MCWA-LSTM with SELU for Text-Based Emotion Classification

Bhargavi Vemala*[ID], M. Humera Khanam[ID]

Department of Computer Science and Engineering, Sri Venkateswara University, Tirupati 517502, India

Corresponding Author Email: vemalabhargavi2700@gmail.com

ABSTRACT

Text-based emotion classification involves determining text to categorize emotions like sadness, anger, fear, happy, and so on. It employs Natural Processing Language (NLP) techniques for understanding sentiment and emotional tone behind words. This technique is widely employed in social media, customer feedback analysis, etc., However, accurately classifying emotions from text remains challenging because of sarcasm, ambiguity, and contextual nuances of human language leads to incorrect emotional responses. This research proposes Monotonic Chunk Wise Attention Long Short-term Memory with Scaled Exponential Linear Unit (MCWA-LSTM with SELU) for mulri-label text based emotion classification. In traditional LSTM, MCWA is incorporated to focus on relevant chunks of input sequentially which minimize noise from irrelevant parts and captures significant context effectively. LSTM capture long-term dependencies and contextual information which makes effective for emotion classification whereas SELU improves learning by managing self-normalizing properties that enhance training and model stability. Therefore, MCWA-LSTM with SELU achieves high accuracy of 98.66%, 98.32% on SemEval-2018 Task1-C, GoEmotion datasets for multi-class which is 37.46% and 27.12% higher compared to Universal Conceptual Cognitive Annotation-Graph Attention Network (UCCA-GAT). The proposed method obtains high f1-score of 98.21% for binary classification on TEL-NLP dataset which is 15.21% higher than existing Bidirectional Encoder Representations from Transformer (BERT) and Clipped Asymmetric Loss (ASL).

1. INTRODUCTION

Natural Language Processing (NLP) is a field of Machine Learning (ML) or Deep Learning (DL) aimed at enabling machines for interpreting and processing human language as it is spoken or written. Unlike programming languages with clear and formal syntax, natural language is highly flexible with word meanings and structures that vary significantly depending on context [1, 2]. With widespread employment of social media and growing availability of online news articles, huge amounts of text are generated for evaluation. Sentiment analysis, questions and answer systems, translation systems, text generation system, information retrieval, and others are prominent accomplishments in NLP field under active investigation [3]. Human emotion is considered as complicated physiological and psychological phenomenon due to arises from interaction of brain activity, hormonal responses, and personal experiences. Emotion classification is one of basic of human and computer emotional interaction that impacts development of Artificial Intelligence (AI) technology which is utilized to analyze text especially on social media to gain insights into people's feelings about happenings or events [4, 5]. This process is employed widely across numerous domains like marketing, psychology, and political science for analyzing people's attitudes. Researchers recognize areas of concern by categorizing text into distant emotional classes using AI [6, 7]. During analysis, generated information assist public health officials that establish targeted method and efficient communication scheme to solve disease outbreaks. It is imperative to closely analyze emotions in order to gain a deeper and more accurate understanding of people's sentiments due to emotional nuances often provide critical insights that go beyond surface-level expressions [8]. The emotions identified are sadness, fear, happiness, disgust, anger, and surprise which gradually enriched with phenomena like embarrassment, shame, pride, and excitement [9].

Emotion is a permanent personality trait and majorly evaluates dynamics of experiencing emotions, i.e., depth of experience, sensitivity, constancy of emotions, duration, and appropriateness of emotional reactions to scenarios. From a historical point of view, emotion is understood as reaction on scenarios and with assistance of factor analysis, emotion is determined as factor saturating 2/3 of primary factors acquired from questionnaires and from assessment of respondents behaviour [10]. Based on classes or dimensions, researchers in visual emotion analysis categorize emotions differently and accurately which involves descriptions like category distribution and dominant categories [11, 12]. However, semantic content (semantic and contextual information) in utterance data has always been regarded as significant sentiment information carrier [13, 14]. Unlike basic text classification concentrates on extracting broad features from text, emotion analysis needs an extra psychological dimension that is presently lacking in existing research [15]. In text-based

emotion detection, NLP methods are used for extracting patterns from text data by using ML to infer user's emotions [16]. However, ML struggled in capturing linguistic context and nuances often based heavily on handcrafted features which limits its ability in generalizing across different emotion expressions [17]. On the other hand, DL methods automatically learn hierarchical representations from raw text that makes better understanding of semantics and context [18-20]. Hence, DL achieves high accuracy and robustness in analysing subtle and varied emotional cues.

1.1 Problem statement

Text-based emotion classification is an NLP field that focuses on identifying and classifying emotions in text. However, accurately classifying emotions from text remains challenging because of ambiguity and contextual nuances of human language leads to incorrect emotional responses. Sarcastic statements convey opposite sentiments that literal meaning while ambiguity in varied sentence structure and word usage struggles emotional interpretation. Moreover, context-dependent needs deeper semantic understanding that existing models fail to capture effectively.

1.2 Objective

The main objective of this research is to develop an efficient DL method which accurately categories emotions from text by utilizing contextual and sequential information. This research proposes Monotonic Chunk Wise Attention Long Short-term Memory with Scaled Exponential Linear Unit (MCWA-LSTM with SELU) for concentrating on emotionally important text chunks while maintaining structure of temporal unit. Furthermore, SELU activation function enhance convergence and stability via self-normalization. By integrating these methods, a robust and context-aware framework is developed that enhance emotion classification tasks over TEL-NLP, SemEval-2018 Task1-C, and GoEmotion datasets.

1.3 Contributions

The main contribution of this research is as follows:

- **Semantic and contextual enrichment:** To extract the features, Word2vec and Term Frequency-Inverse Document Frequency (TF-IDF) are used which enhance feature representation through capturing semantic relationships among words. This makes model for better understanding context and distinguish emotional expressions.
- **Enhanced emotion focusses via MCWA:** A integration of MCWA with LSTM makes the model to selectively process emotionally significant parts in sequential manner that helps to capture context-dependent emotional cues and decrease distraction from irrelevant text.
- Improved training stability and convergence: SELU activation ensures self-normalization in network that stabilize learning through managing consistent activation across layers increase convergence and model's ability in learning complex emotional patterns.

2. LITERATURE SURVEY

Marreddy et al. [21] introduced a supervised graph reconstruction approach, Multi-Task Text Graph Convolutional Network (MT-Text GCN) for Telugu language in single and multi-task like Emotion Identification (EI), Sentiment Analysis (SA), Sarcasm Detection (SAR), and Hate-Speech (HS). MT-Text GCN learned sentence graph embeddings and low-dimensional word from word-sentence graph reconstruction by employing Graph AutoEncoder (GAE) and utilize multi-task text classification utilizing latent sentence graph embeddings. The word embeddings were generated by employing random walk-based methods for Telugu language. However, MT-Text GCN struggled with inaccurate performance due to rich morphology and complex script which affect effective graph construction and token-level representation.

Ramakrishnan and Babu [22] developed a Bidirectional Encoder Representations from Transformer (BERT) and Clipped Asymmetric Loss (ASL) to classify multi-label emotion through prioritizing minority classes while decreasing dominance of frequent classes. A loss function prioritizes minority classes which enable better performance for rare emotions without compromising accuracy. By combining ASL with BERT, contextual understanding and transformer-based language model were leveraged to improve classification performance. This integration enhances F1-score significantly across various emotion accurately. Nevertheless, BERT face difficulty in capturing subtle emotion overlaps in multi-label settings because of fixed context windows and limited label dependencies.

