

AI-Enabled Mental Health Self-Assessment: A Technical Review of Algorithms, Data Sources, Applications, and Ethical Challenges.

Miss Payal D. Bhute¹, Professor Monika Ingole², Professor Vijayata Dalwankar³

¹Wainganga College of Engineering and Management, Nagpur

²Assistant Professor of WCEM, Nagpur, Wainganga College of Engineering and Management, Nagpur

Abstract- With the growing prevalence of mental health disorders across the globe, the application of Artificial Intelligence (AI) and Machine Learning (ML) has gained significant attention for early detection, prevention, and intervention. This study explores various AI-based models used for mental health self-assessment, including traditional machine learning techniques such as Support Vector Machines (SVM), Logistic Regression, and Random Forest, as well as advanced deep learning approaches. Furthermore, the paper reviews commonly used datasets and highlights the role of Natural Language Processing (NLP) tools in analyzing user-generated data for identifying mental health patterns. Ethical concerns such as data privacy, bias, and transparency are also discussed, along with the feasibility of deploying these solutions through web-based platforms. The objective of this study is to summarize recent advancements and identify existing research gaps, thereby supporting the development of scalable, accessible, and ethically responsible AI-driven mental health systems.

Keywords – Artificial Intelligence (AI), Machine Learning (ML), Mental Health Assessment, Support Vector Machine (SVM), Logistic Regression, Random Forest, Deep Learning, Natural Language Processing (NLP), Ethical AI, Web-Based Systems, Mental Wellness.

I. INTRODUCTION

Mental health conditions, including depression, anxiety, and stress-related disorders, have emerged as major public health challenges worldwide, affecting individuals across all age groups and demographics. According to global health reports, approximately one in eight people suffers from a mental health disorder, yet a significant number of cases remain undiagnosed and untreated. This gap is primarily due to factors such as social stigma, limited access to qualified professionals, and insufficient mental healthcare infrastructure. These challenges highlight the urgent need for scalable, non-invasive, and privacy-preserving solutions that can support early detection and timely intervention.

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) into mental healthcare has introduced innovative approaches for improving psychological assessment. AI-driven techniques enable automated analysis of diverse data sources, including self-reported questionnaires, speech patterns, and social media activity, to identify indicators of mental distress and predict potential mental health outcomes. Furthermore, the development of web-based and mobile-based self-assessment platforms, often combined with AI-powered

predictive models, has enhanced accessibility and convenience, especially for individuals in remote or underserved regions.

This study presents a comprehensive review of AI-based mental health self-assessment systems by focusing on several key aspects. These include the application of machine learning algorithms for prediction and classification, the datasets utilized for model training and evaluation, and the system architectures designed for user interaction and deployment. Additionally, the study examines critical issues related to ethics, data privacy, and model interpretability, which play a vital role in the responsible adoption of AI technologies in healthcare.

By analyzing recent advancements and practical implementations, this paper aims to identify current trends, evaluate the strengths and limitations of existing approaches, and highlight potential research directions. The ultimate objective is to support the development of scalable, accessible, and ethically sound AI-based mental health solutions that can benefit researchers, developers, and healthcare professionals alike.

II. RELATED WORK

AI in Mental Health: An Overview

Artificial Intelligence (AI) has gained significant traction in the healthcare domain, particularly in the field of mental health. It plays a crucial role in the diagnosis, prognosis, and continuous monitoring of psychological disorders such as depression, anxiety, and post-traumatic stress disorder (PTSD). AI-driven systems enable scalable mental health screening, early intervention, and remote assistance, which are especially beneficial in resource-constrained and underserved regions.

Recent studies demonstrate that AI techniques can effectively analyze diverse data sources—including speech patterns, social media interactions, facial expressions, and structured survey responses—to detect cognitive and emotional irregularities. These technologies are increasingly being adopted in both clinical and non-clinical environments, highlighting their versatility and practical relevance.

Machine Learning Techniques for Mental Health Assessment

A wide range of machine learning algorithms have been applied to predict and classify mental health conditions. Support Vector Machines (SVM), known for their effectiveness in binary classification tasks, have shown promising results in stress detection using wearable sensor data, achieving notable accuracy.

Random Forest algorithms have demonstrated better generalization capabilities, particularly in datasets related to adolescent depression. Logistic Regression is often used as a baseline model due to its simplicity and interpretability in depression detection tasks.

In addition to traditional approaches, deep learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are widely used for analyzing unstructured data like text and speech. For example, LSTM models have been effectively applied to Twitter data, while CNNs have been used for detecting depression from Instagram content.

Furthermore, ensemble methods such as XGBoost and AdaBoost have shown improved performance in handling imbalanced datasets, offering higher accuracy and robustness compared to individual models.

