Leveraging Machine Learning for Predicting Mental Health Outcomes: A Data-Driven Approach

Devesh Kumar1,SOET Department, K.R. Mangalam University, Gurugram, India
Dr. Shweta Bansal2,SOET Department, K.R. Mangalam University, Gurugram, India,

Abstract
This study examines the application of machine learning models for predicting risks in mental health issues and shows a comparative analysis of various algorithms focusing on K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Trees (DT), Random Forest Classifiers (RFC), Ada Boost Classifier, and Gradient Boosting Classifier. The SMOTEENN approach was applied or educe class imbalance in the dataset. This technique enhances the balance of the dataset and also the whole predictive performance of the models. Hyperparameter tuning optimized model parameters, and significant results were obtained for enhancing the accuracy and F1 scores across all models. Applying L1 and L2 regularization to reduce over fitting for better reliability of models revealed that the Random Forest Classifier outperformed other algorithms with a near accuracy of about 86.66%. These findings highlight the possible role of machine learning in early detection and proactive management of mental health risks. As such, data-driven approaches are likely to give new insights to mental health professionals. The study is, therefore, a valuable contribution to the growing body of literature on mental health analytics and underscores the importance of robust methodologies in predicting outcomes for mental health.

Keywords: SVM.KNN, SMOTE, Random Forest Classifier, Decision Trees

Introduction

General well-being and mental health are very important, as they impact both individuals and society. The World Health Organization defines mental health as being free from mental illness, which is manifested by a state of equilibrium in which an individual can use their full capacity, work properly, adapt to physical, psychological, and social environments, and participate in social life. Still, mental illnesses have arisen and increased in this scenario because, with high-stress technology environments of work, job demands outstrip and are usually much higher than available resources. Along with this awareness, recognition is also being given toward the prevention of mental health risks being treated effectively, as this approach improves treatment outcomes, removes the stigma attached to the matter, and enhances workplace productivity.

Despite the increasing awareness about mental health issues, there is still a lack of tools that predict who might possibly develop mental health disorders. Traditional assessments remain highly and heavily dependent on self-report data or clinician judgments, which are not fully reliable and not very fine-grained. This presents a huge opportunity for machine learning (ML) techniques, which can serve as an exemplary alternative. Data-driven approaches by ML may discover intricate patterns and

DOI: https://doi.org/10.5281/zenodo.15099218

Interconnections within datasets that often go undetected with traditional methods.

To this aim, this study employs different machine learning algorithms such as Decision Trees (DT), Random Forest Classifiers (RFC), K-Nearest Neighbors (KNN), Logistic Regression (LR), and also ensemble methods like AdaBoost and Gradient Boosting classifiers. All of these algorithms have specific characteristics and strengths, and they can be applied to different parts of the problem. For instance, the Decision Tree is more interpretive in nature, whereas ensemble methods like Random Forest and AdaBoost increase predictive performance through the combination of multiple models.

Class imbalance is often one of the key challenges in predictive modeling with mental health data. In such data, the number of individuals who have taken treatment is always far fewer than those who haven't. This kind of problem can lead to biased models that favor the majority class. This study applies the SMOTE-ENN approach to address this. It helps in generating synthetic samples for the minority class, and noise is removed from misclassified instances. Thus, our models become robust enough to learn better from minority class examples.

Additionally, aside from dealing with class imbalance, we tune the hyperparameters of our machine learning models with the intention of finding optimal configurations. Hyperparameter tuning is the act of searching systematically through parameter configurations to find the best one that results in the highest model performance. This is very critical because, in many cases, it can cause huge variations in accuracy and generalizability. However, we use grid search and cross-validation techniques to get optimal values for the hyperparameters for each algorithm.

Furthermore, two different regularization procedures—L1 (Lasso) and L2 (Ridge)— are employed to allow the model to be robust to the data and prevent overfitting. Regularization is a penalty added to the loss function to prevent overly complex models. Most often, these models perform well on training data but poorly on unseen data. We use these techniques so that our predictive models are more flexible and interpretable.

