Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 05, May-2025

A Review of AI-Based Self-Assessment Systems for Mental Health: Models, Datasets, Applications and Ethical Concerns

Shresth Gautam, Divyendra Singh, Anisha Chaudhary,
Under Graduate Student, Department of Computer Science and Applications,
School of Engineering and Technology, Sharda University, Greater Noida 201310, India.

Dr. Anand Pandey, Dr. Md.Tajammul
Associate Professor, Department of Computer Science and Applications,
School of Engineering and Technology, Sharda University, Greater Noida 201310, India

Abstract— **Given the increasing mental health issues worldwide, many have adopted artificial intelligence (AI) and machine learning (ML) for prevention and intervention. We examine different AI models for mental health self-assessment that are driven by artificial intelligence including Support Vector Machines (SVM), Logistic Regression, Random Forest, and deep learning. We review popular datasets, NLP-based tools, ethical issues and the prospects for web-based deployment. Thus, we aim to sum up the most recent findings and point out research gaps that could drive the design and development of scalable, accessible, and ethically sound AI solutions for mental wellness.**

Keywords—**Artificial Intelligence, Machine Learning, Mental Health Assessment, Support Vector Machine, Logistic Regression, Random Forest, Deep Learning, Natural Language Processing, Ethical AI, Web-Based Deployment**

1. INTRODUCTION

Conditions of mental health, including depression, anxiety, and stress-related disorders, have become important issues of public health affecting populations regardless of age or demographics. Today 1 in 8 people worldwide live with mental health disorder, and millions remain ill and untreated for reasons of social stigma, limited access to professional care and inadequate mental health infrastructure, according to the World Health Organization. These challenges call for scalable, non-invasive, and privacy-preserving approaches, which will enable early diagnosis and timely interventions.

Recently, the field of mental health has seen an uptick in the use of artificial intelligence (AI) and machine learning (ML), bringing with it novel ways to automate and improve psychological assessment. AI and mental health tech — from self-reported surveys to speech recognition to social media activity — have found promising ways to identify symptoms of mental distress and possible mental health outcomes. The emergence of web and mobile-driven self-assessment tools, frequently in tandem with AI-fused predictive frameworks, has also contributed to a more convenient means of detecting and addressing mental health concerns, particularly in under-resourced or isolated communities.

This review provides a thorough examination of current AI-based mental health self-assessment systems by focusing on:

- Prediction and classification with machine learning algorithms commonly used for it,
- All datasets used in model development,
- Platforms and architecture for user interaction implementation
- And the ethical, privacy, and interpretability implications that come from having these technologies.

This study highlights current trends, assesses the advantages and disadvantages of different approaches, and suggests possible avenues for further research by combining findings from recent literature with practical applications. The objective is to encourage the creation of inclusive, moral, and clinically useful self-assessment tools while educating academics, developers, and medical practitioners about the changing field of AI in mental health.

2. RELATED WORK

2.1 Introduction to AI in Mental Health

In the medical field, particularly mental health, artificial intelligence (AI) has gained widespread acceptance. It helps with the diagnosis, prognosis, and monitoring of disorders like PTSD, anxiety, and depression [1], [2]. Scalable mental health screening, early intervention, and remote support are made possible by AI and are particularly important in underprivileged areas. By examining speech [3], social media data [4], facial expressions [5], and survey responses [6], studies have demonstrated AI's capacity to identify cognitive and emotional abnormalities. Both clinical and non-clinical settings are now utilizing these technologies [7].

2.2 Machine Learning Approaches in Mental Health Assessment

Several machine learning methods have been applied to the prediction of mental health: With wearable data, Support Vector Machines (SVM), which are used for binary classification, were able to detect stress with 83% accuracy [1].

In datasets pertaining to adolescent depression, Random Forest was found to generalize more effectively than SVM [8].

A common baseline approach for detecting depression is logistic regression [9].

- Deep Learning: Used with unstructured inputs such as voice and text (CNNs, LSTMs). LSTMs were utilized for Twitter analysis [11], and CNNs were utilized for Instagram depression identification [10].
- In unbalanced datasets, ensemble models (such as XGBoost and AdaBoost) perform better than conventional models [12].

2.3 Use of Survey-Based Datasets

Research on mental health prediction frequently uses survey datasets:

- Kaggle Mental Health in Tech (2014): Predicts depression by using demographic and employment data [9].
- Student Mental Health Dataset: Used to predict students' levels of stress and anxiety [6].

ML-based mental health systems use the widely validated PHQ-9 and GAD-7 [13], [14].

