



OPEN

An enhanced social emotional recognition model using bidirectional gated recurrent unit and attention mechanism with advanced optimization algorithms

Taghreed Ali Alsudais¹, Muhammad Swaileh A. Alzaidi², Majdy M. Eltahir³, Mukhtar Ghaleb⁴, Hanan Al Sultan⁵, Abdulsamad Ebrahim Yahya⁶✉, Mohammed Alshahrani⁷ & Mohammed Yahya Alzahrani⁸

Social-emotional learning (SEL) is gradually becoming a region of attention for defining children's school readiness and forecasting academic success. It is the procedure of incorporating cognition, behaviour, and emotion into daily life. School structure contains systemic practices to integrate SEL into teaching and learning so that kids and adults construct social- and self-awareness, acquire the ability to handle their specific and other's feelings and behaviour, make reliable decisions, and build positive relations. Recent school-based programs have demonstrated that SEL greatly improves mental and physical health, academic success, moral judgment, citizenship, and motivation. This paper proposes a Deep Representation Model with Word Embedding and Optimization Algorithm for Social Emotional Recognition (DRMWE-OASER) methodology. The DRMWE-OASER methodology primarily aims to develop an effectual method for detecting social-emotional learning using advanced techniques. At first, the text pre-processing stage is applied at various levels to clean and convert text data into a meaningful and structured format. Moreover, the word embedding process is implemented using the TF-IDF method. Furthermore, a bidirectional gated recurrent unit with attention mechanism (BiGRU-AM) method is employed for classification. Finally, the improved whale optimizer algorithm (IWOA)-based hyperparameter selection process is utilized to optimize the classification results of the BiGRU-AM method. A wide range of experiments using the DRMWE-OASER approach is performed under emotion detection from text dataset. The experimental validation of the DRMWE-OASER approach portrayed a superior accuracy value of 99.50% over existing models.

Keywords Word embedding, Social emotional recognition, Deep learning, Improved whale optimization algorithm, Text Pre-Processing

Supporting children in developing emotional and social skills is crucial for their long-term growth. One effective method to foster these skills in children is involving them in daily conversations¹. This paper underlines key features to consider aside from the number of languages children may hear. Consider, in detail, how various forms of emotion and mental status language might impact emotional and social learning, in addition to the significance of studying the context in which the language takes place and the child's wider surroundings². Owing to this framing, aspiration to encourage children's emotional and social learning over daily language.

¹Department of Teaching and Learning, College of Education and Human Development, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. ²Department of English Language, College of Language Sciences, King Saud University, P. O. Box 145111, Riyadh, Saudi Arabia. ³Department of Information Systems, Applied College at Mahayil, King Khalid University, Mahayil, Saudi Arabia. ⁴College of Computing and Information Technology, University of Bisha, Bisha 61922, Saudi Arabia. ⁵Department of English, College of Arts, King Faisal University, Hofuf, Al-Ahsa, Saudi Arabia. ⁶Department of Information Technology, College of Computing and Information Technology, Northern Border University, Arar, Saudi Arabia. ⁷Computer Department, Applied College, Najran University, Najran 66462, Saudi Arabia. ⁸Saudi Electronic University, Riyadh, Saudi Arabia. ✉email: Abdulsamad.qasem@nbu.edu.sa

Accordingly, in the first years, children participate in emotional and social learning, allowing them to improve necessary skills related to healthy long-term growth³, in association with the most critical advances in developing SEL. This change is not only about language proficiency; it includes the complete development of students, highlighting their emotional capacity, ability, and resilience to communicate with different cultures and peers⁴. An artificial intelligence (AI)-improved SEL infrastructure deals with these challenges by expressing good learning opportunities, giving real-time responses, and adopting an understanding learning environment. By utilizing natural language processing (NLP), adaptive learning technologies, and data analytics, trainers can form targeted approaches that improve student commitment and support emotional stability⁵. SEL programs are a possible solution not only to improve academic and behavioural results for students but also to affect healthy personal relationships, social trust, and warmth⁶.

Nevertheless, these solutions have not been derived from minority students because of some possible flaws in the current SEL programs. Conventional teaching models frequently avoid the social and emotional dimensions of learning, resulting in separation and adversarial psychological effects on students' complete welfare⁷. SEL has developed as essential in generating empowering learning environments that encourage emotional intelligence or academic success. Advanced search proof verifies that SEL skills are educated and determined, support assured growth, minimize unwanted behaviours, and enhance students' health-related behaviours, academic performance, and citizenship⁸. Besides, these skills forecast important lively outcomes such as finishing high school on time, attaining a university degree, and getting a permanent job⁹. Current empirical observation proving that SEL encourages students' academic, career success, and life has resulted in local, federal, and state policies that help a country's younger generation's academic, social, and emotional development¹⁰. Supporting the development of emotional and social abilities in early childhood is crucial for overall well-being and future success. These skills can be developed in children by engaging them in every interaction. Understanding the impact of diverse emotional expressions and mental states on learning enables the development of more effective educational strategies. Additionally, recognizing the environment and context surrounding the child improves the impact of these interactions. This highlights the significance of creating models that can accurately interpret social and emotional cues to support children's growth better.

This paper proposes a Deep Representation Model with Word Embedding and Optimization Algorithm for Social Emotional Recognition (DRMWE-OASER) methodology. The DRMWE-OASER methodology primarily aims to develop an effectual method for detecting social-emotional learning using advanced techniques. At first, the text pre-processing stage is applied at various levels to clean and convert text data into a meaningful and structured format. Moreover, the word embedding process is implemented using the TF-IDF method. Furthermore, a bidirectional gated recurrent unit with attention mechanism (BiGRU-AM) method is employed for classification. Finally, the improved whale optimizer algorithm (IWOA)-based hyperparameter selection process is utilized to optimize the classification results of the BiGRU-AM method. A wide range of experiments using the DRMWE-OASER approach is performed under emotion detection from text dataset. The significant contribution of the DRMWE-OASER approach is listed below.

- The DRMWE-OASER model applies comprehensive multi-level text pre-processing to clean and organize raw data, enhancing its quality and structure for more accurate analysis. This step ensures that subsequent models receive well-prepared inputs, improving overall system performance. It plays a significant role in facilitating effective feature extraction and classification.
- The DRMWE-OASER approach employs the TF-IDF method to generate meaningful word embeddings that capture the significance of terms within the text, improving feature representation. This approach enhances the model's capability to distinguish relevant patterns and semantic relationships. It provides robust input features that assist in accurate and efficient classification.
- The DRMWE-OASER methodology implements the BiGRU-AM model to effectively capture contextual data in sequential text data, improving classification accuracy. Integrating an AM allows the model to concentrate on the most relevant features, improving the handling of long-term dependencies and overall performance.
- The DRMWE-OASER method implements the IWOA model to fine-tune model parameters, effectively improving convergence speed and accuracy. This optimization approach assists in avoiding local minima and enhances overall model performance. It ensures optimal parameter selection for robust and reliable results.
- The DRMWE-OASER technique integrates BiGRU with an AM model, improved by IWOA-based tuning, to present a novel framework that significantly enhances social-emotional recognition accuracy. This incorporation effectively captures contextual dependencies and adaptive optimization of model parameters. The approach balances deep sequential learning with advanced metaheuristic optimization to present robust and precise results.

The article is structured as follows: Sect. 2 presents the literature review, Sect. 3 outlines the proposed method, Sect. 4 details the results evaluation, and Sect. 5 concludes the study.

Literature review

Mahendar et al.¹¹ developed a spatial attention network using the methodology of CNN (SAN-CNN). This technique underscores the saliency characteristics and spatial importance between neighbouring pixels. Initially, a median contour filter is employed for the pre-processing process. Furthermore, mask-assisted ROI is utilized to segment features. Additionally, CNNs, landmark localization, and head position approximation based on spatial attention networks are implemented for classification. Zhang et al.¹² introduced a deep learning (DL) model, namely a multimedia-based ideological and political education system utilizing DL (MIPEDLT). This approach evaluates the behaviours and the capacity of students doing higher studies in collecting data and understanding the impacts of spreading innovations in philosophical and governmental instruction. This

protocol flow is performed by utilizing multimedia models. Sowmya et al.¹³ implemented a DL methodology, namely BiLSTM. At present, a substantial segment of research focuses on textual categorization depending on sentiments, with a trivial segment concentrating on emotion recognition, specifically in business applications. The primary objective of the presented technique is to reduce the gap between the business establishments and the consumers by evaluating the reviews depending on emotion classification. This assists in furnishing institutions with a systematic approach to understanding consumer feelings, giving a more accurate product performance analysis. Kousika et al.¹⁴ introduced an advanced web application that enables users to submit various forms of content, comprising images, texts, videos, and audio, to evaluate hate content. This also utilizes bidirectional encoder representations from transformers (BERT) CNN and VGG16 CNN approaches to perform a detailed review of given data in diverse structures. Accentuating the emotion detection in textual content and hate speech, the model extends to integrate facial emotion detection for imageries. Wang et al.¹⁵ presented a methodology by incorporating pitch acoustic factors with DL-based features to evaluate and comprehend emotions exhibited during hotline communications. To understand its clinical significance, this methodology is utilized in assessing the frequency of negative feelings and the emotional alterations in the dialogue, associating the individuals with suicidal thoughts with those without. Blazhuk et al.¹⁶ developed a methodology to recognize emotional components and communication aims of textual messages utilizing NLP tools for anticipation of emotions and forming an expert opinion concerning the communication intentions depending on the dominant emotional factor with justification in the form of a list of emotionally coloured words and phrases. The approach also employed bidirectional learning and a multi-headed AM. After analyzing various animal emotion detection models.

