



OPEN

Empowering people with intellectual disabilities using integrated deep learning architecture driven enhanced text-based emotion classification

Mohammed Abdullah Al-Hagery^{1✉}, Hechmi Shili², Nasser Aljohani³ & Ishfaq Yaseen^{4,5}

Emotion recognition is an important research field including psychology, healthcare, and human-computer interaction (HCI). However, conventional techniques mainly rely on textual analysis and facial expressions, and they also have potential flaws, making them unreliable. Textual language is the most common carrier of human emotions and its analysis relies on the available data. In natural language processing (NLP), textual emotion recognition (TER) has become a significant area of research due to its essential commercial and academic applications. With the growth of deep learning (DL) technologies, TER has seen a growing interest and has undergone considerable upgrades recently. This paper proposes an Intelligent Emotion Recognition from Text Using a Hybrid Deep Learning Model and Word Embedding Process (IERT-HDLMWEP) model. The aim is to develop a DL-based system for accurate text emotion recognition to support communication for people with disabilities. Initially, the text pre-processing stage involves several typical steps to develop the analysis and minimize the dimensionality of the input data. The IERT-HDLMWEP method creates a hybrid feature representation by integrating pre-trained Word2Vec vectors weighted by TF-IDF category distribution and enriched with Part-of-Speech features to improve emotion detection in text. Finally, the hybrid of a convolutional neural network and a bidirectional gated recurrent unit with an attention mechanism (C-BiG-A) technique is employed for the classification process. A comprehensive simulation was implemented to verify the performance of the IERT-HDLMWEP method in emotion detection from the Text dataset. The empirical outcomes indicated that the IERT-HDLMWEP methodology emphasized improvement over other existing techniques.

Keywords Textual emotion recognition, Disabled people, Human computer interaction, Natural language processing, Hybrid deep learning

Disability is one of the significant issues which remains a challenge. Residual effects of disability is seen as barriers in the form of mental or physical impairments that affects the development and participation of an individual¹. As a result, intense attempts are made to eliminate this kind of obstruction. Individuals with disabilities rely on others to fulfil their needs. Artificial intelligence (AI) is a field of computer science that aims to devise intelligent computer systems, which display qualities associated with human intelligence, including learning, problem-solving, language comprehension, and decision-making². One significant role of AI has persisted in NLP, which combines linguistic and computational methods to enable computers to facilitate HCI and human language understanding. Sentiment analysis (SA), text summarisation, machine translation, and speech recognition fall under the research fields in the area of NLP³. Emotions are characterized into six kinds: sadness, joy, surprise, fear, disgust, and anger. Additionally, emotions are also defined in numerous types, including optimism and

¹Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia. ²Department of Computer Science, Haql University College, University of Tabuk, Tabuk, Saudi Arabia. ³Department of Information Systems, Faculty of Computer and Information Systems, Islamic University of Madinah, Medina 42351, Saudi Arabia. ⁴Department of Computer and Self-Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia. ⁵King Salman Centre for Disability Research, Riyadh 11614, Saudi Arabia. ✉email: hajry@qu.edu.sa

love⁴. In contrast to facial expressions and speech recognition, text sentences often fail to convey themselves as they are bland. Due to the difficulty and vagueness, it is a challenging task to accurately identify the emotions expressed in the text. Determining the emotion of the provided text is a complex process, as each word may have a distinct morphological style and meaning⁵. Emotional interaction is a prevalent cognitive event in humans' everyday life. Precise emotion recognition is the principle of efficient human communication, decision-making, and interaction. With the growth of AI and big data, emotion recognition has become a standard research project in both industry and academia⁶. Since the straightforward and fundamental medium, textual data arising from emotional communication, is frequently employed to understand emotional conditions. TER for people with disabilities is dedicated to automatically recognizing emotional modalities in textual terminologies, like Angry, Sad, and Happy. TER is a more detailed examination when compared to SA, and has attained substantial interest from academic circles⁷. For humans, emotions are identified rapidly depending on their emotional states. At the same time, for automated TER systems, calculation approaches should be constantly created and enhanced to attain more precise emotion prediction. AI implemented in NLP leverages computational and linguistic methods to enable machines to comprehend basic phenomena, such as emotions or sentiments, from texts⁸. Thus, the primary objective is to examine thoughts, ideas, and opinions through the polarities task, encompassing both positive and negative perspectives. In the NLP domain, most analyses regarding TER are performed at a coarse-grained sentiment level and are based on conventional machine learning (ML) methods⁹. While they review advanced methodologies of TER, DL-driven TER methods are only concisely stated without a detailed and thorough outline. Currently, DL models have rectified their reliance on manual feature extraction and achieved good results. The DL approaches have proven their extraordinary performance, frequently achieving advanced outcomes in TER tasks¹⁰.

This paper develops an Intelligent Emotion Recognition from Text Using a Hybrid Deep Learning Model and Word Embedding Process (IERT-HDLMWEP) model. The key contributions of this manuscript are mentioned below:

- A unique IERT-HDLMWEP model is designed for a TER system for people with disabilities.
- The text pre-processing step involves various standard stages designed to enhance analysis and reduce input data dimensionality, thereby supporting better feature extraction and improving overall model performance by filtering noise and standardizing input formats.
- The IERT-HDLMWEP model adopts a hybrid feature representation by integrating Word2Vec, TF-IDF-CDW, and POS encoding to improve emotion detection in textual data. This integration captures semantic meaning, contextual weight, and syntactic structure, enriching feature quality. It enables the model to detect subtle emotional cues across varied linguistic expressions and improves classification robustness and generalization across diverse emotional contexts.
- The hybrid C-BiG-A technique, integrating a CNN and BiGRU with an AM model, is utilized for the final classification. This architecture captures both local and long-range dependencies in the text, improving contextual understanding. The AM also refines its focus on emotion-relevant words, thereby improving interpretability. Overall, it enhances the model's capability to distinguish complex emotional patterns in textual data.
- This IERT-HDLMWEP methodology uniquely integrates convolutional and BiGRU with an AM to capture multi-level contextual features. By effectively incorporating spatial and temporal data, it improves emotion detection beyond conventional methods. The AM dynamically highlights crucial emotional cues, enhancing model focus and accuracy. This novel integration advances the handling of intrinsic linguistic patterns in emotional text classification.

Related works on TER

Kumar et al.¹¹ implemented a multimodal emotion recognition method, Visual Spoken Textual Additive Net (VISTANet), for classifying emotions displayed by input images, text, and speech into separate types. A novel interpretability approach, K-Average Additive exPlanation (KAAP), is established, which classifies significant textual, spoken, and visual features, resulting in the forecasting of a specific emotion type. The VISTANet merges data from text, speech, and image conditions, utilizing a hybrid of late and intermediate fusion. It autonomously alters the intermediate output's weights by calculating the weighted average. Di Luzio et al.¹² explored explainability methods for binary deep neural network (DNN) frameworks in the model of emotion recognition via video analysis. The authors investigated the input features for binary classifiers that recognize emotions, utilizing facial behaviour analysis and an enhanced form of the Integrated Gradients explainability technique. Fu et al.¹³ proposed a spectral domain reconstruction graph neural network (SDR-GNN) model for unfinished multimodal learning in conversational emotion recognition. This model creates an utterance semantic interaction graph utilizing a sliding window that relies on context and speaker relations. Li et al.¹⁴ proposed a new emotion recognition system depending on a curriculum learning (CL) approach (ERNetCL). This approach integrates a spatial encoder (SE), a temporal encoder (TE), and a CL loss. To mitigate the severe effects of emotional change and provide a framework for individuals to learn a curriculum from simple to complex, employ the CL concept in the task of ERC to enhance the network's parameters continually. Kusal et al.¹⁵ presented a hybrid DL network depending on the convolutional-recurrent network utilized for detecting the individual's emotions, depending on the conversational text. A convolutional network can extract local dependencies and patterns and is intrinsically shift-invariant. Simultaneously, the recurrent network extracts long-term relationships in sequential data. Feng et al.¹⁶ developed a multimodal speech emotion recognition technique that relies on multiscale MFCCs and a multiview AM, which can capture numerous audio emotional features and effectively merge emotion-specific features from two features. Under various attention configurations and audio input states, it is observed that the finest emotion recognition precision is achieved by mutually leveraging four attention modules and three diverse scales of MFCCs. Zhang et al.¹⁷ proposed a novel technique for building emotion classification

labels using language resources and density-based spatial clustering of applications with noise (DBSCAN). Moreover, incorporate the spatial features and frequency area of emotional EEG signals and pass those features to the serial network, which integrates LSTM and CNN for EEG emotion learning and identification. Omarov and Zhumanov¹⁸ proposed an innovative Bi-LSTM methodology for analyzing emotions in textual content. This technique implements the strength of RNNs to extract both past and text context, providing a detailed interpretation of emotional content. By incorporating the backwards and forward layers of LSTM model, this model efficiently learns the semantic representation of words and their dependencies within sentences.

Hicham and Nassera¹⁹ proposed the stacked DL models integrating a robustly optimized Bert pretraining approach (RoBERTa) with gated recurrent unit (GRU), long short-term memory (LSTM), bidirectional GRU (BiGRU), and bidirectional LSTM (BiLSTM). These hybrid architectures are optimized using adaptive moment estimation (Adam) models. Mahajan, More, and Shah²⁰ developed a novel multilabel dataset and evaluate various ML and DL techniques comprising logistic regression (LR), support vector machine (SVM), naïve bayes (NB), random forest (RF), LSTM, BiLSTM, GRU, and convolutional neural network (CNN) to capture both single and mixed emotional states. Zhu et al.²¹ developed a reliable medical question-answering system by utilizing knowledge embedding and a transformer-based architecture. The model also integrates a knowledge understanding layer and an answer generation layer to improve both the accuracy and ethical quality of responses. Khan et al.²² improved violence detection in surveillance videos by utilizing a two-stream DL model that integrates 3D convolutional networks with depth-wise convolutions. The model also utilizes RGB frame analysis with background suppression and optical flow to accurately capture violent actions while maintaining computational efficiency suitable for edge devices. Arumugam et al.²³ proposed an Audio, Visual, and Text Emotions Fusion Network (AVTEFN) model that utilizes graph attention networks (GAT), hybrid wav2vec 2.0 with CNN, and bidirectional encoder representations from transformers (BERT) combined with bidirectional gated recurrent units (Bi-GRU). Khan et al.²⁴ proposed an advanced Multimodal Emotion Recognition (MER) approach by utilizing a Joint Multi-Scale Multimodal Transformer (JMMT) with Recursive Cross-Attention. Alyoubi and Alyoubi²⁵ presented an optimized multimodal emotion recognition framework that integrates Bidirectional Encoder Representations from Transformers (BERT)/Robustly Optimized BERT Pretraining Approach (RoBERTa) for text, wav2vec 2.0 for speech, and Residual Networks (ResNet50)/Visual Geometry Group Network (VGG16) for facial expressions. The model also utilizes a transformer-based cross-modal attention mechanism and Shapley Additive Explanations (SHAP) to improve both classification accuracy and interpretability. Vani et al.²⁶ introduced Text Fusion+, an integrated application that utilizes optical character recognition (OCR), NLP, and Text-to-Speech (TTS) technologies. The model also employs DL-based summarization and an NLP-driven question-answering module. Khan et al.²⁷ presented a technique by utilizing sequence learning techniques, including Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), with their advanced forms such as bi-directional and multi-layer architectures, to enhance auditory emotion recognition. Ghous, Najam, and Jalal²⁸ detected emotional states in individuals with cognitive disabilities by applying advanced ML methods. Initially, bandpass filtering (BF) and downsampling is used for pre-processing. The model also incorporates an adaptive automated feature selection and transformation (AAFST) technique with a multi-class SVM to improve the accuracy and reliability of emotion recognition. Patil et al.²⁹ enabled early prediction of learning disabilities, specifically dyslexia, by employing handwritten text recognition and DNN. Mishra et al.³⁰ introduced a model by utilizing MLP and transformer techniques such as GPT-4 and BERT. The model utilizes sentiment analysis, emotion classification, and is trained using the adaptive moment estimation (Adam) optimizer to achieve high accuracy and generalization. Table 1 presents a comparative analysis of existing textual emotion recognition systems for individuals with disabilities.

Existing studies exhibit various limitations and research gaps, despite their advancements. Multi-model approaches, such as VISTANet and SDR-GNN, are computationally intensive, which restricts their real-time applicability. Explainability methods, while improving interpretability, often concentrate on binary classification or limited modalities. Most models do not adequately address the complexity of mixed or multilabel emotions, particularly in conversational or code-mixed contexts. Additionally, few studies integrate curriculum learning or adaptive fusion mechanisms to improve the robustness of the model. There is also a research gap in utilizing lightweight yet effective hybrid DL architectures for scalable emotion recognition. Addressing these gaps can improve generalizability and efficiency across diverse datasets and applications.

Research design and methodology

In this work, a new IERT-HDLMWEP model is developed for emotion identification using textual data. The proposed model aims to enhance the accuracy of a DL-based TER system in identifying and interpreting emotions in text, thereby improving communication support for individuals with disabilities. It encompasses text pre-processing, word embedding, and a hybrid classification method. Figure 1 indicates the complete process of the IERT-HDLMWEP model.

Text pre-processing method

Initially, the text pre-processing step involves several typical levels to enhance analysis and reduce the dimensionality of the input data. Information gained from various resources, primarily social media, is often unstructured³¹. Raw data may be noisy and contain grammatical and spelling errors. Therefore, texts need to be cleaned before examining. Considering that several words are insignificant and add nothing to the text (for example, special characters, prepositions, stop words, and punctuation), pre-processing is used to improve the study and reduce the input data dimensionality. Some usual work is involved in the complete process, as shown.

