Proceedings of CONF-CIAP 2026 Symposium: Applied Mathematics and Statistics
DOI: 10.54254/2753-8818/2026.CH30481

Prediction of Mental Health Risk Levels for Social Media Users Based on Multi-head Attention Mechanism Optimization of Bidirectional Gated Recurrent Unit Networks

Qihao Xu[1*], Chang Liu[1], Zehui Li[1], Yifan Jia[1], Yifan Guo[1], Pengkai Wang[2]

[1]*School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, China*
[2]*IFLYTEK CO.LTD, Hefei, China*
**Corresponding Author. Email: xuqihao8101@163.com*

Abstract. In the field of social media mental health monitoring research, machine learning algorithms serve as the core support, establishing a crucial link between multi-source data and mental health status assessment. In view of the inherent flaws of traditional models, this paper constructs a mental health risk prediction model based on the Bidirectional gated Recurrent Unit (BiGRU) network. The research first conducts data distribution analysis and correlation analysis on the dataset, and then introduces multiple machine learning algorithms for classification and comparison. The results show that the proposed model demonstrates a comprehensive leading advantage in all evaluation indicators. Its accuracy rate, recall rate, precision rate and F1 value all reach 95.4%. Not only is it significantly higher than the 82.9%, 82.9%, 83.5%, 82.8% of Naive Bayes, 89.9%, 89.9%, 90.4%, 89.9% of ExtraTrees, and 87.7%, 87.7%, 87.9%, 87.7% of gradient boosting decision trees, It far exceeds support vector machines whose core classification indicators are only in the range of 54.3% to 57.9%, and also has a slight improvement compared to the suboptimal CatBoost. In terms of the AUC metric reflecting the ability to distinguish categories, this model achieved a high score of 99.6%, approaching the perfect classification level. It is slightly better than CatBoost's 99.2% and 99.0%, and significantly outperforms other comparison models, demonstrating superior classification accuracy and category discrimination ability. This model provides an efficient solution for the precise identification of mental health risks on social media, which is of great significance for promoting the practical development of mental health monitoring technology.

Keywords: Multi-head attention mechanism, Bidirectional gated circulation unit, Mental health risk level

1. Introduction

As social media is deeply integrated into People's Daily lives, the text, images, interaction records and other data left by users on the platforms have gradually become an important source for understanding mental health status [1]. In the current field of mental health monitoring, although relevant research has begun to be conducted by leveraging social media data, most of it is limited to

DOI: 10.54254/2753-8818/2026.CH30481

text data from a single platform, failing to fully consider the completeness of users' expressions in different scenarios, and also ignoring the complementary value among multimodal information such as text, visual, and interactive. This simplistic approach to data collection and analysis makes it difficult to comprehensively and truly restore the dynamic characteristics of users' psychological states over time, resulting in often biased mental health assessment results based on this and failing to provide precise support for subsequent interventions [2]. In the current context where the demand for low-cost and non-invasive mental health screening tools in communities, schools and other scenarios is increasingly urgent, the existing research data and analysis models are no longer sufficient to meet the needs of active early warning in practical applications. There is an urgent need for more comprehensive datasets and more efficient analysis methods to break through the bottleneck [3].

In mental health monitoring research, machine learning algorithms play a core role and serve as a key bridge connecting multi-source data with mental health status assessment [4]. On the one hand, machine learning algorithms can efficiently process massive amounts of social media data, including data cleaning and feature extraction, to mine potential information related to mental health from the disordered raw data. On the other hand, by building predictive models, machine learning algorithms can predict the mental health risk level of users based on the extracted features, providing a quantitative basis for the early identification of potential psychological problems. Although traditional machine learning models have achieved the above-mentioned functions to a certain extent, they often encounter problems such as insufficient data fusion and inaccurate capture of key information when dealing with multi-dimensional data, which leads to prediction accuracy being difficult to meet actual needs and limits their further application in the field of mental health monitoring [5].

To address the shortcomings of traditional models, this paper proposes a mental health risk prediction model based on the Bidirectional gated Recurrent Unit (BiGRU) network. The bidirectional structure of this model can fully capture the contextual semantic associations of social media texts and avoid semantic understanding deviations caused by one-way information transmission. Meanwhile, the gating mechanism can effectively filter out key time series information and eliminate the interference of redundant data on the model's prediction results. On this basis, the model innovatively introduces a multi-head attention mechanism, using multiple parallel attention heads to separate different feature data.

