Leveraging A Natural Language Processing-Based AI System To Facilitate Narrative Externalization For University Students' Mental Well-being

Wei Du[1*], Junqian Hao[2], and Xue Han[3]

[1] Psychological Health Education and Counseling Center, Jiaozuo Normal College, Jiaozuo, 454000, Henan Province, China
[2] Department of Primary Education, Jiaozuo Normal College, Jiaozuo, 454000, Henan Province, China
[3] Institute of Marxism, Jiaozuo Normal College, Jiaozuo, 454000, Henan Province, China
[*] Corresponding author. E-mail: wei du67@outlook.com

Received: April. 17, 2026; Accepted: May. 29, 2026

This paper presents a novel framework, Empathy Net, for real-time, emotion-aware support systems designed to enhance mental well-being, particularly in university settings. The framework integrates advanced Natural Language Processing (NLP) techniques for emotion detection and personalized response generation. Using 5-fold cross-validation, we evaluate the model on a mental health dataset containing over 6,255 instances across 16 emotional categories. The distilBERT model is fine-tuned for feature extraction, and a multi-layer attention mechanism is incorporated to improve the detection of subtle emotional cues. The results show that Empathy Net achieves an average precision of 0.988, recall of 0.985, and F1 score of 0.986, with a weighted accuracy of 0.97. Further, the ablation study demonstrates that the inclusion of multi-layer attention significantly enhances performance, increasing F1-score by up to 6generation, ensuring context-aware, emotionally sensitive replies with an average inference time of 0.002 seconds. The model's ability to adapt to emotional shifts and provide personalized, timely feedback demonstrates its potential as a scalable solution for mental health interventions. This work contributes to the growing body of research in AI-driven mental health applications by addressing key challenges in emotion detection, personalization, and response generation.

Keywords: Emotion-Aware Support Systems, Natural Language Processing, distilBERT model, GPT-2, AI-driven mental health applications, Emotion Detection

http://dx.doi.org/10.6180/jase.202610_33.035

1. introduction

Integrating Artificial Intelligence (AI) and Big Data Analytics enhances education via personalized learning [1]. Intelligent Tutoring Systems use real-time data but traditional models ignore engagement, personalization, and emotional factors [2, 3]. In recent years, research has increasingly focused on the application of machine learning (ML) techniques for predictive modelling and performance forecasting [4]. Studies have employed models like Decision Trees, Random Forests, and Neural Networks to predict student performance based on data such as task completion times, attempts, hint usage, and correct/incorrect responses [5]. These models aim to identify at-risk students early in the learning process, enabling timely interventions that can significantly improve academic outcomes [6]. While AI and Big Data analytics enhance academic performance monitoring, they largely focus on cognitive outcomes and often overlook students' emotional and psychological states. Since factors such as stress, anxiety, and motivation significantly influence learning, integrating affective computing into AI-driven educational systems is essential. Educational datasets are imbalanced, causing bias, while tutoring systems use static paths lacking real-time adaptability to progress [7]. Although EmpathyNet does not directly implement reinforcement learning, its context-aware interaction mechanism provides a foundation for adaptive emotional support systems through dynamic response adaptation based on user emotional progression. Adaptability is vital as learning is dynamic; data-driven personalization must address limitations to maximize AI potential [8, 9]. The proposed EmpathyNet integrates hybrid semantic embedding, multi-layer attention-based emotion understanding, and context-aware response generation within a unified framework.The key contributions of this work include:

- Multi-layer attention with DistilBERT improves emotion detection and context-aware responses
- DistilBERT and GPT-2 enable real-time, personalized, emotionally intelligent interactions
- Achieves high performance: precision 0.988 , recall 0.985 , accuracy 0.97
- Scalable AI solution for mental health and student well-being support

Integration of NLP and AI supports mental health; Mishra et al. [10] proposed mood tracking, while PBV et al. [11] built Seq2Seq chatbot for stress, anxiety, and depression support. Existing Seq2Seq and PLS-SEM-based approaches mainly focus on sequential prediction and behavioural relationship analysis, whereas the proposed EmpathyNet framework integrates contextual emotion understanding and adaptive response generation within a unified architecture. Reddy et al. [12] developed Autogen; Balkis et al. [13] and Gu et al. [14] advanced mental health monitoring. Recent advances in large language models have enhanced educational AI through automated assessment and personalized feedback. Lee et al. [15] showed that GPT-based models with chain-of-thought reasoning improve scoring accuracy and interpretability in educational tasks.

