

A MULTI-MODEL SOFT-VOTING HYBRID ENSEMBLE DEEP LEARNING FRAMEWORK FOR ADVANCED TEXT CLASSIFICATION

D.JASMINE GUNASUNDARI¹, J.ARUNADEVI²✉

¹Ph.D Research Scholar (Part – Time), Department of Computer Science,
Alagappa University, Tamilnadu, India

²✉ Assistant Professor, Department of Computer Science,
Raja Doraisingam Govt. Arts College, Sivaganga, Alagappa University, Tamilnadu, India

✉ Corresponding author : J.Arunadevi, E-mail : arunardm2015@gmail.com

ABSTRACT

This study addresses a specific difficulty in automated emotion recognition from social media text: informal spelling, slang, elongated words, emojis, and hashtags routinely degrade classifier accuracy. A Multi-Model Deep Learning Architecture for Social Media Emotion Detection (MMDLA-SMED) is proposed, combining a dedicated normalization pipeline with a soft-voting ensemble classifier. Raw posts are processed through cleaning, hashtag segmentation, slang expansion, emoji-to-text conversion, word-elongation normalization, and lightweight spelling correction. The normalized text is encoded using FastText, whose sub-word character n-gram representation embeds out-of-vocabulary and misspelled terms common in social media posts. Classification combines three heterogeneous deep learning architectures, an improved temporal convolutional network (TCN), a residual gated recurrent unit (ResGRU), and a hierarchical attention network (HAN), trained independently and merged through soft voting so convolutional, recurrent, and attention-based decision patterns jointly determine the predicted emotion. MMDLA-SMED is evaluated on two benchmark tweet-emotion datasets, one with 12 classes and 39,173 samples and another with 6 classes and 16,000 samples, under a 70:30 train-test split. On the 12-class dataset, the model attains a testing accuracy of 98.72%, precision of 87.95%, recall of 73.15%, and F1-score of 74.82%; on the 6-class dataset it attains a training-phase accuracy of 98.38%, precision of 94.28%, recall of 91.56%, and F1-score of 92.81%, outperforming Random Forest, Decision Tree, LightGBM, GCN, GraphSAGE+Pooling, GoLA, DIMINANT, GAAN, 1D-CNN, LSTM, Bi-LSTM, CNN-LSTM, CNN-BiLSTM, MAPLE, and DSGN on every measure. Class-wise results show minority categories such as boredom and anger are recognized with markedly lower recall than dominant categories, reflecting class imbalance in both datasets rather than an architectural limitation. These findings indicate that pairing subword-aware embeddings and informal-text normalization with a soft-voting ensemble of complementary deep learning models is effective for emotion classification in noisy, user-generated social media text, while pointing to class imbalance as the main remaining bottleneck for rare-emotion detection.

Keywords: *Emotion Identification, Social Media Data, Word Embeddings, Deep Learning, Ensemble Classifier*

1. INTRODUCTION

The rapid growth of social networking platforms such as Instagram, Twitter (X), Reddit, and Facebook has changed the way people communicate and has reshaped public discourse [1]. These platforms function as open spaces where users share opinions, personal experiences, and social or political viewpoints in real time [2]. As a result, everyday activity on these platforms generates large volumes of unstructured text that form a valuable resource for studying individual and collective opinion. Because such posts are informal and spontaneous, they are also particularly

useful for capturing sentiment across diverse contexts, communities, and time periods [3]. This growth of user-generated text has made social media a central focus of computational linguistics and data-driven decision-making. Sentiment Analysis (SA), also referred to as opinion mining, is the sub-field of computational linguistics concerned with automatically identifying and categorizing the opinions, evaluations, and emotional orientation expressed in text [4], [29]. Conventionally, SA assigns text to positive, negative, or neutral polarity classes, whereas Emotion Detection, the specific problem addressed

in this paper, extends this idea by assigning text to a finer set of discrete affective categories such as joy, sadness, anger, fear, and surprise, grounded in established psychological models of emotion [30].

Natural Language Processing (NLP), the sub-field of Artificial Intelligence (AI) concerned with enabling machines to parse, interpret, and generate human language, provides the core tool-set for extracting insight from the ever-growing volume of social media data [5]. When applied to social media text, NLP methods can uncover linguistic patterns that reflect the emotional state of users [6]. Emotion analysis, a major application area of NLP, evaluates such patterns to obtain a better understanding of the psychological well-being of users at scale [7]. Classical approaches to sentiment and emotion classification relied on shallow supervised classifiers such as Support Vector Machines (SVMs) and Naive Bayes (NB), which depend on hand-crafted linguistic features [8]. Deep Learning (DL), a family of neural network methods that automatically learns hierarchical feature representations directly from raw data without manual feature engineering, has since become the dominant paradigm for this task because it captures the complex, non-linear semantic structure of text more effectively than shallow models. SA is generally studied at three levels of granularity: document-level SA (DLSA), which assigns a single sentiment label to an entire document; sentence-level SA (SLSA), which separately classifies each sentence and additionally distinguishes subjective from objective content; and aspect-based SA (ABSA), which identifies sentiment toward specific entities or attributes mentioned in the text and is generally regarded as more fine-grained than both DLSA and SLSA, since it must resolve the relationship between an opinion phrase and its target [9], [10]. Every minute, an enormous number of comments, blog posts, and micro-blog updates are added to social platforms, and SA techniques are used to make sense of this continuous stream of public opinion [11]. A further motivation for automated emotion recognition is emotional contagion, the phenomenon whereby exposure to emotionally expressive content can shape the mood and behavior of other users, often without their conscious awareness.

To enhance the reliability of emotion recognition from social media content, this paper presents a Multi-Model Deep Learning Architecture for Social Media Emotion Detection (MMDLA-SMED). The proposed MMDLA-SMED model

encompasses a comprehensive step for better classification of emotions, which consists of preprocessing of input text, a FastText-based word embedding approach, and an ensemble deep learning-driven emotion classification approach. The effectiveness of the MMDLA-SMED model is inspected using benchmark datasets across diverse measures. The core contribution of this manuscript is as follows:

- Employ the FastText model as a robust word embedding strategy, which converts social platform data into meaningful vector representations.
- Design a soft-voting deep learning-based ensemble classifier, combining an improved temporal convolutional network (TCN), residual gated recurrent unit (ResGRU), and hierarchical attention network (HAN), achieving robust and precise emotion detection.
- Conduct extensive simulation and comparative analyses using benchmark datasets to evaluate the MMDLA-SMED model, demonstrating superior performance over other techniques and models across different evaluation measures.