2. Data from data analysis

The dataset was collected from cross-platform user data of mainstream short-video platforms and the instant messaging tool Moments, containing dynamic records of 4,756 users, including 12 independent variables and 1 predictor variable. Independent variables include basic user attributes (age, gender, occupation), text features (length of posted content, frequency of negative emotion keywords, semantic coherence), visual features (proportion of emoji types, color tone tendency of pictures, complexity of picture elements), and interaction features (interaction frequency, number of interaction objects, emotional tendency of comments, content Posting time period). The predictor variable is the user's mental health risk level evaluated by professional scales, which is divided into three levels: low, medium and high. Statistical analysis was conducted on the data to obtain the mean, variance and median of each variable, as shown in Table 1.

DOI: 10.54254/2753-8818/2026.CH30481

Table 1. Some of the data

Variable	Median	Mean	Variance
Age	42	41.814971	268.727481
Content_Length	256	253.830109	19957.48575
Negative_Keyword_Frequency	1	1.038267	2.062468
Semantic_Coherence	0.7	0.698455	0.021945
Emoji_Ratio	0.3	0.307368	0.036153
Visual_Complexity	0.5	0.499407	0.039359
Interaction_Frequency	11	12.260934	82.044413
Number_of_Interaction_Objects	3	3.807611	10.826701
Comment_Sentiment	0.2	0.198242	0.151912

Correlation analysis was conducted on each variable, and a correlation heat map was drawn, as shown in Figure 1.

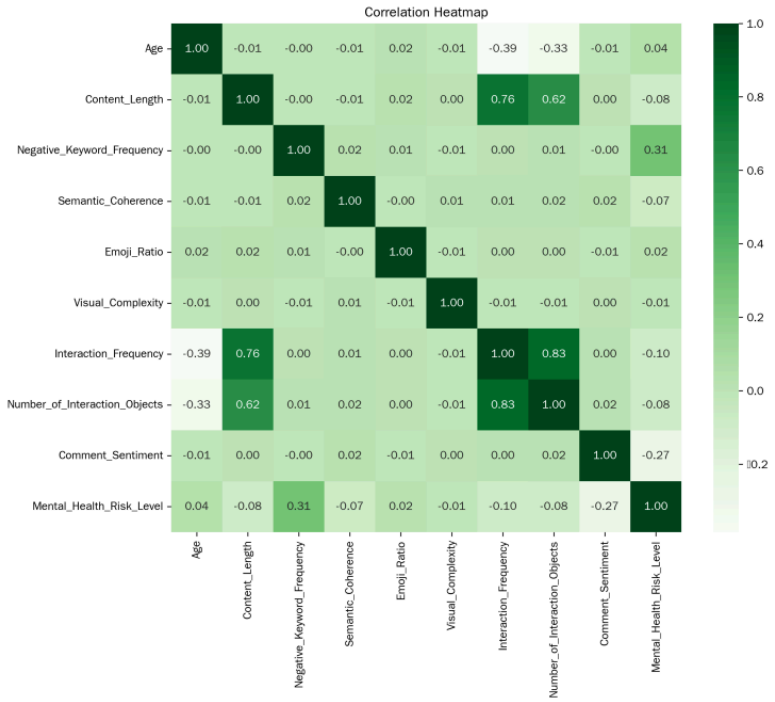

Figure 1. Correlation heat map

Draw the violin plots of each variable and observe the distribution of the data, as shown in Figure 2.