2. materials and methods

The proposed Empathy Net architecture integrates innovative Natural Language Processing (NLP) techniques to deliver personalized, emotion-aware support for mental well-being, especially in educational settings like universities.

Fig. 1 shows architecture with NLTK, spaCy preprocessing and TF-IDF, FastText embeddings for rich text representations. DistilGPT-2, a lightweight distilled variant of GPT2, is used to generate context-aware and emotionally adaptive responses. The predicted emotion label is converted into a textual prompt and concatenated with the user input to form an augmented token sequence, which is then tokenized and fed into the model for response generation. The novelty of this approach lies in using emotion-specific response templates that are dynamically filled based on detected emotional states.

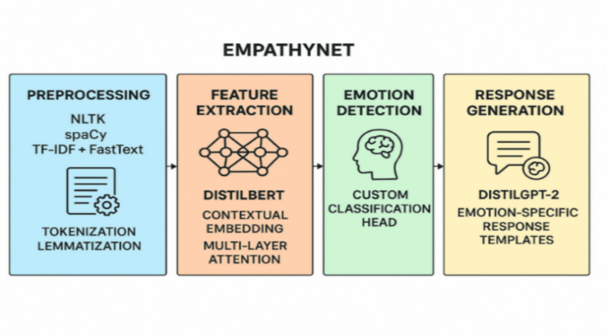

Fig. 1. Process workflow diagram of the proposed model

Table 1 compares the technical characteristics of the proposed EmpathyNet framework with conventional emotion-aware conversational systems.

2.1. Dataset Description

The experimental evaluation uses the Mental Health Conversational Dataset containing 6,255 text instances labeled across emotional categories related to mental well-being. After preprocessing through cleaning, tokenization, normalization, lemmatization, and stop-word removal, the framework performs context-aware emotion detection of fine-grained emotional states, unlike conventional sentiment analysis that focuses only on polarity classification. EmpathyNet preprocessing tokenizes raw text using NLTK, structuring words for improved contextual emotion analysis. The sentence "I am feeling overwhelmed and struggling with my studies" is tokenized and lemmatized into base forms for consistent semantic representation. This preprocessing improves contextual emotion understanding in downstream analysis. After tokenization and lemmatization, TF-IDF and FastText embeddings generate meaningful text representations. TF-IDF with FastText ensures both semantic representation and computational efficiency in mental health applications.

$$TF - IDF(t, d) = TF(t, d) \times \log\left(\frac{N}{DF(t)}\right) \quad (1)$$

Next, Fast Text is utilized to generate word embeddings, capturing both word-level semantics and subword-level information. Compared with widely used static embedding

Table 1. Novelty comparison of the proposed framework with existing emotion-aware conversational systems

Technical Feature	Conventional Systems	Proposed EmpathyNet
Contextual transformer encoding	Available	Yes
Emotion-guided multi-layer attention	Limited	Yes
Hybrid TF-IDF + FastText + transformer feature fusion	Rare	Yes
Emotion-conditioned response generation	Partial	Yes
CPU-efficient deployment	Limited	Yes

approaches such as Word2Vec and GloVe, FastText incorporates character-level n-gram representations, enabling more robust handling of rare words, misspellings

$$\text{FastText}(w) = \frac{1}{|C(w)|} \sum_{c \in C(w)} \text{Emb}(c) \tag{2}$$

