World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

WJAETS

(Research Article)

Check for updates

Identifying Critical Mental Health Indicators Using Ensemble and Explainable AI Techniques

Erin Jahan Meem [1], Md Fakrul islam Polash [2], Mohammed Imam Hossain Tarek [3], Mehedi Hasan [3] and Mostafizur Rahman Shakil [4,*]

[1] *Department of Computer Science, Pacific States University, Los Angeles, CA 90010, USA.*
[2] *Department of Engineering Project Management, Westcliff University, Irvine, CA 92614, USA.*
[3] *Department of Business Administration, International American University, Los Angeles, CA 90010, USA.*
[4] *Department of Engineering Management, Westcliff University, Irvine, CA 92614, USA.*

World Journal of Advanced Engineering Technology and Sciences, 2025, 17(01), 186-199

Publication history: Received on 30 August 2025; revised on 04 October 2025; accepted on 07 October 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.17.1.1389

Abstract

Mental health issues have considerably increased in recent years. We need to devise innovative and effective means, such as ensemble models, for the timely and accurate diagnosis of depression. The models provide accurate predictions, but it is crucial to understand the model prediction behavior to ensure transparency and trust. This study uses ensemble techniques to predict depression and leverage Explainable Artificial Intelligence (XAI) techniques to address the black-box nature of these algorithms and enhance interpretability. A comprehensive publicly available depression dataset is used in this study, and the findings reveal that ensemble and explainable techniques provide a robust, reliable and transparent prediction. The study emphasizes the value of interpretability in AI-powered mental health applications, providing physicians with an open and reliable instrument for making decisions. The visualization techniques adapted in this article provide a comprehensive view towards the explainability of the model. The results will aid practitioners in distinguishing the contributing factors in mental health prediction, thereby improving trust in the classification models developed. The study summarizes that concentration and suicidal ideation are the most decisive factors and assist doctors in the accurate and timely prediction of the mental health of patients.

Keywords: Explainable AI; Depression; Mental health prediction; Ensemble; Feature importance; SHAP; LIME

1. Introduction

Depression is a psychological condition that can be affected by a number of factors such as daily stress, physical activity, and medical conditions. It coexists with symptoms, such as persistent depression, insomnia, and suicidal thoughts. People's feelings fluctuate according to circumstances, time of day, and weather [1]. Mental health is essential to avoid depression and preserve general well-being. Accurately anticipating various scenarios is essential for the healthcare industry. Therefore, it is essential to acknowledge and continue managing each person's circumstances to provide mental health treatment [2].

Many internal and external causes contribute to the serious health problems and psychological stress that people face today. Maintaining mental health lowers the risk of depression and aids in the development of emotional resilience. Accordingly, depression is a mental illness that affects people of all ages [2]. Patients with mental illnesses frequently conceal their conditions as mental illness is stigmatized in society. Furthermore, symptoms, such as dysthymia, lack of excitement, and depression, are common in most people, making them easy to ignore. Improving older individuals' access to social support networks may help to lower their rates of despair and suicidal thoughts. Being in committed

* Corresponding author: Mostafizur Rahman Shakil

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

romantic relationships lowers the risk of mental health issues and depression [3]. Losing relationships can also have a negative impact on well-being, increasing the risk of emotional distress and even premature mortality [4]. Because they provide people with a social support system, assist them in overcoming fears, and create possibilities for personal development, romantic partnerships are believed to enhance well-being [5].

Machine learning models can be used in mental health to identify persons who are at risk of developing mental health illnesses. Text data from social media posts and online support forums were analyzed to identify signals of distress, emotional instability, and other indicators of mental health disorders. Machine learning models are also used for early detection by identifying tiny linguistic clues or changes in language patterns that may signal the onset of mental health concerns [6]. Researchers have experimented with how machine learning (ML) might predict teenage mental health problems and identify major predictors such as genetic and environmental factors [6]. Their study underlines the importance of early prognosis in providing timely assistance. Parent-reported mental health issues such as impulsivity, inattention, and emotional symptoms were major predictors. Register data on neighborhood quality, parity, and gestational age at birth were other prominent features [7]. Overall, these studies suggest that while machine learning shows potential [8], there are still obstacles to achieving accurate and generalized outcomes across varied populations.

Data-empowered decision making often faces questions regarding transparency, despite their acknowledged performance and acceptability. Especially when handling sensitive data such as medical data [9], [10], the right to explainability becomes a serious concern, as a single diagnostic error can cost life. The same has been stressed by EU data protection directives, which raises concerns about making all data-based models equipped with explanatory modules [11]. AI models are black boxes and lack prediction transparency. While modelling based on sensitive data, such as medical financial data, it creates trust issues, and the result validity will be reserved [12]. In addition, irrelevant and redundant features in the dataset impact the model predictions and trustworthiness. Explainability is critical, as these models may also be used by physicians to elucidate causal relationships or as a prototype for decision making [13]. The lack of explainability often forces practitioners to ignore the predictions given by the models and opt for alternative solutions [14]. This has also been criticized as a factor that adds complexity to the model in terms of the cost of accuracy [15], [16]. The final aim is to balance these contradictory objectives, such as enforced explainability, improved accuracy, and reduced complexity. This generates the requirement for a model that provides a local or global explanation of the outcomes without the cost of complexity.

XAI provides techniques that offer insights into these models. XAI models analyze the features of the model and their contribution to classification/prediction. They exhibit the characteristics of feature selection and support analysts in defining the explainability of the model [17]. Prolonged sleep deprivation can affect emotional regulation, making it more difficult to handle stress, escalating anger, and increasing the risk of depression [18]. Additionally, excessive sleep, which is frequently linked to melancholy, may exacerbate sad thoughts and cause sensations of sluggishness and loss of motivation [19]. Depression is believed to be a common cause of a variety of symptoms, including psychomotor agitation and sleeplessness [20]. Sleep disturbance exacerbates the negative effects of depression on cognitive functions such as memory, focus, and decision-making. Feelings of inadequacy and frustration may worsen as a result of this mental fog [21].

Depression is frequently accompanied by hunger, which can decrease or increase appetite. These alterations are frequently linked to physiological, psychological, and emotional elements [22]. The literature also notes that depressed individuals struggle with drive and active participation, as well as a loss of delight and flattening of emotion. They lack a feeling of community and belonging, and begin to doubt their identity [23]. Across cultural boundaries, self-worthlessness/inadequacy is a unique and universal sign of serious depression [24]. Most individuals with depression experience hopelessness and restlessness [25]. Depression can cause severe emotional anguish, and people may feel hopeless. Suicidal thoughts may surpass the ongoing emotional suffering.

