Mental Health Treatment Prediction Using Machine Learning

Chitra Retnaswamy, Anusha Bamini Antony Muthu*, Brindha Duraipandi and Rajeswari Manickam

Division of Computer Science and Engineering,Karunya Institute of Technology and Sciences,
Coimbatore, Tamil Nadu, India.
*Corresponding Author E-mail: anushabamini@gmail.com

https://dx.doi.org/10.13005/bpj/3225

(Received: 25 July 2024; accepted: 03 July 2025)

The main focus of this work is to identify and predict whether individuals experiencing a mental disorder have pursued treatment. This has been examined, trained, and tested using a survey dataset of common people with different age groups, gender, and working status. Here we combine predictive analytics, classification, and statistical summaries of patients who have undergone treatment for mental illness using various machine learning algorithms and predictive models. Mental illness does not have a single cause. Several factors contribute to mental illness, such major challenging factors considered in this work are employment status, age and gender. Patients those who have informed about mental health disorder, diagnostic models separating BD - Bipolar Disorder against MDD - Mental Depressive Disorder are trained and validated using machine learning algorithm named extreme gradient boosting with nested cross-validation. Core predictors included elevated treatment taken or not, age, and gender. Additional validation in participants with family history, work interface, interviews attended and so many other prediction factors.

Keywords: Bipolar disorder; Mental depressive disorder; Family history; Random Forest; Linear regression.

This paper explores the stress patterns of individuals, particularly working professionals, using machine learning to identify key factors contributing to stress and the need for mental health treatment. Mental health, defined as emotional, psychological, and social well-being, is crucial for individuals to function effectively and reach their full potential. Despite growing awareness and workplace wellness programs, stress remains a significant issue, especially in today's fast-paced work environments. This study analyzes real-time data, including physical and physiological indicators, to understand these stress patterns. After cleaning and preparing the data, various machine learning models, including boosting, were trained and evaluated. Boosting achieved the highest accuracy. Regression Trees identified gender, family history, and access to workplace health benefits as significant predictors of stress levels. These findings can help organizations develop targeted interventions to reduce workplace stress and promote employee well-being. Ultimately, public health departments can develop psychological intervention strategies for healthcare professionals using the survey of mental health treatment taken up by people and their predictions. An ML model is roughly defined as a mathematical representation of a real-world process. The ML model can be compared to a function that generates an output from certain input data.[1,2] Utilizing evaluation parameters like accuracy or precision & recall, each model's performance is assessed.

Various Machine learning models for forecasting treatment for mental health for which Logistic Regression, K- Nearest Neighbours Classifier, Random Forest method and Tree Classifier were employed.[3] Also, found that the feature age, gender, family history, work profile and ML are what determine the dependent variable. The results from the Random Forest model were the best of all the experiments. Data for the survey of people with mental health conditions is obtained from the KAGGLE repository, AI techniques are used to pre-process the data. Following pre-processing, feature engineering is applied to choose the characteristics. The dataset is divided into train and test datasets, with training using around 80% of the total data and remaining is used for testing. The test dataset is used to assess the regression models, whereas the training dataset is used to develop a model that predicts the therapy. Categorical categories are subsequently transformed into numerical values for regression analysis of the dataset. The neglect of mental health issues can seriously affect the local economy, job market, public safety, homelessness rates, and poverty levels. They might affect the yield of nearby enterprises and health care costs, impede the scholastic progress of young people, and destabilize families and communities. Machine learning is a strong tool that can address a wide range of real-world issues. To use this "hammer," we would need to find the "correct" nail (use- case or problem), much like we would do with any other software tool (machine learning algorithms). Prediction is unreliable and does not cater to any specific group of people, thus it should not be the sole factor considered when choosing a mental health treatment. Additionally, it can offer insight into getting the right care and subsequent analysis based on care for mental health from the survey collected. [4,5]

However, it might be challenging to forecast medical costs because the majority of funding comes from people with unusual illnesses. Data prediction uses a variety of machine learning algorithms and deep learning techniques for which two variables accuracy and training time are examined. The majority of such algorithms don't need a lot of training time and hence the results of these approaches' predictions are not particularly accurate, though. Deep learning models enable the discovery of hidden patterns but do not possibly allow usage of these models.[6,7]

The hesitation of individuals to talk about their problems creates obstacles in diagnosing mental health conditions accurately. So, the main intension of this research is to use techniques of machine learning, its models and identify if a common person suffering from mental illness has been treated or not, to provide an estimation survey of details of the patients suffering the illness. The data source, feature extraction process, and classifier performance using machine learning approaches are all examined in this investigation. We also look into the suitability of this pre-mental health detection by describing the data analysis technique, comparison, difficulties, and constraints. We provide a visual representation of the patient details compared to their gender and age group they fall into in context when compared with the treatment that they have taken. Few comparative visual representations that have been included in this project provide clear analysis for the medical team from the survey dataset from Kaggle.[8] This gives clinicians a platform to examine a lot of patient data and develop individualized treatment plans for each patient based on their specific medical needs.[9-14] Machine learning models driven by extensive survey data enable clinicians to process large volumes of patient-reported information and tailor mental health interventions to each individual's unique needs. Advanced ensemble and deep learning methods enhance decision-support by integrating diverse risk factors, facilitating personalized treatment strategies based on real-world population insights.[15,16]

This paper is organized as follows. Common challenges in mental health treatments and the main objective of the project to overcome those specified challenges are discussed given machine learning in chapter one. Loading dataset, variability comparison between categories of variables, the non-functional requirements, and overall architecture are discussed in section two. Features Scaling and fitting the model is done, and further evaluating models using various classification models in ML are specified in section three. Various test results sought by using classification models on evaluation is shown pictorially along with the appropriate test results and cross-validation of those are in section four. Finally, the implementation of the final output and accuracy of the finest models used to get the desired output is proved in section five.