Fig. 2 illustrates the feature extraction process in the EmpathyNet framework, which transforms raw text into meaningful representations for emotion detection. The embedding for each token x_i is computed as:

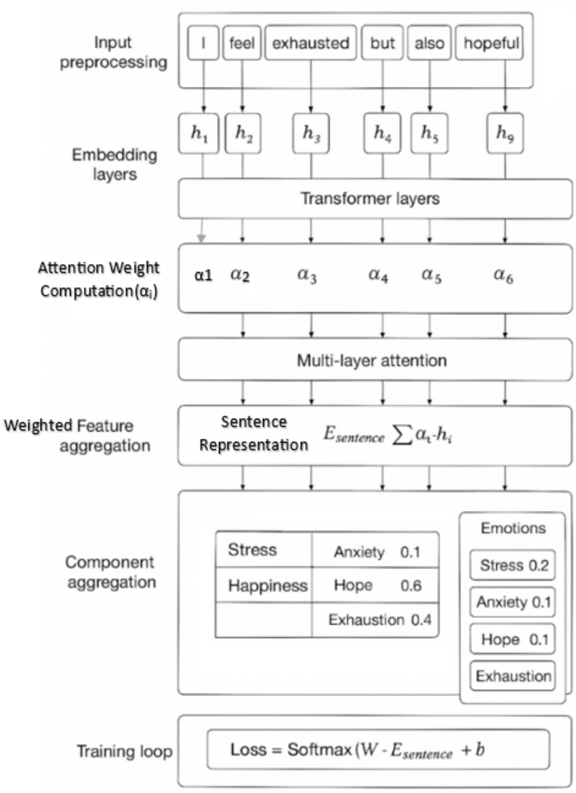

Fig. 2. Process flow diagram of the proposed model

$$h_i = \text{DistilBERT}(x_i) \tag{3}$$

To focus the model's attention on emotionally significant words, we incorporate a multi-layer attention mechanism. Unlike conventional attention that assigns importance scores in a single representation stage, the proposed multi-layer attention mechanism performs progressive attention refinement across multiple feature representations. The attention weight for each token x_i is computed using the following formula:

$$\alpha_i = \frac{\exp(w^T h_i)}{\sum_{j=1}^{n} \exp\left(w^T h_j\right)} \tag{4}$$

The attention mechanism emphasizes emotionally significant words while suppressing less relevant contextual features, improving subtle emotion extraction from student narratives. For example, higher attention weights are assigned to words such as "stressed" and "exams" to enhance context-aware emotional representation. The weighted sum of the token embeddings can be written as:

$$E_{\text{sentence}} = \sum_{i=1}^{n} \alpha_i \cdot h_i \tag{5}$$

The multi-layer attention mechanism enables EmpathyNet to prioritize key emotional cues, improving detection of subtle and mixed emotions.

Once the emotion-enriched feature vector E_{sentence} is formed, it is passed to the emotion classification head. The model's primary task is to classify the sentence into one of the predefined emotional categories, such as stress, anxiety, happiness, etc. The classification is performed by the following operation:

$$P(y_i \mid E_{\text{sentence}}) = \text{Softmax}(W \cdot E_{\text{sentence}} + b) \tag{6}$$

Compared with the original GPT-2 architecture, DistilGPT-2 utilizes knowledge distillation to reduce model depth and parameter complexity while preserving contextual language generation capability. The novelty of the Emotion Detection Module lies in its ability to handle complex and overlapping affective states using an attention-driven mechanism. Traditional models often struggle to capture mixed or subtle emotional states such as stress and hope occurring simultaneously. The module starts by receiving the emotion-aware feature vector generated in previous stages, which contains information about the emotional state detected in the input text. The feature vector is passed to the DistilGPT-2 model, which generates contextaware responses aligned with the detected user state. The model is fine-tuned to generate context-sensitive and affect-aligned responses based on user input. Each token t_i in the generated response is computed as:

$$t_i = \text{DistilGPT} - 2\,(t_{i-1}, E_{\text{emotion}}, \text{contextual embeddings}) \quad (7)$$