- Substitute negative words. Negations are words like never, not yet, and no that specify the unrelated meaning of phrases or words. The SA aims to substitute the negation with an antonym. For instance, the word not

Author	Objective	Method	Dataset	Result
Kumar et al. ¹¹	A hybrid-fusion-based, innovative, and interpretable multimodal emotion detection method, VISTANet, is presented to categorize an input comprising an image, equivalent text, and speech into discrete emotion segments.	KAAP	IIT-R MMEmoRec dataset	Accuracy of 80.11%
Di Luzio et al. ¹²	A novel explainable AI model to recognize vital facial movements and distinctive emotional feelings.	DNN	Extended Cohn-Kanade dataset (CK+)	–
Fu et al. ¹³	A novel structure for handling inadequate conversational information in the task of MERC that deliberates the higher-order data of modality and multi-frequencies, and fully employs the semantical dependencies.	GRU and Spectral Domain Reconstruction Graph Neural Network	IEMOCAP, CMU-MOSI, and CMU-MOSEI datasets	The Time of SDR-GNN _{mini} is 7.52s
Li et al. ¹⁴	An innovative ERNetCL model incorporates RNN- and MHA techniques in a simplified method to acquire spatial and temporal contextual data.	TE, SE, and CL	MELD, IEMOCAP, EmoryNLP, and DailyDialog dataset	Weighted-F1 of 66.31%, 69.73%, 39.71%, and 53.09% of Four Datasets
Kusal et al. ¹⁵	Presents a hybrid DL model that depends on the convolutional-recurrent network employed to identify the emotions of individuals, depending on the conversational text.	Neural Network Language Model (NNLM), CNN, Recurrent Neural Network (RNN)	Empathetic Dialogues dataset	Accuracy of 73.62%
Feng et al. ¹⁶	Establish a multimodal technique that combines text and speech information to capture the full benefits of emotion-relevant data, surpassing the application of a multiscale MFCC multiview AM.	Multiscale MFCC and Multi-Head Attention (MHA)	IEMOCAP and MSP-IMPROV dataset	IEMOCAP is 0.754 in WA and 0.742 in UA
Zhang et al. ¹⁷	An automated emotion study method is utilized to allow the machine to understand the emotional inference transferred by the person's EEG signals.	CNN and LSTM	DEAP dataset	Accuracy of 92.98%
Omarov and Zhumanov ¹⁸	Projects an innovative Bi-LSTM technique for emotion study and identification in textual content, able to acquire either preceding or upcoming context to enhance performance.	Bi-LSTM	Kaggle Emotion Detection Dataset	Weighted Average of 90% Precision, 90% Recall, and 90% F-Score
Hicham and Nassera ¹⁹	To develop a stacked DL technique for efficient multilingual opinion mining.	RoBERTa-GRU, RoBERTa-LSTM, RoBERTa-BiGRU, RoBERTa-BiLSTM, Adam optimizer, AEDA, SMOTE, GPT	French, English, Arabic, Cohen's kappa, ROC-AUC, Accuracy, MCC, K-fold	High Efficacy, Improved Classification
Mahajan, More, and Shah ²⁰	To develop and evaluate models for recognizing single and mixed emotions in multilingual, code-mixed YouTube comments.	LR, SVM, NB, RF, LSTM, BiLSTM, GRU, CNN	13,000 multilabel YouTube comments, Accuracy, F1 score	SVM Highest Accuracy
Zhu et al. ²¹	To develop an accurate and efficient medical question-answering system using advanced AI techniques.	Knowledge Embedding, Transformer-based Architecture, Knowledge Understanding Layer, Answer Generation Layer	MCMLE and USMLE Datasets	82.92% on MCMLE and 64.02% on USMLE
Khan et al. ²²	To develop an efficient DL method for accurate violence detection in surveillance videos using a two-stream approach.	Two-Stream Architecture, 3D Convolution Network, Background Suppression, Optical Flow Analysis, Depth-Wise 3D Convolutions	RWF2000, RLVS	Accuracy, Efficiency
Arumugam et al. ²³	To develop a novel multimodal by integrating audio, visual, and text inputs for accurate emotion recognition.	AVTEFN, Hybrid Wav2Vec 2.0 + CNN, BERT with Bi-GRU, Attention-Based Fusion	Benchmark Dataset	Accuracy at 98.7%, Precision at 98.2%, Recall, at 97.2%, and F1-score of 97.49%
Khan et al. ²⁴	To enhance multimodal emotion recognition by capturing inter- and intra-modal relationships using a joint transformer-based model.	MER, JMMT	IEMOCAP, MELD	Accuracy, F1-Score
Alyoubi and Alyoubi ²⁵	To develop an optimized transformer-based multimodal emotion recognition framework for accurate emotion classification.	BERT/RoBERTa for Text, wav2vec 2.0 for Speech, ResNet50/VGG16 for Visuals, Cross-Modal Attention, SHAP Explainability	Multimodal EmotionLines (MELD) Dataset	Accuracy, Explainability
Vani et al. ²⁶	To develop and evaluate Text Fusion + through advanced text analysis and audio output.	OCR, NLP, TTS, DL Summarization, NLP-based Q&A Module	Standard Dataset	Summarization Accuracy, User Accessibility
Khan et al. ²⁷	To explore sequence learning techniques for accurate auditory emotion recognition using advanced RNNs.	LSTM, GRU, Bi-Directional LSTM, Deep/Multilayer Architectures	Emotion Dataset	Accuracy, Model Robustness
Ghous, Najam, and Jalal ²⁸	To detect emotional states in individuals with cognitive disabilities using EEG data and advanced ML models.	BF, Downsampling, AAFST, Multi-Class SVM	SEED-IV	Accuracy, Emotion Detection
Patil et al. ²⁹	To predict learning disabilities using handwritten text analysis and intelligent systems.	DNN, Character Confusion Detection, Pattern Extraction Techniques	IAM Handwriting Dataset	Accuracy, Scalability
Mishra et al. ³⁰	To develop a transformer-based NLP model to support mental health monitoring.	Transformer Architecture, GPT-4 and BERT, Adam Optimizer	English Twitter Dataset	94% Accuracy

Table 1. A comparative study of the reviewed techniques.

good is substituted with bad which is the opposite of good. Therefore, a sentence like “The car is not good” is converted into “The car is bad. “Still, particular negative words, like never, not, and no, and negative reductions, like doesn't, mustn't, and couldn't, are frequently a part of stopword lists. Therefore, replacing each negative word and contraction with not, and then correcting this problem after stop-word removal as part of the spelling corrections.

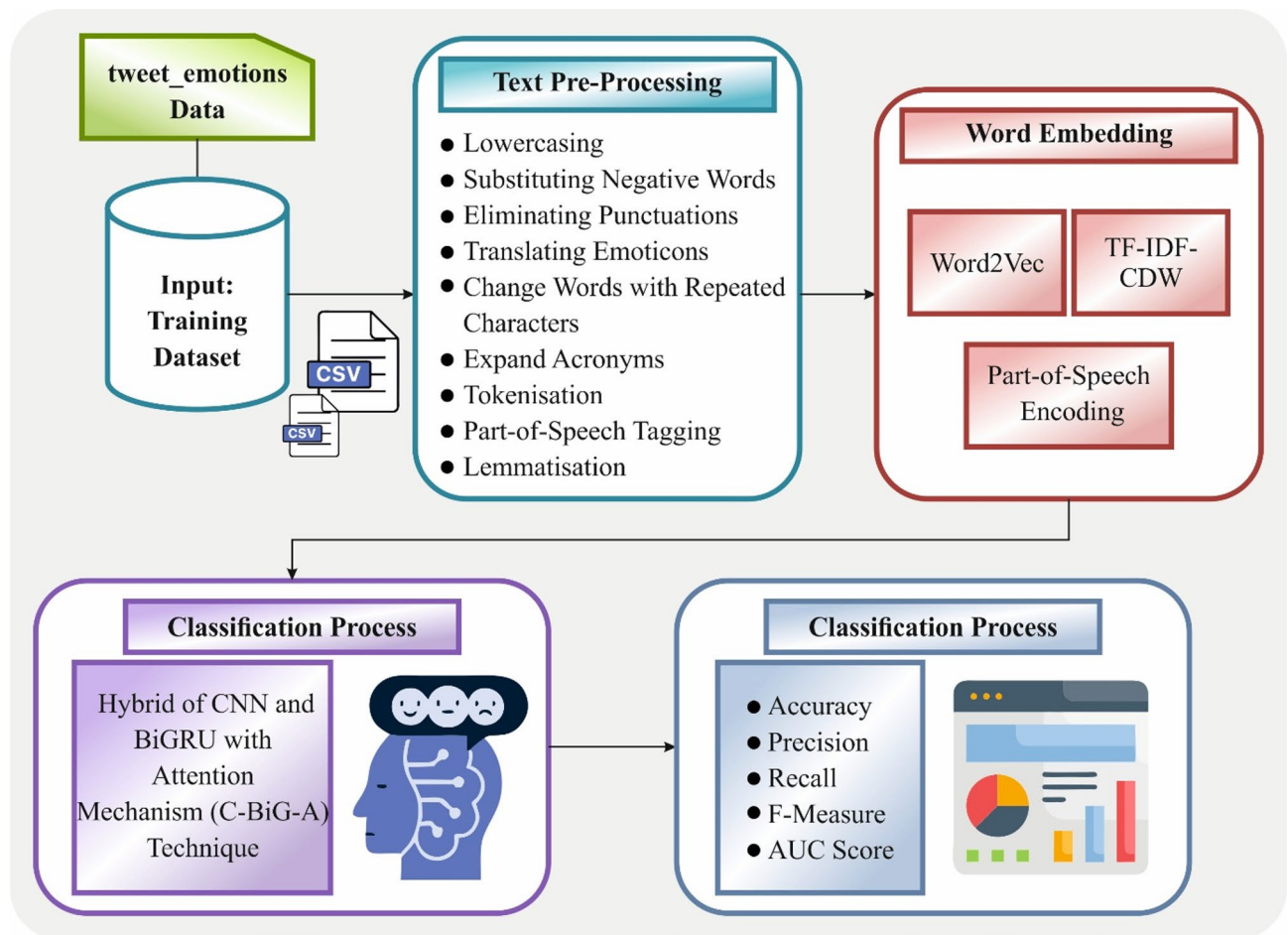


Fig. 1. Overview of the IERT-HDLMWEP model, comprising data preprocessing, word embedding, and a hybrid DL with AM. Evaluation metrics include accuracy, precision, recall, F-measure, and AUC score.

- **Lowercasing:** It converts each character in the dataset to lowercase, as opposed to capitalizing correct words, names, and initials at the beginning of the sentence. Capitalization is challenging on Twitter, as users often avoid capital letters, which can create a tone of ease in the chat, making the messages more conversational. Exchanging each of the texts' letters for a similar circumstance. The word 'Ball', for example, will turn out to be 'ball'.
- **Converting emoticons.** Now, users use emoticons to explain their emotions, thoughts, and feelings. Therefore, converting each emoticon to consistent words will give better outcomes.
- **Removing redundant information** [comprising hashtags (#), additional spaces, punctuation, special characters (\$, &, %, . . .), @username, URL references, stop words (e.g. 'the', 'is', 'at', etc.), non-ASCII English letters and numbers are neglected to preserve the individuality of the information encoding in English]. This type of information fails to predict the emotions expressed by users accurately.
- **Expanded acronyms** to their new words, utilizing the acronym dictionary. Slang and abbreviations are improperly organized words that are often utilized on Twitter. They should be restored to their original form.
- **Change words with repetitive characters** to their English roots. Each frequently uses words with repetitive letters (for example: 'coooooo!') to express their emotions.
- **PoS tagging.** Numerous constructive modules in the text, such as nouns, verbs, adjectives, and adverbs, are recognized in this phase.
- **Tokenization.** It deconstructs text into smaller textual units (for example, documents into expressions, sentences into words).
- **Lemmatization.** There are derivationally related families of words with comparable meanings, like presidential, president, and presidency. This model aims to lower derivational and inflectional forms of the word to the standard base. It uses morphological and vocabulary studies to eliminate inflectional ends and return the dictionary or base form of the word, such as the lemma. If lemmatization were applied to the word 'saw', it would attempt to return both 'see' and 'saw' according to whether it was used as a noun or a verb. It is the procedure of decreasing a specific word to its simpler method, equal to the stem, but it preserves word-related data such as PoS tags.

Word embedding-based Word2Vec approach

For the word embedding process, the IERT-HDLMWEP model employs a hybrid feature representation that combines Word2Vec, TF-IDF-CDW, and POS Encoding for enhanced emotion detection in textual data³². This model is chosen for its superior capability in capturing semantic relationships between words by representing them as continuous vector spaces, which conventional bag-of-words models fail to do. The context-dependent word representations are effectively learned by the Word2Vec model, enabling it to comprehend subtle variations and similarities in language usage. This results in richer feature representations that enhance downstream tasks, such as emotion detection. This model is appropriate for handling extensive textual datasets and is computationally efficient and scalable to large corpora. The robustness of the model is enhanced by its ability to capture both syntactic and semantic data, outperforming simpler encoding techniques in capturing complex language patterns. Overall, Word2Vec presents a balance of effectiveness and efficiency, making it an ideal choice over conventional or more resource-intensive embedding methods.

Word2Vec

Word2Vec is a widely used static embedding model that encodes words as dense vectors in a fixed semantic space, where semantically relevant terms are located. It utilizes two different training mechanisms: CBOW and Skip-gram. It focuses on inferring context words from specified centre words, while CBOW forecasts the central term by combining data from its neighbouring words. To leverage the CBOW model to develop word embeddings, CBOW calculates the likelihood of the centre word w_n , given its adjacent context words w_c , depending on the embeddings of numerous surrounding words w_c .

$$p(w_n|w_c) = \frac{\exp(w_n h_n)}{\sum_{w' \in \text{corpus}} \exp(w' h_n)} \quad (1)$$

Now, h_n depicts an average embedded vector of the nearby contextual window, w_n , and w_c represents the context words. Figure 2 illustrates the architecture of the Word2Vec model.