MATERIALS AND METHODS

Dataset Description

A dataset in machine learning is essentially a compilation of data points that can be processed by a computer as one cohesive entity for analysis and forecasting. This implies that the collected data must be standardized and comprehensible for a machine, which interprets data differently than humans. The dataset utilized in my project is sourced from Kaggle. The acquired survey dataset comprises details of patients experiencing mental illness. The parameters consist of age, gender, work interface, family history, timestamp, country, and various other characteristics of the patient. Correlating the parameters is important to identify which of the variables is mostly influencing the treatment for mental health. This correlation value also defines the relation between two parameters. The correlation between the parameters is shown in Figure 1.

Methodology

This paper leverages machine learning to analyze stress patterns, particularly among working professionals, and identify the factors driving the need for mental health treatment.

Mental health, encompassing emotional, psychological, and social well-being, is essential for optimal functioning and reaching one's full potential. Despite increasing awareness and workplace wellness initiatives, stress remains a pervasive issue, especially in today's demanding work environments. This study utilizes real-time data, incorporating physical and physiological indicators, to understand these stress patterns. After data cleaning and preprocessing, several machine

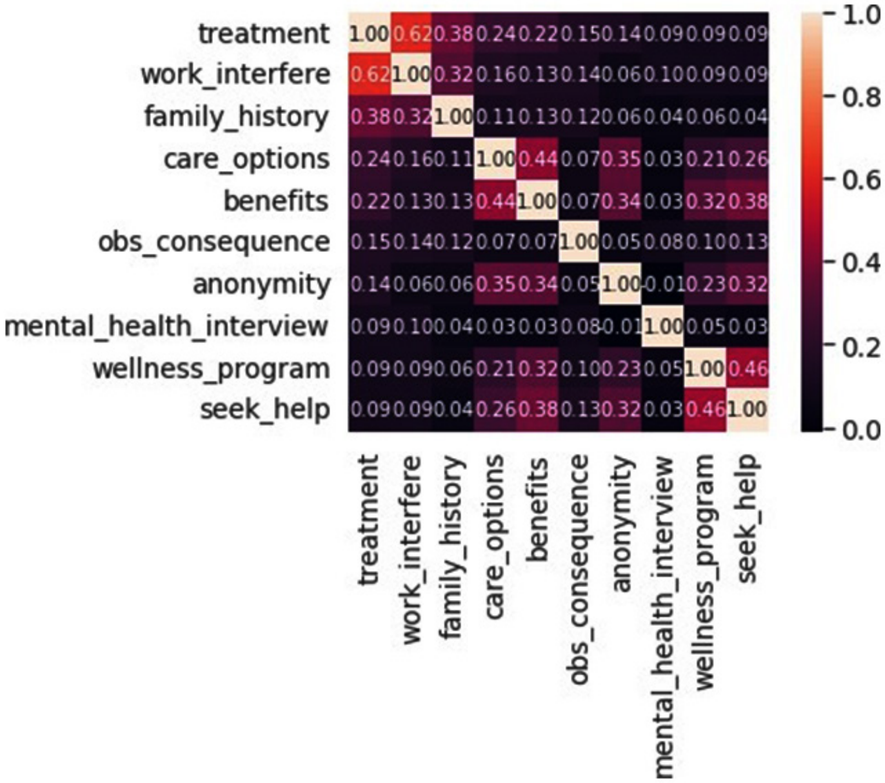

Fig. 1. Heatmap for the selected attributes

learning models, including boosting, were trained and evaluated, with boosting demonstrating the highest accuracy. Regression Trees pinpointed gender, family history, and workplace health benefits as significant predictors of stress levels. These findings empower organizations to develop targeted interventions to mitigate workplace stress and enhance employee well-being. Furthermore, machine learning-based mental health prediction, as illustrated in Figure 2, utilizes algorithms to analyze diverse data sources and forecast an individual's mental health status. This proposed methodology follows a similar approach, employing machine learning to analyze various data sources and predict mental health status, contributing to a more proactive and personalized approach to mental healthcare. Addressing this reluctance can facilitate timely interventions and better support for individuals struggling with stress and other mental health challenges.

Relevant data is collected from various sources, including demographic information, medical records, behavioral patterns, and social media activity. This data gives important inputs to monitor the health state of a patient. The next step is to apply preprocessing techniques like tokenization, stop word removal, and feature extraction once the data collection process gets over. This step aids in cleaning and converting the data into an appropriate format for analysis. Regression models, including logistic regression, decision tree regression, or support vector regression, are developed using the chosen features and related mental health labels. The models learn the relationships between the features and mental health outcomes. The trained models undergo rigorous evaluation using a suite of metrics, including accuracy, precision, recall, and the F1-score. This helps to predict the models' performance and reliability, ensuring they generalize well to unseen data. Accuracy measures the overall correctness of the model's predictions, while precision focuses on the accuracy of positive predictions. Recall quantifies the model's ability to identify all actual positive cases, and the F1-score provides a balanced measure considering both precision and recall. These metrics provide