Benedict and Subair¹⁷ proposed a DL-assisted edge-enabled serverless architecture. Moreover, a study on the cost impact of integrating serverless-based techniques is accomplished for the detection models. Furthermore, the study provided directions for developing edge-enabled serverless architectures that improve socioeconomic situations while averting human-animal conflicts. Khodaei et al.¹⁸ presented a model to detect emotions in Persian text by preparing a labelled dataset covering six basic emotions (JAMFA) and evaluating multiple models including FastText-BiLSTM, Convolutional neural network (CNN)-bidirectional encoder representations from transformers (CNN-BERT), recurrent neural network (RNN)-BERT (RNN-BERT), and an optimized BERT-BiLSTM (OBB) approach. Guo et al.¹⁹ improved emotion recognition in panoramic audio and video virtual reality content by utilizing the XLNet-Bidirectional Gated Recurrent Unit with Attention (XLNet-BiGRU-Attention) approach and the CNN with Bidirectional Long Short-Term Memory (CNN-BiLSTM) model. Sharma et al.²⁰ proposed a model by utilizing the Word2Vec (W2Vec) Skip-gram model for feature representation, integrated with gated recurrent unit (GRU), long short-term memory (LSTM), bidirectional RNN (Bi-RNN), and AM models for classification. N-fold cross-validation is employed to ensure robust evaluation of the models. Sachdeva et al.²¹ enhanced emotion analysis in NLP by employing bidirectional GRU (BiGRUs) within a sequential deep learning framework. The model captures bidirectional dependencies in text by integrating embedding, dropout, batch normalization, and softmax-activated BiGRU layers and achieves high classification accuracy. Pushpa et al.²² presented a technique incorporating BERT, GRU, and CNN techniques. Makhmudov, Kultimuratov, and Cho²³ presented a multimodal emotion recognition framework incorporating speech and text analysis using CNNs on Mel spectrograms and BERT for textual inputs. An attention-based fusion mechanism incorporates both modalities to enhance accuracy in emotional state prediction across benchmark datasets. Das, Kumari, and Singh²⁴ presented a radial basis function-GRU (RBF-GRU) model for classifying five facial expressions using the newly developed Emotional Facial Recognition (EPR) dataset. Duong et al.²⁵ introduced

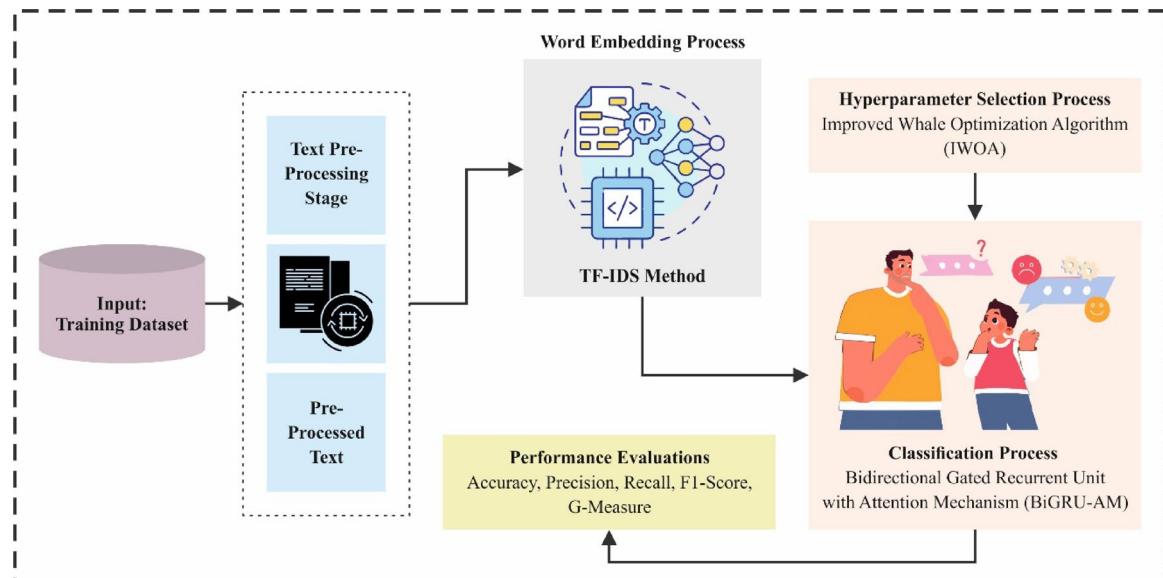


Fig. 1. Overall flow of DRMWE-OASER technique.

a novel emotion recognition in conversation (ERC) framework using residual relation-aware attention (RRAA) with positional encoding and GRU to model complex speaker relationships. By using a fully connected directed acyclic graph to represent inter- and intra-speaker ties, the model enhances contextual emotion understanding across conversations. Pamungkas et al.²⁶ improved EEG-based emotion classification by optimizing the GRU model through feature selection and architectural improvements. The stacked GRU model illustrated superior performance over existing models by utilizing statistical and entropy-based features from alpha and beta EEG bands.

Sherin, SelvakumariJeya, and Deepa²⁷ presented a model by integrating a fuzzy sentiment emotion extractor (FEE), an ensemble bidirectional LSTM-GRU (Bi-LSTM-GRU), and an enhanced Aquila optimizer (EAQ) techniques. Hossain et al.²⁸ proposed EmoNet, a deep attentional recurrent CNN (DAR-CNN) model, integrating LSTM, CNN, and GRU with an advanced AM for accurate emotion classification from social media text. Kumar et al.²⁹ presented a bidirectional GRU (BiGRU)-based model for emotion classification in NLP, utilizing bidirectional sequence learning, dropout, and batch normalization to effectively capture contextual dependencies. Anam et al.³⁰ presented a hybrid GRU-bidirectional LSTM (GRU-BiLSTM) method with Synthetic Minority Over-sampling Technique (SMOTE) for improved emotion detection. Fu et al.³¹ introduced a Multi-Modal Fusion Dialogue Graph Attention Network (MM DialogueGAT) technique that integrates BiGRU and multi-head attention for fusing text, video, and audio modalities, while using graph attention networks (GAT) for effectively capturing temporal context. Yan et al.³² proposed a Multi-branch CNN with Cross-Attention (MCNN-CA) methodology for accurate emotion recognition, utilizing multimodal data from EEG and text sources, and evaluated on SEED, SEED-IV, and ZuCo datasets to improve feature fusion and classification performance. Jiang et al.³³ introduced the ReliefF-based Graph Pooling Convolutional Network with BiGRU Attention Mechanisms (RGPCN-BiGRUAM) method for EEG-based emotion recognition, incorporating graph and recurrent architectures with attention and feature selection to improve classification accuracy. Moorthy and Moon³⁴ presented a Hybrid Multi-Attention Network (HMATN) technique for multimodal emotion recognition, incorporating audio and visual data using a collaborative cross-attention and Hybrid Attention of Single and Parallel Cross-Modal (HASPCM) mechanism for capturing both intermodal and intramodal features. Zhang et al.³⁵ presented a multi-modal emotion recognition model integrating speech and text features using extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) with wavelet transformation and BERT-RoBERTa with GRU, improved by BiLSTM and AMs. Ma et al.³⁶ presented a multi-scale convolutional bidirectional long short-term memory with attention (MSBiLSTM-Attention) model for automatic emotion recognition from raw EEG signals. Qiao and Zhao³⁷ presented a Spatial-Spectral-Temporal Convolutional Recurrent Attention Model (SST-CRAM) methodology for EEG-based emotion recognition, integrating power spectral density and differential entropy features with CNN, convolutional block attention module (CBAM), ECA-Net, and BiLSTM for extracting comprehensive data.

Despite crucial improvements in emotion recognition, existing models often suffer from overfitting due to limited and imbalanced datasets, specifically in low-resource languages or underrepresented emotion classes. Various methods, such as CNN, GRU, and BiLSTM, concentrate on a single modality (text, image, or speech), resulting in mitigated contextual understanding. The integration of multimodal data is still at an early stage, with challenges in aligning semantic and temporal features. Though widely adopted, attention mechanisms are not uniformly optimized across diverse architectures. Most existing techniques depend on specific modalities, restricting their generalizability across diverse datasets and real-world scenarios. Moreover, various models encounter threats with computational complexity and real-time applicability due to deep architectures and attention mechanisms. Furthermore, handling severe class imbalance and noisy multimodal data remains an ongoing issue affecting model robustness. A key research gap is the lack of a unified framework combining multiple data modalities with adaptive attention tuning and optimized learning. Additionally, minimal work addresses real-time emotion detection with minimal latency in edge-enabled environments. There is also requirement for more efficient, scalable models that can seamlessly integrate heterogeneous multimodal data while maintaining robustness against imbalance and noise. Also, improving interpretability and real-time performance in emotion recognition systems is still underexplored.

Proposed methodology

This paper proposes a novel DRMWE-OASER model. The main objective of the model is to develop an effective method for detecting SEL using advanced techniques. To accomplish that, the model uses text pre-processing, word embedding, a classification process, and parameter tuning. Figure 1 represents the entire flow of the DRMWE-OASER technique.

Text Pre-processing

At first, the text pre-processing stage is applied at various levels to clean and convert text data into a meaningful and structured format. In DL, data should be in numerical form³⁸. Before encoding text into numeric representations, an essential step called pre-processing text is important. This includes numerous phases, like:

- Remove null values from the data set.
- Keep only the “polarity” and “sentiment” columns, removing some unnecessary columns.
- Eliminate duplicate entries in the dataset.
- Transform values like “ambiguous,” “mixed,” “negative,” “positive,” or “neutral” into numeric models.
- Tokenization of the text column, generating individual lists utilizing NLTK’s transferred kenize.
- Eliminate stop words from the lists.
- Eliminate characters in the text.
- Use word of the NLTK tokenize to transform the lists into tokens.

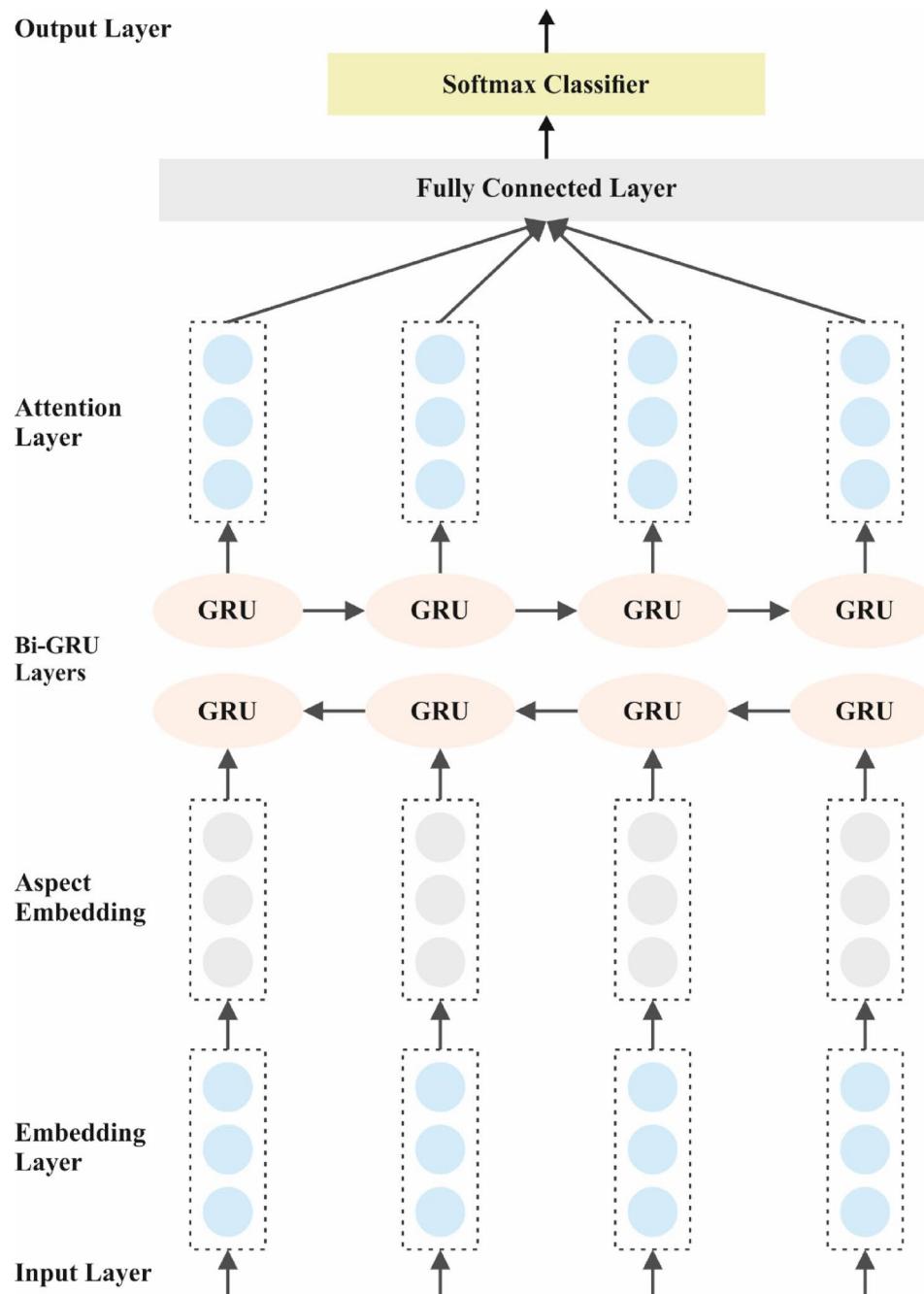


Fig. 2. Architecture of BiGRU-AM method.