2. REVIEW OF RELATED LITERATURE

This section briefly reviews prior related works regarding emotion detection from social media content. Kusal et al. [12] intended to examine the extent to which the utilization of emojis is involved in the automatic identification of emotions in text data, with a focus on online social communications. This emoji-aware hybrid DL method is used for multimodal emotion identification. In [13], an application of a multi-class SA technique for the Amharic language was presented. This system uses a method called integrated DL. Integration of the layers of a CNN with BiLSTM leads to the expansion of this multi-class SA architecture, which is considered the bidirectional convolutional ENN (BCRNN). The abilities of the BiLSTM to collect the sequential dependency could be integrated with the robustness of the CNN for gaining the local attributes. According to these dual variables, the method has a superior ability to gain fine-grained data and contextual data that can be essential to perform an SA on Amharic inscribed texts. Zhou and Mohd [14] developed an innovative architecture for depression identification that combines cutting-edge text preprocessing with a DL method. The preprocessing pipeline comprises dual main modules, where the emojis have been transformed

into texts through emoji normalization, and after slang expressions can be substituted by a customized dictionary. Emotion scores assist as supplementary elements for identifying the depression-based emotional indications in the texts. This system integrates BERT's contextual embeddings [35] with the sequential processing ability of BiLSTM for efficiently representing the emotional contents in social networking site updates.

Khemani et al. [15] implemented an improved graph convolutional network (IGCN), which uses the network architecture of social media texts for improving emotion recognition. Moreover, an attention mechanism (AM) optionally highlighted the contextually important words, thus boosting the classification correctness and interpretability. In [16], the DL-based sentiment and emotion classification (SENTI-EMO) was inspected for providing the correct outcomes. Primarily, the pre-processing of twitter messages, like normalization, correction, tokenization and noise removal, is accomplished to increase the classification accuracy. In the noise removal stage, the stop word method has been utilized for learning about the context of tweets for enhancing the classification outcomes. Later, the pre-processing method, SA classification is completed by utilizing the attention-based GRU that comprises various layers for extracting the higher-level features. In the attention layer, only crucial features have been measured to decrease the redundancy, which improves the precision of SA classification. The extracted features are classified into sentiments like strongly positive, positive, neutral, and negative. These categorized sentiments have been integrated by employing the spatiotemporal optics (STO) model that creates cluster-based time stamp Twitter ID, and geographical position.

Shwetha and Pushpalatha [17] considered the multi-MentalBERT, a hybrid DL architecture for performing the two-stage multi-class classification by using the monolingual and code-mixed datasets. This model incorporates the multilingual contextual attributes originated from MuRIL with psychological health-related embeddings of MentalBERT by employing the cross-attention fusion, maintained by hierarchical attention and deep joint autoencoder (AE) for making a combined learning representation. Lee et al. [18] presented an innovative technique for improving emotion detection through changing SNS informal texts into normal language via ChatGPT, which provides a generative large

language model (LLM). The 4-transformer architecture, XLM-RoBERTa, BERT, DistilBERT, and RoBERTa are trained with the integrated dataset that exceeded such trained texts and denoted that converting the SNS vernacular texts into a common language increases the emotion identification.

3. PROPOSED MODEL ARCHITECTURE

This study proposes the MMDLA-SMED model for improving the accuracy of emotion recognition from social media texts. Fig. 1 showcases the end-to-end architecture of the MMDLA-SMED model. As exhibited in Fig. 1, the process begins with text preprocessing, where cleaning, hashtag segmentation, slang expansion, emoji conversion, word elongation normalization, and lightweight spelling correction techniques are applied to prepare the input for analysis. Subsequently, the FastText architecture is deployed for converting the text into meaningful numerical vectors, capturing semantic information. Besides, the classification process is executed using a soft-voting ensemble approach that combines improved TCN, ResGRU, and HAN techniques. Each model predicts class probabilities independently, and the final emotion label is determined by averaging these probabilities. This strategy improves overall classification stability and performance compared to a single model.

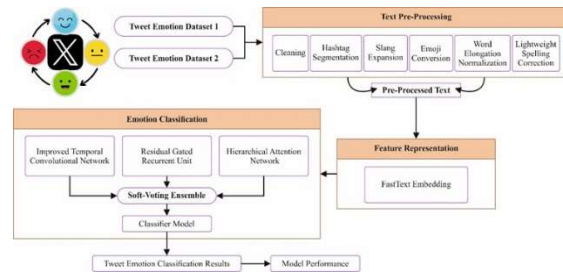


Figure 1: End-to-end workflow of the MMDLA-SMED architecture

3.1. Text Pre processing Pipeline

Text preprocessing ensures the dataset is reliable, proper, and accurate, resulting in higher performance of the model. In this study, a vital normalization pipeline is utilized for cleaning and standardizing the textual data whereas maintaining the emotional semantics [19]. Primarily, the basic cleaning process is executed, such as redundant whitespace, elimination of URLs, lowercasing, and user mentions (@user). Hashtags have been processed by removing the hash symbol (#) and executing the hashtag segmentation in the representation of meaningful essential words. This

stage allows the model to gain latent semantic data embedded in the multiple hashtags.

In order to overcome the informal language utilization, the slang terms and abbreviations must be extended by applying the predetermined lexical dictionary (for example, “omg” → “oh my god”, “tbh” → “to be honest”). Emojis are transformed into their respective textual descriptions to maintain the emotional signs exhibited using pictographic symbols. Also, extended words are utilized for emotional highlights (like “soooo happy”) that can be normalized by decreasing the recurrent characters to their normal formations of words. In conclusion, the lightweight spelling correction depends on the edit-distance has been used for correcting small typographical errors. This important preprocessing model decreases the noise while preserving emotional integrity and context of the texts, thus enhancing the efficacy of consequent embedding and classification phases.

3.2. Semantic Feature Representation using FastText

FastText is employed to transform social media text into meaningful semantic vector representations for effective feature learning. The FastText is a subclass of Word2Vec designed by Facebook’s AI Research (FAIR) lab, increases word embedding by taking into account of sub-word character n -gram details [33], as demonstrated in related word-embedding-based classification studies [20]. As an alternative to learning word vectors, FastText directly indicates every individual word as a bag of character n -grams. This methodology enables FastText to produce embeddings for out-of-vocabulary terms while identifying internal word structures. The loss function can be utilized for FastText in which same as the Word2Vec but integrates the subword information.

$$J_{FastText} = - \sum_{w_i \in context} \log \sigma(v_{w_i}^T v_{w_t}) \quad (1)$$

Here the parameter v_{w_i} denotes context word’s embedding vector and v_{w_t} refers to target word’s embedding vector. With the usage of subword details, FastText has achieved increased efficiency on tasks that involve unique and complex morphological languages. This embedding method are chosen due to provide a balance among simplicity, performance and interpretability that can be more crucial for tasks.

