Psychological Wellbeingness Prediction Using Machine Learning with Hybrid Hyparameter Tuning Techniques

Puvanenthira Rajah
Dept. of Computer Engineering
Marwadi University
Rajkot, India
puvanenthirarajah.sathasivam114232@marwadiuniversity.ac.in

Masruk Habib
Dept. of Computer Engineering
Marwadi University
Rajkot, India
masrukhabib.114214@marwadiuniversity.ac.in

Ravikumar R N
Dept. of Computer Engineering
Marwadi University
Rajkot, India
rnravikumar.cse@gmail.com

Sivakumar Nadarajan
Dept. of Computer Science and Engineering
Alliance University
Bangalore, India
drsivakumar.nadarajan@gmail.com

Bhagyashree Amit Chougule
Dept. of Computer Engineering
Tatyasaheb Kore Institute of Engineering and Technology
Maharashtra, India
bhagya.magadum004@gmail.com

Kirubhakaran M
Dept. of AI & DS
Suriya Engineering College
Erode, India
kirupaa.m@gmail.com

2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit) | 979-8-3503-7971-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/GLOBALAISUMMIT62156.2024.10947775

***Abstract*—This study investigates the application of machine learning (ML) techniques in predicting Psychological Well-being outcomes, emphasizing the use of ensemble methods like AdaBoost and Random Forest for enhanced accuracy. With Psychological well-being issues affecting a significant portion of the global population, traditional assessment methods are challenged by issues of accessibility, stigma, and subjectivity. Leveraging data from various surveys, this research compares different ML algorithms, including Logistic Regression and Support Vector Machines, in predicting psychological well-being conditions. The results highlight the superior performance of AdaBoost and Random Forest, particularly when hyperparameters are finely tuned, achieving accuracies up to 0.996. The study underscores the potential of ML in improving psychological well-being interventions by offering precise, unbiased, and accessible diagnostic tools. Future research directions include the integration of multimodal data and the development of models that can suggest personalized treatment plans while ensuring data privacy. This paper advocates for the integration of advanced computational techniques with clinical insights to revolutionize psychological well-being diagnostics and care.**

Keywords— Psychological Well-being, Machine Learning (ML), Ensemble Methods, Hyperparameter Tuning, Diagnostic Tools, Personalized Treatment Plans, Computational Techniques, Mental Health Diagnostics.

I. Introduction

Psychological well-being, a crucial component of overall well-being, affects an individual's cognitive, emotional, and social functioning. According to the World Health Organization (WHO), Psychological well-being is a state of well-being where a person recognizes their abilities, manages everyday stressors, works effectively, and contributes to their community. Psychological well-being diagnosis impacts daily activities and has repercussions for families, communities, and economies. In 2022, 23.1% of adults in the US experienced mental illness, with 6% having a major mental illness. Globally, among 193 million and 246 million human beings, or more or less 28% of the populace, had depressive signs, while anxiety disorders affected 298 million to 374 million people, marking a 25% increase. In the US, young adults (18 to 25 years old) had the highest rates of mental illness (36.2%), followed by adults aged 26 to 49 (29.4%) and those 50 and over (13.9%). Traditionally, mental health assessment and intervention relied heavily on subjective evaluations by professionals, which are often labor-intensive, resource-intensive, and biased. This issue is exacerbated by the stigma surrounding psychological well-being and the difficulty in accessing services, preventing many from receiving timely help. Addressing these challenges requires more objective, accessible, and unbiased methods for Psychological Well beingness diagnosis and treatment [1].

