frequently it appears in the matrix. As a result, a feature index is given to every distinct token. The TF-IDF weight calculation is expressed in (1).

$$\mathrm{W(d,t) = TF(d,t) * \log\left(\frac{N}{df(t)}\right)} \quad (1)$$

In which d denotes documents, t denotes terms in the document, and N represents the total number of documents.

3.3.2. Word embeddings

Various methods of word embedding have become common; they are GloVe, Word2vec, fastText, and ELMo, although this study is based on Word2vec. Word2vec is trained on a large text corpus to learn word associations via a neural network and presents words (and sentences) as numerical vectors, which can be used to capture semantic relationships [26]. Due to the variation in sentence length, padding is used for standardizing input size, allowing the model to process text uniformly, where a common length of sequence is 50 tokens. The study creates an embedding matrix of the vocabulary terms (size-18210, 300) into a 300-dimensional continuous vector space with a pre-trained embedding to enhance semantic representation and overall model performance.

3.4. Deep learning models

Artificial intelligence (AI) systems that enable ML as well as evolution without the need for explicit programming are known as DL models [27]. For training the model as well as predicting emotions on the test dataset, the preprocessed training dataset is fed into a number of components in this research, such as TF-IDF transformer, CountVectorizer, and MLClassifier. The study makes use of pre-built DL and ML models. An MLP is used in ML for predicting emotions. On the other hand, DL uses architectures like BiGRU, CNN, and GRU for emotion prediction tasks [27], [28]. The three DL components used in the proposed framework are described in the following subsections.

3.4.1. Gated recurrent unit

The vanishing gradient problem that traditional recurrent neural networks (RNNs) face is resolved by the GRU [29]. Since they function similarly in various situations and share structural similarities with long short-term memory (LSTM) networks, GRUs are frequently regarded as LSTM variants. GRU is implemented as a single layer in the suggested model. The embedding layer of dimensions (18210, 300) is used as input for GRU model after feature extraction. The GRU model then uses the training vector as input for the prediction of emotions throughout the dataset.

3.4.2. Bidirectional gated recurrent unit

A sequence processing paradigm known as BiGRU [30] consists of two GRUs running simultaneously, one backward and one forward. The output and input gates are the only ones utilized in such a bidirectional RNN. BiGRU is implemented as a single layer in our suggested model. The embedding layer of dimensions (18210, 300) is used as input for BiGRU model after feature extraction. BiGRU model then uses the training vector as input for predicting emotions throughout the dataset.

3.4.3. Convolutional neural network

In the context of DL, the CNN is a deep neural network (DNN) architecture that can also be applied to text classification [31]. The CNN will form one layer in the proposed model. The CNN model consumes an embedding layer of parameters (18210, 300) once the features are extracted, and then it takes the training vector to predict the emotions horizontally in a dataset.

3.5. Multi-head attention

Multi-head attention represents a core component of modern DL—especially transformer-based NLP models—because it can focus on different parts of an input sequence in parallel and integrate information across representation layers. This makes it extremely efficient in the capture of the complex dependencies and long-range relationships in tasks like text classification and machine translation [32]. Multi-head attention is applied in this work to combine the deep features extracted from three neural architectures, namely GRU, CNN, and BiGRUs, that captures the various elements of sequential text, including local patterns, temporal dependencies, and bidirectional context. Both of these complementary representations give the model a deeper understanding of the input and enhance its capacity to identify both global and local contextual cues, resulting in improved NLP performance. The other benefit is its computational efficiency: multi-head attention is capable of modeling complex patterns and processing multiple input components at once, helping to address long-range dependency and decreasing issues such as vanishing gradients associated with deep networks. Given that attention is computed in parallel with several heads, training and inference are more scalable and faster, hence multi-head attention is central at the core of state-of-the-art NLP systems [32].