- Divided the data into testing and training groups.

It is applied in the data pre-processing phase, where tokenization becomes an important task. Tokenization, a crucial stage in NLP, involves breaking down a text into different elements, whether phrases, tokens, words, or symbols. This procedure is a fundamental component of DL methods, allowing them to process and examine textual data skillfully.

Word embedding

In this section, the TF-IDF method implements the word embedding process³⁹. This method is chosen for its simplicity, interpretability, and efficiency in representing textual data in a structured numerical form. Unlike neural embeddings like Word2Vec or GloVe, TF-IDF does not require extensive training data and computational resources, making it appropriate for smaller or domain-specific datasets. It highlights the significance of rare but significant terms by down-weighting common words, thus improving the model's focus on informative features. Moreover, TF-IDF retains the actual vocabulary, allowing easy traceability and relevance scoring. Its

deterministic nature ensures consistent results across diverse runs, which is valuable in repeatable NLP pipelines. This method is specifically advantageous when transparency and lightweight computation are required.

It is a numerical standard utilized in NLP and data recovery to measure the importance of keywords in particular documents. The incidence rates of expressions in documents can be signified as TF and computed as:

$$TF(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}, \quad (1)$$

Whereas $f_{t,d}$ denotes the frequency word t and $\sum_{t' \in d} f_{t',d}$ signifies the number of frequencies of every word in a similar document. Greater TF values specify phrases studied that are related to the particular document context.

Alternatively, document frequency (DF) specifies the incidence frequencies of a particular word through the complete set of documents. Words along with higher DF values absent significance owing to their frequent incidence. Subsequently, greater IDF values specify infrequent words in the complete set, thus improving their full importance.

Classification using the BiGRU-AM model

Furthermore, the DRMWE-OASER model utilizes the BiGRU-AM method for the classification model⁴⁰. This model was chosen for its ability to effectively capture sequential dependencies from both past and future contexts in text data. Unlike standard RNNs or unidirectional GRUs, the bidirectional structure processes input in both directions, improving context understanding. The AM model additionally improves performance by dynamically focusing on the most relevant parts of the sequence, which is crucial for tasks like emotion or sentiment analysis. Compared to LSTM, GRU is computationally lighter while still retaining long-term dependencies. This integration provides a balance of efficiency, accuracy, and interpretability. The model is appropriate for applications where subtle contextual cues must be identified precisely. Figure 2 depicts the infrastructure of the BiGRU-AM technique.

BiGRU is a sophisticated technique that originated from the conventional RNN, intended to effectively address restrictions of RNN by handling gradient vanishing and long-term memory concerns. As its essential component, GRU dynamically regulates the data flow at every step by presenting the reset and update gate instruments and sequentially forgetting and retaining significant data, simplifying the data transferring process and overwhelming the gradient vanishing. BiGRU covers this source by processing the series' backwards or forward data over a bidirectional framework. Forward GRU progressively processes time-step aspects from the initial to the end of the sequence. In contrast, backward GRU processes reverse aspects from the end to the initial sequence and eventually incorporate the hidden layer (HL) of both to form a bidirectional semantical representation. Presuming the HL created by the forward GRU and the backward GRU, the outcome of BiGRU is denoted as $h_t^{BiGRU} = [\vec{h}_t; \overleftarrow{h}_t]$. With the Bi-directional framework, this method can concurrently take the sequence's front and back textual data to more extensively create effective contextual feature representations that are specifically appropriate to deal with composite semantic 360 relationships in text.

Moreover, the AM focuses on overwhelming the boundaries of conventional sequence methods, such as LSTM and RNN, in longer sequence processing, specifically the efficacy concern in taking longer-range dependency. Whereas LSTM and RNN might be undermined by longer-range dependency owing to the step-by-step process of sequences, the AM dynamically modifies the input element weights over the global viewpoint, considerably enhancing the task's longer sequence efficacy and precision. The basic concept is to assign weights to every input element depending on the conditions and assess its significance in the existing task, allowing the method to aim at key aspects while restraining sound and unnecessary data. Structurally, the AM functions over the interaction of keys, values, and queries. The relationship between a query vector Q and every key vector K is computed utilizing the dot product, resulting in attention scores e_t that measure the significance of each key to the query. Afterwards, the scores are normalized by employing the softmax function to create attention weights α_t , guaranteeing that the weight values are inside a numerical range and add to 1. These attention weights are used to summation the value vectors V to create the output vector of context c . The computation method of AM can be given:

Calculation of attention scores:

$$e_t = \tanh (W_e h_t + b_e) \quad (2)$$

Using the more common scaled dot-product form:

$$e_{ij} = Q_i \cdot K_j^T \quad (3)$$

Normalization of attention weights:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^T \exp(e_i)} \quad (4)$$

Generation of context vectors:

$$c = \sum_{i=1}^T \alpha_i h_i \quad (5)$$

Sentiment	No. of Record
Sadness	5000
Enthusiasm	700
Neutral	5000
Worry	5000
Surprise	2000
Love	3000
Fun	1500
Hate	1300
Happiness	5000
Boredom	150
Relief	1500
Anger	100
Total Records	30,250

Table 1. Details of the dataset.

S.no	Sentiment	Content
1	Sadness	“I haven’t seen Muffin in two whole days!”
2	Enthusiasm	:I want another tatt”
3	Neutral	“fun in the sun hmmm bell no it’s cold”
4	Worry	“I need skott right now”
5	Surprise	“No churchill downs for me tomorrow”
6	Love	“there’s really no Android Twitter app of Tweetie’s caliber”
7	Fun	“RIP leonardo. You were a great mini fiddler crab”
8	Hate	“I’m so very tired...and have insomnia.”
9	Happiness	“english class! Working on interactive orals”
10	Boredom	“I’m so tired”
11	Relief	“I’m at work”
12	Anger	“Working But it’s Fidaaaayyyyy”

Table 2. Samples of text.

Here, h_t is the method’s HL at the existing moment, functioning as the input to AM. W_e specify the learned weight matrix employed to adapt the significance of every input feature. b_e represents the biased term for linear transformation, permitting modifications to the linear transformation output.

This work incorporates the AM and BiGRU, focused on utilizing the intensities of these dual prevailing DL models to more effectively take contextual and semantical data inside complicated sequential data. With its bidirectional framework, BiGRU takes contextual data from either backward or forward directions, giving accurate contextual aspects suitable for handling Chinese text with complex semantic and syntactic relationships. Nevertheless, despite its capability to demonstrate bidirectional dependency, BiGRU shows restrictions in managing longer-term dependency. Consequently, this paper presents the AM that captures the HL created by BiGRU as input. Using normalization and dynamic computation, the AM allocates weight to every input element, recognizing crucial input data depending on the task context. With parallel computation and global attention, the AM effectively takes longer-term dependency and evades the restrictions of sequential computation. Finally, the extracted features are transferred to the linear layer to output the prediction outcomes.

IWOA-based parameter tuning

Finally, the IWOA-based hyperparameter selection process is applied to optimize the classification results of the BiGRU-AM model⁴¹. This model was chosen for its superior capability of effectively exploring and exploiting the search space. This model introduces mechanisms like adaptive weights and reverse learning to avoid local minima and accelerate convergence, unlike conventional optimization techniques such as grid or random search. These improvements ensure more precise parameter selection, which directly enhances model performance. Its population-based strategy makes it robust against noisy and high-dimensional search spaces common in DL tasks. Thus, IWOA provides a more reliable and efficient solution for optimizing the BiGRU-AM model than conventional or less adaptive metaheuristic methods. The humpback whale’s foraging behaviour stimulates the WOA model. Once prey, humpback whales approach the prey in a spiral motion, encircling it while discharging a bubble net for searching. The foraging behaviour of the whale includes three ways: surrounding prey, bubble-net attacking, and hunting for prey.

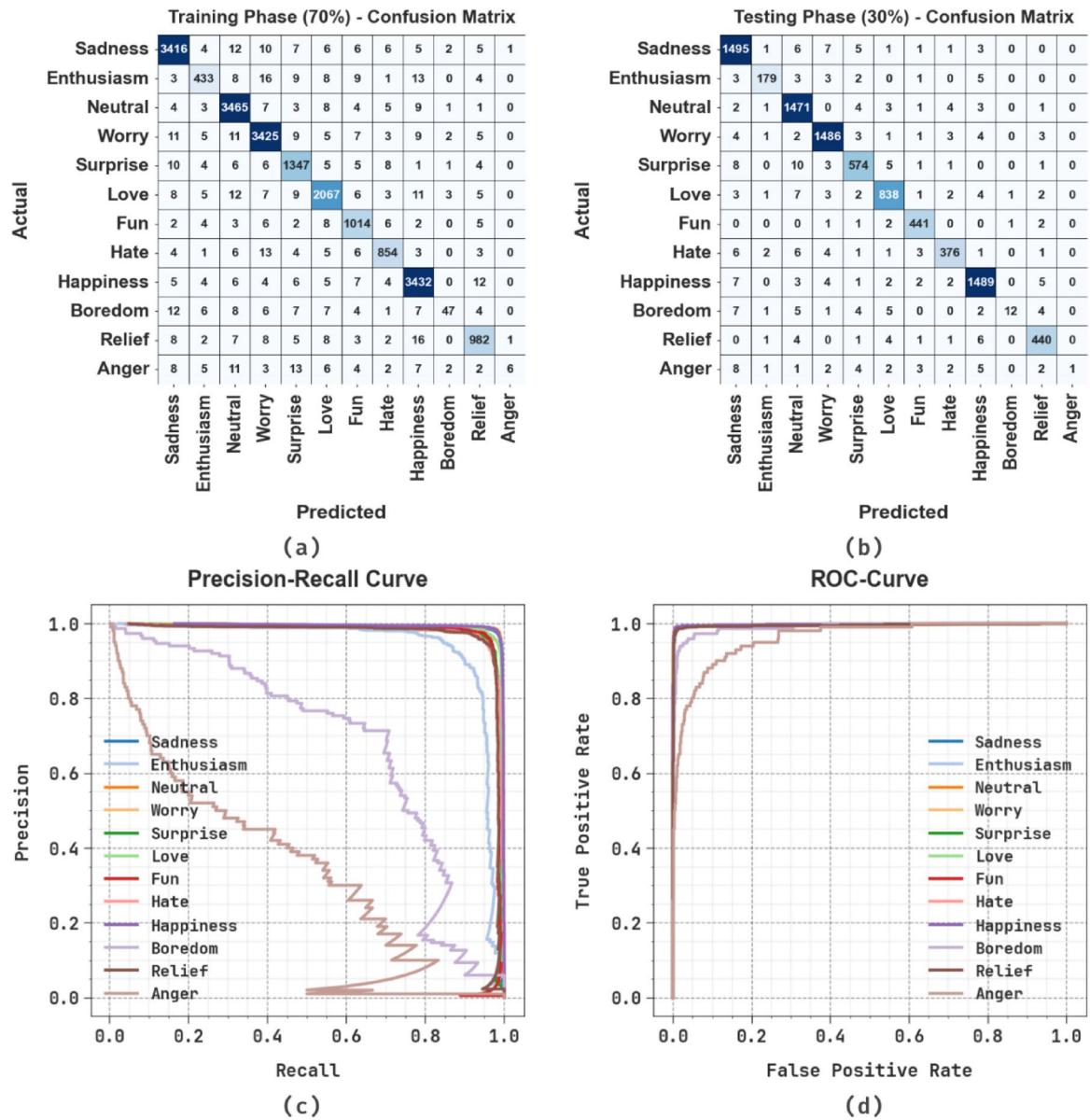


Fig. 3. Classifier outcomes of (a-b) 70%TRPH and 30%TSPH of the confusion matrix and (c-d) curves of PR and ROC.

