

EmoNet: Deep Attentional Recurrent CNN for X (Formerly Twitter) Emotion Classification

Madasu Sai Nikhita¹, Ravi Kumar Tirandasu^{2*}

¹Master's Student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India – 522502

Email: sainikhitamadasu@gmail.com

²Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India – 522302

Email: ravi.tirandasu@gmail.com, travikumar@kluniversity.in

Phone: 7507976799

ORCID: 0000-0002-8233-7816

***Corresponding Author:** Ravi Kumar Tirandasu, Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India – 522302

Email: travikumar@kluniversity.in

a) ABSTRACT

It's essential to classify emotions found on social media platforms, such as X (Twitter), in order to understand human behavior, public opinion and market dynamics. The Twitter Emotion Dataset with more than 500,000 labelled tweets is a rich yet hard source of data because of its unstructured and noisy nature. To address these challenges, novel DL architectures, fusing convolution, recurrent and attention networks, have been applied to achieve the ability to model local and sequential dependencies in text. Different models such as BERT, RoBERTa, DistilBERT, attention based LSTM-CNN, CNN-GRU, hybrid attention based CNN-LSTM-GRU architectures and BERT with LSTM and GRU layers have been experimented. We studied optimisation strategies such as Adam, RMSprop, Nadam, Adagrad, and SGD. Experimental results suggest that the model of BERT + LSTM + GRU has the maximum performance with 95.3% accuracy, 95.6% precision, 95.3% recall and 95.3% F1-score. Interpretability is obtained by applying XAI approaches (LIME) by focusing on the influential aspects. The model's efficiency is shown on large-scale social media data of high dimensional emotion recognition.

Keywords: Emotion classification, social media analysis, deep learning, convolutional neural networks, recurrent neural networks, attention mechanism, BERT, explainable AI.

How to cite this article: Sai Nikhita M, Tirandasu RK., EmoNet: Deep Attentional Recurrent CNN for X (Formerly Twitter) Emotion Classification. Int J Drug Deliv Technol. 2026;16(86s): 1945-1951; DOI: 10.25258/ijddt.16.86s.176

Source of support: Nil.

Conflict of interest: None

II. INTRODUCTION

Social media has grown to a tremendous amount and has changed fundamental communication in the digital sphere, providing people with a space to convey their ideas, perspectives and emotions as they happen. The X platform (previously Twitter) is one of the most widely used social media platforms in 2023, boasting more than 421 million active users, and has produced many short, informal, and emotionally charged pieces of text. One of the most popular social media platforms in 2023 is X (formerly Twitter), which has a large user base of over 421 million active users and generates a considerable amount of short, informal, and emotionally charged texts. It offers relevant information about public attitude, emerging trends and behaviours among the various populations and events around the world. In this respect, emotion identification is particularly important to understand social dynamics, track psychological issues, foresee political results and analyse customer-brand interactions [2], [3].

Emotions are important to human communication, impacting decisions, social interactions and the wider public debate. With the rapid development of big data, automatic emotion recognition from text data becomes a hot research field, which can help academics and organisations to extract useful information from the huge amount of unstructured data [4]. However, emotion detection in social media is a

challenging task because emotions are not always clearly expressed, context dependent, and can be expressed in a diverse fashion within short messages. These algorithms are based on manually engineered features, and are unable to detect fine-grained nuances and relationships in the data.

DL has achieved tremendous progress in the performance of emotion classification tasks. The architectures that combine CNNs, RNNs and attention mechanisms have demonstrated their ability to learn local semantic patterns and sequential dependencies in text, which is essential for more accurate emotion recognition [7] and [8]. Also, transformer-based models such as BERT and RoBERTa have improved the representation of textual information and the ability to capture the high-dimensional contextual linkages [9]. These are combined with hybrid architectures and optimised by powerful algorithms, providing state-of-the-art performance with large-scale social media data sets [10].

The ultimate goal of this research is to build a strong DL model (EmoNet) based on attentional recurrent CNNs and transformer-based models for the purpose of correctly categorizing emotions on X. The system is designed to effectively detect delicate emotional cues, provide better

explainability via explainable AI techniques and offer scalable solutions for real-time emotion analysis in social media.