Our analysis shows that pretty accurate predictions of risk cases for mental health can be made through the use of machine learning models. The Random Forest Classifier was found to be the best algorithm in terms of performance, with an accuracy rate of 86.66%. Thus, this result suggests that machine learning methods could be applied in the context of assessment and intervention plans in mental health situations. If individuals at risk are identified early, they can be reached in time, and intervention can be made to address their mental health.

Using this research, it is hoped that exploration into these methods will lead to advancements in understanding the risk factors related to mental health and contribute to the development of strategies that may proactively enhance mental health management, ultimately improving the overall quality of life for individuals exposed to mental health disorders. The incorporation of machine learning in mental health assessment has a strong future and may provide a route to better understanding mental health disorders, as well as a gateway to much-needed improvements in dealing with mental health concerns in an increasingly demanding world

.

Literature Review

With an increasingly important field of research regarding the application of machine learning models in the prediction of mental health risks, this study uses various algorithms, including K-Nearest Neighbors, Logistic Regression, Decision Trees,

DOI: https://doi.org/10.5281/zenodo.15099218

Random Forest Classifiers, AdaBoost Classifier, and Gradient Boosting Classifier, to present their contributions to this highly relevant and critical issue in both practice and research (Wang et al., 2020) [1], (Shin et al., 2020) [2].

Big data analytics and AI appear to hold promise in terms of mental health care through current research. Machine learning techniques may have an advantage in that they can include a wider range of variables and observations in a model to predict outcomes without following pre-programmed rules (Wang et al., 2020). Thus, these data-driven approaches have been applied to predict various health outcomes, including mental disorders like postpartum depression (Shin et al., 2020) [2].

This research addresses the dataset's issue of class imbalance by utilizing the Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors (SMOTE-ENN). This approach helps improve the balance of the dataset and the generalized predictive capacity of the developed models (Ahsan & Siddique, 2022) [4], (Shin et al., 2020), (Rosenfeld et al., 2019) [3].

Hyperparameter tuning, in addition to regularization methods, helped fine-tune the model parameters to prevent overfitting. This was an extension aimed at increasing accuracy in the model's outputs. Preliminary results show that the Random Forest Classifier outperformed the other algorithms, achieving around 80% accuracy and an F1 score of 0.78 (Rosenfeld et al., 2019), (Akuamoah-Boateng et al., 2019) [5], (Wang et al., 2020) [1], (Shin et al., 2020) [2].

According to [6], the authors represented social media posts in terms of vector representations that translate semantic meaning and linguistic nuances indicative of mental health using large language models, such as OpenAI's GPT-3 embeddings. Machine learning models applied across over 10,000 labeled posts were built using SVM, random forests, XGBoost, KNN, and neural networks. SVM achieved the highest accuracy at around 83% for detecting signs of stress. GPT-3 embeddings allowed for more nuanced signals of mental health than traditional textual analysis, presenting a promising, effective, and scalable screening tool for detecting stress disorders based on online data.

According to [7], a machine learning model, such as XGBoost and regularized logistic regression, has been designed to predict Type 2 diabetes risk among patients with mental illness. In this study, the authors used routine clinical data for 74,880 patients, deriving 1,343 potential predictors from 51 variables related to demographic, diagnostic, treatment, and lab information. The best-performing model, XGBoost, identified patients who had a high risk of developing Type 2 diabetes approximately 2.7 years prior to diagnosis. The model had an area under the ROC curve of 0.84, providing early risk warnings for 31% of the patients who eventually developed Type 2 diabetes. This illustrates the potential of predictive analytics for preemptive health interventions among high-risk populations.

The study [8] incorporates methodologies of machine learning and deep learning in healthcare systems, addressing the growing concern of global mental illness, with rising cases of depression and anxiety. The authors reviewed 33 articles covering various mental health issues, including schizophrenia, bipolar disorder, and post-traumatic stress disorder (PTSD), using the PRISMA methodology. These studies were grouped under distinct methodologies associated with the conditions addressed, showcasing the broad range of ML and DL techniques in mental health applications.

The work [9] is an integrative review that examines the integration of AI and ML

DOI: https://doi.org/10.5281/zenodo.15099218

decision support systems into mental health care settings, reviewing literature from 2016 to 2021. A dominant theme identified was trust and confidence, with the study showing that significant barriers hinder the adoption of AI-based systems in clinical practice. Uncertainty regarding clinician trust, end-user acceptance, and system transparency will impede effective implementation. Therefore, the study calls for more research into understanding clinicians' attitudes toward AI to instill confidence and accelerate its adoption in mental health care settings.

