Education Insights

Article

Sentiment Analysis of Mental Health among Chinese College Students Using Hybrid Modeling

Xieyu Chen [1,2,*], Yumei Deng [1,2], Na Tang [1,2], Xi Xiong [2,3], Weiping Deng [1,2] and Tongyu Wu [1,2]

[1] School of Business, Geely University of China, Chengdu, 610100, China
[2] Sichuan Provincial Research Center for Mental Health Education, Chengdu, 611756, China
[3] School of Cybersecurity, Chengdu University of Information Technology, Chengdu, 610225, China
* Correspondence: Xieyu Chen, School of Business, Geely University of China, Chengdu, 611756, China; Sichuan Mental Health Education Research Center, Chengdu, 611756, China

Abstract: As mental health issues among Chinese college students have evolved into a significant social concern within the higher education domain, this study addresses the urgent need for mental health monitoring by proposing a sentiment analysis technical solution that combines deep learning with traditional machine learning. Using "college students" as the scenario keyword, 4,201 relevant posts were collected from the Sina Weibo platform between 2023 and 2025. After data cleaning, word segmentation, and annotation, a tri-classification dataset containing 3,950 effectively labeled entries was constructed. The study first employed the TF-IDF method to extract text features, revealing that academic stress-related vocabulary had the highest weights, reflecting that academic burden is the primary psychological stressor. In model evaluation, traditional machine learning performed best with Random Forest (accuracy: 0.792), while the SVM model exhibited overfitting. In contrast, the hybrid deep learning model CNN-BiLSTM-Attention demonstrated comprehensive advantages (accuracy: 0.813, F1-score: 0.829), particularly excelling in identifying neutral and negative sentiments. Its training and testing losses were also significantly lower than those of other machine learning models. Therefore, in complex contexts, deep learning models achieve higher recognition accuracy than traditional machine learning models. Finally, the study provides a technical solution for college mental health monitoring that balances timeliness and accuracy, aiming to offer precise, real-time, and compliant support for university psychological service systems.

Keywords: college students; mental health; sentiment analysis; hybrid modeling

Published: 08 September 2025

1. Introduction

In recent years, mental health issues among college students in China have shown a high incidence trend, becoming a significant social concern affecting the quality of higher education. According to the 2022 Report on the Mental Health Status of Chinese College Students released by the Chinese Academy of Sciences, over 35% of students experience psychological distress such as anxiety and depression, with the incidence of severe psychological crises increasing by 62% compared to five years ago. The COVID-19 pandemic exacerbated the concealment of psychological issues due to social isolation and online teaching models. Traditional psychological assessment methods (e.g., scale screenings, face-to-face counseling) revealed critical limitations, including delayed response times and insufficient coverage [1]. China faces a critical shortage of professional mental health workers in universities, with only 8.6 professionals per 100,000 students, which is significantly below the global median of 13 [2]. This systemic gap has led to an unconventional reliance on academic counselors as frontline mental health monitors and pseudo-counselors. However, these counselors simultaneously shoulder academic advising, career guidance, and disciplinary roles, creating a conflict of interest that deters students from disclosing psychological distress [2,3]. The lack of dynamic monitoring further contributes to missed early intervention opportunities. Current empirical research on college students' mental health primarily relies on scale-based questionnaire surveys as the main methodological approach [2-4]. This traditional data collection is constrained by temporal factors and, due to operational and participation barriers, results in limited sample sizes, consequently failing to achieve real-time monitoring of dynamic psychological states. The static nature of survey implementation hinders continuous assessment of emotional fluctuations. Additionally, traditional assessments often overlook early linguistic signals of seasonal affective disorder, such as abrupt increases in negative vocabulary frequency [5]. The evolution of artificial intelligence has facilitated novel pathways for mental health monitoring through textual sentiment recognition technologies. It has been suggested that within digital environments, short texts published by students can be utilized for early identification of potential negative emotions, preventing their progression into extreme psychological states [6]. Enhanced capability in managing student crises correlates with heightened precision in monitoring online discourse, thereby enabling educators to discern patterns of emotional fluctuation during public emergencies or individual psychological crises. This facilitates the formulation of more targeted intervention strategies. To address the shortage of counseling resources, sentiment analysis technology can identify emotional tendencies in student-generated texts, predict at-risk populations for early screening, and support decision-making with data-driven insights.

