

A Study on Deep Learning Techniques for Emotion Classification

Dhanasekaran K¹, Sudhakar K²

¹Post-doctoral researcher, Lincoln University College (LUC), 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia; ²Department of Artificial Intelligence and Data Science, Nitte Meenakshi Institute of Technology (NMIT), Nitte (Deemed-to-be University), Bengaluru, Karnataka, India
jcedhana16@gmail.com

Abstract: Emotion classification has become a major concern in recent years. Because emotion affects both mental health and physical health. Accurate and early detection of mental stress is very important for timely intervention and mental well-being. Emotion data collected from social media has been widely used for emotion classification. Social media platforms like Reddit and Twitter provide text data for emotion research. In recent years, many studies have been carried out on stress classification using social media data. However, there is still a challenge in capturing semantic and aspect-level correlations, utilizing dependencies, and improving efficiency to develop domain-specific emotion detection systems. Unlike traditional machine learning methods, the advanced deep learning techniques based on transformer utilizes self-attention mechanism to effectively capture complex dependencies within social media data. This study presents related works in deep learning techniques for mental health analysis, stress classification and provides insights into challenges, applications, and emotion classification methods.

Keywords: transformer; stress; emotion; attention; deep learning

Introduction

Emotion analysis has become an essential component of healthcare services worldwide. Recently, automatic stress detection through the transformer approach has emerged as a popular framework to provide support for early health care intervention. Mental stress usually affects cognitive performance, emotional stability, and overall health. Automatic stress detection systems have gained much attention due to advances in artificial intelligence and sensor technologies.

In recent years, Transformer-based deep learning techniques have shown significant promise for capturing long-range dependencies and complex patterns in physiological and social media data. This literature review discusses key research studies based on transformer-models for mental stress classification. The rest of the sections of this paper include the literature review and conclusion.

Related work

One of the topic modelling methods discussed by Adesokan et al. [1] aimed to detect and extract the themes and trends from tweets, for identifying hate speech. Stress impacts health conditions a lot, and the way individuals behave in a particular situation. Kumar et al. [2] presented a hybrid deep learning

model that combines RoBERTa, adapter layers, BiLSTM and addressed the class imbalance problem through the focal loss optimization.

Digital learning platforms provide learners with opportunities to learn flexibly and have enabled the delivery of content to a larger audience, but the challenge lies in detecting emotions to improve learner engagement. Transfer learning improves the performance of hierarchical BERT-based classification on low-frequency emotions [3].

In the digital era, Twitter has become a popular platform for individuals to search for information. The sentiment analysis using BERT has shown better performance on emotion analysis from tweets [4].

Social media plays a very important role in tracking disease conditions. Mental health mention poses a challenge due to nuanced texts. One of the existing works on health mention classification utilized affective knowledge injection, task fine-tuning, and a feature fusion strategy for health mention prediction [5].

Suhartono et al. discussed transformer-based models such as BERT and RoBERTa for stress classification using Twitter data. Their approach applied preprocessing on social media posts and pretrained transformer embeddings to detect stress patterns in texts. Experimental results showed BERT with an accuracy of 84.8% and an F1-score of 0.847. It indicates that transformer models can effectively analyze contextual semantic features in text-based stress detection tasks [6].

Lin et al. presented an ensemble approach that uses multiple pretrained transformer models to detect self-reported chronic stress from Twitter data. The study fine-tuned several transformer models, which include BERT, RoBERTa, ALBERT, and XLNet, and merged their predictions using a majority-vote ensemble strategy. This approach improved robustness and classification results compared with other models. Experimental evaluation on the SMM4H shared task dataset achieved an F1-score of 0.73 for stress classification. It demonstrates that ensemble transformer models can effectively capture linguistic patterns related to stress in social media text [7].

Depression affects mental health severely, which is leading to death majorly among people in the age group of 10-24. To address this issue, emotion analysis in Spanish has been explored to compare emotions related to mental health disorders [8]. Table 1 shows the analysis of the performance of the existing approaches.

Table 1. Analysis of results

Dataset	Reference	Year	Method	F1
GoEmotions	[1]	2023	BERT	0.50
			RoBERTa	0.62
			BERT	0.24
			RoBERTa	0.21
			BERT	0.40
			RoBERTa	0.51
	[2]	2025	RoBERTa with Adapter-based Mental Health Analyzer (RAMHA)	0.87
			RoBERTa	0.79
			RoBERTa+BiLSTM	0.80

Dataset prepared by authors	[8]	2024	Encoder-only MarIA (Ensemble model)	60.4
			Spanish-GPT-2 (encoder-decoder)	77.2

Kancharapu and Ayyagari investigated transformer architectures such as DistilBERT, ALBERT, and hybrid BERT-CNN models for analyzing mental health signals from Twitter posts. The study collected and preprocessed tweets using sentiment analysis and keyword-based filtering techniques before feeding them into transformer models for classification [9].

Another study presented lightweight deep learning architectures, utilizing BERT embeddings to identify stress expressions in social media posts. The research compared multiple deep learning models, such as CNN, BiLSTM, and BiGRU, and integrated them with transformer-based representations. Experimental results showed that BERT-ELECTRA based models showed approximately 85.67% accuracy on Reddit datasets. But the CNN-based architectures achieved better accuracy on large Twitter datasets. The results demonstrate that transformer models can effectively capture contextual information in social media text, which can enable accurate stress classification [10].