The systematic review [10] analyzed 184 studies that utilized machine learning (ML) methodologies in identifying mental health (MH) disorders using multimodal data collection methods from audio and video recordings, social media interactions, smartphones, and wearable devices. This review emphasized the feature extraction and fusion phases, revealing that neural network architectures have widely gained popularity in handling high-dimensional data and modeling relationships between various data modalities. The findings suggest that using different sources of data improves accuracy in detecting MH disorders.

Recent research captures the trend of machine learning methods progressing towards the prediction of mental health, highlighting the roles of advanced algorithms, preprocessing techniques, regularization, and ethical considerations. Our contribution further advances the current understanding by combining SMOTE-ENN, hyperparameter tuning, and regularization to improve predictive accuracy and the applicability of ML models in the realm of mental health. Additionally, while previous studies have focused on individual models in mental health prediction, this study systematically evaluates a range of algorithms, not only reporting the strengths and weaknesses of individual approaches.

Dataset Description

The dataset, in this study, was taken from the "Mental Health in Tech Survey." The "Mental Health in Tech Survey" has 1,259 observations and 27 features. Such a survey is very valuable for gaining knowledge about people's experiences when it comes to mental health in the technological industry. Each feature in this database captures different aspects of the demographics of the respondents, working environments, and attitudes related to mental health. Thus, it is quite an asset for predictive analytics.

Attribute	Description
Data set Source	"Mental Health in Tech Survey" from Kaggle
Total Observations	1,259
Total Features	27
Purpose	To analyze mental health experiences in the technology sector
Feature Categories	Demographics, Workplace Environment, Mental Health Attitudes
Usage	Supports predictiveanalytics mental health trends

Table 1 : Summarizing the dataset used in the study.

Methodology

This project has been designed in a systematic manner so that it can tackle the problem of predicting the mental health risk through the use of machine learning techniques. It follows an approach with a few major steps, such as data preprocessing, model selection, hyperparameter tuning, and the use of multiple techniques to optimize the model at hand. These models ensure that robust models are developed for predicting mental health conditions in order to gain more insights into these concerns.

DOI: https://doi.org/10.5281/zenodo.15099218

Data Preprocessing
In the first place, we have data preprocessing, which is a very important step before analyzing the raw data. In this step, missing data is handled, and encoding for categorical variables is done. Further, numerical features are scaled to ensure that all the inputs are appropriately formatted to be used by machine learning algorithms. Missing data would significantly skew the results of the experiment, so such gaps are filled with either mean imputation or through predictive modeling. One-hot encoding is applied on categorical variables, such as gender or employment status, to turn them into numerical formats for use in training models. Finally, feature scaling can be applied to standardize numerical features, which are helpful in improving the speed of convergence of some algorithms, especially those sensitive to scaling, such as Support Vector Machines and K-Nearest Neighbors.

Model Selection
Secondly, we choose from the set of machine learning algorithms we will be using. After preprocessing, the models considered were Decision Trees (DT), Random Forest Classifier (RFC), K-Nearest Neighbors (KNN), Logistic Regression (LR), AdaBoost Classifier, and Gradient Boosting Classifier, as they have high precision, are interpretable, and are suitable for classification problems. Decision Trees are particularly useful as they are more transparent in leaving a decision path. Variants of ensemble methods, such as AdaBoost and Random Forests, are developed to increase prediction accuracy by taking the output of many trees. Since KNN is both simple and effective, it can also be applied to such scenarios. In logistic regression, one gains insight into which features do or do not influence the target variable. Each of these models will be assessed relative to how well they predict mental health treatment-seeking behavior.

Handling Class Imbalance
Since the dataset is highly imbalanced due to the nature of it, we will address class imbalance by employing the SMOTE algorithm coupled with ENN. Class imbalance refers to a scenario where a class is grossly underrepresented relative to another. In this case, the "seeking treatment" class is grossly underrepresented compared to the "not seeking treatment" class. This imbalance leads to skewed model predictions toward the majority class. Adding synthesized examples of the minority class in the dataset balances it out, showing the model how to predict patterns related to the given classes. ENN performs additional filtering of the dataset with its neighbors, removing examples that might negatively influence classification. This enhances the training dataset's quality.