3.3. Multi-Model Emotion Classification Framework

The classification process is carried out through using a soft-voting ensemble mechanism that integrates three DL architectures, such as improved TCN, ResGRU, and HAN.

3.3.1. Soft-Voting Ensemble

In order to increase the strength of emotion classification, a soft voting ensemble approach is used for integrating the classification of Improved TCN, Residual GRU, and Hierarchical Attention Network (HAN). Every base model generates a probability distribution using the emotion classes. Rather than choosing the majority class (hard voting), soft voting averages the projective possibilities from whole models and preserves the confidence data. The last class label could be evaluated by selecting the class with the maximum combined probability. In the mathematical manner, the final classification can be measured by,

$$\hat{y} = \arg \max_k \left(\frac{1}{M} \sum_{m=1}^M p_{m,k} \right) \quad (2)$$

Here the parameter M refers to the number of models and $p_{m,k}$ indicates the probability projected at the m^{th} model for classification.

3.3.2. Improved TCN

The TCN is dependent upon the conventional one-dimensional fully convolutional neural network (1DFCNN) method, and integrates the residual block architecture, dilated convolution and causal convolution [21], [34]. Accordingly, the model has the capability for feature representation and the achievement of classification. Then, efficiently overcome the performance degradation issue of deep networks in the network training process. The dilated causal convolution has been utilized for feature representation, the weight norm controlled hidden layer’s (HL) inputs to address the gradient explosion complexity in the network architecture, the ReLU provides non-linear mapping, and Dropout implements regularization through randomly dropping neurons, thus it is prevented the overfitting complexity and boosted the training speed of model. Given input x , the residual block produces,

$$o = Activation(x + F(x)) \quad (3)$$

Whereas the mathematical term activation refers to the activation function and $F(x)$ remains

the residual. AS the residual $F(x)$ cannot be 0 in real-time, the stacked layers in the DL architecture is learned new features, thus the learning outcome of the deep network cannot be decreased. Casual Convolutions are initially presented within the WaveNets architecture. The receptive field expansion of NN neurons in conventional causal convolution requires several network layers, as stacked or larger convolution kernels are utilized. Moreover, TCN integrates the dilated convolution with causal convolution, and utilizes the dilated causal convolution (DCC) for improving the receptive field of neurons with no major rise in computational expenses. The 1D DCC model can be represented by,

$$F(s) = \sum_{i=0}^{k-1} f(i)x_{s-di} \quad (4)$$

Here, the parameter $f(i)$ represents the filter, otherwise called as convolution kernels, x refers to the input sequence, k denotes the dimension of convolution kernels, d indicates the dilation parameter, $s - di$ is confirmed that only previous inputs must be measured.

Modeling human cognitive focus, the attention mechanism (AM) assigns distinct weights to features within the HL, thus highlighting the effect of significant features. The self-attention mechanism (SAM) is an enhancement of the AM that intentions to gain the internal relationship of the data and increase the capability classification model. The SAM comprises the V (Value), K (Key), and Q (Query) vectors, and these have been acquired through the multiplication of the input text data using 3 parameters W_v, W_k , and W_q . The mathematical terms W_q, W_k , and W_v can be the conversions of text data, and the attention matrix has been applied for determining the feature attention.

$$attention(Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d_K}} \right), \quad (5)$$

Now, the parameter d_K represents the dimension of K and T means the matrix transpose.

The addition of self-attention earlier as every convolutional module in the TCN residual block architecture, and the improved TCN residual block model is adopted, where the deep TCN architecture contains an input layer for text data, dual stacked enhanced residual blocks for increasing model depth, and an output fully connected layer to compute the classified values.

According to the improved residual block model, the architecture enhances training suitability and effectively generates the output vector from the TCN framework for classification. Fig. 2 demonstrates the architecture of the improved TCN.

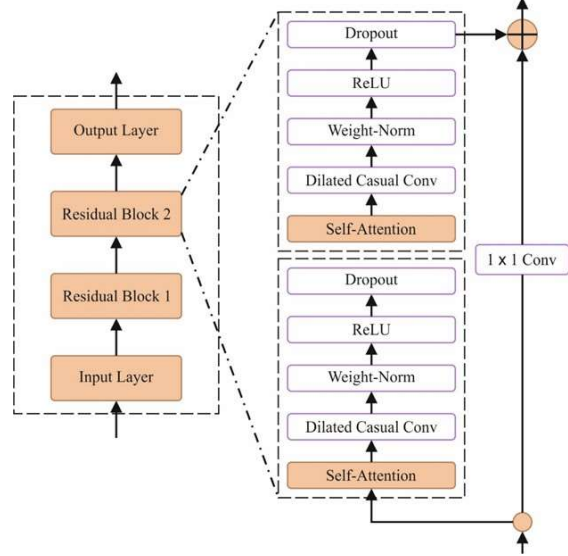


Figure 2: Architecture of the improved TCN

3.3.3. ResGRU

ResGRU is established by implementing the residual block in the GRU architecture for feature representation with the specified text data. The GRU method has been utilized for managing the extended memory and the sequential complexity in the network architecture [22], [31]. The HL in the network can be exchanged by utilizing the gate model for evaluating the important text data and preventing redundant data. The input representations must be identified by employing the reset gate s_t , with the help of the update gate y_t and preceding memory. The text data is obtained on the basis of the maximal value of y_t . The quantity of data is evaluated by applying the reset gate in the node g_t . Also, the HL should be measured in the existing movements through the update gate. The mathematical form is calculated by,

$$y_t = \sigma(X_y z_t + V_y g_{t-1} + c_z) \quad (6)$$

$$s_t = \sigma(X_s z_t + V_s g_{t-1} + c_s) \quad (7)$$

$$g_t = \tanh(X_g z_t + V_g (s_t \otimes g_{t-1}) + c_g) \quad (8)$$

$$g_t = (1 - y_t) \otimes g_{t-1} + y_t \otimes g_t \quad (9)$$

From this expression, the symbol $\otimes$ refers to the Hadamard product. The bias value is

represented by the terms c_y, c_s, c_g . The existing movement HL output is denoted as g_t , and the weight matrices must be specified as $X_y, X_s, X_g, V_y, V_s, V_g$.

