Mental Health Prediction Using Machine Learning

Rina Kushwaha
dept. of Computer Science and Engineering
KIET Group Of Institutions
Ghaziabad, India
kushwaha123rina@gmail.com

Varsha Dubey
dept. of Computer Science and Engineering
KIET Group Of Institutions
Ghaziabad, India
dubeyvarsha122@gmail.com

Aman khan
dept. of Computer Science and Engineering
line 3: *KIET Group Of Institutions*
Ghaziabad, India
amankhan02001@gmail.com

Akash kumar Verma
dept. of Computer Science and Engineering
KIET Group Of Institutions
Ghaziabad, India
akash96218@gmail.com

***Abstract*—Experts can treat patients more effectively and enhance their quality of life when mental health issues are identified early. A person's mental health encompasses their thoughts, feelings, and social well-being, all of which impact their thoughts, feelings, and behaviors. It's vital throughout life, from child-hood to adulthood. This study examines the role that machine learning can play in forecasting mental health problems. In the proposed work the ability of five distinct machine learning techniques to recognize mental health dis-orders is implemented. These techniques include Decision Tree Classifier, K-Nearest Neighbors (K-NN) Classifier, and Logistic Regression. Each method's accuracy was evaluated based on a number of factors, including overall accuracy, precision, and recall. The work evaluates these models' ad-vantages and disadvantages in order to determine the best strategy for early mental health problem detection.**

Keywords—Depression, Anxiety, Insomnia, Post traumatic stress disorder, Machine Learning.

I. INTRODUCTION

The psychological state of a person as well as the general context of their surroundings are referred to as their mental well-being. Unbalances in brain chemistry are frequently the cause of mental diseases. The state of someone's mental health can be a clue as to how best to treat their medical conditions [1]. The mental health profiles of different groups must be monitored in order to predict any health-related anomalies. These demographics comprise high school pupils, college students, and working professionals. It is widely believed that people of all ages and backgrounds are impacted by stress and sadness [2].

It is crucial to evaluate the mental health of various demographic groups at suitable intervals in order to prevent serious disorders. In the future, medical will have to take patients' mental health characteristics into account in order to provide better therapies and encourage quicker healing [3].

Major mental health diseases, including schizophrenia, bipolar disorder, and chronic conditions [4], usually develop gradually and show early-stage diagnosable symptoms. If abnormal mental states are identified early, prompt treatment and care can frequently prevent or improve the management of many diseases. It is a difficult psychological endeavor that has not yet been fully automated to assess someone's mental state based only on their behavior or looks. Despite the existence of screening tests, their cost and time constraints make them impractical for large populations. Additionally, diagnosis-based approaches frequently deter people from getting treatment. As a result, many psychological problems remain unrecognized or untreated.

II. LITERATURE REVIEW

A person's emotional, psychological, and social functioning are all impacted by their mental health, which is a crucial aspect of total wellbeing. Interest in prediction models and data-driven strategies for early identification and intervention has grown in recent years due to heightened awareness of mental health disorders [5].

A wealth of research has examined the use of artificial intelligence (AI) and machine learning approaches to anticipate mental health diseases, providing important insights into how mental health outcomes might be predicted, controlled, and enhanced.

i. Early Detection Early Detection and Importance of Mental Health Monitoring

The Value of Mental Health Monitoring and Early Detection Mental health illnesses such schizophrenia, bipolar disorder, depression, and anxiety must be identified early in order to stop them from developing into more severe and long-lasting forms. Many mental problems present with early signs that, if detected and treated promptly, can considerably lessen the severity of the condition [6]. According to the literature,

regular mental health monitoring of a range of population groups, such as students, working professionals, and adolescents, is essential for successful early intervention [7].

ii. Predictive Models in Mental Health

The creation of predictive models for mental health has advanced significantly as a result of recent developments in AI and machine learning. Predictive research on mental health frequently uses algorithms including logistic regression, decision trees, K-Nearest Neighbors (K-NN), support vector machines (SVM), and neural networks [8]. These methods look at big data sets to find trends and connections between different elements that could influence the development of mental health conditions, like environment, lifestyle, and genetic predispositions.

iii. Challenges in Mental Health Prediction

Predictive models provide potential, but there are still a number of issues. First, based only on demographic information or outward behaviors, it is challenging to forecast mental health disorders because to the complexity of human conduct and mental states [9]. In order to diagnose mental health issues, qualified specialists typically conduct clinical evaluations and interviews, which require intricate psychological analysis that is challenging to completely automate [10].