2.4 NLP-Based Tools and Social Media Analysis

Techniques for natural language processing (NLP) have showed a lot of promise:

- Reddit/Twitter Mining: The accuracy of NLP models trained on r/depression and r/anxiety forums in identifying discomfort was greater than 80% [4], [15].
- Facebook Analysis: Depression has been identified using linguistic cues and topic modeling [16].
- Sentiment & Emotion Detection: In tasks involving the classification of mental health, transformer models such as BERT have demonstrated great precision [17].

2.5 Mobile Apps and Digital Interventions

Digital tools offer real-time, remote mental wellness support:

- MindLAMP: Gathers passive and active information to monitor mood [18].
- Woebot and Wysa are AI-powered cognitive behavioral therapy chatbots that reduce anxiety and stress [7], [19].
- Youper: an AI-powered app for controlling emotions [20].

2.6 Gaps in Existing Systems

Gaps still exist despite progress:

- Privacy Risks: AI models frequently use invasive data, such chat records or GPS [21].
- Transparency: Interpretability is sometimes lacking in deep learning models [22].

Accessibility: Certain systems demand a lot of resources or are monetized [7].

- Clinical Validation: Real-world testing is lacking in many models [23].

2.7 Our Contribution in Context

These issues are addressed by our system via:

- Privacy-First Design: No tracking or logging of cloud data.

Transparent SVM Modeling: Simple to understand and describe.

- Accessible Web Deployment: a Flask-based user interface that is freely available.
- Scalable Input Mechanism: Evaluation only needs a quick survey.

This strategy is in line with contemporary ethical and responsible AI in healthcare ideas [24].

3. METHODOLOGY

AI-powered mental health self-assessment systems are usually developed using a structured pipeline that includes feature engineering, dataset selection, preprocessing, model training, evaluation, and system deployment. Although the precise techniques used in different studies differ, a number of best practices and common tendencies have been identified. A comparison of the approaches used in recent works is given in this section

3.1 Dataset Preprocessing

Cleaning and preprocessing user-generated survey or behavioral data is the first step in the majority of research. Among the methods are:

- Managing Null Values: Statistical techniques are frequently used to eliminate or impute missing entries.

Standardization and normalization are crucial for distance-based algorithms such as SVM and KNN.

- Categorical Features Encoding: Features such as location, occupation, or gender are encoded using one-hot or label encoding.

Text normalization is another tool that some academics utilize when working with NLP data, including postings from social media.

3.2 Feature Selection

Selecting features is essential for enhancing the interpretability and performance of the model. Typical methods consist of:

- Correlation Matrix Analysis
- RFE, or recursive feature elimination

Clinical relevance-based manual selection and Principal Component Analysis (PCA)

Emotional state, coping strategies, sleep quality, and subjective self-worth are among the characteristics that are commonly evaluated.

3.3 Model Selection and Training

Thorough training and assessment are essential for any model. During the training phase, the dataset is divided into training and testing subsets, usually using a 70/30 or 80/20 split, or k-fold cross-validation methods. By evaluating the model's performance on unseen data, these techniques aid in determining how generalizable it is.

A crucial first step in maximizing any model's performance is hyperparameter tuning. Two popular techniques for choosing the ideal set of hyperparameters are grid search and randomized search. Whereas randomized search examines a predetermined amount of random choices, grid search attempts every possible combination of a given set of hyperparameters. When combined with cross-validation, both techniques make sure that the selected hyperparameters don't cause overfitting, enhancing the model's resilience and predictive ability.

3.4 Evaluation Metrics

Performance is commonly measured using:

- Accuracy
- Precision
- Recall (Sensitivity)
- F1-score
- Confusion Matrix
- ROC-AUC Score (especially for binary classifiers)

Researchers emphasize **high recall** to minimize false negatives, which is crucial for mental health applications.

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

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

3.5 System Deployment and User Interface

A number of systems have been deployed as:

- Web applications (using Flask, Django)
- Mobile apps (Android/iOS with TensorFlow Lite)
- Chatbots (Woebot, Wysa using NLP engines)

These systems prioritize:

- User privacy (e.g., no data logging)
- Accessibility (mobile responsiveness, multi-language support)
- Ease of use (simple UI, clear guidance)

4. SYSTEM DESIGN AND ARCHITECTURE

Real-time analysis, smooth data flow, and intuitive user engagement are all features of the AI-based mental health self-assessment system. In order to make predictions about mental health condition, this section explains the system architecture, emphasizing the underlying machine learning pipeline.

4.1 Machine Learning Workflow

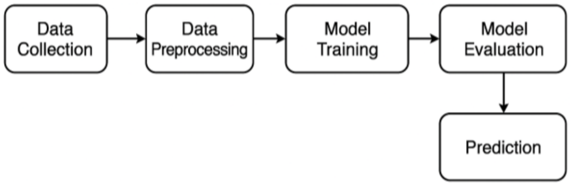

Fig. 1. Machine learning pipeline used in the mental health prediction system.