TF-IDF-CDW

TF-IDF is a commonly endorsed term-weighting model in NLP intended for assessing the importance of words in specific documents. Term frequency (TF) imitates how frequently a term t_i arises inside a document d_j :

$$TF(t_i, d_j) = \frac{n_{i,j}}{|d_j|} \quad (2)$$

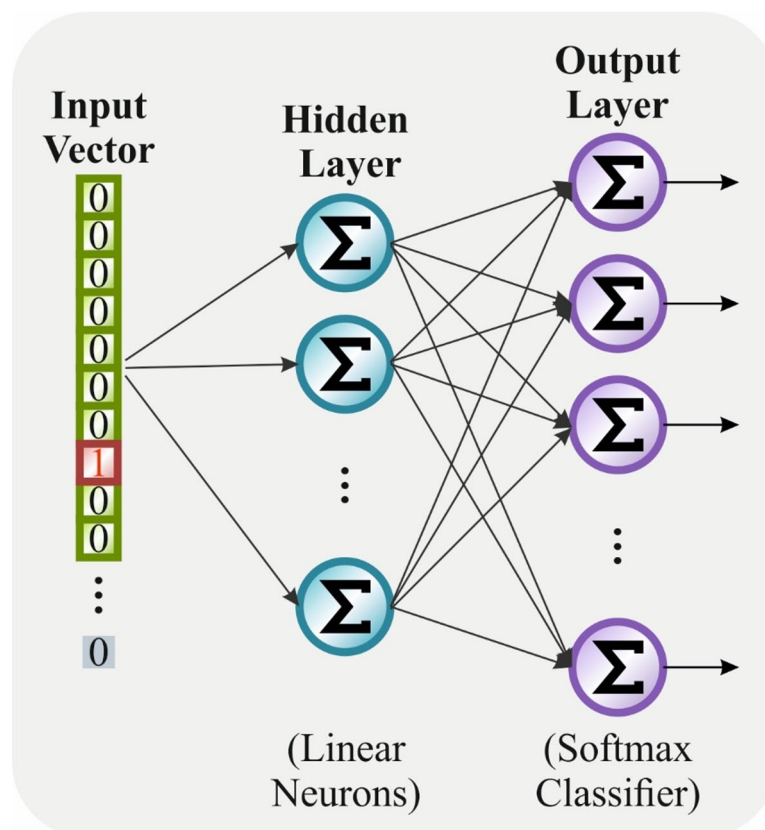


Fig. 2. Architecture of the Word2Vec method consisting of an input vector, a hidden layer with linear neurons, and an output layer with a softmax classifier.

Now, $n_{i,j}$ represents the TF of t_i in document d_j , and $|d_j|$ refers to the overall word count in d_j . IDF assesses the rarity of the term through the overall corpus.

$$IDF(t_i) = \log \frac{|D|}{1 + |\{d_j : t_i \in d_j\}|} \quad (3)$$

Here $|D|$ depicts the total number of documents inside the corpus, and $|\{d_j : t_i \in d_j\}|$ indicates document counts comprising term t_i . To multiply TF and IDF , the TF-IDF score is attained to depict the importance of term t_i in document d_j :

$$TP-IDF(t_i, d_j) = \frac{n_{i,j}}{|d_j|} \times \log \frac{|D|}{1 + |\{d_j : t_i \in d_j\}|} \quad (4)$$

Consequently, some terms that occur regularly within particular classes but rarely elsewhere can receive inadequately low weights. To tackle this restriction, project an improved metric named CDW .

$$CDW(t_i) = 1 + \frac{\sum_{c \in C} P(c|t_i) \log P(c|t_i)}{\log |C| + 1} \quad (5)$$

Now, $P(c|t_i)$ depicts the proportion of documents including t_i which belong to class c , and $|C|$ refers to overall class counts. This model is designed to identify terms that are consistently distributed across classes, while highlighting those with category-specific preferences. CDW is primarily efficient at recognizing terms strongly associated with specific classes. To combine CDW , the methodology allocates a higher weight to category-specific terms. At last, the score of TF-IDF- CDW is calculated to multiply the value of TF-IDF with an equivalent value of CDW , thus acquiring either document- or category-level term significance:

$$TP-IDF-CDW(t_i, d_j) = TP-IDF(t_i, d_j) \cdot IDF(t_i) \cdot CDW(t_i) \quad (6)$$

This formulation enables the method to assign higher significance to terms that contribute more significantly to the distinction class, thereby enhancing the performance of classification and improving the quality of semantic representation.

POS encoding.

POS acts as a basic syntactic indicator in language study, depicting the grammatical part of a word within the sentence. Integrating POS data could enhance the understanding of the grammatical framework and the semantic context. Specifically, in tasks of text classification, POS tags such as verbs and nouns tend to be more effective for classifier outcomes. Therefore, integrating POS aspects enhances the syntactic awareness of the methodology and the performance of classification.

To accept an arbitrary vector-based POS encoding approach. In particular, every dissimilar POS class is allocated an arbitrary 10-D vector that can adjust over training. Then calculate the TF-IDF- CDW weight for every word $TIC_{1:n} = \{tic_1, tic_2, \dots, tic_n\}$ as described in the equation mentioned above. POS encoding vectors $POS_{1:n} = \{pos_1, pos_2, \dots, pos_n\}$ are derived utilizing the method. At last, the improved embedding vector $V_{1:n}$ for every word is developed to scale the Word2Vec vector pv_i with its equivalent TF-IDF- CDW weight tic_j and aimed at a 10-D POS encoding vector pos_i , producing a 310-D representation:

$$V_{1:n} = \{pv_1 \cdot tic_1 + pos_1, pv_2 \cdot tic_2 + pos_2, \dots, pv_n \cdot tic_n + pos_n\} \quad (7)$$

An innovative improved word embedding vector reflects the significant differences of similar words through diverse texts and further acquires POS data. The created embedding matrix emphasizes key semantic components while reducing noise interference in the classification method.

Hybrid classification process

At last, the hybrid of the C-BiG-A technique is employed for the classification process. This hybrid model is chosen for its ability to effectively capture both local patterns and long-range dependencies in textual data. The Bi-GRU model processes sequential data in both forward and backwards directions, and CNN outperforms in extracting spatial features and local n-gram patterns, capturing contextual relationships. The AM model further enhances the model by selectively focusing on the most relevant parts of the input, thereby improving both interpretability and performance. This fusion model effectually balances complexity and efficiency, presenting superior emotion classification accuracy, particularly for complex and mixed emotions compared to standalone models. Its architecture is appropriate for handling varying text lengths and diverse linguistic structures, making it a robust choice over simpler or less adaptive models. Figure 3 specifies the structure of the C-BiG-A technique.

CNN can well remove the features and information³⁵. By selecting the convolutional kernel across various input sequence positions, it can successfully capture the changing local features and patterns of the sequence data. The convolutional layer takes local features from the time series data by using a sliding window. While the RNN has the benefit of being able to remove contextually related data, it may admit a minimum range of contextual information and suffer from the long-term dependency problem. Then, it is essential to introduce a threshold mechanism into the RNN architecture to maintain the required data. It can not only be an efficient solution to the problems of gradient explosion and gradient vanishing, but also partially address the disadvantages of transferring information over longer distances. The generally applied RNN frameworks with gate mechanisms are the LSTM and GRU structures, which possess temporal solid processing abilities. Still, if GRU fails to

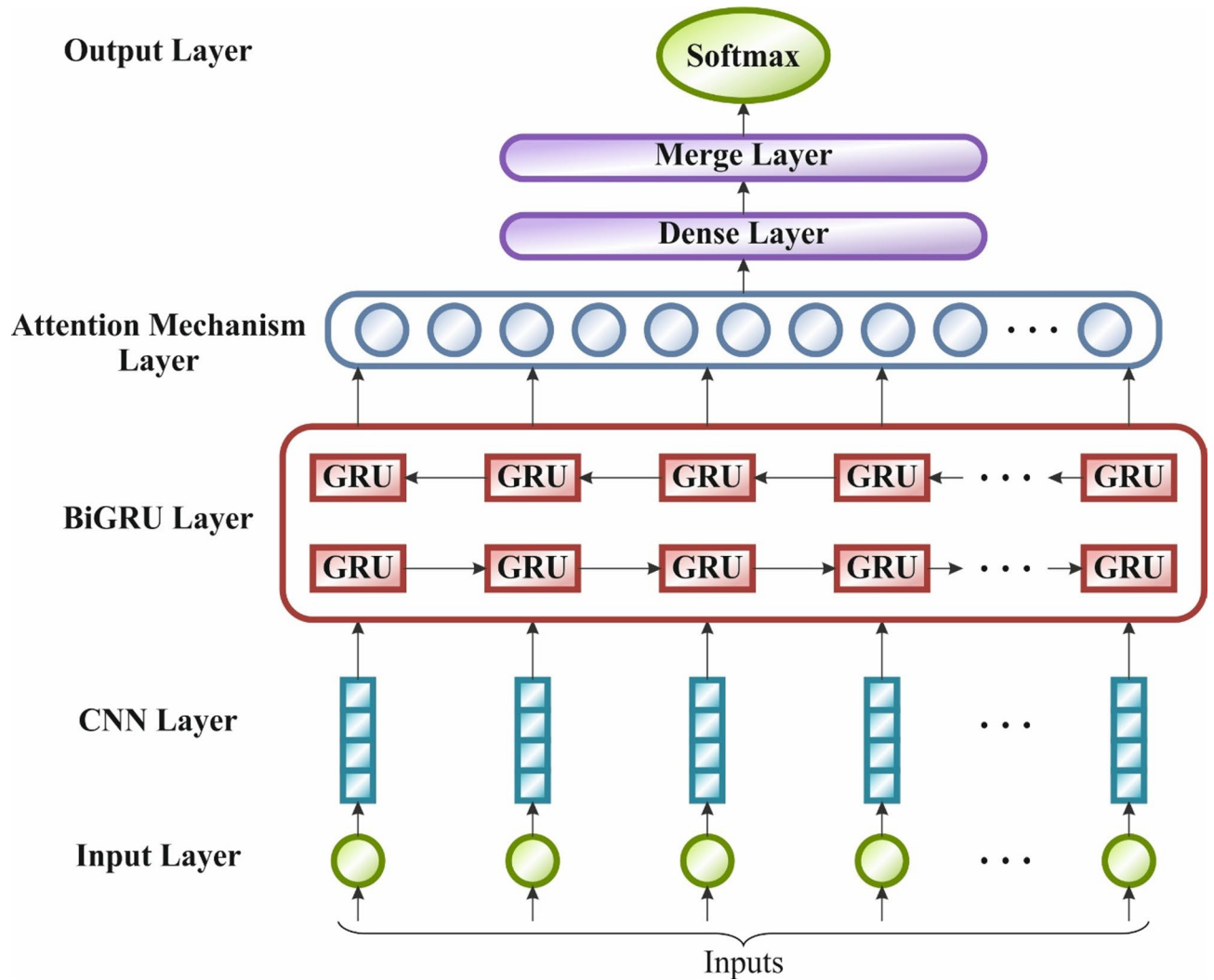


Fig. 3. Framework of C-BiG-A technique consisting of input, convolutional, recurrent, attention, and dense layers resulting in the softmax-based output.

maintain memory, the internal architecture of the network has only dual gates that prevent the memory area, decrease the training parameters, and outcomes at the fastest training speed. The fundamental structure of a GRU method, and the forward GRU computational equation, is as shown:

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \quad (8)$$

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \quad (9)$$

$$\tilde{h}_t = \tanh(W_h[r_t * h_{t-1}, x_t] + b_h) \quad (10)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (11)$$

$$y_t = \text{softmax}(W_0 \cdot h_t + b_0) \quad (12)$$

whereas, x_t characterizes the input vector of the present time step; W_r , W_z , and W_h characterize the weighted matrices; and b_r , b_z , and b_h symbolize the biased vectors.

The Bi-GRU model can learn either past or future information. It may learn the input sequence information more broadly to prevent missing information after processing longer sequences. g'_t and g_t represent the consistent outputs for the Backwards and Forward GRU layers at moment t , correspondingly presented as shown:

$$g_t^{\text{forward}} = \text{GRU}(x_t, h_{t-1}^{\text{forward}}) \quad (13)$$

$$g_t^{\text{backward}} = \text{GRU}(x_t, h_{t+1}^{\text{backward}}) \quad (14)$$

The Bi-GRU output O_t is presented as shown:

$$O_t = \overrightarrow{W}g_t + \overleftarrow{W}g'_t \leftarrow + b_t \tag{15}$$

whereas \overleftarrow{W} and \overrightarrow{W} represent the weighted matrices for the backwards and forward GRU structures, individually, while b_t symbolizes the output layer's bias vector.

They present a linear generalized attention mechanism, which achieves near-linear growth in both time and space complexity. This model significantly reduces training time, particularly after processing longer sequences. The attention matrix $A \in \mathbb{R}^{L \times L}$ as demonstrated:

$$A(i, j) = K(q_i^T, k_j^T) \tag{16}$$

Here, k_j and q_i characterize the j th and i th row vectors of the key K and query Q , respectively. The kernel K is represented as demonstrated:

$$K(x, y) = \mathbb{E} \left[\varphi(x)^T \varphi(y) \right] \tag{17}$$

Now, $\varphi(x)$ signifies mapping functions. If $Q', K' \in \mathbb{R}^{L \times p}$, their row vectors are embodied as $\varphi(q_i^T)$ and $\varphi(k_j^T)$, individually. The efficient attention, depending on the kernel description, is formulated as shown:

$$\widehat{Att} \leftrightarrow (Q, K, V) = \widehat{D}^{-1} (BV) \tag{18}$$

whereas, $B = Q' (K')^T$ and $\widehat{D} = diag(B1_L)$. Now, \widehat{Att} represents the estimated attention, while the brackets specify the computing sequence. Table 2 specifies the key hyperparameters of the C-BiG-A technique.