III. RELATED WORK

All of the recent progress of emotion detection via text has been attributed to the rise of social media platforms and the development of large scale of emotion annotated datasets. The social media platforms such as X (Twitter) are constantly changing and offer opportunities and challenges for emotion classification. This is because content produced by users is brevity, informal and context-specific. There have been two types of approaches: the rule-based systems and the lexicon-driven approaches. Yet these methods typically do not attend to subtle forms of emotional content in informal content [11]. With the development of ML and NLP, increasingly advanced methods appeared, mixing statistical models with manual feature extraction to increase classification performance [12].

With the advent of DL, emotion detection has been revolutionized, since models can automatically learn hierarchical feature representations directly from the raw text input. RNNs and particularly LSTM networks have been found to be useful for modelling sequential dependencies and context in text, and thus, to be better for capturing emotions in text that span multiple words or sentences [13]. To improve the understanding of the context, the text is read in both directions, by using bidirectional LSTMs [14] and [15]. Attention mechanisms are used to focus the model's attention on specific text segments that are most relevant to sentiment analysis [14,15].

Hybrid networks that combine CNNs and RNN networks has been explored to model short-range and long-range dependencies in text. CNNs are quite good at finding important phrases and n-gram characteristics, however, RNNs keep the temporal information over the sequence [16]. The use of attention layers in these hybrid architectures has been demonstrated to increase performance, by giving more weight to relevant features and boosting both accuracy and interpretability [17]. The issue of several emotions co-occurring or overlapping has been proposed to be addressed by several approaches in brief texts, such as transformer embeddings and clustering or topic modelling [14].

Pre-trained contextual embeddings on huge corpora have set new standards for emotion detection by transformer-based models like BERT, RoBERTa and its derivatives [12]. The models contain large semantic information, and these have been successfully fine-tuned with domain-specific emotion data. To capture the sequential dependency which is not captured completely by transformers, BERT-based models have been combined with downstream models such as LSTM, GRU or Attention layers [13] [17]. Multi-feature fusion strategies that combine handcrafted lexical features with DL representations have been proved to increase the classification performance, showing the complimentary strengths of the classical and neural approaches [18].

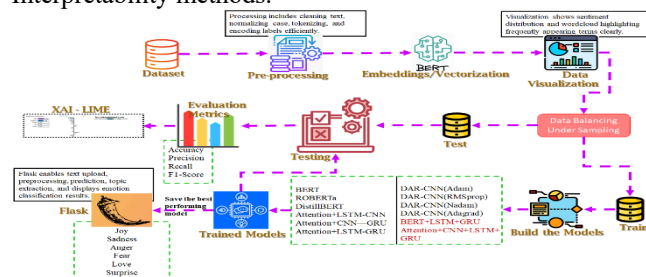
Several studies have demonstrated that multimodal methods are relevant for good recognition of emotions. To enhance the accuracy of detection, it is suggested that the networks be used in conjunction with additional data (topic traits, temporal patterns, even physiological information such as EEG [19, 20]). These techniques address issues with text only systems particularly in disambiguating nuanced emotions or nuanced emotional expression. In addition, the interpretability has been improved by the growing use of

attention processes and XAI techniques, which enable researchers and practitioners to understand which aspects are responsible for model predictions, and hence trust the automated emotion detection systems [17], [18].

Despite these improvements, there are still hurdles in the emotion categorisation on social media data. These challenges include class imbalance, loud or unclear language and developing linguistic patterns such as emojis, hashtags and slang. To overcome these challenges, researchers have explored new optimization techniques, data augmentation and transfer learning approaches to improve the robustness and generalisability of the models in different datasets [11] and [12]. Furthermore, benchmark studies have shown that ensemble or hybrid models incorporating various architectures can often achieve superior performance benefits by leveraging the strengths of each individual model and mitigating their weaknesses [13] and [15].

IV. MATERIALS AND METHODS

The proposed approach aims to build a robust framework for categorizing emotions in X (Twitter) text with advanced DL and transformer based models. The framework employs transformer embeddings of BERT, RoBERTa and DistilBERT to ensure rich contextual representations and fuses with the hybrid attention-based architectures such as Attention-based LSTM-CNN, Attention-based CNN-GRU, and the proposed Attention-based CNN-LSTM-GRU model. This architecture consists of convolutional layers for local features extraction, LSTM layers for modelling long-term dependencies, and GRU layers for fast sequential learning, which guarantees robust identification of complicated emotional patterns [21], [23]. Also, the system utilises different optimisation algorithms (Adam, RMSprop, Nadam, Adagrad, and SGD) in an effort to improve the convergence of training and adapt to noisy social media data [24]. Moreover, we use explainable AI techniques such as LIME to identify relevant textual components, further improving interpretability and providing insights into the model's reasoning [25]. The suggested Emotion Recognition system is: Transformer embeddings, Hybrid DL architectures and Interpretability methods.