A supervised machine learning pipeline that analyzes input data and produces predictive results powers the system's intelligence. The pipeline starts with the gathering of information from surveys or standardized questionnaires pertaining to mental health, as shown in Fig. 1. Preprocessing of this data includes feature selection, normalization, and management of missing values.

In order to identify patterns and relationships in the responses, a Support Vector Machine (SVM) algorithm is trained using the data after preprocessing. Following training, the model is assessed using relevant performance metrics, including F1-score, recall, accuracy, and precision. The model can make predictions using fresh user input after it has been validated.

4.2 System Architecture

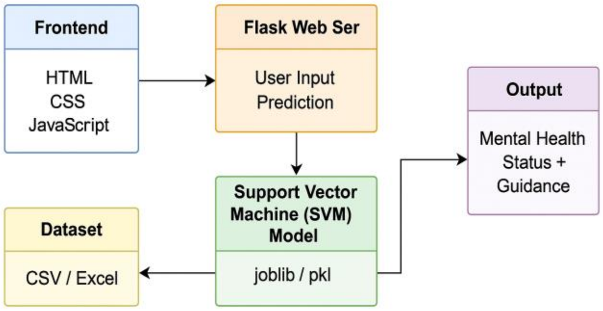

Fig. 2. Architecture of the AI-based mental health self-assessment system.

Fig. 2 displays the deployed system's architecture. Its frontend, which was created with HTML, CSS, and JavaScript, enables users to communicate with the system and enter their answers. A Flask web server serves as a mediator between the UI and the prediction engine, sending these inputs to the backend.

A Support Vector Machine (SVM) model that has already been trained and serialized using joblib or.pkl formats receives the data from the server. After processing the input, the model provides the anticipated mental health condition as well as any advice or suggestions. The output module then presents the findings on the user interface, including recommendations for self-care or expert consultation along with insights like "mild stress," "severe anxiety," or "normal mental health."

In order to ensure compatibility and ease of preprocessing during development, the dataset used to train the SVM model is usually in CSV or Excel format.

5. RESULTS AND EVALUATION

The evaluation metrics for AI-based self-assessment systems in mental health that have been documented in the literature are compared in this section. In order to forecast conditions like stress, anxiety, and depression, a variety of machine learning models have been applied to structured survey data, social media text, speech, and behavioral logs. Usually, metrics like accuracy, precision, recall, F1-score, and ROC-AUC are used to gauge their performance.

5.1 Comparative Analysis of Model Performance Across Modalities

Among the models that have been assessed, conventional techniques such as SVM and Random Forest exhibit high accuracy and interpretability when applied to structured datasets. Despite being a commonly used baseline, logistic regression exhibits relatively low recall. CNN and LSTM, two deep learning models, perform better on unstructured data, such as postings from social media. Notably, on Reddit datasets, BERT performs better than any other model, attaining the best accuracy and recall. It is particularly useful for identifying mental health issues in text-rich situations because of its comprehension of contextual language. These findings emphasize how crucial it is to choose models for mental health assessments according to the type of data and the application setting.

Model	**Input Type**	**Dataset / Source**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
SVM	Wearable Sensor Data	Stress Detection Dataset	83%	0.82	0.85	0.83
Random Forest	Survey (Teen Depression)	Adolescent Dataset	86%	0.84	0.88	0.85
Logistic Regression	Structured Survey	Kaggle Mental Health in Tech	79%	0.78	0.77	0.77
CNN	Instagram Image Features	Instagram-Based Depression Study	88%	0.86	0.89	0.87
LSTM	Twitter Posts	CLEF eHealth Twitter Dataset	84%	0.83	0.82	0.82
BERT	Reddit Text (NLP)	r/depression & r/anxiety Forums	90%	0.89	0.91	0.90

Table II: Comparative Performance of Machine Learning Models in Mental Health Prediction.

6. CONCLUSION

6.1 Conclusion

This study offers an AI-based self-assessment tool designed for mental health screening that is portable, private, and easily accessible. The system maintains interpretability and user trust while providing real-time predictions by utilizing a linear SVM classifier trained on structured survey data. Our strategy places a higher priority on ethical deployment, scalability, and simplicity than many other models that rely on invasive data sources or black-box algorithms. As a result, it is well-suited for implementation in community and educational contexts.

The conclusions and architectural layout presented here make a significant contribution to the expanding corpus of research on responsible AI in mental health. The significance of creating inclusive and human-centered AI solutions in psychological health domains is highlighted by this work, which addresses prevalent constraints like lack of transparency, restricted access, and data privacy issues.