Fatigue can drain motivation, making simple tasks challenging. This lack of vigor frequently magnifies depressive and frustrating emotions . It was also observed that anger and depression can be produced by unfavorable experiences that have a strong affective impact [26]. People experiencing frequent panic attacks may experience depression . Individuals who are depressed may find it difficult to focus on the activity at hand because their minds frequently wander to unfavorable ideas or concerns. When depressed, people say that they do not feel as energetic as they used to [27]. Technology use may lead to a less physically demanding lifestyle. Additionally, individuals may be at risk for mental disorders if they are under persistent pressure. Peer pressure, heart attacks, despair, and many other consequences are examples of these vulnerabilities [28]. It is clear from previous studies that there is a great deal of potential for applying data mining and machine learning [29], [30], [31], [32]. They used unsupervised learning techniques to group responses and then applied various machine learning algorithms, such as support vector machines, decision trees, naïve Bayes classifier, KNN classifier, and logistic regression, to identify the state of mental health in a target group.

Researchers have also explored the efficacy of single and ensemble classifiers for mental health prediction. The role of machine-learning mechanisms in different psychological disorders was also examined. This study revealed that ensemble approaches outperformed single classifiers in terms of accuracy and resilience when predicting mental health. A systematic assessment of machine learning approaches used for mental health diagnostics, emphasizing the relevance of feature selection and data quality, was also presented [33]. It is clear from previous studies that there is a great deal of potential for applying machine learning to various facets of psychology and mental health. Although there are a considerable number of studies on predicting mental health using supervised machine-learning algorithms, we found that there is a lack of extant studies on explainable models. The primary goal of this study was to select the best ML model among the various models with good explainability.

Explainable AI has been experimented with for predictions in healthcare applications [34]. The decision Forests-based classifier, based on an explainable AI model, was found to be better than deep learning-based models for COVID-19 predictions [35]. Another study explored various machine learning techniques, such as KNN, SVM, and Multilayer Perceptron (MLP), along with feature extraction methods, Recursive Feature Elimination (RFE), and Extra Trees (ET), for predicting levels of stress, depression, and anxiety. This study used a dataset that encompassed sociodemographic factors, underlying diseases, and mental health attributes, and SVM was found to be better at predicting the mental health conditions of recovered COVID-19 patients.

In the context of this research, we propose the use of a series of ML models to predict mental health and prove the results genuinely by explaining model predictions using different XAI techniques. This facilitates physicians to understand, evaluate, and trust the decisions made by the models and generates a high rate of prediction accuracy. We evaluated the dataset using three ML models and explained the results using explainable AI techniques. This study disclosed the interdependencies of the attributes in the dataset under consideration, and the classification results were evaluated. This research contributes to the knowledge base by identifying the attributes on which practitioners can build trust in predicting the results.

2. Methodology and Experimental Setup

Recent works have improved therapeutic outcomes. Techniques in precision wound healing utilize regenerative therapies and advanced technologies for personalized recovery strategies. Hybrid nanoconjugates of temozolomide enhance drug stability and effectiveness in glioblastoma treatment, addressing limitations of traditional therapies [36] [37]. The introduction of molecular erasers for protein degradation is reprogramming cancer immunity and opening new avenues in immuno-oncology [38]. Simultaneously, artificial intelligence (AI) is transforming various fields. In precision agriculture [39], AI enhances efficiency through automated farming systems. In computer vision, attention mechanisms are advancing road segmentation for autonomous vehicles [40]. In healthcare, hybrid models combining CNNs and SVMs with enhanced MRI data show promise in diagnosing Alzheimer’s disease [41]. Furthermore, multimodal deep learning frameworks like MultiSenseNet [42] help predict machinery failures in industries, while deep stacking models improve leaf disease identification in plant pathology [43]. These developments highlight how data-driven solutions work together to tackle complex challenges across critical sectors.

2.1. Dataset Description

This dataset included patient survey responses regarding a range of depression-related symptoms. The dataset used in this study was obtained from Kaggle. The responses to the 14 questions to which each patient had to respond were coded from 1 to 6 according to how frequently each symptom was experienced. The dataset also contained a column that showed a patient's general level of depression. The independent variables measure attributes such as sleep disturbances, changes in appetite, loss of interest in activities, feelings of fatigue and worthlessness, difficulty in concentrating, physical agitation and aggression, Suicidal Ideation and Panic Attacks, Feelings of hopelessness and restlessness. These independent variables predict the mental health of a patient and classify them into depression states: “No Depression”, “Mild”, “Moderate” and ”Severe”.

2.2. Experimental Setup

Choosing the best experimental setup improves the reliability of the explainability, as it ensures the reproducibility of the results. The experiments were conducted on a machine with 16GB RAM and an Intel Core Processor with Windows 11 operating system. The classification models were implemented using libraries in Python version 3.11.5. For the Explainable AI experiments, we used SHAP 0.46.0 and LIME 0.2.0.1 versions.

2.3. Explainable AI techniques

Explainable AI provides both global and local explanations for a model. By providing a global explanation, which defines the entire model behavior, and a local explanation for intrinsic definitions, explainability sheds light on model predictions. This research analyzes the explainability of these prediction models using Local Interpretable model–agnostic explanations (LIME) and Shapley Additive explanations (SHAP).

SHAP identifies the feature importance in the prediction for each class of predictions. The Shapley Additive exPlanations) method provides a transparent method for defining the feature importance for prediction. This model is based on cooperative game theory, which uses Shapley values to interpret predictions considering their contributions. SHAP provides more reliable local and global explanations of predictions and is faster when implementing tree-based models. This demonstrates feature weighting with SHAP values and contributes to the explainability of the complex model. It is a flexible tool that offers insights into model prediction using Shapley values and attribute marginal contributions. It can handle high-dimensional complex datasets and display global explanations of the model. In this study, we analyzed the models using SHAP summary plots and waterfall plots.

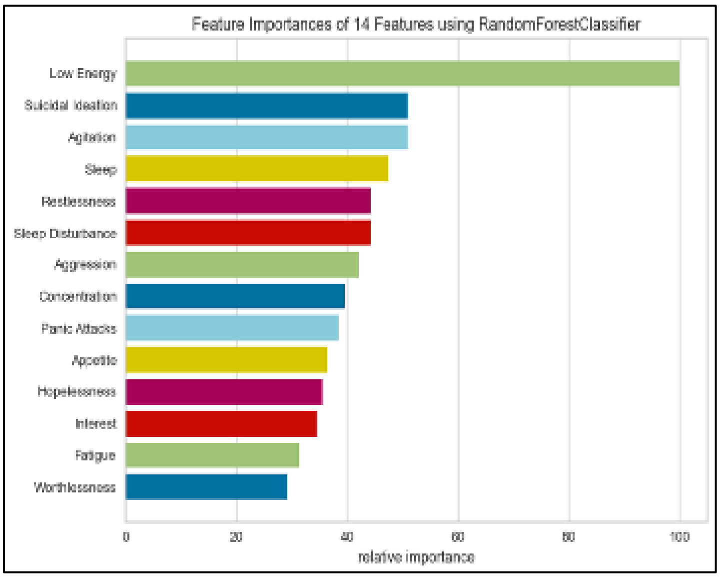
The LIME model explains the predictions by locally approximating the contributions. LIME provides a local explanation of the model and a granular analysis of the attribute importance. Granular analysis highlighted the individual predictions of each classification. LIME provides reliable explanations of the model, irrespective of its features and complexity. It provides explanations of individual prediction, feature importance, and quantitative measures of the feature impact on prediction. LIME explanations are described using feature contribution plots.