DOI: 10.54254/2753-8818/2026.CH30481

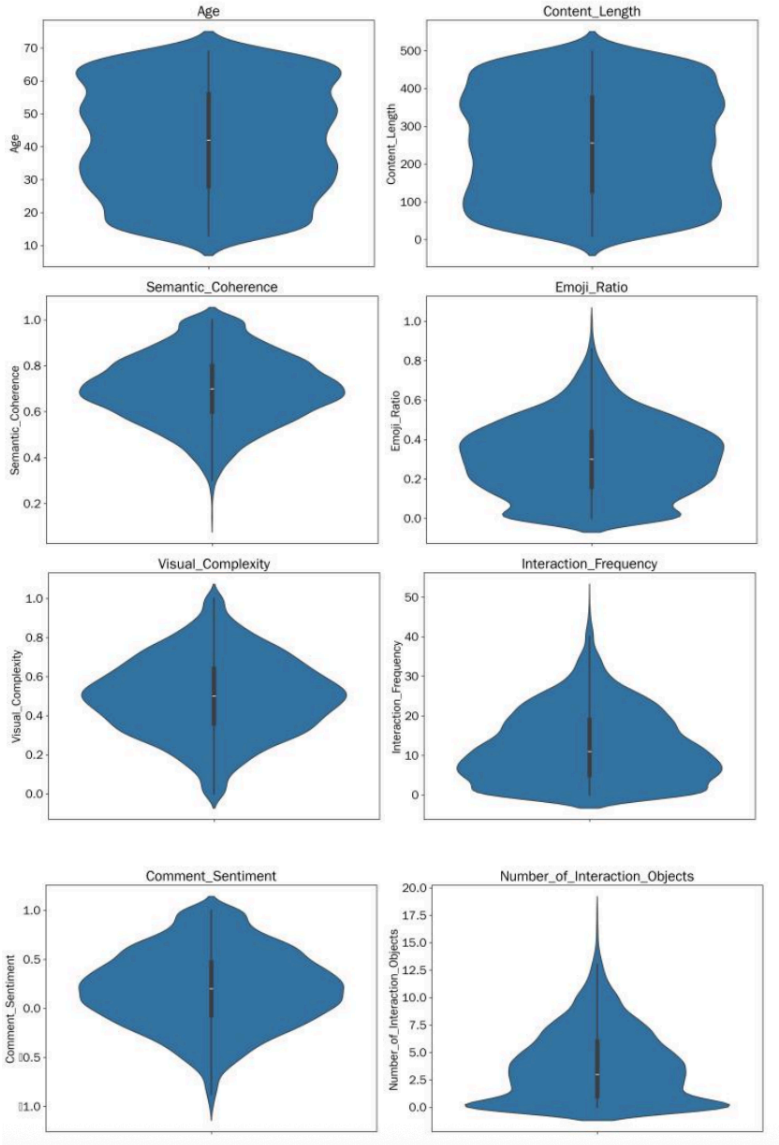

Figure 2. Violin plots of key indicators

3. Method

3.1. BiGRU

BiGRU (Bidirectional Gated Recurrent Unit) is an improved timing model based on GRU (Gated Recurrent Unit), with the core being to capture the context information of the sequence through a bidirectional structure. GRU controls the transmission of information through update gates and reset gates. The update gate determines the proportion of historical information to be retained, while the reset gate decides whether to ignore historical information and focus on the current input, thereby solving the vanishing gradient problem of traditional RNNS [6]. BiGRU consists of a forward GRU and a backward GRU. The forward GRU calculates backward from the starting point of the sequence, capturing the dependency relationship of future directions. Backward GRU calculates from the end of the sequence forward to capture the dependencies of past directions. Finally, the hidden states of the two directions are concatenated by dimension to obtain a feature representation

DOI: 10.54254/2753-8818/2026.CH30481

that simultaneously contains the information of the context [7]. The network structure of BiGRU is shown in Figure 3.

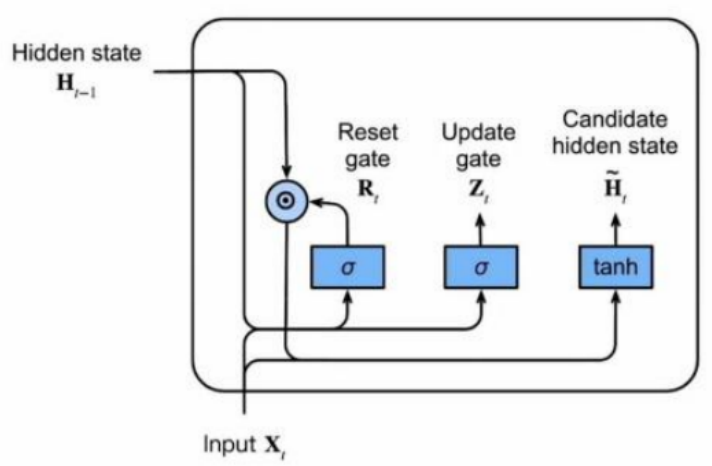

Figure 3. The network structure of BiGRU

3.2. The multi-head attention mechanism

The multi-head attention mechanism is an extension of the attention mechanism. Its core is to capture the correlation information of different dimensions in the sequence through parallel computing of multiple independent attention heads [8]. The traditional attention mechanism obtains the attention weight by calculating the similarity of the query (Q), key (K), and value (V), and then sums the weighted sum V to obtain the attention output. It can only capture the dependency from a single Angle. The network structure of the multi-head attention mechanism is shown in Figure 4.

