Statistical Models and Machine Learning in Depression Detection: Evaluating XGBoost

MARÍA BARRERA[1], CRISTIAN INCA[2], ZILMA DIAGO[3], EVELYN INCA[4], NELSÓN VILLARROEL[5], JOSÉ RUBÉN LEÓN PÉREZ[6], LAURA MUÑOZ[3], ROLANDO SANCHÉZ[2]

[1]Universidad Estatal de Milagro,
Milagro,
ECUADOR

[2]Escuela Superior Politécnica de Chimborazo (ESPOCH),
Riobamba,
ECUADOR

[3]Universidad Nacional de Chimborazo,
Riobamba,
ECUADOR

[4]Independent Researcher,
Riobamba,
ECUADOR

[5]Universidad de las Fuerzas Armadas ESPE,
Latacunga,
ECUADOR

[6]Universidad Autónoma de Tlaxcala,
MEXICO

Abstract: - The research compares the performance of XGBoost with traditional logistic regression and decision tree approaches for identifying and predicting depression. It utilizes the Behavioral Risk Factor Surveillance System, which includes questions about mental and physical health, as well as some sociodemographic variables. XGBoost achieved 75.82% accuracy with an F1 score of 0.5385 and an AUC-ROC of 0.7649, demonstrating superior performance in the depression classification task, particularly with complex and dissociated data structures. Furthermore, the research highlights the importance of timely depression recognition for effective treatment and argues that XGBoost has the potential to improve current diagnostic practices by reducing current time and costs. The research also addresses the necessary future provisions for this approach and its ethical implications in healthcare.

Key-Words: - Machine learning, early detection, accuracy, prediction, XGBoost, Statical models. Depression.

Received: March 12, 2026. Revised: April 6, 2026. Accepted: May 13, 2026. Published: August 4, 2026.

1 Introduction

Depression is considered a common mental disorder that affects people around the globe. According to the World Health Organization (WHO), 322 million people have depression, a figure that increases each year, [1], [2].

The Patient Health Questionnaire (PHQ) is a tool developed by the American Psychological Association that integrates DSM-IV guidelines within a self-report format to systematically identify and diagnose depression. The disorder is characterized by lengthy episodes of withdrawal and sorrow, which impede the ability to carry out daily tasks and maintain social contacts, [3], [4], [5].

Even the non-specialized physicians find identifying depression is a challenge. The PHQ-9, which is one of the PHQ instruments, has been recognized as one of the key instruments for depression screening, particularly due to its specificity, sensitivity, and accuracy, [6], [7], [8]. Most traditional evaluations of the symptoms rely on recollections of the time and tend to miss the rapid changes in behavior and mood that are present right before a depressive episode, [6].

Statistical methods, as well as frameworks in machine learning, have recently expanded the research frontier on the recognition of depression. Specifically, the use of multimodal frameworks that integrate audio, written, and affective components of voice and text, as well as the computation of emotions in text, shows promise for automated depression detection. One of the works that successfully combined a bidirectional long-term memory model with a closed recurrent unit model was, [7]. The use of a specific method for text analysis described in [8] was able to achieve an F1 score of 0.83, which demonstrates the positive impact of machine learning in the area described in this paper.

The opportunity to observe and record symptoms continuously and in real time provides an opportunity to manage and prevent relapses in major depressive disorder, [9]. Greater recurrences in patients with major depressive disorder are largely driven by the absence of real-time capture of the depressive symptomatic state [9], as automated data collection enables detection of mood shifts and depression onset signs much earlier, [10].

Sophisticated predictive models that take into account the element of time should be integrated so that basic information is provided to the healthcare professionals. Despite these advancements, issues associated with the interpretability of the learning models and data biases arise, and these are considered critical hurdles that need to be addressed to enhance the relevance and reliability of the methods in clinical settings, [11], [12].

Between the two, machine learning methods and ensemble learning strategies are highly regarded for their ability to handle intricate data sets and optimize the outcomes of classification. The use of Random Forest and XGBoost models has also been mentioned in the literature as superior to previous techniques in extracting depression-related features, [13], [14].

Furthermore, the incorporation of machine learning technologies, specialized tools, and software allows for the fusion of various data types, such as demographics, biomarkers, and clinical data, thereby improving the accuracy of diagnostic tools. These improvements become invaluable when the primary concern is the early identification of depression for rapid and optimal interventions. It has been claimed that the Depression Stacking Generalization (DESGM) Ensemble model, which combines several machine learning techniques, produces significantly superior results compared to those generated by each of the independent models, [15], [16].