Ameer et al. [23] suggested a semantic and syntactic aware graph attention network to categorize emotions from text with numerous labels. A semantic information was integrated in syntactic information in form of dependency trees and Graph Attention Network (GAT) in form of Universal Conceptual Cognitive Annotation (UCCA). Then, the feature and adjacency matrices from semantic and syntactic representations were extracted. At last, the matrices were used in graph attention to classify multi-label. However, semantic and syntactic aware graph attention struggled with scalability while managing texts with numerous emotion labels because of increased graph complexity.

Ameer et al. [24] established a Long Short-Term Memory (LSTM) and fine-tuning of transformer network via transfer learning with single attention network and RoBERTa-Multi Attention (RoBERTa-MA) to classify multi-label emotions. During pre-processing, eight different normalization phases namely Uniform Resource Locator (URL), username, Email, currency, number, phone, data, and time normalization were established on Tweets for better understanding. Then, multiple attention mechanism was added to transformer model output and fine-tuned the model on multi-label emotion classification. Nevertheless, LSTM processes sequences linearly which makes less effective at modelling parallel or overlapping emotional cues present across different parts of sentence.

Ni and Ni [25] presented an Emotion Correlation-enhanced Sentiment Analysis Model (ECO-SAM) to classify multi-label sentiment analysis. ECO-SAM employed a pre-trained BERT to acquire semantic embeddings of input texts and later self-attention method was used to model semantic correlation among emotions. Also, ECO-SAM employed a text emotion matching neural network to enable sentiment analysis for input texts. However, ECO-SAM overemphasize predefined

emotion correlations that leads to biased performance while emotions co-occur in unexpected patterns.

Gamage et al. [26] developed an AI model for robust, adaptable, and explainable detection of multi-granular assembles of emotion. The developed method employed lexicon generation and finetuned Large Language Model (LLM) for formulating multi-granular assembles of two, eight, and fourteen emotions. This model was robust to ambiguous emotion expressions which were implied in conversation, assembles were explainable, and adoptable to domain-specific semantic emotion utilizing constituent terms and intensity. Nevertheless, the developed model lack robustness to noisy or ambiguous inputs that results in misclassification at fine-grained levels.

From the overall analysis, existing methods had limitations like difficulty in capturing subtle emotion overlaps, overemphasize predefined emotion correlations, lack robustness to noisy or ambiguous inputs, sarcasm, ambiguity, and contextual nuances of human language which leads to incorrect emotional responses. To solve this issue, this research proposes MCWA-LSTM with SELU to classify text-based emotions accurately by focusing on emotionally appropriate chunks of text. LSTM captures long-term dependencies and nuanced emotional patterns that improves model's ability in identifying subtle emotion overlaps. Moreover, SELU ensures self-normalization that stabilize training and increase convergence. Together, these process generate robust and context-aware model by handling sarcasm and ambiguity more accurately.

3. PROPOSED METHODOLOGY

In this research, MCWA-LSTM with SELU is proposed to classify the text-based emotions accurately. Initially, TEL-NLP, SemEval-2018 Task1-C, and GoEmotion are used to determine model performance. During pre-processing, tokenization, stopword removal, and stemming are employed to eliminate the irrelevant words that enhance contextual understanding. Then, Word2vec and TF-IDF are performed to extract the features effectively while proposed MCWA-LSTM with SELU is used for emotion classification. Figure 1 depicts workflow of entire process.
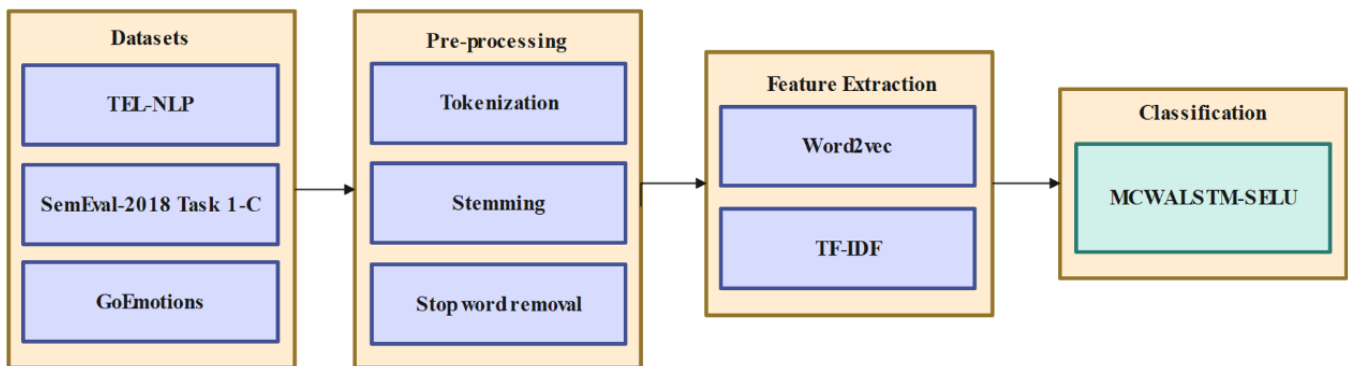

Figure 1. Workflow for text classification using feature extraction and classification

3.1 Datasets

This research evaluates model performance using TEL-NLP [21], SemEval-2018 Task1-C, and GoEmotions datasets. These datasets support both binary and multi-class emotion classification tasks. A detailed description about these datasets are explained below.

TEL-NLP: An annotated Telugu dataset called TEL-NLP is created for four NLP tasks namely SA, EI, HS, and SAR. Each annotator accurately annotates at least 80% sentences from sample dataset to evaluate and complete annotation process. After 2 rounds of sanity checks, overall of 16,234 labeled sentences are obtained for SA, EI, HS, and SAR. The SA tasks use "Negative" and "Positive" labels, EI tasks include "Angry", "fear", "happy", and "sad" labels, while SAR and HS tasks are labeled as "yes" or "no" respectively.

SemEval-2018 Task1-C: It involves 10,983 tweets in English subset with 11 different emotions called disgust, anger, anticipation, joy, fear, love, pessimism, optimism, trust, surprise, and sadness. This dataset act as significant benchmark for testing and developing emotion recognition models especially in a context of multi-label emotion classification.

GoEmotions: It is a large multi-label dataset with 58,000 comments which are manually labeled with 27 emotions and neutral class. Emotion categories are amusement, admiration, annoyance, anger, caring, curiosity, approval, confusion, disappointment, desire, disgust, disapproval, excitement, gratitude, embarrassment, fear, love, grief, nervousness, joy, pride, optimism, relief, realization, sadness, remorse, and surprise. The obtained data are passed through pre-processing stage for further process.

3.2 Pre-processing

After obtaining text, tokenization, stopword removal, and stemming [27] are used to convert text into structured and analyzable format. These methods unify word forms and focus on emotion-relevant terms enhance contextual understanding and efficiency of emotion classification models. A description for these methods is explained in detail to highlight individual roles in preparing text data.

Tokenization: It breaks down text into meaningful units (tokens) like phrases and words that enable model to understanding word boundaries and process input uniformly. The data is split into individual tokens by utilizing specific delimiters like spaces, commas, hash symbols (#), and at symbol (@) respectively.

Stemming: This method reduces inflected or derived words in root form that assist in generalizing words. Minimizing vocabulary size minimize model complexity and enhance generalization by grouping similar emotional expressions.