3. Result Analysis and Discussion

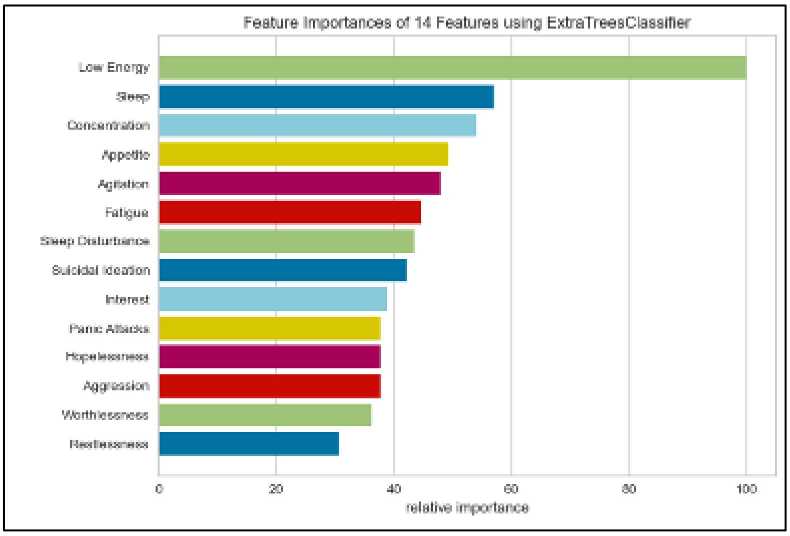
In ensemble-based medical recommendation systems, the major attribute that defines the model's usability is trust. Appropriate predictions and the usage of correct attributes are highly recommended to make the model convincing. Three algorithms, Random Forest Classifier (accuracy: 51.39), Extra Trees classifier (accuracy: 51.62), and Decision Tree Classifier (accuracy: 51.62), are considered, and the results are explained using LIME, SHAP, waterfall, and yellow brick models. The aim was to conduct an interpretable analysis of the results of this classification and facilitate model evaluation and reliability.

3.1. Feature Importance using yellow brick model

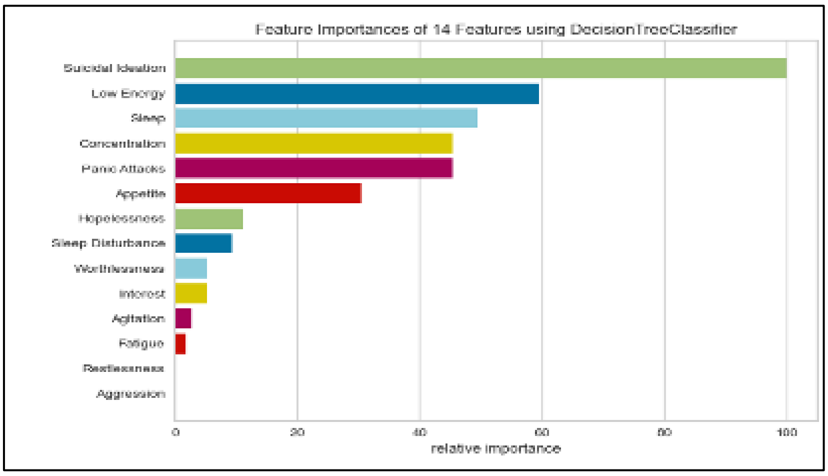
The explainability of all models was first explored using the yellow brick model for feature importance. It visualizes the attribute ranks and their relative importance by plotting the attribute coefficient on the x-axis and sorting them according to their importance on the y-axis.



(a) Random Forest Classifier



(b) Extra Trees Classifier



(c) Decision Tree Classifier

Figure 1 Feature Importance using yellow brick model

Owing to the inherent characteristics of the models, the feature contributions differed according to the coefficients of the predictors. Feature importance using the yellow brick model evaluates a classification model to describe the causal impact of features in predicting the classifier. It allows visual diagnostics of the model and can re-evaluate features with transparency. The attribute contributions exhibited different trends for different prediction models. Fig 1 (a), (b), and (c) show the feature importance with reference to the three models under consideration. As shown in Fig 1(a) and (b), "low energy" contributes the most to mental health prediction using Random Forest and Extra Tree classifiers. However, Decision tree predictions were highly dependent on "Suicidal Ideation, followed by "Low Energy. To help the practitioner summarize and decode the important features, in this study, we sieve out those features that are uniformly selected by all algorithms and summarize them.

Features	Random Forest	Extra Trees	Decision Tree
Sleep	4	2	3
Appetite	10	4	6
Interest	12	9	10
Fatigue	13	6	12
Worthlessness	14	13	9
Concentration	8	3	4
Agitation	3	5	11
Suicidal Ideation	2	8	1
Sleep Disturbance	6	7	8
Aggression	7	12	13
Panic Attacks	9	10	5
Hopelessness	11	11	7
Restlessness	5	14	13
Low Energy	1	1	2

Figure 2 Feature Importance – Summarized

The observations on the importance of the 14 features in the context of all three algorithms are summarized in a color-coded plot, as shown in Figure 2. Red indicates the weightage of the feature in each algorithm, and the darker the color, the larger is the weightage. It reveals highly relevant features and claims that sleep and low energy are identified as the most important features of all three algorithms, and agitation, suicidal ideation, appetite, and concentration by at least two. This plot also highlights the explainability of the model by highlighting the relative importance of each feature in the classification process. Analyzing the feature contribution makes the model's recommendations convincing and explainable to common users.

3.2. Interpretation Based on LIME

As models become more complex, their accuracy increases, but non-interpretability emerges. The LIME model realizes the interpretability of complex models by carefully evaluating instances and their predictions. LIME explains the model by using perturbations of each sample point and training the black-box model using the newly generated data point. Regardless of the classification model, the LIME model interpreted the classification of the results. The features identified by each model vary, and the contributions are analyzed as positive or negative. This helps us to understand the essential attributes that make a poor contribution to the classification. LIME provides the prediction probabilities for each class in the classifier. Fig 3 (a) shows the prediction probabilities for the RandomForestClassifier for all four classes 0.20: Mild, 0.27 for Moderate, 0.29 for No depression and 0.23 for Severe. Similar explanations can be observed for the ExtraTreesClassifier in Fig 3(b) and the DecisionTreeClassifier in Fig 3(c).