Another area of feature selection implemented in DESGM is exploring how Random Forest parameters can identify psychological and demographic characteristics associated with depression, which in turn helps identify individuals with potential depression in primary care. The numerous risk factors and diverse assessments required for depression make these algorithms suitable for its diagnosis, [17], [18].

Recent studies focusing on predicting mental health problems emphasize the incorporation of stacking techniques. These studies suggest that using stacked models for depression prediction greatly improves accuracy due to the collaborative effect of different algorithms, [19], [20]. Furthermore, Random Forest models identify critical factors such as educational level and social support that are strongly correlated with depression. These findings support the use of integrated methodologies within the operational framework of health systems for the early detection of individuals who are likely to be underserved.

Thanks to advances in artificial intelligence, algorithms such as Random Forest and XGBoost achieve remarkable accuracy in recognizing depression. Research indicates that these models achieve an accuracy rate of 84%, which is considerably higher than older techniques, such as support vector machines, which operate at lower levels of precision, [21], [22].

The reliability of depression screening initiatives is further strengthened by integrating advanced algorithms with both datasets, as seen in the DAIC-WOZ dataset, where the BiLSTM units, recurrent gates, and care coverage model achieved an F1 score of 0.83, [23], [24]. The confidence in the aforementioned studies is reflected in the literature, which reports a sensitivity range of 0.35 to 0.94 and a specificity range of 0.39 to 0.91. This implies that well-articulated evaluations of model performance under varied conditions are needed, [25], [26].

In relation to research on the automation of depression assessment, the construction of EATD-Corpus datasets containing audio interviews and their corresponding written transcripts is particularly valuable, [27], [28]. This type of research allows practitioners to effectively validate and train their models, thus aiding in the automation of depression assessment. By incorporating different types of data to improve overall accuracy, this research increases the number of methodologies developed for depression screening.

This research explores a central question: What strategies enable the rapid and cost-effective identification of individuals suffering from depression so they can receive support? To investigate this topic further, we turned to the XGBoost tool, a type of AI that analyzes numerous documents and recognizes correlations linked to patterns associated with depression. This research aims to address how this technology overcomes the limitations of conventional methods that struggle with the scope of the problem.

To effectively address this issue, it was necessary to integrate disparate variables related to mental and physical health, depression, and the social determinants of health, such as socioeconomic status, age, gender, and depression itself. This demonstration clearly demonstrated the contextual variables that influence the experience of illness. For a problem as simple as this, it follows that the more effectively information is captured to identify individuals with the potential to benefit from rapid interventions, the more comprehensive and accurate the data collection process should be.

The crucial nature of this stems from the fact that depression is not merely an individual problem, but a global public health issue affecting millions worldwide. The ability to identify such conditions early, within a short timeframe, determines whether an individual has a future full of opportunities or a life of suffering and relentless isolation. By using XGBoost, we hope to improve the accuracy of predictions to facilitate timely and effective interventions within the therapeutic window. It is during the most critical periods of their lives that people need the support we can provide.

2 Materials and Methods

2.1 Data Collection

The 2015 BRFSS database is a rich source of public health information, tracking a range of health behaviors, chronic conditions, and health care access and utilization within the U.S. population, [5]. The BRFSS is a telephone survey system that gathers information on health risk behaviors and health conditions and health services access and utilization. Along with the vast 2015 BRFSS database, the substantial variables integrated within the database include.

- Health-related physical activities
- Health impact risk behaviors (tobacco, alcohol)

The demographic variables were obtained from the telephone interviews conducted via mobile and traditional telephones (landlines). They were essential to describe the population aged 18 and over, living outside institutionalized settings, to conduct public health planning. These filters aid in health objective setting, health program evaluation, and health system performance monitoring at the regional or local levels.

2.2 Data Analysis

This dataset analysis assesses and contrasts various machine learning approaches on depression risk prediction using data from the 2014 Behavioral Risk Factor Surveillance System (BRFSS). Initially, the data is imported and organized, focusing on the following variables: mental and physical health, level of exercise, general health, alcohol consumption, smoking, income, and marital status. The main variable depression (ADDEPEV2) is recoded to indicate whether a person exhibits (1) or does not exhibit (0) depression symptoms, up to the point of singular depression detection and, in the process, ranks the depression-related factors in the dataset. This recoding offers the dataset a more appropriate structure for the prediction task.