Stopword removal: It removes frequently employed words like "is", "the", "and" assist to minimize noise and dimensionality in text data allows models for concentrating on more meaningful words. This enhances processing efficiency and assist the classifier to focus on emotion-rich content like love, angry, and so on. These pre-processed texts are the transformed into numerical features using TF-IDF and Word2vec methods.

3.3 Extracting features using word embeddings

The pre-processed text is fed as input into feature extraction process using TF-IDF and Word2vec which preserves semantic relationships between words for deeper contextual understanding. A detailed process for these methods is discussed below for extracting features effectively.

Word2Vec: This method captures semantic relationships and contextual similarity between words by representing dense vectors in continuous space assist deeper emotional understanding. Word2vec [28] effectively encodes word similarities which enhance feature quality for emotion classification. Word2vec contains two techniques namely skip-gram and Continuous Bag of Words (CBOW). In text-based emotion classification, skip-gram is employed to determine surrounding words given a target word assist to capture emotional context of words in sentence. Meanwhile, CBOW determine target word depending on surrounding context that efficiently encode overall emotion. Hence, this method makes better understanding of word relationships and enhance feature representation for emotion classification.

TF-IDF: It effectively highlights significant words in text by assigning terms depending on frequency in document relative to frequency across all documents that assist in identifying emotionally significant words. This method decreases the impact of common, fewer informative words, and enhance distinction among various emotional expressions. TF measures frequency of particular word appearing in document while IDF computes logarithmic inverse probability of documents. The mathematical formula for TF-IDF [29] is expressed in Eqs. (1) to (3).

$$W_{mn} = tf_{mn} \times idf_n \tag{1}$$

$$idf_n = log \frac{D}{df_n} \tag{2}$$

$$W_{mn} = tf_{mn} \times log \frac{D}{df_n} \tag{3}$$

where, W_{mn} refers to m document for n word, tf_{mn} denotes occurrence of word count in specific document, idf_n determines inverse document frequency, D represents total number of documents, and df indicates number of documents that involve specific word t_n. Thus, TF-IDF captures significance of words by weighting frequency based on entire corpus which highlights emotionally significant terms. Word2vec captures semantic relationships by encoding words into dense vectors makes model for understanding context and emotional nuances. Integrating both enhance emotion classification by leveraging accurate keyword importance and rich contextual meaning. Then, extracted features are passed through classification process to categorize emotions.

3.4 Classification

The extracted features are fed as input to classification process using MCWA-LSTM with SELU to classify text-based emotions. LSTM [30] is used for emotion classification due to has the ability in modeling sequential data and capture long-range dependencies in text. Emotions in language depend not just on individual words but on context build over a whole sentence or paragraph. LSTM is especially designed for retaining significant information over earlier in sequence which makes better understanding the nuanced and context dependent emotion expressions. In traditional LSTM, MCWA is incorporated to enhance model's ability in processing emotional content more effectively. MCWA enable LSTM to focus selectively on emotionally salient chunks without being distracted by fewer relevant parts of sequence. Moreover, SELU automatically normalizes output of each neuron to have zero mean and unit variance during training. This process provides stable activation across layers which minimize vanishing or exploding gradient issue and generates rapid convergence. As shown in Figure 2, Convolutional Neural Network (CNN) based character embedding is utilized to capture subword and morphological features. LSTM consider these inputs to determine relations between words based on contextual resemblance. Then, MCWA has been used over LSTM's output on the assumption that phrases are created via contextual words. The acquired output from each attention is fed into LSTM before computing final label distribution by SoftMax. The proposed MCWA-LSTM with SELU is described in detail below for emotion classification.

3.4.1 CNN-based character embedding

CNN for character embedding acquires the information of local context by considering filter with different window sizes from 2 to 5 are utilized for learning contextual character representations. Smaller filters like 2 and 3 effectively capture short-term dependencies which are significant for identifying suffixes, prefixes, and subword patterns. Larger filters like size 4 and 5 enable model for learning longer character sequences and morphological sequences which are crucial to understand semantic variations. By using multiple window sizes, CNN extracts rich hierarchical features which enhance robustness of contextual character representations. A maximum pooling process is employed over convoluted features and then concatenation of all filter output is represented as contextual word representation. The generated features are fed into three stacked Fully Connected (FC) layers. The resultant vectors from penultimate layer contains character level information for i^{th} word. Later, to produce a final representation of word vector, these character vectors are stacked with word embedding vector. An intra-word dependency is computed through temporal sequence method using LSTM by converting vectors into hidden states.

3.4.2 Monotonic Chunk Wise Attention Long Short-term Memory with Scaled Exponential Linear Unit (MCWA-LSTM with SELU)

Unlike traditional attention mechanisms consider entire input sequence at once, MCWA operates in monotonic that concentrates on chunks of text. Moreover, monotonic nature of MCWA enforces a left-to-right progression via input which aligns effectively with natural language of human language and assist in capturing temporal dynamics of emotional cues. Overall, combining LSTM with MCWA leads to model that is not only more efficient but also better at isolating and interpreting emotional content embedded in text sequences. The term monotonicity determines computation of attention

depending on hidden states on timesteps. These timesteps are subset of entire state sequence produced based on input window length called chunks. A hidden state is indicated by s_{j-1} to compile un-normalized energy function e_{ji} which is formulated in Eq. (4)

$$e_{ji} = MonotonicEnergy(s_{j-1}, h_i) \qquad (4)$$

The probability selection p_{ji} is calculated by employing logistic sigmoid function using Eq. (5)

$$p_{ji} = \sigma(e_{ji} + \epsilon); \epsilon \; \epsilon \; N(0,1) \qquad (5)$$

where, ϵ represents unit variance Gaussian distribution with mean 0. A state selection for window is based on Bernoulli that is utilized for content vector generation. The probability of monotonic attention is computed using selection probability using Eq. (6). This ChunkEnergy is utilized to determine attention score by employing Eqs. (7) and (8).

$$\alpha_j, := p_j, : . \; cumprod(1 - p_j, :). \; cumsum(\frac{\alpha_{j-1,:}}{cumprod(1-p_j,:)}) \qquad (6)$$

$$u_{j,i} = ChunkEnergy \; (s_{j-1}, h_i) \qquad (7)$$

$$ChunkEnergy \; (s_{j-1}, h_i) = V^T . tanh(W. h_i + W. s_{j-1} + b) \qquad (8)$$

where, V, b, and W denotes learning parameters. Softmax function normalizes ChunkEnergy across chosen chunk of W that is computed by MovingSum. A probability distribution of MCWA $\beta_j,:$ is determined using Eqs. (9) and (10).

