Human Interaction and Emerging Technologies (IHIET-FS 2025), Vol. 196, 2025, 123–133
https://doi.org/10.54941/ahfe1005961

Transforming Mental Health Assessment: Machine Learning for Early Detection and Personalized Care Among College Students

Madhav Theeng Tamang, Mhd Saeed Sharif, and Seyed Ali Ghorashi

School of Architecture, Computing and Engineering, UEL, University Way, London, E16 2RD, UK

ABSTRACT

The growing global incidence of mental health disorders underlines the urgent need for improved tools to enable early diagnosis and intervention. This study investigates the potential of machine learning models to predict mental health issues among college students by utilizing a dataset that includes a variety of demographic and behavioural characteristics. This study employs several Machine learning models, including Logistic Regression, Random Forest, Decision Tree, and XGBoost, using a dataset comprising demographic, behavioural, and self-reported mental health information. Data preprocessing involved cleansing, normalization, and feature selection to optimize model performance. Models were trained and validated using cross-validation, and their performance was measured using metrics such as accuracy, precision, and ROC-AUC scores. Machine learning models, particularly Logistic Regression, show significant potential for improving mental health assessments by providing early, accurate, and scalable predictions. This study is significant in addressing the rising mental health challenges among college students by leveraging machine learning (ML) for early detection and personalized care. Traditional diagnostic methods, often time-consuming and subjective, are enhanced by ML's ability to process large datasets for faster, accurate, and scalable assessments. The Logistic Regression model achieved an accuracy of 85% and a precision of 81%, demonstrating its reliability for general mental health predictions. By integrating demographic, behavioural, and physiological data, the study promotes tailored interventions while emphasizing ethical considerations like privacy and transparency. Its findings can guide institutions and policymakers in developing data-driven mental health programs, fostering healthier academic environments and advancing mental health care.

Keywords: Machine learning, Mental health, Logistic Regression, Data preprocessing

INTRODUCTION

Traditional mental health assessments often rely on clinical tests and self-report questionnaires. However, with advancements in machine learning (ML), mental health evaluations have become faster and more accurate

(Burger & Scholz, 2018). ML, a branch of artificial intelligence (AI), enables algorithms to learn from data and solve complex problems. While humans may overlook patterns in large datasets, ML models can detect these insights, helping to predict mental health issues, track their progression, and even recommend personalized treatments. By analyzing social media activity, electronic health records (EHRs), and data from wearable devices, ML offers a new approach to mental health care.

Developing a predictive model for mental health evaluation involves several crucial steps, including data collection and preparation, feature selection, model training, validation, and deployment. Collecting patient mental health data is particularly challenging due to privacy concerns and the sensitive nature of the information. While ML has the potential to enhance mental health care, its implementation must follow strict ethical guidelines and ensure robust data security. Machine learning offers several advantages for mental health assessments. Predictive algorithms could help identify at-risk individuals early, leading to timely and personalized interventions that improve overall treatment outcomes. However, practical and ethical concerns—such as transparency, fairness, and accountability in model development and implementation—must be carefully addressed in this sensitive field.

In conclusion, ML-driven predictive models represent a promising innovation in mental health evaluation. By leveraging modern data analytics and computational techniques, these models can enhance the accessibility, efficiency, and precision of mental health care, ultimately benefiting both individuals and society as a whole.

COMPREHENSIVE OVERVIEW OF EXISTING LITERATURE

Machine learning (ML) has become an essential tool in mental health prediction and diagnosis. Various studies have explored its potential in improving early detection, treatment, and risk assessment for different mental health conditions. Vaishnavi et al. (2022) examined five ML methods—Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), and Stacking to predict mental health disorders. The study aimed to improve early detection, highlighting that current screening methods are costly and time-consuming. Similarly, Mutalib, Shafiee, and Abdul-Rahman (2021) explored mental health prediction models using machine learning in higher education institutions. They found that Decision Tree, Support Vector Machine (SVM), and Neural Networks were the most effective models for detecting stress, depression, and anxiety, particularly among students. The study also emphasized the link between mental health disorders and social isolation, financial hardship, and academic pressure, underscoring the need for targeted interventions.