After this, the following five machine learning techniques: logistic regression, decision tree, XGBoost, LightGBM, and neural networks, are implemented. Each of the models is trained on a balanced dataset generated using the Synthetic Minority Oversampling Technique (SMOTE) owing to the lower prevalence of depression cases in the public health records. Evaluating the models is done based on a set of standard metrics, i.e., accuracy, AUC-ROC and F1 Score, classification reports, and confusion matrices to provide a detailed understanding of the predictive power of the models.

The information shown in Figure 1 is crucial for understanding the depression prediction results. The bar charts present a simplified comparison of multiple models and their corresponding performance metrics for effective depression prediction.

A confusion matrix was constructed for each model, and each matrix was then illustrated as a heat map. This allowed for distinguishing patterns caused by correct and incorrect identifications.

These tools are vital for stakeholders and the iterative refinement of the model.

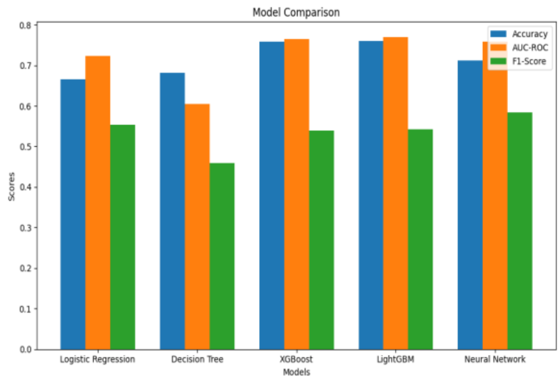

Fig. 1: Used Model Comparison
Sources: Created for authors

This was the first evaluation utilizing multiple machine learning approaches to analyze the predictive capabilities of depression based on behavioral risk factors. The results of this study can inform the early stages and help develop efficient public health interventions in targeting high-risk groups for the potential development of depressive disorders.

3 Results and Discussion

Table 1 illustrates the XGBoost (Extreme Gradient Boosting) method's efficiency in detecting depression and improving mental health evaluations. Among decision tree-based supervised learning methods, this technique has outperformed other predictive methods in the literature by a considerable margin.

Table 1. Metrics of the models used

Model	Accuracy	AUC-ROC	F1-Score	Accuracy (0/1)	Recall (0/1)
Logistic Regression	0.6619	0.7240	0.5542	0.82 / 0.47	0.65 / 0.68
Decision Tree	0.7273	0.7488	0.5707	0.81 / 0.55	0.79 / 0.59
Gradient Boosting (XGBoost)	0.7582	0.7649	0.5385	0.79 / 0.65	0.89 / 0.46
LightGBM	0.7600	0.7693	0.5436	0.79 / 0.66	0.89 / 0.46
Neural network	0.7171	0.7596	0.5837	0.83 / 0.53	0.75 / 0.64

Sources: Created for authors

The confusion matrix illustrated in Figure 2 pertains to one binary classification model developed with the XGBoost approach. The confusion matrix indicates that 41,676 'No's and 9,526 'Yes' responses were correctly predicted by the model. This demonstrates the model's ability to distinguish between those who have depression and those who do not.

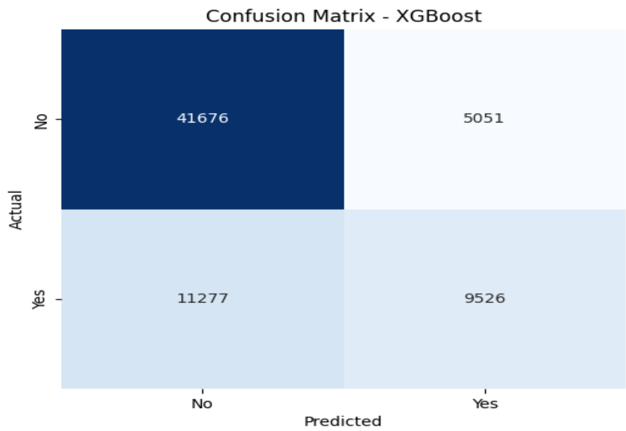

Fig. 2: Correlation between variables
Sources: Created for authors

The results indicate that the model effectively captures instances classified as "No." However, the concern regarding the 11,277 false negatives means that too many true "Yes" instances were also classified as "No." There are also 5,051 cases where the model incorrectly recognized "Yes" instances that were actually "No." These are false positives.