3.6. Feature selection with whale optimization algorithm

The WOA is a metaheuristic optimization method, which has been based on the cooperative hunting behavior of the humpback whales. It searches for optimal solutions through iteratively updating a population of candidate solutions, balancing exploration and exploitation with encircling prey and bubble-net feeding techniques where promising regions of the search space [33]. Optimization-based feature-selection methods have been applied to textual and sentiment-analysis tasks to retain informative features and reduce redundant attributes [34], [35]. The WOA model is used in this study to find the most useful deep features proposed by multi-head attention with the purpose of preserving informative features and losing redundant or irrelevant features [34], [35]. This reduces the feature-space size and improves the efficiency, interpretability, and performance of later learning tasks. Through the combination of WOA with multi-head attention, the method enables a thorough search of high-dimensional feature spaces—common in DL—while remaining scalable and adaptable, helping produce more robust and generalizable feature representations [33].

3.7. Hybrid model

As seen in Figure 3, the suggested hybrid model combines DL algorithms for predicting emotions. BiGRU and CNN are examples of DL components, whereas an MLP classifier is used in ML [28]. Word2vec is used to transmit the input datasets via the word embedding layer at the beginning of the procedure. The resultant embedding vector is then sent into CNN and BiGRU, two DL algorithms. The BiGRU and CNN models only serve as encoders because the last layer was eliminated [28]. Furthermore, for the provided input embedding vector, both of such encoders produce a latent vector. These latent vectors are then sent into the MLP classifier after being concatenated. The emotion regarding the input words is after that predicted by the MLP classifier. Because classification models from both the DL and ML domains were carefully chosen, the suggested hybrid model shows better results in terms of F1-score and accuracy. BiGRU and CNN, two DL model results, performed well individually. In the same way, MLP stood out as a strong performer among ML algorithms [36]. To improve overall performance, the hybrid model integrates the top classifiers from the two categories. The best DL models are combined to create the hybrid model, which produced the highest accuracy as well as F1-score. The top-performing ML models, which are renowned for their high accuracy in emotion prediction tasks, were then fed the latent vector that was extracted from such models [37]. Figure 3 shows how the DL and ML models are combined to produce better results by utilizing the advantages of the two fields.

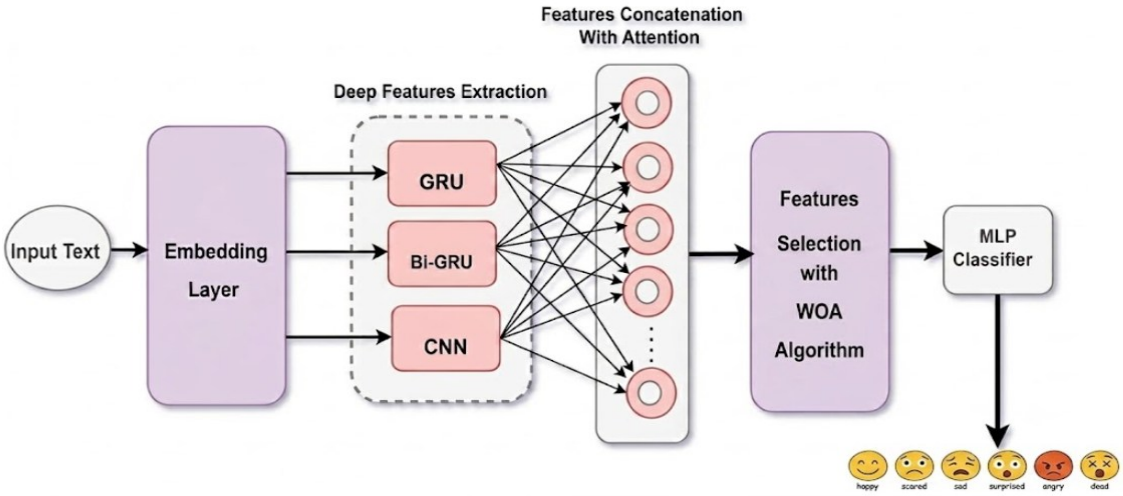

Figure 3. CNN model for detecting emotions from text

4. RESULTS AND DISCUSSION

To understand the best performance of the suggested hybrid model, a thorough sequence of experiments with the highest number of DL models and testing approaches was performed. The analysis was based on commonly used binary measures in classification: precision, recall, F1-score, and accuracy built on the confusion matrix [38]. The metrics were calculated by evaluating each of the classes of emotion separately and then getting the macro-averaged accuracy of overall model performance across all classes. To this end, the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) were harvested by matching model-generated predictions against the expert-annotated ground truth on each of six distinct emotion categories. The calculation of the items of performance was made by (2) to (5).