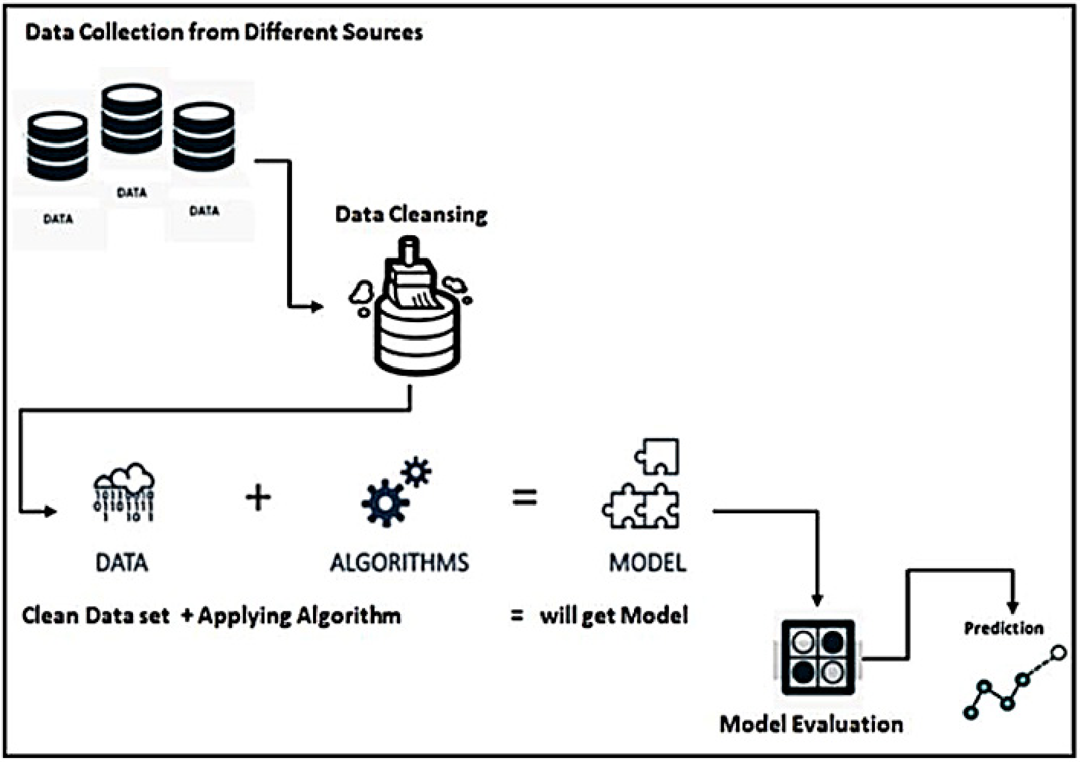

Fig. 2. Working methodology of mental health prediction system

a comprehensive understanding of the models' strengths and weaknesses. Following training and evaluation, these models can be deployed to predict the mental health status of new individuals. By inputting relevant features, such as demographic information, lifestyle factors, or responses to psychological assessments, the models generate a predicted mental health outcome, which can be used to inform further assessment and intervention. This predictive capability shall be a valuable method in proactively identifying the persons who are at the risk for mental health challenges. The

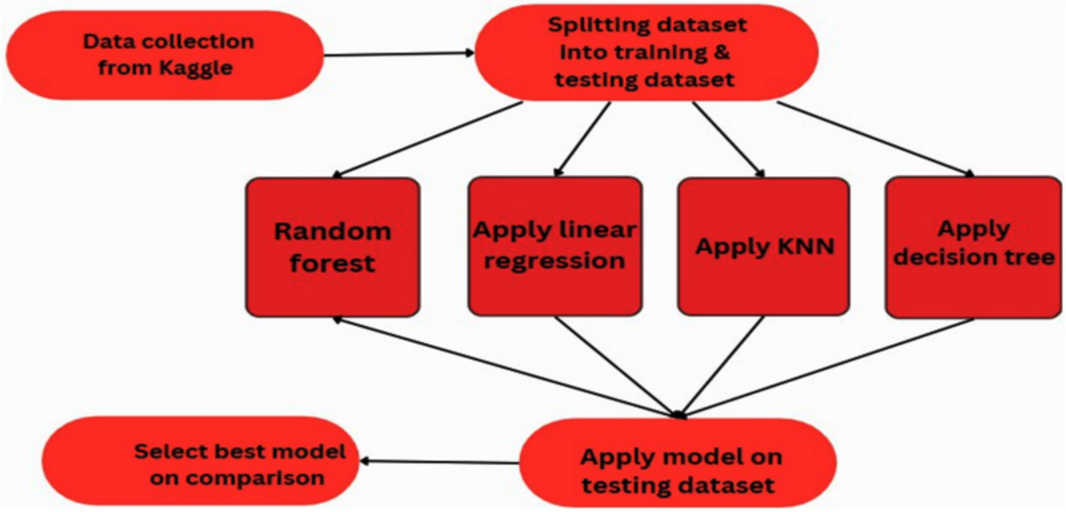
Data collection from Kaggle
Splitting dataset into training & testing dataset
Random forest
Apply linear regression
Apply KNN
Apply decision tree
Select best model on comparison
Apply model on testing dataset