A. Symptoms of Hypothyroid

Psychological well-being disorders encompass a wide range of conditions, each with its own unique set of symptoms. However, some common symptoms that may indicate the presence of a Psychological Well beingness concern include:

- **Mood swings:** Persistent melancholy, hopelessness, irritability, or mood changes disrupting daily tasks.
- **Anxiety:** Excessive worry, fear, or nervousness affecting daily activities or relationships.
- **Behavioral changes:** Increased anxiety, aggression, social withdrawal, or altered sleep and eating habits.
- **Difficulty concentrating:** Trouble focusing, remembering, making decisions, or completing tasks.
- **Fatigue:** Persistent tiredness or lack of energy despite adequate rest.
- **Sleep disturbances:** Difficulty falling or staying asleep, leading to daytime fatigue.
- **Social withdrawal:** Avoiding social activities and difficulty connecting with others.
- **Intrusive thoughts:** Persistent, distressing thoughts or images causing anxiety, guilt, or shame.

The primary contributions of our work are to:

- To evaluate and compare the predictive performance of various machine learning algorithms, including Logistic Regression, Support Vector Machines, AdaBoost, and Random Forest, in predicting psychological well-being outcomes based on survey data.

- To enhance the accuracy of psychological well-being predictions by employing ensemble methods like AdaBoost and Random Forest, and optimizing their hyperparameters to achieve the highest possible accuracy rates.
- To address the challenges of accessibility, stigma, and subjectivity associated with traditional psychological well-being assessments by developing and validating machine learning models that provide precise, unbiased, and accessible diagnostic tools.
- To explore future research directions that include the integration of multimodal data sources and the development of machine learning models

II. Literature Review

This research uses 26-country Global School-based Student Health Survey data to investigate ML methods for predicting and assessing adolescent suicide thoughts and actions. Compare Support Vector Regression, Multilayer Perceptron, Random Forest, Logistic Regression, Ridge Regression, K-Nearest Neighbors, and Extreme Gradient Boosting. Variable error metrics show Support Vector Regression is the most efficient model. It shows that ML can improve suicide prevention and that predictive modeling may improve mental health treatments [2]. A 2023 article uses data from the 2019 Open Sourcing Mental Illness (OSMI) in Tech Survey and the Support Vector Machine (SVM) algorithm to diagnose mental health. Radial basis function (RBF) kernels represent complicated interactions between characteristics and mental health outcomes well, followed by polynomial, sigmoid, Bessel, and ANOVA kernels. SVM has the potential for mental health analysis, with a 95% accuracy rate. Future research might improve model performance by examining parameter adjustment and kernel types [3].

Machine learning's prediction of mental health illnesses is examined in this 2023 research using the 2016 Open Sourcing Mental Illness (OSMI) Mental Health in Tech Survey. It compares NB, KNN, DT, RF, SVM, LR, NN, Stochastic Gradient Descent, and Cross Gradient Booster and finds the latter accurate at 79.8%.The study suggests model optimization and interpretability methods to better comprehend feature effect on predictions by emphasizing historical mental health illnesses [4]. A 2020 research employed data mining to predict mental health issues in working people using the Open Sourcing Mental Illness (OSMI) survey dataset. Research used machine learning methods such as Decision Tree, Random Forest, and Naïve Bayes. The best accurate algorithm was Decision Tree at 82%. This research showed machine learning's capacity to identify and raise employee mental health awareness. It advised greater research into mental disease forecasts and significant data collecting to increase accuracy and applicability [5] [6].

This 2022 study provides a private federated deep learning method for detecting Cerebellar Ataxia (CA). Data from four clinics in three Australian states is transformed using recurrence plots, melspectrograms, and Poincaré plots with motion capture sensor data. MobileNetV2, ResNet101V2, DenseNet, and VGG16 were examined. MobileNetV2 had the greatest validation accuracy of 86.69%. Diagnostic accuracy without feature engineering and data privacy are improved by this strategy, making clinical deployment possible. Additional pattern transformations and blockchain-secured federated learning for heterogeneous datasets are future goals [7]. The study explores hyperparameter optimization for machine learning models in menstrual mental health monitoring using a Reddit dataset. Sentiment analysis was conducted to understand women's emotions during premenstrual dysphoric disorder (PMDD) and the menstrual cycle. Various algorithms were tested, including Support Vector Machine, Decision Tree Classifier, Logistic Regression, Random Forest, AdaBoost Classifier, and Boosting Classifier. The Random Forest model achieved the highest performance with 94% accuracy classification [8] [9].