2. Review of Applications of Sentiment Analysis Technology

Sentiment analysis technology constitutes a computational method that systematically extracts and quantifies subjective emotional information from texts through natural language processing and text mining means. Its core objective is to identify emotional polarity, intensity, and targets from unstructured texts, providing data support for decision-making [7]. This technology has deeply penetrated multiple domains, including social media public opinion monitoring and mental health early warning, demonstrating an evolutionary trend from single-text analysis to multimodal fusion. For instance, the LLaMA3.2-Vision fine-tuning framework has been shown to effectively assist government public opinion monitoring in social media scenarios, significantly enhancing the accuracy and efficiency of netizen sentiment recognition [8]. Similarly, J. P. Nayinzira and M. Adda employed a structured mental health counseling dialogue dataset to construct the SentimentCareBot framework, systematically assessing the efficacy of varied Retrieval-Augmented Generation (RAG) methodologies [9]. Through comparative analysis of two prominent large language models (OpenAI and MistralAI), the study validated that integrating sentiment analysis modules within the RAG infrastructure substantially improves the operational performance and contextual accuracy of mental health support chatbots. Soria et al. released a multimodal cross-platform dataset encompassing social media images and texts, voice recordings, and wearable device data, which supports multidimensional analysis of depression tendencies and enables automated monitoring of psychological states through integration with deep learning algorithms [10]. Additionally, in the mental health domain, traditional machine learning models demonstrate high practicality. Juanita et al. developed a machine learning framework analyzing Indonesian online health consultations, demonstrating Gradient Boosting's superior performance in classifying anxiety responses, with optimal results using hybrid optimization techniques [11]. However, traditional machine learning models generally underperform deep learning architectures in both emotion classification and predictive tasks. Hinduja et al. developed a proactive mental health monitoring framework by implementing comparative analysis of LSTM deep learning models versus conventional machine learning approaches on Twitter data [12]. Empirical results demonstrate the LSTM model's superior performance in mental health prediction, particularly excelling in time-series analysis. Mainstream deep learning models such as the CNN-BiLSTM combined model can effectively capture long-distance dependencies and contextual information in texts, improving sentiment classification accuracy [13]. Therefore, this paper integrates traditional machine learning models and deep learning models to address problems, including scenario incompatibility and slow response speed. It conducts real-time, dynamic analysis of mental health from a holistic perspective, reducing the impact of delayed discovery, and provides precise, real-time, and compliant technical support for university psychological service systems.

3. Research Methods and Design

The research primarily constructs a sentiment classification model based on sentiment analysis technology to effectively identify students' mental health states on social platforms. The study mainly consists of the following four components.

3.1. Short Text Data Collecting

The first component involves collecting substantial short text data from target users, primarily covering personal opinions, emotional state descriptions, and daily experience sharing. Before model analysis, systematic data cleaning steps are required to eliminate irrelevant interference information such as advertisements, emojis, symbols, and low-quality texts, thereby enhancing data validity and relevance to the research topic.

3.2. Text Annotation

Prior to model training, the cleaned text data requires manual annotation. The purpose of annotation is to achieve structured sentiment classification, improving the accuracy of subsequent model analysis [14]. The annotation rules are detailed in Table 1 below.

Table 1. Sentiment Annotation Rules.

Sentiment Category	Annotation Rules
Positive	Expressions conveying positive commendatory emotions such as happiness, hope, satisfaction, or gratitude, without implicit negative connotations.
Neutral	Statements of fact or questions devoid of discernible emotional tone.
Negative	Expressions containing explicit derogatory terms or metaphors indicating sadness, anger, anxiety, despair, loneliness, or similar negative states.

3.3. Text Feature Extraction

After completing data preprocessing and annotation tasks, the JIEBA word segmentation tool, specifically designed for Chinese text, was used to perform word segmentation processing. Subsequently, stop words, punctuation marks, and numbers were removed according to a stop word dictionary [15]. Feature extraction was conducted based on the TF-IDF (Term Frequency-Inverse Document Frequency) method, where words with higher weights were extracted and annotated according to their TF-IDF values, and then passed to the model, as shown in Equation (1).

$$TF\text{-}IDF(w, d, D) = \underbrace{F(w, D)}_{TF} \times \underbrace{\log \frac{N}{|\{d: w \in d\}|}}_{IDF} \quad (1)$$

Where w represents a single word, d represents a single document, and D represents the entire document collection. $F(w, D)$ denotes the number of occurrences of w in document D, i.e., term frequency. N represents the total number of documents in the entire collection, and $|\{d: w \in d\}|$ represents the number of documents containing word w. IDF is the inverse document frequency index: the smaller $|\{d: w \in d\}|$ is, the larger IDF becomes, indicating that the word has strong representativeness. Through this process, the original text data is transformed into high-dimensional numerical feature vectors, thereby effectively capturing key semantic information in the text, such as topic distribution and sentiment tendency.

