

AI-Integrated Mental Health Support for University Students: A Dual-Study on Conversational Triage and AI-Optimized Appointment Management

Md. Arif Bin Hashem Mahim

Dept. of Computer Science and Engineering

International Islamic University Chittagong (IIUC)

Chattogram, Bangladesh — ID: C231137

Abdullah Al Shaimon

Dept. of Computer Science and Engineering

International Islamic University Chittagong (IIUC)

Chattogram, Bangladesh — ID: C231139

Abstract—University students worldwide are experiencing an unprecedented rise in mental health challenges, including anxiety, depression, academic burnout, and crisis-related behaviors. Existing campus counseling services are frequently constrained by limited professional resources, social stigma, and inefficient administrative workflows, resulting in delayed intervention and unmet student needs. This research proposes an integrated Artificial Intelligence (AI)-driven mental health support framework that jointly addresses early risk identification and service utilization efficiency through a dual-study approach.

Study A focuses on conversational mental health triage using a transformer-based Natural Language Processing (NLP) model fine-tuned on student-centric data. Leveraging the MentalBERT architecture, the system classifies user inputs into Routine, Moderate, and Crisis severity levels, enabling scalable, confidential, and non-judgmental preliminary assessment. A safety-aware escalation mechanism ensures immediate human intervention when high-risk thresholds are exceeded. Study B addresses operational inefficiencies by employing machine learning-based predictive models, including Random Forest and Gradient Boosting, to estimate appointment no-show probabilities and optimize counselor scheduling.

Primary data collected from 53 students at the International Islamic University Chittagong (IIUC) is augmented with a statistically consistent synthetic dataset of 1,100 records to support robust model training while preserving privacy. The proposed framework targets a triage accuracy range of 74–86% and a projected appointment no-show reduction of 10–22%. By integrating technical performance evaluation with ethical governance considerations, this research provides a holistic and deployable AI-assisted mental health support model for university environments.

Index Terms—Artificial Intelligence, Mental Health, Transformer Models, Machine Learning, Triage Accuracy, Behavioral Analytics.

I. INTRODUCTION

Mental health concerns among university students have emerged as a significant global public health issue. Academic pressure, competitive environments, financial uncertainty, social isolation, and the transition to independent living contribute to elevated levels of stress, anxiety, and depressive symptoms. Empirical studies indicate that a substantial proportion of students experience psychological distress during their academic tenure, often peaking in the later years of

study when academic and career-related expectations intensify [4], [5]. Despite this growing demand, institutional mental health services frequently remain under-resourced and unable to provide timely, individualized support.

Traditional counseling infrastructures rely heavily on manual triage, self-reporting, and appointment-based consultations. These mechanisms introduce delays in identifying at-risk individuals and often fail to scale with increasing student populations. Furthermore, stigma associated with seeking mental health support discourages many students from initiating professional contact, resulting in untreated or escalating conditions. Consequently, universities face rising dropout rates, academic disengagement, and compromised student well-being.

Recent advances in Artificial Intelligence (AI), particularly in Natural Language Processing (NLP) and machine learning, present a promising opportunity to augment existing mental health services. Conversational AI systems can provide anonymous, always-available interaction channels that lower psychological barriers to help-seeking. Transformer-based architectures, such as BERT and its domain-specific variants, have demonstrated strong capability in extracting psychological indicators from unstructured text, enabling preliminary risk assessment at scale [3]. However, most existing systems are designed in isolation and focus solely on classification accuracy, without addressing downstream care delivery challenges.

In parallel, inefficiencies in appointment scheduling further strain counseling services. High appointment no-show rates waste limited counselor availability and extend waiting times for other students. Predictive analytics applied to scheduling data has shown potential in healthcare settings for identifying high-risk no-show cases and enabling proactive intervention [4]. Yet, the application of such techniques within university counseling environments remains underexplored.