$$Precision = \frac{truely\ positive\ classifications}{all\ positive\ classifications} = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{truely\ positive\ classifications}{all\ positive\ examples} = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times (TP+FP+FN)} \tag{4}$$

$$Accuracy = \frac{true\ classifications}{all\ classifications} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

The testing was done on a multi-domain textual dataset that consists of a mixture of tweets, formal sentences, and dialogues. This dataset includes ISEAR, WASSA, and Emotion-Stimulus, with all of them covering a broad area of linguistic expressions and emotional stimuli. The pre-trained word embeddings have been used for initializing input features. In particular, it referred to the embedding layer that took in a padded version of the input vector (18210, 300) to allow consistent representation of data on text of a different length. As can be observed in Table 2, the proposed hybrid model, based on GRU, BiGRU, CNN, WOA, and MLP, performs better than the baseline models in all of the evaluation metrics. The hybrid model has a precision of 85.34%, a recall of 83.98%, and an F1-score of 84.34% that culminate in overall accuracy of 83.76%. The base model (CNN+BiGRU+SVM) achieved a precision of 82.39%, recall of 80.40%, F1-score of 81.27%, and accuracy of 80.11% [28]. These incremental improvement evidence indicates how well the hybrid framework can be used in the description of emotive details in a superior way as compared with individual or simpler composite modeling. Previous studies have also investigated deep-learning and hybrid architectures for text-based emotion recognition, including several deep-learning and hybrid approaches reported in [28], [36], [37], [39]. Additional related work has examined emotion recognition through facial images, EEG signals, rule-based implicit-text analysis, conversational transfer learning, multimodal audio-text modeling, graph-based conversation modeling, and broad reviews of text-based emotion detection [40]–[46].

Table 2. Evaluation results of different models

Model	Precision	Recall	F1-score	Accuracy
GRU	78.37	78.94	78.65	78.02
BiGRU	80.62	79.64	80.09	79.46
CNN	82.12	79.92	80.76	79.32
CNN+BiGRU+SVM (Base article [28])	82.39	80.40	81.27	80.11
GRU+BiGRU+CNN+WOA+MLP (Proposed method)	85.23	83.98	84.34	83.76

The proposed model works well as it is a combination of complementary elements of DL, GRU, and BiGRU help to capture the sequential and contextual dependencies in a text, and CNN layers extract local semantic patterns. The attention mechanism emphasizes highlights emotion-relevant tokens, and WOA is used for feature selection that are most informative and decrease redundancy. The selected features are then classified through an MLP to give stable emotion predictions. The comparative bar charts (Figure 4) prove its superiority because the improvement is always observed in the evaluation metrics. The findings show excellent empirical functionality, as well as the prospects of multi-component designs to advance TER, especially in real-world applications that would benefit from subtle knowledge of affective language, including sentiment-sensitive chatbots, mental health analysis, and opinion-finding systems.