Furthermore, when used on huge populations, current machine learning algorithms have scalability issues. Financial, time, and resource limitations some-times make screening big groups impractical [11]. In addition, diagnosis-based methods could deter involvement because people are frequently reluctant to get treatment because mental health issues are stigmatized..

iv. Potential for Early Intervention

Predictive models, however, may aid in the early detection of mental health problems if they are optimized. This would enable medical professionals to take action sooner and lessen the burden of chronic mental health disorders.

According to Kumar et al. (2020), early detection may minimize the detrimental effects of mental health issues by enabling more prompt and individualized treatments. Improvements in data accessibility, including through social media, wearable technology, and medical records, can improve these models' efficacy and accuracy (Mishra et al., 2021 Real-time data gathering approaches coupled with machine learning techniques can offer a potent tool for tackling mental health crises before they completely materialize.

The way medical professionals manage and intervene in mental health issues could be completely transformed by the capacity to forecast and monitor mental health trends across various populations [12]..

III. DESIGN

Data collection, cleaning to address inconsistencies or missing values, encoding the data into a numerical format for machine learning algorithm compatibility, calculating a covariance matrix to identify relationships between variables, fine-tuning parameters to optimize the model's performance, and evaluating multiple models, including logistic regression, decision tree, random forest, and KNN, are all part of the structured process for developing a machine learning model. The final results are obtained by evaluating the accuracy of each model and using the best model to make predictions on the test data. Fig 1 shows the complete design procedure performed in the proposed work.

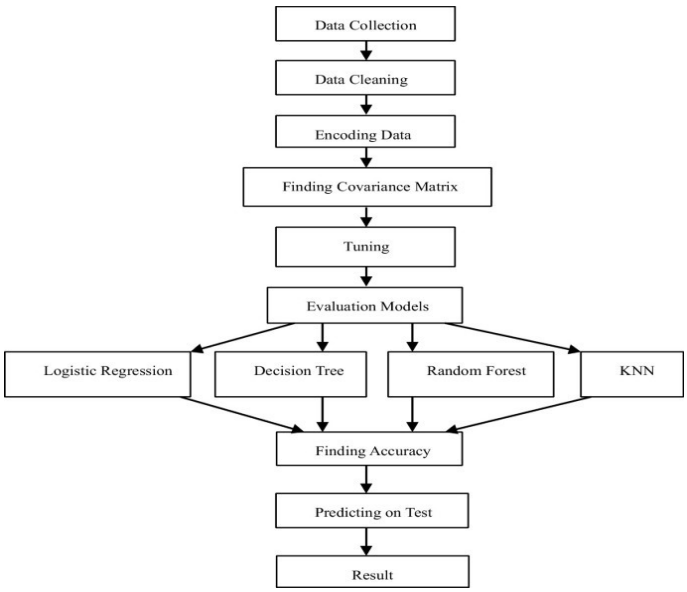

Fig-1 Block Diagram of proposed Approach

IV. MACHINE LEARNING ALGORITHMS

Machine learning algorithms must be adjusted with crucial parameters according to the tasks and data type in order to perform at their best. To improve the mod-el's accuracy and usefulness, certain parameters must be changed. In order to maximize performance, we have made sure in our research that every algorithm is adjusted with the right parameter values. We used a dataset with information on depression, anxiety, PTSD, and insomnia to test the machine learning algorithms after determining the optimal parameters. 80% of the dataset was used for training, and 20% was used for testing in order to avoid overfitting and assess the accuracy of the model. The algorithms listed below were used to determine the most precise strategy.

i. K-Nearest Neighbors (KNN)

KNN is a straightforward, non-parametric method that effectively classifies data points for mental health prediction classification tasks by determining how close they are to known data points.

ii. Logistic Regression

The logistic regression approach is frequently applied to situations involving binary categorization. It delivers results that are easy to under-stand and determines the likelihood of a mental health issue based on the dataset's attributes.