DOI: https://doi.org/10.5281/zenodo.15099218

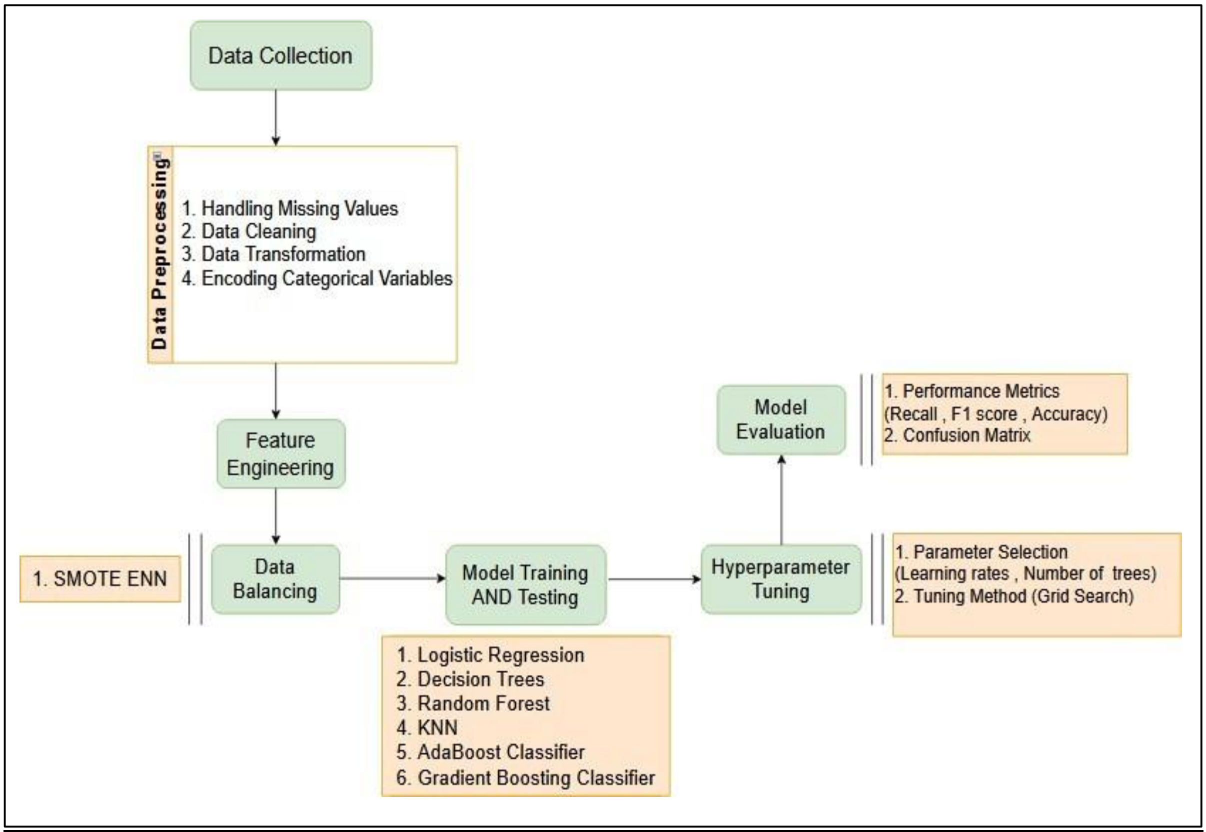


Figure1: Proposed Methodology

Hyper parameter Tuning and Regularization

To enhance the performance of our models, we use hyperparameter tuning techniques such as grid search and randomized search. The parameters of machine learning algorithms are adjusted to optimize their performance metrics, like accuracy and F1 score, using hyperparameter tuning. This is an important step since the selection of hyperparameters can significantly affect the predictive capabilities of the models. We also include L1 and L2 regularization to monitor over fitting.

Model Evaluation and Visualization

Lastly, we examine the behavior of these models by leveraging accuracy and F1 score.