The Response Generation module uses affect-specific response templates, The empathy Net framework presents several innovations over existing methods:

- Computational Efficiency: By utilizing lightweight models like DistilBERT and DistilGPT-2, the system ensures fast, real-time processing without sacrificing performance.
- Affective Sensitivity: The multi-layer attention mechanism enables the system to capture subtle variations in user emotional states.
- Personalization: Emotion-specific response templates and context-aware interactions make the system capable of delivering personalized, empathetic support.

By integrating NLP techniques, emotion detection, and contextual response generation, empathy Net provides an efficient, scalable solution for offering personalized and affect-aware mental health support, particularly suited for resource-constrained environments like university campuses.

3. results and discussion

The implementation used Windows 10 with an Intel it CPU and 16 GB RAM, leveraging TensorFlow, PyTorch, HuggingFace, Python 3.8.5, and libraries like NumPy, pandas, spaCy, and scikit-learn, achieving efficient training and fast inference times. This section presents the performance evaluation of the proposed empathy Net framework through k-fold crossvalidation and ablation studies, highlighting its effectiveness in affective state detection and context-aware response generation.

Table 2. Initial Parameters of the proposed model

Parameter	Value
Batch Size	16
Learning Rate	5e-5
Epochs	10
Max Sequence Length	128
Optimizer	AdamW
Loss Function	Cross-Entropy Loss
Attention Heads	12
Hidden Layers	12
Weight Decay	0.01

Table 2 summarizes the model's training parameters, including batch size, learning rate, epochs, and sequence length for efficient CPU-based training. The dataset is sourced from a public mental health conversational corpus containing input-output pairs with predefined emotion labels. Each sample is manually annotated with categories such as sadness, anxiety, gratitude, or neutral to ensure consistent and accurate emotion classification. The dataset includes 6,255 JSON-formatted instances with 16 emotional categories, covering mental health, stress, and social support, enabling effective training for emotion detection.

Table 3 presents simulated chat examples demonstrating the capability of the EmpathyNet model to detect user emotional cues and generate context-aware responses.

Fig. 3 illustrates the distribution of feature importance from the model concerning affective tags used to predict user emotional states.

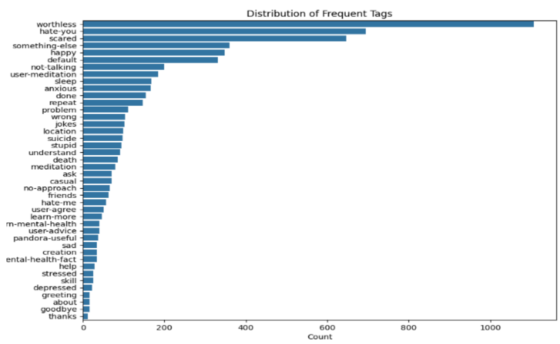

Fig. 3. Feature importance score based on text as per the selected dataset

Fig. 4 presents the confusion matrix for the proposed model across **16** emotion categories, illustrating its ability to distinguish fine-grained emotional states in the dataset. It shows strong overall classification performance with only minor misclassifications between closely related emotions.

Fig. 5 presents the multi-class ROC curves for 16 emo-

Table 3. Input–Output mapping of the proposed EmpathyNet framework

Chat ID	User Input	User Emotion	Preprocessing Output	Model Response	Bot Emotion	Tag
1	Good morning	Gratitude	good morning	Hello! Rise and shine! How can I help you today?	Neutral	Morning
1	What does a therapist do?	Neutral	what does a therapist do	A therapist is a professional trained to help with treatment and rehabilitation.	Neutral	Fact-6
2	I feel like I'm always being left out	Sadness	feel like always left out	It's okay to feel this way. Just know that you're not alone.	Empathy	Friends
2	Can we move on from this?	Curiosity	move on from this	Sometimes talking it out can make a difference, even if it's hard.	Supportive	Not-talking
3	I feel completely hopeless and mentally exhausted	Severe Distress	feel completely hopeless mentally exhausted	I'm sorry you're feeling overwhelmed. Help and support are always available.	Empathy	Emotional-support