This research addresses these gaps by proposing an integrated AI-based mental health support ecosystem comprising two complementary studies. Study A develops a conversational AI triage system capable of assessing preliminary mental health risk and recommending appropriate actions while maintaining ethical safeguards and human oversight. Study B focuses on AI-driven appointment management through predic-

tive modeling to reduce no-show rates and optimize counselor workload distribution. Unlike prior single-focus studies, this dual-study framework evaluates the complete student support pathway—from initial self-disclosure to sustained engagement with professional services.

The primary contributions of this work are threefold: (i) the design and evaluation of a transformer-based conversational triage model tailored to student mental health contexts, (ii) the application of machine learning techniques for appointment adherence optimization in university counseling services, and (iii) an integrated ethical framework addressing privacy, transparency, and safety in AI-assisted mental healthcare. Together, these contributions aim to provide a scalable, deployable, and ethically grounded solution for improving student mental health support systems.

II. LITERATURE REVIEW

A. Conversational AI for Mental Health Support

The application of conversational AI in mental health has gained considerable attention due to its potential to deliver scalable and accessible psychological support. Transformer-based models, including BERT and GPT variants, have demonstrated strong performance in identifying emotional states, depressive symptoms, and self-harm indicators from free-text inputs [3]. Benchmark datasets such as MentalChat16K and eRisk have reported triage and classification accuracies ranging from 74% to 86%, highlighting the feasibility of automated preliminary assessment [6].

Despite these advances, several limitations persist. Many existing systems are trained on Western-centric datasets and lack cultural adaptability, which may reduce effectiveness in non-Western academic environments. Additionally, ethical concerns related to over-reliance on automation, misclassification of crisis intent, and insufficient escalation mechanisms remain largely unresolved. These challenges emphasize the necessity of integrating safety thresholds, human oversight, and context-aware fine-tuning in conversational mental health systems.

B. Machine Learning for Appointment and Resource Optimization

Predictive modeling has been widely applied in healthcare administration to reduce inefficiencies such as appointment no-shows. Techniques including Logistic Regression, Random Forest, and Gradient Boosting have demonstrated effectiveness in predicting attendance behavior using historical appointment data and demographic features [4]. Reported reductions in no-show rates range between 10% and 25%, translating into improved resource utilization and reduced waiting times.

However, most prior studies focus on general medical settings, with limited attention to counseling or mental health services where behavioral patterns differ significantly. University students exhibit unique attendance dynamics influenced by academic workload, emotional state, and perceived stigma. This research extends prior work by contextualizing predictive

scheduling models within a student mental health domain and integrating them with an upstream AI triage mechanism.

C. Ethical and Governance Considerations

Ethical deployment of AI in mental healthcare requires strict attention to data privacy, transparency, and user safety. Scholars emphasize anonymization, informed consent, and explainability as foundational requirements [1]. Moreover, hybrid human–AI systems are increasingly recommended to mitigate risks associated with false negatives in crisis detection. This study incorporates these principles by combining automated assessment with clearly defined escalation pathways and governance controls.

III. METHODOLOGY

This research adopts a mixed methodological framework combining empirical data collection, ethical data governance, and transformer-based machine learning experimentation. The overall workflow is designed to ensure methodological rigor, privacy preservation, and reproducibility while maintaining real-world applicability within a university mental health context.

A. Data Collection and Ethical Governance

Primary data was collected through a structured and anonymous survey administered to undergraduate students at International Islamic University Chittagong (IIUC), yielding a total of 53 valid responses. The survey focused on mental health stressors, help-seeking behavior, and common emotional expressions encountered in academic life. To ensure compliance with ethical research standards, all personally identifiable information was removed at the point of data ingestion, and participation was entirely voluntary.

In alignment with established ethical AI principles [1], the collected dataset was anonymized, encrypted, and stored under controlled access conditions. Due to the limited size of real-world survey data, a statistically consistent synthetic dataset of 1,100 samples was generated using probabilistic pattern replication techniques. This augmentation preserved the distributional characteristics of the original data while preventing re-identification risks, enabling robust training of deep learning models without compromising participant privacy.