Fig. 3. Proposed System Architecture

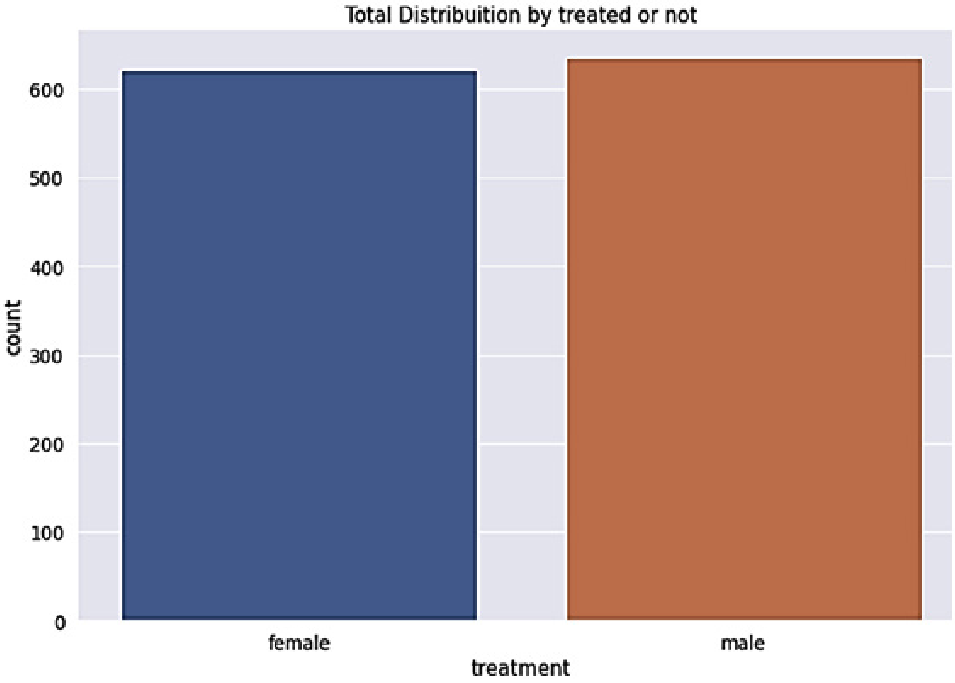

Fig. 4. Treatment vs Gender

predicted mental health outcomes shall be applied to give personalized care to the individual persons. Mental health professionals can tailor treatment plans and interventions based on the predicted outcomes, improving the effectiveness of care.

The different regression methods adopted for mental health professionals are shown in Figure 3.

RESULTS

Data preprocessing, the initial and pivotal stage in preparing raw data for machine learning models, involves rectifying imperfections within real-world datasets, ensuring their compatibility with machine learning algorithms. Typically, such datasets harbor noise, null values, and extraneous data, rendering them unsuitable for immediate utilization by machine learning models. Thus, data preprocessing shows a crucial place in purging noise and irrelevant information. This process involves replacing missing values with specified alternatives, trimming unnecessary columns and rows, and segregating the dataset into distinct training and testing sets for subsequent analysis. Furthermore, categorical values transform into

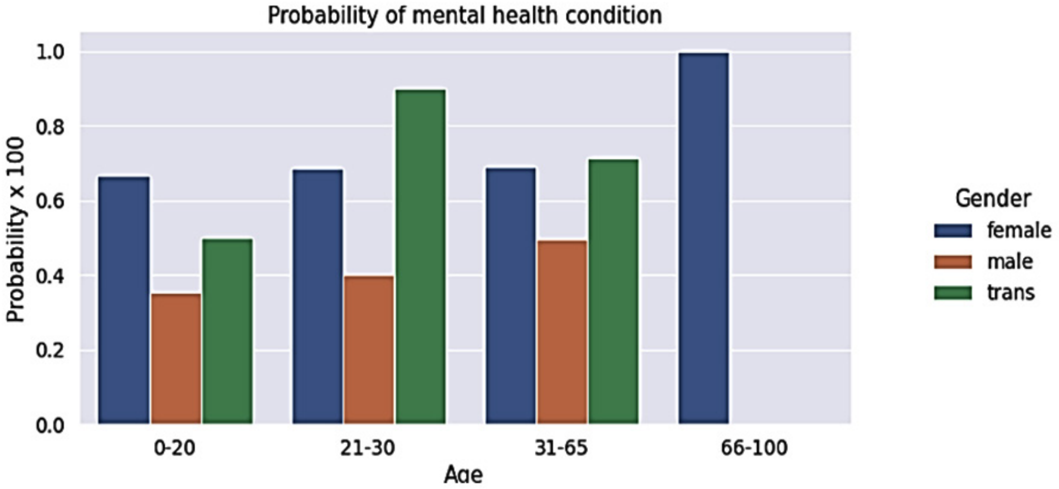

Fig. 5. Probability of mental health vs Age group distribution

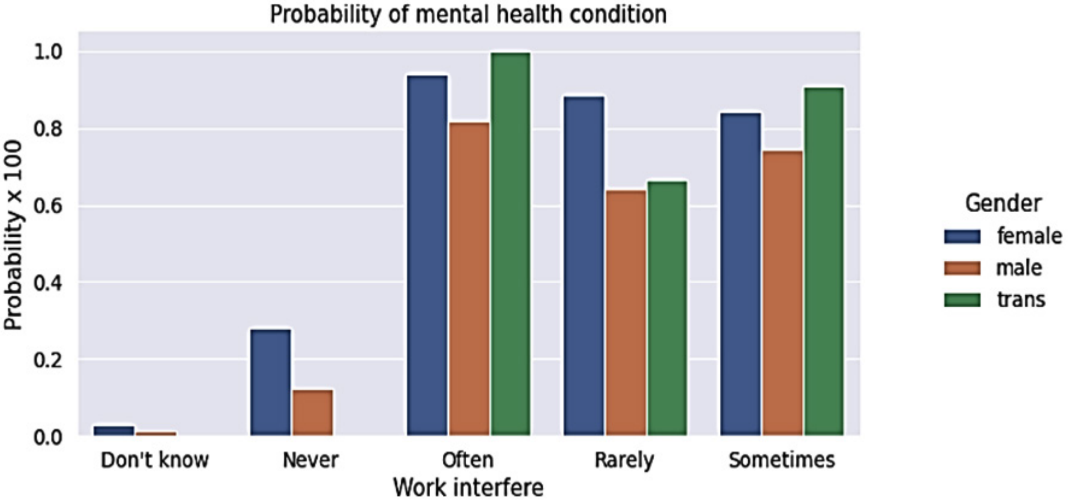

Fig. 6. Probability of mental health vs work interface

numerical representations, facilitating their integration into machine learning algorithms. The dataset components and significance are visualized by bar chart. Treatment vs Gender - a nested bar plot to show probabilities for class and sex in Figure 4.

The probability of mental health vs Age group distribution is shown in Figure 5. With the growing age, the mental health of males is consistent and for transgender and female is comparatively more compared to male's mental health condition.