$$\beta_j, := exp(u_j,:). \; MovingSum(\frac{\alpha_j,:}{MovingSum(exp(u_j,:),w,1)}),1,w \qquad (9)$$

$$MovingSum(x,b,f)_n = \sum_{m=n-b+1}^{n+f-1} x_m \qquad (10)$$

At last, expected value of context vector for j_{th} timestamp is determined via normalizing to each hidden state of input sentence vector using Eq. (11)

$$c_j = \sum_{i=1}^{n} \beta_{j,i}. h_i \qquad (11)$$

This context vector is integrated with previous hidden state of prior timestamp passed into LSTM. Compared to RNN and GRU, LSTM is preferred due to captures long-term dependencies in sequential data by efficiently solving vanishing gradient issue. While RNN faces challenges to remember information over long sequences, GRU provide a simpler gating mechanism, and LSTM employs more sophisticated memory cell and gating structure that makes to retain information more accurately. Therefore, this makes LSTM particularly useful for emotion classification where understanding over longer text spans is crucial. Overall, LSTM provide enhanced learning stability and accuracy for complex sequence modeling. LSTM involves three basic parts, input layer, hidden layer, and output layer. A number of neurons in input layer is determined by number of features for given sample of data whereas output layer is based on number of target features which enable LSTM a better candidate for

managing multivariate forecasting. Hidden layers contain three types namely forget f_t, output O_t, input i_t gates, and t represents timestep. Respective gates are utilized by memory gates s_t which allows the information to be stored in network. Information in cell state s_{t-1} is evaluated for deletion whereas inputs are generated through sigmoid activation function in [0,1] range. The mathematical formula for forget gate f_t is formulated in Eq. (12). Candidate values $\tilde{s_t}$ and activation function i_t for each input gate is expressed in Eqs. (13) and (14). New cell state s_t is formulated in Eq. (15). An output of h_i is evaluated by utilizing sigmoid and SELU activation functions in Eqs. (16) and (17).

$$f_t = \sigma(W_{f,x} x_t + W_{f,h} h_{t-1} + b_f) \qquad (12)$$

$$\tilde{s} = SELU(W_{\tilde{s},x} x_t + W_{\tilde{s},h} h_{t-1} + b_{\tilde{s}}) \qquad (13)$$

$$i_t = SELU(W_{i,x} x_t + W_{i,h} h_{t-1} + b_i) \qquad (14)$$

$$s_t = f_t \odot s_{t-1} + i_t \tilde{s_t} \qquad (15)$$

$$i_t = \sigma(W_{o,x} x_t + W_{o,h} h_{t-1} + b_o) \qquad (16)$$

$$h_t = o_t \odot SELU(s_t) \qquad (17)$$

where, σ represents sigmoid activation function, $W_{f,x}, W_{f,h}, W_{\tilde{s},x}, W_{\tilde{s},h}, W_{i,x}, W_{i,h}, W_{o,x}, W_{o,h}$ indicates weight matrices in forget, input, output gate, output of prior timestep $t - 1$ is denoted as $h + t\_1$, $b_f, b_{\tilde{s}}, b_i, b_o$ illustrates bias vector in forget, input, output gate, $\odot$ denotes Hadamard product. SELU enable self-normalizing neural networks by automatically maintaining mean and variance during training that assist to prevent vanishing gradients, leads to rapid convergence, and enhanced training stability. Like ELU, SELU enable negative values in its output that assist network to learn complex patterns by maintaining zero-centered activation which is formulated in Eq. (18).

$$SELU = 1.05 \times (max(0, x) + min(0,1.67 \times (exp(x) - 1))) \qquad (18)$$
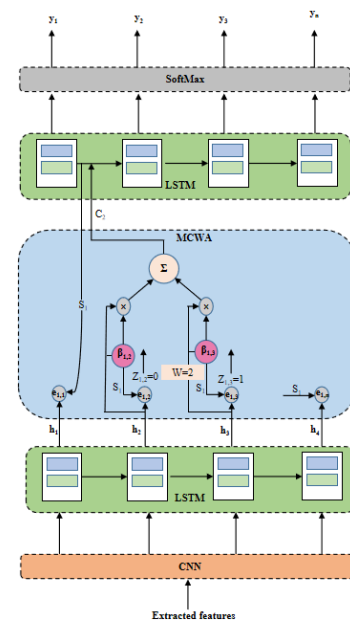

Figure 2. Architecture of MCWA-LSTM for classification

The hyperparameter of proposed method are considered as 0.01 learning rate for stable convergence, 32 batch size for efficient training, 0.5 dropout to minimize overfitting, SELU activation function for self-normalizing property, and Adam optimizer for adaptive learning rate. The integration of LSTM captures long-term dependencies in emotional context whereas MCWA enable monotonic attention flow which is appropriate for sequential data. Moreover, SELU generates self-normalization that enhance convergence speed and training stability. The respective architecture illustrated in Figure 2. Together, this method ensures accurate and robust emotion classification from intricate and nuanced textual inputs.

4. EXPERIMENTAL RESULTS

This research is simulated using NVIDIA GeForce RTX 3090 GPU (Graphics Processing Unit) CUDA version 12.4. Accuracy, F1-score, recall, and precision are employed to evaluate model's performance using Eqs. (19) to (22).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (19)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \qquad (20)$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \times 100 \qquad (21)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \qquad (22)$$

where, TP demonstrates True Positive, FP illustrates False Positive, TN denotes True Negative, and FN determines False Negative.

4.1 Performance analysis

Tables 1 to 3 show a performance evaluation of different classifiers on TEL-NLP, SemEval-2018 Task1-C, and GoEmotion datasets. While compared to existing methods like Deep Belief Network (DBN), BERT, RNN, Transformer with Extra Hop attention (Transformer-XH), and Text-to-Text-Transfer Transformer (T5), LSTM obtains high accuracy of $98.66\% \pm 0.04$, $98.32 \pm 0.06\%$ on SemEval-2018 Task1-C, GoEmotion datasets for multi-class and f1-score of $98.21 \pm 0.02\%$ for binary classification on TEL-NLP datasets. Due to LSTM has the ability in capturing long-term dependencies and contextual information, this method obtains high accuracy in sequential data. LSTM retain appropriate past information while filtering out irrelevant input which is crucial for understanding the emotional tone in text. Unlike traditional RNN, LSTM solves vanishing gradient issue that makes to learn deeper patterns and gating mechanism ensures precise control over information to remember or forget. Moreover, Transformer-XH requires heavy computation and large data whereas T5 is prone to overfitting. Therefore, this makes LSTM well-appropriate for modeling intricate emotional expressions across text.

Table 1. Analysis of different classifiers on SemEval-2018 Task1-C and GoEmotion

Data Sets	Methods	Accuracy (%)	Recall (%)	Precision (%)	Macro -F1 (%)	Micro-F1 (%)
	DBN	79.54	80.15	78.15	87.18	80.69
	BERT	82.36	79.47	82.69	78.69	82.48
	RNN	86.19	90.36	84.19	83.19	89.10
SemEval-2018 Task1-C	Transformer-XH	90.85	89.25	91.36	89.27	92.34
	T5	91.99	88.0	90.45	90.10	91.04
	LSTM	**98.66**	**96.54**	**96.32**	**95.53**	**96.24**
	DBN	85.16	86.7	84.69	86.37	90.15
	BERT	89.47	89.47	87.30	81.78	89.47
	RNN	91.05	91.26	92.48	84.39	94.79
GoEmotion	Transformer-XH	93.59	92.36	92.48	90.84	89.63
	T5	92.04	92.78	93.69	91.48	95.09
	LSTM	**98.32**	**98.12**	**97.04**	**98.23**	**98.05**

Table 2. Performance evaluation of different classifiers on TEL-NLP dataset

Methods	Accuracy (%)			
	SA	E1	HS	SAR
DBN	89.36	91.54	89.07	84.17
BERT	87.49	87.26	91.68	89.30
RNN	90.65	89.30	87.36	87.36
Transformer-XH	91.25	89.36	89.24	92.64
T5	92.08	90.24	91.06	94.35
LSTM	**97.95**	**96.15**	**97.10**	**97.56**