The paper aims to enhance cardiovascular prediction using deep learning (DL) and advanced machine learning (ML) techniques. By merging UCI and Kaggle datasets, the study improves data characteristics for more accurate predictions. Algorithms tested include Multilayer Perceptron (MLP), Artificial Neural Network, K-Nearest Neighbor, Support Vector Machine, Decision Tree, Adaptive Boosting, and Extreme Gradient Boosting. The MLP executed the best accuracy of 98.54%, accompanied closely by DT and an AdaBoost-XGBoost combination at 98.53% [8] [10]. The paper, addresses the detection of Post-Traumatic Stress Disorder (PTSD) in mental health using machine learning algorithms. It utilizes various datasets, including a significant one consisting of over 243,000 tweets from PTSD sufferers, to explore the effectiveness of different algorithms such as random forest, gradient boosting, convolutional neural networks, support vector machines, and linear discriminant analysis. The study found that RF achieved the highest prediction accuracy of 97% for PTSD [11] [6].

In 2022, two separate studies enhanced the use of machine learning (ML) in mental health diagnostics using various methodologies and datasets. One study, accepted in June 2022, introduced the CASTLE framework, which combines techniques like MOON for network embedding and SMOTE for label imbalance. It recognizes students' mental health issues by integrating statics on social existence, instructional overall performance. and physical appearance [12]. The study explores the potential of ML for forecasting outcomes related to mental health using MRI scans from Alzheimer's patients and questionnaires from mental health patients. It evaluated algorithms such as LR, DT, KNN, AB, and RF, employing pre-processing techniques like halt word elimination and lemmatization [13] [14]. The researchers used machine learning (ML) to predict public psychological well-being using medical Internet of Things (mIoT) data. They applied algorithms like K-NN, RF, DT, and LR. The study found significant correlations between lifestyle factors (low physical activity, high smoking, excessive drinking) and increased mental distress, particularly among females [15].

Published in 2022, the paper explores the shift towards noninvasive monitoring of mental health issues like depression, stress, and anxiety through objective, real-time data collection [16] [11]. The study predicts stress levels using electrodermal activity (EDA) data from wrist-worn wearables. It evaluates five ML classifiers across multiple EDA databases, finding that Support Vector Machine (SVM) performs best. Gender analysis shows significant differences in stress classification accuracy between males and females [17] [18].

III. Methodology

The preprocessing data, dataset analysis, Feature scaling, Comparison of results and performance assessment are all steps in the process. This methodology ensures that the algorithm will be used in a methodical manner to predict the progression of Mental health, allowing for a comprehensive and informed analytical process.

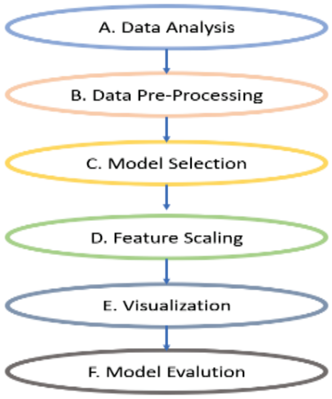

Fig.1. Work Flow

A. Data Analysis

We have introduced a new column named "class," calculated as the sum of parameters including 'Gender', 'family history', 'benefits', 'care options', 'anonymity', 'leave', 'work interfere', and 'treatment'. Individuals with a sum below 9 are classified as not affected by psychological health issues, while those with a sum of 9 or above are classified as affected. The 'age' column was excluded from the analysis due to its wide range and limited data records. Values in the 'class' column below 9 are assigned a value of 0, while values at or above 9 are assigned a value of 1 [19].