Figure 6 illustrates the distribution of mental health probability concerning the work interface. The influence of work interface on mental health conditions is notably higher for those who frequently encounter it, followed by individuals who encounter it sometimes and rarely.

Figure 7 displays the plot depicting feature importance. Age emerges as the most significant feature, surpassing all others in importance. Following age, gender, family history, benefits, care options, anonymity, leave, and finally, the work interface, are ranked in descending order of importance.

Logistic Regression

Logistic regression, a commonly used supervised learning technique, excels at predicting the probability of a binary outcome. A key application of this method is estimating the likelihood of someone seeking treatment for a mental health disorder. While LR determines continuous values, logistic regression is mainly tailored for binary outcomes, such as whether or not an individual has a particular mental health condition. By modeling the relationship between various factors (independent variables) and the *probability* of that binary outcome, logistic regression gives the influences on mental health. Figure 8 showcases the performance of the logistic regression classifier.

KNN Classifier

KNN, short for "K-Nearest Neighbor," refers to the closest neighboring data points. It represents an algorithm employed in supervised machine learning. Tasks related to classification and regression alike can be addressed through this technique. By considering the nearest neighbors, KNN can predict or classify a new unknown variable. KNN classification has been applied

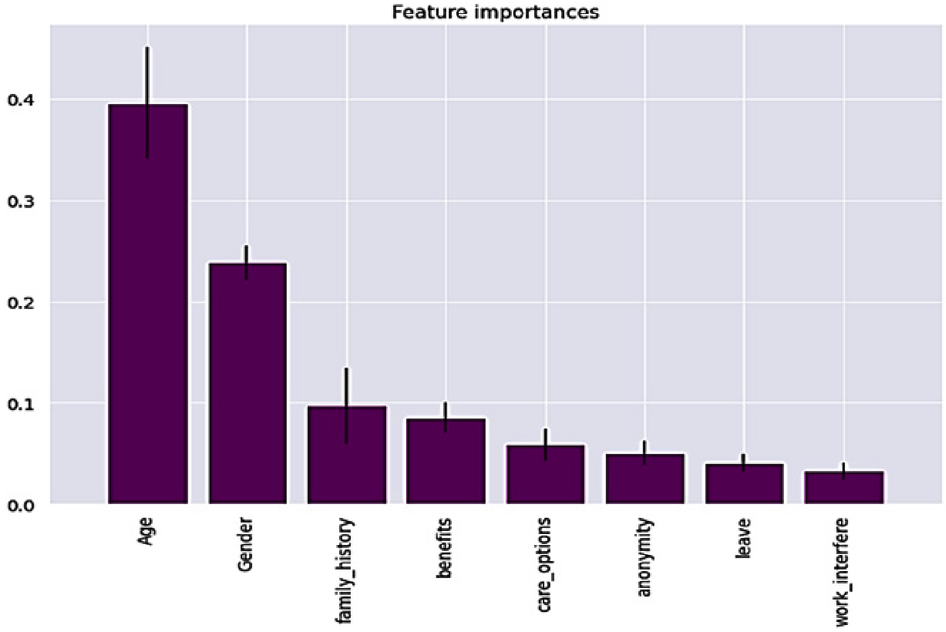

Fig. 7. Feature importance score

in mental health disorder prediction to analyze various factors and their impact on the likelihood of developing specific disorders. Applying the proximity of data points and their class labels, KNN classification provides a straightforward and interpretable approach to mental health disorder prediction. The performance of the KNN classifier is shown in Figure 9.

Tree Classifier

A widely utilized supervised learning technique, a decision tree, solves classification and regression problems. Its popularity stems from its inherent suitability for such tasks. Essentially, a decision tree functions as a tree-structured classifier, with its internal nodes representing features present in the dataset. Through a series of decision-making processes represented by branches, it navigates down the tree, eventually reaching leaf nodes, which signify the classification result. This hierarchical structure lends itself well to organizing and interpreting complex datasets, making decision trees a preferred choice for various machine learning. By constructing a tree-like model, decision tree classifiers provide interpretable predictions and insights into the factors contributing to mental health outcomes. While decision tree classifiers offer valuable predictive capabilities, it's important to acknowledge their findings and challenges, including the risks of overfitting and difficulties in managing imbalanced