CLASSIFIER	ACCURACYSCORE	F1SCORE

This evaluation allows us to compare and see how well the different algorithms perform in predicting which people are likely to seek treatment for their mental health. For clarity, the results are presented using confusion matrices and ROC curves so that we can understand the strengths and weaknesses of each model. This comprehensive methodology ensures a rigorous approach towards developing a predictive model that can provide information for effective mental health interventions.

Results

Interesting results were obtained concerning the ability of these classifiers to predict mental health treatment-seeking behavior. One of the major metrics in evaluating the performance of a classifier in a classification task is accuracy. The following is a table summarizing the accuracy scores achieved from each classifier during testing.

DOI: https://doi.org/10.5281/zenodo.15099218

LOGISTIC REGRESSION	0.8293	0.8375
K-NEAREST NEIGHBORS (KNN)	0.7813	0.7771
DECISION T CLASSIFIER	0.8533	0.8614
RANDOM FOR CLASSIFIER	0.8586	0.8684
ADA BOOST CLASSIFIER	0.8213	0.8337
GRADIENT BOOST CLASSIFIER	0.8426	0.8543

Table 2: Representing Accuracy score of different classifiers

Among all the classifiers that were used, the Random Forest Classifier had the best score of 0.8586 for accuracy. After completing hyperparameter tuning, the accuracy of the Random Forest Classifier further improved to 86.66% when the parameters were adjusted as follows:

- **n_estimators**: 159
- **min_samples_split**: 5
- **min_samples_leaf**: 1
- **max_depth**: 70
- **bootstrap**: False

Below is the **Confusion Matrix** of the Random Forest Classifier before and after tuning. The confusion matrix is a direct way to observe what the model predicts for outputs. It can help explain findings to stakeholders who may not be familiar with more complicated metrics. Additionally, patterns of misclassification can be identified through the confusion matrix.

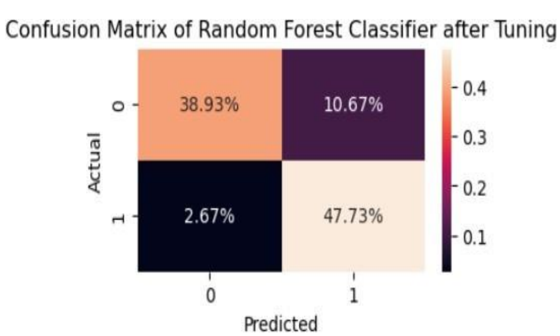


Figure 2: Confusion Matrix of Random Forest Classifier before Hyperparameter Tuning

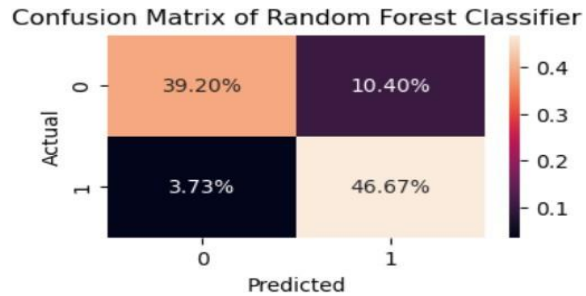
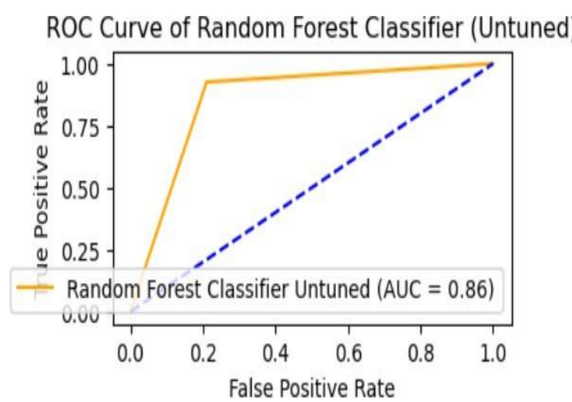


Figure 3: Confusion Matrix of Random Forest Classifier after Hyperparameter Tuning

Below is the **ROC Curve** for the Random Forest classifier before and after hyperparameter tuning. The **ROC AUC curve** denotes how well the model might classify a person who could be in need of a visit to mental health facilities against those who are probably not in need, with the positive class versus the negative class.