“Fig. 1. System Architecture”

The flowchart "Fig. 1" illustrates a ML workflow that involves training and evaluating models, likely for a natural language processing application. The process starts with a “Dataset” that goes through several “Pre-processing” steps such as data processing, vectorisation, visualisation, balancing and feature extraction. The pre-processed data will then be split into “Train” and “Test”. The sets are used to train different “Trained Models” shown in the figure, such as BERT, RoBERTa, DistilBERT, and several attention based architectures that includes LSTM, CNN and GRU with different optimizers. Finally, the Performance Metrics, or metrics, such Accuracy, Precision, Recall and F1-Score are

RESEARCH PAPER

used to evaluate the performance of these models. The results of these measures are represented in the bar graph.

A) Dataset Collection

The data used in this study is Twitter Emotion Dataset. Over 500,000 tweets are annotated with one of the six major emotions: Joy, Sadness, Love, Anger, Fear and Surprise. We preprocessed each tweet by removing URLs, mentions, hashtags, emoticons and punctuation, resulting in clean textual input that is ideal for D models. They were categorized with each emotion's name and then their distribution in the group was analysed and visually presented in a systematic way, which revealed a substantial imbalance of the emotion categories. This dataset contains a wealth of challenging and rich real-world social media data that is unstructured, and is suitable for training and testing the most sophisticated transformer-based emotion recognition models and hybrid DL architectures [29].

Unnamed: 0		text	label
0	0	i just feel really helpless and heavy hearted	4
1	1	ive enjoyed being able to slouch about relax a...	0
2	2	i gave up my internship with the dmrg and am f...	4
3	3	i dont know i feel so lost	0
4	4	i am a kindergarten teacher and i am thoroughl...	4

“Fig.2 Dataset Collection”

B) Pre-Processing

Emotion categorisation is a crucial step in the processing of information from social media sources, since it converts raw data into a tidy and organized structure appropriate for DL models. In this process there are several steps, such as converting to lower case, eliminating HTML tags, URLs and non-printable characters, and selecting whether to keep punctuation based on the model requirements. Preprocessing the data for modelling involves tokenisation and label encoding.

i) *Data Processing*: Images were carefully prepared for model training by loading, resizing and normalising them, to be compatible with different detection architectures. The annotation has been carefully validated and converted to the target formats and the bounding boxes have been verified. In this way, the models learn correct Item position and classes during the training and evaluation. However, when capturing images of crowds in the real world, the scale and angle of the photographs are diverse, and a few mistakes or inconsistencies can occur in the model. Proper preprocessing can help minimize the errors and ensure the consistency and reliability of the model for real-world crowd photographs.

ii) *Embeddings / Vectorization*: Embeddings / Vectorisation: conversion of textual data into dense number representations, which represent semantic and syntactic relations between words. In this system BERT embeddings are used, using a pre-trained transformer to generate contextualised word vectors. Unlike regular static embeddings, the context-dependent meanings of BERT play an important role in conveying nuanced emotions in tweets. These embeddings are fed into hybrid DL architectures which allow the models to recognise complicated patterns and dependencies. Although the vectorised representation does not lose the richness of language, it is a very convenient representation for being used as inputs to the next layers such as Attention-based CNN-LSTM-GRU for accurate emotion categorisation.

iii) *Data Visualization*: Visualising data helps to understand the distribution and properties of the data set and patterns, class imbalances and trends. The method comprises plotting the distribution of each emotion label to have a better understanding of the dataset and to identify over/under-represented classes. We create word clouds for the top most frequent words in each mood, giving important textual information that can affect the classification. Visualisation is not just part of EDA, but also has an impact on preprocessing, feature selection and modelling methods. Visualising data allows investigators to check that the data is well balanced, to spot noise or abnormalities, and to communicate insights effectively to inform model development and evaluation.