Krishnan et al. (2024) developed a predictive machine learning model to assess mental health issues in higher education students due to COVID-19 using the HADS assessment. Iyortsuun et al. (2023) provided a review of machine learning and deep learning approaches for mental health diagnosis. They further evaluated ML and deep learning (DL) models using accuracy, F1-Score, precision, recall, and AUC metrics, revealing that some models achieved up to 90% accuracy. However, the study highlighted key challenges, including data privacy concerns, model interpretability, and the need for extensive datasets. The authors recommended training and validating models across different datasets to improve robustness and avoid overfitting. Similarly, Shafiee and Mutalib (2020) investigated the prediction of mental health problems among higher education students using machine learning. They explored the prevalence of mental health issues among Malaysian college students. Their review identified academic stress, socioeconomic struggles, and lack of social support as major contributors to mental health challenges. The study found that supervised ML methods, particularly SVM, demonstrated predictive accuracy ranging from 70% to 96%.

Several studies have focused on ML and DL applications in diagnosing disorders such as schizophrenia, depression, anxiety, bipolar disorder, PTSD, anorexia, and ADHD. Organisciak et al. (2022) introduced RobIn, a robust interpretable deep network for schizophrenia diagnosis. This framework was able to achieve the F1-Score of 98.56% in schizophrenia diagnosis. Bentley et al. (2022) examined the implementation of machine learning models for suicide risk prediction in clinical practice through a focus group study with hospital providers. Despite concerns about liability, alert fatigue, and system demand, clinicians were open to ML tools that tracked patient risk indicators and streamlined workflows. Similarly, Mohamed et al. (2023) developed a hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms. They developed a hybrid ML model to classify anxiety levels among individuals in conflict zones, achieving a maximum accuracy of 98.13% with Random Forest. The model also provided tailored recommendations for mental health providers.

In another study, Nanath et al. (2022) developed a mental health index using a machine learning approach to assess the impact of mobility and lockdown during the COVID-19 pandemic. They analysed over three million tweets, measuring the psychological impact of COVID-19 lockdowns. Their findings indicated that restrictions on workplace mobility significantly affected mental health, whereas limitations on access to parks and grocery stores had a minimal impact. Prout et al. (2020) used a machine learning approach to identify predictors of psychological distress during COVID-19. Papini et al. (2023) developed and validated a machine learning prediction model for posttraumatic stress disorder (PTSD) following military deployment. Army personnel, emphasizing the need for objective, data-driven tools to complement traditional diagnostic methods.

Opoku Asare et al. (2021) explored the use of machine learning methods, hyperparameter optimization, and feature importance analysis to predict depression from smartphone behavioural markers. Over a 22-day study involving 629 participants, researchers analysed behavioural markers like screen time and internet usage, finding that ML models could accurately detect depressive symptoms. This highlights the potential for passive, scalable mental health monitoring using mobile technology.

Caldirola et al. (2022) used a machine learning approach to predict new-onset psychiatric disorders throughout the COVID-19 pandemic. A Gradient Boosted Decision Tree model achieved 70% sensitivity and 73% specificity, identifying resilience levels and pandemic-related stress as key predictors. Czyz et al. (2023) developed machine learning-based decision tools to predict short-term suicidal thoughts in adolescents at a daily level following hospitalization. Using Classification and Regression Trees (CARTs), the model achieved an AUC of 0.86, with self-efficacy, burdensomeness, and mood changes as strong predictors.

Finally, Koutsouleris et al. (2021) explored multimodal machine learning workflows for predicting psychosis in patients with clinical high-risk syndromes and recent-onset depression across multiple European sites. By integrating clinical, neurocognitive, genetic, and MRI data, the study achieved an 85.5% balanced accuracy rate, demonstrating ML's potential for scalable, real-world applications in mental health diagnostics. Collectively, these studies underscore the promise of ML in mental health assessment, while also highlighting challenges related to data privacy, ethical considerations, and model interpretability. Addressing these issues will be crucial to ensuring ML's responsible integration into mental health care.

METHODOLOGY

This study employs a machine learning-based approach to predict mental health conditions using a publicly available dataset. The methodology consists of data acquisition, preprocessing, feature selection, model development, training, evaluation, and deployment to ensure robustness and accuracy.