Encircling prey

Search agents discover the global area to find and encircle the best solution. Other agents upgrade their locations near this optimal candidate. The mathematic representation of this method is as shown:

$$\vec{D} = \left| \vec{C} \times \vec{X}_b(t) - \vec{X}(t) \right| \quad (6)$$

$$\vec{X}(t+1) = \vec{X}_b(t) - \vec{A} \times \vec{D} \quad (7)$$

Whereas $\vec{X}_b(t)$ denotes the local ideal solution, $\vec{X}(t)$ signifies the location of the individual vector, and t signifies the count of present iterations. \vec{A} and \vec{C} are co-efficient vectors, and the mathematic formulations of \vec{A} and \vec{C} are as shown:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (8)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (9)$$

$$\vec{a} = 2 - 2 \frac{t}{iter_{max}} \quad (10)$$

Class Labels	Accu _y	Prec _n	Recal	F1 _{score}	G _{measure}
TRPH (70%)					
Sadness	99.34	97.85	98.16	98.01	98.01
Enthusiasm	99.46	90.97	85.91	88.37	88.40
Neutral	99.36	97.47	98.72	98.09	98.09
Worry	99.28	97.55	98.08	97.82	97.82
Surprise	99.41	94.79	96.42	95.60	95.60
Love	99.34	96.68	96.77	96.72	96.72
Fun	99.53	94.33	96.39	95.35	95.35
Hate	99.59	95.42	94.99	95.21	95.21
Happiness	99.36	97.64	98.48	98.06	98.06
Boredom	99.66	81.03	43.12	56.29	59.11
Relief	99.48	95.16	94.24	94.70	94.70
Anger	99.69	75.00	08.70	15.58	25.54
Average	99.46	92.82	84.17	85.81	86.88
TSPH (30%)					
Sadness	99.20	96.89	98.36	97.62	97.62
Enthusiasm	99.71	95.21	91.33	93.23	93.25
Neutral	99.27	96.90	98.72	97.81	97.81
Worry	99.45	98.15	98.54	98.35	98.35
Surprise	99.37	95.35	95.19	95.27	95.27
Love	99.43	96.99	96.99	96.99	96.99
Fun	99.76	96.71	98.44	97.57	97.57
Hate	99.55	95.92	93.77	94.83	94.84
Happiness	99.35	97.83	98.28	98.06	98.06
Boredom	99.66	85.71	29.27	43.64	50.09
Relief	99.57	95.44	96.07	95.76	95.76
Anger	99.67	100.00	03.23	06.25	17.96
Average	99.50	95.93	83.18	84.61	86.13

Table 3. Emotion detection of DRMWE-OASER model under 70%TRPH and 30%TSPH.

Whereas the vector \vec{a} reduces linearly from (2 - 0), \vec{r} signifies an arbitrary coefficient vector in the interval of $[0,1]$, and $iter_{max}$ symbolizes the maximal iteration counts.

Bubble-net attacking

Whales further attack their victim over a bubble-net approach that contains either shrinking around or spiral upgrades. If $-1 < A < 1$, the searching agent agrees to the finest whale in the present state, upgrading the location of the whale's group based on Eq. (7).

$$\vec{X}(t+1) = \vec{D} \cdot e^{ql} \cdot \cos(2\pi l) + \vec{X}_b(t) \quad (11)$$

$$\vec{D} = \left| \vec{X}_b(t) - \vec{X}(t) \right| \quad (12)$$

Whereas \vec{D} characterizes the absolute value for the distance between the present whale and the most appropriate whale; q refers to the constant demonstrating the spiral shape; and l stands for arbitrary number belonging in the interval¹.

$$\vec{X}(t+1) = \begin{cases} \vec{X}_b(t) - \vec{A} \cdot \vec{D}, & \text{if } p < 0.5 \\ \vec{D} \cdot e^{ql} \cdot \cos(2\pi l) + \vec{X}_b(t), & \text{if } p \geq 0.5 \end{cases} \quad (13)$$

On the other hand, p denotes a randomly generated number between zero and one .

Hunting for prey

During this WOA, parameter A controls the whale searching behaviour. If $|A| \geq 1$, whales explored by moving near an arbitrary whale.

$$\vec{D} = \left| \vec{C} \times \vec{X}_{rand}(t) - \vec{X}(t) \right| \quad (14)$$

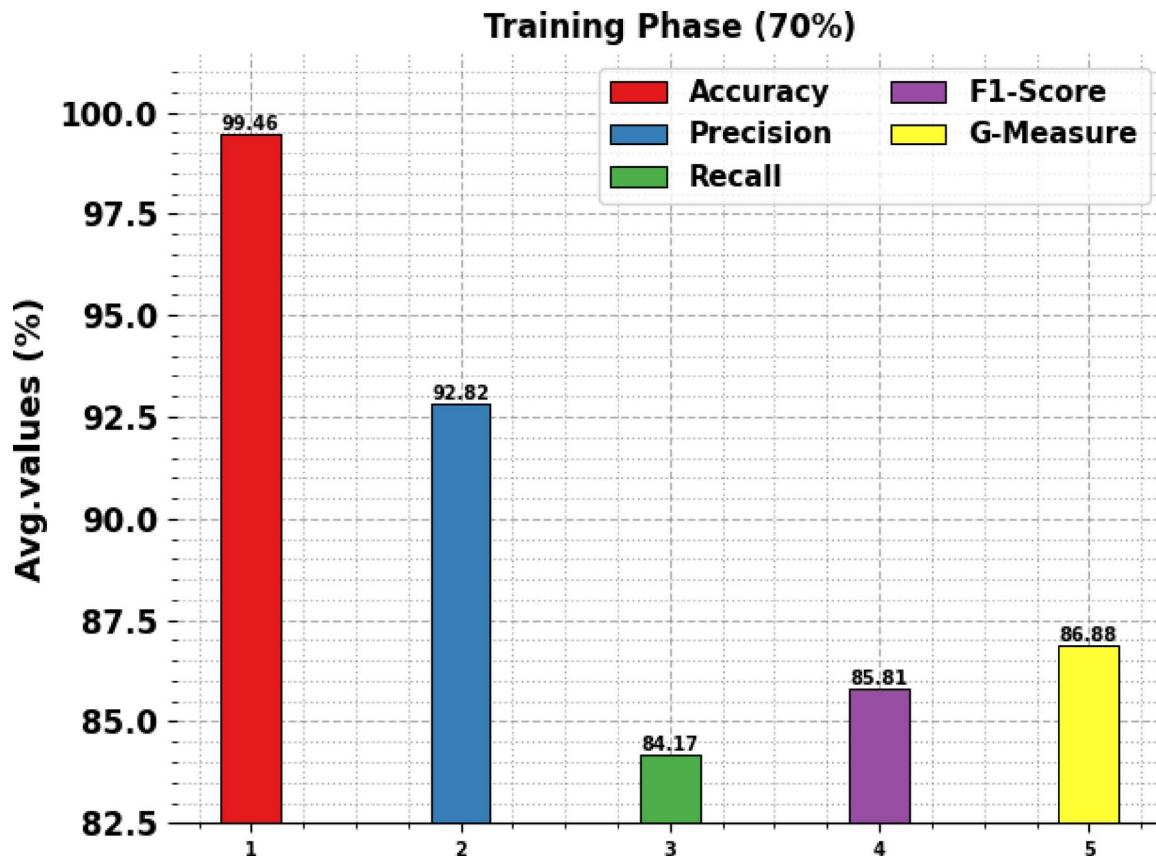


Fig. 4. Average of DRMWE-OASER model under 70%TRPH.

$$\vec{X}(t+1) = \vec{X}_{rand}(t) - \vec{A} \times \vec{D} \quad (15)$$

Here, $\vec{X}(t)$ signifies the position vector of the randomly chosen whale.

The standard WOA suffers from a slower convergence speed and an inclination to drop into the local bests. To deal with these shortcomings, the IWOA presents the succeeding main developments: population initialization with random difference mutation and reverse learning approaches for preventing the model from being stuck in the local bests, in addition to nonlinear convergence features incorporated with adaptive weight approaches to dynamically fine-tune the searching process and improve the speed of convergence. The computing efficiency of the IWOA is generally affected by the iteration counts (T), problem dimension (D), population size (N), and complexity of the fitness function (FF) (C_f). In the population initialization stage, the complexity originates from making the first and reverse populations $O(2N \times D)$, assessing the fitness of each individual $O(2N \times C_f)$, and choosing the best N individuals $O(2N \log(2N))$. Consequently, the complete complexity for initialization is $O(2N \times D + 2N \times C_f + 2N \log(2N))$. During this location upgrade stage, the complexity comprises upgrading the location of all individuals $O(N \times D)$, using random differential mutation $O(N \times D)$, assessing the fitness of each individual $O(2N \times C_f)$, and choosing the optimal individuals $O(2N \log(2N))$, leading to a complete complexity of $O(2N \times D + 2N \times C_f + 2N \log(2N))$. For the iteration stage, the complete complexity becomes $O(T \times (2N \times D + 2N \times C_f + 2N \log(2N)))$, while T denotes iteration counts. Joining each stage, the total computing cost of the IWOA is $O(2N \times D + 2N \times C_f + 2N \log(2N)) + O(T \times (2N \times D + 2N \times C_f + 2N \log(2N)))$.