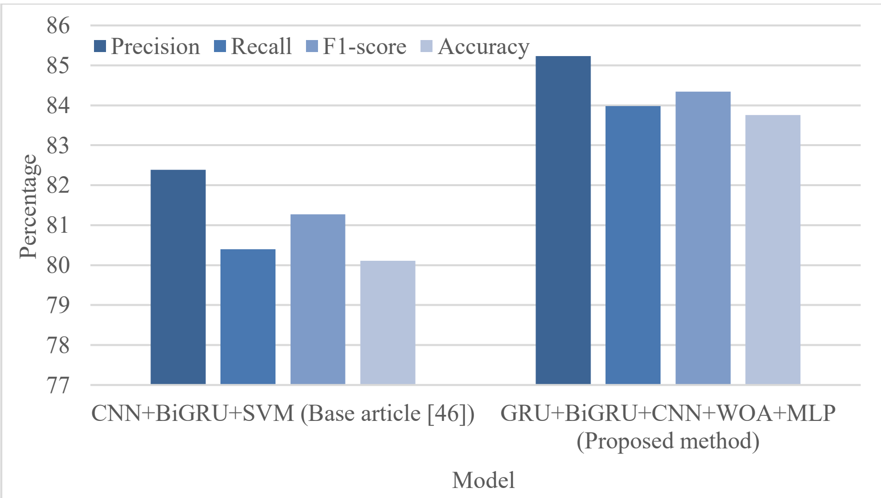

Figure 4. Evaluation results of different models

5. CONCLUSION

This paper is a hybrid framework that uses ML and DL to develop a text-based emotion recognition model. The three datasets, which are included in the proposed model, are Emotion-stimulus dataset, WASSA, and ISEAR. Its adaptability ensures that the model is effective in processing a wide range of text inputs (tweets, dialogues, keywords, sentences, and lexicon words) that are associated with emotions. Comparison of findings in the proposed method and baseline model, as shown in the given table, indicates significant differences in the performance measures. In particular, the given model has better accuracy, recall, and precision as well as F-measure compared to the baseline base article model. This significant advancement highlights the suitability of the developed method in text-based emotion recognition. The amalgamation of several neural network architectures and machine learning methods in the suggested model seems to provide synergistic advantages, as one might gain a more in-depth insight and classification of emotional content within textual data. Moving forward, we will consider the use of other classifiers or ensemble approaches for improving performance even further. In the area of DL, the combination of the BiGRU, CNN, and LSTM architectures will be explored. Moreover, the differences in the structure of the sentence will be explored, and the analysis of the regional languages will be included. Given the prevalent use of text-based communication channels such as text messaging, social media, and online reviews, there exists an opportunity to develop a real-time emotion recognition model capable of discerning people's emotions and moods from vast data streams.

FUNDING INFORMATION

This research was funded by Faculty Income Generation Grant (FTM1) of Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Tabarek Alwan Tuib	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	
Mohammad-Reza Feizi-Derakhshi		✓	✓		✓	✓	✓		✓		✓			
Yaqdhan Mahmood Hussein	✓	✓		✓		✓		✓		✓				
Fahad Taha Al-Dhief	✓		✓	✓	✓		✓	✓	✓	✓			✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] S. Zhou and M. Mohd, “Mental health safety and depression detection in social media text data: a classification approach based on a deep learning model,” *IEEE Access*, vol. 13, pp. 63284–63297, 2025, doi: 10.1109/ACCESS.2025.3559170.

[2] S. K. Hamed, M. J. A. Aziz, and M. R. Yaakub, “Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users’ comments,” *Sensors*, vol. 23, no. 4, Feb. 2023, doi: 10.3390/s23041748.

[3] S. Zainudin, “Sentiment analysis on Indonesian COVID-19 pandemic based on 2020 Twitter data,” *Asia-Pacific Journal of Information Technology and Multimedia*, vol. 14, no. 02, pp. 438–452, Dec. 2025, doi: 10.17576/apjitm-2025-1402-06.

[4] C. R. Chopade, “Text-based emotion recognition: a survey,” *International Journal of Science and Research*, vol. 4, no. 6, pp. 409–414, 2015.

[5] P. Ekman, “Basic emotions,” in *Handbook of Cognition and Emotion*, Chichester, United Kingdom: Wiley, 1999, pp. 45–60, doi: 10.1002/0470013494.ch3.

[6] W. Wundt, *Outlines of Psychology*, vol. 1, no. 5445. Leipzig, Germany: Wilhelm Engelmann, 1907, doi:10.1136/bmj.1.5445.1302-b.

[7] R. Plutchik, “*The nature of emotions*,” American Scientist, vol. 89, no. 4, pp. 344–350, 2001, doi: 10.1511/2001.28.344.

[8] M. M. Kataa and W. Kaur, “Recognizing facial emotion in real-time using MuWNet a novel deep learning network,” *Asia-Pacific Journal of Information Technology and Multimedia*, vol. 13, no. 01, pp. 01–20, Jun. 2024, doi: 10.17576/apjitm-2024-1301-01.

[9] A. A. Nafea *et al.*, “A brief review on preprocessing text in Arabic language dataset: techniques and challenges,” *Babylonian Journal of Artificial Intelligence*, vol. 2024, pp. 46–53, May 2024, doi: 10.58496/BJAI/2024/007.

[10] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.