TABLE I. FEATURE DESCRIPTION

Feature Name	Explanation
Gender	The gender identity of the participant
Self Employed	Declares whether or not the person works for themselves
Family History	Whether the participant has a family history of mental health problems is shown by this.
Treatment	Indicates whether the participant is currently undergoing mental health treatment.
Work Interfere	Indicates how much work interference the participant perceives due to mental health issues.
Number Employees	The quantity of workers at the participant's place of employment.
Remote Work	Indicates whether the participant's job allows remote work.
Tech Company	Indicates whether the participant's employer is in the tech industry
Care Options	Indicates the availability of care options for mental health issues at the participant's workplace
Wellness Program	Indicates whether the participant's workplace offers a wellness program
Seek Help	Indicates if the individual is at ease asking for help at work for mental health difficulties.
Anonymity	Indicates whether the participant believes seeking help for mental health issues would be anonymous
Leave	Indicates the ease of taking leave for mental health issues at the participant's workplace
Mental Health Consequence	The perceived consequence of disclosing a mental health issue at work
Phys Health Consequence	The perceived consequence of disclosing a physical health issue at work
Coworkers	The participant's assessment of their coworkers' level of support for mental health concerns
Supervisor	The participant's assessment of their supervisor's level of support for mental health concerns
Mental Health Interview	This indicates if the interviewee has ever been questioned about mental health concerns during a job interview.
Phys Health Interview	This indicates if the interviewee has ever been questioned during a job interview about their physical health.
Mental vs Physical	The participant's opinion on whether mental health and physical health are given equal weight at work.
Obs Consequence	This shows if the participant believes there would be negative consequences if they disclosed a mental health problem at work.

B. Data Pre-Processing

We enhanced our dataset through various data cleaning and preprocessing methods. Irrelevant features were removed, retaining only relevant variables. Gender discrepancies were resolved by normalizing entries into distinct categories. Missing age values were substituted with the mean age, and age was categorized into intervals for detailed trend analysis.

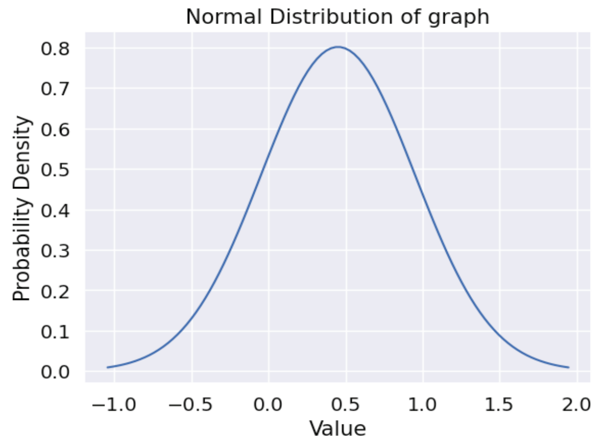

Fig.1.2. Normal distribution graph

C. Model Traning

The models listed below are put to the test:

- **SVM:** This sophisticated supervised learning model for regression and classification finds the ideal hyperplane to optimize class margins and provide solid classification bounds.
- **Random Forest**: Combining several decision trees improves accuracy, stability, and overfitting for regression and classification problems in this ensemble learning technique.
- **Gradient Boosting:** Gradient boosting, ideal for regression and classification, incrementally builds models to correct previous errors, refining weak models into a single strong predictor.
- **LightGBM:** Modern gradient boosting frameworks are fast and efficient, particularly with big data sets. Tree-based learning techniques for distributed and fast training make it suited for big and complicated datasets.
- **LR:** Logistic Regression is a prevalent statistical technique for modeling binary outcomes. By applying

a logistic function, it estimates probabilities that are then used to classify data into binary categories, making it ideal for binary classification problems.

- **eXtreme Gradient Boosting:** Scalable and performant, XGBoost optimizes gradient boosting. It excels at handling different data kinds and sizes in machine learning contests, demonstrating its accuracy and efficiency.
- **Adaptive Boosting:** Notable boosting algorithm AdaBoost helps poor learners become strong. It automatically modifies the weights of erroneously categorized examples to concentrate future models on difficult situations, improving predicting performance.