iv) *Data Balancing*: Considering the class imbalance issue in emotion-labeled datasets could lead to the result of influentially being biased towards the majority class, which might adversely affect the model performance on the minority class, data balancing is a technique to address such issue. In this context, the under-sampling methods are used to reduce the number of samples of the major classes, and thus to obtain a more balanced distribution of feelings. This approach aims at avoiding overfitting towards the majority classes and maintains a sufficient number of data for learning. Other approaches like over-sampling or synthetic data generation may also be used, depending on the nature of the data. Balancing helps to ensure that the model is trained fairly to capture all the emotions, hence, improve generalisation and make correct predictions from the entire range of sentiment types from social media content.

v) *Feature Extraction*: Transforming the raw text data into meaningful features that capture important patterns for emotion classification is referred to as feature extraction. This approach uses convolutional layers to capture n-gram local features and recurrent layers (e.g., LSTM, GRU) to capture sequential information along the word sequence. Another type of attention mechanism is to highlight words or phrases that are emotionally relevant, which enhances feature representation. BERT embeddings provide a vector representation of the context, and can be used as an informative basis for features. All these retrieved features are combined together to form the semantic, syntactic and contextual information, important for accurate categorisation. A good feature extraction reduces this noise, fine tunes the model training and guarantees that these subtleties, complexities and multiple emotions in social media content can be captured by DL architectures.

C) Train & Test:

The data set is divided into 80% training and 20% testing to determine the model's performance in a solid manner. 80% of the data is used in the training set to train the DL models to identify patterns, dependencies and context of the text. The other 20 % is test set, which the model is tested on its ability to generalise to unseen data. This separation means that a large portion of data is used to train our model, while still having a good amount of independent data to test the model, thereby reducing the chances of overfitting and providing useful information about the efficacy of our model in real world applications.

D) Algorithms

BERT: BERT generates good context embeddings from a bidirectional context. It captures semantic and syntactic nuances and enhances emotion categorisation in social media

RESEARCH PAPER

data. It takes advantage of pre-trained linguistic information to minimize the need for labeled data to improve generalisation, robustness, and the interpretation of the sentiment in the text.

RoBERTa: In RoBERTa, RoBERTa eliminates the task of next sentence prediction and removes the dynamic masking from BERT. Provides high quality contextual embeddings. It does a good job at catching the sarcasm and changes in context. The large-scale pretraining enables the model to have high semantic comprehension and the hybrid models can achieve better accuracy and scalability for emotion recognition.

DistilBERT: DistilBERT is a compact version of BERT that is faster in terms of performance and has a lower memory usage while still maintaining the ability to understand the language. It efficiently produces contextual embeddings, thereby allowing real-time large-scale emotion categorisation with scalability and speed without sacrificing much predictive performance as compared to BERT.

Attention-based LSTM + CNN: A hybrid model which leverages LSTM for sequential feature extraction and CNN for detecting local patterns. Highlighting of attention words gives them special meaning that enhances interpretation and accuracy. It effectively captures subtle and complex emotions from unstructured social media datasets, by modelling contextual dependencies together with n-gram characteristics.

Attention-based CNN + GRU: CNN to learn local patterns of the textual information, GRU to learn the sequential dependencies, attention to learn the importance of emotional tokens. This architecture is interpretable and achieves accuracy on the par with the LSTM based models with lesser computation requirement, which can be advantageous for efficient classification on large size sentiment datasets with subtle expressions from a heterogeneous source.

Attention-based LSTM + GRU: This is the model that has LSTM for long term dependency capturing and GRU for fast sequential processing. Attention mechanisms are designed to attend to tokens of emotional relevance, while reconciling the compromises between accuracy and efficiency. It is good for dealing with big volume unstructured data, finding out subtle context dependant emotions and complicated patterns.

Proposed Attention-based CNN + LSTM + GRU: CNN is used for local features extraction, LSTM is used for long-term dependencies modelling, GRU is used for efficient learning of sequences, and attention is used for influential words highlighting. Together, they model spatial, temporal and sequential patterns, and serve as a powerful, interpretable and scalable emotion recognition on large-scale, complex social media information.

Adam Optimizer: Adam optimiser is a hybrid of momentum with RMSprop, that adapts learning rates for each parameter. It increases the rate of convergence, controls sparse gradients and enhances the generalisation of the model. Its adaptive property guarantees its stability and makes it suitable for training complicated hybrid architectures on large-scale textual emotion datasets.