Within the context of this study, compared to the type of analysis that captures model performance, the specific context surrounding the model remains important. In situations where substantial detection gaps exist, such as in the case of "Yes" responses from potentially depressed individuals, the lack of detection becomes a risk that must be mitigated, especially regarding the risk of false negatives. Depending on the context and tolerance, false positives can be equally significant in the overall outcome, especially where the false positive result is less substantial.

The overall confusion matrix demonstrates the model's ability to identify instances classified as "No" and the need for model tuning to improve the detection of "Yes" instances. Suggested improvements to the model include hyperparameter adjustments, the addition of new predictive features, and new data sampling techniques to achieve better overall class balance.

A ROC curve, presented in Figure 3, evaluates the discriminative ability of the XGBoost classification model.

This illustrates the relationship between the true positive rate (sensitivity) and the false positive rate (1 - specificity) for different classification thresholds. The model's performance, represented by the orange curve, is compared to the blue dashed line, which represents the line of no discrimination for random classifiers.

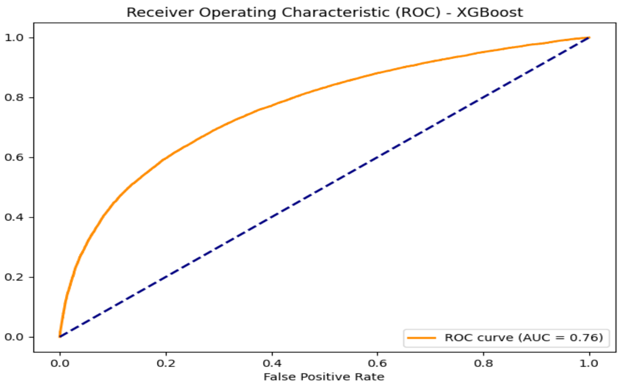

Fig. 3: XGBoost ROC Curve
Sources: Created for authors

Calculating the area under the curve (AUC) yields a value of 0.76. This means the model demonstrates a strong proficiency in differentiating between positive and negative instances. An AUC of 0.76 indicates that the model has a 76% likelihood of accurately determining if a randomly chosen case is positive or negative.

In relation to the ROC curve, the model achieves a satisfactory level of true positives while incurring a minimal level of false positives, which is a favorable outcome within a wide range of use cases. The ROC curve, however, indicates that the true positive rate is not responding to the changes in the false positive rate. This warns that in some segments, the model may not be configuring an optimal equilibrium among sensitivity, specificity, or both.

This acceptable performance, combined with the findings in the literature, indicates that adjusting the hyperparameters, model features, or more sophisticated ensembles to improve discrimination power [29], [30] would provide a measurable increase.

In the study that aimed to improve depression diagnosis with XGBoost, the various model metrics (balanced accuracy, accuracy, recall, and F1 score) exceeded 0.90 in balanced samples (with oversampling and combined oversampling with undersampling techniques)2. Therefore, XGBoost can be said to, with high confidence, diagnose depression from biomarkers, which is a finding congruent with the literature [31], [32] and is potentially a useful adjunct streamlining the current interviews, which are standardized.

In its multitude of optimization processes, XGBoost's performance and accuracy can be remarkably enhanced, [33] ,[34]. To mention a few, there is regularization for overfitting, branch pruning in decision trees, and parallel learning in blocks of data, not to mention sparse data techniques, which deal with overfitting and strong pruning control, [35], [36].

XGBoost's ability to perform well on class-imbalanced data gives it an advantage over other algorithms on both classification and regression problems. This is particularly true for mental health data. Regarding depression detection, the algorithm has identified target biomarkers and predicted depression cases with excellence, partly due to the under-identified cases of depression within the data, [37].

For mental health, and especially for depression prediction, the use of XGBoost will likely foster further improvements within the field and in clinical practice. The rapid and reliable classification of depression using biomarker data will facilitate XGBoost's assistance in traditional depression diagnoses, saving time and money currently spent on structured clinical interviews, [38], [39].

XGBoost's ability to determine the significance of features and manage unbalanced data sheds light on the biological aspects of depression. Although XGBoost has excelled in classification tasks, we must emphasize the importance of understanding the consequences of placing any model in clinical use. This implies further validation that assesses the ethics of using algorithms for automated mental health surveillance and care, [40].

To conclude, XGBoost stands out as vital for improving the speed and efficiency of treatment for the identification and assessment of depression. In the field of mental health, the evolution of research has had positive effects on treatment and the well-being of patients as a whole, [41]. Understanding the possible applications in healthcare requires the analysis of the strengths and weaknesses of each predictive model. Each model has strengths and weaknesses that could greatly impact a clinician's choice. For example, highly predictive models could, at the same time, be uninterpretable, resulting in a model that a clinician is unwilling, or unable, to use, [42].