D. Feature Engineering

In feature engineering, we encoded data, handled missing values, scaled features, and created new characteristics. Imputation eliminated missing values; scaling normalized data distribution. We explored categorical variable interactions using a covariance matrix. We separated the data and created additional model training variables to aid feature selection for key components. People under 30 were classified as 0, and those 30 and over as 1. This allowed us to analyze age-related variances for more rational and unambiguous findings.

E. Hyper-parameter Tuning

Selecting a range for hyperparameters (n_estimators: [50, 100, 200] and learning_rate: [0.01 to 1.0]) enables a thorough search for ideal settings. For instance, the optimal setup is {{'learning_rate': 0.606850157946487, 'n_estimators': 100}}, balancing model quantity and correction effect. An AdaBoost Classifier with 0.996031746031746 accuracy after tuning shows how careful tweaking improves model accuracy. This shows how hyperparameter adjustment makes AdaBoost a powerful predictive modeling tool with high accuracy.

F. Visualization

The comprehension of algorithms and their composite characteristics is significantly enhanced through the visualization of datasets and the analysis of their outcomes.

1) Correlation Matrix:

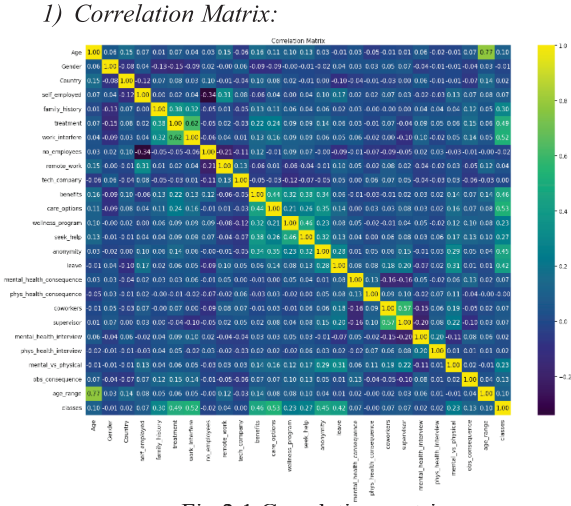

Fig.2.1 Correlation matrix

2) Distribution of age:

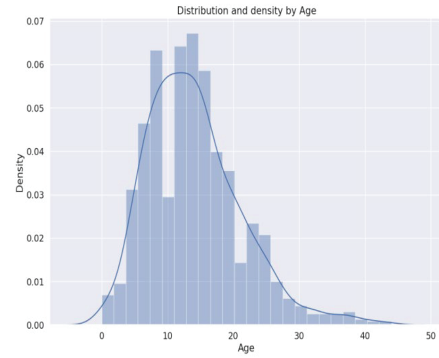

Fig.2.2. Age distribution of across the dataset

3) Distribution of male and female:

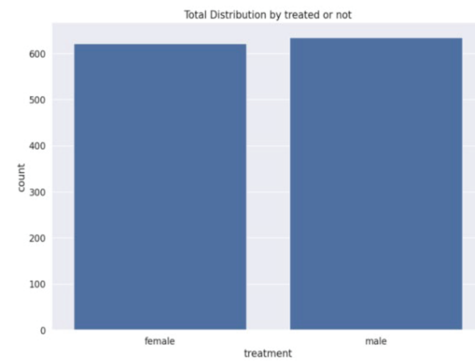

Fig.2.3 Distribution of male and female

4) Features :

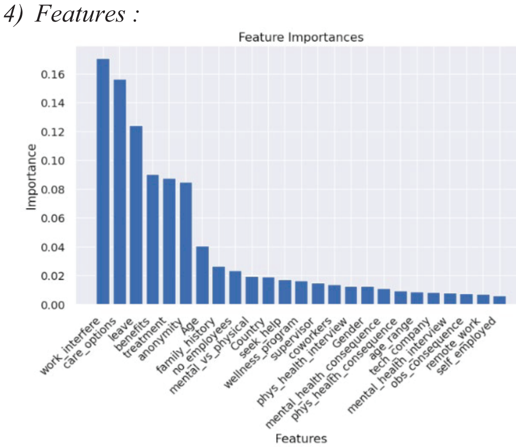

Fig.2.4. Features

5) Probability of mental health condition based on care option:

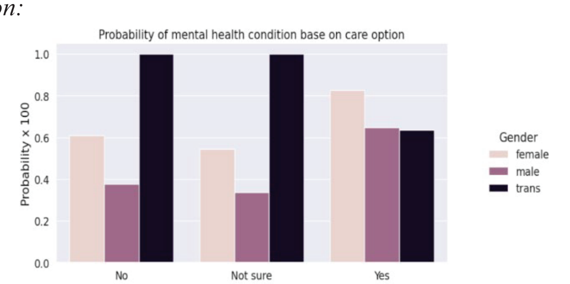

Fig.2.5 Mental health condition based on care option

6) Probability of mental health condition based on family history

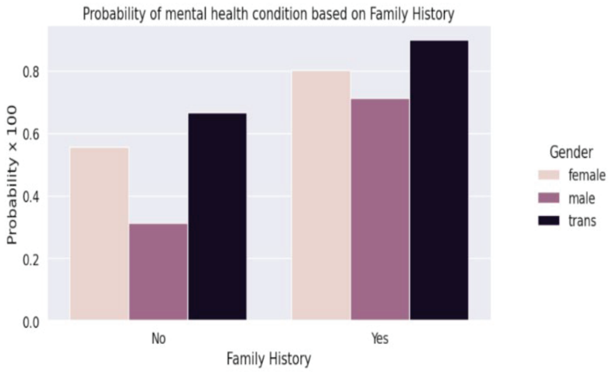

Fig.2.5. Mental health condition based on family

7) Probability of mental illness that interferes with job

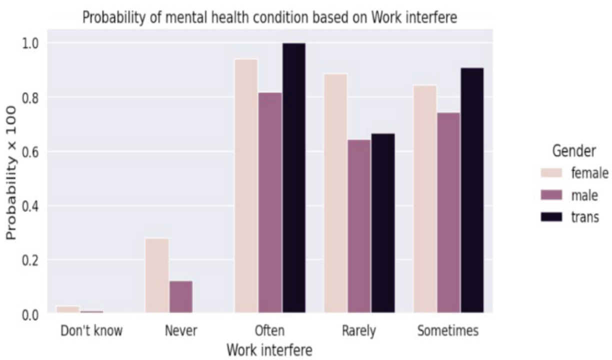

Fig.2.6. mental illness that interferes with job

G. Model Evalution

During the rigorous evaluation phase of the algorithm, we employed a robust suite of critical performance metrics—accuracy, precision, recall, and the F1 score. These metrics are indispensable tools for assessing the effectiveness and predictive power of each developed model, providing a detailed

IV. Result and Discussion

TABLE II. RESULTS TABLE WITH HYPER PARAMETER TUNING

Models	Precision	Recall	F1-Score	Accuracy
RF	0.927	0.929	0.928	0.929
SVM	0.935	0.928	0.931	0.933
LGBM	0.962	0.967	0.964	0.964
GB	0.963	0.965	0.964	0.964
XGB	0.970	0974	0.972	0.972
AB	0.992	0.992	0.992	0.992
AB+RF (HPT))	**0.996**	**0.995**	**0.996**	**0.996**

AdaBoost has the greatest accuracy at 0.992, according to bar chart fig.1.1. Both LightGBM and Logistic Regression score around 0.964, indicating that both complicated ensemble approaches and simpler models work well for the data. Gradient Boosting trails with about 0.964, while Random Forest and SVM have 0.929 and 0.933, respectively. Ensemble approaches often minimize bias and variation, improving predictions, therefore they may be better for the dataset. The inferior performance of Random Forest and SVM may need parameter adjustment and data preparation to improve accuracy. Boosting algorithms work across varied datasets and follow machine learning trends. This review emphasizes the need of evaluating processing resources, model interpretability, and dataset features when choosing a method.