The degradation outcome is addressed by employing the residual network. The identity mapping has been used through the accumulation layer after the residual will be represented as 0. Totalling the final GRU outputs mirrors residual mapping. The residual block can be indicated in this mathematical form,

$$z_R = \text{relu}(Bn_{\gamma, \beta}(z) + h(y_t)) \quad (10)$$

Where the batch normalization (BN) and ReLU activation operations have been denoted by $BN(\cdot)$ and $\text{relu}(\cdot)$. The learnable variables can be represented as γ and β . The relationship among the text data and the semantic information is recollected by utilizing the ResGRU. It helps to increase the identification efficiency.

3.3.4. HAN

A multi-level attention framework is combined with two path encoders and a multi-head attention (MHA) fusion mechanism for social media emotion classification [23], building on the original hierarchical attention formulation [32].

The two-path encoder comprises a detailed encoder path and a fast-encoding path. The fast path utilizes a lightweight multi-layer perceptron (MLP) for processing the textual feature representations $s_{feature}$:

$$h_{fast} = \text{ReLU}(W_2 \cdot \text{ReLU}(W_1 \cdot s_{feature} + b_1) + b_2), \quad (11)$$

Here, the mathematical form $W_1 \in \mathbb{R}^{64 \times d}$, $W_2 \in \mathbb{R}^{32 \times 64}$, and b_1, b_2 describe biases. This path quickly extracts the main emotional features from the input representation.

$$h_{detailed} = \text{Transformer}(\text{concat}(s_{raw}, s_{semantic}, s_{context})). \quad (12)$$

The MHA fusion model balances the involvement of the dual encoder paths through effectively changing their weights with respect to textual feature complexity. The mathematical form is derived by,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (13)$$

$$h_{fused} = \text{MultiHead}(h_{fast}, h_{detailed}) \quad (14)$$

$$\alpha = \text{sigmoid}(W_f \cdot s_{feature} + W_c \cdot s_{complexity} + b) \quad (15)$$

$$h_{final} = \alpha \cdot h_{fast} + (1 - \alpha) \cdot h_{detailed}. \quad (16)$$

The gating coefficient α is learned dynamically instead of fixed. This can be actively shifted attention among the fast path ($\alpha \approx 1$ for higher noise) and the detailed path ($\alpha \approx 0$ for intricate semantics), thus accelerating the classification effectiveness.

In order to provide multi-scale feature extraction, a symmetric dual-path encoder model is developed, and a balance is accomplished between computation effectiveness and semantic depth. Particularly, the fast path employs a lightweight MLP for quickly extracting the significant data from features $s_{feature}$,

$$h_{fast} = \text{ReLU}(W_2 \cdot \text{ReLU}(W_1 \cdot s_{feature} + b_1) + b_2), \quad (17)$$

Here the parameters W_1 and W_2 describes weight matrices. On the other hand, the detailed path utilizes a stacked hierarchical attention layer to gain the intricate dependencies in the combined the semantic contexts $[s_{ra}, s_{semantic}, s_{context}]$,

$$h_{detailed} = \text{Transformer}(\text{concat}(s_{raw}, s_{semantic}, s_{context})). \quad (18)$$

The fusion model accomplishes fast path balancing through swift data component as the query for a thorough context dataset ($K = V$). A variable weighting value σ , calculated from the data's complexity and the merging of features, then regulates this process.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (19)$$

$$\alpha = \alpha \cdot h_{fast} + (1 - \alpha) \cdot h_{detailed}, \quad (20)$$

Here, the mathematical value $\alpha \approx 1$ is prioritized the fast path, if $\alpha \approx 0$ highlights comprehensive semantics, thus assuring adaptive feature weighting.

Moreover, a dual network framework has been utilized for decoupling the state values from the benefits, thus improving the stability,

$$y = \text{Softmax}(W_o \cdot h_{final} + b_o). \quad (21)$$

For assuring constant convergence, a multi-stage training approach is utilized and developed from a large-scale social media dataset training together with mini-batch optimization. This method is used to decrease the categorical cross-entropy loss,

$$L = - \sum_{i=1}^K y_i \log(\hat{y}_i) \quad (22)$$

The hierarchical attention mechanism estimates contextual importance weights at the

word-level and sentence-level instead of computing them directly from raw sequences. The model is pre-trained on processed textual corpora and continuously updated during supervised training. In order to overcome contextual variability in social media text, a sliding window (SW) mechanism with a fixed length has been utilized. A first-in-first-out (FIFO) approach confirms out-of-date contextual representations are rejected, thus permitting the model to adapt dynamically to modern data distribution in social media streams.

The prediction probabilities attained from the Improved TCN, ResGRU, and HAN are integrated through a soft-voting strategy. The final emotion label is determined depending on the averaged class probabilities, thus improving overall classification robustness and stability.

4. EXPERIMENTAL RESULT ANALYSIS

In this section, the performance validation the MMDLA-SMED method is inspected under dual benchmark datasets like Emotion Detection from Text (Tweet Emotions Dataset 1)[24] and Tweet Emotions (Tweet Emotions Dataset 2) [25].

4.1. Result Analysis on Tweet Emotions Dataset 1

Emotion detection from text is major challenging issue in NLP. The purpose is the inaccessibility of labeled dataset and the multi-class types of the problem. Humans have a variation of emotions and it is complex to gather enough records for every emotion and hence the complexity of class imbalance rises. Now, we have a labeled data for emotion detection and the major intention is to make an efficient model to identify the emotion. The dataset encompasses 39173 under 12 samples as represented in Table 1. Table 2 shows the sample texts of tweet emotions dataset 1.