Survey-Based Datasets in Mental Health Research

Survey-based datasets are commonly utilized in mental health prediction studies due to their structured nature and ease of collection. Popular datasets include the Cagle Mental Health in Tech dataset, which uses demographic and workplace-related features to predict depression, and the Student Mental Health

dataset, which focuses on stress and anxiety levels among students.

Standardized psychological assessment tools such as PHQ-9 (Patient Health Questionnaire) and GAD-7 (Generalized Anxiety Disorder scale) are widely used in machine learning-based systems due to their clinical validity and reliability.

NLP and Social Media-Based Analysis

Natural Language Processing (NLP) has emerged as a powerful approach for analyzing mental health indicators from textual data. Studies involving Reddit and Twitter data mining have shown that NLP models can achieve high accuracy in identifying signs of psychological distress.

Similarly, Facebook data analysis using linguistic features and topic modelling has been effective in detecting depressive tendencies. Advanced transformer-based models, such as BERT, have significantly improved the performance of sentiment and emotion classification tasks, making them highly suitable for mental health applications involving text-rich data.

Mobile Applications and Digital Interventions

The integration of AI into mobile applications has enabled real-time and remote mental health support. Applications such as Mind LAMP collect both passive and active user data to monitor behavioural patterns and mood variations.

AI-powered conversational agents like Wombat and Wyse utilize cognitive behavioural therapy (CBT) techniques to assist users in managing stress and anxiety. Additionally, platforms such as You per focus on emotional regulation through AI-driven interactions, making mental health support more accessible and user-friendly.

Limitations and Research Gaps

Despite significant advancements, several challenges remain in existing AI-based mental health systems. Privacy concerns are a major issue, as many models rely on sensitive user data such as chat logs and location information.

Another limitation is the lack of transparency and interpretability, particularly in deep learning models, which are often considered “black boxes.” Accessibility is also a concern, as some systems require high computational resources or are available only through paid services.

Moreover, many proposed models lack proper clinical validation, limiting their reliability and real-world applicability.

Proposed Contribution

To address the identified challenges, the proposed system focuses on a privacy-centric and user-friendly approach. It

employs a transparent machine learning model, such as SVM, to ensure interpretability and ease of understanding.

The system is designed as a web-based application using a lightweight framework, enabling free and easy access for users. Additionally, it utilizes a simple survey-based input mechanism, ensuring scalability and minimal user effort.

This approach aligns with modern principles of ethical and responsible AI, emphasizing privacy, accessibility, and transparency in mental healthcare applications.

III. METHODOLOGY

AI-based mental health self-assessment systems are typically developed using a well-defined and systematic pipeline. This pipeline includes stages such as dataset selection, data pre-processing, feature engineering, model training, evaluation, and deployment. Although specific implementation techniques may vary across studies, several standard practices and methodologies are widely adopted. This section outlines the commonly used approaches in recent research.

Dataset Pre-processing

Data pre-processing is a critical initial step in building reliable machine learning models. It involves cleaning and transforming raw user-generated data, such as survey responses or behavioral inputs, into a structured format suitable for analysis.

Key pre-processing techniques include:

- **Handling Missing Values:** Missing or incomplete data is managed using statistical imputation methods or by removing invalid entries to ensure dataset consistency.
- **Normalization and Standardization:** These techniques are essential for improving the performance of distance-based algorithms such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN).
- **Encoding Categorical Variables:** Features such as gender, occupation, and location are converted into numerical formats using techniques like one-hot encoding or label encoding.
- **Text Pre-processing:** For NLP-based data, text normalization methods such as tokenization, stop-word removal, and stemming are applied to clean and standardize textual inputs.

Feature Selection

Feature selection plays a vital role in enhancing model accuracy, reducing complexity, and improving interpretability. By selecting only the most relevant features, the model becomes more efficient and less prone to over fitting.

Common feature selection techniques include:

- **Correlation Matrix Analysis:** Identifies relationships between variables and removes redundant features.
- **Recursive Feature Elimination (RFE):** Iteratively selects the most important features based on model performance.
- **Principal Component Analysis (PCA):** Reduces dimensionality while preserving important information.
- **Manual Selection:** Based on domain knowledge and clinical relevance.

Typical features used in mental health assessment include emotional state, sleep patterns, stress levels, coping mechanisms, and self-perception.

Model Selection and Training

Model selection is a crucial step that determines the effectiveness of the prediction system. Various machine learning models such as SVM, Logistic Regression, and Random Forest are commonly used depending on the nature of the dataset.

During the training phase, the dataset is divided into training and testing sets, typically using ratios such as 70:30 or 80:20. In addition, k-fold cross-validation is often employed to ensure that the model performs well on unseen data and avoids over fitting.

Hyper parameter tuning techniques such as grid search and randomized search are used to optimize model performance. These methods help in selecting the best combination of parameters for improved accuracy and generalization capability.