Identify applicable funding agency here. If none, delete this text box.

iii. *Decision Tree*

A decision tree is a data classification tool that employs a tree-like structure of decisions. Both categorical and numerical data may be handled and interpreted with ease, which makes it helpful for forecasting mental health disorders.

iv. *Random Forest*

A decision tree is a method for classifying data that uses a structure of judgments that resembles a tree. It is easy to manage and analyze both numerical and categorical data, which makes it useful for predicting mental health conditions.

V. Implementation

The project's initial phase was gathering data. We made our own dataset by creating survey forms for each mental health ailment (e.g., depression, anxiety, PTSD, insomnia) because there was no easily accessible dataset that catered to our needs. Fig 2 shows the probability graph of mental health condition. The survey, which asked situational and objective questions, was distributed both online and offline. We also included answers from those who were receiving treatment for mental health issues at the time.

Once there were sufficient answers, the information was transformed into numerical values that, depending on the question type, ranged from 0 to 3 or 0 to 4. Training and testing datasets are the two subsets of the gathered data.

Next, we performed data preprocessing, where missing values were filled or modified to improve data quality. After preprocessing, the data was used for feature extraction, which helps in identifying important patterns or factors. Finally, the machine learning models were applied to predict mental illnesses based on the extracted features.

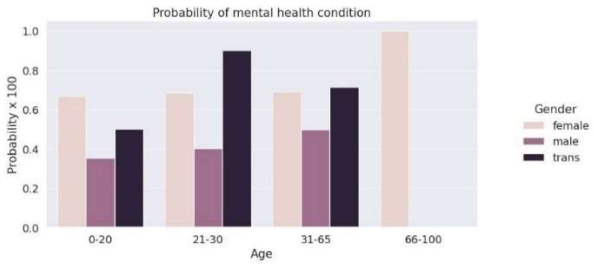

Fig. 2 probability graph of mental health condition

VI. Result

To achieve high accuracy in our model, we ensured the data was thoroughly cleaned and pre-processed for optimal performance. We utilized Python libraries such as NumPy, pandas, and matplotlib for this purpose. In our effort to get the best results, we applied several machine learning algorithms, including logistic regression, support vector machines (SVM), random forest, and k-nearest neighbors (KNN). For example, in predicting anxiety, we obtained accuracies of 97.27%, 94%, 81%, and 80%, respectively, using the above algorithms. The same approach was applied for predicting other mental health conditions, each yielding different accuracy levels. Ultimately, we selected the algorithm that provided the highest and most reliable accuracy. Additionally, we fine-tuned hyperparameters to see if further improvements in accuracy were possible.

VII. Conclusion and Future Work

Heading In this study, five machine learning models (Decision Tree, Logistic Regression, Random Forest, K-Nearest Neighbors, and Support Vector Machines) were implemented to predict mental health conditions such as anxiety, depression, PTSD, and insomnia. Logistic Regression emerged as the most accurate model, achieving the highest reliability in specific cases, such as anxiety prediction with 97.27% accuracy, due to its simplicity and effective handling of binary classifications. While Random Forest and Decision Trees provided robust performance, they did not surpass the precision of Logistic Regression, and K-Nearest Neighbors was efficient but less accurate overall. These results highlight the potential of machine learning for mental health prediction, with Logistic Regression proving to be a promising model for practical applications.

In future work, efforts can focus on addressing scalability challenges, enhancing data diversity by incorporating real-time data from wearable devices and social media, and improving model robustness against biases. Expanding the dataset and including more mental health disorders will further validate the model's applicability in real-world scenarios. Additionally, integrating interpretability tools will make these predictive systems more reliable for clinical decision-making.