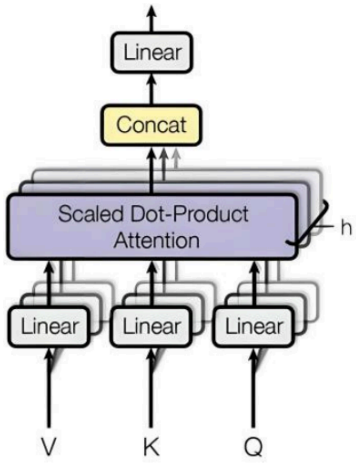

Figure 4. The network structure of the multi-head attention mechanism

Multi-head attention projects Q, K, and V into multiple subspaces through different linear transformations. Each subspace corresponds to an attention head, and the attention weights and outputs are independently calculated. Then, all the outputs of the attention heads are concatenated by dimension and integrated into the final result through linear transformation [9].

DOI: 10.54254/2753-8818/2026.CH30481

3.3. BiGRU with multi-head attention optimization

The BiGRU classification algorithm optimized by multi-head attention is a hybrid model that combines the advantages of both. The core is to enhance the feature expression ability of BiGRU with multi-head attention and improve the classification accuracy [10]. The algorithm process is divided into three steps: Firstly, BiGRU is used to process the input sequence, and the temporal features containing the context are initially extracted by using the bidirectional structure to solve the gradient problem of traditional RNNS; Then, the multi-head attention mechanism is introduced. The hidden states output by BiGRU are simultaneously regarded as Q, K, and V. Through parallel computing of multiple sets of attention heads, the correlation of key features in the sequence is focused, and redundant information is filtered out. Finally, the output of the multi-head attention is sent to the fully connected layer, and the classification result is obtained through the activation function.

4. Result

In terms of parameter Settings for this project, the proportion of the training set to the dataset is 0.7. The maximum number of iterations is 150, the batch size is 128, the initial learning rate is 0.001, and segmented learning rate scheduling is adopted. The learning rate decline factor is 0.1. After 1200 training sessions, the learning rate is updated. At each epoch, the dataset is shuffled and a training progress curve is plotted. The LSTM layer is set with 10 units and outputs the result of the last time step. The dimension of the self-attention layer is consistent with the feature dimension, the dimension of the input layer is the feature dimension, and the dimension of the fully connected layer is the number of categories. Subsequently, the softmax layer and the classification layer are connected.

Naive Bayes, CatBoost, SVM, ExtraTrees, CatBoost and GBDT models were respectively used as comparative experiments, and the experimental results are shown in Table 2. The ROC curve during the model training process is shown in Figure 5.

Table 2. The results of each comparative experiment

Model	Accuracy	Recall	Precision	F1	AUC
Naive Bayes	0.829	0.829	0.835	0.828	0.95
CatBoost	0.934	0.934	0.936	0.934	0.992
SVM	0.543	0.543	0.579	0.548	0.675
ExtraTrees	0.899	0.899	0.904	0.899	0.98
CatBoost	0.933	0.933	0.935	0.933	0.99
GBDT	0.877	0.877	0.879	0.877	0.973
Our model	0.954	0.954	0.954	0.954	0.996