Table 3. F1-score metric analysis on different classifiers

Methods	F1-score (%)			
	SA	E1	HS	SAR
DBN	90.25	91.68	87.65	84.94
BERT	87.06	91.06	89.36	83.19
RNN	92.78	92.78	92.03	90.46
Transformer-XH	91.36	92.15	87.12	93.54
T5	89.54	93.68	90.15	92.34
LSTM	**97.54**	**96.32**	**98.21**	**97.06**

Table 4. Performance evaluation of different attention mechanisms on TEL-NLP dataset

Methods	Accuracy (%)				F1-score (%)			
	SA	E1	HS	SAR	SA	E1	HS	SAR
LAM [31]	87.36	89.24	87.15	85.69	82.49	82.69	89.45	89.45
GAM [32]	89.15	92.15	89.67	83.96	86.34	84.16	92.16	93.68
MHSA [33]	93.48	91.48	92.08	90.15	92.16	89.17	92.08	95.16
MCWA	**97.95**	**96.15**	**97.10**	**97.56**	**97.54**	**96.32**	**98.21**	**97.06**

Table 5. Performance evaluation of different activation functions on TEL-NLP dataset

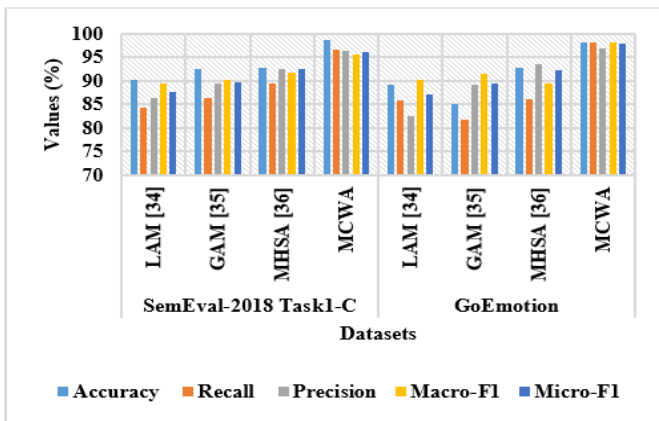| Methods	Accuracy (%)				F1-score (%)			
	SA	E1	HS	SAR	SA	E1	HS	SAR
ReLU	86.18	84.29	89.67	92.46	83.61	84.06	92.39	87.16
LReLU	84.96	87.63	92.46	89.48	86.49	84.29	93.67	89.37
ELU	85.05	89.71	92.39	87.69	89.48	90.10	90.16	92.16
SELU	**97.95**	**96.15**	**97.10**	**97.56**	**97.54**	**96.32**	**98.21**	**97.06**

Figure 3. Graphical representation of different attention mechanisms on SemEval-2018 Task1-C and GoEmotion

Figure 3 and Table 4 represent performance analysis of ablation study for different attention mechanisms. The proposed MCWA achieves a high accuracy of 98.66%, 98.32% on SemEval-2018 Task1-C, GoEmotion datasets for multi-class classification and f1-score of 98.21% for binary classification on TEL-NLP datasets by effectively focusing on meaningful chunks of input text. It balances local and global context by attending to fixed-size elements that assist the model in capturing subtle emotional cues in those chunks. This structured attention minimizes distraction from inappropriate words and increases the understanding of emotion. By processing text in organized chunks, MCWA enhances context alignment and obtains high classification performance compared to existing methods like Local Attention Mechanism (LAM) [31, 32], Global Attention Mechanism (GAM) [33], and Multi-Head Attention Mechanism (MHAM) [34] respectively.

Figure 4 and Table 5 determine performance analysis of different activation functions. Compared to existing activations like Rectified Linear Unit (ReLU), Leaky ReLU (LReLU), and ELU, SELU obtains high accuracy of 98.66%, 98.32% on SemEval-2018 Task1-C, GoEmotion datasets for multi-class classification and f1-score of 98.21% for binary classification on TEL-NLP. SELU outperforms ReLU and LReLU especially when managing negative sentiments due to SELU introduce self-normalizing properties that maintain activations around zero mean and unit variance which ensures stable learning over layers. Unlike ReLU eliminates negative value, LReLU which only performs retaining but SELU

conserves significant negative activations which are crucial to model subtle emotional like anger, sadness, and etc. It exponential component captures intricate and non-linear emotional patterns effectively whereas self-normalizing enhances convergence and minimize vanishing gradient. This makes model for better capturing ambiguity of language which results more robust and obtain better performance.
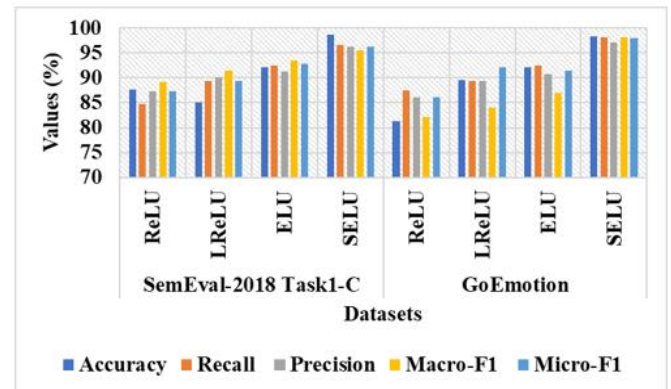

Figure 4. Graphical representation of different activation functions on SemEval-2018 Task1-C and GoEmotion

Table 6 represents performance analysis of time complexity and memory consumption. The proposed MCWA-LSTM with SELU achieves a less time complexity of 197s, 145s, and 166s. The proposed method consumed 326MB, 308MB, and 312MB of Graphic Processing Unit (GPU) memory during training on TEL-NLP, SemEval-2018 Task1-C and GoEmotion datasets due to effective attention mechanism and self-normalizing activations. MCWA processes input in fixed chunks that minimize the requirement of computing attention over whole sequence which consumes time and computational resources. Moreover, LSTM has sequential learning efficiency while SELU removes need for batch normalization that minimize overhead. Hence, this integration streamlines inference and training that makes rapid convergence with less parameters.

Table 7 demonstrates performance evaluation of k-fold validation with respect to accuracy on three datasets. The k=5 in k-Nearest Neighbors (k-NN) obtains high accuracy due to provides balanced view of local neighborhood that minimize sensitivity to noise and avoids over smoothing. The k=2 and 3 is too reactive to outliers that results in misclassification whereas k=7 dilute influence of appropriate neighbors. With k=5, model captures sufficient variation in data that maintains

generalization. Moreover, model efficiently averages local patterns without biasing toward minority or noisy patterns. Therefore, this balance leads to enhanced classification accuracy.

Figures 5 and 6 show a performance analysis of Receiver Operating Characteristics (ROC) curve for three datasets.

ROC curve is a graphical representation of classifier's performance by indicating True Positive Rate (TPR) against False Positive Rate (FPR). A model achieves high accuracy that rises sharply toward top-left corner which represents a high TPR and low FPR. A high ROC curve close to 0.96 reflects better classification ability and high robustness.