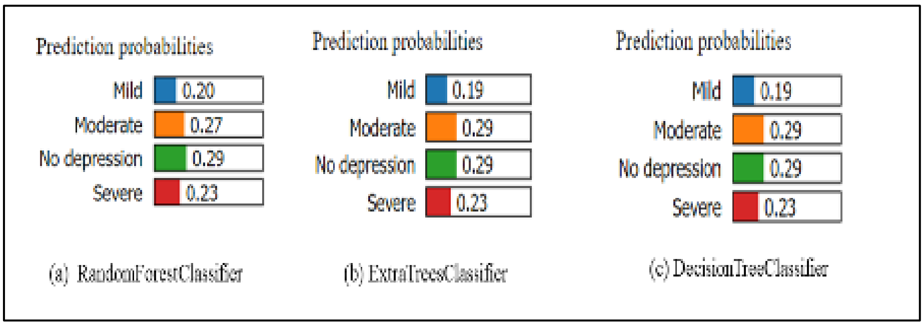


Figure 3 Prediction Probabilities (a) RandomForestClassifier (b) ExtraTreesClassifier (c) DecisionTreeClassifier

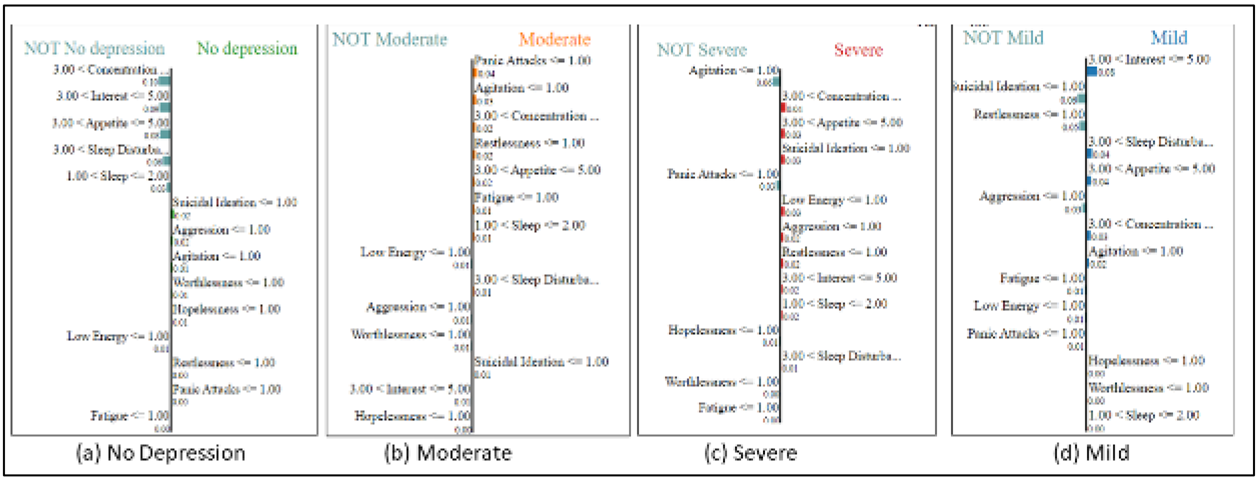


Figure 4 Feature contribution – Random Forest

The LIME plot provides an explanation for the contribution of each feature to each class. Figure 4 to Fig 6 show the feature contributions of each model for all four classes: no depression, Moderate, Severe and Mild as (a), (b), (c), and (d), respectively. The contributions are defined using the magnitude and direction plotted on the x-axis, and the y-axis sorts it according to the contribution impact. The positive stretch of the beam indicates that an increasing value of the feature will positively impact the classification, and the negative stretch shows a negative impact.

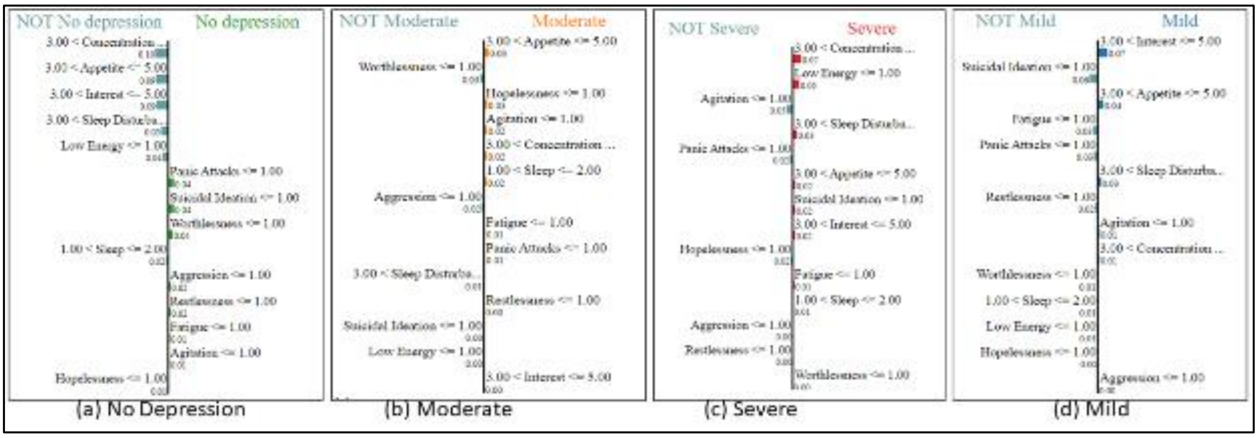


Figure 5 Feature contribution – Extra Tree Classifier

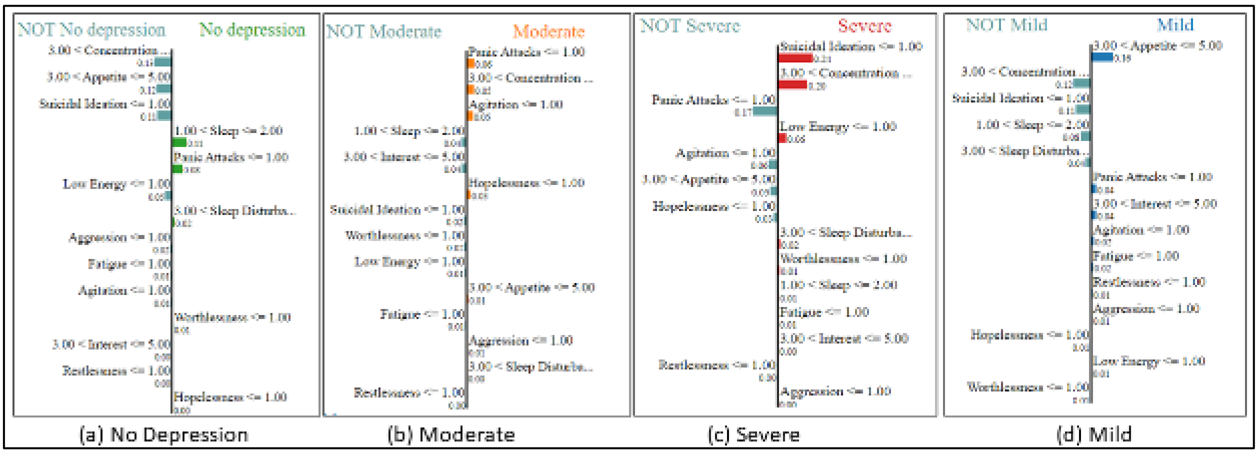


Figure 6 Feature contribution – DecisionTreeClassifier

While analyzing the Random Forest feature contribution LIME, we see that attributes such as concentration, appetite, and interest have a negative impact on the No Depression class (see Fig 4(a)), whereas panic attack, agitation, concentration, and appetite have a positive impact on the Moderate class (Refer Fig 4(b)). A similar type of analysis can be derived from other classifiers and classes. (Fig 5 and Fig 6), respectively. The LIME feature contribution also provides the magnitude of contribution of each attribute to the classification class, such as -0.13, as the contribution of attribute concentration towards the class No Depression in the DecionTreeClassifier. (Fig 6 (a)).