On the other hand, a model may be parsimonious, highly interpretable, and clinically relevant, but at a cost to precision. Hence, the importance of the application in real-world healthcare to achieve positive outcomes for patients and enhanced levels of care is the assessment of these instances where compromises arise, [43].

More than predictiveness and interpretability, the implementation of these predictive models in practice is crucial. Patient, condition, and resource levels vary greatly, and all impact the effectiveness of models in practice, [44].

A model may work well in a research setting, but it may not work in a different clinical setting because of different patient characteristics or different workflows. Therefore, a model should have contextual validation so that it can be operationalized and it can be functionally validated in real-world usage, [45].

On top of that, an adequate description of a model's output is important in building trust and communication between the healthcare professionals and customers. For example, clinicians must be able to explain the reasoning behind a model's prediction, especially in cases where the prediction is used to make a critical clinical decision, [46].

A lack of explanatory power in clinical models can create complexity and ambiguity, which can lead to erosion of trust on the part of clinicians and patients. Weak trust in the model may adversely affect the therapeutic relationship. Trust model development is important in order to not just in clinical judgment but also in the patient's involvement and compliance in the treatment plan.

4 Conclusions

This investigation has shown that the Gradient Boosting method (XGBoost) has been powerful in recognizing and forecasting depression utilizing an extensive body of statistical usable data, including mental and physical health as well as demographic data. XGBoost has been shown to significantly exceed the performances of legacy models of Logistic Regression and Decision Trees in capturing the criteria of accuracy, AUC-ROC, f1-score, and many other performance measures. For example, XGBoost has registered an accuracy of 75.82%, AUC-ROC of 0.7649, and f1-score of 0.5385, demonstrating the improved reliability and correctness of the aforementioned measures.

The data derived from the confusion classes and the ROC curve provide supporting evidence on the aforementioned claim, having high confidence in true positives of the predicted depression cases, and showing that the positive and negative cases of depression have predictive separability. Clearly, XGBoost has illustrated that it is a robust tool in predicting depression and has immense predictive value to help mitigate the clinical implications of depression.

Upon closer examination of performance metrics, it becomes clear that XGBoost not only outperforms older, traditional methods in terms of accuracy but also performs exceptionally well on unbalanced and challenging datasets. Most notably, it is adept at identifying and leveraging key predictive variables and accurately estimating the number of predicted depressive cases, even in datasets with extremely low prevalence.

However, XGBoost, as promising as it is in this study, still requires further validation and consideration of the ethical implications of using AI in mental health diagnosis, particularly in clinical practice. While clinical psychiatry in practice involves recognizing depression during the "golden hour" to optimize patient care and well-being, the overall impact on the healthcare system is significant.

When integrating machine learning models into clinical practice, many more hurdles must be addressed to demonstrate claims of accuracy, fairness, and transparency. That said, XGBoost has the potential to revolutionize the diagnosis of depression, and this study certainly paves the way for improved early public mental health interventions.

References:

[1] Dhara, S., Thakur, J., Pandey, N., Mozumdar, A., & Roy, S. (2024). Prevalence of major depressive disorder and its determinants among young married women and unmarried girls: Findings from the second round of UDAYA survey. PloS one, 19(7), e0306071. https://doi.org/10.1371/journal.pone.0306071.

[2] Gilbert, W., Eltanoukhi, R., Morin, A. J., & Salmela-Aro, K. (2025). Achievement goals as mediators of the links between self-esteem and depressive symptoms from mid-adolescence to early adulthood. Journal of youth and adolescence, 54(1), 103-120. https://doi.org/10.1007/s10964-024-02045-z.

[3] Ortiz Morán, M., Vásquez Vega, J., Correa Aranguren, I. G., Reyes Rodríguez, J. A., Laura Barraza, E. W., & Segovia, J. L. (2024). Cuestionario de Salud del Paciente-9 (PHQ-9): una revisión sistemática y un metaanálisis de la generalización de la confiabilidad. Revista de Neuro-Psiquiatría, 87(3), 273-288. https://doi.org/10.20453/rnp.v87i3.5264.