**Table 6.** Performance analysis of time complexity and memory consumption

| Methods	Datasets	Time Complexity (s)	Memory Consumption (MB)	t-test
MCWA-DBN with SELU		269	389	0.019
MCWA-BERT with SELU		258	346	0.014
MCWA-RNN with SELU		269	367	0.013
MCWA-Transforer-XH	TEL-NLP	256	359	0.011
T5		269	345	0.009
MCWA-LSTM with SELU		**197**	**326**	**0.005**
MCWA-DBN with SELU		198	346	0.015
MCWA-BERT with SELU		196	320	0.009
MCWA-RNN with SELU		150	369	0.007
MCWA-Transforer-XH	SemEval-2018 Task1-C	154	378	0.009
T5		159	359	0.007
MCWA-LSTM with SELU		**145**	**308**	**0.004**
MCWA-DBN with SELU		187	354	0.006
MCWA-BERT with SELU		179	397	0.010
MCWA-RNN with SELU		172	345	0.008
MCWA-Transforer-XH	GoEmotion	172	378	0.012
T5		175	369	0.008
MCWA-LSTM with SELU		**166**	**312**	**0.003**
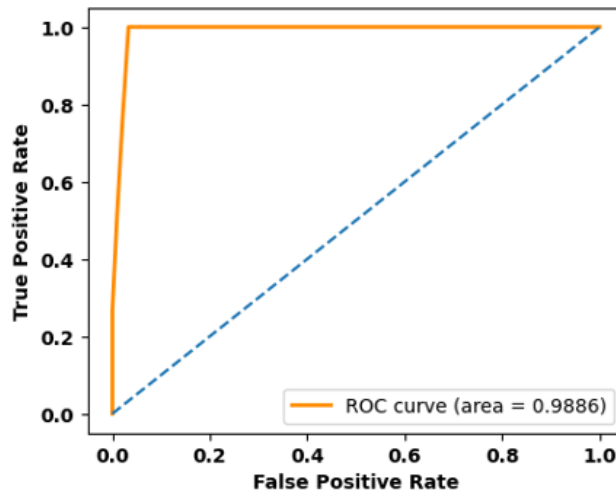
a) SA
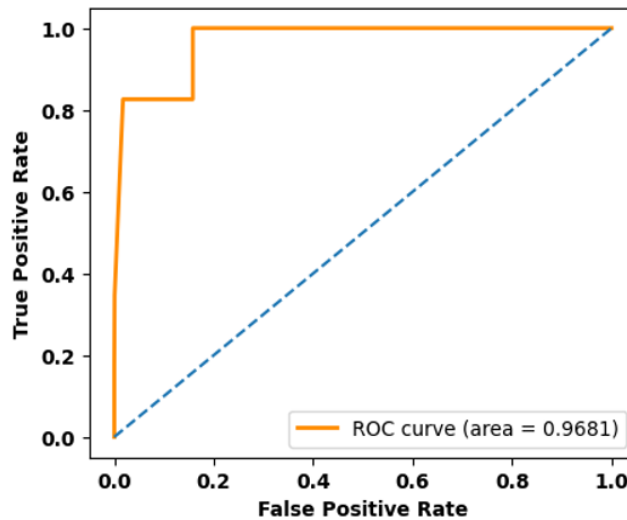
b) EI
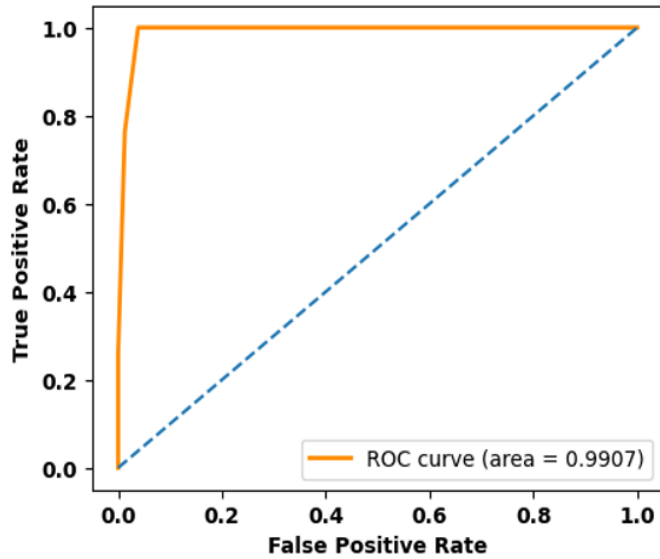
c) HS
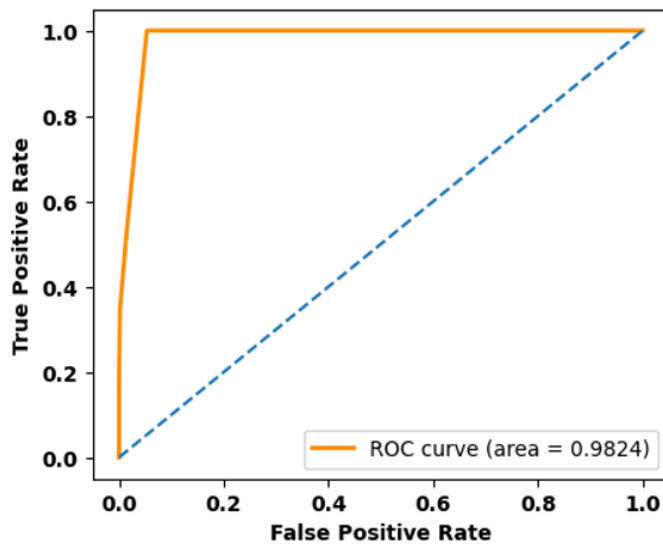
d) SAR

Figure 5. Performance analysis of ROC curve on TEL-NLP dataset: a) SA; b) EI; c) HS; d) SAR
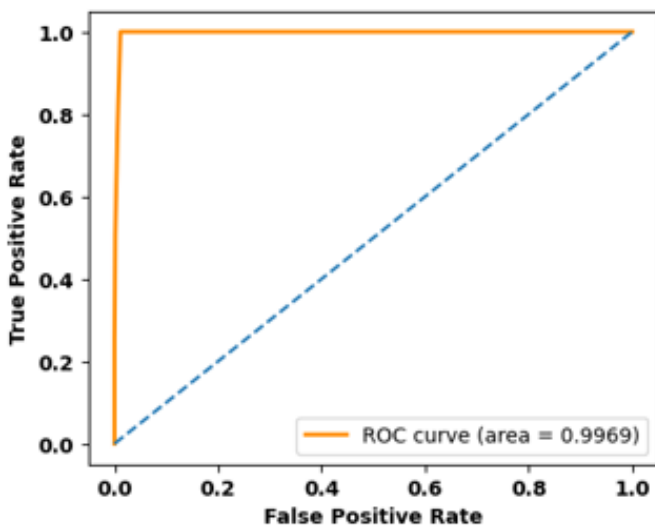
SemEval-2018 Task1-C
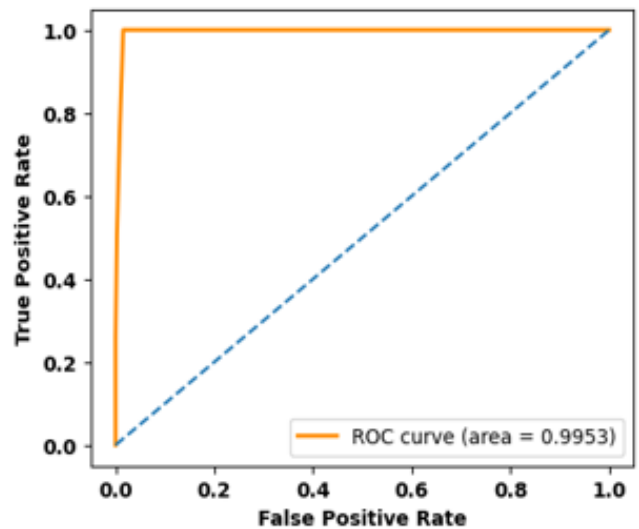
GoEmotion