RMSprop Optimizer: RMSprop adjusts the learning rate using recent magnitudes of the gradients to help stabilise training of deep models. It is suitable for sequential architectures like LSTM and GRU, it can learn temporal

dependencies, it can prevent the oscillation, and it enhances classification accuracy on complicated emotional patterns.

Nadam Optimizer: Nadam is a combination of Adam and Nesterov momentum, which allows for flexible learning rates and speedier convergence. It can quickly process sequential data with sparse data, and enhances the stability of hybrid models during training. The optimiser improves accuracy, generalisation and resilience and boosts the ability to capture subtle emotions in textual data sets.

Adagrad Optimizer: Adagrad adapts the learning rate to the parameters, doing larger updates for infrequent and smaller updates for frequent parameters. It is helpful for emotion classification with the text, as it overcomes the issue of imbalance on word occurrence, offers an effective feature learning from sparse data and enhances the performance accuracy of a hybrid system.

BERT with LSTM + GRU: BERT embeddings can give semantic richness to the model, LSTM can model long-term dependencies, and GRU can enable efficient sequential learning. They complement each other to boost the resilience, scalability and accuracy of emotion categorisation, enabling the efficient analysis of large-scale unstructured text while maintaining a sophisticated understanding of the context.

Attention based CNN + LSTM + GRU Optimiser CNN extracts the local features, LSTM models the long term dependency, GRU is used to process the sequence and attention identifies the important token in the sequence. The model can properly classify emotions on large-scale heterogeneous textual samples in an accurate, interpretable and stable way when controlling the parameters of SGD.

V. EXPERIMENTAL RESULTS

Accuracy: A test that is accurate and correctly differentiates between patient and healthy cases. The accuracy of a test should be defined as the percentage of true positive and true negative of all cases tested. In mathematical notation, this means that:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} (1)$$

Precision: Precision is the fraction of accurately classified instances or samples out of the ones that are classified as positive. Hence the formula for precision is:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$$

Recall: Recall is a ML measure of the ability of the ML algorithm to recover all instances of a particular class. It is the proportion of positive cases correctly predicted to the total number of positive cases. It explains the effectiveness of a model to represent every item in a class.

$$\text{Recall} = \frac{TP}{TP + FN} (3)$$

F1-Score: F1 score is a metric to evaluate ML. It assesses the correctness of a model. It combines the accuracy and sensitivity of the models. Accuracy metric calculates what is the number of times a model produced a valid prediction over the whole data set.

$$\text{F1 Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} * 100 (1)$$

“Table.1 Performance Evaluation”

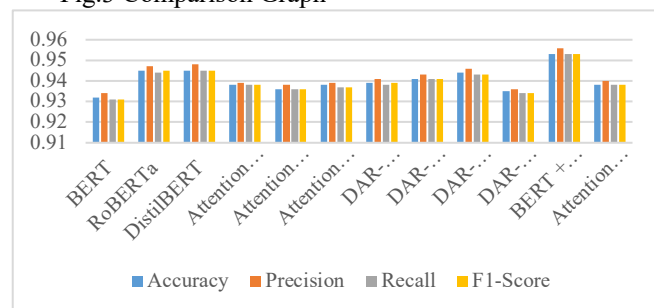
ML Model	Accuracy	Precision	Recall	F1-Score
BERT	0.932	0.934	0.931	0.931
RoBERTa	0.945	0.947	0.944	0.945

RESEARCH PAPER

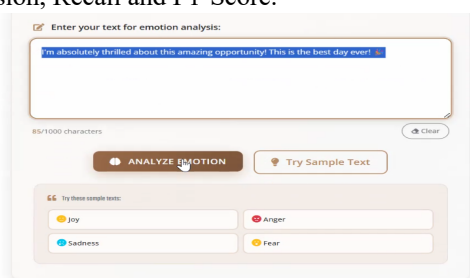
DistilBERT	0.945	0.948	0.945	0.945
Attention + LSTM-CNN	0.938	0.939	0.938	0.938
Attention + CNN-GRU	0.936	0.938	0.936	0.936
Attention + LSTM-GRU	0.938	0.939	0.937	0.937
DAR-CNN (Adam)	0.939	0.941	0.938	0.939
DAR-CNN (RMSprop)	0.941	0.943	0.941	0.941
DAR-CNN (Nadam)	0.944	0.946	0.943	0.943
DAR-CNN (Adagrad)	0.935	0.936	0.934	0.934
BERT + LSTM + GRU	0.953	0.956	0.953	0.953
Attention + CNN + LSTM + GRU	0.938	0.940	0.938	0.938

Several models are compared in terms of their performances in Table.1. The best accuracy and F1-score is obtained by BERT + LSTM + GRU.