VIII. References

[1] A. Rahman, K. Omar, S.A.M. Noah, M.S.N.M. Danuri, and M.A. Al-Garadi, “Application of machine learning methods in mental health detection: A systematic review,” IEEE Access, vol. 8, pp. 183952–183963, 2020.

[2] S. Gurjar, C. Patil, R. Suryawanshi, M. Adadande, A. Khore, and N. Tarapore, “Mental health prediction using machine learning,” Int. Res. J. Eng. Technol. (IRJET), vol. 9, no. 12, pp. 1117–1122, 2022.

[3] J. Chung and J. Teo, “Mental health prediction using machine learning: Taxonomy, applications, and challenges,” Appl. Comput. Intell. Soft Comput., vol. 2022, Article ID 9970363, 19 pages, 2022.

[4] U. Madububambachu, A. Ukpebor, and U. Ihezue, “Machine learning techniques to predict mental health diagnoses: A systematic literature review,” Clin. Pract. Epidemiol. Ment. Health, vol. 20, Article ID e17450179315688, 2024.

[5] S. Mutalib, N.S. Mohd Shafiee, and S. Abdul-Rahman, “Mental health prediction models using machine learning in higher education,” Turkish J. Comput. Math. Educ. (TURCOMAT), vol. 12, no. 5, pp. 1782–1792, 2021.

[6] M. Kavitha, M. Pingili, M. Spurthi, S. Fayaz, K. Kirthan, and S. Santhosh, “Classification algorithm-based mental health prediction

using data mining," Turkish J. Comput. Math. Educ. (TURCOMAT), vol. 13, no. 2, pp. 1168–1175, 2022.

[7] S.S. Shahapur, P. Chitti, S. Patil, et al., "Decoding minds: Estimation of stress level in students using machine learning," Indian J. Sci. Technol., vol. 17, no. 19, pp. 2002–2012, 2024.

[8] V. Pathak, K. Dwivedi, and M.K. Tiwari, "Mental health prediction using machine learning algorithms," TEJAS J. Technol. Hum. Sci., vol. 3, no. 3, pp. 18–26, 2024.

[9] H. Abdul Rahman, M. Kwicklis, M. Ottom, et al., "Machine learning-based prediction of mental well-being using health behavior data from university students," Bioengineering, vol. 10, no. 5, p. 575, 2023.

[10] M.R. Sumathi and B. Poorna, "Prediction of mental health problems among children using machine learning techniques," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 1, pp. 552–557, 2016.

[11] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine learning techniques to predict mental health diagnoses: A systematic literature review," Clin. Pract. Epidemiol. Ment. Health, vol. 20, Article ID e17450179315688, 2024.

[12] J.A. Renjit, A.S.M.J., S.V.D., and S.D.D., "Prediction of mental health using machine learning," J. Emerg. Technol. Innov. Res. (JETIR), vol. 9, no. 5, pp. i634–i640, 2022.

[13] M.M. Islam, S. Hassan, S. Akter, et al., "A comprehensive review of predictive analytics models for mental illness using machine learning algorithms," Healthcare Analytics, vol. 6, Article ID 100350, pp. 1–12, 2024.

[14] V.G. Kshirsagar, S. Kumar, and N. Karande, "Detection of mental illness using machine learning and deep learning," NeuroQuantology, vol. 20, no. 16, pp. 2606–2613, 2022.

[15] K. Vaishnavi, U.N. Kamath, B.A. Rao, and N.V.S. Reddy, "Predicting mental health illness using machine learning algorithms," J. Phys.: Conf. Ser., vol. 2161, Article ID 012021, pp. 1–7, 2022.

[16] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: A scoping review," Transl. Psychiatry, vol. 10, Article 116, pp. 1–26, 2020.

[17] A. Thieme, D. Belgrave, and G. Doherty, "Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems," ACM Trans. Comput.-Hum. Interact. (TOCHI), vol. 27, Article 34, pp. 1–53, 2020.

[18] B. Sahu, J. Kedia, V. Ranjan, et al., "Mental health prediction in students using data mining techniques," Open Bioinform. J., vol. 16, Article ID e187503622307140, 2023.