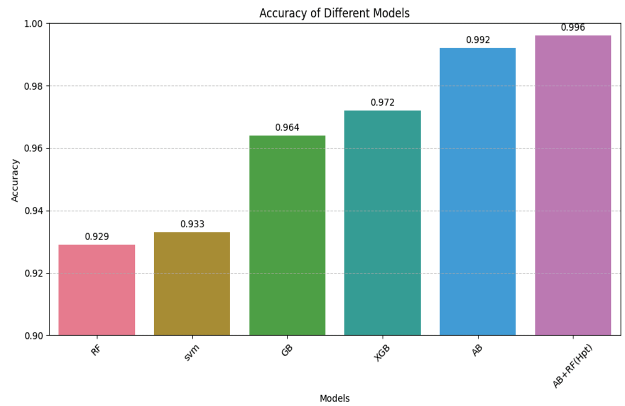

Fig.4.1 Result graph

V. Future Directions

The study's findings provide many intriguing avenues for additional research. This is the most crucial because multimodal data integration might improve mental health projections' accuracy and customization. Mental health diseases are complicated, requiring sophisticated deep learning approaches and algorithmic parameter optimization. Transforming patient care may need models that predict outcomes and offer customized treatment strategies. Following study, machine learning projections may be linked to clinical applications to ensure that technology advances improve healthcare. Since personal health data is being used more, data privacy will remain a primary focus, which might lead to safer federated learning systems.

VI. Conclusion

This study shows how machine learning improves mental health diagnosis and treatment. Ensemble approaches, like AdaBoost, predict mental health outcomes better than other models. This emphasizes the need of using complicated algorithms and multimodal data to treat mental health illnesses. A comprehensive strategy using modern computational tools and clinical insights to adapt patient treatment is recommended by the research. Future research should bridge the gap between prediction models and actual implementations to ensure that machine learning advances improve healthcare. The ethical considerations of exploiting personal health data emphasize the need for safe and private technology. Overall, ML has the potential to change psychological well-being care by making therapy more tailored, efficient, and accessible.

Acknowledgment

This work is an outcome of "Student Research Internship Training (SRIT)" from Marwadi University, Department of Computer Engineering and a collaborative effort from Alliance University, TKIET, and Surya Engg. College.

REFERENCES

[1] "Forbes Health." [Online]. Available: https://www.forbes.com/health/mind/mental-health-statistics

[2] F. Faisal, M. M. Nishat, K. R. Raihan, A. Shafiullah, and S. Ali, "A Machine Learning Approach for Analyzing and Predicting Suicidal Thoughts and Behaviors," *Int. Conf. Ubiquitous Futur. Networks, ICUFN*, vol. 2023-July, pp. 43–48, 2023, doi: 10.1109/ICUFN57995.2023.10201075.

[3] R. Chahar, A. K. Dubey, and S. K. Narang, "A Mental Health Performance Assessment using Support Vector Machine," *2023 3rd Int. Conf. Intell. Technol. CONIT 2023*, pp. 1–7, 2023, doi: 10.1109/CONIT59222.2023.10205772.

[4] Y. Li, "Application of Machine Learning to Predict Mental Health Disorders and Interpret Feature Importance," *2023 3rd Int. Symp. Comput. Technol. Inf. Sci. ISCTIS 2023*, pp. 257–261, 2023, doi: 10.1109/ISCTIS58954.2023.10213032.

[5] V. Laijawala, A. Aachaliya, H. Jatta, and V. Pinjarkar, "Classification algorithms based mental health prediction using data mining," *Proc. 5th Int. Conf. Commun. Electron. Syst. ICCES 2020*, no. Icces, pp. 1174–1178, 2020, doi: 10.1109/ICCES48766.2020.09137856.