3.3. Interpretation Based on SHAP

The integration of SHAP into the classification algorithm sheds light on the decision-making process of the model. The calculated SHAP values can also be represented using waterfall plots and SHAP summary plots to explore the global importance of the selected features in predicting mental health. The SHAP summary plot represents the feature contribution to the model prediction. By assigning a value to each attribute for one prediction, SHAP explains the root causes behind each prediction, thereby augmenting the transparency and reliability of the model. The features are arranged based on their contributions on the y-axis and SHAP values on the x-axis. A higher SHAP value indicates a greater influence on the model prediction. The SHAP waterfall model defines local interpretability and global interpretability using summary plots. The summary plots summarize the feature impact for each class and visualize the relative importance using color codes. The waterfall model displayed individual feature contributions at the granular level, enhancing model explainability.

The SHAP Waterfall plot provides insights into the local interpretability of the model. The SHAP Waterfall model comprehensively interprets the contribution of each attribute to the decision-making process. In the plots, E[f(X)] represents the expected output value, and f(X) indicates the prediction. They depict individual predictions with contributions as positive(red) and negative(blue) for each feature and represent the features that drive and drag the prediction, respectively. The length and direction of the bar also provide valuable perceptions of the influence of attributes. The attribute weightage calculated from the SHAP value is placed on the x-axis, which quantifies the contribution, and the y-axis displays the features in the order of their contribution to the prediction value.

The figures (Fig 7 Fig 9) explain all three models with four different classes and their attribute contributions using waterfall plots.

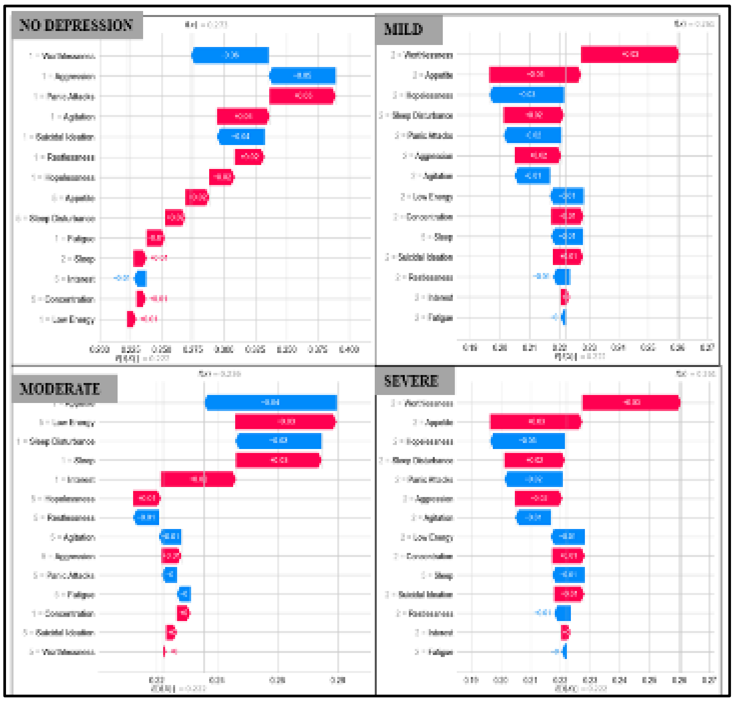


Figure 7 Waterfall model - Random Forest

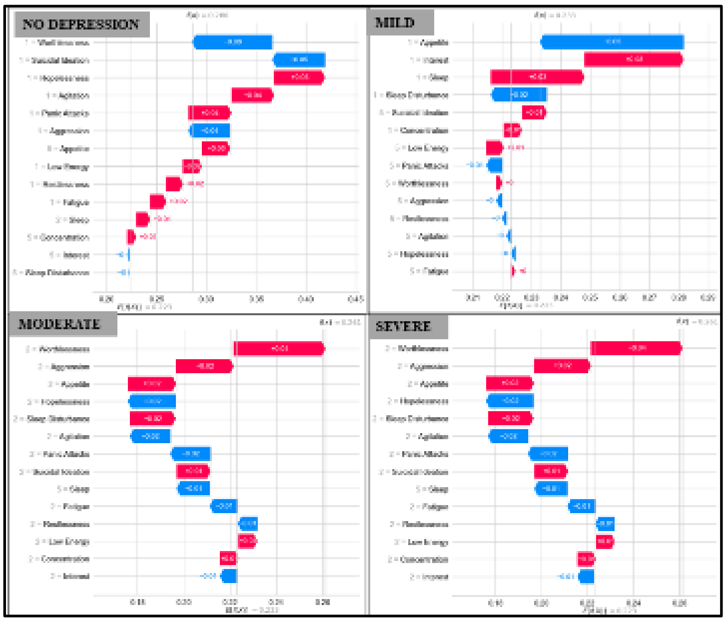


Figure 8 Waterfall model - Extra Tree Classifier

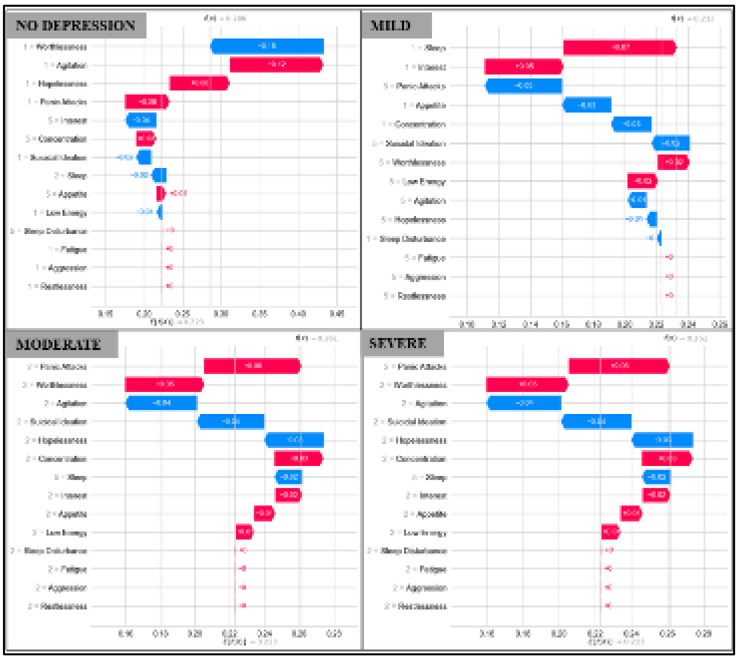


Figure 9 Waterfall model –DecisionTreeClassifier

As seen in Fig 7, in RandomForestClassifier, the attribute "Worthlessness" negatively impacts Class "No depression" (-0.05), positively impacting class "Mild" (+0.03) and "Severe"(+0.03) and showing negligible influence on class "Moderate" (+0.0). Similar interpretations can be derived from Fig 8 and Fig 9 for the four classes and influential attributes. Attributes such as Hopelessness, Worthlessness and Appetite were influential in all three models under consideration for all four classes. This indicates the dependency of the model on these attributes in terms of classification and prediction. The attributes showing negligibly smaller SHAP values (either positive or negative) in the plots indicate that the models do not depend on these features for prediction. For example, "Interest" and "Fatigue" were not influential in any of the three models considered, as suggested by the waterfall plot. We can explain the model's global predictions using summary plots, as shown in Figures 10–12. It provides a comprehensive explanation of attribute importance and its relative contributions to the prediction. The SHAP values are plotted on the x-axis and provide a measure of the impact on the model output, regardless of whether it is positive or negative. The features are ordered based on their significance on the y axis. The bar length corresponding to each feature indicates the influence of the feature on each class colored as follows: blue represents no depression, purple for mild, green for moderate, and red for severe. This also helps us to determine which class is more associated with each feature and prediction.