Data Acquisition and Preprocessing

The dataset used in this study is sourced from Kaggle a widely recognized repository (OSMI, 2023). It contains self-reported survey responses from individuals in the tech industry, covering mental health status, workplace conditions, and demographic attributes. The dataset is pre-processed to ensure data integrity and compatibility with machine learning models. Data anonymization is implemented to protect user privacy, ensuring compliance with ethical standards and data protection regulations. Handling missing values is a critical step, where imputation techniques such as median, mode, or mean replacement are applied based on the nature of the missing data. Categorical variables are converted into numerical representations using one-hot encoding and label encoding, while numerical data is standardized through min-max scaling to maintain uniformity. Outlier detection and removal techniques are applied to minimize data noise, and feature selection methods such as Principal Component Analysis (PCA), correlation analysis, and Recursive Feature Elimination (RFE) are used to enhance model efficiency by reducing redundant variables. The processed dataset is then split into training (70%), validation (15%), and testing (15%) subsets to ensure robust model generalization and prevent overfitting. These preprocessing techniques optimize the dataset for accurate, fair, and interpretable mental health predictions.

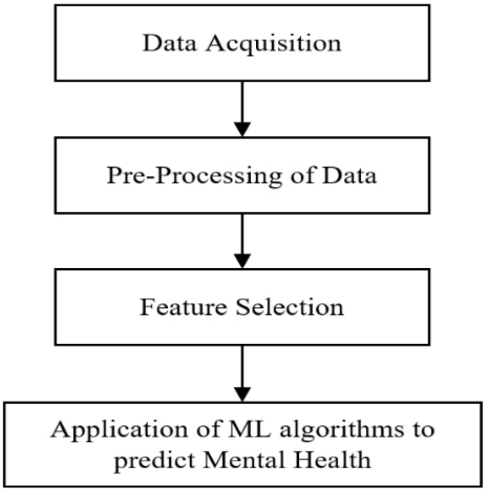

Figure 1: Stages of this study.

IMPLEMENTATION

This study developed and implemented a data analysis approach based on machine learning. A comprehensive literature review was conducted on the best machine learning algorithm. Logistic Regression, Random Forest, Decision Tree, and XGBoost were selected to execute the data analysis.

Similarly, Receiver Operating Characteristic (ROC) curves and the confusion matrix were utilized to assess the model's performance. ROC curves provide a comprehensive visual representation of a classification model's ability to distinguish between classes at various threshold levels, offering valuable insights into its predictive power (Kumar & Indrayan, 2011). Likewise, the confusion matrix was employed to evaluate and summarize the effectiveness of the selected machine learning techniques (Patro & Patra, 2014). The confusion matrix serves as a detailed analytical tool, presenting the number of correct and incorrect predictions for each class. By breaking down these results, it facilitates the assessment of the model's accuracy, precision, and overall classification reliability.

Logistic Regression (LR)

It is a widely used statistical model for binary classification, including mental health prediction. It analyses input features like physiological data, survey responses, and demographics to predict conditions such as anxiety and depression. LR's simplicity and interpretability help medical professionals understand how factors like sleep patterns and social activity influence mental health outcomes. Using this algorithm, we were able to achieve an accuracy of 85%. The Figure 2 illustrates the performance of this model using the ROC/AUC curve and Confusion matrix.

Figure 3 presents the confusion matrix for the binary classification model. It compares the model's predicted labels with the actual outcomes, providing insights into classification performance. The matrix highlights true positives, true negatives, false positives, and false negatives, enabling an assessment of the model's accuracy.

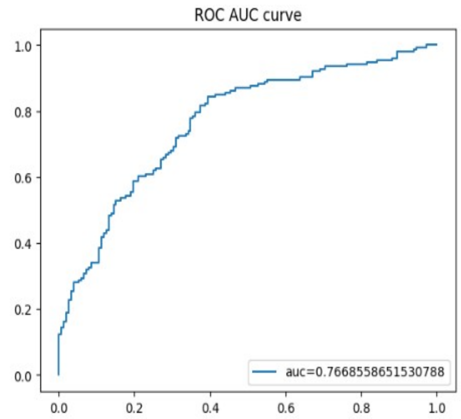

Figure 2: ROC/AUC curve for Logistic Regression.