[6] S. G. Singh, K. Kandoi, S. K. Singh, R. Bediya, and K. Mishra, "Improving Cardiovascular Prediction with Advanced Machine Learning and Deep Learning Approaches," 2023.

[7] E. Naveen Paul and S. Juliet, "Comparative Analysis of Machine Learning Techniques for Mental Health Prediction," *Proc. 8th Int. Conf. Commun. Electron. Syst. ICCES 2023*, pp. 1–6, 2023, doi: 10.1109/ICCES57224.2023.10192763.

[8] T. Ngo *et al.*, "Federated Deep Learning for the Diagnosis of Cerebellar Ataxia: Privacy Preservation and Auto-Crafted Feature Extractor," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 803–811, 2022, doi: 10.1109/TNSRE.2022.3161272.

[9] L. Zhu *et al.*, "Stress Detection Through Wrist-Based Electrodermal Activity Monitoring and Machine Learning," *IEEE J. Biomed. Heal. Informatics*, vol. 27, no. 5, pp. 2155–2165, 2023, doi: 10.1109/JBHI.2023.3239305.

[10] T. Chawla, S. Mittal, and K. Srinivasan, "Hyperparameter Optimization of Machine Learning Models for Monitoring Menstrual Mental Health," *Proc. IEEE InC4 2023 - 2023 IEEE Int. Conf. Contemp. Comput. Commun.*, vol. 1, pp. 1–5, 2023, doi: 10.1109/InC457730.2023.10263184.

[11] M. Nouman, S. Y. Khoo, M. A. P. Mahmud, and A. Z. Kouzani, "Recent Advances in Contactless Sensing Technologies for Mental Health Monitoring," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 274–297, 2022, doi: 10.1109/JIOT.2021.3097801.

[12] K. Nagarajaiah, M. H. Krishnappa, and K. R. Asha, "Machine Learning based Detection of Post Traumatic Stress Disorder of Mental Health," *2023 4th Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2023 - Proc.*, pp. 1456–1460, 2023, doi: 10.1109/ICESC57686.2023.10193036.

[13] P. Basak Upama, M. Valero, H. Shahriar, M. Syam, and S. Iqbal Ahamed, "Mental Health Analysis during Pandemic: A Survey of Detection and Treatment," *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 2023-June, pp. 1530–1538, 2023, doi: 10.1109/COMPSAC57700.2023.00236.

[14] R. R N, S. Jain and M. Sarkar, "SMOTE and Hyperparameter Optimization: A Dual Machine Learning Strategy for Enhancing Coupon Recommendation in Vehicular Contexts," 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/SMARTGENCON60755.2023.10442306.

[15] R. N. Ravikumar, M. Banga and M. Sarkar, "An Experimental Study on Hyper-parameter Optimization for Breast Cancer Risk Prediction using IBM Snap Machine Learning Techniques," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2022, pp. 1-5, doi: 10.1109/ICCSEA54677.2022.9936467.

[16] S. S. Nigatu, P. C. R. Alla, R. N. Ravikumar, K. Mishra, G. Komala and G. R. Chami, "A Comparitive Study on Liver Disease Prediction using Supervised Learning Algorithms with Hyperparameter Tuning," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2023, pp. 353-357, doi: 10.1109/InCACCT57535.2023.10141830.

[17] 1Ravikumar . R N., S. . Jain, and M. . Sarkar, "Efficient Hybrid Movie Recommendation System Framework Based on A Sequential Model", Int J Intell Syst Appl Eng, vol. 11, no. 9s, pp. 145–155, Jul. 2023.

[18] Ravikumar R N, Sanjay Jain and Manash Sarkar, "LSTM-GNOG: A New Paradigm to Address Cold Start Movie Recommendation System using LSTM with Gaussian Nesterov's Optimal Gradient" International Journal of Advanced Computer Science and Applications(ijacsa), 15(6), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150675

[19] Datasetlink: https://www.kaggle.com/datasets/shariful07/student-mental-health