“Fig.3 Comparison Graph”

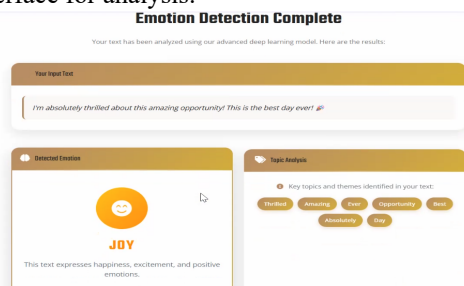


This Fig. 3 bar chart shows the performance comparison of different NLP models in terms of their Accuracy, Precision, Recall and F1-Score.



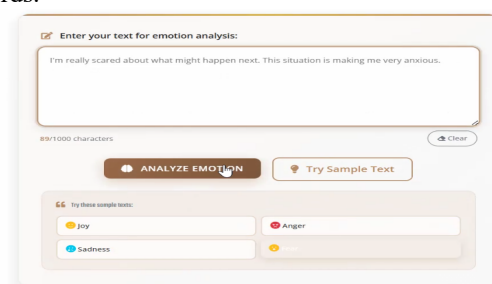
“Fig.4 Upload Input Text”

Figure 4 is a screen shot of an emotion analysis application with a positive example text being entered in the user interface for analysis.



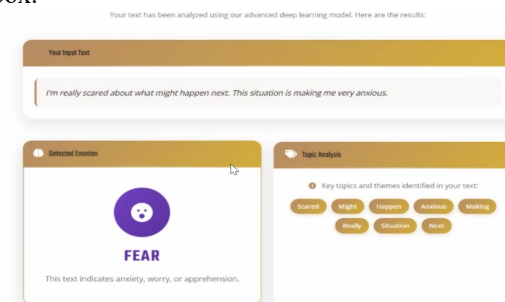
“Fig.5 Predict Result”

The results of an emotion analysis tool are displayed in this, Fig. 5. The emotion detected is JOY and the related keywords.



“Fig.6 Upload another Input Text”

The image, Fig. 6, is a user interface with text input and emotion analysis based on text input anxiety and terror in input box.

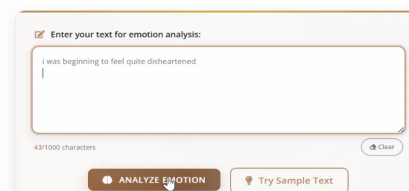


“Fig.7 Final Outcome”

Here's a screen shot (Fig. 7: The results of an emotion analysis, highlighting the emotion detected as “FEAR” and key terms that were extracted from the text input.

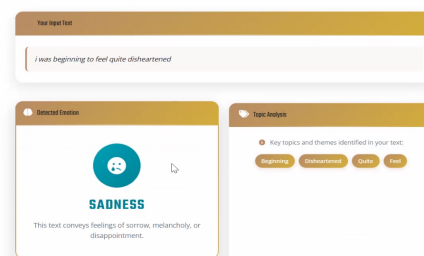
Analyze Your Text for Emotional Content

Enter any text below to discover the emotional undertones using our advanced deep learning model. Perfect for analyzing tweets, comments, or any written content.



“Fig.8 Upload Input Text to Analyze Emotion”

The screenshot from Fig. 1 is an image of the. Figure 8. User interface for analyzing an emotion analysis tool that receives a text for analysis, disheartened emotions.



“Fig.9 Result”

The results of the emotion analysis conducted are displayed in this screenshot (Fig. 9). Detects the sentiment as ‘ SADNESS ’ and the associated words.