Component	Hyperparameter	Typical Values/Description
Input	Input Shape	(sequence_length, embedding_dim) (e.g., (100, 300) for NLP tasks)
	Padding	'same', 'valid'
	Embedding Type (if any)	Pretrained (e.g., GloVe, Word2Vec) or Trainable
	Embedding Dimension	100, 200, 300 (depending on task and resource availability)
CNN Block	Number of Conv1D Filters	64, 128, 256
	Kernel Size	3, 5, 7
	Stride	1
	Activation Function	ReLU, LeakyReLU
	MaxPooling Size	2, 3
	Dropout (CNN Output)	0.2–0.5
Bi-GRU Block	GRU Units	64, 128, 256
	Layer Numbers	1–2 (stacked GRUs)
	Bidirectional	True
	Return Sequences	True (required for attention layer)
	Recurrent Dropout	0.1–0.3
Attention Layer	Attention Type	Bahdanau, Self-Attention
	Attention Units (if used)	64, 128
	Output Aggregation	Context vector (weighted sum over time steps)
Fully Connected	Dense Layer Units	64, 128, 256
	Activation Function	ReLU
	Dropout Rate	0.2–0.5
Output Layer	Output Units	Depends on task: 1 (binary), n (multi-class)
	Activation Function	Sigmoid (binary), Softmax (multi-class)
Training	Optimizer	Adam, AdamW
	Learning Rate	1e-3, 1e-4, 5e-4
	Batch Size	32, 64, 128
	Epoch Number	10–100 (depending on convergence and early stopping)
	Loss Function	Binary Crossentropy, Categorical Crossentropy
	Metrics	Accuracy, F1-score, Precision, Recall
	Early Stopping	Patience: 5–10 epochs

Table 2. Key hyperparameters of the C-BiG-A model.

Sentiment	No. of samples
"Sadness"	"5165"
"Enthusiasm"	"759"
"Neutral"	"8638"
"Worry"	"8459"
"Surprise"	"2187"
"Love"	"3842"
"Fun"	"1776"
"Hate"	"1323"
"Happiness"	"5209"
"Boredom"	"179"
"Relief"	"1526"
"Anger"	"110"
Total Samples	39,173

Table 3. Details of the dataset.

Sentiment	Sample text
Sadness	"I'm having a problem with my photo here on Twitter amf!!!!...can't see my face!"
Enthusiasm	"Counting the minutes when my dad's g-friend is leaving"
Neutral	"So tense.so emotional...so not me."
Worry	"Feels like I'm going to cough up a lung"
Surprise	"No churchill downs for me tomorrow"
Love	"Looking at old pictures... I love it & miss the good old days"
Fun	"I really want a Blackberry my sidekick is hella wack. Night"
Hate	"My stomach is killing me do i can't sleep"
Happiness	"Is dreading going to work BUT...its friiiiday!! whoop!!!"
Boredom	"Don't know what to do today. amazing weather and i'm sitting here, great"
Relief	"At least my wish came true, it is not hot anymore...just rainy"
Anger	"The Internet is painfully slow today"

Table 4. Sample text.

Experimental validation

The performance assessment of the IERT-HDLMWEP method is examined under the Emotion detection from the Text dataset³⁴. The technique runs on Python 3.6.5 with an i5-8600k CPU, 4GB GPU, 16GB RAM, 250GB SSD, and 1 TB HDD, using a 0.01 learning rate, ReLU, 50 epochs, 0.5 dropout, and batch size 5. This dataset consists a total of 39,173 samples, categorized into 12 sentiments, as outlined in Table 3 below. Table 4 represents the sample text.

Figure 4 shows the confusion matrices formed by the IERT-HDLMWEP approach at 80:20 and 70:30 ratios of TRPHE to TSPHE. The results state that the IERT-HDLMWEP technique effectively detects and identifies each class.

Table 5; Fig. 5 present the textual emotion detection outcome of the IERT-HDLMWEP technique at 80:20. Under 80% TRPHE, the IERT-HDLMWEP technique attains an average $accu_y$ of 99.64%, $prec_n$ of 94.95%, $recal$ of 89.19%, $F_{Measure}$ of 91.22%, AUC_{Score} of 94.49%, and Kappa of 94.56%. Also, on 20% TSPHE, the IERT-HDLMWEP model obtains an average $accu_y$ of 99.67%, $prec_n$ of 93.61%, $recal$ of 88.54%, $F_{Measure}$ of 89.97%, AUC_{Score} of 94.18%, and Kappa of 94.25%.

Table 6; Fig. 6 portray the textual emotion detection outcome of the IERT-HDLMWEP method on 70:30. Based on 70% TRPHE, the IERT-HDLMWEP method attains an average $accu_y$ of 99.31%, $prec_n$ of 90.71%, $recal$ of 80.81%, $F_{Measure}$ of 82.83%, AUC_{Score} of 90.12%, and Kappa of 90.28%. Similarly, on 30% TSPHE, the IERT-HDLMWEP technique obtains an average $accu_y$ of 99.38%, $prec_n$ of 93.88%, $recal$ of 80.87%, $F_{Measure}$ of 82.93%, AUC_{Score} of 90.26%, and Kappa of 90.32%.

Figure 7 exemplifies the training (TRAIN) $accu_y$ and validation (VALID) $accu_y$ of an IERT-HDLMWEP approach at an 80:20 ratio over 25 epochs. Primarily, both TRAIN and VALID $accu_y$ rise quickly, representing efficient pattern learning from the data. Around the epoch, the VALID $accu_y$ slightly exceeds the training accuracy, signifying good generalization without over-fitting. As training advances, it reflects higher performance and a lower performance gap between TRAIN and VALID. The close alignment of both curves during training implies that the model is well-regularised and well-generalized. This shows the approach's stronger capability in learning and retaining beneficial features across both seen and unseen data.

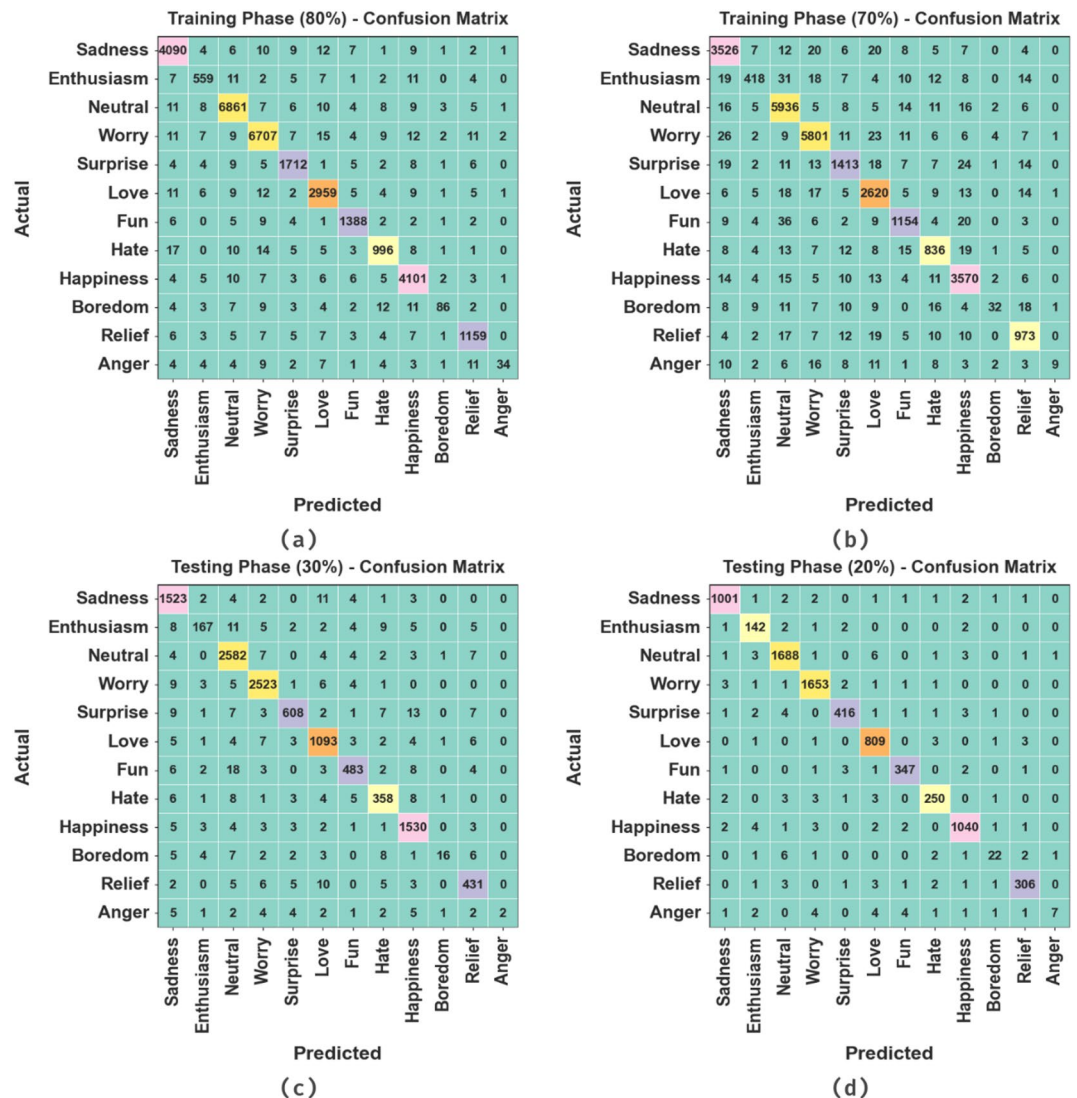


Fig. 4. Confusion matrices of (a–c) TRPHE of 80% and 70% and (b–d) TSPHE of 20% and 30%.

Figure 8 demonstrates the TRAIN and VALID losses of the IERT-HDLMWEP technique at 80:20 over 25 epochs. Initially, both TRAIN and VALID losses are higher, showing that the method begins with a partial understanding of the data. As training evolves, both losses continually decrease, indicating that the technique is efficiently learning and optimizing its parameters. The close alignment between the TRAIN and VALID loss curves in training implies that the model hasn't overfitted and retains good generalization to unseen data. This reliable and steady decrease in loss shows a well-trained, stable, and consistent DL model.

In Fig. 9, the precision-recall (PR) inspection study of the IERT-HDLMWEP methodology on the 80:20 dataset provides insights into its performance by charting Precision against Recall for all classes. The figure illustrates that the IERT-HDLMWEP technique consistently achieves increased PR values across diverse classes, indicating its potential in maintaining a significant share of true positive predictions among all positive predictions (precision), while also capturing a substantial portion of actual positives (recall). The steady improvement in PR results across each class depicts the efficiency of the IERT-HDLMWEP technique during the classifier process.

In Fig. 10, the ROC analysis of the IERT-HDLMWEP technique is examined under an 80:20 ratio. The results indicate that the IERT-HDLMWEP method achieves elevated ROC values across all classes, demonstrating a significant ability to differentiate between class labels. This consistent pattern of increased values of ROC across numerous class labels indicates the effective results of the IERT-HDLMWEP method in class prediction, underscoring the robust nature of the classification process.

Table 7; Fig. 11 demonstrate the comparative analysis of the IERT-HDLMWEP method with current techniques under various metrics^{19,20,35,36}. The outcomes emphasized that the IERT-HDLMWEP model got higher *accu_y*, *prec_n*, *recal_i*, and *F_{Measure}* of 99.67%, 93.61%, 88.54%, and 89.97%, respectively. While the existing methodologies, namely RoBERTa, BiGRU, RF, U-Net, Improved DBN-SVM, Bi-GRU, XLNet, DAN, Bi-LSTM, and SVM, have shown worse performance under various metrics.

Class Labels	$Accu_y$	$Prec_n$	$Recal_l$	$F_{Measure}$	AUC_{Score}	Kappa
TRPHE (80%)						
Sadness	99.53	97.96	98.51	98.23	99.10	99.16
Enthusiasm	99.70	92.70	91.79	92.24	95.82	95.89
Neutral	99.50	98.78	98.96	98.87	99.31	99.37
Worry	99.43	98.66	98.69	98.68	99.16	99.24
Surprise	99.69	97.11	97.44	97.27	98.63	98.69
Love	99.55	97.53	97.85	97.69	98.79	98.87
Fun	99.77	97.13	97.75	97.44	98.80	98.88
Hate	99.63	94.95	93.96	94.45	96.89	96.94
Happiness	99.55	97.88	98.75	98.31	99.21	99.28
Boredom	99.77	86.00	60.14	70.78	80.05	80.10
Relief	99.68	95.71	96.02	95.86	97.93	97.99
Anger	99.82	85.00	40.48	54.84	70.23	70.30
Average	99.64	94.95	89.19	91.22	94.49	94.56
TSPHE (20%)						
Sadness	99.69	98.82	98.82	98.82	99.32	99.38
Enthusiasm	99.69	89.87	94.67	92.21	97.23	97.29
Neutral	99.50	98.71	99.00	98.86	99.32	99.40
Worry	99.66	98.98	99.40	99.19	99.56	99.64
Surprise	99.71	97.88	96.74	97.31	98.31	98.37
Love	99.60	97.35	98.90	98.12	99.29	99.36
Fun	99.76	97.20	97.47	97.34	98.67	98.74
Hate	99.68	95.42	95.06	95.24	97.45	97.53
Happiness	99.60	98.58	98.48	98.53	99.13	99.19
Boredom	99.73	75.86	61.11	67.69	80.51	80.58
Relief	99.71	96.84	95.92	96.38	97.90	97.97
Anger	99.73	77.78	26.92	40.00	63.45	63.53
Average	99.67	93.61	88.54	89.97	94.18	94.25

Table 5. Textual emotion detection outcome of IERT-HDLMWEP model under 80%:20%.