The existing emotion detection of many robots is not intelligent enough and lacks an understanding of the semantic representation. Those methods rely on keyword search and prior knowledge.

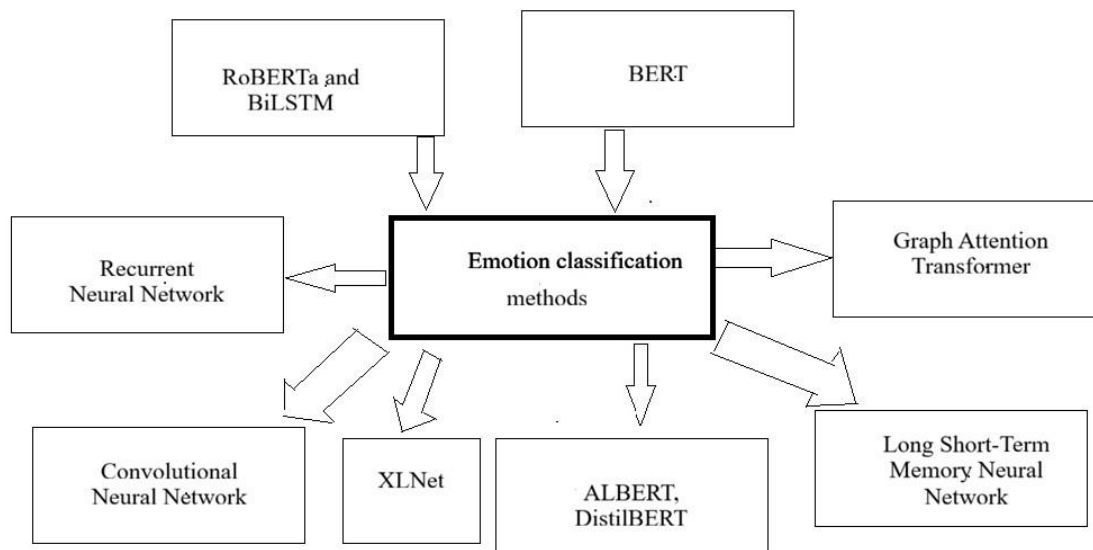


Figure 1. Emotion classification methods

Although many recent studies worked with the GoEmotions dataset, there is still a challenge in capturing fine-grained emotion details, semantic and aspect-level correlation, to enable the development of domain-specific mental health risk prediction systems.

Conclusions

Detecting emotion is a major mental health challenge globally. In this literature review, several recent research works were studied to analyze emotion classification techniques using deep learning methods. The previous mental health prediction methods have shown that individuals with mental health problems show negative emotions, which cause serious health issues. This research presented the literature review to understand the challenges, applications, and deep learning approaches of emotion analysis and stress detection.

References

1. Adesokan A, Madria S, Nguyen L, HatEmoTweet: low-level emotion classifications and spatiotemporal trends of hate and offensive COVID-19 tweets, *Social network analysis and mining*, Vol. 13, issue 136, pp. 1-31, 2023.
2. Kumar M, Khan L, Choi A, RAMHA: A hybrid social text-based transformer with adapter for mental health emotion classification, *mathematics*, Vol. 13, 2918, pp.1-26, 2025.
3. Abd-Ellatif L.A, Shamaki A.M, Hierarchical BERT for emotion detection and engagement anomaly analysis in online learning forums, *IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Padang, Indonesia, pp. 1-8, Dec 2025.
4. Sai kumar T. S, Arunaggi Pandian K, Thabasum Aara S, A reliable technique for sentiment analysis on Tweets via Machine Learning and BERT, *Asian conference on innovation in technology (ASIANCON)*, Pune, August 28-29, pp. 1-5, 2021.
5. Aduragba O.T, Yu J, Cristea A.I, Incorporating emotions into health mention classification task on social media, *IEEE International conference on Big Data*, pp. 4834-4842, 2023.
6. D. Suhartono, I. F. Saputra, A. R. Pratama, and G. Nathaniel, "Psychological stress detection using transformer-based models," *ComTech*, vol. 15, no. 1, pp. 65–71, 2024.
7. T.-M. Lin, C.-Y. Chen, Y.-W. Tzeng, and L.-H. Lee, "Classification of self-reported chronic stress on Twitter using ensemble pre-trained transformer models," in *Proc. 7th Workshop Social Media Mining for Health Applications (SMM4H)*, Gyeongju, South Korea, 2022, pp. 62–64.
8. Salmeron-Rios A, Garcia-Diaz J.A, Pan R, Valencia-Garcia R, Fine grain emotion analysis in Spanish using linguistic features and transformers, *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.1992, pp. 1-29, 2024.
9. R. Kancharapu and S. N. Ayyagari, "Depression detection: Unveiling mental health insights with Twitter data and BERT models," *International Journal of Education and Management Engineering*, vol. 14, no. 4, pp. 1–14, 2024.

10. M. Al-Fahad et al., "Lightweight advanced deep-learning models for stress detection on social media," *Engineering Applications of Artificial Intelligence*, vol. 140, 2025.