3.4. Model Evaluation

The preprocessed dataset will be partitioned into training and testing sets according to a predetermined ratio. During the model training phase, the study employs supervised learning methods, utilizing annotated feature vectors to train the classification model. By optimizing the objective function, the model autonomously learns complex mapping relationships between textual features and mental health labels, thereby constructing high-precision classification decision boundaries. The trained model acquires automated classification capabilities for newly collected short texts, enabling effective identification of potential mental health risk indicators among student populations and providing data support for early warning and intervention. For the classification task of mental health problem detection, model performance is typically comprehensively evaluated based on four fundamental discrimination outcomes on the test set: True Positive (TP), denoting instances where the actual class is positive and correctly predicted as positive; True Negative (TN), referring to instances where the actual class is negative and correctly predicted as negative; False Positive (FP), representing instances where the actual class is negative but erroneously predicted as positive; and False Negative (FN), indicating instances where the actual class is positive but erroneously predicted as negative. Based on these four outcome categories, the following four core performance metrics can be further calculated.

Accuracy, used to measure the overall proportion of correct classifications by the model, as detailed in Equation (2).

$$\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{2}$$

Precision reflects the proportion of true positive instances among all samples predicted as positive, as detailed in Equation (3).

$$\frac{TP}{TP+FP} \tag{3}$$

Recall represents the proportion of actual positive instances correctly identified by the model, as detailed in Equation (4).

$$\frac{TP}{TP+FN} \tag{4}$$

F1-score, as the harmonic mean of precision and recall, is used to comprehensively evaluate the model's balanced performance in positive class identification. Its value is calculated as twice the product of precision and recall divided by their sum, as detailed in Equation (5).

$$2\times\frac{Precision \times Recall}{Precision \times Recall} \tag{5}$$

4. Model Construction

4.1. Data Collection

Sina Weibo (weibo.com), as one of China's most influential social media platforms, has formed a rapidly disseminating information network environment due to its massive user base and high-frequency interaction patterns. Compared to traditional media, Weibo users tend to express opinions in real-time, straightforward, and emotionally distinctive ways. This unique expression mechanism makes it a valuable data resource for social sentiment monitoring and public emotion trend analysis. Therefore, this study selected Weibo as the data collection platform, setting the data collection period from 00:00 on January 1, 2023, to 22:00 on August 26, 2025. To better investigate the causes of mental health crises among college students, we used Python 3.13 as the tool, employed the requests library to send HTTP requests, and simulated browser behavior to crawl a total of 4,201 posts under topics such as "CET-4/6", "civil service exams", "postgraduate entrance exams", "employment", "college students", and "depression".

4.2. Data Preprocessing

1) Data Cleaning and Annotation

The crawled data underwent preprocessing, including the removal of duplicate posts, null values, and noise information unrelated to the content, such as advertisements, URLs, and emojis. Subsequently, the segmented Weibo posts were annotated, with annotation examples detailed in Table 2 below. Finally, the JIEBA segmentation tool was used for word segmentation, with the Harbin Institute of Technology stopword list applied, resulting in 3,950 effectively annotated data entries.

Table 2. Annotation Examples.