Table 1: Details on tweet emotions dataset 1

Emotions	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	39173

Table 2: Sample text on tweet emotions dataset 1

S. No	Sample Text	Emotions
1	The storm is here and the electricity is gone	Sadness
2	So, I need to make a lot of money tomorrow	Enthusiasm
3	Chocolate milk is so much better through a straw. I lack said straw	Neutral
4	Choked on her retainers	Worry
5	OMGididnt tweet all that much today.... *sadness*	Surprise
6	Pug just woke me from an incredible sleep!	Love
7	RIP leonardo. You were a great mini fiddler crab	Fun
8	It is so annoying when she starts typing on her computer in the middle of the night!	Hate
9	good lord, just seen the agency work for our upcoming Open Day. Money for Jam	Happiness
10	why did i agree to work a double shift	Boredom
11	I'm at work	Relief
12	Working But it's Fridaaaaayyyyy	Anger

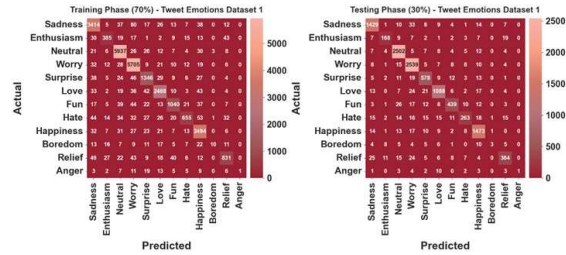


Figure 3: Confusion matrices of MMDLA-SMED method with 70:30 under tweet emotions dataset 1

Fig. 3 exemplifies the confusion matrices created by the MMDLA-SMED method at 70:30 of TRAP/TESP. The simulation outcome denoted that the MMDLA-SMED technique has efficient detection and identification of all 12 classes exactly.

Table 3 and Fig. 4 display the classifier outcomes of MMDLA-SMED system with 70:30 on tweet emotion dataset 1. The experimental outcome value suggests that the MMDLA-SMED model appropriately identified the samples. According to 70% TRAP, the MMDLA-SMED technique offers average $accu_r$, $preci_n$, $recal_l$, F_{score} and MCC of 98.70%, 88.69%, 73.15%, 74.66%, and 75.53%, correspondingly. Meanwhile, with 30% TESP, the MMDLA-SMED method provides average $accu_r$, $preci_n$, $recal_l$, F_{score} and MCC of 98.72%, 87.95%, 73.15%, 74.82%, and 75.85%.

Table 3: Classifier outcomes of MMDLA-SMED method with 70:30 on tweet emotion dataset 1

Classes	Accur _y	Preci _n	Recal _l	F1 _{score}	MCC
Training Phase (70%)					
Sadness	98.01	91.63	93.56	92.58	91.44
Enthusiasm	99.09	79.22	72.10	75.49	75.11
Neutral	98.50	95.70	97.65	96.66	95.71
Worry	98.10	93.89	97.45	95.64	94.46
Surprise	98.56	86.73	87.74	87.23	86.47
Love	98.58	92.57	92.78	92.68	91.89
Fun	98.76	88.51	83.53	85.95	85.34
Hate	98.61	86.87	69.83	77.42	77.20
Happiness	98.29	92.02	95.44	93.70	92.73
Boredom	99.56	71.43	07.81	14.08	23.53
Relief	98.67	85.67	78.62	81.99	81.38
Anger	99.71	100.00	01.23	02.44	11.09
Average	98.70	88.69	73.15	74.66	75.53
Testing Phase (30%)					
Sadness	98.39	93.34	94.26	93.80	92.87
Enthusiasm	99.25	84.42	74.67	79.25	79.02
Neutral	98.43	95.13	97.81	96.45	95.46
Worry	97.99	93.72	97.47	95.56	94.29
Surprise	98.58	86.27	88.51	87.38	86.63
Love	98.45	92.52	92.05	92.28	91.42
Fun	98.72	88.33	82.67	85.41	84.79
Hate	98.54	84.29	68.31	75.47	75.16
Happiness	98.34	92.47	95.16	93.79	92.85
Boredom	99.57	60.00	05.88	10.71	18.70
Relief	98.55	84.85	77.61	81.07	80.40
Anger	99.76	100.00	03.45	06.67	18.55
Average	98.72	87.95	73.15	74.82	75.85

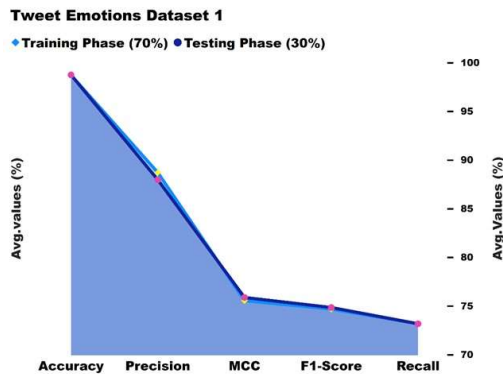


Figure 4: Average outcomes of MMDLA-SMED method with 70:30 on tweet emotion dataset 1

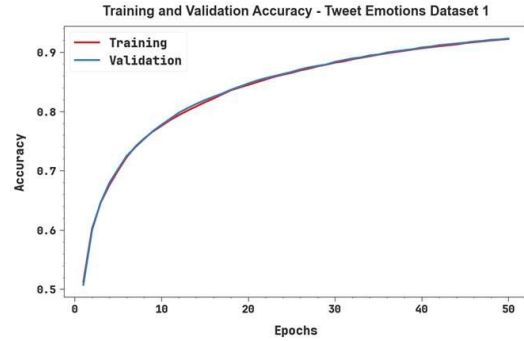


Figure 5: Accur_y curve of MMDLA-SMED method on tweet emotions dataset 1

Fig. 5 exhibits the training (TRAIN) and validation (VALID) accuracy curves of the MMDLA-SMED technique on tweet emotions dataset 1 and 300 epochs. TRAIN accuracy is gradually upsurged and gained the higher level, thus signifying useful learning and convergence. The VALID accuracy is closely noticed the training tendency with a smaller and consistent gap, as well as it shows lesser overfitting and greater generality.

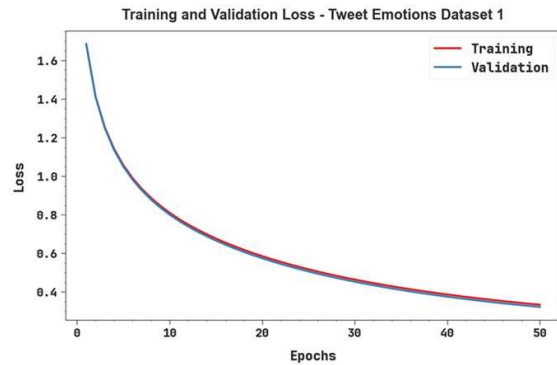


Figure 6: Loss curve of MMDLA-SMED method with tweet emotions dataset 1

Fig. 6 shows the TRAIN and VALID loss curves of the MMDLA-SMED technique with tweet emotions dataset 1 by using the 300 epochs. Dual losses are majorly reduced at the preliminary epochs and it denotes the fast learning, and later progressively minimized in a smooth and constant way, thus considering the efficient convergence. The closer arrangement among training and validation loss, with only a minor gap, thereby suggesting the declined overfitting and strong generalization ability.