Population initialization with reverse learning

The typical WOA model uses arbitrary initialization that frequently leads to higher randomness and cannot ensure enough variety in the primary population. Reverse learning includes describing the variable's boundary range and originating its equivalent reverse solution according to particular rules, thus guaranteeing superior diversities in the early population. During this WOA, assume the population size is N and the search area is $2M$ -dimensional, the location of the i th whale in the $2M$ -dimensional is stated as shown: $\vec{X}_i(t) = \{X_i^1(t), X_i^2(t), \dots, X_i^{2M}(t)\}$ ($i = 1, 2, \dots, N$),

$X_i^j(t) \in [0, 0 + |V|]$ ($j = M + 1, M + 2, \dots, 2M$) $X_i^j(t) \in [0, 1]$ ($j = 1, 2, \dots, M$), $a_i^j(t)$, and $b_i^j(t)$ represents upper and lower limits of $X_i^j(t)$, correspondingly. The inverse solution equivalent to $X_i^j(t)$ is demonstrated in Eq. (16):

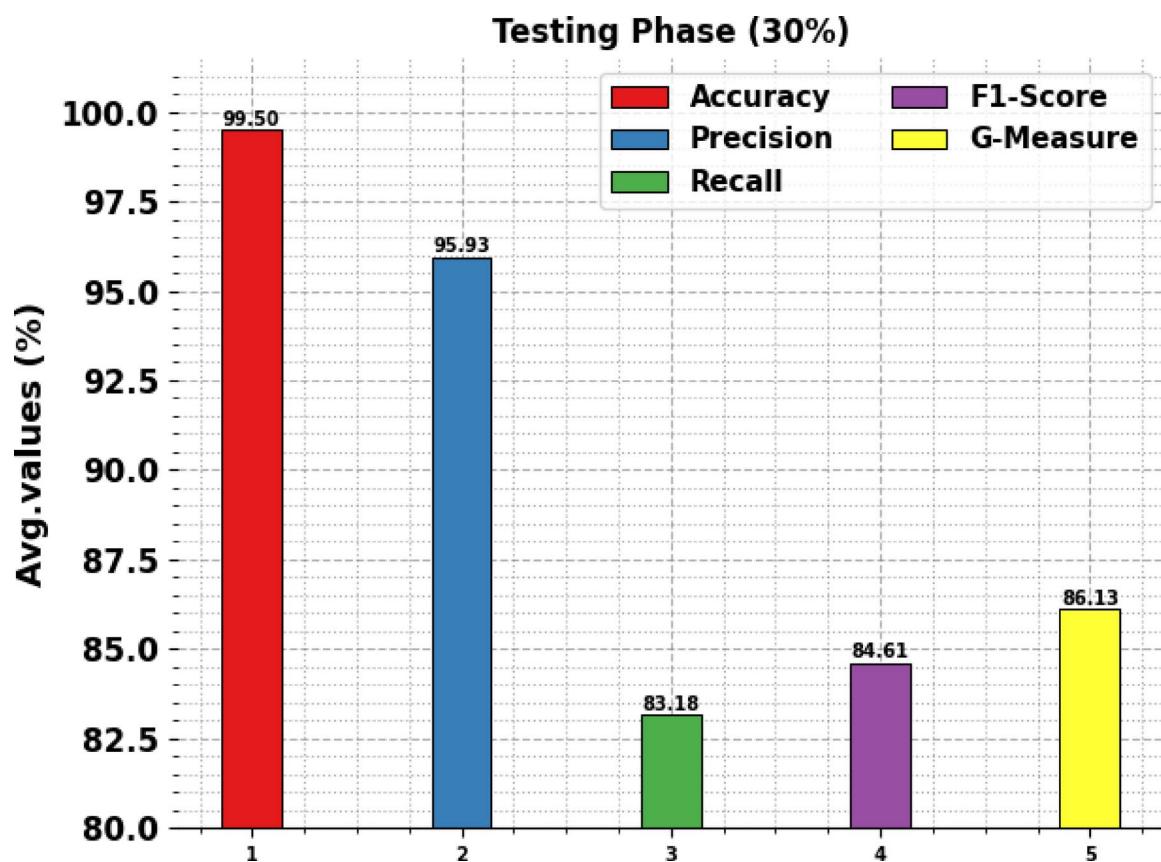


Fig. 5. Average of DRMWE-OASER model under 30%TSPH.

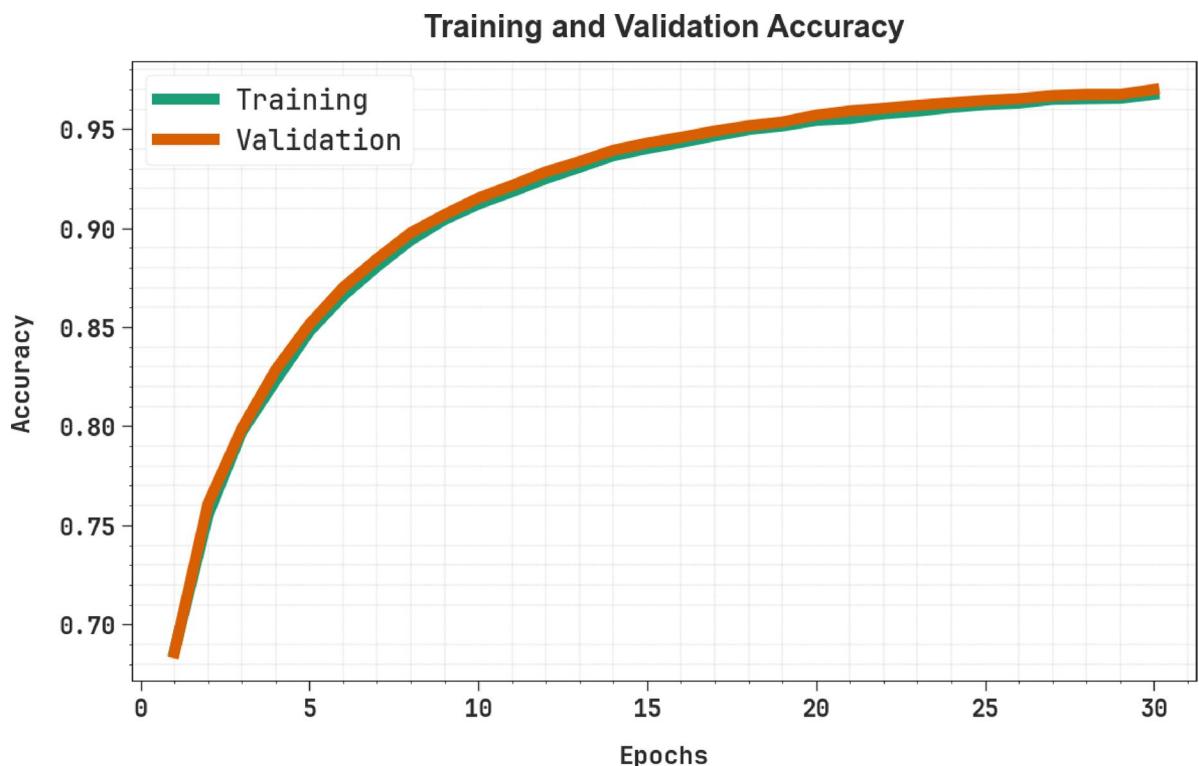


Fig. 6. $Accu_y$ curve of DRMWE-OASER model.

Training and Validation Loss

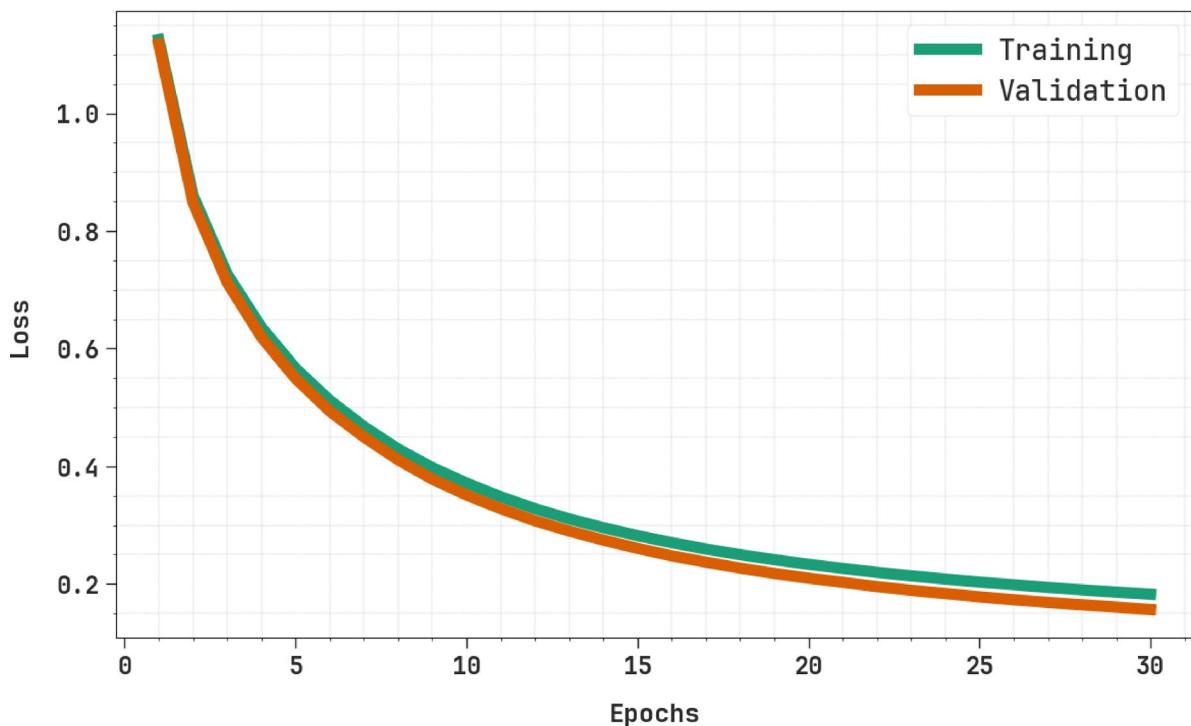


Fig. 7. Loss curve of DRMWE-OASER model.

Methodology	Accu _y	Prec _n	Recal _t	F1 _{score}
XLNET	98.23	91.49	75.66	75.25
RoBERTa	89.55	93.33	80.56	78.39
DistilBERT	90.54	95.09	77.99	73.36
BiLSTM	89.59	93.62	77.27	81.11
GRU Model	96.03	89.07	77.53	80.01
Lexicon	99.16	89.31	71.47	80.24
SVM (TF-IDF)	94.42	95.18	73.00	81.49
XLNet-BiGRU-Attention	91.08	95.66	78.68	73.92
CNN-BiLSTM	90.09	94.12	77.82	81.82
Bi-RNN	96.79	89.65	78.28	80.73
DRMWE-OASER	99.50	95.93	83.18	84.61

Table 4. Comparative analysis of the DRMWE-OASER model with existing methods^{19,20,43–45}.

$$X_{new,i}^j(t) = a_i^j(t) + b_i^j(t) - X_i^j(t) \quad (16)$$

If N individuals of the model population are initialized, N reverse individuals are made, and all reverse individuals parallel one of the N primary individuals. Assumed $2N$ candidate solutions (N reversals and N primary individuals), the optimal N is preserved as the population while the reverse evolution model begins. All reverse of N individuals is estimated in all generations, and in these $2N$ candidate solutions, the optimal N individuals are retained for the following production. The equation to select N superior individuals is as shown:

$$fitness\left(\vec{X}_i(t)\right) < fitness\left(\vec{X}_{new,i}(t)\right) ? \vec{X}_i(t), \vec{X}_{new,i}(t) \quad (17)$$

Whereas *fitness* refers to FF, $\vec{X}_{new,i}$ and $\vec{X}_i(t)$, reverse learning individuals and current random individuals, correspondingly.