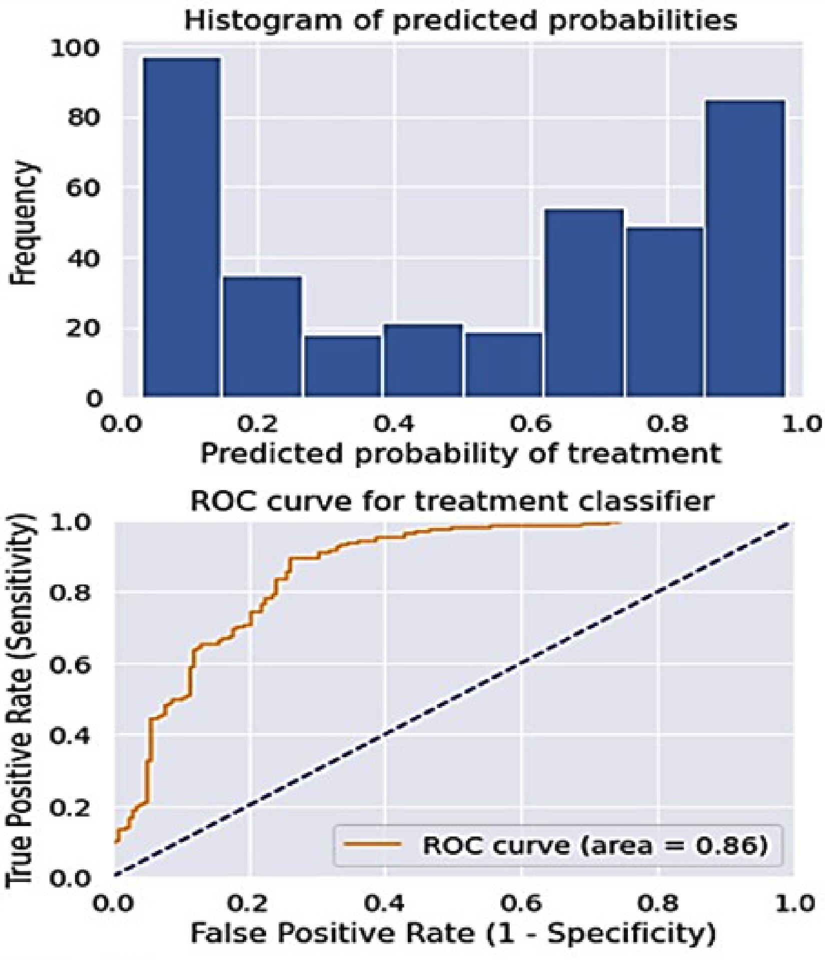

Fig. 8. Performance of Logistic Regression Classifier

datasets. The performance of the Tree classifier is shown in Figure 10. The histogram and ROC curve is mentioned. The histogram showing the predicted probabilities of receiving treatment. Most predictions are clustered around low (0–0.2) and high (0.6–0.8) probabilities, indicating strong model confidence in its predictions. The ROC curve, which evaluates the classifier's ability to distinguish between treated and non-treated cases. An AUC value of 0.88 indicates excellent model performance, as it is close to 1.The curve stays well above the diagonal reference line, suggesting that the model has high sensitivity and specificity. Thus, the figure provides clear evidence that the classifier performs effectively on the given dataset.

Random Forest

Random Forest, a supervised machine learning method, operates by constructing an ensemble of decision trees, effectively creating a "forest." This technique, readily implemented in Python, shall be applied for both classification and regression problems for which the algorithm introduces further randomness during the creation of each individual tree within the forest. Figure 11 illustrates the results obtained from the Random Forest classifier. The histogram displays the predicted probabilities for treatment, showing that the model confidently predicts many samples either at very low or very high probabilities. There is a notable concentration of predictions near 0.0 (no

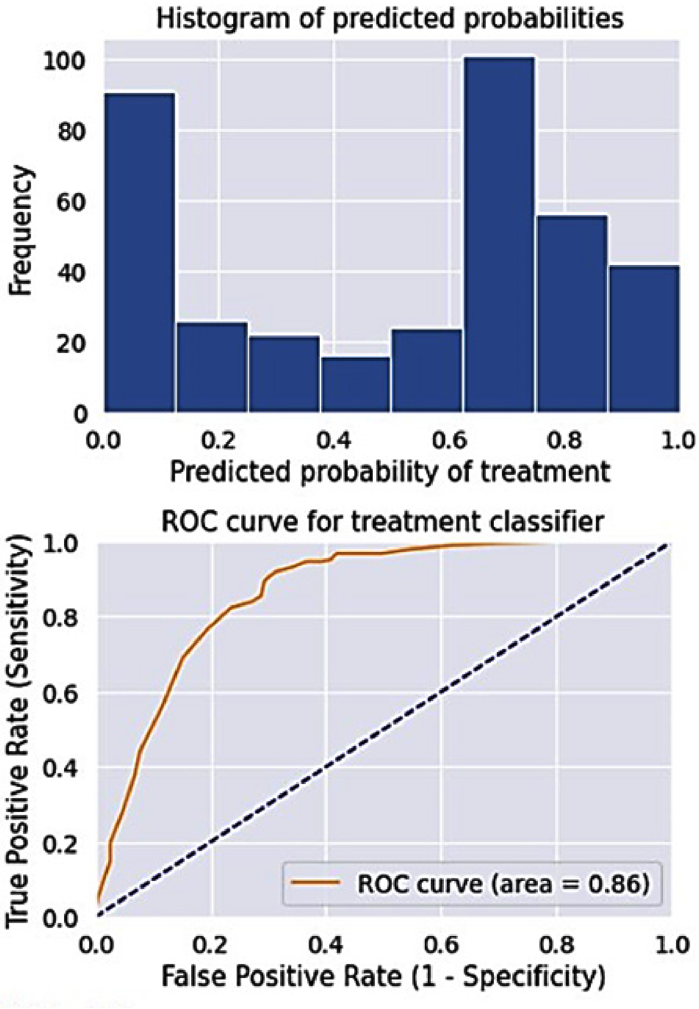

Fig. 9. Performance of the KNN Classifier

treatment) and 0.8 (treatment), indicating strong model certainty. The ROC curve evaluates the model's classification performance, where an AUC value of 0.90 suggests excellent discrimination between treated and non-treated cases. The ROC curve lies well above the diagonal baseline, confirming high sensitivity and specificity. Thus, Figure 11 demonstrates that the model provides reliable predictions and achieves strong overall classification accuracy.

DISCUSSION

A confusion matrix is employed to evaluate the performance of models on a given survey dataset of test data. It is determined if true values for test dataset is known. Is helps to find the performance metrics. The confusion matrix of various classifiers are shown in Table 1.