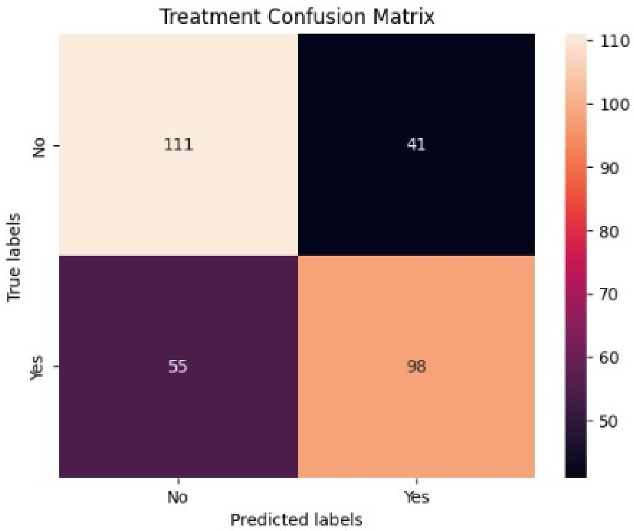

Figure 3: Confusion matrix to evaluate the performance of Logistic Regression.

Random Forest

It is an ensemble learning method that enhances prediction accuracy and mitigates overfitting by aggregating multiple decision trees. In mental health prediction, RF excels at managing noisy data and handling large, feature-rich datasets, ensuring robust and reliable predictions. The Figure 4 and Figure 5 illustrates the performance of this model using the ROC/AUC curve and Confusion matrix.

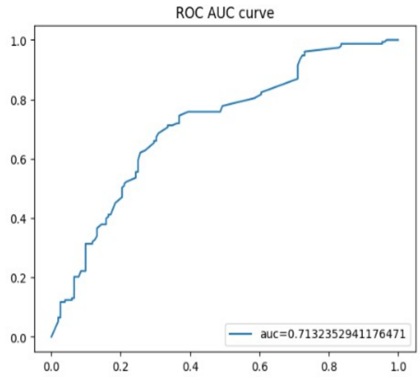

Figure 4: ROC/AUC curve to illustrate the performance of Random Forest.

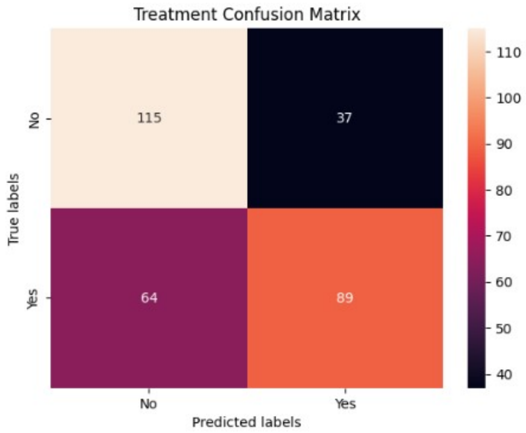

Figure 5: Confusion matrix to evaluate the performance of Random Forest.

XGBoost Classifier

It is a high-performance gradient boosting algorithm known for its accuracy and efficiency. By iteratively enhancing weak learners like decision trees, it refines model performance. In mental health prediction, XGBoost effectively processes large, complex datasets and uncovers intricate feature relationships for improved predictive accuracy. The Figure 6 and Figure 7 illustrates the performance of this model using the ROC/AUC curve and Confusion matrix.

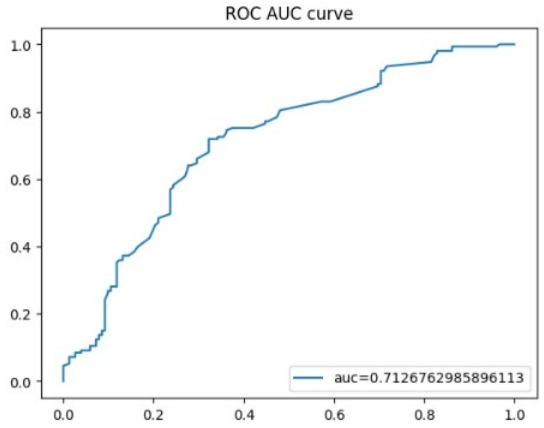

Figure 6: ROC/AUC curve to illustrate the performance of XGBoost.