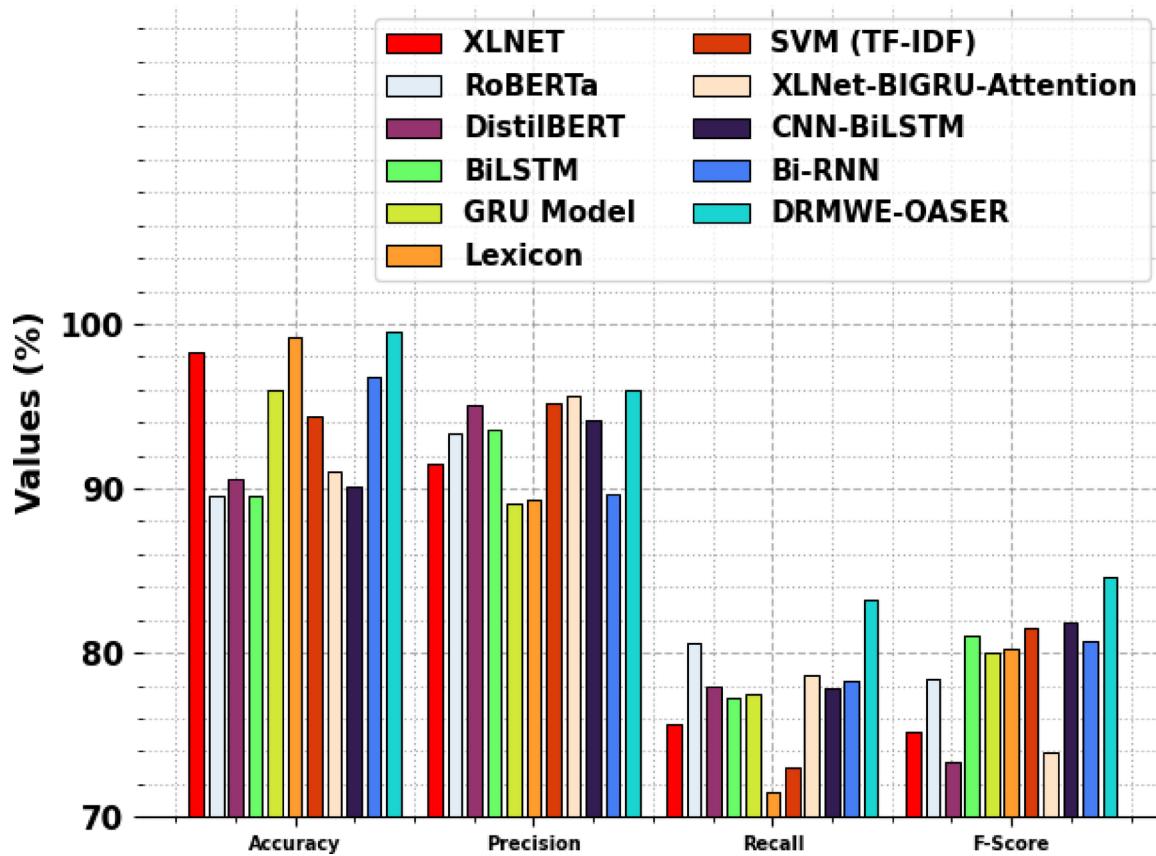


Fig. 8. Comparative analysis of the DRMWE-OASER model with existing methods.

Methodology	CT (sec)
XLNET	24.26
RoBERTa	24.59
DistilBERT	16.85
BiLSTM	21.17
GRU Model	19.80
Lexicon	11.19
SVM (TF-IDF)	24.76
XLNet-BiGRU-Attention	24.46
CNN-BiLSTM	13.95
Bi-RNN	9.31
DRMWE-OASER	7.11

Table 5. CT evaluation of the DRMWE-OASER technique with existing models.

The IWOA model originates a fitness function (FF) for achieving an enriched classification performance. A positive number signifies better efficiency for the candidate solution. Here, the classification rate of error minimization is reflected as FF. Its mathematical formulation is given in Eq. (18).

$$\begin{aligned}
 \text{fitness}(x_i) &= \text{ClassifierErrorRate}(x_i) \\
 &= \frac{\text{no. of misclassified samples}}{\text{Total no. of samples}} \times 100
 \end{aligned} \tag{18}$$

Performance validation

The performance evaluation of the DRMWE-OASER model is examined under emotion detection from text dataset⁴². This dataset contains 30,250 records under 12 sentiments, as depicted in Table 1. Table 2 represents the sample texts.

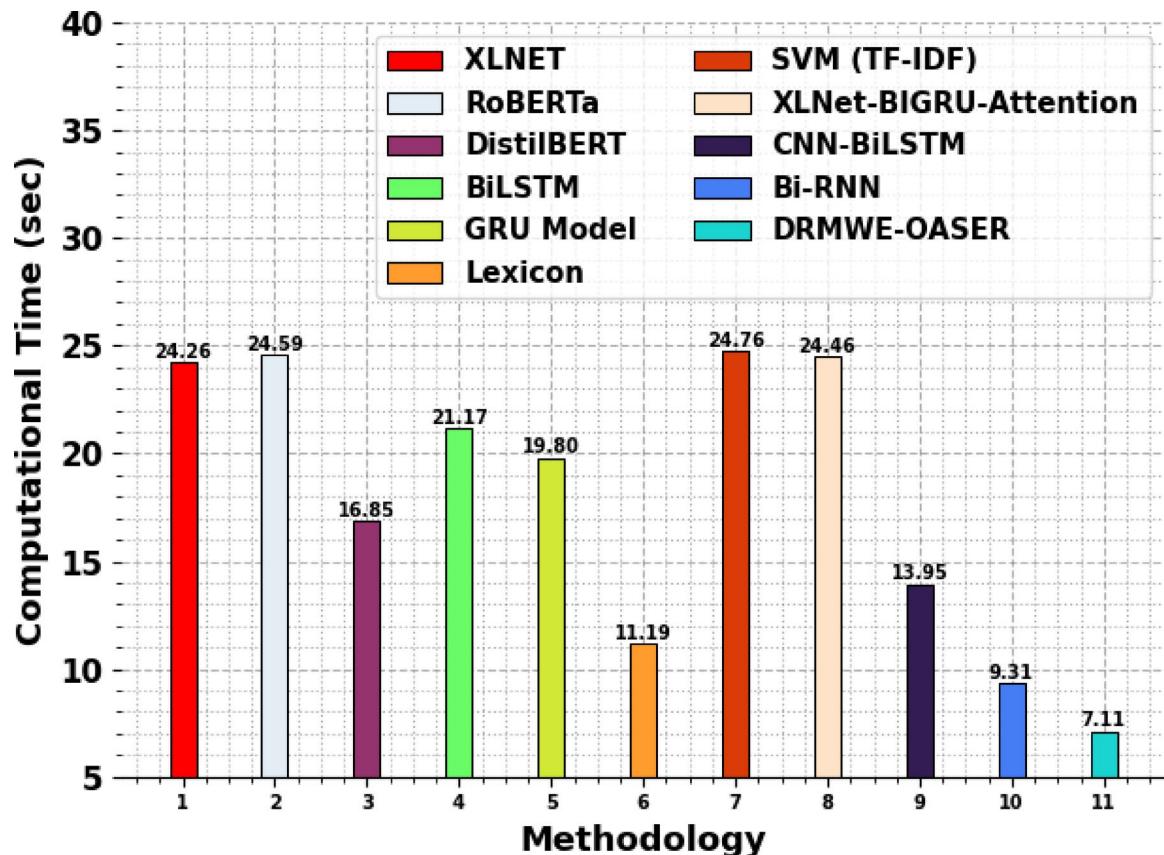


Fig. 9. CT evaluation of the DRMWE-OASER technique with existing models.

Methodology	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_t</i>	<i>F1_{score}</i>
XLNET	1.77	8.51	24.34	24.75
RoBERTa	10.45	6.67	19.44	21.61
DistilBERT	9.46	4.91	22.01	26.64
BiLSTM	10.41	6.38	22.73	18.89
GRU Model	3.97	10.93	22.47	19.99
Lexicon	0.84	10.69	28.53	19.76
SVM (TF-IDF)	5.58	4.82	27.00	18.51
XLNet-BiGRU-Attention	8.92	4.34	21.32	26.08
CNN-BiLSTM	9.91	5.88	22.18	18.18
Bi-RNN	3.21	10.35	21.72	19.27
DRMWE-OASER	0.50	4.07	16.82	15.39

Table 6. Error analysis of DRMWE-OASER methodology with existing techniques.

Methodology	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_t</i>	<i>F1_{score}</i>
IWOA	98.35	94.78	81.76	83.06
BiGRU-AM	98.92	95.33	82.45	83.86
DRMWE-OASER	99.50	95.93	83.18	84.61

Table 7. Result analysis of the ablation study of DRMWE-OASER methodology with existing techniques.

Figure 3 presents the classifier performances of the DRMWE-OASER method below 70%TRPH and 30%TSPH. Figure 3a and b reveals the confusion matrix through precise classification and identification of all distinct classes. Figure 3c exhibits the PR inspection, which notified superior outcomes across all class labels.