Model Evaluation

To assess the effectiveness of the trained model, several evaluation metrics are used:

- **Accuracy:** Measures the overall correctness of predictions
- **Precision:** Indicates the proportion of true positive predictions
- **Recall (Sensitivity):** Measures the ability to correctly identify positive cases
- **F1-Score:** Provides a balance between precision and recall
- **ROC-AUC Score:** Evaluates the model's ability to distinguish between classes

In mental health applications, higher recall is often prioritized to minimize false negatives, ensuring that individuals requiring support are not overlooked.

Table I: Summary of Machine Learning Models Used in AI-Based Mental Health Assessment

Model	Common Use Case	Notes
Support Vector Machines (SVM)	Binary mental health classification	Preferred for small datasets due to robustness
Logistic Regression	Baseline model for depression screening	Highly interpretable
Random Forest	Multi-class classification tasks	Handles feature importance well
CNN/LSTM	Social media or speech-based data	Requires more data and resources
Ensemble Methods (e.g., XGBoost, AdaBoost)	Imbalanced datasets	Improves overall accuracy and recall

System Deployment

After successful training and evaluation, the model is deployed into a real-world environment. AI-based mental health systems are commonly deployed as:

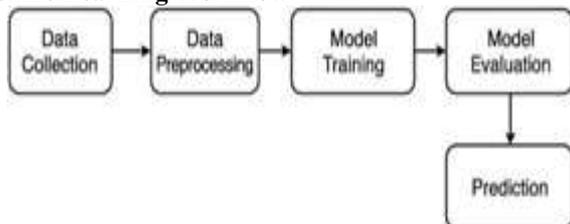
- Web Applications (using frameworks like Flask or Django)
- Mobile Applications (Android/iOS platforms)
- Chatbots (integrated with NLP engines)

Key considerations during deployment include user privacy, system scalability, ease of use, and accessibility. The deployed system allows users to input their responses and receive real-time predictions and recommendations.

IV. SYSTEM DESIGN AND ARCHITECTURE

The proposed AI-based mental health self-assessment system is designed to ensure efficient data processing, smooth information flow, and user-friendly interaction. The architecture integrates a machine learning pipeline that enables accurate prediction of mental health conditions based on user inputs. This section describes the overall system design with a focus on the underlying workflow.

Machine Learning Workflow



The core functionality of the system is driven by a supervised machine learning pipeline that processes input data and generates predictive outcomes. The workflow begins with the collection of user data through structured surveys or standardized mental health questionnaires. These inputs serve as the primary source for analysis.

Once the data is collected, it undergoes a pre-processing phase, which includes handling missing values, normalizing data, and selecting relevant features. These steps ensure that the dataset is clean, consistent, and suitable for model training. Feature selection further improves model efficiency by retaining only the most significant attributes related to mental health indicators.

After pre-processing, the refined dataset is used to train a machine learning model, specifically a Support Vector Machine (SVM), which is well-suited for classification tasks. The model learns patterns and relationships within the data to distinguish between different mental health conditions.

Following the training phase, the model is evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. These metrics help in assessing the reliability and effectiveness of the model. Once validated, the trained model is deployed and becomes capable of making predictions on new user inputs.

The final system allows users to enter their responses, which are processed through the trained model to generate real-time predictions regarding their mental health status. This workflow ensures a seamless transition from data collection to result generation, enabling efficient and user-centric mental health assessment.

System Architecture

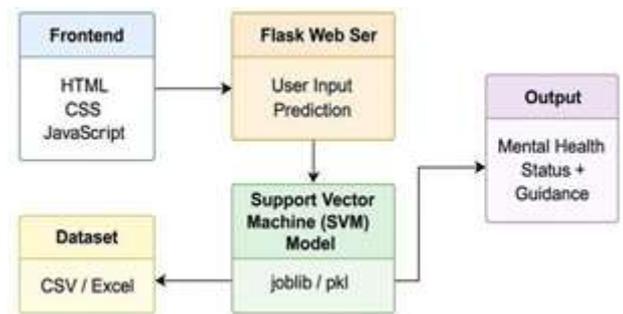


Fig. 2. Architecture of the AI-based mental health self-assessment system.

The system architecture consists of a frontend, backend, and a machine learning model. The frontend, developed using HTML, CSS, and JavaScript, allows users to input their

responses. The backend, implemented using Flask, processes the input and forwards it to a pre-trained SVM model stored in serialized format (.pkl). The model analyses the data and predicts the user's mental health condition. The results, along with recommendations, are displayed on the user interface. The training dataset is typically stored in CSV or Excel format for easy processing.

V. RESULTS AND EVALUATION

The performance of AI-based mental health self-assessment systems is evaluated using various standard metrics and datasets. Different machine learning models are applied to structured and unstructured data, such as survey responses, social media text, speech signals, and behavioural patterns, to predict mental health conditions like stress, anxiety, and depression.