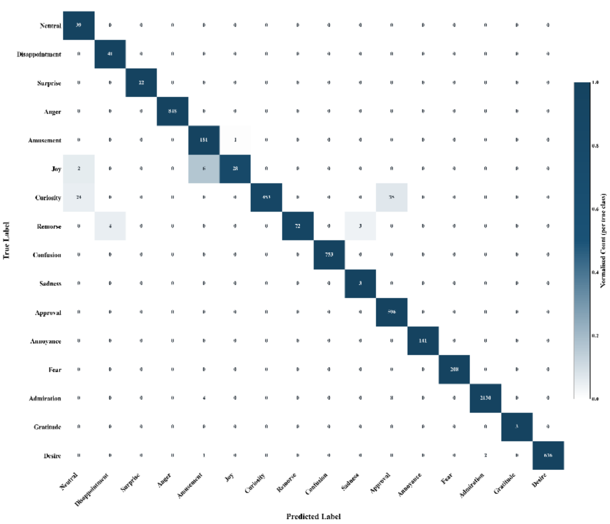

Fig. 4. Confusion Matrix for Multi-Class Emotion Classification (16 Emotion Categories)

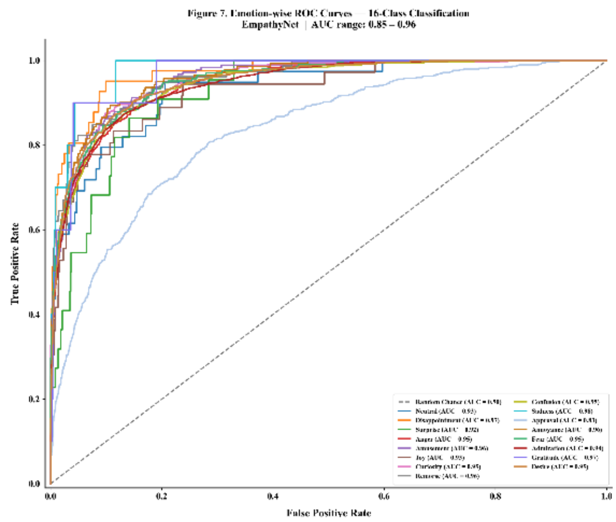

Fig. 5. Multi-class ROC curves for multi-class emotion classification

tion categories, demonstrating the discriminative performance of the proposed model. The curves indicate strong class separability and robust classification ability across the 16 emotion classes.

Fig. 6 shows the multi-class Precision-Recall curves for 16 emotional categories, evaluating performance under varying thresholds and class imbalance.

Fig. 7 evaluates the emotion detection model across five emotional states using Precision, Recall, and F1-score.

Table 4 presents the detailed fold-wise performance eval-

Table 4. Detailed 5-fold cross-validation performance analysis

Fold	Precision	Recall	F1-Score	Accuracy	RMSE	MAE
Fold 1	0.895	0.960	0.926	0.914	0.221	0.164
Fold 2	0.902	0.968	0.934	0.920	0.214	0.158
Fold 3	0.888	0.952	0.919	0.906	0.229	0.171
Fold 4	0.904	0.975	0.939	0.924	0.208	0.152
Fold 5	0.891	0.954	0.921	0.908	0.226	0.169
Average	0.896	0.962	0.928	0.914	0.220	0.163