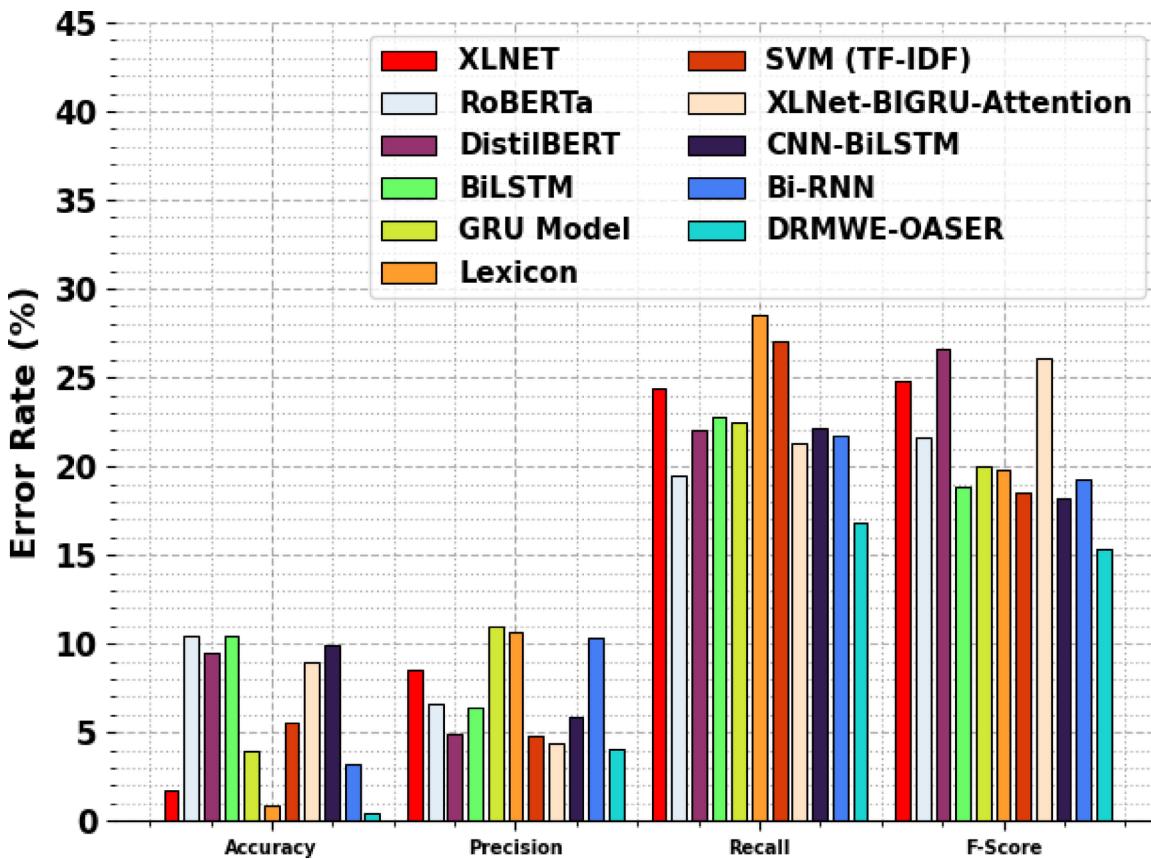


Fig. 10. Error analysis of DRMWE-OASER methodology with existing techniques.

At last, Fig. 3d exemplifies the ROC outcome, which signifies capable solutions using high ROC values for dissimilar class labels.

Table 3 demonstrates the emotion detection of the DRMWE-OASER model below 70%TRPH and 30%TSPH.

Figure 4 states the average solution of the DRMWE-OASER model below 70%TRPH. The performances declare that the DRMWE-OASER technique accurately recognized the samples. On 70%TRPH, the DRMWE-OASER technique delivers average $accu_y$, $prec_n$, $recal$, $F1_{score}$, and $G_{measure}$ of 99.46%, 92.82%, 84.17%, 85.81%, and 86.88%, respectively.

Figure 5 provides the average solution of the DRMWE-OASER method below 30%TSPH. The results establish that the DRMWE-OASER approach suitably recognized the samples. Using 30%TSPH, the DRMWE-OASER approach delivers average $accu_y$, $prec_n$, $recal$, $F1_{score}$, and $G_{measure}$ of 99.50%, 95.93%, 83.18%, 84.61%, and 86.13%, respectively.

Figure 6 shows the training (TRA) $accu_y$ and validation (VAL) $accu_y$ solutions of the I DRMWE-OASER technique. The values of $accu_y$ are computed across a period of 0–30 epochs. The figure underscored that the values of TRA and VAL $accu_y$ present a growing tendency, indicating the capacity of the DRMWE-OASER method with maximum performance through multiple repetitions. Moreover, the TRA and VAL $accu_y$ values remain close across the epochs, notifying decreased overfitting and expressing the improved performance of the DRMWE-OASER method, ensuring reliable calculation on unseen samples.

In Fig. 7, the TRA loss (TRALOS) and VAL loss (VALLOS) graph of the DRMWE-OASER technique is showcased. The loss values are computed over a time period of 0–30 epochs. The values of TRALOS and VALLOS demonstrate a declining tendency, indicating the capacity of the DRMWE-OASER methodology to equalize a tradeoff between generalization and data fitting. The subsequent dilution in values of loss and securities results in a higher performance of the DRMWE-OASER methodology and tuning the calculation results afterwards.

The comparative study of the DRMWE-OASER method with existing models is depicted in Table 4; Fig. 8^{19,20,43–45}. The model performance identified that the DRMWE-OASER technique outperformed greater performances. The existing approaches, namely XLNET, RoBERTa, DistilBERT, BiLSTM, GRU, Lexicon, SVM, XLNet-BiGRU-Attention, CNN-BiLSTM, and Bi-RNN models, reached lesser results. While the DRMWE-OASER technique has got enhanced $accu_y$, $prec_n$, $recal$, and $F1_{score}$ of 99.50%, 95.93%, 83.18%, and 84.61%.

Table 5; Fig. 9 specifies the computational time (CT) analysis of the DRMWE-OASER technique with existing models. The results highlight the efficiency of the DRMWE-OASER technique, which completes its task in just 7.11 s outperforming other models. Among the rest, Bi-RNN is the next most efficient at 9.31 s, followed by Lexicon at 11.19 s and CNN-BiLSTM at 13.95 s. The GRU Model takes 19.80 s, BiLSTM requires 21.17 s, and transformer-based models like DistilBERT, XLNet, XLNet-BiGRU-Attention, RoBERTa, and SVM

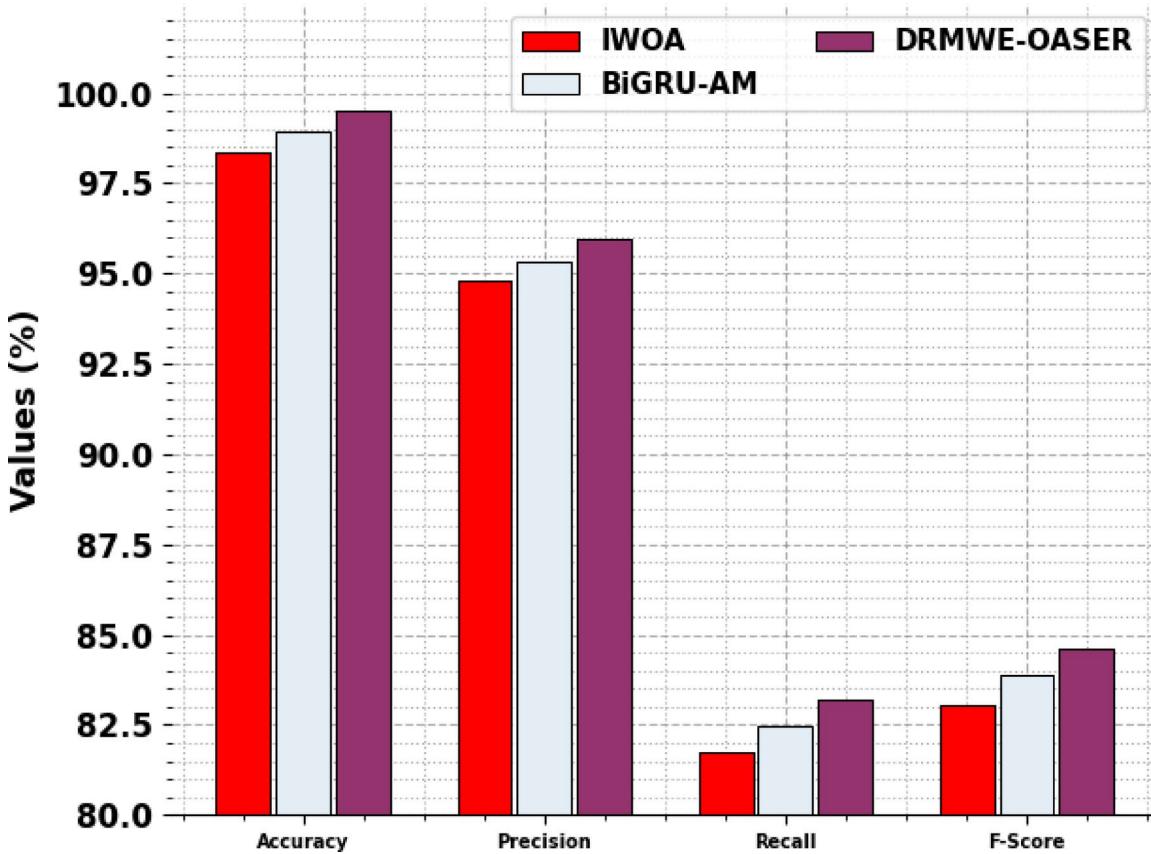


Fig. 11. Result analysis of the ablation study of DRMWE-OASER methodology with existing techniques.

with TF-IDF range between 16.85 and 24.76 s. These results position the DRMWE-OASER model as the most computationally efficient choice for text-based analysis, ideal for real-time applications without compromising on classification performance.

Table 6; Fig. 10 denotes the error analysis of the DRMWE-OASER methodology with the existing techniques. The error analysis of various text classification methodologies exhibits performance inconsistencies across key evaluation metrics. Models like RoBERTa and BiLSTM show relatively higher $accu_y$ at 10.45% and 10.41% respectively, while DistilBERT and CNN-BiLSTM follow closely with 9.46% and 9.91%. The Lexicon-based method records the highest $reca_l$ at 28.53% but with a low $accu_y$ of 0.84%, indicating its robust retrieval capability but limited precision. XLNET performs poorly with just an $accu_y$ of 1.77% and a $reca_l$ of 24.34%, while the proposed DRMWE-OASER methodology illustrates the lowest performance overall with an $accu_y$ of 0.50%, $prec_n$ of 4.07%, $reca_l$ of 16.82%, and 15.39% $F1_{score}$. These findings suggest the DRMWE-OASER model struggles to generalize effectively, and improvements in feature representation or learning strategies may be necessary for enhanced performance.

Table 7; Fig. 11 portrays the ablation study of the DRMWE-OASER approach. The ablation study analysis highlights the comparative effectiveness of different models in terms of classification performance. The IWOA method highlights efficient outcomes with $prec_n$ of 94.78%, $reca_l$ of 81.76%, and $F1_{score}$ of 83.06%. The BiGRU-AM model shows further improvement, achieving $prec_n$ of 95.33%, $reca_l$ of 82.45%, and $F1_{score}$ of 83.86%. The proposed DRMWE-OASER model outperforms the others, with $prec_n$ of 95.93%, $reca_l$ of 83.18%, and $F1_{score}$ of 84.61%, indicating a robust balance between detection capability and correctness of predictions, making it more robust than existing techniques.

Conclusion

In this paper, a novel DRMWE-OASER methodology is proposed. The main objective of the DRMWE-OASER methodology is to develop an effective method for detecting SEL using advanced techniques. At first, the text pre-processing stage is applied at various levels to clean and convert text data into a meaningful and structured format. Moreover, the word embedding process is implemented using the TF-IDS method. Furthermore, the DRMWE-OASER technique employs the BiGRU-AM method for classification. Finally, the IWOA-based hyperparameter selection is performed to optimize the classification results of the BiGRU-AM model. A wide range of experiments with the DRMWE-OASER approach is performed under emotion detection from text dataset. The experimental validation of the DRMWE-OASER approach portrayed a superior accuracy value of 99.50% over existing models. The limitations of the DRMWE-OASER approach comprise dependency on specific datasets, which may limit generalizability across diverse contexts. Additionally, the performance of the

model could be affected by imbalanced data and noisy inputs. The performance of the model may be affected by the variability and ambiguity inherent in natural language, which can challenge accurate emotion detection. Also, real-time application might be restricted by computational demands of the BiGRU-AM and optimization processes. Future work may concentrate on incorporating more diverse datasets to enhance generalization, thus improving data augmentation techniques to increase training robustness, and integrating explainability methods to provide clearer insights into model decisions. These efforts aim to strengthen the transparency and performance of the model across varied real-world scenarios.