DOI: 10.54254/2753-8818/2026.CH30481

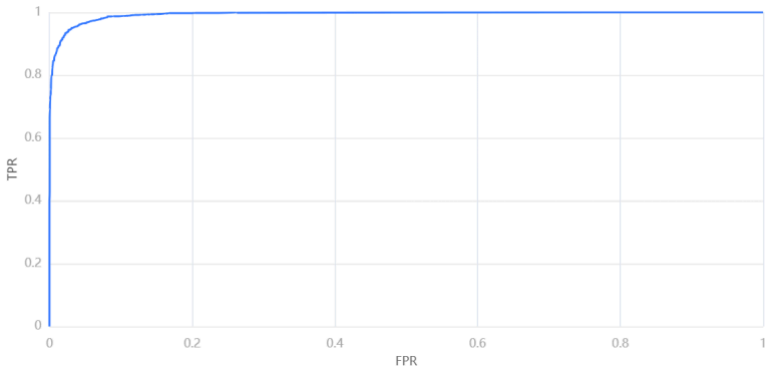

Figure 5. The ROC curve

The bar charts comparing the indicators of each comparison model are shown in Figure 6. From the perspective of overall classification performance, Our model shows a comprehensive leading advantage in all evaluation indicators. Its accuracy rate, recall rate, precision rate and F1 value all reach 95.4%, which is not only significantly higher than the 82.9%, 82.9%, 83.5% and 82.8% of Naive Bayes, The 89.9%, 89.9%, 90.4%, 89.9% of ExtraTrees and 87.7%, 87.7%, 87.9%, 87.7% of GBDT are far superior to the worst-performing SVM, whose core classification index is only between 54.3% and 57.9%. It has a huge gap compared with Our model, and at the same time, it has a slight improvement compared with the suboptimal CatBoost. In terms of the AUC metric reflecting the category discrimination ability of the model, Our model achieved a high score of 99.6%, approaching the perfect classification level, slightly better than CatBoost's 99.2% and 99.0%. It is significantly ahead of 98.0% of ExtraTrees, 97.3% of GBDT, 95.0% of Naive Bayes and 67.5% of SVM. This means that Our model can not only complete the classification task more accurately, but also has a better ability to distinguish different categories. The overall performance comprehensively outperforms other comparison models.

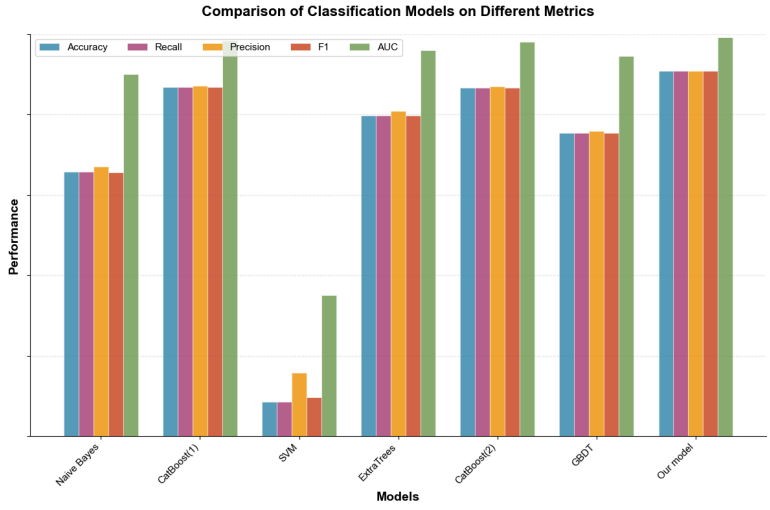

Figure 6. The bar charts comparing the indicators of each comparison model

Output the confusion matrix of the test set of Our model, as shown in Figure 7.