Figure 6. Performance analysis of ROC curve on SemEval-2018 Task1-C and GoEmotion dataset

Table 7. Performance evaluation of k-fold validation based on accuracy on different datasets

| k-fold | TEL-NLP | | | | SemEval-2018 Task1-C | Go Emotion |
	SA	E1	HS	SAR		
k=2	81.06	85.16	83.10	86.12	89.48	87.94
k=3	85.96	89.40	86.17	90.14	92.49	89.67
k=5	**97.95**	**96.15**	**97.10**	**97.56**	**98.66**	**98.32**
k=7	92.69	92.47	89.32	95.36	92.34	92.49

4.2 Performance comparison analysis among classifiers

Tables 8 to 10 indicate comparative analysis of existing methods on TEL-NLP, SemEval-2018 Task1-C, and GoEmotion datasets. The proposed method is compared with existing methods like BERT-ASL [22], UCCA-GAT [23], RoBERTa-MA [24], and ECO-SAM [25]. The existing methods values of the studies [21, 22, 25] are presented in decimal form which is converted into percentage. Compared to these existing methods, proposed method obtains superior

accuracy of 98.66%, 98.32% on SemEval-2018 Task1-C, GoEmotion datasets for multi-class classification and f1-score of 98.21% for binary classification on TEL-NLP datasets. MCWA concentrates on appropriate chunks of text while LSTM captures long-term dependencies effectively which is crucial for understanding emotional context. Moreover, SELU ensures self-normalization that generates rapid convergence and consistent learning across layers. Hence, the model learns complex patterns effectively which results in superior classification performance.

Table 8. Comparative analysis of existing methods on TEL-NLP dataset

| Methods | F1-score (%) | | | |
	SA	E1	HS	SAR
BERT-ASL [22]	84	55	83	63
Proposed MCWA-LSTM with SELU	**97.54**	**96.32**	**98.21**	**97.06**

Table 9. Comparative analysis of existing methods on SemEval-2018 Task1-C dataset

Methods	Accuracy (%)	Recall (%)	Precision (%)	Macro-F1 (%)	Micro-F1 (%)
BERT-ASL [22]	N/A	54	67	59	N/A
UCCA-GAT [23]	61.2	N/A	N/A	60.0	66.1
RoBERTa-MA [24]	62.4	N/A	N/A	60.3	74.2
Proposed MCWA-LSTM with SELU	**98.66**	**96.54**	**96.32**	**95.53**	**96.24**

Table 10. Comparative analysis of existing methods on GoEmotion dataset

Methods	Accuracy (%)	Recall (%)	Precision (%)	Macro-F1 (%)
BERT-ASL [22]	N/A	54	56	N/A
UCCA-GAT [23]	71.2	N/A	N/A	63.9
ECO-SAM [25]	N/A	58.08	65.81	N/A
Proposed MCWA-LSTM with SELU	**98.32**	**98.12**	**97.04**	**98.23**

4.3 Discussion

The advantage of proposed MCWA-LSTM with SELU and limitations of existing methods are discussed in this section in detail. The limitations of existing methods like MT-Text GCN [21] struggled with inaccurate performance due to rich morphology and complex script which affect effective graph construction and token-level representation. BERT [22] face difficulty in capturing subtle emotion overlaps in multi-label settings because of fixed context windows and limited label dependencies. Semantic and syntactic aware graph attention [23] struggled with scalability while managing texts with numerous emotion labels because of increased graph complexity. ECO-SAM [24] overemphasizes predefined emotion correlations that leads to biased performance while emotions co-occur in unexpected patterns. The proposed MCWA-LSTM with SELU overcomes these existing method limitations by capturing both temporal dependencies and focused attention on emotionally significant chunk words. MCWA aligns attention progressively that preserves sequential structure and highlights significant emotions. LSTM effectively models long-range dependencies for better

understanding context in different emotions. SELU ensures self-normalizing and stable training that minimize vanishing gradient issue and enhance convergence. This integration results in highly discriminative and context-aware model that enhance classification performance.

5. CONCLUSION

This research proposed MCWA-LSTM with SELU to classify the different emotions accurately. LSTM captures long-term dependencies which is crucial to understand sequential flow and emotional nuances. By integrating MCWA, model selectively process emotionally salient text chunks that makes model to capture contextually rich emotional expressions. Moreover, SELU enhance model performance by managing self-normalizing properties that leads to rapid convergence and enhanced training stability. Experimental results show that proposed MCWA-LSTM with SELU achieves better performance compared to existing methods on three datasets. Overall, proposed method provides a context-aware, highly discriminative, and computationally

effective solution for text-based emotion classification tasks. Compared to existing methods like BERT-ASL, proposed MCWA-LSTM with SELU achieves a high accuracy of 98.66%, 98.32% on SemEval-2018 Task1-C, GoEmotion datasets for multi-class classification and f1-score of 98.21% for binary classification on TEL-NLP datasets respectively. Moreover, MCWA achieves less time complexity and memory consumption due to weighted attention which selectively emphasizes most significant features and minimize redundant computations. Hence, MCWA results in rapid convergence and less time complexity. However, the limitation of this research is potential bias in GoEmotion dataset as cultural and linguistic nuances which affects the way emotions are interpreted and expressed accurately which minimize model generalization. In future, imbalanced issue will be solved through oversampling method to further enhance model performance.

REFERENCES

[1] Zhang, D., Song, Y., Xiang, Q., Wang, Y. (2025). IMCMK-CNN: A lightweight convolutional neural network with multi-scale kernels for image-based malware classification. Alexandria Engineering Journal, 111: 203-220. https://doi.org/10.1016/j.aej.2024.10.055

[2] Talaat, A.S. (2023). Sentiment analysis classification system using hybrid BERT models. Journal of Big Data, 10(1): 110. https://doi.org/10.1186/s40537-023-00781-

[3] Kumar, T., Mahrishi, M., Sharma, G. (2023). Emotion recognition in Hindi text using multilingual BERT transformer. Multimedia Tools and Applications, 82(27): 42373-42394. https://doi.org/10.1007/s11042-023-15150-1

[4] Hosseini, S.S., Yamaghani, M.R., Poorzaker Arabani, S. (2024). Multimodal modelling of human emotion using sound, image and text fusion. Signal, Image and Video Processing, 18(1): 71-79. https://doi.org/10.1007/s11760-023-02707-8

[5] Talaat, F.M., El-Gendy, E.M., Saafan, M.M., Gamel, S.A. (2023). Utilizing social media and machine learning for personality and emotion recognition using PERS. Neural Computing and Applications, 35(33): 23927-23941. https://doi.org/10.1007/s00521-023-08962-7

[6] Al-Tameemi, I.K.S., Feizi-Derakhshi, M.R., Pashazadeh, S., Asadpour, M. (2023). Interpretable multimodal sentiment classification using deep multi-view attentive network of image and text data. IEEE Access, 11: 91060-91081. https://doi.org/10.1109/ACCESS.2023.3307716

[7] Glenn, A., LaCasse, P., Cox, B. (2023). Emotion classification of Indonesian tweets using bidirectional LSTM. Neural Computing and Applications, 35(13): 9567-9578. https://doi.org/10.1007/s00521-022-08186-1