Future research could involve testing in clinical settings, incorporating multilingual survey capabilities, and extending model capabilities through semi-supervised learning for more generalization**.**

6.2 Future Work

Although the existing system performs admirably in terms of ethical and easily available mental health screening, there are still a number of areas that might be improved:

• Integration of Multilingual Capabilities: By adding support for many languages, the questionnaire will be more widely used by a variety of user groups, particularly in multilingual nations.

• Clinical Validation: The model's diagnostic reliability and medical credibility will be established by working with mental health specialists to test and improve it in clinical settings.

• Hybrid Model Enhancements: To increase accuracy without compromising transparency, future iterations may integrate sophisticated deep learning techniques with interpretable models like SVM.

• Adaptive Learning: Personalized predictions and ongoing model improvement over time may be made possible by integrating user feedback into the model's training loop.

• Mobile App Deployment: Users looking for on-the-go support may find the current web-based system more accessible and convenient if it is extended to mobile platforms.

• Expanded Feature Set: By preserving privacy, adding optional passive data (such as usage trends or sentiment analysis from journaling) can enhance insights and strengthen the model's resilience.

7. REFERENCES

[1] A. B. Shatte, D. Hutchinson, and P. Teague, "Machine learning in mental health: A scoping review of methods and applications," Psychol. Med., 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 literature review," J. Med. Internet Res., vol. 22, no. 7, e20766, 2020.

[3] N. Cummins et al., "A review of depression and suicide risk assessment using speech analysis," Speech Commun., vol. 71, pp. 10–49, 2015.

[4] D. Guntuku et al., "Detecting depression and mental illness on social media: An integrative review," Curr. Opin. Behav. Sci., vol. 18, pp. 43–49, 2017.

[5] J. M. Girard et al., "Social risk and depression: Evidence from facial expressions and behavior," J. Nonverbal Behav., vol. 38, no. 2, pp. 161–183, 2014.

[6] H. Tariq, M. Abbas, S. Z. Abbas, and F. Rizwan, "A machine learning approach for student mental health prediction," Procedia Comput. Sci., vol. 192, pp. 693–702, 2021.

[7] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot)," JMIR Ment. Health, vol. 4, no. 2, e19, 2017.

[8] T. Nguyen et al., "Comparative performance of Random Forest and SVM for adolescent mental health prediction," Int. J. Med. Inform., vol. 137, 104099, 2020.

[9] S. Saeb et al., "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," J. Med. Internet Res., vol. 17, no. 7, e175, 2015.

[10] A. Chancellor, S. Lin, and M. L. De Choudhury, "Quantifying mental health signals in Instagram," in Proc. CHI, 2016, pp. 1171–1182.

[11] F. Orabi, N. Buddhitha, M. Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," in Proc. CLEF eHealth Workshop, 2018.

[12] A. Bashar and R. Nayak, "An ensemble approach to mental health classification," in Proc. ICMLA, 2021.

[13] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9: Validity of a brief depression severity measure," J. Gen. Intern. Med., vol. 16, no. 9, pp. 606–613, 2001.

[14] R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe, "A brief measure for assessing generalized anxiety disorder," Arch. Intern. Med., vol. 166, no. 10, pp. 1092–1097, 2006.

[15] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in Proc. CHI, 2013.

[16] H. A. Schwartz et al., "Toward assessing changes in degree of depression through Facebook," in Proc. Workshop Comput. Linguist. Clin. Psychol., 2014, pp. 118–125.

[17] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing BERT for emotion classification from text," Informatics, vol. 5, no. 3, p. 40, 2018.

[18] J. Torous, A. Staples, and M. Onnela, "Realizing the potential of mobile mental health: New methods for new data in psychiatry," Curr. Psychiatry Rep., vol. 20, no. 8, p. 61, 2018.

[19] B. Inkster, T. Sarda, and P. Subramanian, "An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being," J. Med. Internet Res., vol. 20, no. 11, e10148, 2018.

[20] G. Lucas et al., "Improving mental health care with conversational agents: A preliminary investigation," J. Med. Internet Res., vol. 22, no. 10, e16794, 2020.

[21] N. Papernot et al., "The limitations of deep learning in adversarial settings," in Proc. IEEE EuroS&P, 2016.

[22] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.

[23] D. C. Mohr, M. Zhang, and S. M. Schueller, "Personal sensing: Understanding mental health using ubiquitous sensors and machine learning," Annu. Rev. Clin. Psychol., vol. 13, pp. 23–47, 2017.

[24] E. Vayena, A. Blasimme, and I. Cohen, "Machine learning in medicine: Addressing ethical challenges," PLoS Med., vol. 15, no. 11, e1002689, 2018.