The comparison analysis of MMDLA-SMED technique with existing model are shown in Table 4 [26-28]. The simulation outcome values are defined that the MMDLA-SMED method outperformed higher performances. The experimental result established that the proposed

MMDLA-SMED approach realizes superior solutions with $accu_r$ of 98.72%, $prec_i$ of 87.95%, $recal_i$ of 73.15% and $F1_{score}$ of 74.82%, correspondingly. Similarly, the GraphSAGE+Pooling, RF, DT, and GCN methods are acquired lower outcomes. Also, the Light GBM, GoLA, DIMINANT and GAAN techniques are provides inferior performance.

Table 4: Comparison analysis of MMDLA-SMED method with existing models under tweet emotions dataset 1

Algorithms	$Accu_r$	$Prec_i$	$Recal_i$	$F1_{score}$
Random Forest	90.01	78.39	72.89	66.22
Decision Tree	90.03	73.19	71.51	70.44
Light GBM	78.32	70.72	69.34	70.88
GCN	90.05	68.90	71.49	66.05
GoLA	69.00	76.02	68.45	79.47
DIMINANT	71.00	74.08	68.31	73.15
GraphSAGE+Pooling	92.00	70.73	67.40	70.53
GAAN	49.00	56.15	59.61	52.30
MMDLA-SMED	98.72	87.95	73.15	74.82

4.2. Result Analysis on Tweet Emotions Dataset 2

This dataset is developed for Facial Emotion Recognition (FER) tasks and comprises labeled face images that indicates a range of human emotions. FER is a main issue in CV and sentimental computation with applications in human-computer interaction (HCI), surveillance systems, mental health analysis, and AI-driven user experience improvement. The dataset is formed to assist the image classification and DL techniques, thus allowing the researchers and experts for training, evaluating, and benchmark models for emotion identification from facial expressions. The dataset comprises 16000 samples under 6 classes as represented in Table 5. Table 6 implies the sample text under tweet emotions dataset 2.

Table 5: Details on tweet emotions dataset 2

Emotions	No. of Samples
Sadness (0)	4666
Joy (1)	5362
Love (2)	1304
Anger (3)	2159
Fear (4)	1937
Surprise (5)	572
Total Samples	16000

Table 6: Sample texts of tweet emotions datasets 2

S. No	Sample Text	Emotions
1	I feel so inhibited in someone elses kitchen like im painting on someone elses picture	Sadness (0)
2	I have the feeling she was amused and delighted	Joy (1)
3	I am feeling horny so I ask her that lets go home	Love (2)
4	I stopped feeling cold and began feeling hot	Anger (3)
5	I had stated to her the reason I feel so fearful is because I feel unsafe	Fear (4)
6	I want to hold this feeling of shocked awe and wonder forever	Surprise (5)

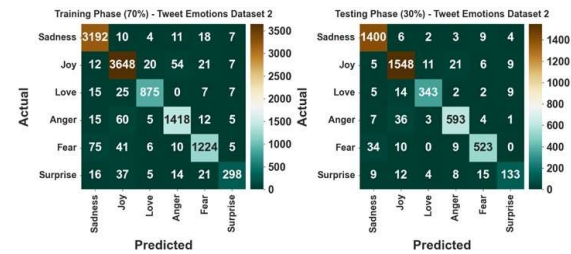


Figure 7: Confusion matrices of MMDLA-SMED method with 70:30 under tweet emotions dataset 2

Fig. 7 demonstrates the confusion matrices created by the MMDLA-SMED method at 70:30 of TRAP/TESP. The simulation outcomes indicate that the MMDLA-SMED technique has efficient classification and identification of all 6 classes correctly.

Table 7 shows the classifier outcomes of MMDLA-SMED method with 70:30 on tweet emotions dataset 2. The experimental values are described that the MMDLA-SMED method suitably identified the samples. As stated by 70% TRAP, the MMDLA-SMED method gains average $accu_r$ of 98.38%, $prec_i$ of 94.28%, $recal_i$ of 91.56%, F_{score} of 92.81%, and MCC of 91.84%, correspondingly. Additionally, with 30% TESP, the MMDLA-SMED system acquires average $accu_r$ of 98.19%, $prec_i$ of 92.94%, $recal_i$ of 90.48%, F_{score} of 91.63%, and MCC of 90.54%, respectively.

Table 7: Classifier outcomes of MMDLA-SMED method with 70:30 on tweet emotions dataset 2

Classes	Accur _y	Preci _n	Recal _l	F1 _{score}	MCC
TRAP (70%)					
Sadness	98.37	96.00	98.46	97.21	96.07
Joy	97.44	95.47	96.97	96.22	94.28
Love	99.16	95.63	94.19	94.90	94.45
Anger	98.34	94.09	93.60	93.85	92.89
Fear	98.07	93.94	89.93	91.89	90.83
Surprise	98.89	90.58	76.21	82.78	82.54
Average	98.38	94.28	91.56	92.81	91.84
TESP (30%)					
Sadness	98.25	95.89	98.31	97.09	95.85
Joy	97.29	95.20	96.75	95.97	93.94
Love	98.92	94.49	91.47	92.95	92.38
Anger	98.04	93.24	92.08	92.66	91.53
Fear	98.15	93.56	90.80	92.16	91.12
Surprise	98.52	85.26	73.48	78.93	78.40
Average	98.19	92.94	90.48	91.63	90.54

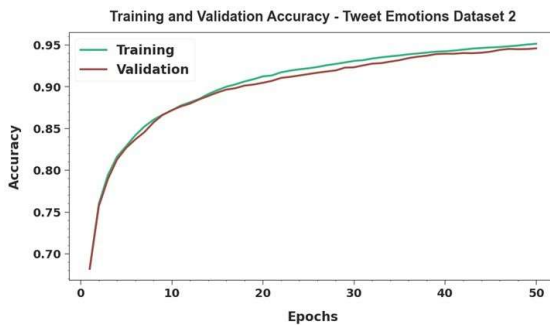


Figure 8: Accur_y curve of MMDLA-SMED approach on tweet emotions dataset 2

Fig. 8 displays the training (TRAIN) and validation (VALID) accuracy curves of the MMDLA-SMED method under tweet emotions dataset 2 over 50 epochs. The TRAIN accuracy is consistently elevated and gained, which can be integrated optimum learning and convergence. The VALID accuracy has been following closely the training tendency with a constant gap, thereby representing the higher generalization and decreased overfitting.