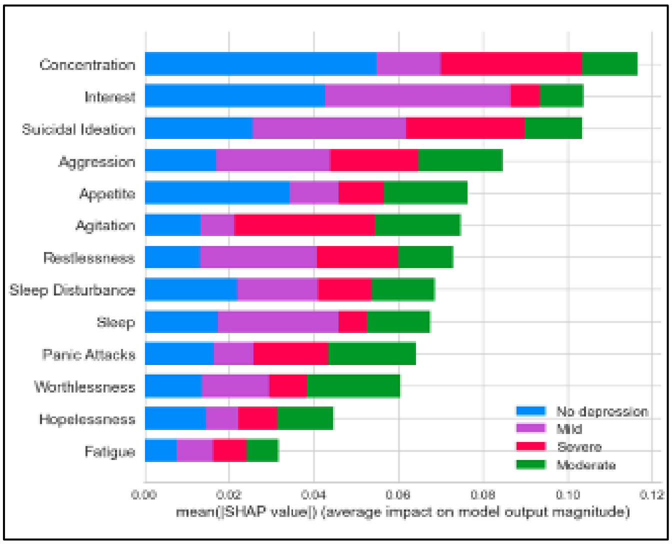


Figure 10 Summary Plot – Random Forest

The SHAP summary plot is interpreted as follows: In Fig 10, the bluer color on attribute concentration indicates that the attribute explains no depression prominently in the Random Forest classifier. It has a lesser contribution towards the class "Mild" and "Moderate" and a considerable contribution to the class "Severe." Similarly, the “Interest” attribute contributes to "No Depression" and "Mild" classes equally and is not a determinant factor for the "Severe" and “Moderate" classes. Similar explanations can be derived from Fig 11 and Fig 12 for the Extra Tree and Decision Tree Classifiers, respectively.

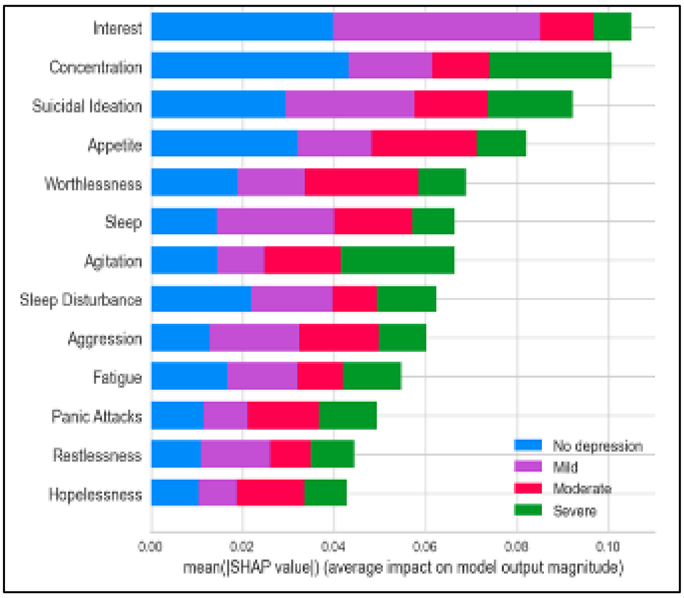


Figure 11 Summary Plot –Extra Tree Classifier

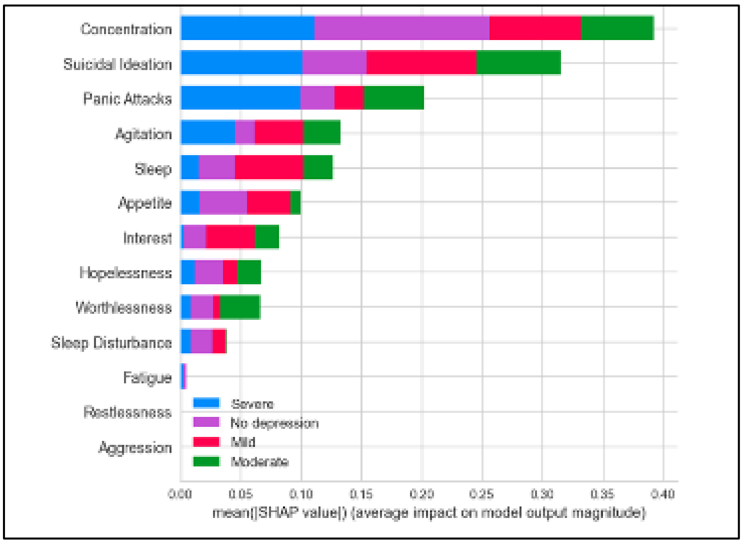


Figure 12 Summary Plot –DecisionTreeClassifier

Some attributes show significant contributions to the model across classes and across different models. This indicates their relative importance in prediction and forces the practitioner to ensure that these symptoms are addressed during diagnosis. The variations in the impact value specify the diversity that the algorithms provide and necessitate the proper consideration of the attributes. The analysis discloses that attributes "Concentration" and "Suicidal Ideation" have a higher influence on the classification in all the models considered, and "Fatigue" and "Restlessness" contribute weakly to the classification. It is also noted that the attributes "Restlessness" and "Aggression" do not at all contribute to the classification, and "Fatigue" shows a negligible contribution in the case of the Decision Tree Classifier. (Fig 12). Most of the attributes of the model show a noticeable influence on classification based on Random Forest and Extra Trees, and other than the top three attributes, all attributes show a uniform distribution of contribution. (Fig 10 and Fig 11), respectively. However, in the case of the Decision Tree Classifier, other than "Concentration" and "Suicidal Ideation," no attributes could define the classification. (Fig 12) This lack of strong feature contributions may also indicate complex interdependencies between features.