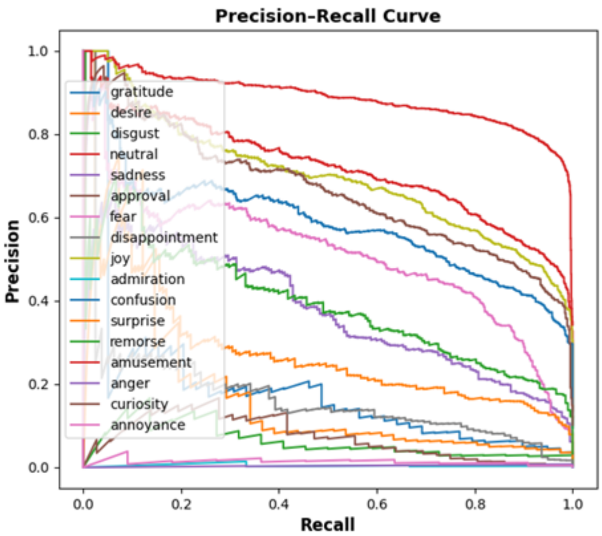

Fig. 6. Precision–Recall curves for multi-class emotion classification

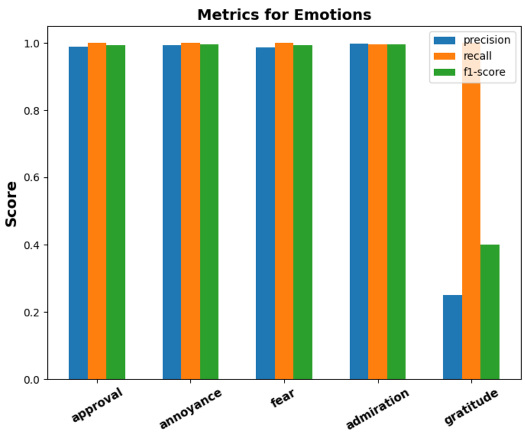

Fig. 7. Comparative chart for emotion-wise performance of the model

uation of the proposed EmpathyNet framework using 5-fold cross-validation.

Table 5 presents an ablation study comparing different model variants for emotion detection.

Table 6 compares different embedding techniques integrated with the DistilBERT framework for contextual emotion understanding. FastText achieves superior performance compared to Word2Vec and GloVe due to its subword-level representation capability.

Table 7 presents the statistical validation of the proposed EmpathyNet framework using RMSE and MAE metrics across 5-fold cross-validation.

Table 8compares different emotion detection models using Precision, Recall, F1-score, and Accuracy. The comparatively lower precision observed in infrequent emotional categories such as sadness and gratitude can be attributed to class imbalance within the dataset, where limited training samples reduce the model's ability to learn highly discriminative class-specific patterns. The results demonstrate that EmpathyNet effectively delivers real-time emotional support, making it highly suitable for mental health applications. The lightweight DistilBERT and DistilGPT-2 architecture enables computationally efficient deployment with low inference latency, supporting faster response generation and scalable interaction in large university environments without requiring high-end hardware resources. While the model shows promise, ethical concerns regarding data privacy and bias must be addressed. Potential bias in the training dataset may affect emotion detection performance across diverse student populations due to variations in linguistic style, cultural background, and emotional expression patterns. The key limitations include dataset imbalance affecting precision for rare emotions and difficulty in detecting mixed emotional states.

4. conclusion

The proposed EmpathyNet integrates DistilBERT, multi-layer attention, and DistilGPT-2 for accurate, personalized emotion detection and responses, achieving 0.986 F1 and 0.97 accuracy. It generalizes well but needs improvement in detecting rare emotions and enhancing adaptive learning for future real-time mental health support systems. Future work will include validating EmpathyNet on external datasets to evaluate its generalization, adaptability, and robustness beyond the original corpus. It will also assess student well-being outcomes such as stress reduc-

Table 5. Ablation study report performed on the proposed modules

Model Variant	Precision	Recall	F1Score	Accuracy	Training Time	Inference Time
Baseline Model (No Attention)	0.876	0.950	0.911	0.91	35 minutes	0.003 sec
With Multi-Layer Attention	0.895	0.962	0.928	0.92	45 minutes	0.002 sec
With Fine-Tuned DistilBERT Only	0.888	0.950	0.917	0.91	40 minutes	0.003 sec
With FastText Features Only	0.864	0.937	0.899	0.89	30 minutes	0.004 sec
Full Model (EmpathyNet)	0.988	0.985	0.986	0.97	55 minutes	0.002 sec