Fig. 9 exemplifies the TRAIN and VALID loss curves of the MMDLA-SMED technique on tweet emotions dataset 2 over 50 epochs. Dual losses are declined exactly in the preliminary epochs and signified the fast learning, and later progressively weakening in an even manner, deliberating efficient convergence. The closer arrangement among training and validation loss, with only a smaller consistent gap, thus implying minor overfitting and improved ability of the generalization.

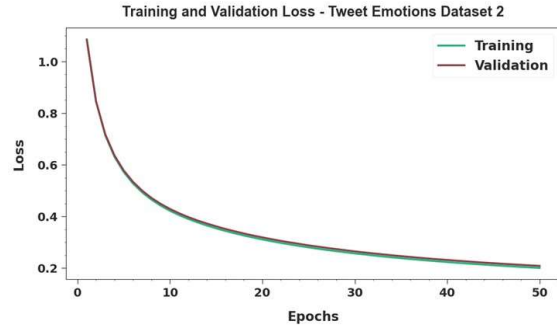


Figure 9: Loss curve of MMDLA-SMED method with tweet emotions dataset 2

Table 8 shows the experimental outcomes of MMDLA-SMED technique with existing models under tweet emotions dataset 2. The experimental outcome value revealed that the 1D-CNN method gets unsuccessful performance with accur_y of 78.13%. Simultaneously, the LSTM, Bi-LSTM, CNN-LSTM, and CNN-BiLSTM models provides somewhat lesser performance with accur_y of 80.62%, 81.07%, 82.85%, and 83.21%, respectively. Furthermore, the MAPLE and DSGN approaches acquire considerable outcomes with accur_y of 93.00% and 90.00%. In conclusion, the proposed MMDLA-SMED system accomplishes excellent performance with accur_y of 98.38%, preci_n of 94.28%, recal_l of 91.56% and F1_{score} of 92.81%.

Table 8: Comparison outcomes of MMDLA-SMED method with other existing model under tweet emotions dataset 2

Algorithms	Accur _y	Preci _n	Recal _l	F1 _{score}
MAPLE	93.00	89.69	90.00	89.80
DSGN	90.00	89.24	89.12	89.22
1D-CNN	78.13	78.12	78.14	76.16
LSTM	80.62	80.62	80.68	80.57
Bi-LSTM	81.07	81.06	81.07	81.03
CNN-LSTM	82.85	82.85	82.81	82.83
CNN-BiLSTM	83.21	83.23	83.21	83.19
MMDLA-SMED	98.38	94.28	91.56	92.81

4.3. Discussion, Critical Analysis, and Limitations

Although the average metrics reported in Tables 3 and 7 indicate strong overall performance, a class-wise reading of the results reveals a more nuanced picture. On tweet emotions dataset 1, the minority classes Boredom (179 samples) and Anger (110 samples) obtain markedly lower recall and F1-score (07.81% and 01.23% recall, respectively, in the training phase) than majority classes such as Neutral and Worry, which each exceed 8,000

samples. A similar, though less severe, pattern appears on tweet emotions dataset 2, where Surprise, the smallest class with 572 samples, records the lowest recall (76.21% in the training phase) among the six emotion categories. This behavior is consistent with the well-documented tendency of soft-voting ensembles, and of deep learning classifiers more generally, to bias predictions toward classes that dominate the training distribution [8]. It indicates that the high overall accuracy of MMDLA-SMED is driven largely by its performance on frequent classes, and that class imbalance, rather than a limitation specific to the improved TCN, ResGRU, or HAN components individually, is the principal factor constraining recall on rare emotions.

The comparison in Tables 4 and 8 shows that MMDLA-SMED consistently exceeds graph-based methods (GCN, GraphSAGE+Pooling, GAAN), tree-based methods (Random Forest, Decision Tree, LightGBM), and sequential deep learning baselines (LSTM, Bi-LSTM, CNN-LSTM, CNN-BiLSTM) on both datasets. Because these baseline results are drawn from the original published works rather than re-executed under an identical experimental protocol in this study, the comparison should be read as indicative of relative standing rather than as a controlled statistical test; a useful direction for follow-up work is to re-implement the strongest baselines under the same data splits and report significance tests, such as a paired t-test or McNemar's test, alongside the accuracy gains. It is also worth noting that the two benchmark datasets differ considerably in size, class count, and label difficulty, so the consistency of the improvement across both settings offers stronger support for the generality of the soft-voting ensemble than either result would alone.

Two further limitations should be acknowledged. First, both benchmark datasets are English-language Twitter corpora, so the reported results do not directly establish how MMDLA-SMED performs on multilingual, code-mixed, or dialectal social media text, despite this being one of the motivations raised in the introduction. Second, combining three deep learning architectures through soft voting increases training and inference cost relative to a single model, which is a practical consideration for deployment on high-volume, real-time social media streams. Extending the evaluation to multilingual and code-mixed data, reporting statistical significance against re-implemented baselines, and exploring lighter-weight or

knowledge-distilled variants of the ensemble are left for future work.

5. CONCLUSION

In this manuscript, the MMDLA-SMED model has been presented for precisely classifying user emotions from social media data by leveraging a data-driven DL technique. Primarily, a text preprocessing pipeline is implemented, including cleaning, hashtag segmentation, slang expansion, emoji conversion, word elongation normalization, and lightweight spelling correction to improve text quality for effective emotion detection. For feature representation, the proposed approach utilized FastText embedding to convert social media text into a meaningful vector representation. Moreover, emotion classification is carried out using a soft-voting ensemble model, incorporating improved TCN, ResGRU, and HAN for enhancing classification performance. The result analysis of the MMDLA-SMED method is conducted on the benchmark datasets, and the outcomes are examined under different aspects. The simulation analysis highlighted the effectiveness of the proposed MMDLA-SMED method, with an accuracy of 98.38% compared to existing models. Thus, the proposed system is found to be effective and reliable for effectual classification of user emotions from social media data.

DECLARATIONS

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Considerations

This study relies exclusively on publicly available, de-identified benchmark tweet-emotion datasets; no new data collection involving human participants was undertaken, and no ethical approval was therefore required.

