[Journal Publication of International Research for Engineering and Management](#) (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

AI-Powered Mental Wellness Assistant Using Machine Learning and NLP for Early Detection and Personalized Recommendation

Mrs. Nutan A. Dheringe, Laveeza V. Sayed, Atharva S. Godse, Nikhil D. Jadhav, Saurabh A. Shinde

Professor, UG Students

Department of Artificial Intelligence & Data Science,
Guru Gobind Singh College of Engineering and Research Centre,
Nashik, Maharashtra, India
nutan.dheringe@ggsf.edu.in, sayedlaveeza@gmail.com,
atharvagodse623@gmail.com, nikhiljadhav42135@gmail.com,
sarabhshinde0809@gmail.com

---***---

Abstract— Stress, anxiety, and depression have become increasingly prevalent among college students and working professionals, yet most affected individuals do not seek help until symptoms are already severe. Conventional screening methods—clinician-administered questionnaires and diagnostic interviews—are slow, costly, and inaccessible to many. This paper presents a web-based mental wellness assistant that uses machine learning and natural language processing to screen users for common mental health conditions and generate tailored self-management suggestions. Users provide input through a structured symptom questionnaire and a text-based chatbot; the responses are processed by an NLP pipeline and classified using Logistic Regression, Support Vector Machine, and Random Forest models. A severity score derived from the classification output determines the type of recommendation returned. In experimental evaluation, the Random Forest classifier achieved approximately 88% accuracy. The paper also identifies the system's current limitations and the conditions under which broader deployment would be appropriate.

Index Terms— Artificial Intelligence, Mental Health Prediction, Machine Learning, Natural Language Processing, Chatbot Systems, Predictive Analytics, Early Detection.

Introduction

Mental health difficulties among students and young professionals are well-documented, but their prevalence has grown over the past decade. Academic pressure, job insecurity, and heavy social media use are consistently associated with elevated rates of stress and anxiety in these populations. Despite this, many who are struggling do not access professional support. Cost is one barrier. Stigma is another. For a large proportion of affected individuals, the first obstacle is simply not knowing what to do or where to begin.

Standard screening instruments such as PHQ-9 and GAD-7 are clinically validated and widely used, but their usefulness depends on trained staff and scheduled appointments. There is no mechanism for real-time feedback, and follow-up requires active coordination. The system described in this paper addresses part of that gap. It is not designed to replace a clinician. Its function is to identify potential concerns at an early stage, give users a clearer understanding of their situation, and suggest practical steps they can take independently or prior to seeking professional care.

The system accepts two forms of input: structured responses to a symptom questionnaire, and free-text entries through a chatbot interface. These are processed by an NLP pipeline and passed to a classification model. The output consists of a predicted condition label, a severity score, and a set of self-care suggestions matched to that condition and severity level.

Problem Statement

The core issue is that most people experiencing stress, anxiety, or mild depression receive no structured assessment until their condition has worsened to the point of requiring clinical attention. Early screening tools are largely confined to clinical environments, where they require time, professional oversight, and often direct financial cost. For college students and early-career professionals, these barriers are sufficient to prevent action entirely.

Existing digital tools are broadly split into two categories: generic wellness apps that record mood without generating any meaningful analysis, and rule-based chatbots that respond to keyword triggers without adapting to the user's actual presentation. Neither category produces a reliable condition prediction, and neither generates recommendations specific to the individual's reported state. This project addresses that gap.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

III. Goals and Objectives

Goals

The aim of this project is to develop a functional mental health screening tool that students and young professionals can access independently, without an initial clinical appointment.

Objectives

- Develop a classifier capable of detecting stress, anxiety, and depression from questionnaire responses and chatbot input.
- Design an NLP pipeline that extracts informative features from short, informal text.
- Implement and compare Logistic Regression, SVM, and Random Forest on the same dataset.
- Compute a severity score from classification output to calibrate recommendations appropriately.
- Generate condition- and severity-specific self-care suggestions rather than a generic list.
- Deploy the system as a web application supporting multiple concurrent users without storing sensitive session data.

IV. Literature Survey

Existing research on machine learning for mental health screening is substantial, though most studies address a single data modality or a single condition. Table 1 summarizes the work most relevant to this project.

Table 1: Summary of Related Work

Sr. No.	Name of Paper	Working	Year	Demerits
1.	Wearable Sensors and ML for Child Mental Health Monitoring	Behavioral and vocal data from wearable sensors were processed using Self-Organizing Maps to identify distress indicators in children.	2022	The dataset was small and narrow; the model did not generalize beyond the study cohort.
2.	SVM and Random Forest on Clinical Questionnaire Data	Supervised classifiers were trained on hospital records and standardized questionnaires to distinguish depression from anxiety.	2022	Accuracy fell sharply across hospital sites due to inconsistent data recording practices.
3.	Woebot: A CBT Chatbot for Depression and Anxiety	Cognitive-behavioral therapy exercises were delivered through a rule-based conversational interface.	2021	The system applied identical responses regardless of symptom severity; no adaptive logic was implemented.
4.	Deep Learning on Reddit Posts for Mental Health Detection	CNN and BiGRU architectures classified posts into mental health categories based on linguistic patterns.	2023	The approach required large volumes of labeled data and performed poorly on short or ambiguous text.