[11] D. Seal, U. K. Roy, and R. Basak, “Sentence-level emotion detection from text based on semantic rules,” in *Information and Communication Technology for Sustainable Development*, Singapore: Springer Nature, 2020, pp. 423–430, doi: 10.1007/978-981-13-7166-0_42.

[12] K. R. Scherer and H. G. Wallbott, “Evidence for universality and cultural variation of differential emotion response patterning,” *Journal of Personality and Social Psychology*, vol. 66, no. 2, pp. 310–328, 1994, doi: 10.1037/0022-3514.66.2.310.
[13] F. M. Alotaibi, “Classifying text-based emotions using logistic regression,” *VAWKUM Transactions on Computer Sciences*, pp. 31–37, Apr. 2019, doi: 10.21015/vtcs.v16i2.551.
[14] P. Xu, A. Madotto, C. S. Wu, J. H. Park, and P. Fung, “Emo2Vec: learning generalized emotion representation by multi-task training,” in *9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018, pp. 292–298, doi: 10.18653/v1/W18-6243.
[15] W. Ragheb, J. Azé, S. Bringay, and M. Servajean, “LIRMM-advanse at SemEval-2019 Task 3: attentive conversation modeling for emotion detection and classification,” in *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019, pp. 251–255, doi: 10.18653/v1/S19-2042.
[16] M. Suhasini and S. Badugu, “Two step approach for emotion detection on Twitter data,” *International Journal of Computer Applications*, vol. 179, no. 53, pp. 12–19, Jun. 2018, doi: 10.5120/ijca2018917350.
[17] M. Hasan, E. Rundensteiner, and E. Agu, “Automatic emotion detection in text streams by analyzing Twitter data,” *International Journal of Data Science and Analytics*, vol. 7, no. 1, pp. 35–51, Feb. 2019, doi: 10.1007/s41060-018-0096-z.
[18] A. Rodriguez, Y. L. Chen, and C. Argueta, “Framework for detection and integration of unstructured data of hate speech on Facebook using sentiment and emotion analysis,” *IEEE Access*, vol. 10, pp. 22400–22419, 2022, doi: 10.1109/ACCESS.2022.3151098.
[19] L. Cao, S. Peng, P. Yin, Y. Zhou, A. Yang, and X. Li, “A survey of emotion analysis in text based on deep learning,” in *2020 IEEE 8th International Conference on Smart City and Informatization (iSCI)*, Dec. 2020, pp. 81–88, doi: 10.1109/iSCI50694.2020.00020.
[20] F. A. Acheampong, C. Wenyu, and H. N. Mensah, “Text-based emotion detection: advances, challenges, and opportunities,” *Engineering Reports*, vol. 2, no. 7, Jul. 2020, doi: 10.1002/eng2.12189.
[21] A. S. Navarrete, C. M. Araneda, C. V. Castro, and C. R. Manzano, “A novel approach to the creation of a labelling lexicon for improving emotion analysis in text,” *The Electronic Library*, vol. 39, no. 1, pp. 118–136, 2021, doi: 10.1108/EL-04-2020-0110.
[22] K. Sailunaz and R. Alhajj, “Emotion and sentiment analysis from Twitter text,” *Journal of Computational Science*, vol. 36, Sep. 2019, doi: 10.1016/j.jocs.2019.05.009.
[23] S. Mohammad and F. B. Marquez, “WASSA-2017 shared task on emotion intensity,” in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017, pp. 34–49, doi: 10.18653/v1/W17-5205.
[24] D. Ghazi, D. Inkpen, and S. Szpakowicz, “Detecting emotion stimuli in emotion-bearing sentences,” in *Computational Linguistics and Intelligent Text Processing*, Cham, Switzerland: Springer, 2015, pp. 152–165, doi: 10.1007/978-3-319-18117-2_12.
[25] A. Estabrooks and N. Japkowicz, “A mixture-of-experts framework for learning from imbalanced data sets,” in *Advances in Intelligent Data Analysis*, Berlin, Germany: Springer, 2001, pp. 34–43, doi: 10.1007/3-540-44816-0_4.
[26] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *27th International Conference on Neural Information Processing Systems (NIPS ’13)*, 2013, pp. 3111–3119.
[27] S. K. Bharti *et al.*, “Text-based emotion recognition using deep learning approach,” *Computational Intelligence and Neuroscience*, pp. 1–8, Aug. 2022, doi: 10.1155/2022/2645381.