DOI: 10.54254/2753-8818/2026.CH30481

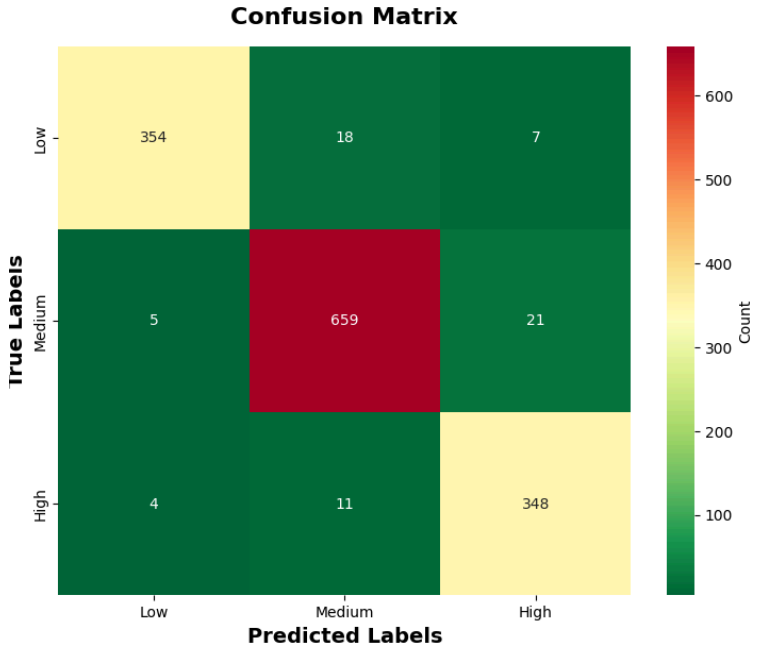

Figure 7. The confusion matrix of the test set of our model

5. Conclusion

In the field of social media mental health monitoring, machine learning algorithms play an indispensable core role and become a key link connecting multi-source data with mental health status assessment. To make up for the performance shortcomings of traditional models, this paper constructs a mental health risk prediction model with the bidirectional gated recurrent unit (BiGRU) network as the core. During the research process, the analysis of data distribution characteristics and the exploration of variable correlations were first carried out, and then multiple mainstream machine learning algorithms were introduced for classification experiments. The experimental results show that the proposed model (Our model) demonstrates all-round performance advantages. Its accuracy rate, recall rate, precision rate and F1 value all reach 95.4%, which is not only significantly higher than the 82.9%, 82.9%, 83.5% and 82.8% of Naive Bayes, ExtraTrees' 89.9%, 89.9%, 90.4%, 89.9% and GBDT's 87.7%, 87.7%, 87.9%, 87.7% far exceed the bottom-performing SVM - whose core classification index is only in the range of 54.3% to 57.9%, with a significant gap between the two. At the same time, it has also achieved a slight improvement compared to the suboptimal CatBoost. In terms of the AUC metric for measuring the ability to distinguish categories, this model achieved a high score of 99.6%, approaching the perfect classification level, slightly better than CatBoost's 99.2% and 99.0%. It is significantly ahead of 98.0% of ExtraTrees, 97.3% of GBDT, 95.0% of Naive Bayes and 67.5% of SVM. This fully proves that it not only has precise classification ability, but also has excellent category discrimination performance. Its overall performance is significantly better than other comparison models. It provides efficient technical support for the early screening and intervention of mental health risks in social media scenarios, helping to enhance the accuracy and timeliness of mental health services.

6. Summary and future work

This paper focuses on the prediction of mental health risk levels for social media users. In response to the problems of insufficient data fusion and inaccurate capture of key information in traditional

DOI: 10.54254/2753-8818/2026.CH30481

models, a BiGRU prediction model integrating a multi-head attention mechanism is proposed. The research utilized cross-platform multimodal data, covering 12 types of independent variables and mental health risk level predictor variables from 4,756 users. The modeling basis was optimized through data distribution and correlation analysis. The experiments compared multiple mainstream algorithms. The core indicators of the proposed model, such as accuracy and recall rate, all reached 95.4%, and the AUC value was 99.6%, approaching perfect classification. It comprehensively outperformed the comparison models, providing an efficient solution for the precise identification of mental health risks and promoting the practical development of monitoring technology.

In the future, the sample size and coverage of social media platforms can be further expanded, and more modal data such as voice can be incorporated to enrich the information dimension. Optimize the model structure and parameter scheduling strategy to enhance the deployment efficiency on mobile devices and other terminals; Conduct longitudinal tracking research based on actual application scenarios to explore the adaptability of the model in long-term mental health intervention.

References

[1] Salvi, Utkarsh , et al. "Decoding Mental Health: Leveraging Machine Learning for Social Media Analysis." #i{Doctoral Symposium on Human Centered Computing} Springer, Singapore, 2025.

[2] Purohit, Shardha , et al. "Analyzing the Impact of Social Media Usage on Mental Health: A Machine Learning Approach." #i{2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)} (2024): 1-6.

[3] Azzolina, Danila , et al. "Editorial: Machine learning approaches for monitoring mental health and substance abuse using social media data." #i{Frontiers in Public Health} (2025).

[4] Pachava, Vengalarao , et al. "Machine Learning Analysis of Social Media's Impact on Mental Health of Indian Youth." #i{International Research Journal of Multidisciplinary Scope} 5.2(2024): 623-635.

[5] Kim, Jina. , D. Lee , and E. Park . "Authors' Reply to: Bibliometric Studies and the Discipline of Social Media Mental Health Research Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study"." #i{Journal of medical Internet research} 23.6(2021): e29549.

[6] Resnik, Philip , et al. "Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study" (Preprint)." #i{Journal of medical Internet research} (2021).

[7] Gao, Yan. "Research on the Application of Chinese Language Teaching System based on Artificial Intelligence Technology." 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE). IEEE, 2024.

[8] Sharan, Roneel V. , et al. "Macro-Sleep Staging With ECG-Derived Instantaneous Heart Rate and Respiration Signals and Multi-Input 1-D CNN& #x2013; BiGRU." #i{IEEE Transactions on Instrumentation and Measurement} (2024): 73.

[9] Tang, Yikai , et al. "Degradation Prediction of Proton Exchange Membrane Fuel Cell Based on Multi-Head Attention Neural Network and Transformer Model." #i{Energies (19961073)} 18.12(2025).

[10] Wang, Yuxiao , C. Suo , and Y. Zhao . "Multi-head attention-based variational autoencoders ensemble for remaining useful life prediction of aero-engines." #i{IOP Publishing Ltd} (2024).