In Table 8; Fig. 12, the computational time (CT) of the IERT-HDLMWEP methodology is compared with that of existing techniques. The IERT-HDLMWEP methodology presents a lower CT of 8.06 s while the RoBERTa, BiGRU, RF, U-Net, Improved DBN-SVM, Bi-GRU, XLNet, DAN, Bi-LSTM, and SVM methodologies attained superior CTs of 16.99 s, 13.00 s, 17.90 s, 18.56 s, 21.82 s, 28.89 s, 15.36 s, 21.64 s, 12.59 s, and 10.43 s, respectively.

Table 9; Fig. 13 demonstrates the error analysis of the IERT-HDLMWEP technique with existing methods. The error analysis of various models exhibit significant differences in performance across evaluation metrics. RoBERTa achieved the highest overall results with $accu_y$ of 10.25%, $prec_n$ of 13.19%, $recal_l$ of 13.27%, and $F_{Measure}$ of 17.07%. U-Net followed closely with $accu_y$ of 11.03%, $prec_n$ of 13.80%, $recal_l$ of 13.95%, and $F_{Measure}$ of 17.63%. In contrast, conventional models such as RF and SVM performed poorly, with RF illustrating an $accu_y$ of 1.87%, $prec_n$ of 6.97%, $recal_l$ of 17.86%, and $F_{Measure}$ of 17.84%, while SVM had $accu_y$ of 1.97%, $prec_n$ of 8.93%, $recal_l$ of 13.58%, and $F_{Measure}$ of 15.67%. BiGRU and Bi-GRU showed low $accu_y$ of 5.21% and 2.61% respectively, but relatively high recall values above 17%, indicating a bias in detecting positives. XLNet balanced its metrics well, achieving $accu_y$ of 3.43%, $prec_n$ of 7.82%, $recal_l$ of 19.25%, and the highest $F_{Measure}$ of 18.78%. The IERT-HDLMWEP model attained an $accu_y$ of 0.33%, $prec_n$ of 6.39%, $recal_l$ of 11.46%, and $F_{Measure}$ of 10.03%, indicating severe issues in both $prec_n$ and $recal_l$. These results indicate that as more advanced representations and architectural enhancements are introduced, both $prec_n$ and $recal_l$ improve consistently, resulting in more balanced and reliable classification performance.

Table 10; Fig. 14 indicates the ablation study of the IERT-HDLMWEP approach. The ablation study illustrates the progressive improvement in performance as diverse components are integrated into the model. Starting with Word2Vec, the baseline model attained an $accu_y$ of 97.06%, $prec_n$ of 90.92%, $recal_l$ of 85.81%, and $F_{Measure}$ of 87.06%. The TF-IDF-CDW approach slightly improved performance with $accu_y$ of 97.81%, $prec_n$ of 91.46%, $recal_l$ of 86.57%, and $F_{Measure}$ of 87.75%. Incorporating the PSE technique further enhanced the metrics to $accu_y$ of 98.50%, $prec_n$ of 92.23%, $recal_l$ of 87.30%, and $F_{Measure}$ of 88.51%. The C-BiG-A model significantly outperformed previous variants, reaching $accu_y$ of 99.03%, $prec_n$ of 92.98%, $recal_l$ of 87.80%, and $F_{Measure}$ of 89.28%. The final IERT-HDLMWEP model attained the best results across all metrics, with an $accu_y$ of 99.67%, $prec_n$ of 93.61%, $recal_l$ of 88.54%, and $F_{Measure}$ of 89.97%, confirming the efficiency of the complete hybrid architecture.

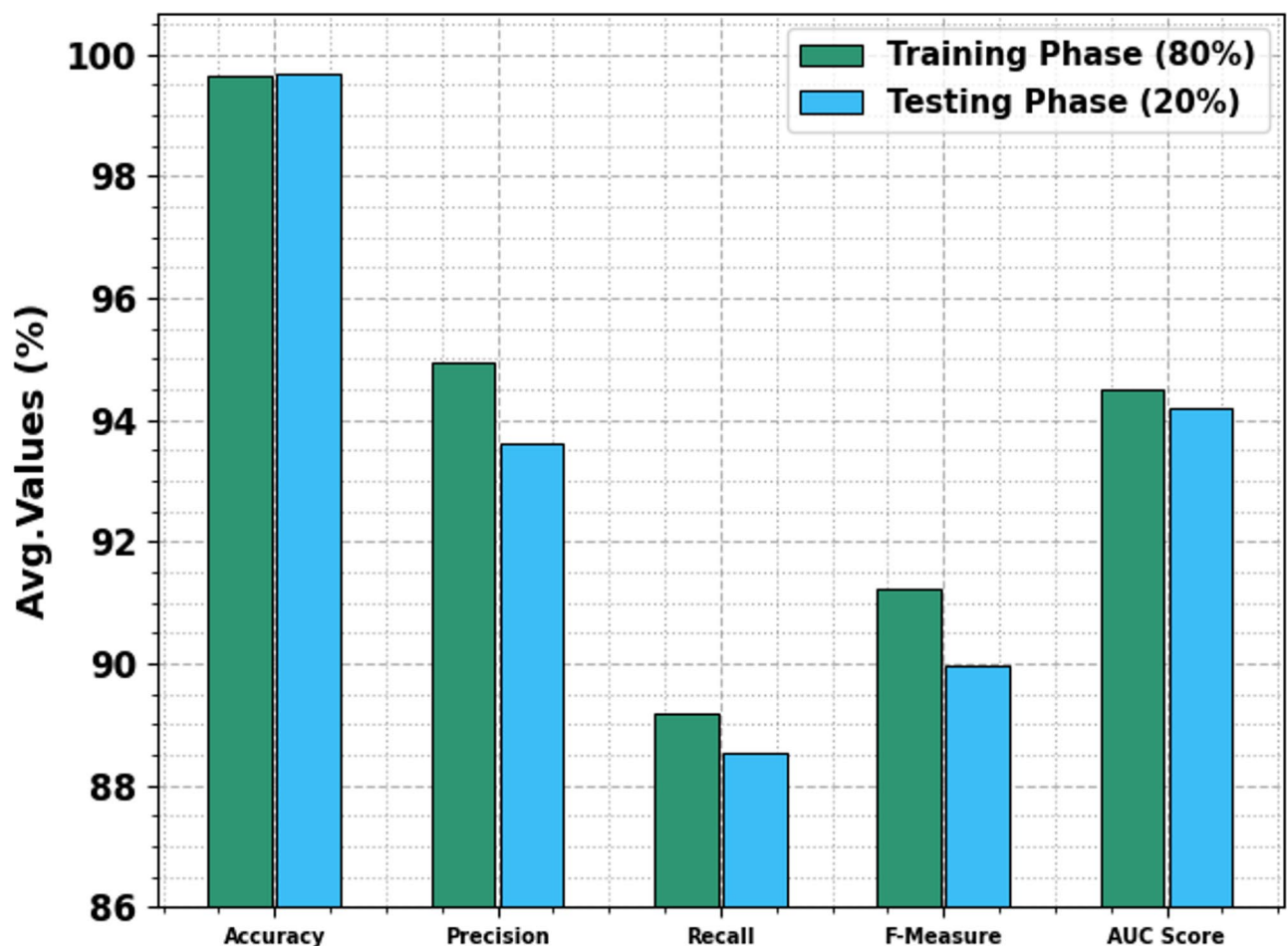


Fig. 5. Average values of the IERT-HDLMWEP model under 80%:20%.

Conclusion

In this study, a novel emotion detection model, named the IERT-HDLMWEP method, is developed to accurately identify and interpret emotions in text, thereby enhancing communication support for people with disabilities. Primarily, the text pre-processing stage involves several typical levels to increase the study and minimize the dimensionality of input data. For the word embedding process, the IERT-HDLMWEP method generates a hybrid feature representation by integrating pre-trained Word2Vec vectors weighted using a TF-IDF-CDW and enriched with POS feature vectors to enhance emotion detection in textual data. At last, the hybrid of the C-BiG-A technique is employed for the classification process. A comprehensive simulation was implemented to verify the performance of the IERT-HDLMWEP methodology. The empirical results indicated that the IERT-HDLMWEP methodology emphasized development over other recent techniques. The limitations of the IERT-HDLMWEP methodology include its dependence on general textual data. Though disability-specific data is not required due to the text-based nature of the approach, the present model may not capture all contextual variations relevant to specific user groups and may not generalize well to other languages or domains with diverse linguistic styles or informal expressions. The performance may also be affected when dealing with highly informal or noisy text, which is typical of some real-world communications. Future work should focus on enhancing the system's robustness to diverse linguistic styles and expanding its adaptability to various domains. Incorporating user feedback mechanisms could additionally personalize emotion recognition. Additionally, integrating this text-based model with other assistive technologies may improve overall support for people with intellectual disabilities. Finally, developing real-time processing capabilities and user-friendly interfaces will be crucial for practical deployment.

Class Labels	$Accu_y$	$Prec_n$	$Recall_l$	$F_{Measure}$	AUC_{Score}	Kappa
TRPHE (70%)						
Sadness	99.17	96.21	97.54	96.87	98.48	98.55
Enthusiasm	99.38	90.09	77.26	83.18	88.55	88.63
Neutral	99.03	97.07	98.54	97.80	98.85	98.92
Worry	99.17	97.96	98.21	98.08	98.82	98.87
Surprise	99.25	93.95	92.41	93.18	96.03	96.11
Love	99.15	94.96	96.57	95.76	98.00	98.07
Fun	99.37	93.52	92.54	93.03	96.12	96.18
Hate	99.30	89.41	90.09	89.75	94.86	94.94
Happiness	99.22	96.49	97.70	97.09	98.58	98.65
Boredom	99.62	72.73	25.60	37.87	62.78	62.86
Relief	99.34	91.19	91.88	91.53	95.76	95.84
Anger	99.73	75.00	11.39	19.78	55.69	55.76
Average	99.31	90.71	80.81	82.83	90.21	90.28
TSPHE (30%)						
Sadness	99.23	95.97	98.26	97.10	98.82	98.89
Enthusiasm	99.41	90.27	76.61	82.88	88.22	88.27
Neutral	99.09	97.18	98.78	97.97	98.98	99.03
Worry	99.39	98.32	98.86	98.59	99.20	99.27
Surprise	99.38	96.35	92.40	94.34	96.10	96.17
Love	99.28	95.71	96.81	96.26	98.18	98.25
Fun	99.38	94.71	91.30	92.97	95.53	95.58
Hate	99.34	89.95	90.63	90.29	95.14	95.20
Happiness	99.34	96.65	98.39	97.51	98.94	99.00
Boredom	99.64	80.00	29.63	43.24	64.80	64.86
Relief	99.35	91.51	92.29	91.90	95.97	96.04
Anger	99.75	100.00	06.45	12.12	53.23	53.31
Average	99.38	93.88	80.87	82.93	90.26	90.32

Table 6. Textual emotion detection outcome of IERT-HDLMWEP model under 80%:20%.

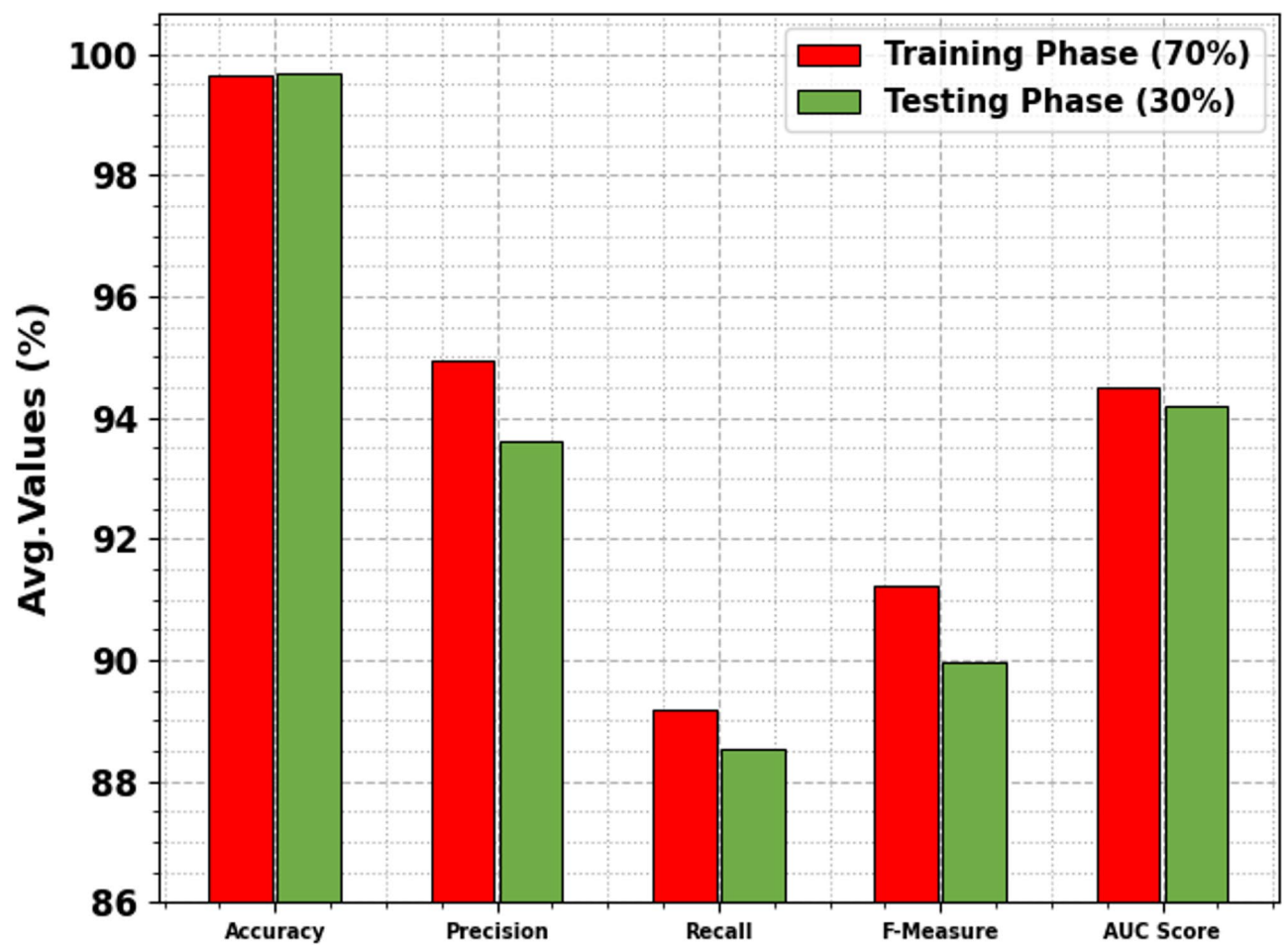


Fig. 6. Average values of the IERT-HDLMWEP model under 70%:30%.