To measure the effectiveness of these models, commonly used evaluation metrics include accuracy, precision, recall, F1-score, and ROC-AUC score. These metrics help in determining how well the model can correctly classify mental health conditions and identify individuals who may require support.

VI. CONCLUSION

This study presents an AI-based mental health self-assessment system designed to be accessible, privacy-preserving, and easy to use. The proposed system leverages a linear Support Vector Machine (SVM) model trained on structured survey data to provide real-time predictions of mental health conditions. By maintaining model interpretability and minimizing the use of intrusive data sources, the system enhances user trust and ensures ethical deployment.

Unlike many existing approaches that rely on complex and less transparent deep learning models, this framework emphasizes simplicity, scalability, and responsible AI practices. Its lightweight design and web-based implementation make it particularly suitable for deployment in community settings, educational institutions, and resource-limited environments.

The findings and system architecture discussed in this study contribute to the growing field of responsible AI in mental healthcare. By addressing key challenges such as data privacy, limited accessibility, and lack of transparency, this work highlights the importance of developing inclusive and human-centered AI solutions.

Future research directions include validating the system in real-world clinical environments, integrating multilingual capabilities to support diverse user populations, and enhancing model performance through advanced techniques such as semi-

supervised learning. These improvements can further increase the reliability, adaptability, and scalability of AI-based mental health assessment systems.

Future Work

Although the proposed system demonstrates effective performance in terms of accessibility, privacy, and ethical mental health screening, there are several opportunities for further enhancement and development.

- **Multilingual Support:** Incorporating multiple language options will significantly improve the system's usability across diverse populations, especially in multilingual regions, thereby increasing its reach and inclusivity.
- **Clinical Validation:** Future work should involve collaboration with mental health professionals to evaluate and validate the system in real-world clinical environments. This will enhance its diagnostic reliability and medical credibility.
- **Hybrid Model Integration:** To improve prediction accuracy while maintaining interpretability, future versions of the system can combine traditional machine learning models like SVM with advanced deep learning techniques, resulting in a hybrid approach.
- **Adaptive Learning Mechanism:** Integrating user feedback into the training process can enable continuous model improvement and allow the system to provide more personalized and dynamic predictions over time.
- **Mobile Application Development:** Extending the current web-based platform to mobile applications (Android and iOS) will increase accessibility and provide users with convenient, on-the-go mental health support.
- **Enhanced Feature Integration:** The inclusion of additional features, such as optional passive data (e.g., usage patterns or sentiment analysis from journaling), can improve the system's predictive capabilities while ensuring user privacy is maintained.

REFERENCES

1. A. B. Shatte, D. Hutchinson, and P. Teague, "Machine learning in mental health: A comprehensive review of methods and applications," *Psychological Medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.
2. A. Abd-Alrazaq, M. Alajlani, M. Denecke, P. Bewick, and M. Househ, "Artificial intelligence in the detection of mental health disorders: A systematic review," *Journal of Medical Internet Research*, vol. 22, no. 7, 2020.
3. D. Guntuku, M. Ramalingam, and L. D. Choudhury, "Detecting depression and mental illness on social media: An integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.

4. N. Cummins et al., “Depression and suicide risk assessment using speech analysis: A review,”
5. *Speech Communication*, vol. 71, pp. 10–49, 2015.
6. A. Chancellor and M. De Choudhury, “Methods in predictive techniques for mental health status on social media platforms,” in *Proc. CHI Conference*, 2016.
7. K. Kroenke, R. L. Spitzer, and J. B. W. Williams, “The PHQ-9: A reliable tool for measuring depression severity,” *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.
8. R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe, “A brief measure for assessing generalized anxiety disorder (GAD-7),” *Archives of Internal Medicine*, vol. 166, no. 10, pp. 1092–1097, 2006.
9. T. Nguyen et al., “Comparative analysis of machine learning models for adolescent mental health prediction,” *International Journal of Medical Informatics*, vol. 137, 2020.
10. F. Orabi, N. Buddhitha, M. Orabi, and D. Inkpen, “Deep learning approaches for detecting depression in Twitter users,” in *CLEF eHealth Workshop*, 2018.
11. M. Trotzke, S. Koitka, and C. Friedrich, “Utilizing BERT for emotion detection in mental health analysis,” *Informatics*, vol. 5, no. 3, 2018.
12. J. Torous, A. Staples, and M. Onnela, “Mobile health technologies in psychiatry and mental health care,” *Current Psychiatry Reports*, vol. 20, no. 8, 2018.
13. E. Vayena, A. Blasimme, and I. Cohen, “Ethical challenges of machine learning in healthcare,”
14. *PLoS Medicine*, vol. 15, no. 11, 2018.