Sentiment	Label	Example Content	Annotation Basis
Positive	0	"College life is colorful, and so am I."	Contains the positive adjective "colorful," expressing a positive emotional tendency toward college life
Neutral	1	"Being lazy and indulgent suits me best."	Describes a life state without a clear emotional tendency, meeting the definition of neutral
Negative	2	"The more I study, the more I collapse—teacher certification"	Contains the word "collapse", indicating negative emotions

2) Feature Extraction

Based on the annotated data, the study retained "college students" as the scenario keyword and extracted the top 20 feature words ranked by TF-IDF values, as shown in Table 3. The overall numerical distribution (0.19–0.58) indicates that the feature words have good discriminative power. Academic pressure-related vocabulary dominated, while emotional states and daily topics served as significant supplements. Specifically, academic development constituted the primary source of psychological pressure for college students, with feature words not only occupying 12 of the top 20 positions but also exhibiting the highest weights. "Major courses" (0.5856), "postgraduate entrance exams" (0.5677), and "CET-4/6" (0.4421) ranked highest, directly reflecting the immense psychological burden caused by academic workload and competitive exams. Closely related were high-frequency words such as "anxiety" (0.4234), "collapse" (0.2877), and "pressure" (0.2456), confirming that academic pressure has translated into emotional distress. Meanwhile, career-planning keywords like "job hunting" (0.4110) and "civil service exams" (0.3756) also ranked highly, revealing widespread anxiety about uncertain futures. Notably, leisure and lifestyle terms such as "celebrity chasing" (0.2118) and "express delivery" (0.2340) appeared, reflecting college students' attempts to regulate emotions through entertainment and daily life. This feature distribution visually depicts the complex emotional landscape formed by contemporary college students under academic pressure, career development, and socialization, providing a feature foundation for constructing a fine-grained mental state recognition model.

Table 3. Top 20 Feature Words.

Rank	Feature Word	TF-IDF	Rank	Feature Word	TF-IDF
1	Major courses	0.5856	11	Dormitory	0.3057
2	Postgraduate exams	0.5677	12	Collapse	0.2877
3	CET-4/6	0.4421	13	Success	0.2775
4	Anxiety	0.4234	14	Pressure	0.2456
5	Job hunting	0.4110	15	Life	0.2431
6	Math	0.3876	16	School	0.2357

7	Civil service exams	0.3756	17	Express delivery	0.2340
8	Double degree	0.3378	18	Teacher	0.2217
9	Academic Advisor	0.3228	19	Celebrity chasing	0.2118
10	College students	0.3119	20	Scream	0.1947

Note: All Chinese words have been translated into their corresponding English words.

5. Model Evaluation

After extracting text features using TF-IDF, the study constructed a three-category test set containing 2,756 data entries, with sentiment distributions of 0 (positive), 1 (neutral), and 2 (negative) at 40%, 20%, and 40%, respectively. The evaluation results are shown in Table 4. Traditional machine learning methods, including Support Vector Machine (SVM), Logistic Regression, Multinomial Naive Bayes, and Random Forest, were compared with the deep learning method CNN-BiLSTM-Attention. Traditional machine learning models are mostly lightweight but rely heavily on manual features, making it difficult to capture deep semantic relationships. In contrast, CNN-BiLSTM-Attention automatically extracts deep features and enhances interpretability through attention mechanisms. The hybrid model reduces reliance on a single model and improves overall system stability and generalization capability [16].

Table 4. Model Evaluation Results.

Model	Sentiment	Precision	Recall	F1-score	Accuracy
SVM	0	0.927	0.772	0.843	0.742
	1	0.674	0.715	0.694	
	2	0.762	0.681	0.719	
Logistic Regression	0	0.919	0.801	0.856	0.763
	1	0.692	0.738	0.714	
	2	0.751	0.703	0.737	
Multinomial Naive Bayes	0	0.882	0.842	0.862	0.781
	1	0.731	0.786	0.758	
	2	0.773	0.702	0.762	
Random Forest	0	0.896	0.851	0.873	0.792
	1	0.758	0.792	0.775	
	2	0.784	0.721	0.777	
CNN-BiLSTM-Attention	0	0.907	0.879	0.893	0.813
	1	0.781	0.825	0.802	
	2	0.802	0.783	0.792	

The model evaluation results in Table 4 and the loss comparison in Figure 1 clearly reveal the performance differences and characteristics of various machine learning methods in sentiment classification tasks. Among traditional machine learning models, Random Forest performed best (accuracy: 0.792), with its ensemble learning mechanism effectively improving generalization ability. As shown in Figure 1, its loss value was also the lowest compared to other machine learning models. Multinomial Naive Bayes (accuracy: 0.781) achieved the highest F1 score in positive sentiment identification due to its probabilistic modeling, but was limited by the feature independence assumption. Logistic Regression (accuracy: 0.763) showed relatively stable performance but had lower precision in neutral sentiment classification. Overall, Logistic Regression and Multinomial Naive Bayes exhibited ideal fitting states, while SVM (accuracy: 0.742) achieved extremely high precision in positive sentiment (0.927) but the lowest recall (0.772), indicating overfitting. This is also validated in Figure 1, where the training and test set losses diverge significantly, reflecting insufficient generalization capability. All machine learning models showed relatively low recall for negative sentiment, suggesting limitations in recognizing metaphorical expressions and sarcastic contexts.