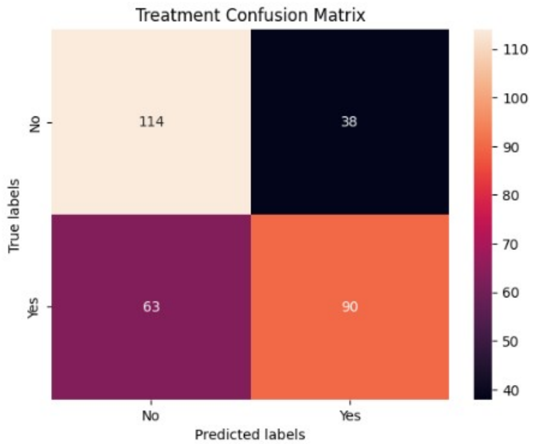

Figure 7: Confusion matrix to evaluate the performance of XGBoost.

Decision Tree

Decision Trees are a supervised learning algorithm that predicts outcomes by applying decision rules derived from data features. In mental health prediction, DTs provide an interpretable model that visualizes the decision-making process. For instance, a decision tree can illustrate how factors like stress, poor sleep quality, and high social media use contribute to an increased risk of depression, aiding in targeted intervention strategies. The Figure 8 and Figure 9 illustrates the performance of this model using the ROC/AUC curve and Confusion matrix.

In the Figure 9 above presents the confusion matrix for the binary classification model predicting "Treatment," comparing actual labels to model-predicted labels. The model correctly identified 116 cases where treatment was not needed (True Negative), while 36 cases were incorrectly predicted as requiring treatment when they did not (False Positives), indicating an overestimation of care necessity. Additionally, 70 cases where treatment was needed were misclassified as not requiring treatment

(False Negative), highlighting potential under-detection. Conversely, the model accurately predicted treatment necessity in 83 cases (True Positives), confirming its ability to identify individuals requiring mental health support.

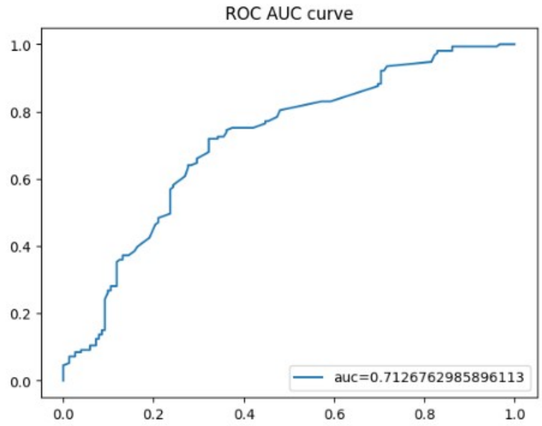

Figure 8: ROC/AUC curve to illustrate the performance of Decision Tree.

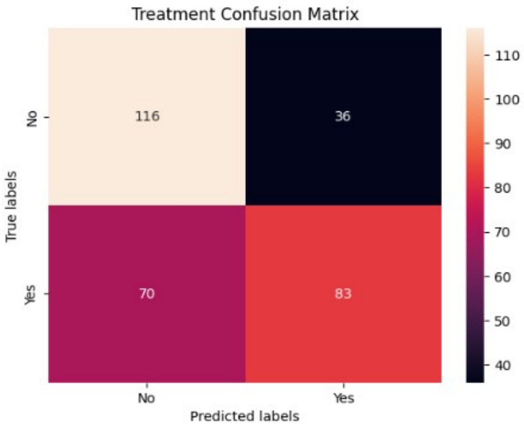

Figure 9: Confusion matrix to evaluate the performance of Decision Tree.

CONCLUSION

In this research, machine learning algorithms have been a significant advancement in mental health care, particularly for early detection and personalized care among college students. Several machines learning models, including Random Forest, Logistic Regression, XGBoost, and Decision Trees, were employed to predict mental health conditions based on survey data. Among these, Logistic Regression (LR) achieved an accuracy of 85% and a precision of 81%, demonstrating its effectiveness as an interpretable model for general mental health assessment. Similarly, the AUC score obtained from the Random Forest, XGB Classifier and Decision Tree are 0.713, 0.712 and 0.713 respectively.

Our findings emphasize the importance of data preprocessing, including feature selection, normalization, and handling missing values, in improving model performance. Properly curated data significantly contributed to the accuracy and reliability of predictions. Nonetheless, challenges such as dataset diversity, model interpretability, and ethical considerations remain. The reliance on self-reported data and potential biases in survey responses highlight the need for more representative datasets to ensure fairness across different demographic groups.