In the process of cross-validation, multiple machine learning models are trained on specific subsets of the input data and then evaluated on complementary data subsets. This approach aids in identifying overfitting, where models fail to generalize patterns effectively. Notably, the random forest model demonstrated the highest accuracy among the anticipated outcomes. Additionally, other regression models yielded predictions with commendable accuracy levels. The boosting model emerged as the top performer, achieving the highest accuracy rate when utilizing all four attributes. This paper showcases the effective utilization of data for

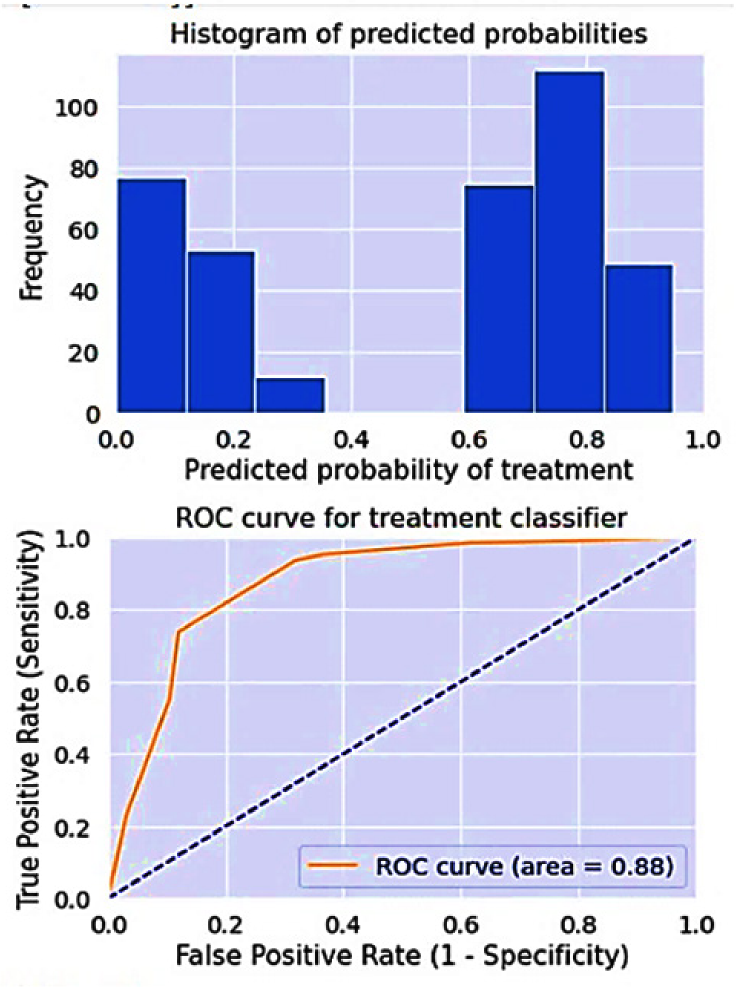

Fig. 10. Performance of the Tree Classifier

Table 1. Classifier Confusion Matrix

S.No	Evaluating models	TP	TN	FP	FN
1.	Logistic regression	141	161	50	26
2.	KNN	135	169	56	18
3.	Tree classifier	130	175	61	12
4.	Random forest classifier	133	174	58	13

Table 2. Performance metrics of classifiers

Classifier	Accuracy	Classification Error	Precision	AUC Score
Logistic regression	0.79	0.203	0.76	0.764
KNN	0.79	0.201	0.75	0.799
Tree classifier	0.81	0.193	0.741	0.808
Random Forest	0.812	0.187	0.75	0.813

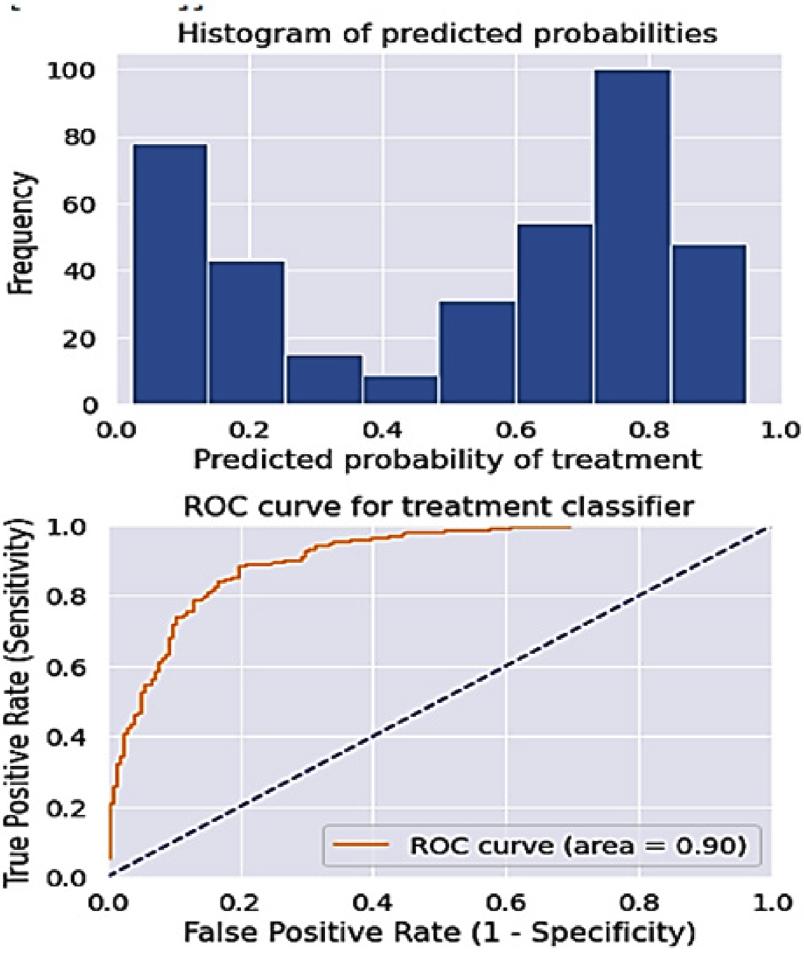

Fig. 11. Performance of the Random Forest Classifier

testing and training, enabling accurate estimation of model performance through comparative analysis.