[4] Hao, Y., Zhang, J., Yang, L., Zhou, C., Yu, Z., Gao, F., ... & Yu, J. (2023. British Journal of Clinical Pharmacology, 89(9), 2714-2725. http://dx.doi.org/10.1111/bcp.15734

[5] Gyimah, L., Agyepong, I. A., Owiredu, D., Awini, E., Yevoo, L. L., Ashinyo, M. E., ... &

Danso-Appiah, A. (2024). Tools for screening maternal mental health conditions in primary care settings in sub-Saharan Africa: systematic review. Frontiers in Public Health, 12, 1321689. https://doi.org/10.3389/fpubh.2024.1321689.

[6] Mahmood, S., Tan, X., Chen, B., & Tor, P. C. (2024). The influence of age on ECT efficacy in depression, mania, psychotic depression, and schizophrenia: A transdiagnostic analysis. Journal of Psychiatric Research, 177, 203-210. https://doi.org/10.1016/j.jpsychires.2024.07.012.

[7] Quang-Anh, N. D., Ha, M. H., Dinh, T. K., Pham, M. D., & Ninh, N. V. (2024). Emotional Vietnamese Speech-Based Depression Diagnosis Using Dynamic Attention Mechanism. arXiv preprint arXiv:2412.08683. http://dx.doi.org/10.48550/arXiv.2412.08683.

[8] Wu, C. S., Chen, C. H., Su, C. H., Chien, Y. L., Dai, H. J., & Chen, H. H. (2023). Augmenting DSM-5 diagnostic criteria with self-attention-based BiLSTM models for psychiatric diagnosis. Artificial Intelligence in Medicine, 136, 102488. https://doi.org/10.1016/j.artmed.2023.102488.

[9] Frey, M., Smigielski, L., Tini, E., Fekete, S., Fleischhaker, C., Wewetzer, C., ... & Egberts, K. M. (2023). Therapeutic Drug Monitoring in Children and Adolescents: Findings on Fluoxetine from the TDM-VIGIL Trial. Pharmaceutics, 15(9), 2202. http://dx.doi.org/10.3390/pharmaceutics15092202.

[10] Minervini, G., Marrapodi, M. M., Siurkel, Y., Cicciù, M., & Ronsivalle, V. (2024). Accuracy of temporomandibular disorders diagnosis evaluated through the diagnostic criteria for temporomandibular disorder (DC/TDM) Axis II compared to the Axis I evaluations: a systematic review and meta-analysis. BMC Oral Health, 24(1), 299. http://dx.doi.org/10.1186/s12903-024-03983-7.

[11] Shivaprasad, S., Chadaga, K., Sampathila, N., Prabhu, S., Chadaga P, P. R., & KS, S. (2024). Explainable machine learning methods to predict postpartum depression risk. Systems Science & Control Engineering, 12(1), 2427033. http://dx.doi.org/10.1080/21642583.2024.2427033.

[12] Feng, W., Wu, H., Ma, H., Yin, Y., Tao, Z., Lu, S., ... & Liu, Y. (2025). Deep learning based prediction of depression and anxiety in patients with type 2 diabetes mellitus using regional electronic health records. International Journal of Medical Informatics, 196, 105801. https://doi.org/10.1016/j.ijmedinf.2025.105801

[13] Meyer, T. N., Andreeva, O., Weiss, R. D., Ding, W., Shen, I., Wang, C., ... & Dagnew, T. M. (2025). Enhancing neuromolecular imaging classification in low-data regimes with generative machine learning: A case study in HDAC PET/MR imaging of alcohol use disorder. Neuroscience Informatics, 100225. https://doi.org/10.1016/j.neuri.2025.100225.

[14] Shen, L., Xu, X., Yue, S., & Yin, S. (2024). A predictive model for depression in Chinese middle-aged and elderly people with physical disabilities. BMC psychiatry, 24(1), 305. http://dx.doi.org/10.1186/s12888-024-05766-4.

[15] Benacek, J., Lawal, N., Ong, T., Tomasik, J., Martin-Key, N. A., Funnell, E. L., ... & Bahn, S. (2024). Identification of predictors of mood disorder misdiagnosis and subsequent help-seeking behavior in individuals with depressive symptoms: gradient-boosted tree machine learning approach. JMIR mental health, 11, e50738. https://doi.org/10.2196/50738.

[16] Wadhvani, B., Mehta, J., Singh, K., Oza, A., Kumar, S., Gupta, S., & Sobti, R. (2025). Leading-edge sentiment analysis: A survey of application context, challenges, and advanced techniques. Recent Advances in Computer Science and Communications, 18(6), E26662558312258. http://dx.doi.org/10.2174/012666255831225824091111028.