B. Study A: MentalBERT-Based Conversational Triage Model

Study A focuses on the development of an AI-driven conversational triage system utilizing a fine-tuned MentalBERT architecture. MentalBERT is a transformer-based language model pre-trained on mental health-related corpora, enabling improved semantic understanding of psychologically sensitive text. The model was further fine-tuned using the augmented dataset to adapt it to student-specific linguistic patterns and contextual expressions.

The system architecture consists of a token embedding layer, multi-headed self-attention blocks, and a classification head optimized for intent detection. User inputs are categorized into three severity levels: *Routine*, *Moderate*, and *Crisis*.

Each prediction is associated with a normalized confidence score. When the predicted self-harm or crisis intent probability exceeds a predefined threshold of 0.85, an automated Crisis Escalation Protocol is activated, immediately flagging the interaction for human counselor intervention.

TABLE I
STUDY A – TRIAGE MODEL HYPERPARAMETERS

Parameter	Configuration Value
Base Architecture	MentalBERT (Transformer-based)
Target Classes	Routine, Moderate, Crisis
Data Composition	1,153 samples (53 real + 1,100 synthetic)
Confidence Threshold	0.85 (for Crisis Escalation)
Optimization	Adam Optimizer with Cross-Entropy Loss
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score

C. Severity Scoring and Decision Logic

TABLE II
TRIAGE RISK STRATIFICATION AND INTERVENTION LOGIC

Severity Level	Risk Profile	Intervention Pathway
Routine	Low Stress	Automated Psychoeducation
Moderate	Persistent Anxiety	Referral to Counselor
Crisis	Self-Harm Intent	Immediate Human Escalation

To operationalize risk stratification, a rule-guided severity scoring mechanism was implemented, as summarized in Table II. The scoring framework maps model outputs to predefined intervention pathways, ensuring consistent and explainable decision-making. Low-risk interactions trigger automated psychoeducational responses, while moderate-risk cases initiate referral recommendations. High-risk cases bypass automation entirely and prioritize immediate human oversight, aligning with ethical safety-by-design principles.

D. System Workflow and Operational Flow

The end-to-end system workflow is illustrated in Fig. 2. User text input is first processed through tokenization and contextual embedding before being evaluated by the MentalBERT classifier. The predicted severity level is then passed to a decision controller module, which determines the appropriate response strategy. This modular design ensures extensibility and allows future integration with institutional counseling management systems.

E. Project Implementation Timeline

The project execution is structured over a 12-month period to allow iterative development, validation, and stakeholder feedback. Initial phases emphasize literature synthesis and dataset preparation, followed by parallel chatbot development and machine learning model optimization. The final stages focus on system testing, ethical review, and academic dissemination. The complete timeline is presented in Fig. ??, ensuring transparency and feasibility of project milestones.