The higher the **AUC**, the better the model is at correctly classifying a person into the group requiring treatment.

Attention. This is particularly very crucial for the mental health sector because such early diagnosis boosts the chances of being



cured.

Figure 4: ROC Curve of Random Forest Classifier before Tuning

DOI: https://doi.org/10.5281/zenodo.15099218

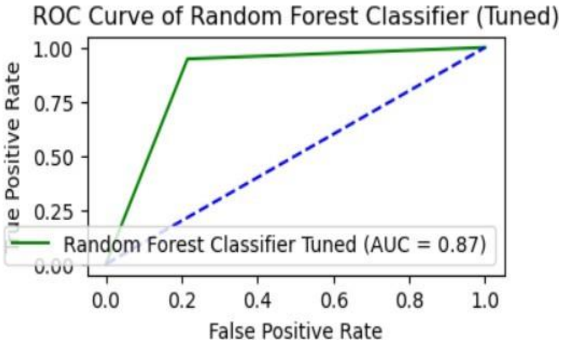


Figure 5: ROC Curve of Random Forest Classifier after Tuning

In summary, these results indicate the strength of ensemble methods, especially the AdaBoost algorithm, for mental health analytics. Moreover, the analysis here calls attention not only to appropriate model selection but also hyperparameter tuning to enhance mental health outcome prediction models' performance.

Conclusion and Future Work

Furthermore, a better performance was achieved by the Random Forest Classifier, which attained an accuracy of 86.66% after hyperparameter tuning. This improvement in accuracy must be taken into account in order to enhance the prediction capability of the model. Above, we have predominantly discussed the ways in which machine learning algorithms are being applied to mental health analyses, and such capabilities could be meaningfully relevant to mental health research and practice for both researchers and practitioners.

Some avenues for future work might include expanding this dataset to better represent the population, making it more generalizable and robust for a wider audience. Additionally, it is likely that deep learning and natural language processing can offer further insights into the dynamics of mental health. Mental health professionals should be able to make real-time predictions and interventions with tools or applications that are user-friendly. The potential for such advanced analytics in managing mental health in the future is promising, especially for early detection and intervention across diverse populations, leading to better mental health outcomes.

References

[1] Wang, W., Kiik, M., Peek, N., Curcin, V., Marshall, I. J., Rudd, A. G., ... & Bray, B. (2020). A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PloS One*, 15(6), e0234722.

[2] Shin, D., Lee, K. J., Adeluwa, T., & Hur, J. (2020). Machine learning-based predictive modeling of postpartum depression. *Journal of Clinical Medicine*, 9(9), 2899.

[3] Rosenfeld, A., Benrimoh, D., Armstrong, C., Mirchi, N., Langlois-Therrien, T., Rollins, C., ... & Yaniv-Rosenfeld, A. (2019). Big data analytics and AI in mental healthcare. *arXiv Preprint* arXiv:1903.12071.

[4] Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, 128, 102289.

[5] Akuamoah-Boateng, K., Banguti, P., Starling, D., Mvukiyehe, J. P., Moses, B., Tuyishime, E., ... & Bethea, A. (2020). 1383: Effect of implementing a fundamental critical care support course in emerging critical care systems. *Critical Care Medicine*, 48(1), 668.

[6] Radwan, A., Amarneh, M., Alawneh, H., Ashqar, H. I., AlSobeh, A., & Magableh, A. A. A. R. (2024). Predictive analytics in mental health leveraging LLM embeddings and machine learning models for social media analysis. *International Journal of*

Web Services Research (IJWSR), 21(1), 1-22.
[7] Bernstorff, M., Hansen, L., Enevoldsen, K., Damgaard, J., Hæstrup, F., Perfalk, E., ... & Østergaard, S. D. (2024). Development and validation of a machine learning model for prediction of type 2 diabetes in patients with mental illness. *Acta Psychiatrica Scandinavica*.
[8] Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In *Healthcare* (Vol. 11, No. 3, p. 285). MDPI.
[9] Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, 32(4), 966-978.
[10] Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). Machine learning for multimodal mental health detection: A systematic review of passive sensing approaches. *Sensors*, 24(2), 348.