[17] Guayasamin, M., Depaauw-Holt, L. R., Adedipe, I. I., Ghenissa, O., Vaugeois, J., Duquenne, M., ... & Murphy-Royal, C. (2025). Early-life stress induces persistent astrocyte dysfunction associated with fear generalisation. Elife, 13, RP99988. https://doi.org/10.7554/elife.99988.

[18] Liu, L., Jia, H., Qiu, B., Zhang, A., & Zhang, Q. (2025). The relationship between mindfulness and depression: examining the chain mediating role of shyness and core self-evaluation. BMC Psychology, 13(1), 1-14. http://dx.doi.org/10.1186/s40359-025-02774-1.

[19] Srinivasan, R., Lewis, G., Roiser, J., & Rice, F. (2024). Irritability and Risk-Taking Behaviour on the Cambridge Gambling Task (CGT) in Adolescents With a Family History of Depression. BJPsych Open, 10(S1), S83-S83. http://dx.doi.org/10.1192/bjo.2024.250.

[20] Ikäheimonen, A., Luong, N., Baryshnikov, I., Darst, R., Heikkilä, R., Holmen, J., ... & Aledavood, T. (2024). Predicting and

monitoring symptoms in patients diagnosed with depression using smartphone data: Observational study. Journal of Medical Internet Research, 26, e56874. https://doi.org/10.2196/56874.

[21] Arora, A., Chakraborty, P., & Bhatia, M. P. S. (2023). Identifying digital biomarkers in actigraph-based sequential motor activity data for assessment of depression: a model evaluating SVM in LSTM extracted feature space. International Journal of Information Technology, 15(2), 797-802. http://dx.doi.org/10.1007/s41870-023-01162-5.

[22] Xie, Y., Zheng, H., Gan, W., Su, C., Shams, M., & Yang, J. (2025). The performance of machine learning models in predicting postpartum depression: a meta-analysis and systematic review. Journal of Reproductive and Infant Psychology, 1-18. https://doi.org/10.1080/02646838.2025.2517103.

[23] Thekkekara, J. P., Yongchareon, S., & Liesaputra, V. (2024). An attention-based CNN-BiLSTM model for depression detection on social media text. Expert Systems with Applications, 249, 123834. https://doi.org/10.1016/j.eswa.2024.123834.

[24] Zhang, Z., Zhang, S., Ni, D., Wei, Z., Yang, K., Jin, S., ... & Wang, J. (2024). Multimodal sensing for depression risk detection: Integrating audio, video, and text data. Sensors, 24(12), 3714. http://dx.doi.org/10.3390/s24123714.

[25] Dixit, A., Gupta, A., Chaplot, N., & Bharti, V. (2024). Emoji support predictive mental health assessment using machine learning: unveiling personalized intervention avenues. International Academic Publishing House, 42, 228-240. http://dx.doi.org/10.52756/ijerr.2024.v42.020.

[26] Parsons, M., Qiu, L., Levis, B., Fan, S., Sun, Y., Amiri, L. S., ... & Thombs, B. D. (2024). Depression prevalence of the Geriatric Depression Scale-15 was compared to Structured Clinical Interview for DSM using individual participant data meta-analysis. Scientific Reports, 14(1), 17430. https://doi.org/10.1038/s41598-024-68496-3.

[27] Li, Y., Kumbale, S., Chen, Y., Surana, T., Chng, E. S., & Guan, C. (2025). Automated Depression Detection from Text and Audio: A Systematic Review. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.1109/jbhi.2025.3570900.

[28] Hamidi, H., & Ghelichi, M. (2025). A hybrid deep learning technique for depression detection of Persian tweets using using Bi-directional LSTM-CNN model. Multimedia Tools and Applications, 1-31. http://dx.doi.org/10.1007/s11042-025-21092-7.

[29] Ayodele, A., Adetunla, A., & Akinlabi, E. (2024). Prediction of Depression Severity and Personalised Risk Factors Using Machine Learning on Multimodal Data. International Journal of Online & Biomedical Engineering, 20(9). http://dx.doi.org/10.1109/IS61756.2024.10705185.

[30] Khan, S., & Alqahtani, S. (2024). Hybrid machine learning models to detect signs of depression. Multimedia Tools and Applications, 83(13), 38819-38837. http://dx.doi.org/10.1007/s11042-023-16221-z.

[31] Ghosal, S., & Jain, A. (2023). Depression and suicide risk detection on social media using FastText embedding and XGBoost classifier. Computer Science, 218, 1631-1639. https://doi.org/10.1016/j.procs.2023.01.141.