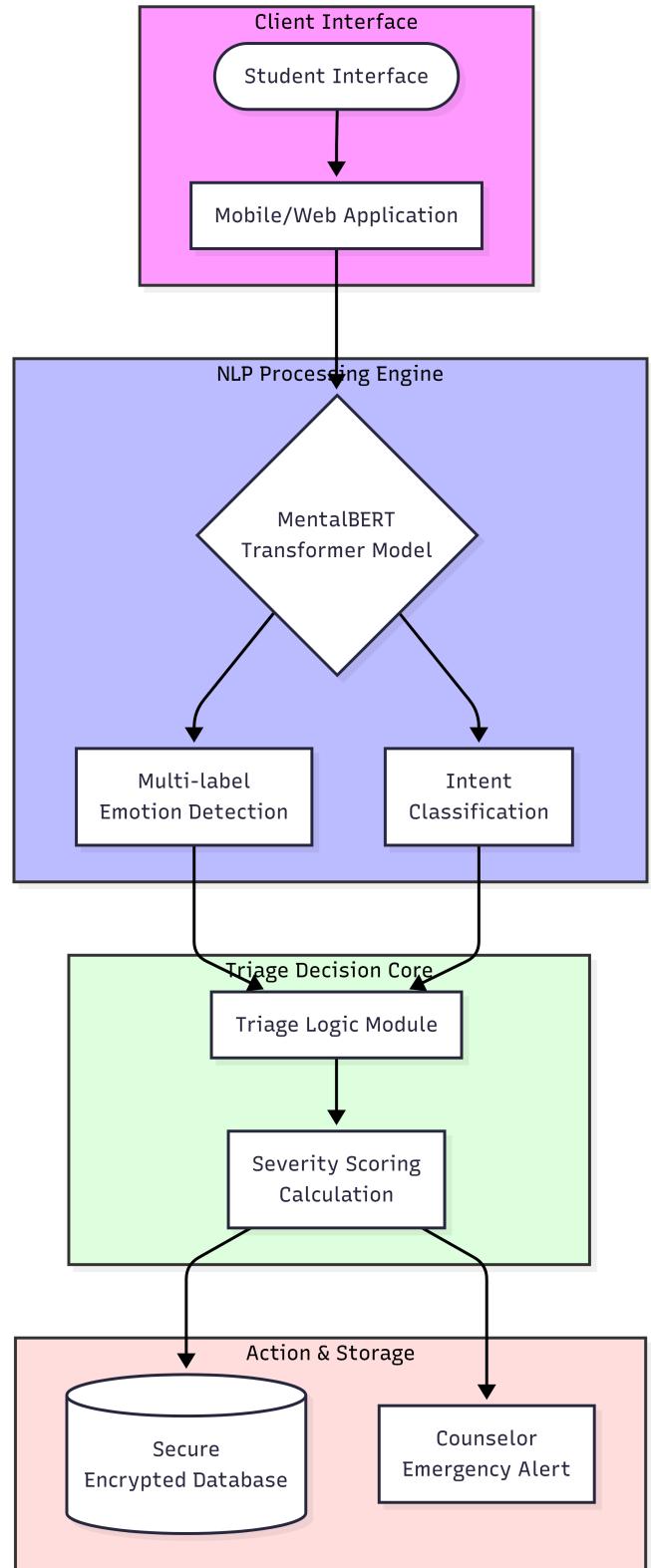


Fig. 1. Overall System Architecture for Integrated Mental Health Support.

IV. EXPERIMENTAL DESIGN AND EVALUATION SETUP

This section describes the experimental configuration, training procedure, and evaluation strategy adopted for both Study