VI. CONCLUSION

The BERT + LSTM + GRU achieved the highest classification accuracy of 95.3% for emotions in the large-scale Twitter Emotion Dataset, having a precision of 95.6%, a recall of 95.3%, and an F1-score of 95.3%. The method combines the contextual embeddings of BERT with LSTM

and GRU layers, which capture the sequential modelling ability, to effectively capture the semantic and temporal patterns of unstructured textual data. Its ability to handle noisy and high dimensional social media text data while maintaining good generalisation ability is shown in comparative comparison with the baseline models and other hybrid architectures. The introduction of attention methods provides greater interpretability and reliability of predictions in hybrid architectures by guiding the attention of the model towards influential words. Explainable AI approaches, especially LIME, facilitate discovering the important textual elements that lead to classification decisions, thus providing transparency and useful insights. The assessment demonstrates its efficiency, its ability to scale and its robustness to data complexity, and thus makes it suitable for practical application like sentiment monitoring, behaviour analysis, automatic emotion detection on various social media. The results have shown that the BERT + LSTM + GRU network offers an advanced solution for achieving precise and informative emotion classification applications with high performance.

Going forward, emotion classification can be expanded to multilingual social media datasets to enable the model to be trained on various linguistic and cultural contexts. Using cutting-edge transformer models such as DeBERTa or GPT embeddings can provide improved context-aware understanding and classification results. Investigating semi-supervised or self-supervised learning approaches may help to alleviate the dependency on huge labelled datasets, as well as better tackle the problem of class imbalance. Additionally, multi-modal inputs like photographs, emoticons and videos can result in increased expression of user emotions. SHAP or attention visualisation techniques could be further developed to increase the transparency and trust in predictions. Lastly, lightweight implementations of the model can be developed for edge devices or real-time social media monitoring systems, enhancing its applicability and scalability.

REFERENCES

- [1] Babu, N. R., & Viswanathan, V. (2025). MFENet: A Multi-Feature Extraction Network for Enhanced Emotion Detection Using EEG and STFT. *IEEE Access*.
- [2] Yan, J., Pu, P., & Jiang, L. (2025). Emotion-RGC net: A novel approach for emotion recognition in social media using RoBERTa and Graph Neural Networks. *Plos one*, 20(3), e0318524.
- [3] Anzum, F., & Gavrilova, M. L. (2024, May). EmoBlend Fusion: Leveraging Handcrafted and Deep Features for Emotion Detection. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)* (pp. 1–6). IEEE.
- [4] Hu, G., Zhu, Z., Hershovich, D., Hu, L., Seifi, H., & Xie, J. (2024). Unimeec: Towards unified multimodal emotion recognition and emotion cause. *arXiv preprint arXiv:2404.00403*.
- [5] Huang, X., Xu, J., Zheng, W., Mao, Q., & Dhall, A. (2024). A Survey of Deep Learning for Group-level Emotion Recognition. *arXiv preprint arXiv:2408.15276*.
- [6] Kher, D. (2021). Multi-label emotion classification using machine learning and deep learning methods. Ph.D. dissertation, School Eng. Comput. Sci., Laurentian Univ., Sudbury, ON, Canada.
- [7] Sankari, S., & Priscila, S. S. (2024, May). EmoNet deep learning for emotion classification in social media content analysis. In *Proc. Int. Conf. Adv. Comput., Commun. Appl. Informat. (ACCAI)* (pp. 1–7).
- [8] Kusno, J. W., Setiawan, R. C., Iswanto, I. A., Andangsari, E. W., & Chowanda, A. (2023, December). Deep learning for text based emotion classification from social media. In *Proc. Int. Workshop Artif. Intell. Image Process. (IWAIP)* (pp. 1–4).
- [9] Cosma, O.-G., Gota, D., Ciobotaru, A., Misaros, M., Stan, O., Valean, H., & Miclea, L. (2024, July). End-to-end deep learning based emotion classification solution using ResNet-50 model. In *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)* (pp. 1–6).
- [10] Tanwar, V. (2024, August). Deep learning-based emotion detection on Twitter: A study of six primary emotions. In *Proc. 2nd Int. Conf. Intell. Cyber Phys. Syst. Internet Things (ICoICI)* (pp. 1232–1237).
- [11] Singh, G., Singh, D., Sharma, R., & Bhardwaj, K. (2024, June). Potential approach for text-based emotion detection using NLP coupled with deep learning of sentiment analysis. In *Proc. 15th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)* (pp. 1–5).
- [12] Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artificial Intelligence Review*, 54(8), 5789–5829.
- [13] Kodati, D., & Dasari, C. M. (2024). Negative emotion detection on social media during the peak time of COVID-19 through deep learning with an auto regressive transformer. *Engineering Applications of Artificial Intelligence*, 127, 107361.
- [14] Ding, F., Kang, X., Nishide, S., Guan, Z., & Ren, F. (2020, October). A fusion model for multi-label emotion classification based on BERT and topic clustering. In *Proc. Int. Symp. Artif. Intell. Robot.* (p. 36).
- [15] Suhasini, M., & Badugu, S. (2018). Two step approach for emotion detection on Twitter data. *International Journal of Computer Applications*, 179(53), 12–19.
- [16] Anusha, V., & Sandhya, B. (2015). A learning based emotion classifier with semantic text processing. In *Advances in Intelligent Informatics* (pp. 371–382). Springer.
- [17] Wang, S., Zhu, Y., Gao, W., Cao, M., & Li, M. (2020). Emotion-semantic enhanced bidirectional LSTM with multi-head attention mechanism for microblog sentiment analysis. *Information*, 11(5), 280.
- [18] Anzum, F., & Gavrilova, M. L. (2024, May). Emoblend fusion: leveraging handcrafted and deep features for emotion detection. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)* (pp. 1–6). IEEE.
- [19] Babu, N. R., & Viswanathan, V. (2025). MFENet: A Multi-Feature Extraction Network for Enhanced Emotion Detection Using EEG and STFT. *IEEE Access*.
- [20] Hu, G., Zhu, Z., Hershovich, D., Hu, L., Seifi, H., & Xie, J. (2024). Unimeec: Towards unified multimodal emotion recognition and emotion cause. *arXiv preprint arXiv:2404.00403*.
- [21] Yan, J., Pu, P., & Jiang, L. (2025). Emotion-RGC net: A novel approach for emotion recognition in social media