[8] Olusegun, R., Oladunni, T., Audu, H., Houkpati, Y.A.O., Bengesi, S. (2023). Text mining and emotion classification on monkeypox Twitter dataset: A deep learning-natural language processing (NLP) approach. IEEE Access, 11: 49882-49894. https://doi.org/10.1109/ACCESS.2023.3277868

[9] Peng, S., Zeng, R., Cao, L., Yang, A., Niu, J., Zong, C., Zhou, G. (2023). Multi-source domain adaptation method for textual emotion classification using deep and broad learning. Knowledge-Based Systems, 260: 110173. https://doi.org/10.1016/j.knosys.2022.110173

[10] Machová, K., Szabóova, M., Paralič, J., Mičko, J. (2023). Detection of emotion by text analysis using machine learning. Frontiers in Psychology, 14: 1190326. https://doi.org/10.3389/fpsyg.2023.1190326

[11] Deng, S., Wu, L., Shi, G., Xing, L., Jian, M., Xiang, Y., Dong, R. (2024). Learning to compose diversified prompts for image emotion classification. Computational Visual Media, 10(6): 1169-1183. https://doi.org/10.1007/s41095-023-0389-6

[12] Sams, A.S., Zahra, A. (2023). Multimodal music emotion recognition in Indonesian songs based on CNN-LSTM, XLNet transformers. Bulletin of Electrical Engineering and Informatics, 12(1): 355-364. https://doi.org/10.11591/eei.v12i1.4231

[13] Peng, S.X., Chen, K., Tian, T., Chen, J.Y. (2024). An autoencoder-based feature level fusion for speech emotion recognition. Digital Communications and Networks, 10(5): 1341-1351. https://doi.org/10.1016/j.dcan.2022.10.018

[14] Milintsevich, K., Sirts, K., Dias, G. (2023). Towards automatic text-based estimation of depression through symptom prediction. Brain Informatics, 10(1): 4. https://doi.org/10.1186/s40708-023-00185-9

[15] Li, Y., Chan, J., Peko, G., Sundaram, D. (2024). An explanation framework and method for AI-based text emotion analysis and visualisation. Decision Support Systems, 178: 114121. https://doi.org/10.1016/j.dss.2023.114121

[16] Anzum, F., Gavrilova, M.L. (2023). Emotion detection from micro-blogs using novel input representation. IEEE Access, 11: 19512-19522. https://doi.org/10.1109/ACCESS.2023.3248506

[17] Bird, J.J., Ekárt, A., Faria, D.R. (2023). Chatbot interaction with artificial intelligence: Human data augmentation with T5 and language transformer ensemble for text classification. Journal of Ambient Intelligence and Humanized Computing, 14(4): 3129-3144. https://doi.org/10.1007/s12652-021-03439-8

[18] Gou, Z., Li, Y. (2023). Integrating BERT embeddings and BiLSTM for emotion analysis of dialogue. Computational Intelligence and Neuroscience, 2023(1): 6618452. https://doi.org/10.1155/2023/6618452

[19] Zhang, X., Wu, Z., Liu, K., Zhao, Z., Wang, J., Wu, C. (2023). Text sentiment classification based on BERT embedding and sliced multi-head self-attention Bi-GRU. Sensors, 23(3): 1481. https://doi.org/10.3390/s23031481

[20] Liu, X., Xu, Z., Huang, K. (2023). Multimodal emotion recognition based on cascaded multichannel and hierarchical fusion. Computational Intelligence and Neuroscience, 2023(1): 9645611. https://doi.org/10.1155/2023/9645611

[21] Marreddy, M., Oota, S.R., Vakada, L.S., Chinni, V.C., Mamidi, R. (2022). Multi-task text classification using graph convolutional networks for large-scale low resource language. In 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, pp. 1-8. https://doi.org/10.1109/IJCNN55064.2022.9892105

[22] Ramakrishnan, S., Babu, L.D. (2025). Improving multi-label emotion classification on imbalanced social media data with BERT and clipped asymmetric loss. IEEE Access, 13: 60589-60601. https://doi.org/10.1109/ACCESS.2025.3557091

[23] Ameer, I., Bölücü, N., Sidorov, G., Can, B. (2023).

Emotion classification in texts over graph neural networks: Semantic representation is better than syntactic. IEEE Access, 11: 56921-56934. https://doi.org/10.1109/ACCESS.2023.3281544

[24] Ameer, I., Bölücü, N., Siddiqui, M.H.F., Can, B., Sidorov, G., Gelbukh, A. (2023). Multi-label emotion classification in texts using transfer learning. Expert Systems with Applications, 213: 118534. https://doi.org/10.1016/j.eswa.2022.118534

[25] Ni, Y., Ni, W. (2024). A multi-label text sentiment analysis model based on sentiment correlation modeling. Frontiers in Psychology, 15: 1490796. https://doi.org/10.3389/fpsyg.2024.1490796

[26] Gamage, G., De Silva, D., Mills, N., Alahakoon, D., Manic, M. (2024). Emotion AWARE: An artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis. Journal of Big Data, 11(1): 93. https://doi.org/10.1186/s40537-024-00953-2

[27] Kukkar, A., Mohana, R., Sharma, A., Nayyar, A., Shah, M.A. (2023). Improving sentiment analysis in social media by handling lengthened words. IEEE Access, 11: 9775-9788.
https://doi.org/10.1109/ACCESS.2023.3238366

[28] Erkan, A., Güngör, T. (2023). Analysis of deep learning model combinations and tokenization approaches in sentiment classification. IEEE Access, 11: 134951-134968.
https://doi.org/10.1109/ACCESS.2023.3337354

[29] Holla, L., Kavitha, K.S. (2024). An improved fake news detection model using hybrid time frequency-inverse document frequency for feature extraction and adaboost ensemble model as a classifier. Journal of Advances in Information Technology, 15(2): 202-211. https://doi.org/10.12720/jait.15.2.202-211

[30] Asqolani, I.A., Setiawan, E.B. (2023). Hybrid deep learning approach and Word2VEC feature expansion for cyberbullying detection on Indonesian Twitter. Ingenierie des Systemes d'Information, 28(4): 887. https://doi.org/https://doi.org/10.18280/isi.280410

[31] Xue, H., Guo, C., Dong, G., Zhang, C., Lian, Y., Yuan, Q. (2025). Prediction of runoff in the upper reaches of the Hei River based on the LSTM model guided by physical mechanisms. Journal of Hydrology: Regional Studies, 58: 102218. https://doi.org/10.1016/j.ejrh.2025.102218

[32] Tang, J., Xu, L., Wu, X., Chen, K. (2024). A short-term forecasting method for ionospheric TEC combining local attention mechanism and LSTM model. IEEE Geoscience and Remote Sensing Letters, 21: 1-5. https://doi.org/10.1109/LGRS.2024.3373457

[33] Zhou, K., Tong, Y., Li, X., Wei, X., Huang, H., Song, K., Chen, X. (2023). Exploring global attention mechanism on fault detection and diagnosis for complex engineering processes. Process Safety and Environmental Protection, 170: 660-669. https://doi.org/10.1016/j.psep.2022.12.055

[34] Li, Z., Li, L., Chen, J., Wang, D. (2024). A multi-head attention mechanism aided hybrid network for identifying batteries' state of charge. Energy, 286: 129504. https://doi.org/10.1016/j.energy.2023.129504