[32] Rumahorbo, B. N., Nanggala, K., Elwirehardja, G. N., & Pardamean, B. (2023). Analyzing important statistical features from facial behavior in human depression using XGBoost. Commun. Math. Biol. Neurosci., 2023, Article-ID. https://doi.org/10.28919/cmbn%2F7916.

[33] Khurshid, M. R., Manzoor, S., Sadiq, T., Hussain, L., Khan, M. S., & Dutta, A. K. (2025). Unveiling diabetes onset: Optimized XGBoost with Bayesian optimization for enhanced prediction. PloS one, 20(1), e0310218. https://doi.org/10.1371/journal.pone.0310218.

[34] Sim, G. H., Park, M. G., Cha, K. H., Yoo, Y., Kim, Y., & Lee, D. (2024). XGBoost-based anomaly detection for quality management of spent fuel safety information. Journal of Nuclear Science and Technology, 1-13. http://dx.doi.org/10.1080/00223131.2024.2434566.

[35] Cao, Y., Dai, J., Wang, Z., Zhang, Y., Shen, X., Liu, Y., & Tian, Y. (2025). Machine Learning Approaches for Depression Detection on Social Media: A Systematic Review of Biases and Methodological Challenges. Journal of Behavioral Data Science, 5(1), 67-102. http://dx.doi.org/10.35566/jbds/caoyc.

[36] Pandya, S. (2023). A Machine Learning Framework for Enhanced Depression Detection in Mental Health Care Setting. Int. J.

Sci. Res. Sci. Eng. Technol, 356-368. https://doi.org/10.32628/ijsrset2358715.

[37] Du, W. (2022). Application of improved smote and XGBoost algorithm in the analysis of psychological stress test for college students. Journal of Electrical and Computer Engineering, 2022(1), 2760986. https://doi.org/10.1155/2022/2760986.

[38] Madububambachu, U., Ukpebor, A., & Ihezue, U. (2024). Machine learning techniques to predict mental health diagnoses: A systematic literature review. Clinical Practice and Epidemiology in Mental Health: CP & EMH, 20, e17450179315688. https://doi.org/10.2174/0117450179315688240607052117.

[39] Han, Y., Wei, Z., & Huang, G. (2024). An imbalance data quality monitoring based on SMOTE-XGBOOST supported by edge computing. Scientific Reports, 14(1), 10151. http://dx.doi.org/10.1038/s41598-024-60600-x.

[40] Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. Technologies, 13(3), 88. https://doi.org/10.3390/technologies13030088.

[41] Li, S., Lu, P., Liang, W., Chen, Y., & Da, Q. (2024). Performance evaluation of hybrid PSO-BPNN-AdaBoost and PSO-BPNN-XGBoost models for rockburst prediction with imbalanced datasets. Applied Sciences, 14(24), 11792. https://doi.org/10.3390/app142411792.

[42] Ding, Z., Wang, Z., Zhang, Y., Cao, Y., Liu, Y., Shen, X., ... & Dai, J. (2025). Trade-offs between machine learning and deep learning for mental illness detection on social media. Scientific Reports, 15(1), 14497. https://doi.org/10.1038/s41598-025-99167-6.

[43] Ntam, V. A., Huebner, T., Steffens, M., & Scholl, C. (2025). Machine learning approaches in the therapeutic outcome prediction in major depressive disorder: a systematic review. Frontiers in Psychiatry, 16, 1588963. https://doi.org/10.3389/fpsyt.2025.1588963.

[44] Dang, V. N., Campello, V. M., Hernández-González, J., & Lekadir, K. (2025). Empirical comparison of post-processing debiasing methods for machine learning classifiers in healthcare. Journal of Healthcare Informatics Research, 1-29. https://doi.org/10.1007/s41666-025-00196-7.

[45] Liu, Y., Fang, A., Moriarty, G., Firman, C., Kraut, R. E., & Zhu, H. (2024). Exploring trade-offs for online mental health matching: Agent-based modeling study. JMIR Formative Research, 8, e58241. doi: 10.2196/58241.

[46] Shah, H. H. AI-Enhanced Depression and Anxiety Detection: Integrating EEG Systems, Performance-Cost Trade-Offs, and Optimization Algorithms. Global Journal of Emerging AI and Computing, 1(1), 59-69. http://dx.doi.org/10.1109/ICCCIT62592.2025.10928002.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.