RESEARCH PAPER

- using RoBERTa and Graph Neural Networks. *Plos one*, 20(3), e0318524.
- [22] Huang, X., Xu, J., Zheng, W., Mao, Q., & Dhall, A. (2024). A Survey of Deep Learning for Group-level Emotion Recognition. *arXiv preprint arXiv:2408.15276*.
- [23] Yan, F., Guo, Z., Ilyasu, A. M., & Hirota, K. (2025). Multi-branch convolutional neural network with cross-attention mechanism for emotion recognition. *Scientific Reports*, 15(1), 3976.
- [24] Mehta, A., & Yang, W. (2023, December). NAC-TCN: temporal convolutional networks with causal dilated neighborhood attention for emotion understanding. In *Proc. 2023 7th Int. Conf. Video Image Process.* (pp. 9–16).
- [25] Sun, L., Xu, M., Lian, Z., Liu, B., Tao, J., Wang, M., & Cheng, Y. (2021). Multimodal emotion recognition and sentiment analysis via attention enhanced recurrent model. In *Proc. 2nd on multimodal sentiment analysis challenge* (pp. 15–20).
- [26] Qazi, M. H. E.-u.-H., AboAlsamh, H., & Ullah, I. (2019). Automatic Emotion Recognition (AER) System based on Two-Level Ensemble of Lightweight Deep CNN Models. *arXiv preprint arXiv:1904.13234*.
- [27] Salloum, S., Tahat, K., Mansoori, A., Alfaisal, R., & Tahat, D. (2024, September). EmoNet: Unveiling Affective States Through Convolutional Neural Networks in Textual Emotion Classification. In *Proc. 2024 Int. Conf. Intelligent Computing, Communication, Networking and Services (ICCNIS)* (pp. 302–305). IEEE.
- [28] Labib, F. H., Elagamy, M., & 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), 48.
- [29] Kastrati, M., Kastrati, Z., Shariq Imran, A., & Biba, M. (2024). Leveraging distant supervision and deep learning for twitter sentiment and emotion classification. *Journal of Intelligent Information Systems*, 62(4), 1045–1070.
- [30] Qian, C., Lobo Marques, J. A., de Alexandria, A. R., & Fong, S. J. (2025). Application of multiple deep learning architectures for emotion classification based on facial expressions. *Sensors*, 25(5), 1478.