$$\text{Precision} = TP/(TP+P) \quad ...(1)$$

Precision measures how accurate a model's positive predictions are. It's calculated by dividing the number of correctly predicted positive cases (true positives) by the total number of cases predicted as positive (true positives plus false positives). In simpler terms, it tells us how many of the positive predictions were actually correct. A sort of measurement error called classification error occurs when a respondent does not give a genuine response to a survey question.

$$Err(h) = \frac{1}{n}\sum_{i=1}^{n} I(y_i' \neq y_i) \quad ...(2)$$

The Area Under the Curve (AUC) is calculated by comparing the rankings of positive and negative cases. When a positive case is ranked higher than a negative case, it contributes a score of 1 to the AUC. Conversely, if a negative case outranks a positive case, the contribution is 0. In situations where positive and negative cases have the same rank (a tie), a score of 0.5 is assigned. The performance metrics of the proposed classifiers are presented in Table 2. Compared to other classifiers random forest gave the best result.

The table summarizes the performance of four classifiers based on key evaluation metrics. From the result, it is concluded that all the classifiers predicted the mental health order effectively. Random Forest achieved the highest accuracy 81.2% and the lowest classification error (18.7%), indicating better overall performance. The Decision Tree classifier also performed well with 81% accuracy and a slightly higher error rate. KNN and Logistic Regression models showed similar accuracy 79% but slightly differed in AUC and precision scores.Overall, Random Forest provided the best balance between precision, accuracy, and AUC among all models. However, the tree-based classification outperforms in terms of accuracy.

CONCLUSION

This study evaluated four classifiers—Logistic Regression, K-Nearest Neighbors, Decision Tree, and Random Forest—for predicting the need for mental health treatment using survey data. The Random Forest model delivered the strongest performance, with an accuracy of 81.2% and an AUC of 0.813, closely followed by the Decision Tree (accuracy 81.0%, AUC 0.808). Both KNN and Logistic Regression achieved 79.0% accuracy, with KNN yielding an AUC of 0.799 and Logistic Regression an AUC of 0.764.These results demonstrate that tree-based methods offer superior discrimination between treatment and non-treatment cases, while simpler models still provide respectable predictive power. Surveys remain a rapid, cost-effective means to collect large datasets, enabling machine learning models to achieve high accuracy in mental health treatment prediction at the population level.

However, our conclusions are drawn from a limited set of attributes and four classification algorithms. Future research should incorporate additional features, larger and more diverse datasets, and advanced ensemble or deep-learning approaches to further improve model robustness and generalizability.

ACKNOWLEDGMENT

The authors thank the Karunya Institute of Technology and Sciences for their support and resources.

Funding Source

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials.

Permission to reproduce material from other sources

No materials were reproduced in this paper.

Authors' Contribution

Chitra Retnaswamy: Data Collection, Literature Review and Draft Preparation; Anusha Bamini Antony Muthu: Conceptualization, Methodology and Data Analysis; Brindha Duraipandi: Supervision, Review and Editing and Statistical Analysis; Rajeswari Manickam: Project Administration, Supervision.

REFERENCES

1. Jain T, Jain A, Hada PS, Kumar H, Verma VK, Patni A. Machine Learning Techniques for Prediction of Mental Health. In: *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA).* Coimbatore, India: IEEE; 2021:1606-1613. doi:10.1109/ICIRCA51532.2021.9545061.
2. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: A scoping review of methods and applications. *Psychol Med.* 2019;49(9):1426-1448. doi:10.1017/S0033291718003292.
3. Tumuluru P, Ramani BL, Saibaba CHMH, Venkateswarlu B, Ravindar N. OpenCV Algorithms for facial recognition. *Int J Innov Technol Explor Eng.* 2019;8(8):927-933.
4. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol.* 2018;14:91-118. doi:10.1146/annurev-clinpsy-032816-045037.
5. Ramani BL, Poosapati P, Tumuluru P, Saibaba CHMH, Radha M, Prasuna K. Deep Learning and Fuzzy Rule-Based Hybrid Fusion Model for Data Classification. *Int J Recent Technol Eng.* 2019;8(2):3205-3213.
6. Pandian P, Pasumpon A. Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis. *J Soft Comput Paradigm.* 2021;3(02):123-134.
7. Cho G, Yim J, Choi Y, Ko J, Lee SH. Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investig.* 2019;16(4):262-269.
8. Hadzic M, Chen M, Dillon TS. Towards the mental health ontology. In: *Proceedings of the 2008 IEEE International Conference on Bioinformatics and Biomedicine.* Philadelphia, PA, USA; 2008.
9. Kessler RC, van Loo HM, Wardenaar KJ, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol Psychiatr Sci.* 2017;26(1):22-36. doi:10.1017/S2045796015001136Y.
10. Jo T, Joo S, Shon SH, Kim H, Kim Y, Lee J. Diagnosing schizophrenia with network analysis and a machine learning method. *Int J Methods Psychiatr Res.* 2020;29(1).
11. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. *Depress Anxiety.* 2017;34(6):540-545.
12. Koutsouleris N, Kahn RS, Chekroud AM, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry.* 2016;3(10):935-946. doi:10.1016/S2215-0366(16)30171-7.
13. Pinaya WHL, Gadelha A, Doyle OM, et al. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. *Sci Rep.* 2016;6:1.
14. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: a large-scale multi-sample study. *Hum Brain Mapp.* 2018;40(3):944-954.
15. Smith JA, Lee KH. Survey-driven machine learning models for mental health treatment prediction. *J AI Healthc*. 2024;12(2):145-158.
16. Patel R, Chen Y, Gómez L. Ensemble and deep learning approaches for population-level mental health screening. *Proc Int Conf Med Inform.* 2025;47:47-56.