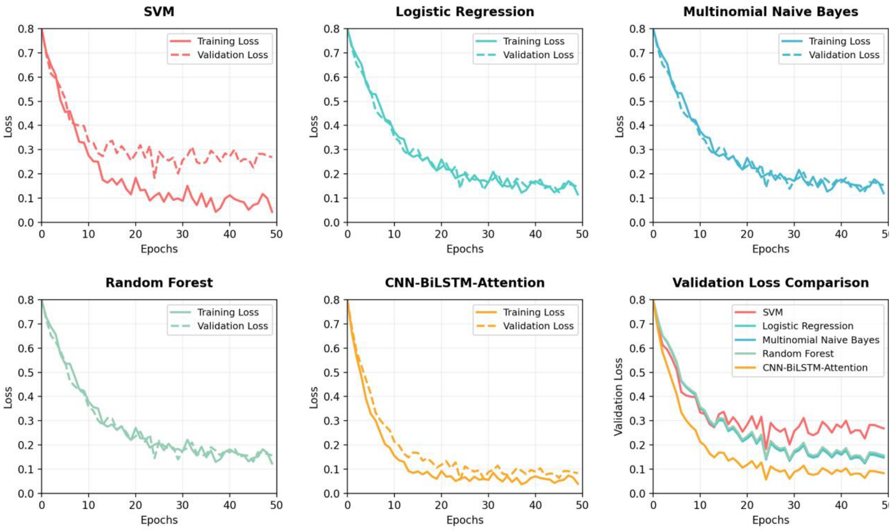

Figure 1. Model Training and Testing Loss Comparison.

In contrast, the CNN-BiLSTM-Attention deep learning model demonstrated comprehensive advantages (accuracy: 0.813), with an F1-score as high as 0.829. It significantly outperformed traditional machine learning models in identifying the most challenging neutral and negative sentiments. As evident from Figure 1, the training and testing losses of this model are significantly lower than those of other machine learning approaches. This performance enhancement stems from its hierarchical architecture: the CNN component captures local semantic features, the BiLSTM module models long-range contextual dependencies, and the attention mechanism dynamically focuses on critical text segments. Consequently, the model achieves a substantial improvement in recall while maintaining high precision, enabling deep understanding and stable classification of complex emotional expressions.

6. Conclusions

This study focuses on sentiment analysis technology for monitoring college students' mental health, innovatively integrating traditional machine learning models with a hybrid deep learning model based on CNN-BiLSTM-Attention. By combining local semantic feature extraction and long-range contextual modeling capabilities, the model achieves precise identification of complex emotional expressions. Feature analysis reveals that academic burden constitutes the core psychological stressor. Experimental results demonstrate that the CNN-BiLSTM-Attention model significantly outperforms traditional machine learning methods, particularly in recognizing neutral and negative emotions, while also exhibiting the lowest loss. In contrast, SVM shows overfitting, which is attributed to the inherent limitations of the model itself. The hybrid model developed in this study enables sentiment recognition of students' social texts; however, the corpus source is relatively homogeneous, and real-time performance remains slightly inadequate. Future work should focus on deploying the hybrid model architecture to achieve real-time text analysis. Secondly, a dynamic lexicon update mechanism should be developed to regularly incorporate emerging internet expressions. Additionally, a multi-source data correlation analysis system should be constructed to integrate social texts with behavioral data. Finally, strict ethical guidelines must be established, including informed consent protocols and manual review processes, to ensure the compliance of technology applications.

Future research should expand into multimodal data analysis and develop personalized adaptation algorithms to further enhance the system's performance in complex scenarios such as sarcasm recognition, thereby providing efficient and reliable technical support for university mental health service systems.

Acknowledgments: This study was supported by the 2025 project "Research on Mental Health Assessment of College Students Based on Sentiment Analysis Technology" (Grant Number: XLJKJY2545B) from Sichuan Key Research Base for Mental Health Education, Southwest Jiaotong University.