This analysis explains the feature contributions across models using different plots. The SHAP values displayed show the impact of each feature on the prediction, and the sign indicates the direction of the prediction as well. The variations in the feature selection clarified the model differences and helped the practitioner select the best model that matched the predictions. This study combines the inputs from multiple plots to comprehend the feature contributions of each model. This makes the decision-making process more reliable and helps the practitioner build trust in model predictions. This study evaluated the efficiency of mental health predictions and reliability of the model. The results are summarized by analyzing the outputs from the yellow brick model, LIME, and SHAP. The yellow brick model summarized Sleep, Appetite, Concentration, Agitation, Suicidal Ideation and Low energy as the most influential attributes in all three models. The LIME model identifies Appetite and Concentration as influential, which matches the Yellow Brick model but shows a negligible influence of Sleep and Low energy. Agitation and Suicidal Ideation showed moderate influence, and attributes of panic attack and interest emerged as influential through LIME analysis. The most important features that emerged after the SHAP waterfall model are that the SHAP model identifies concentration, suicidal ideation, Appetite, Worthlessness and Hopelessness and Interest, and Fatigue and Restlessness, which did not contribute to the classification. This confirms our results in LIME, where most of these features also emerged in the LIME analysis.

4. Conclusion

Mental health is a growing concern, and for maintaining overall wellness, there is a need to deal with depression. Numerous factors lead to depression, and this study revealed the most influential attributes. To make the outcome of this study useful for practitioners, explainable models were considered in this study. We identified Concentration and Suicidal Ideation as the most contributing factors, irrespective of the classification model. It also revealed the considerable influence of attributes of Appetite, Sleep and Agitation on patients' mental health prediction. This research reinforces the association of these features with patients' mental health conditions, highlighting the need for proper diagnosis and remediation action concentrating on these conditions.

Compliance with ethical standards

Disclosure of conflict of interest

There is no conflict of interest.

References

[1] K. Zantvoort, N. Hentati Isacsson, B. Funk, and V. Kaldo, 'Dataset size versus homogeneity: A machine learning study on pooling intervention data in e-mental health dropout predictions', Digit Health, vol. 10, Jan. 2024, doi: 10.1177/20552076241248920.

[2] V. Pendyala and H. Kim, 'Assessing the Reliability of Machine Learning Models Applied to the Mental Health Domain Using Explainable AI', Electronics 2024, Vol. 13, Page 1025, vol. 13, no. 6, p. 1025, Mar. 2024, doi: 10.3390/ELECTRONICS13061025.

[3] A. P. Darko, C. O. Antwi, K. Adjei, B. Zhang, and J. Ren, 'Predicting determinants influencing user satisfaction with mental health app: An explainable machine learning approach based on unstructured data', Expert Syst Appl, vol. 249, p. 123647, Sep. 2024, doi: 10.1016/J.ESWA.2024.123647.

[4] Z. Zhang, 'Early warning model of adolescent mental health based on big data and machine learning', Soft comput, vol. 28, no. 1, pp. 811–828, Jan. 2024, doi: 10.1007/S00500-023-09422-Z/METRICS.

[5] A. Alagic et al., 'Machine Learning for an Enhanced Credit Risk Analysis: A Comparative Study of Loan Approval Prediction Models Integrating Mental Health Data', Machine Learning and Knowledge Extraction 2024, Vol. 6, Pages 53-77, vol. 6, no. 1, pp. 53–77, Jan. 2024, doi: 10.3390/MAKE6010004.

[6] S. Jayanthi, V. Priyadharshini, V. Kirithiga, and S. Premalatha, 'Mental health status monitoring for people with autism spectrum disorder using machine learning', International Journal of Information Technology (Singapore), vol. 16, no. 1, pp. 43–51, Jan. 2024, doi: 10.1007/S41870-023-01524-Z/METRICS.

[7] M. R. Ahmed, R. Haque, S. M. A. Rahman, A. W. Reza, N. Siddique, and H. Wang, 'Vision-audio multimodal object recognition using hybrid and tensor fusion techniques', Information Fusion, vol. 126, p. 103667, Feb. 2026, doi: 10.1016/J.INFFUS.2025.103667.

[8] R. Haque, S. H. Laskar, K. G. Khushbu, M. J. Hasan, and J. Uddin, 'Data-Driven Solution to Identify Sentiments from Online Drug Reviews', Computers 2023, Vol. 12, Page 87, vol. 12, no. 4, p. 87, Apr. 2023, doi: 10.3390/COMPUTERS12040087.

[9] J. Debnath et al., 'LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis', Inform Med Unlocked, vol. 57, p. 101669, Jan. 2025, doi: 10.1016/J.IMU.2025.101669.

[10] Md. R. Ahmed et al., 'Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis', Bioengineering 2025, Vol. 12, Page 651, vol. 12, no. 6, p. 651, Jun. 2025, doi: 10.3390/BIOENGINEERING12060651.

[11] Z. E. Imel et al., 'Outcomes in Mental Health Counseling From Conversational Content With Transformer-Based Machine Learning', JAMA Netw Open, vol. 7, no. 1, pp. e2352590–e2352590, Jan. 2024, doi: 10.1001/JAMANETWORKOPEN.2023.52590.

[12] R. Haque et al., 'Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis', Comput Biol Med, vol. 191, p. 110166, Jun. 2025, doi: 10.1016/J.COMPBIOMED.2025.110166.

[13] M. I. H. Siddiqui et al., 'Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable AI', Inform Med Unlocked, vol. 56, p. 101657, Jan. 2025, doi: 10.1016/J.IMU.2025.101657.

[14] S. Khan et al., 'Ensemble-Based Explainable Approach for Rare Medicinal Plant Recognition and Conservation', 2025 10th International Conference on Information and Network Technologies, ICINT 2025, pp. 88–93, 2025, doi: 10.1109/ICINT65528.2025.11030872.

[15] A. S. U. K. Pranta et al., 'A Novel MaxViT Model for Accelerated and Precise Soybean Leaf and Seed Disease Identification', Computers 2025, Vol. 14, Page 197, vol. 14, no. 5, p. 197, May 2025, doi: 10.3390/COMPUTERS14050197.

[16] A. Al Noman et al., 'ViX-MangoEFormer: An Enhanced Vision Transformer–EfficientFormer and Stacking Ensemble Approach for Mango Leaf Disease Recognition with Explainable Artificial Intelligence', Computers 2025, Vol. 14, Page 171, vol. 14, no. 5, p. 171, May 2025, doi: 10.3390/COMPUTERS14050171.

[17] A. Hossain et al., 'Transformer-Based Ensemble Model for Binary and Multiclass Oral Cancer Segmentation', 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025, 2025, doi: 10.1109/ECCE64574.2025.11012921.

[18] C. Sweeney, E. Ennis, M. D. Mulvenna, R. Bond, and S. O'Neill, 'Insights Derived From Text-Based Digital Media, in Relation to Mental Health and Suicide Prevention, Using Data Analysis and Machine Learning: Systematic Review.', JMIR Ment Health, vol. 11, no. 1, p. e55747, Jun. 2024, doi: 10.2196/55747.

[19] S. Singhal, D. L. Cooke, R. I. Villareal, J. J. Stoddard, C. T. Lin, and A. G. Dempsey, 'Machine Learning for Mental Health: Applications, Challenges, and the Clinician's Role', Curr Psychiatry Rep, vol. 26, no. 12, pp. 694–702, Dec. 2024, doi: 10.1007/S11920-024-01561-W/METRICS.