Overall, this study demonstrates the feasibility of machine learning in mental health prediction, reinforcing its potential for early intervention and personalized care. With continued advancements in model optimization and ethical AI integration, machine learning can play a transformative role in improving mental health care outcomes and fostering proactive mental health support for vulnerable populations.

ACKNOWLEDGMENT

The authors would like to acknowledge the School of ACE at UEL and all those who provided support during this research.

REFERENCES

Bentley, K. H., Zuromski, K. L., Fortgang, R. G., Madsen, E. M., Kessler, D., Lee, H., Nock, M. K., Reis, B. Y., Castro, V. M. and Smoller, J. W., 2022. Implementing machine learning models for suicide risk prediction in clinical practice: Focus group study with hospital providers. JMIR Formative Research, 6(3), p. e30946.

Burger, P. H. and Scholz, M., 2018. Gender as an underestimated factor in mental health of medical students. *Annals of Anatomy-Anatomischer Anzeiger*, *218*, pp. 1–6.

Caldirola, D., Cuniberti, F., Dacco, S., Grassi, M., Torti, T. and Perna, G., 2022. Predicting new-onset psychiatric disorders throughout the COVID-19 pandemic: A machine learning approach. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *34*(3), pp. 233–246.

Czyz, E. K., Koo, H. J., Al-Dajani, N., King, C. A. and Nahum-Shani, I., 2023. Predicting short-term suicidal thoughts in adolescents using machine learning: Developing decision tools to identify daily level risk after hospitalization. Psychological medicine, 53(7), pp. 2982–2991.

Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J. and 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.

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J. and Ruef, A., 2021. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA psychiatry*, *78*(2), pp. 195–209.

Kumar, R. and Indrayan, A., 2011. Receiver operating characteristic (ROC) curve for medical researchers. *Indian pediatrics*, *48*, pp. 277–287.

Mohamed, E. S., Naqishbandi, T. A., Bukhari, S. A. C., Rauf, I., Sawrikar, V. and Hussain, A., 2023. A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms. Healthcare Analytics, 3, p. 100185.

Mutalib, S., Shafiee, N. S. M. and Abdul-Rahman, S., 2021. Mental health prediction models using machine learning in higher education institution. Turkish Journal of Computer and Mathematics Education, 12(5), pp. 1782–1792.

Nanath, K., Balasubramanian, S., Shukla, V., Islam, N. and Kaitheri, S., 2022. Developing a mental health index using a machine learning approach: Assessing the impact of mobility and lockdown during the COVID-19 pandemic. *Technological forecasting and social change*, *178*, p. 121560.

Opoku Asare, K., Terhorst, Y., Vega, J., Peltonen, E., Lagerspetz, E. and Ferreira, D., 2021. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR mHealth and uHealth*, *9*(7), p. e26540.

Organisciak, D., Shum, H. P., Nwoye, E. and Woo, W. L., 2022. RobIn: A robust interpretable deep network for schizophrenia diagnosis. *Expert Systems with Applications*, *201*, p. 117158.

OSMI, 2016. *Mental Health in Tech Survey* [Online]. Available at: https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey.

Papini, S., Norman, S. B., Campbell-Sills, L., Sun, X., He, F., Kessler, R. C., Ursano, R. J., Jain, S. and Stein, M. B., 2023. Development and validation of a machine learning prediction model of posttraumatic stress disorder after military deployment. JAMA network open, 6(6), pp. e2321273–e2321273.

Patro, V. M. and Patra, M. R., 2014. Augmenting weighted average with confusion matrix to enhance classification accuracy. Transactions on Machine Learning and Artificial Intelligence, 2(4), pp. 77–91.

Prout, T. A., Zilcha-Mano, S., Aafjes-van Doorn, K., Békés, V., Christman-Cohen, I., Whistler, K., Kui, T. and Di Giuseppe, M., 2020. Identifying predictors of psychological distress during COVID-19: A machine learning approach. Frontiers in Psychology, 11, p. 586202.

Shafiee, N. S. M. and Mutalib, S., 2020. Prediction of mental health problems among higher education student using machine learning. International Journal of Education and Management Engineering, 10(6), p. 1.

Vaishnavi, K., Kamath, U. N., Rao, B. A. and Reddy, N. S., 2022. Predicting mental health illness using machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012021). IOP Publishing.