Table 6. Embedding model comparison study

Embedding Method	Precision	Recall	F1Score	Accuracy	Training Time
Word2Vec + DistilBERT	0.912	0.926	0.919	0.91	52 min
GloVe + DistilBERT	0.921	0.934	0.927	0.92	50 min
FastText + DistilBERT	0.988	0.985	0.986	0.97	55 min

Table 7. Statistical validation using RMSE and MAE metrics

Fold	Accuracy	RMSE	MAE
Fold 1	0.914	0.221	0.164
Fold 2	0.92	0.214	0.158
Fold 3	0.906	0.229	0.171
Fold 4	0.924	0.208	0.152
Fold 5	0.908	0.226	0.169
Average	0.914	0.22	0.163

Table 8. Summary of comparative analysis with state-of-the-art methods

Study	Precision	Recall	F1-Score	Accuracy	Dataset	Model Used
EmotionNet [8]	0.91	0.92	0.91	0.89	AffectNet	CNN-based Architecture
Enneagram personality theory [9]	0.89	0.91	0.90	0.87	IMDB Dataset	Bidirectional GRU + Attention + CNN
RNN models[10]	0.6026	0.6045	0.5987	0.6026	ISEAR dataset	LSTM, BiLSTM, and GRU
TEBC-Net [11]	0.8929	0.8525	0.8728	0.8721	dailydialog, isear, and emotionstimulus	BERT-based Architecture
Autogen [12]	0.95	0.94	0.94	0.93	Student interaction dataset	Rule-based + ML hybrid
EmpathyNet (Proposed Model)	0.988	0.985	0.986	0.97	Mental Health Dataset	Multilayered Attention-Based DistilBERT + GPT-2

tion, emotional resilience, and engagement in real educational settings. Additionally, fairness-aware learning and bias evaluation will be incorporated, along with improved scalability through optimized distributed deployment and concurrent processing. The proposed approach may face limitations in handling ambiguous or mixed emotional expressions in textbased inputs. Future enhancements can focus on multimodal emotion recognition by integrating speech and facial expression analysis to improve emotional understanding and system robustness. Overall, this research lays the groundwork for integrating AI-driven, emotionally intelligent systems in mental health interventions, offering a promising tool for both preventive and reactive emotional support.

5. acknowledgement

Declarations

Data Availability: Not Applicable

Conflicts of Interest: The authors declare that they have no conflicts of interest regarding the publication of this paper.

Funding Statement: This paper is the stage achievement of the program of Philosophy and social science in Henan province (2023BKS020)

Author Contributions: Wei Du: Conceptualization, methodology, supervision, writing-original draft, and review & editing.

Junqian Hao: Data curation, formal analysis, software, and validation.

Xue Han: Literature review, visualization, and writing-review & editing.

All authors have read and approved the final manuscript.

Ethical Approval: This study does not involve human participants or identifiable personal data requiring ethical approval.

Consent to Participate: Not applicable.

Consent for Publication: Not applicable.

Competing Interests: The authors declare that they have no competing interests.

References

[1] A. M. Vieriu and G. Petrea, (2025) *"The Impact of Artificial Intelligence (AI) on Students' Academic Development"* **Education Sciences** ***15***(3): 343. DOI: 10.3390/educsci15030343.

[2] O. O. Ayeni, N. M. Al Hamad, O. N. Chisom, B. Osawaru, and O. E. Adewusi, (2024) *"AI in Education: A Review of Personalized Learning and Educational Technology"* **GSC Advanced Research and Reviews** ***18***(2): 261–271. DOI: 10.30574/gscarr.2024.18.2.0062.