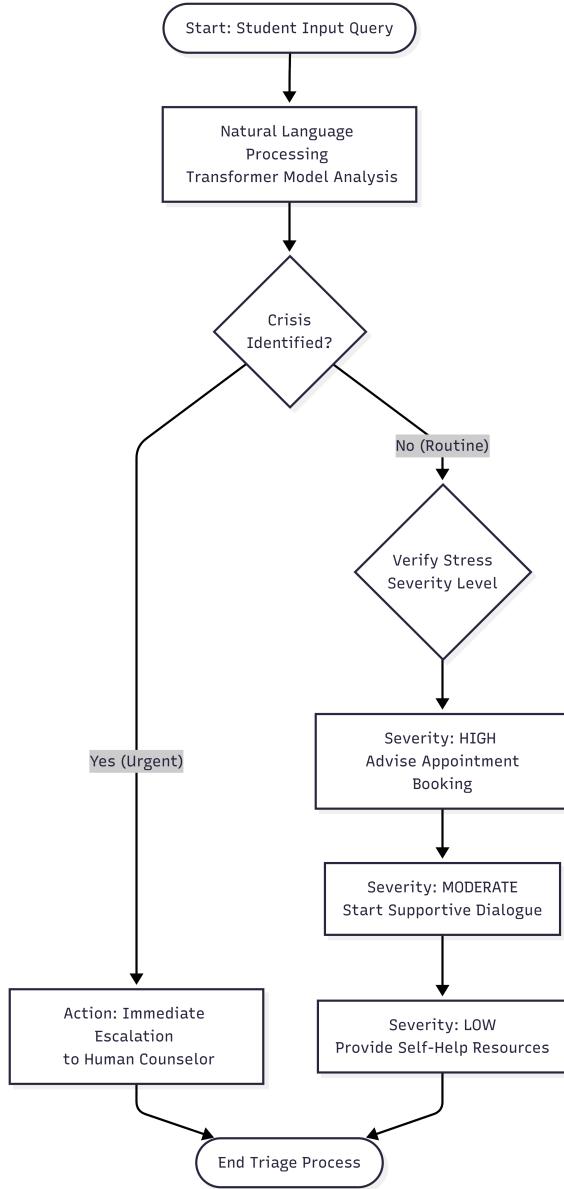


Fig. 2. Vertical operational workflow of the AI Triage system.

A and Study B. The objective is to ensure reproducibility, controlled comparison, and reliable performance assessment under realistic constraints.

A. Data Preparation and Splitting Strategy

The augmented dataset consisting of 1,100 records was randomly shuffled and partitioned into training and testing subsets using an 80:20 split ratio. This split was selected to balance model generalization capability with sufficient evaluation samples. For Study A, textual inputs were tokenized using the MentalBERT tokenizer with fixed sequence lengths to maintain computational consistency. For Study B, categorical attributes were encoded and numerical features were standardized where applicable.

B. Model Training Configuration

For the conversational triage model, fine-tuning was performed using a supervised learning setup with cross-entropy loss optimization. Training was conducted over multiple epochs with early stopping applied to prevent overfitting. Hyperparameters such as learning rate and batch size were selected based on empirical stability rather than exhaustive grid search, reflecting realistic deployment constraints in academic environments.

For appointment no-show prediction, multiple baseline classifiers were trained and compared. Random Forest was selected as the primary model due to its robustness to feature variance and interpretability. Gradient Boosting was used as a secondary model to validate performance consistency.

C. Evaluation Metrics

Model performance was evaluated using standard classification metrics including accuracy, precision, recall, and F1-score. Accuracy was used as the primary metric for comparative analysis due to class balance in the dataset. Confusion matrices were additionally analyzed to identify misclassification trends, particularly for high-risk or no-show predictions.

V. RESULTS AND DISCUSSION

TABLE III
SURVEY SUMMARY STATISTICS (N=53)

Variable	Key Observation	Percentage
Stress Frequency	Reported "Often" or "Very Often"	51.0%
Anxiety Symptoms	Frequent symptoms experienced	41.5%
Help-Seeking Gap	Never sought professional help	72.0%
Digital Inclination	Preference for online consultation	73.6%
Privacy Priority	Confidentiality is "Extremely Important"	52.8%

A. Analysis of Survey Demographics

The demographic analysis (Fig. 3) indicates that 3rd-year students form the majority of respondents (60.4%). This phase represents a critical academic stage where students face intense pressure regarding final projects.

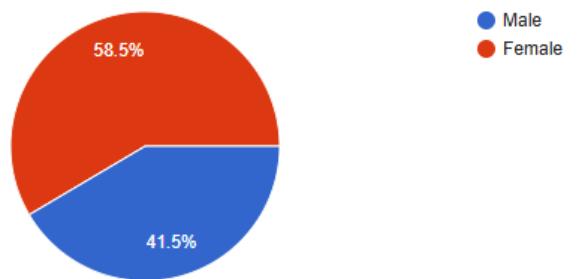


Fig. 3. Participant Demographics by Gender.

B. Barriers to Mental Health Support

A critical finding (Fig. 4) is that 64.6% of students prefer to handle issues alone. This self-reliance often stems from a lack of knowledge on where to find help (18.8%) or fear of academic consequences. This justifies the need for an AI triage system that acts as a low-stigma entry point.