References

1. X. Wang, and Y. Zou, "Psychological research of college students based on online education under COVID-19," *Sustainability*, vol. 15, no. 2, p. 1040, 2023, doi: 10.3390/su15021040.
2. R. Hou, S. Huang, K. P. L. Fung, A. Li, C. Jia, S. Cheng, and J. P. H. Wong, "Who is helping students? A qualitative analysis of task-shifting and on-campus mental health services in China's university settings," *Social Science & Medicine*, vol. 363, p. 117527, 2024.
3. C. Wang, J. Hu, Y. Zhou, L. Jiang, J. Chen, J. Xi, and S. Zhang, "Problematic smartphone usage and inadequate mental health literacy potentially increase the risks of depression, anxiety, and their comorbidity in Chinese college students: A longitudinal study," *Journal of Affective Disorders Reports*, vol. 20, p. 100907, 2025, doi: 10.1016/j.jadr.2025.100907.
4. H. Wang, R. Jia, M. Zhang, and W. Fan, "The influence of stress on mental health among Chinese college students: The moderating role of psychological suzhi," *Heliyon*, vol. 10, no. 5, 2024, doi: 10.1016/j.heliyon.2024.e26699.
5. T. Althoff, K. Clark, and J. Leskovec, "Large-scale analysis of counseling conversations: An application of natural language processing to mental health," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 463-476, 2016, doi: 10.1162/tacl_a_00111.
6. N. Salimi, B. Gere, W. Talley, and B. Irioogbe, "College students mental health challenges: Concerns and considerations in the COVID-19 pandemic," *Journal of College Student Psychotherapy*, vol. 37, no. 1, pp. 39-51, 2023.
7. M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731-5780, 2022, doi: 10.1007/s10462-022-10144-1.
8. H. Du, Q. Yu, and J. Chen, "Research on sentiment analysis of online public opinion based on multimodal big language modeling," *Journal of Computational Methods in Sciences and Engineering*, 2025, doi: 10.1177/14727978251361417.
9. J. P. Nayinzira, and M. Adda, "SentimentCareBot: Retrieval-augmented generation chatbot for mental health support with sentiment analysis," *Procedia Computer Science*, vol. 251, pp. 334-341, 2024, doi: 10.1016/j.procs.2024.11.118.
10. K. M. Soria, and B. Horgos, "Factors associated with college students' mental health during the COVID-19 pandemic," *Journal of College Student Development*, vol. 62, no. 2, pp. 236-242, 2021, doi: 10.1353/csd.2021.0024.
11. S. Juanita, A. H. Daeli, M. Syafrullah, W. Anggraeni, and M. H. Purnomo, "A machine learning approach for text pattern diagnosis in mental health consultations," *Decision Analytics Journal*, vol. 15, p. 100572, 2025, doi: 10.1016/j.dajour.2025.100572.
12. S. Hinduja, M. Afrin, S. Mistry, and A. Krishna, "Machine learning-based proactive social-sensor service for mental health monitoring using twitter data," *International Journal of Information Management Data Insights*, vol. 2, no. 2, p. 100113, 2022, doi: 10.1016/j.jjimei.2022.100113.
13. X. Feng, N. Angkawisittpan, and X. Yang, "A CNN-BiLSTM algorithm for Weibo emotion classification with attention mechanism," *Mathematical Models in Engineering*, vol. 10, no. 2, pp. 87-97, 2024, doi: 10.21595/mme.2024.24076.
14. J. Zhou, Z. Liang, Y. Fang, and Z. Zhou, "Exploring public response to ChatGPT with sentiment analysis and knowledge mapping," *IEEE Access*, vol. 12, pp. 50504-50516, 2024, doi: 10.1109/access.2024.3386362.
15. M. Liang, and T. Niu, "Research on text classification techniques based on improved TF-IDF algorithm and LSTM inputs," *Procedia Computer Science*, vol. 208, pp. 460-470, 2022, doi: 10.1016/j.procs.2022.10.064.
16. L. Shan, Y. Liu, M. Tang, M. Yang, and X. Bai, "CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction," *Journal of Petroleum Science and Engineering*, vol. 205, p. 108838, 2021, doi: 10.1016/j.petrol.2021.108838.