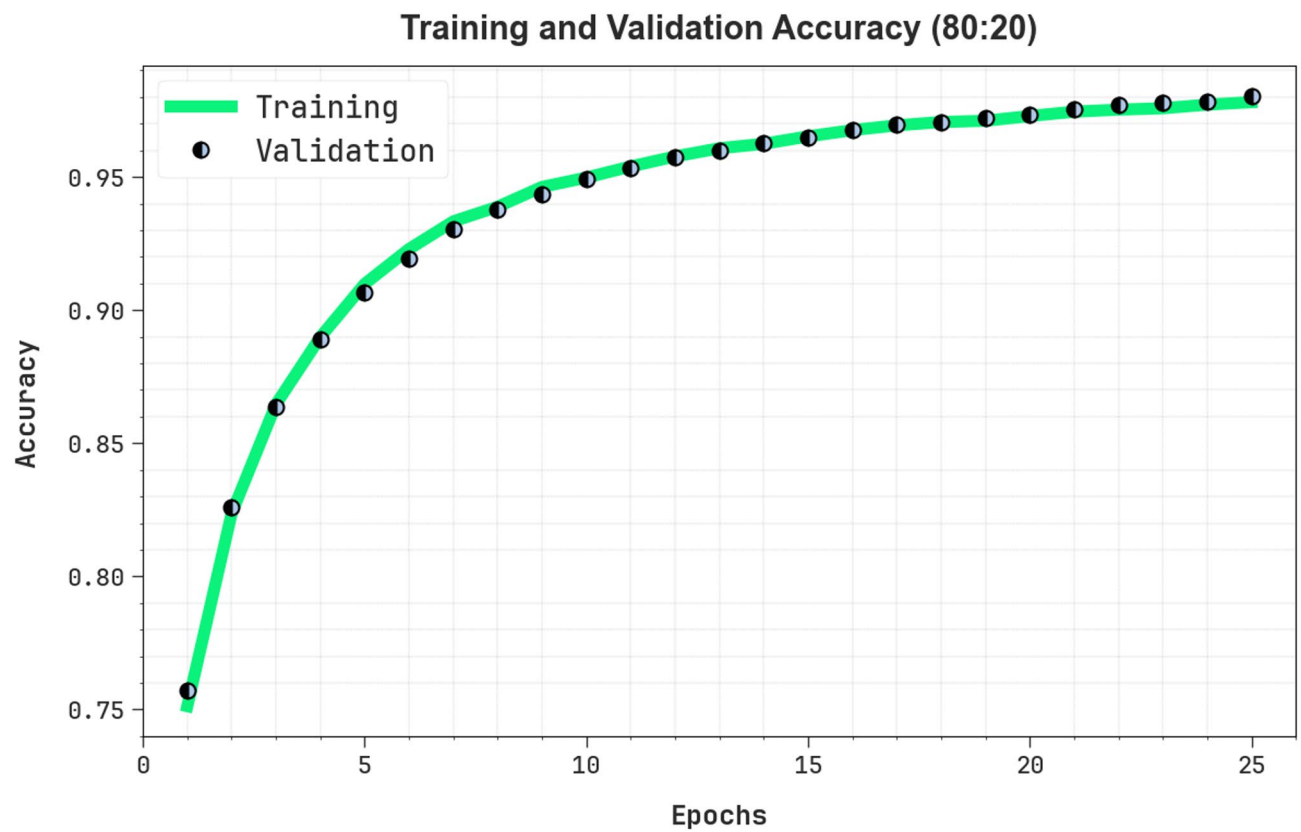


Fig. 7. $Accu_y$ curve of the IERT-HDLMWEP model under 80:20.

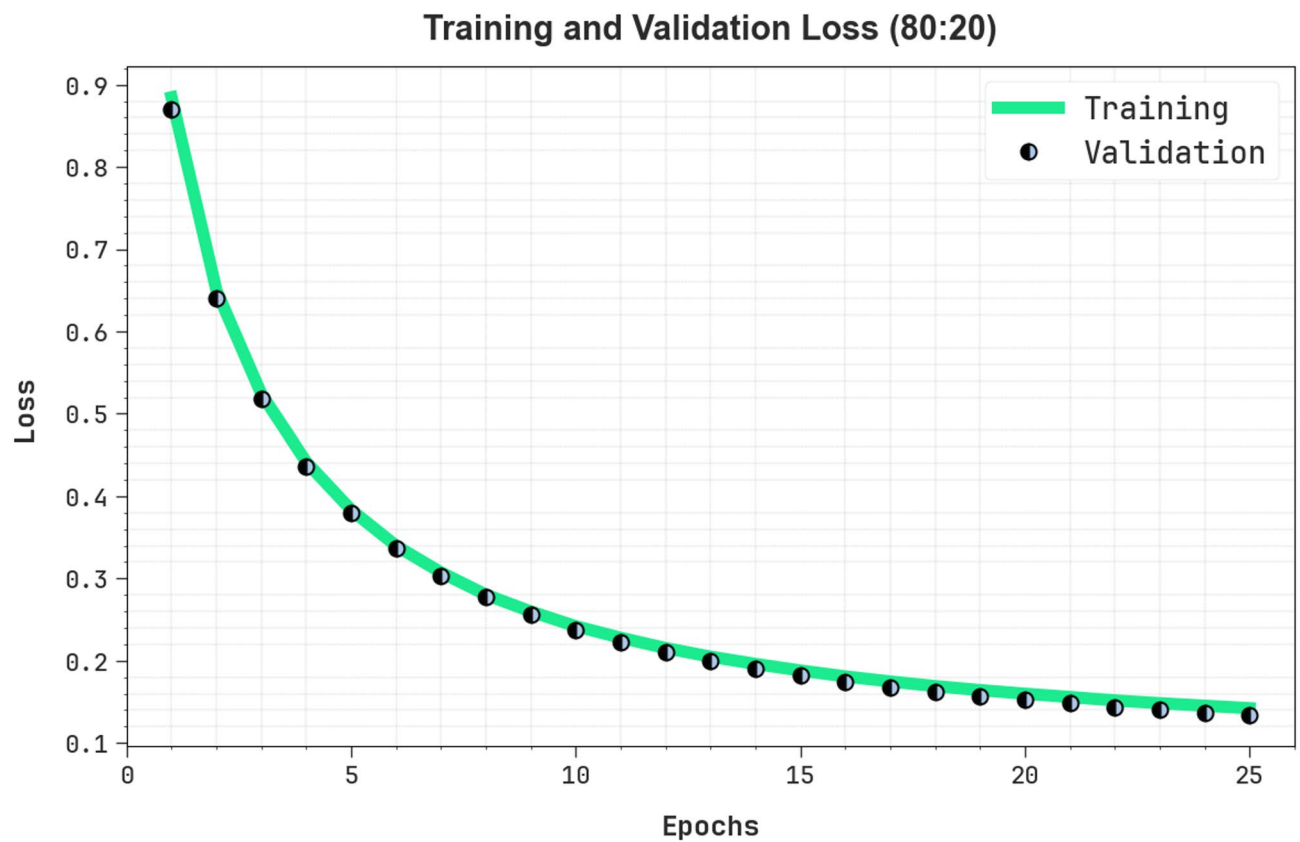


Fig. 8. Loss curve of the IERT-HDLMWEP model under 80:20.

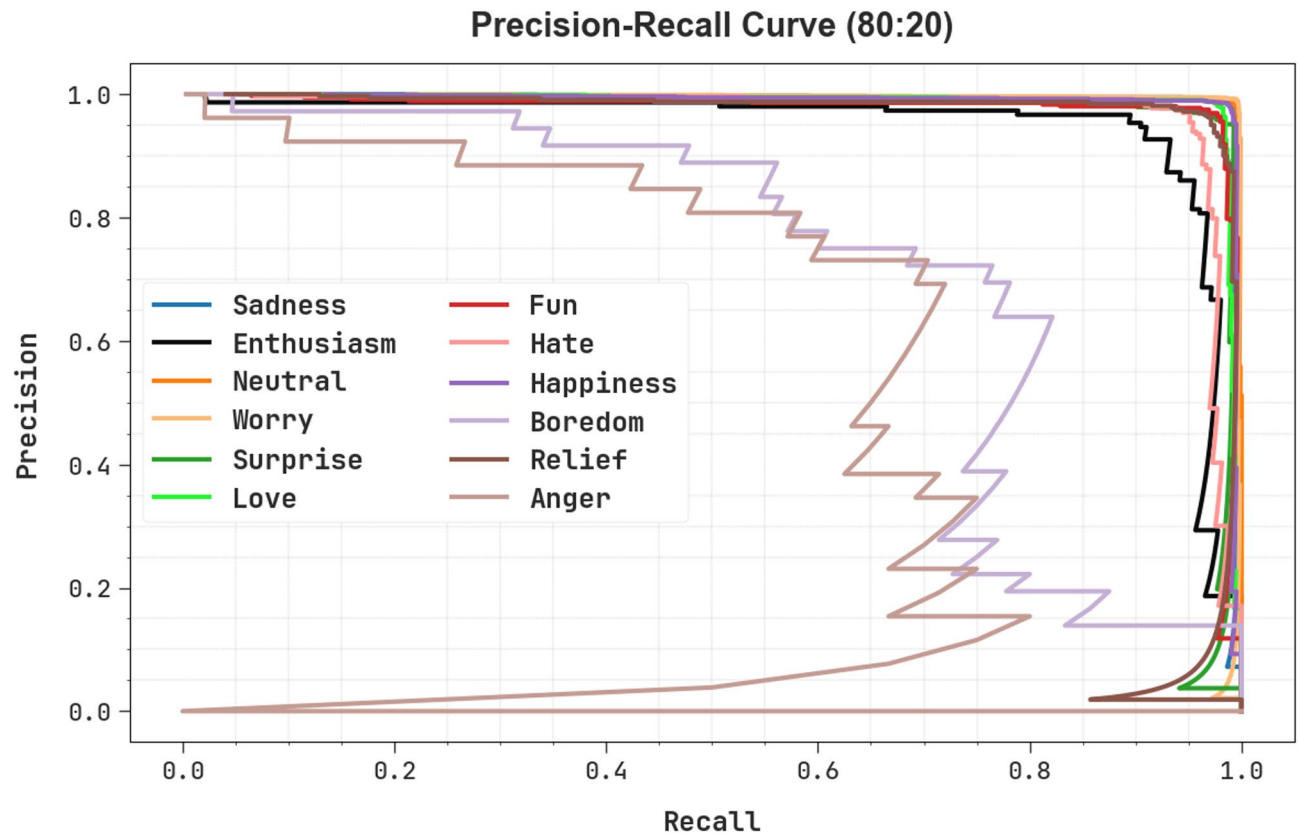


Fig. 9. PR curve of the IERT-HDLMWEP method under 80:20.

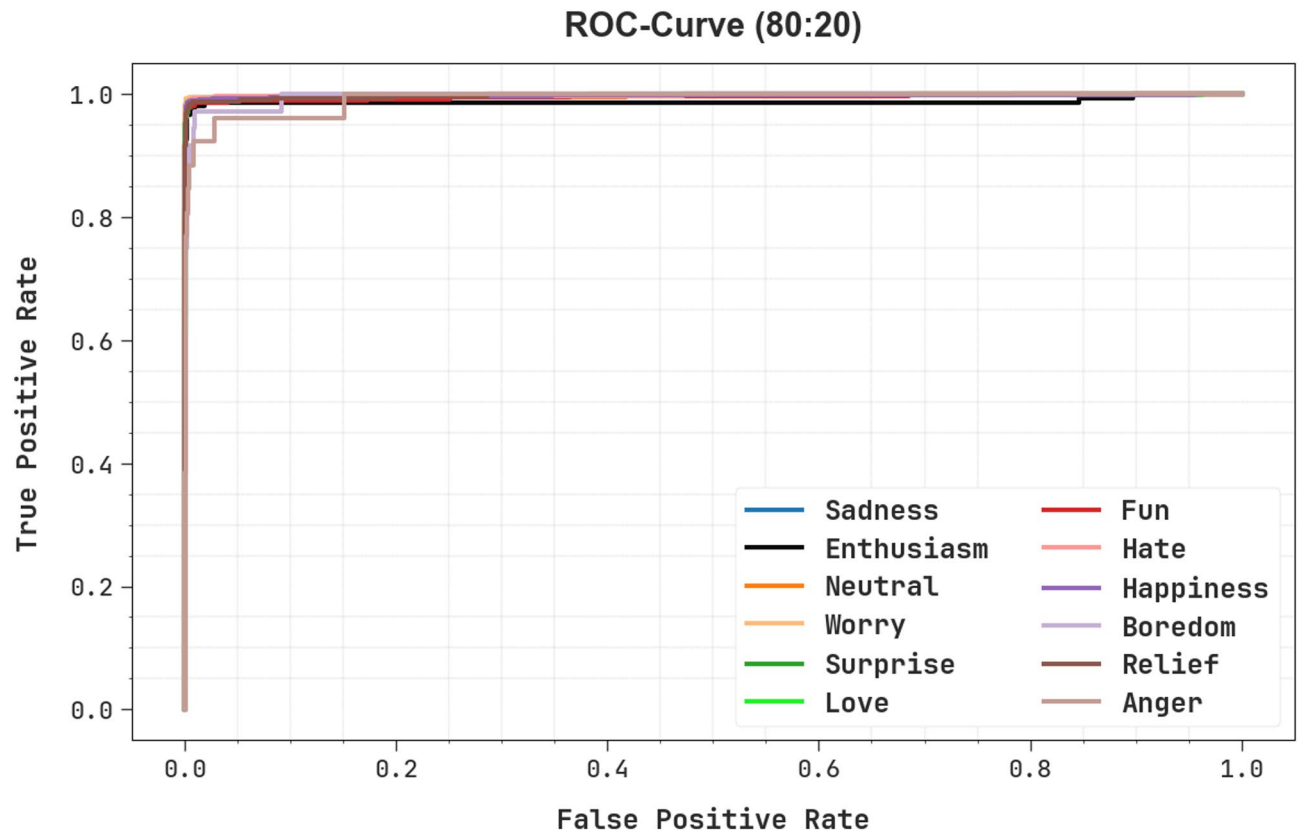


Fig. 10. ROC curve of the IERT-HDLMWEP method at 80:20.