REFERENCES

- [1] Smirnov, I., Aroyehun, S.T., Plener, P. and Garcia, D., 2026. Automatic Classifiers Underdetect Emotions Expressed by Men. arXiv preprint arXiv:2601.04730.
- [2] Kayal, M., Pati, J., Kumar, S. and Singh, L.K., 2026. Parametrically modified BERT for

- emotion prediction through sentiment analysis. *Discover Computing*, 29(1), p.69.
- [3] Wagay, F.A., 2026. An efficient approach to detecting mental illness on online social media platforms by using multidimensional textual features. *ActaPsychologica*, 262, p.106191.
- [4] Labib, F.H., Elagamy, M. and Saleh, S.N., 2025. Emoberta-x: advanced emotion classifier with multi-head attention and des for multilabel emotion classification. *Big Data and Cognitive Computing*, 9(2), p.48.
- [5] Li, Y., Fan, S. and Wang, H., 2025. Research on Cross-lingual Sentiment Analysis Methods for Social Media Based on Feature Optimization. *Academia Nexus Journal*, 4(2).
- [6] Muhammad, D., Ahmed, I., Naveed, K. and Bendeache, M., 2025. Explainable AI models for decoding emotional subtexts on social media. *Complexity*, 2025(1), p.9258956.
- [7] Joshi, D.B. and Agrawal, P., 2025. Emotion Detection Using Transformer Model WithDeepLearning. Available at SSRN 5198476.
- [8] Ramakrishnan, S. and Babu, L.D., 2025. Improving Multi-Label Emotion Classification on Imbalanced Social Media Data With BERT and Clipped Asymmetric Loss. *IEEE Access*.
- [9] Mane, S.S., Kundu, S. and Sharma, R., 2025. A survey on online aggression: Content detection and behavioral analysis on social media. *Acm computing surveys*, 57(7), pp.1-36.
- [10] Vu, B., Keshri, N., Chandna, S., Jalali, M. and Mehraeen, S., 2025. A systematic approach to fine-tuning transformers for emotion detection on the empathetic dialogues benchmark. *International Journal of Information Technology*, 17(7), pp.3895-3912.
- [11] Alloing, C., Fortant, E., Pierre, J., Richert, F. and Palisser, R., 2025. Knowing your users by heart: A critical examination of the scientific research on emotions conducted by social media platforms. *Social Media+ Society*, 11(3), p.20563051251355456.
- [12] Kusal, S., Patil, S. and Kotecha, K., 2025. Multimodal text-emoji fusion using deep neural networks for text-based emotion detection in online communication. *Journal of Big Data*, 12(1), p.32.
- [13] Berhanu, R., Rout, M., Satapathy, S.C. and Jena, J.J., 2026. Amharic Multi-Class Sentiment Analysis Model: Leveraging Deep Learning for Emotion Detection in Text. *Hyper-Intelligent Networks: Exploring the Future of Connectivity for Society 5.0*, pp.349-364.
- [14] Zhou, S. and Mohd, M., 2025. Mental health safety and depression detection in social media text data: A classification approach based on a deep learning model. *IEEE Access*.
- [15] Khemani, B., Patil, S., Malave, S. and Gupta, J., 2025. Improved graph convolutional network for emotion analysis in social media text. *MethodsX*, 14, p.103325.
- [16] Yu, F. and Liu, J.M., 2026. A Deep Learning-Based Framework for Sentiment and Emotion Classification of Social Media Messages During Pandemic Periods. *Journal of Circuits, Systems and Computers*, 35(03), p.2550413.
- [17] Shwetha, C.H. and Pushpalatha, K., 2026. A hybrid deep learning framework for multilingual depression detection and symptom classification from social media text. *International Journal of Information Technology*, pp.1-24.
- [18] Lee, S.J., Lee, H.T. and Lee, K., 2024, August. Enhancing emotion detection through ChatGPT-augmented text transformation in social media text. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)* (pp. 872-879). *IEEE*
- [19] Reddy, M.M.M., 2026. Context-Aware Sentiment Classification of Social Media Text Using Attention-Based Transformer Models. *American Journal of AI Cyber Computing Management*, 6(1), pp.22-29.
- [20] Al-Tarawneh, M.A., Al-irr, O., Al-Maaitah, K.S., Kanj, H. and Aly, W.H.F., 2024. Enhancing fake news detection with word embedding: A machine learning and deep learning approach. *Computers*, 13(9), p.239.
- [21] Hu, Y., Lyu, L., Wang, N., Zhou, X. and Fang, M., 2023. Application of hybrid improved temporal convolution network model in time series prediction of river water quality. *Scientific Reports*, 13(1), p.11260.
- [22] Krishnan, A., Arunachalam, R., Kosuru, S.D., T, S. and Venugopal, S., 2025. An IoT-based healthcare system using blockchain technology and multiscale stacked Residual-GRU for secure data transmission. *Scientific Reports*, 15(1), p.44851.
- [23] Hou, Z., Fu, Q., Li, W., Wang, Y., Dong, Z., Ye, X., Chen, X. and Zhang, F., 2026. Symmetry-Aware Interpretable Anomaly Alarm Optimization Method for Power Monitoring Systems Based on Hierarchical Attention Deep

- Reinforcement Learning. Symmetry, 18(2), p.216.
- [24] <https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text/data>
- [25] <https://www.kaggle.com/datasets/hiteshyadavx/tweet-emotion-recognition/data>
- [26] Kusal, S., Patil, S., Kotecha, K., Aluvalu, R. and Varadarajan, V., 2021. AI based emotion detection for textual big data: Techniques and contribution. *Big Data and Cognitive Computing*, 5(3), p.43.
- [27] Daniel, G.V., Chandrasekaran, K., Venkatesan, M. and Prabhavathy, P., 2025. A multi-Graph Neural Network attention fusion framework for emotion-aware subgraph anomaly detection in social media fake news propagation. *International Journal of Cognitive Computing in Engineering*.
- [28] Sasidhar, T.T., Premjith, B. and Soman, K.P., 2020. Emotion detection in hinglish (hindi+ english) code-mixed social media text. *Procedia Computer Science*, 171, pp.1346-1352.
- [29] Pang, B. and Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), pp.1-135.
- [30] Ekman, P., 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3-4), pp.169-200.
- [31] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1724-1734.
- [32] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. and Hovy, E., 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp.1480-1489.
- [33] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T., 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, pp.135-146.
- [34] Bai, S., Kolter, J.Z. and Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- [35] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp.4171-4186.