Data availability

The data supporting this study's findings are openly available in the Kaggle repository at <https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text>, reference number [42].

Received: 12 May 2025; Accepted: 7 August 2025

Published online: 21 November 2025

References

1. Brackett, M. A., Bailey, C. S., Hoffmann, J. D. & Simmons, D. N. A theory-driven, systemic approach to social, emotional, and academic learning. *Educational Psychol.* **54** (3), 144–161 (2019).
2. Panayiotou, M., Humphrey, N. & Wigelsworth, M. An empirical basis for linking social and emotional learning to academic performance. *Contemp. Educ. Psychol.* **56**, 193–204 (2019).
3. Schonert-Reichl, K. A. Advancements in the landscape of social and emotional learning and emerging topics on the horizon. *Educational Psychol.* **54** (3), 222–232 (2019).
4. Daneshfar, F. & Kabudian, S. J. Speech emotion recognition system by quaternion nonlinear echo state network. *arXiv preprint arXiv:2111.07234*. (2021).
5. Peng, S. et al. A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Networks.* **8** (5), 745–762 (2022).
6. Gueldner, B. A., Feuerborn, L. L. & Merrell, K. W. *Social and Emotional Learning in the Classroom: Promoting Mental Health and Academic Success* (Guilford, 2020).
7. Elias, M. J. What if the doors of every schoolhouse opened to social-emotional learning tomorrow: reflections on how to feasibly scale up high-quality SEL. *Educational Psychol.* **54** (3), 233–245 (2019).
8. El Koshiry, A. M., Eliwa, E. H. I. & Omar, A. Improving Arabic spam classification in social media using hyperparameters tuning and particle swarm optimization. *Full Length Article.* **16**, 08–08 (2024).
9. Jones, S. M., McGarrah, M. W. & Kahn, J. Social and emotional learning: A principled science of human development in context. *Educational Psychol.* **54** (3), 129–143 (2019).
10. Hassan, A. Q. Applied linguistics driven artificial intelligence for automated sentiment detection and classification. *Full Length Article.* **24** (3), 08–08 (2024).
11. Mahendar, M., Malik, A. & Batra, I. Emotion Estimation model for cognitive state analysis of learners in online education using deep learning. *Expert Syst.* **42** (1), e13289 (2025).
12. Zhang, Y., Yan, Y., Kumar, R. L. & Juneja, S. Improving college ideological and political education based on deep learning. *Int. J. Inf. Commun. Technol.* **24** (4), 431–447 (2024).
13. Sowmya, G., Rohan, A., Rao, G. K., Midhilesh, K. & kumar Nayak, A. Real time emotion recognition from text using deep learning and data analysis. In *Humanizing Technology with Emotional Intelligence* (33–46). IGI Global Scientific Publishing. (2025).
14. Kousika, N. et al. April. Enhancing Multimodal Sentiment Analysis with Deep Learning Techniques to Foster Emotional Intelligence. In *2024 10th International Conference on Communication and Signal Processing (ICCP)* (pp. 778–783). IEEE. (2024).
15. Wang, H. et al. Deep Learning-Based Feature Fusion for Emotion Analysis and Suicide Risk Differentiation in Chinese Psychological Support Hotlines. *arXiv preprint arXiv:2501.08696*. (2025).
16. Blazhuk, V., Mazurets, O. & Zalutska, O. *An Approach To Using the mBERT Deep Learning* (Neural Network Model for Identifying Emotional Components and Communication Intentions, 2024).
17. Benedict, S. & Subair, R. Deep Learning-Driven Edge-Enabled serverless architectures for animal emotion detection. *Informatica*, **49**(7). (2025).
18. Khodaei, A., Bastanfar, A., Saboohi, H. & Aligholizadeh, H. A Transfer-Based Deep Learning Model for Persian Emotion Classification. *Multimedia Tools and Applications*, pp.1–29. (2024).
19. Guo, S., Wu, M., Zhang, C. & Zhong, L. Emotion recognition in panoramic audio and video virtual reality based on deep learning and feature fusion. *Egyptian Informatics Journal*, **30**, p.100697. (2025).
20. Sharma, R., Alemu, M. A., Sreenivas, N., Sungeetha, A. & Mahapatra, S. Cyberbullying recognition on social media for an Afro-Asiatic Language content using attention mechanism and Bi-RNN classifier. *Procedia Comput. Sci.* **258**, 2232–2243 (2025).
21. Sachdeva, B., Gill, K. S., Kumar, M. & Rawat, R. S. September. Harnessing Gated Recurrent Units for Enhanced Emotion Recognition in Deep Learning. In *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)* (pp. 71–76). IEEE. (2024).
22. Pushpa, G. et al. An Advanced AI Framework for Mental Health Diagnostics Using Bidirectional Encoder Representations from Transformers with Gated Recurrent Units and Convolutional Neural Networks. *Ingenierie des Systemes d'Information*, **30**(1), p.213. (2025).
23. Makhmudov, F., Kultimuratov, A. & Cho, Y. I. Enhancing Multimodal Emotion Recognition through Attention Mechanisms in BERT and CNN Architectures. *Applied Sciences*, **14**(10), p.4199. (2024).
24. Das, S., Kumari, R. & Singh, R. K. Advancements in computational emotion recognition: a synergistic approach with the emotion facial recognition dataset and RBF-GRU model architecture. *Int. J. Syst. Assur. Eng. Manage.* **16** (2), 734–749 (2025).
25. Duong, A. Q. et al. Residual relation-aware attention deep graph-recurrent model for emotion recognition in conversation. *IEEE Access*, **12**, 2349–2360 (2024).
26. Pamungkas, Y., Pratasik, S., Krisnanda, M. & Crisnapti, P. N. Optimizing gated recurrent unit architecture for enhanced EEG-Based emotion classification. *J. Rob. Control (JRC)*, **6** (3), 1450–1461 (2025).
27. Sherin, A., SelvakumariJeya, I. J. & Deepa, S. N. Enhanced Aquila optimizer combined ensemble Bi-LSTM-GRU with fuzzy emotion extractor for tweet sentiment analysis and classification. *IEEE Access* (2024).
28. Hossain, M. S. et al. EmoNet: Deep Attentional Recurrent CNN for X (formerly Twitter) Emotion Classification. *IEEE Access*. (2025).
29. Kumar, A., Gill, K. S., Upadhyay, D. & Devliyal, S. October. Elevating Deep Learning Models with Gated Recurrent Units for Emotion Detection. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1–5). IEEE. (2024).
30. Anam, M. K. et al. Improved performance of hybrid GRU-BiLSTM for detection emotion on Twitter dataset. *J. Appl. Data Sci.* **6** (1), 354–365 (2025).

31. Fu, R. et al. MM DialogueGAT—A fusion graph attention network for emotion recognition using Multi-Model system. *IEEE Access*, **12**, 150941–150952 (2024).
32. Yan, F., Guo, Z., Iliyasu, A. M. & Hirota, K. Multi-branch convolutional neural network with cross-attention mechanism for emotion recognition. *Scientific Reports*, **15**(1), p.3976. (2025).
33. Jiang, C. et al. EEG Emotion Recognition Employing RGPCN-BiGRUAM: Relieff-Based Graph Pooling Convolutional Network and BiGRU Attention Mechanism. *Electronics*, **13**(13), p.2530. (2024).
34. Moorthy, S. & Moon, Y. K. Hybrid Multi-Attention Network for Audio–Visual Emotion Recognition Through Multimodal Feature Fusion. *Mathematics*, **13**(7), p.1100. (2025).
35. Zhang, S. et al. Multi-modal emotion recognition based on wavelet transform and BERT-RoBERTa: an innovative approach combining enhanced BiLSTM and focus loss function. *Electronics*, **13**(16), p.3262. (2024).
36. Ma, Y. et al. MSBiLSTM-Attention: EEG Emotion Recognition Model Based on Spatiotemporal Feature Fusion. *Biomimetics*, **10**(3), p.178. (2025).
37. Qiao, Y. & Zhao, Q. SST-CRAM: spatial-spectral-temporal based convolutional recurrent neural network with lightweight attention mechanism for EEG emotion recognition. *Cogn. Neurodyn.* **18** (5), 2621–2635 (2024).
38. Ali, A., Khan, M., Khan, K., Khan, R. U. & Aloraini, A. Sentiment analysis of Low-Resource Language literature using data processing and deep learning. *Computers Mater. & Continua*, **79**(1). (2024).
39. Rajagukguk, N., Kencana, I. P. E. N. & Kusuma, I. G. L. W. May. Application of Term Frequency-Inverse Document Frequency in The Naive Bayes Algorithm For ChatGPT User Sentiment Analysis. In *First International Conference on Applied Mathematics, Statistics, and Computing (ICAMSAC 2023)* (pp. 29–40). Atlantis Press. (2024).
40. Luo, X., Zhang, L., Ma, Y., Du, Q. & Bai, L. Revolutionizing Accident Text Classification in the Aec Industry: A Multi-Level Framework Powered by Pre-Trained Models and Deep Learning. Available at SSRN 5110313.
41. Gong, S., Lou, P., Hu, J., Zeng, Y. & Fan, C. An Improved Whale Optimization Algorithm for the Integrated Scheduling of Automated Guided Vehicles and Yard Cranes. *Mathematics*, **13**(3), p.340. (2025).
42. <https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text>
43. Abas, A. R., Elhenawy, I., Zidan, M. & Othman, M. BERT-CNN: A deep learning model for detecting emotions from text. *Computers Mater. & Continua*, **71**(2). (2022).
44. Yohanes, D., Putra, J. S., Filbert, K., Suryaningrum, K. M. & Saputri, H. A. Emotion detection in textual data using deep learning. *Procedia Comput. Sci.* **227**, 464–473 (2023).
45. Sboev, A., Naumov, A. & Rybka, R. Data-driven model for emotion detection in Russian texts. *Procedia Comput. Sci.* **190**, 637–642 (2021).

Acknowledgements

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/206/46. Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R579), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Ongoing Research Funding program, (ORF-2025-714), King Saud University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2025-2248-06. The authors are thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.

Author contributions

Taghreed Ali Alsudais: Conceptualization, methodology, validation, investigation, writing—original draft preparation, fundingMuhammad Swaileh A. Alzaidi: Conceptualization, methodology, writing—original draft preparation, writing—review and editingMajdy M. Eltahir: methodology, validation, writing—original draft preparationMukhtar Ghaleb: software, visualization, validation, data curation, writing—review and editingHanan Al Sultan: validation, original draft preparation, writing—review and editingAbdulsamad Ebrahim Yahya: Project administration, methodology, validation, Conceptualization, writing—review and editingMohammed Alshahrani: methodology, validation, original draft preparation.Mohammed Yahya Alzahrani: validation, original draft preparation, writing—review and editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.E.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025