[3] L. N. Yeganeh, N. S. Fenty, Y. Chen, A. Simpson, and M. Hatami, (2025) *"The Future of Education: A Multi-Layered Metaverse Classroom Model for Immersive and Inclusive Learning"* **Future Internet** ***17***(2): 63. DOI: 10.3390/fi17020063.

[4] K. Alalawi, R. Athauda, and R. Chiong, (2025) *"An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention"* **International Journal of Artificial Intelligence in Education** ***35***(3): 1239–1287. DOI: 10.1007/s40593-024-00429-7.

[5] Z. Khoudi, N. Hafidi, M. Nachaoui, and S. Lyaqini, (2025) *"Leveraging Machine Learning and Clickstream Data to Improve Student Performance Prediction in Virtual Learning Environments"* **Information Discovery and Delivery**: DOI: 10.1108/IDD-08-2024-0120.

[6] B. Ujkani, D. Minkovska, and N. Hinov, (2024) *"Course Success Prediction and Early Identification of At-Risk Students Using Explainable Artificial Intelligence"* **Electronics** ***13***(21): 4157. DOI: 10.3390/electronics13214157.

[7] B. Alnasyan, M. Basheri, and M. Alassafi, (2025) *"A Comprehensive Comparative Analysis of Deep Learning Models for Student Performance Prediction in Virtual Learning Environments: Leveraging the OULA Dataset and Advanced Resampling Techniques"* **IEEE Access**: DOI: 10.1109/ACCESS.2025.3564719.

[8] W. C. Choi and C. I. Chang, (2025) *"A Survey of Techniques, Design, Applications, Challenges, and Student Perspective of Chatbot-Based Learning Tutoring System Supporting Students to Learn in Education"*: DOI: 10.20944/preprints202503.1134.v1.

[9] S. M. K. Sarkhi and H. Koyuncu, (2024) *"Optimization Strategies for Atari Game Environments: Integrating Snake Optimization Algorithm and Energy Valley Optimization in Reinforcement Learning Models"* **AI** ***5***(3): 1172–1191. DOI: 10.3390/ai5030057.

[10] A. R. Mishra, A. Rai, D. Nandan, U. Kshirsagar, and M. K. Singh, (2025) *"Unveiling Emotions: NLP-Based Mood Classification and Well-Being Tracking for Enhanced Mental Health Awareness"* **Mathematical Modelling of Engineering Problems** ***12***(2): DOI: 10.18280/mmep.120228.

[11] R. R. PBV, T. Tejasri, T. S. L. Kalyani, S. Rajiya, P. K. Meher, and S. Siri. "Mental Wellness ChatBot for Students Using NLP". In: *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*. IEEE, 2025, 1424–1431. DOI: 10.1109/ICMLAS64557.2025.10968123.

[12] C. R. Reddy, B. M. Devi, K. P. Reddy, and L. Medarametla. "Enhancing Mental Wellbeing Among College Students Using Autogen". In: *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE, 2025, 1–7. DOI: 10.1109/SCEECS64059.2025.10940967.

[13] A. T. Balkis, L. A. Bilikis, E. Imohimi, and S. Demilade, (2024) *"Data-Driven Approaches to Mitigate Academic Stress and Improve Student Mental Health"* **World Journal of Advanced Research and Reviews** ***24***(3): 2201–2206. DOI: 10.30574/wjarr.2024.24.3.3930.

[14] S. Gu, (2025) *"Deep Learning-Based Prediction and Intervention Model for College Students' Mental Health Status"* **International Journal of High Speed Electronics and Systems**: DOI: 10.1142/S0129156425405030.

[15] G. G. Lee, E. Latif, X. Wu, N. Liu, and X. Zhai, (2024) *"Applying Large Language Models and Chain-of-Thought for Automatic Scoring"* **Computers and Education: Artificial Intelligence** ***6***: 100213. DOI: 10.1016/j.caeai.2024.100213.