[28] N. A. Mohamed, M. A. Zulkifley, A. A. Ibrahim, and M. Aouache, “Optimal training configurations of a CNN-LSTM-based tracker for a fall frame detection system,” *Sensors*, vol. 21, no. 19, Sep. 2021, doi: 10.3390/s21196485.
[29] K. Cho *et al.*, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” in *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.
[30] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
[31] Y. Kim, “Convolutional neural networks for sentence classification,” in *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751, doi: 10.3115/v1/D14-1181.
[32] A. Vaswani *et al.*, “Attention is all you need,” *Advances in Neural Information Processing Systems*, pp. 5999–6009, 2017, doi: 10.1201/9781003561460-19.
[33] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.
[34] S. R. Ahmad, A. A. Bakar, and M. R. Yaakub, “Ant colony optimization for text feature selection in sentiment analysis,” *Intelligent Data Analysis*, vol. 23, no. 1, pp. 133–158, Feb. 2019, doi: 10.3233/IDA-173740.
[35] S. Gite *et al.*, “Textual feature extraction using ant colony optimization for hate speech classification,” *Big Data and Cognitive Computing*, vol. 7, no. 1, Mar. 2023, doi: 10.3390/bdcc7010045.
[36] A. Chowanda, R. Sutoyo, Meiliana, and S. Tanachutiwat, “Exploring text-based emotions recognition machine learning techniques on social media conversation,” *Procedia Computer Science*, vol. 179, pp. 821–828, 2021, doi: 10.1016/j.procs.2021.01.099.
[37] A. R. Abas, I. Elhenawy, M. Zidan, and M. Othman, “BERT-CNN: a deep learning model for detecting emotions from text,” *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2943–2961, 2022, doi: 10.32604/cmc.2022.021671.
[38] D. Krstinić, M. Braović, L. Šerić, and D. B. Štulić, “Multi-label classifier performance evaluation with confusion matrix,” in *Computer Science & Information Technology*, Jun. 2020, pp. 01–14, doi: 10.5121/csit.2020.100801.
[39] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare, G. Yogapriya, and P. Pandey, “An efficient way of text-based emotion analysis from social media using LRA-DNN,” *Neuroscience Informatics*, vol. 2, no. 3, Sep. 2022, doi: 10.1016/j.neuri.2022.100048.
[40] M. I. H. Alzawali *et al.*, “Facial emotion images recognition based on binarized genetic algorithm-random forest,” *Baghdad Science Journal*, vol. 21, no. 2(SI), Feb. 2024, doi: 10.21123/bsj.2024.9698.
[41] T. A. Tuib, B. H. Saoudi, Y. M. Hussein, T. H. Mandeel, and F. T. Al-Dhief, “Convolutional neural network with binary moth flame optimization for emotion detection in electroencephalogram,” *IAES International Journal of Artificial Intelligence*, vol. 13, no. 1, pp. 1172–1178, Mar. 2024, doi: 10.11591/ijai.v13.i1.pp1172-1178.
[42] O. Udochukwu and Y. He, “A rule-based approach to implicit emotion detection in text,” in *Natural Language Processing and Information Systems*, Cham, Switzerland: Springer, 2015, pp. 197–203, doi: 10.1007/978-3-319-19581-0_17.
[43] D. Hazarika, S. Poria, R. Zimmermann, and R. Mihalcea, “Conversational transfer learning for emotion recognition,” *Information Fusion*, vol. 65, pp. 1–12, Jan. 2021, doi: 10.1016/j.inffus.2020.06.005.
[44] X. Liang, Y. Zou, X. Zhuang, J. Yang, T. Niu, and R. Xu, “MMATERIC: multi-task learning and multi-fusion for AudioText emotion recognition in conversation,” *Electronics*, vol. 12, no. 7, Mar. 2023, doi: 10.3390/electronics12071534.
[45] W. Nie, R. Chang, M. Ren, Y. Su, and A. Liu, “I-GCN: incremental graph convolution network for conversation emotion detection,” *IEEE Transactions on Multimedia*, vol. 24, pp. 4471–4481, 2022, doi: 10.1109/TMM.2021.3118881.
[46] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora, and I. Pappas, “A review on text-based emotion detection: techniques, applications, datasets, and future directions,” May 2022, *arXiv*:2205.03235.