Model	$Accu_y$	$Prec_n$	$Recal$	$F_{Measure}$
RoBERTa	89.75	86.81	86.73	82.93
BiGRU	94.79	93.88	82.34	86.28
RF	98.13	93.03	82.14	82.16
U-Net	88.97	86.20	86.05	82.37
Improved DBN-SVM	94.17	93.11	81.63	85.53
Bi-GRU	97.39	92.49	81.54	81.61
XLNet	96.57	92.18	80.75	81.22
DAN	90.06	91.37	82.63	85.55
Bi-LSTM	91.64	89.85	82.61	83.90
SVM	98.03	91.07	86.42	84.33
IERT-HDLMWEP	99.67	93.61	88.54	89.97

Table 7. Comparative analysis of the IERT-HDLMWEP model with existing methods^{19,20,35,36}.

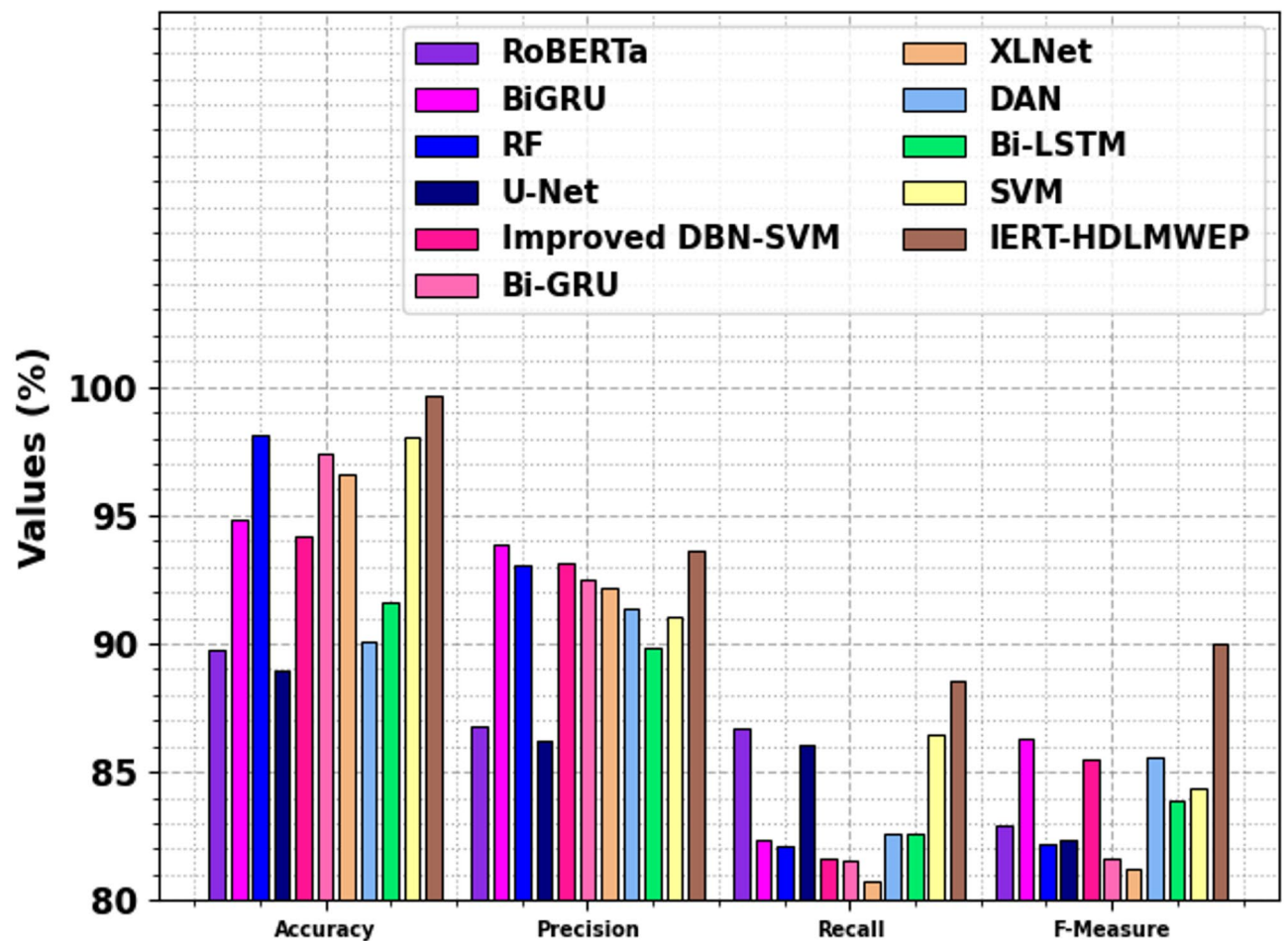


Fig. 11. Comparative analysis of the IERT-HDLMWEP model with existing methods.

Model	CT (sec)
RoBERTa	16.99
BiGRU	13.00
RF	17.90
U-Net	18.56
Improved DBN-SVM	21.82
Bi-GRU	28.89
XLNet	15.36
DAN	21.64
Bi-LSTM	12.59
SVM	10.43
IERT-HDLMWEP	8.06

Table 8. CT outcome of IERT-HDLMWEP methodology with existing models.

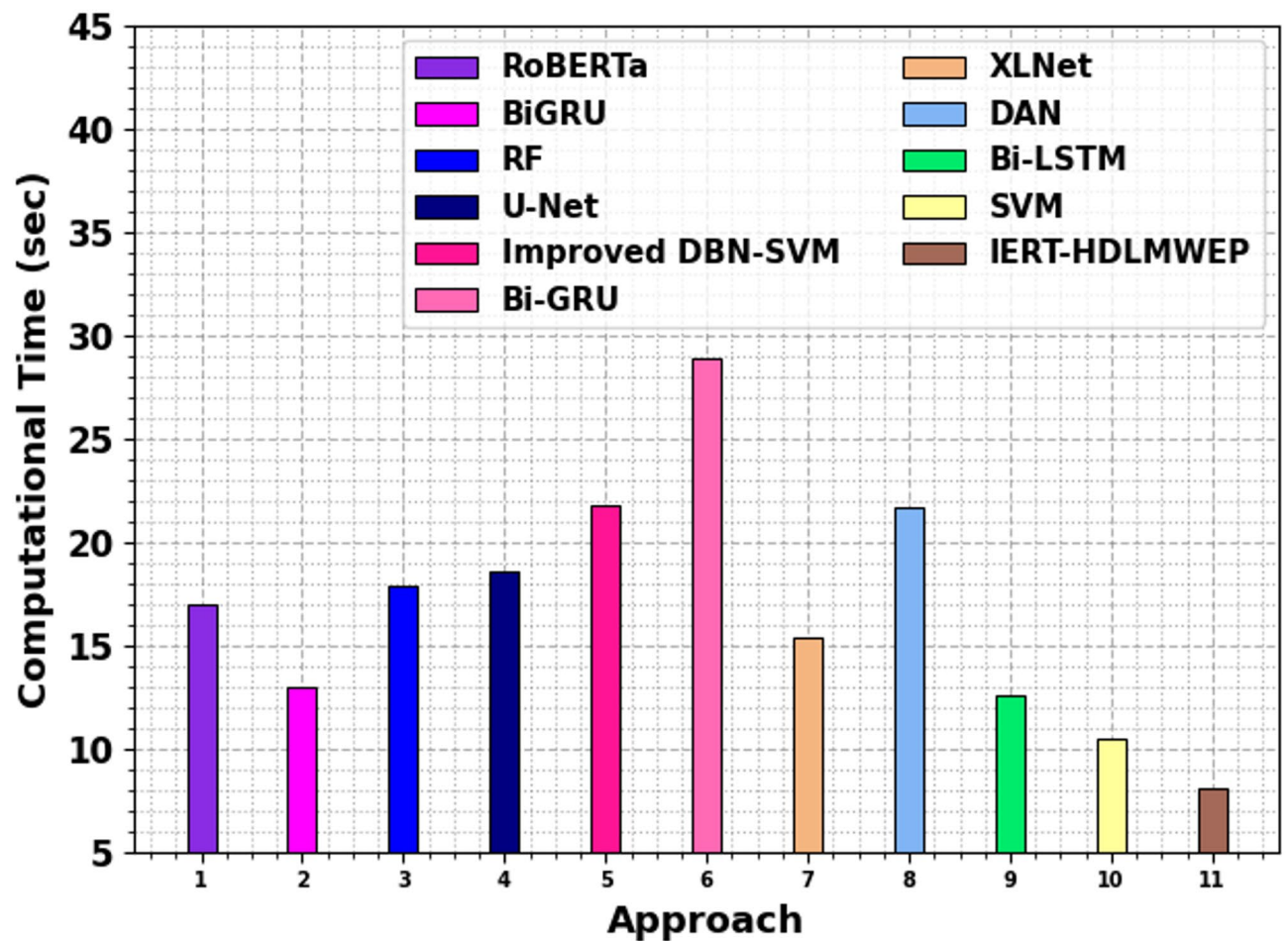


Fig. 12. CT outcome of IERT-HDLMWEP methodology with existing models.

Approach	$Accu_y$	$Prec_n$	$Recal$	$F_{Measure}$
RoBERTa	10.25	13.19	13.27	17.07
BiGRU	5.21	6.12	17.66	13.72
RF	1.87	6.97	17.86	17.84
U-Net	11.03	13.80	13.95	17.63
Improved DBN-SVM	5.83	6.89	18.37	14.47
Bi-GRU	2.61	7.51	18.46	18.39
XLNet	3.43	7.82	19.25	18.78
DAN	9.94	8.63	17.37	14.45
Bi-LSTM	8.36	10.15	17.39	16.10
SVM	1.97	8.93	13.58	15.67
IERT-HDLMWEP	0.33	6.39	11.46	10.03

Table 9. Error analysis of IERT-HDLMWEP technique with existing methods.

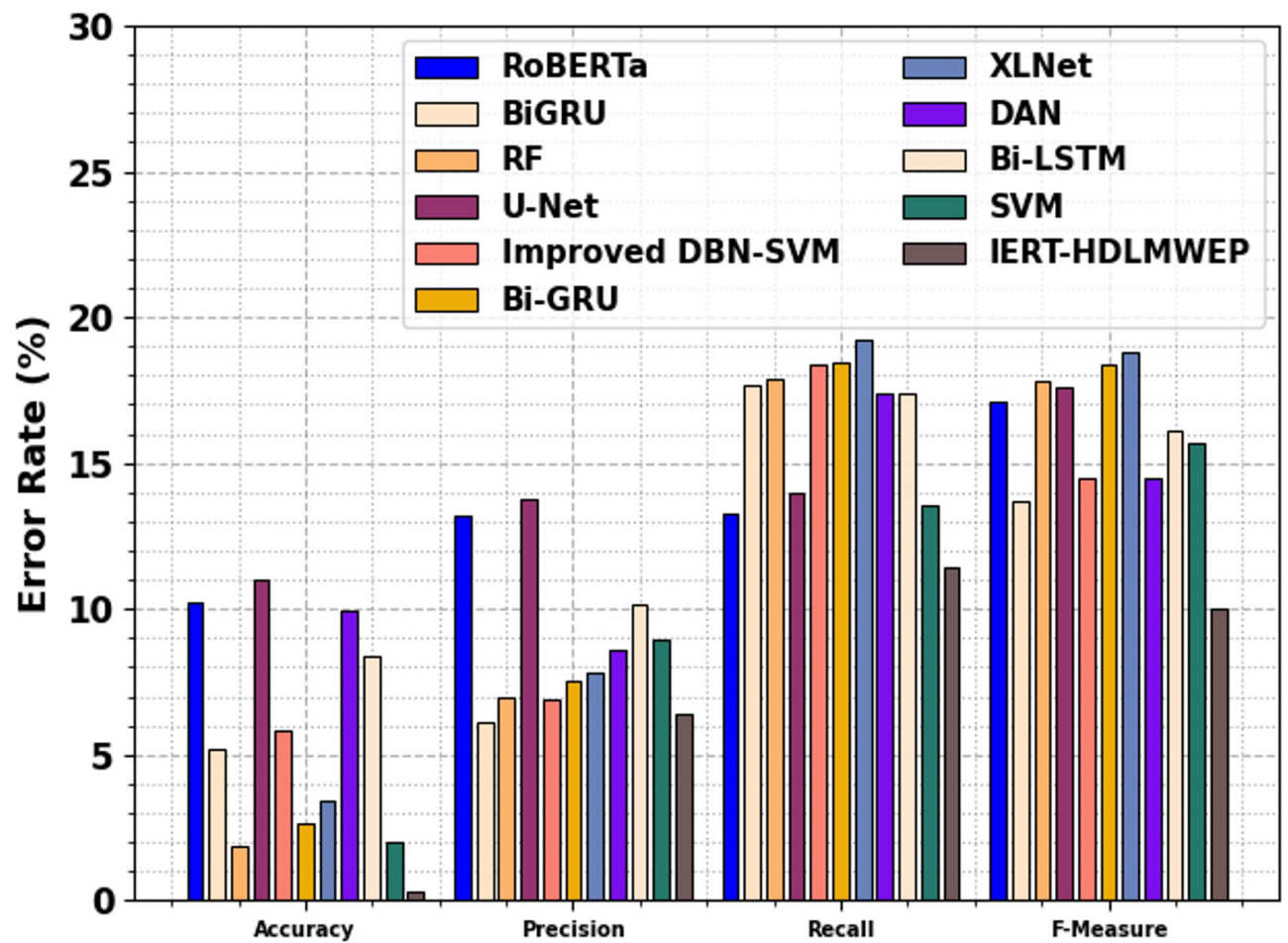


Fig. 13. Error analysis of IERT-HDLMWEP technique with existing methods.