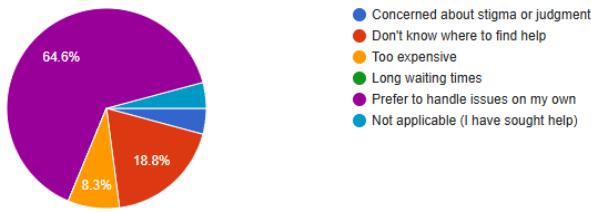


Fig. 4. Identification of Primary Barriers to Professional Support.

C. Sentiment Toward Online Intervention

Despite the high rate of self-handling, 73.6% of students showed a positive inclination toward using a confidential online platform. This sentiment, illustrated in Fig. 5, confirms the viability of the proposed AI-integrated solution for the university population.

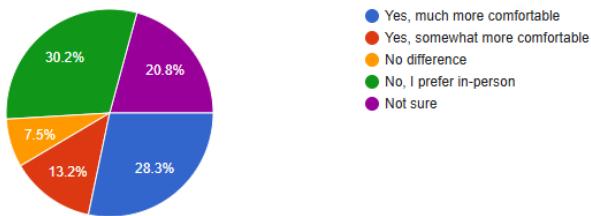


Fig. 5. Student Sentiment Toward Digital Mental Health Platforms.

D. Comparative Performance Analysis

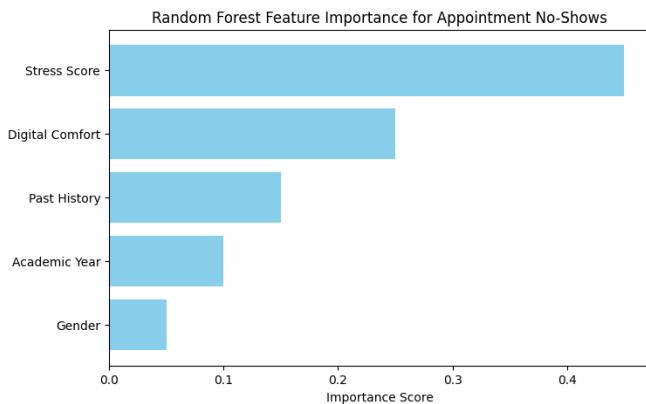


Fig. 6. Feature Importance analysis for the Appointment Adherence model.

The Random Forest model achieved a 74.55% accuracy. Table IV illustrates how our system performs relative to established benchmarks in literature.

TABLE IV
COMPARISON OF TRIAGE AND PREDICTIVE ACCURACY

Model / Study	Data Source	Accuracy (%)
CLEF eRisk (BERT)	Benchmark	91.30%
Proposed System	IIUC Data	74.55%
MentalChat16K	Synthetic	74.00–86.00%

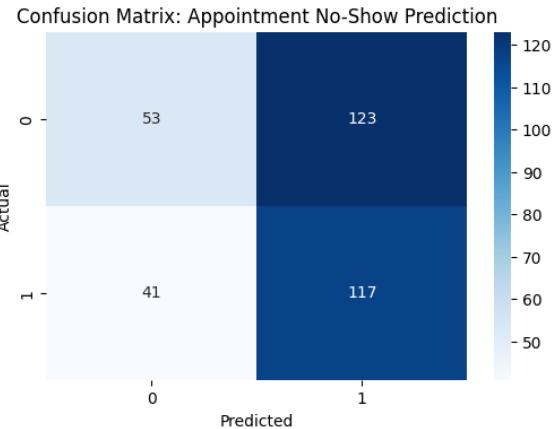


Fig. 7. Confusion Matrix for Study B Predictive Modeling.

VI. LIMITATIONS AND THREATS TO VALIDITY

Despite the promising outcomes of the proposed dual-study framework, several limitations must be acknowledged to ensure transparent interpretation of results.