BIOGRAPHIES OF AUTHORS

Tabarek Alwan Tuib is an emerging researcher specializing in intelligent systems and computational engineering. She has established a robust academic trajectory. She earned her foundational degree, a B.S. in Computer Techniques/Computer Electronics Engineering, from the Electrical Engineering Technical College in 2017-2018. Building on this, she completed an M.S. degree in Computer Engineering, Artificial Intelligence, and Robotics at Islamic Azad University in 2021. She is currently advancing her expertise by pursuing a Ph.D. in Computer Engineering and Artificial Intelligence at the University of Tabriz. Her doctoral research is centered on the core domains of deep learning and machine learning, with a focus on developing novel algorithms and models within the broader scope of artificial intelligence. Her investigative work aims to contribute to the creation of more adaptive, efficient, and innovative technological solutions. She is actively engaged in the academic community. She can be contacted at email: tabarcka@gmail.com.

Mohammad-Reza Feizi-Derakhshi is a distinguished academic in the field of computer science. He earned his B.Sc. in Software Engineering from the University of Isfahan, Iran, followed by both his M.Sc. and Ph.D. in Artificial Intelligence from Iran University of Science and Technology. Currently, he holds the position of full professor in Computer Engineering at the University of Tabriz. His research portfolio is extensive and impactful, focusing on several key areas of modern computing. His primary interests include natural language processing (NLP), where he works on enabling machines to understand human language; optimization algorithms, developing methods to enhance system efficiency and problem-solving; social network analysis, exploring the structure and dynamics of online communities; and intelligent databases, which involve integrating AI to manage and extract knowledge from data more effectively. Through his work, he contributes significantly to advancing both theoretical and applied aspects of artificial intelligence and data systems. He can be contacted at email: mfeizi@tabrizu.ac.ir.

Yaqdhan Mahmood Hussein received the B.S. in Computer Techniques Engineering in 2014-2015 from Islamic University College in Najaf City, and M.S. degrees in Electronic Engineering (Telecommunication System) from Universiti Teknikal Malaysia Melaka, Malaysia, in 2018. He currently studies for a Ph.D. degree in Electronic Engineering at Universiti Teknologi Malaysia in Johor Bahru city. His research focuses on the design of advanced antennas for next-generation communication systems, particularly millimeter-wave (mmWave) and base station antennas using substrate integrated waveguide (SIW) technology. His Ph.D. at Universiti Teknologi Malaysia is centered on developing high-performance antenna systems for 5G and beyond. He has extensive expertise in millimeter-wave antennas, base station antennas, and the integration of substrate integrated waveguide (SIW) technology with butler matrices for beamforming applications. A key and innovative aspect of his current work involves the application of machine learning (ML) and artificial intelligence (AI) to automate and enhance the antenna design process, predict performance, and solve complex electromagnetic problems. His research contributions are documented in several peer-reviewed international journals. Notable publications include a study on a dual-band MIMO antenna for mmWave applications and the design of a compact butlermatrix for 5G networks. He can be contacted at email: yaqthanm79@gmail.com.

Fahad Taha Al-Dhief received the B.S. in Software Engineering from Imam Jaafar Al-Sadiq University, Iraq in 2013 and also received M.S. in Computer Science from Universiti Kebangsaan Malaysia, Malaysia in 2016. He also received Ph.D. in the Faculty of Electrical Engineering, Department of Communication Engineering, Universiti Teknologi Malaysia, Malaysia in 2023. Currently, he is a senior lecturer at Center for Software Technology and Management, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia. He is scientific researcher in different healthcare applications. He is an active member of IEEE and a member of IEEE Communications Society. His research interests are machine learning, artificial neural networks, deep learning, sensor networks, routing protocols, mobile ad-hoc networks, social networks, internet of things, and location-based service. He can be contacted at email: fahad@ukm.edu.my.