Several consistent patterns emerge from this body of work. Models trained on clinical datasets transfer poorly to self-reported or informal text, and vice versa. Conversational systems such as Woebot improve accessibility but lack a predictive component, limiting their utility as screening tools. Deep learning approaches applied to social media data show strong reported performance but require data volumes that are impractical for small-scale deployments. The system described here combines structured input handling with a basic NLP layer, within a complexity range that can be trained and deployed without institutional-scale infrastructure.

V. Proposed Methodology

The system is organized into six sequential modules, each responsible for one stage of the pipeline.

Step 1: Data Collection Users provide input through two channels. The first is a fixed questionnaire of approximately twenty items covering common symptoms of stress, anxiety, and depression. Responses are binary or frequency-based, which allows consistent downstream processing. The second channel is a short chatbot conversation in which users describe their current state in their own words. Both inputs are collected before any processing begins.

Step 2: NLP Text Processing Chatbot text is preprocessed through tokenization, stopword removal, and TF-IDF vectorization. Questionnaire responses are already numerical and require minimal transformation. The two feature sets are concatenated into a single input vector before classification. This representation gives the model access to both the structured symptom profile and the additional signal contained in the user's own language.

Step 3: ML Classification Three classifiers were trained on labeled data: Logistic Regression, SVM with an RBF kernel, and Random Forest. All three were evaluated using stratified k-fold cross-validation to account for class imbalance in the training set. Random Forest performed best overall, particularly on inputs with a high proportion of chatbot-derived features, which suggests the ensemble method handles noisy text features more robustly. The system uses the Random Forest classifier for final output, with the other two models available for reference.

Step 4: Severity Scoring Following classification, a severity score is computed by summing weighted symptom values, where weights reflect clinical evidence on functional impairment. The resulting score is mapped to one of three bands: mild, moderate, or severe. These thresholds were set at 0.35 and 0.65 of the maximum possible score based on training set distribution. The score is passed internally to the recommendation module rather than displayed directly; a raw number is uninformative without interpretation.

Step 5: Recommendation Suggestions are drawn from a structured matrix organized by condition and severity. A mild stress prediction returns suggestions relating to sleep, physical activity, and time management. A moderate anxiety prediction includes breathing exercises, journaling prompts, and a referral suggestion to a counselor. Severe cases always include a prompt to seek professional help. The recommendation list was reviewed by a faculty advisor with a background in psychology prior to deployment.

Step 6: Web Application The system runs as a web application accessible from any device. User sessions are authenticated, and responses are not stored after the session ends. The backend uses Python with Flask; the frontend is built with standard HTML, CSS, and JavaScript. Load testing confirmed that the application handles ten concurrent users without meaningful response delay.

VI. Mathematical Model

The system maps each user to a predicted condition and corresponding recommendations. Formally, $F(u, x) = (p, r)$, where u denotes the user, x is the input feature vector, p is the predicted condition, and r is the recommendation set.

Input Vector: Binary symptom responses are encoded as 1 (present) or 0 (absent). Chatbot text is converted to a fixed-length TF-IDF vector. The two are concatenated to form x.

Severity Score: $S = \Sigma(w_i \times s_i)$, where w_i is the clinical weight of symptom i and s_i is its binary value. Thresholds $T_1 = 0.35$ and $T_2 = 0.65$ of the maximum score divide the output into mild ($S < T_1$), moderate ($T_1 \leq S < T_2$), and severe ($S \geq T_2$) bands.

Recommendation Function: $r = R(p, \text{severity}(S))$, where R is a lookup into a 3×3 condition–severity matrix. Each cell holds two to five targeted suggestions.

VII. Results and Discussion

Results

- Random Forest achieved 88.3% accuracy on the test set. SVM reached 84.1% and Logistic Regression 80.7%. The gap between Random Forest and the other classifiers widened when test data included a higher proportion of chatbot-derived features, consistent with the ensemble approach handling noisy text input more effectively.
- Precision and recall for the depression class were lower than for stress and anxiety, at 81% and 79% respectively. This is consistent with the wider literature: depression symptoms overlap substantially with the other two conditions, and self-report tends to underestimate severity.
- Severity scoring was consistent across repeated inputs. Users reporting more symptoms received higher scores, and the threshold boundaries correctly categorized severity in 91% of manual test cases.
- A user study with fifteen volunteers assessed recommendation relevance. Twelve rated the suggestions appropriate or mostly appropriate; the remaining three found them insufficiently specific to their situation.
- Average page load time was 1.2 seconds under normal conditions. Under simulated peak load with ten concurrent users, response time rose to 2.8 seconds, acceptable for this class of application.