[20] J. Rogan, S. Bucci, and J. Firth, 'Health Care Professionals' Views on the Use of Passive Sensing, AI, and Machine Learning in Mental Health Care: Systematic Review With Meta-Synthesis.', JMIR Ment Health, vol. 11, no. 1, p. e49577, Jan. 2024, doi: 10.2196/49577.

[21] A. Radwan, M. Amarneh, H. Alawneh, H. I. Ashqar, A. AlSobeh, and A. A. A. R. Magableh, 'Predictive Analytics in Mental Health Leveraging LLM Embeddings and Machine Learning Models for Social Media Analysis', https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJWSR.338222, vol. 21, no. 1, pp. 1–22, Jan. 1AD, doi: 10.4018/IJWSR.338222.

[22] M. M. Islam, S. Hassan, S. Akter, F. A. Jibon, and M. Sahidullah, 'A comprehensive review of predictive analytics models for mental illness using machine learning algorithms', Healthcare Analytics, vol. 6, p. 100350, Dec. 2024, doi: 10.1016/J.HEALTH.2024.100350.

[23] C. Halkiopoulos and E. Gkintoni, 'The Role of Machine Learning in AR/VR-Based Cognitive Therapies: A Systematic Review for Mental Health Disorders.', Electronics (2079-9292), vol. 14, no. 6, p. 1110, Mar. 2025, doi: 10.3390/ELECTRONICS14061110.

[24] L. S. Khoo, M. K. Lim, C. Y. Chong, and R. McNaney, 'Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches', Sensors 2024, Vol. 24, Page 348, vol. 24, no. 2, p. 348, Jan. 2024, doi: 10.3390/S24020348.

[25] U. Madububambachu, A. Ukpebor, and U. Ihezue, 'Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review', Clin Pract Epidemiol Ment Health, vol. 20, no. 1, p. e17450179315688, Jul. 2024, doi: 10.2174/0117450179315688240607052117.

[26] A. Al Noman et al., "Monkeypox Lesion Classification: A Transfer Learning Approach for Early Diagnosis and Intervention," 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), pp. 247–254, Sep. 2024, doi: 10.1109/IC3I61595.2024.10828678.

[27] A. B. S. Rahman, H. T. Ta, L. Najjar, A. Azadmanesh, and A. S. Gönul, 'DepressionEmo: A novel dataset for multilabel classification of depression emotions', J Affect Disord, vol. 366, pp. 445–458, Dec. 2024, doi: 10.1016/J.JAD.2024.08.013.

[28] T. Ahmed, S. Ivan, A. Munir, and S. Ahmed, 'Decoding depression: Analyzing social network insights for depression severity assessment with transformers and explainable AI', Natural Language Processing Journal, vol. 7, p. 100079, Jun. 2024, doi: 10.1016/J.NLP.2024.100079.

[29] J. Debnath et al., 'Hybrid Vision Transformer Model for Accurate Prostate Cancer Classification in MRI Images', 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025, 2025, doi: 10.1109/ECCE64574.2025.11013952.

[30] H. Rahman et al., 'Automated Weed Species Classification in Rice Cultivation Using Deep Learning', 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–6, Feb. 2025, doi: 10.1109/ECCE64574.2025.11014047.

[31] M. S. Rahman et al., 'Effective Disease Recognition in Cucumbers: A Web-Based Application Using Transfer Learning Models', 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things, RAAICON 2024 - Proceedings, pp. 59–64, 2024, doi: 10.1109/RAAICON64172.2024.10928353.

[32] M. I. Hossain Siddiqui et al., 'Eggplant Disease Diagnosis Using a Robust Ensemble of Transfer Learning Architectures', 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–6, Feb. 2025, doi: 10.1109/ECCE64574.2025.11013918.

[33] E. Bao, A. Pérez, and J. Parapar, 'Explainable depression symptom detection in social media', Health Inf Sci Syst, vol. 12, no. 1, pp. 1–18, Dec. 2024, doi: 10.1007/S13755-024-00303-9/FIGURES/5.

[34] R. Haque et al., 'Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning', BioMedInformatics 2024, Vol. 4, Pages 966-991, vol. 4, no. 2, pp. 966–991, Apr. 2024, doi: 10.3390/BIOMEDINFORMATICS4020054.

[35] A. Finze, L. Deleanu, C. Weiss, M. Ratliff, and M. Seiz-Rosenhagen, 'Depression in brain tumor patients—early detection and screening', Supportive Care in Cancer, vol. 31, no. 6, pp. 1–8, Jun. 2023, doi: 10.1007/S00520-023-07785-5/TABLES/3.

[36] A. H. Malik and S. Rahman, 'Toward precision wound healing: Integrating regenerative therapies and smart technologies', International Journal of Science and Research Archive, vol. 16, no. 3, pp. 244–257, Sep. 2025, doi: 10.30574/IJSRA.2025.16.3.2492.

[37] A. H. Malik and S. Rahman, 'Hybrid Temozolomide Nanoconjugates: A polymer–drug strategy for enhanced stability and glioblastoma therapy', International Journal of Science and Research Archive, vol. 16, no. 3, pp. 258–268, Sep. 2025, doi: 10.30574/IJSRA.2025.16.3.2493.

[38] A. H. Malik and S. Rahman, 'Molecular erasers: Reprogramming cancer immunity through protein degradation', World Journal of Advanced Engineering Technology and Sciences, vol. 16, no. 3, pp. 277–291, Sep. 2025, doi: 10.30574/WJAETS.2025.16.3.1335.

[39] M. Sabbir Hossain et al., 'Automatic Navigation and Self-Driving Technology in Agricultural Machinery: A State-of-the-Art Systematic Review', IEEE Access, vol. 13, pp. 94370–94401, 2025, doi: 10.1109/ACCESS.2025.3573324.

[40] M. S. Hossain, M. Rahman, M. Ahmed, M. M. Kabir, M. F. Mridha, and J. Shin, 'DeepLabv3Att: Integrating Attention Mechanisms in DeepLabv3 for Enhanced Road Segmentation', 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 711–718, 2024, doi: 10.1109/3ICT64318.2024.10824285.

[41] M. Rahman, M. S. Hossain, A. A. Eva, M. M. Kabir, M. F. Mridha, and J. Shin, 'Alzheimers Disease Classification with a Hybrid CNN-SVM Approach on Enhanced MRI Data', 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 50–57, 2024, doi: 10.1109/3ICT64318.2024.10824633.

[42] M. Rahman, M. S. Hossain, U. Rozario, S. Roy, M. F. Mridha, and N. Dey, 'MultiSenseNet: Multi-Modal Deep Learning for Machine Failure Risk Prediction', IEEE Access, vol. 13, pp. 120404–120416, 2025, doi: 10.1109/ACCESS.2025.3586978.

[43] M. S. Hossain, M. Rahman, M. G. R. Abir, J. Maua, and A. Rahman, 'Plant Leaf Disease Detection Using Deep Stacking: Integrating CNNs and Gradient Boosting for Enhanced Classification Accuracy', Studies in Computational Intelligence, vol. 1202, pp. 15–26, 2025, doi: 10.1007/978-981-96-4520-6_2.