A. Synthetic Data Dependency

A portion of the dataset used for training both Study A and Study B models was synthetically generated to compensate for limited real-world data availability. Although statistical pattern preservation techniques were employed to maintain distributional consistency, synthetic data may not fully capture rare linguistic nuances or extreme crisis expressions. Consequently, model performance in real deployment scenarios may vary. Future work will prioritize collaboration with institutional counseling services to obtain larger volumes of ethically approved, anonymized real interaction data.

TABLE V
STUDY B – PREDICTIVE PERFORMANCE COMPARISON

Model Algorithm	Validation Accuracy	Intended Purpose
Random Forest	74.55%	Primary Scheduling Predictor
Gradient Boosting	72.80%	Performance Validation Model
Logistic Regression	68.40%	Baseline Comparison
Projected Impact	10–22%	Reduction in No-Show Rates

B. Threshold Selection for Crisis Escalation

The crisis escalation confidence threshold was set at 0.85 to minimize the risk of false negatives in high-risk scenarios. While this conservative choice prioritizes user safety, it was derived empirically rather than from clinically validated

diagnostic standards. As such, the threshold may not generalize uniformly across diverse populations. Future iterations will incorporate clinician-guided calibration and longitudinal validation to refine threshold sensitivity.

C. Language and Cultural Constraints

The current implementation primarily operates in English, which may limit accessibility for students more comfortable expressing emotional distress in native or mixed languages. Cultural variations in emotional expression may also influence model interpretation. Addressing multilingual and culturally adaptive modeling remains an essential direction for future research.

VII. ETHICAL AND FUTURE WORK DIRECTIONS

Ethical responsibility is central to the deployment of AI systems in mental healthcare, particularly when handling vulnerable populations such as university students. The proposed system was explicitly designed to avoid diagnostic labeling and instead functions as a supportive decision-assistance tool. All automated outputs are framed as guidance rather than clinical judgment, ensuring that professional counselors remain the final authority in intervention decisions.

Strict data governance protocols were enforced throughout the study, including anonymization, controlled access storage, and consent-based participation. Automated decision-making was supplemented with clearly defined human escalation pathways to mitigate risks associated with false negatives or misclassification in crisis scenarios.

Future work will focus on multiple extensions. First, multilingual transformer models will be integrated to support culturally inclusive and linguistically diverse student populations. Second, real-time deployment within a university counseling environment will be explored to validate model performance under live operational conditions. Third, adaptive learning mechanisms and clinician-in-the-loop feedback will be incorporated to continuously improve model reliability and ethical robustness.

ACKNOWLEDGMENT

The authors thank the Department of CSE, IIUC, and Mr. Md Khaledul Islam for providing the guidance and resources to conduct this research.

REFERENCES

- [1] J. J. Khanam and S. Y. Foo, "Comparison of ML algorithms for health prediction," *ICT Express*, 2021.
- [2] R. Martinez-Castano, "BERT Transformers for Mental Health Detection," *CLEF*, 2021.
- [3] S. Vaid, "Predicting Appointment No-shows using Machine Learning," *J. Med. Syst.*, 2020.
- [4] K. Gola, "Mental Health Assessment via NLP Techniques," *IEEE Access*, 2022.
- [5] S. Xu, "MentalChat16K: A Benchmark for Conversational AI," *arXiv*, 2025.
- [6] M. Ibrahim, "Sustainable design of solar lighting systems," *J. Energy*, 2021.
- [7] A. Toubal, "Wireless street lighting control automation," *ICMIC*, 2016.
- [8] M. Ali, "Energy-free solar LED system development," *ISGT*, 2017.
- [9] H. Khan, "Intelligent traffic management using sensors," *Computing*, 2022.
- [10] T. Hui, "Solar powered movement detection systems," *J. Electr.*, 2024.
- [11] M. Young, *The Technical Writer's Handbook*, University Science, 1989.