Approach	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_i</i>	<i>F_{Measure}</i>
Word2Vec	97.06	90.92	85.81	87.06
TF-IDF-CDW	97.81	91.46	86.57	87.75
PSE	98.50	92.23	87.30	88.51
C-BiG-A	99.03	92.98	87.80	89.28
IERT-HDLMWEP	99.67	93.61	88.54	89.97

Table 10. Ablation study-based comparative analysis of the IERT-HDLMWEP methodology.

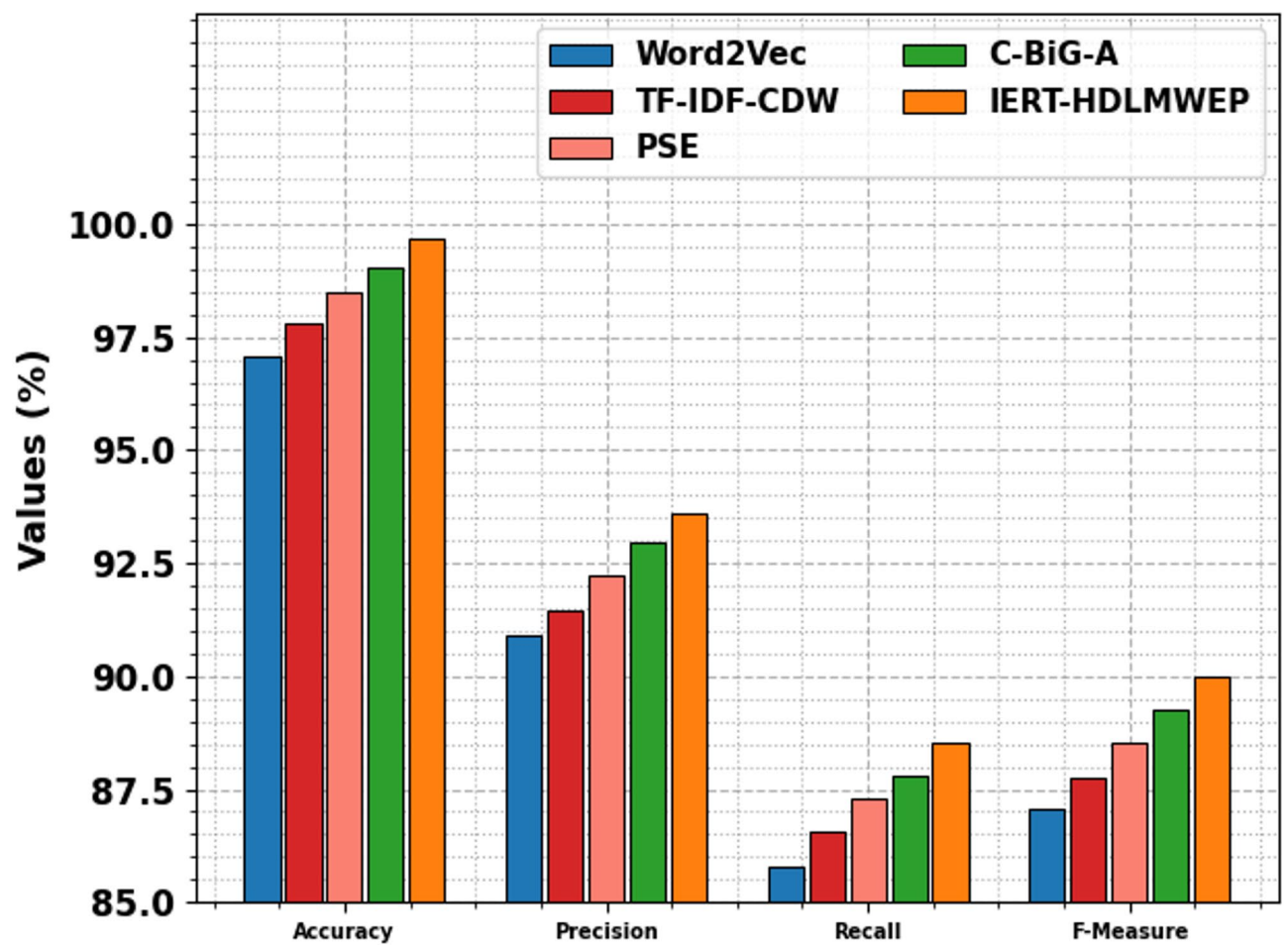


Fig. 14. Ablation study-based comparative analysis of the IERT-HDLMWEP methodology.

Data availability

The data that support the findings of this study are openly available in the Kaggle repository at [<https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text>] (<https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text>), reference number³⁴.

Received: 14 June 2025; Accepted: 29 September 2025

Published online: 04 November 2025

References

- Alshahrani, H. M., Yaseen, I. & Drar, S. Textual emotion analysis-based disabled people talking using improved metaheuristics with deep learning techniques for intelligent systems. *J. Disabil. Res.* **2** (3), 40–47 (2023).
- Bharti, S. K. et al. Text-based emotion recognition using deep learning approach. *Comput. Intell. Neurosci.* **1**, 2645381 (2022).
- Sivakumar, M. et al. Transforming Arabic Text Analysis: Integrating Applied Linguistics with m-Polar Neutrosophic Set Mood Change and Depression on Social Media. *Int. J. Neutrosophic Sci. (IJNS)* **25**, 2 (2025).
- Lyakso, E. et al. November. Recognition of the emotional state of children with down syndrome by video, audio and text modalities: human and automatic. In *International Conference on Speech and Computer* (pp. 438–450). Cham: Springer International Publishing. (2022).
- Mohammad, S. M. Ethics sheet for automatic emotion recognition and sentiment analysis. *Comput. Linguistics.* **48** (2), 239–278 (2022).
- Uddin, M. Z., Dysthe, K. K., Følstad, A. & Brandtzaeg, P. B. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Comput. Appl.* **34** (1), 721–744 (2022).
- Singh, P., Srivastava, R., Rana, K. P. S. & Kumar, V. A multimodal hierarchical approach to speech emotion recognition from audio and text. *Knowl.-Based Syst.*, 229:107316 (2021).
- Subramanian, M., Sathiskumar, V. E., Deepalakshmi, G., Cho, J. & Manikandan, G. A survey on hate speech detection and sentiment analysis using machine learning and deep learning models. *Alexandria Eng. J.* **80**, 110–121 (2023).
- Alrasheedy, M. N., Muniyandi, R. C. & Fauzi, F. October. Text-based emotion detection and applications: a literature review. In *2022 international conference on cyber resilience (ICCR)* (pp. 1–9). IEEE. (2022).
- Sreedhar, P. et al. Automated EEG based emotion detection using bonobo optimizer with deep learning on human computer interaction. *J. Intell. Syst. Internet Things* **12**, 1 (2024).
- Kumar, P., Malik, S., Raman, B. & Li, X. VISTANet: Visual Spoken Textual Additive Net for Interpretable Multimodal Emotion Recognition. *IEEE Trans. Affect. Comput.*
- Di Luzio, F., Rosato, A. & Panella, M. An explainable fast deep neural network for emotion recognition. *Biomed. Signal Process. Control.* **100**, 107177 (2025).
- Fu, F. et al. SDR-GNN: Spectral Domain Reconstruction Graph Neural Network for incomplete multimodal learning in conversational emotion recognition. *Knowl.-Based Syst.* 309:112825. (2025).
- Li, J., Wang, X., Liu, Y. & Zeng, Z. ERNetCL: A novel emotion recognition network in textual conversation based on curriculum learning strategy. *Knowl.-Based Syst.* 286:111434. (2024).
- Kusal, S., Patil, S., Choudrie, J., Kotecha, K. & Vora, D. Transfer learning for emotion detection in conversational text: a hybrid deep learning approach with pre-trained embeddings. *Int. J. Inform. Technol.* **8**, 1–18 (2024).
- Feng, L. et al. Multimodal speech emotion recognition based on multiscale MFCCs and multiview attention mechanism. *Multimedia Tools Appl.* **82** (19), 28917–28935 (2023).
- Zhang, L., Xia, B., Wang, Y., Zhang, W. & Han, Y. A fine-grained approach for EEG-based emotion recognition using clustering and hybrid deep neural networks. *Electronics* **12**(23), 4717 (2023).
- Omarov, B. & Zhumanov, Z. Bidirectional long-short-term memory with attention mechanism for emotion analysis in textual content. *Int. J. Adv. Comput. Sci. Appl.* **14**, 6 (2023).
- Hicham, N. & Nassera, H. Multilingual opinion mining of imbalanced customer review datasets using GPT and an ensemble hybrid deep learning model. *Data Sci. Manag.* (2025).
- Mahajan, R., More, A. S. & Shah, U. Navigating emotion in code-mixed languages: performance of ML and DL models on hindi-english text. *Procedia Comput. Sci.* **258**, 4029–4037 (2025).
- Zhu, X., Khan, M., Taleb-Ahmed, A. & Othmani, A. Advancing medical question answering with a knowledge embedding transformer. *PLoS One.* **20** (8), e0329606 (2025).
- Khan, M., Gueaieb, W., El Saddik, A., De Masi, G. & Karray, F. September. An efficient violence detection approach for smart cities surveillance system. In *2023 IEEE International Smart Cities Conference (ISC2)* (pp. 1–5). IEEE. (2023).
- Arumugam, L., Arumugam, S., Chidambaram, P. & Govindasamy, K. A multi-model deep learning approach for human emotion recognition. *Cogni. Neurodyn.* **19**(1), 123 (2025).
- Khan, M. et al. Joint Multi-Scale multimodal transformer for emotion using consumer devices. *IEEE Trans. Consum. Electron.* (2025).
- Alyoubi, A. A. & Alyoubi, B. A. Interpretable multimodal emotion recognition using optimized transformer model with SHAP-based transparency. *J. Supercomput.* **81**(9), 1044 (2025).
- Vani, G., Kalyan, K., Karthik, S., Reddy, R. N. V. S. & Haritha, P. B. and January. Text Fusion+: Advanced Integrated Image-to-Speech and Text Analysis Systems for Enhanced Accessibility and Interactive Learning. In *International Conference on Advanced Computing Techniques in Engineering & Technology* (pp. 174–184). Cham: Springer Nature Switzerland. (2025).
- Khan, M., Ishaq, M., Swain, M. & Kwon, S. Advanced sequence learning approaches for emotion recognition using speech signals. In *Intelligent Multimedia Signal Processing for Smart Ecosystems* 307–325 (Springer International Publishing, 2023).
- Ghous, G., Najam, S. & Jalal, A. Human Emotion Recognition from EEG-brain Signals using Enhanced Machine Learning Method. In *2025 6th International Conference on Advancements in Computational Sciences (ICACS)* (pp. 1–7). IEEE. (2025).
- Patil, S. P., Apare, R. S., Borhade, R. H. & Mahalle, P. N. July. Handwritten Text-Based Learning Disability Prediction in Children Using Deep Neural Networks. In *International Conference on Data Science and Big Data Analysis* (pp. 265–276). Singapore: Springer Nature Singapore. (2024).
- Mishra, A. R., Rai, A., Nandan, D., Kshirsagar, U. & Singh, M. K. Unveiling emotions: NLP-based mood classification and well-being tracking for enhanced mental health awareness. *Math. Modell. Eng. Probl.* **12**, 2 (2025).
- Al-Tameemi, I. K. S., Feizi-Derakhshi, M. R., Pashazadeh, S. & Asadpour, M. A comprehensive review of visual–textual sentiment analysis from social media networks. *J. Comput. Social Sci.* **7** (3), 2767–2838 (2024).
- Li, C., Xie, Z. & Wang, H. Short text classification based on enhanced word embedding and hybrid neural networks. *Appl. Sci.* **15**(9), 5102 (2025).
- Si, S., Mu, D. & Tang, H. A hybrid CBiGRUPE model for accurate grinding wheel wear prediction. *Sensors* **25**(9), 2935 (2025).
- <https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text>
- Jiang, X., Zhang, Y., Lin, G. & Yu, L. Music emotion recognition based on deep learning: A review. *IEEE Access* (2024).
- Peng, S. et al. A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Networks.* **8** (5), 745–762 (2022).

Acknowledgements

The authors extend their appreciation to the King Salman center For Disability Research for funding this work through Research Group no KSRG-2024- 350.

Author contributions

Mohammed Abdullah Al-Hagery: Conceptualization, methodology, validation, investigation, writing—original draft preparation, Hechmi Shili: Conceptualization, methodology, writing—original draft preparation, writing—review and editing Nasser Aljohani: methodology, validation, writing—original draft preparation Ishfaq Yaseen: software, visualization, validation, data curation, writing—review and editing.

Funding

The authors extend their appreciation to the King Salman Centre for Disability Research for funding this work through Research Group no. KSRG-2024-350.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.A.A.-H.

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