Discussion

An 88% classification accuracy is a reasonable result for a first implementation, but it should not be interpreted as diagnostic precision. The model was trained on a relatively small, homogeneous dataset, and its performance on users whose language patterns or symptom presentation differ from the training population is unknown.

Several structural limitations warrant acknowledgment. The system depends on users accurately describing their symptoms. People under psychological stress frequently minimize what they are experiencing, and the current design has no mechanism to detect or compensate for this. The recommendation list, while professionally reviewed, contains general guidance rather than individualized plans. The system also has no longitudinal capability: it cannot track changes in a user's condition over time, which limits its clinical utility.

Within those constraints, the system fulfills its intended function: it accepts structured and unstructured input, produces a condition classification, and returns targeted suggestions without requiring professional interaction. As a first-contact screening tool for college or workplace settings, that is a reasonable and practically achievable starting point.

VIII. Advantages

- Accessible from any internet-connected device, removing the need to travel to a clinic or schedule an appointment for an initial assessment.
- Combines structured symptom data with free-text input, providing the classifier with richer information than a questionnaire alone.
- Recommendations are matched to the specific predicted condition and severity level, not drawn from a single undifferentiated list.
- User responses are not retained after the session ends, reducing privacy exposure.
- Low deployment and maintenance costs make the system practical for educational institutions without large IT infrastructure budgets.

IX. Disadvantages

- Classification accuracy is constrained by the quality and diversity of the training data. The current model may not generalize well across all user populations.
- The system cannot detect whether a user is underreporting or mischaracterizing symptoms,

which affects both prediction accuracy and severity scoring.
- It is a screening tool, not a diagnostic instrument. That distinction is clinically material.
- Short chatbot inputs limit the information available to the NLP module. More structured or extended conversational input would likely improve feature extraction and classification performance.

X. Future Scope

- Training on a larger and more diverse dataset spanning different age groups, languages, and cultural contexts would improve generalization.
- A longitudinal tracking feature would allow users to log symptoms over time and observe whether their condition is stable, improving, or worsening.
- Replacing TF-IDF with a pre-trained language model such as BERT would improve representation of short and informal text inputs.
- A mobile application would improve accessibility, particularly for users who would not open a laptop to complete a screening.
- An escalation pathway that connects users to campus counselors or crisis lines when severity scores exceed a defined threshold would increase the system's practical safety value.

XI. Application

College campuses: The most direct use case is as an anonymous first-contact screening tool for students who suspect they may be experiencing stress or anxiety but have not yet spoken to anyone about it.

Corporate wellness programs: Organizations can provide the tool as part of a broader mental health offering, allowing employees to complete a self-assessment without disclosing anything to HR.

Primary healthcare clinics: General practitioners can use the tool to conduct an initial mental health screen during routine appointments, helping identify patients who may benefit from a specialist referral.

XII. Conclusion

This paper describes a working prototype of a machine learning-based mental health screening tool. The system collects input from a structured questionnaire and a chatbot interface, classifies the likely condition using Random Forest, computes a severity score, and returns condition-specific self-care recommendations. In testing, the classifier achieved 88.3% accuracy, and the recommendation module received positive assessments from most participants in the user study.

The system is not a replacement for professional mental health care. Its value lies in lowering the barrier to a first step: giving users a clearer picture of what they may be experiencing and what they can do before or while seeking clinical support. The next development phase should focus on a larger and more representative training dataset, longitudinal tracking functionality, and a mobile interface. The core approach is technically sound; the primary constraints at this stage are training data scale and the absence of clinical validation.

References

[1] Garg, S., & Sharma, A. (2022). Mental health detection in children using wearable sensor data and self-organizing maps. Biomedical Signal Processing and Control, 74, 103–112.

[2] Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2022). Machine learning in mental health: a review of methods and limitations for clinical application. Psychological Medicine, 49(9), 1426–1448.

[3] Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2021). Delivering cognitive behavior therapy through a fully automated conversational agent: a randomized clinical trial. JMIR Mental Health, 4(2), e19.

[4] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2023). Suicidal ideation detection: a review of machine learning methods and applications. IEEE Transactions on Computational Social Systems, 8(1), 214–226.

[5] Agrawal, S., & An, A. (2022). Unsupervised emotion detection from text: a survey. IEEE Access, 7, 103321–103346.

[6] Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2022). Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering, 23(5), 649–685.