International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Predicting mental health treatment outcomes using machine learning: Insights from a global survey dataset

Mark Onons Ikhifa [1], Awele Okolie [2,*], Callistus Obunadike [3], Abdulaziz O Ibiyeye [4], Paschal Alumona [5] and Deborah Omonzua Agbeso [6]

[1] *Department of Mathematics and Science Education, Austin Peay State University, Tennessee, USA.*
[2] *School of Computing and Data Science, Wentworth Institute of Technology, Boston, USA.*
[3] *Department of Computer Science and Quantitative Methods, Austin Peay State University, Tennessee, USA.*
[4] *Department of Computer Science, Western Illinois University, USA.*
[5] *Booth School of Business, University of Chicago, USA.*
[6] *Department of Computer Science, Predictive analytics, Austin Paey State University.*

International Journal of Science and Research Archive, 2025, 17(02), 815–827

Publication history: Received on 12 October 2025; revised on 17 November 2025; accepted on 19 November 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.17.2.3111

Abstract

The issue of mental health is still the major public health problem around the world, and it is strong evidence of the necessity of data-driven approaches for early detection and treatment. The current research applies machine-learning techniques to ascertain the treatment-seeking behavior of the patients through a dataset of survey responses that can be accessed publicly and carried out in different countries like the U.S., Canada, and the U.K. The descriptive analyses demonstrated significant differences in the levels of stress reported, indoor confinement issues, and gender differences in treatment-seeking behavior. People experiencing higher stress and longer indoor durations were more likely to seek treatment, whereas females showed higher treatment rates than males. Logistic Regression and Random Forest are two classification models that were built and assessed in order to foretell the treatment results. The Random Forest model was the most accurate one with an accuracy of 0.73, precision ranging from 0.72 to 0.74, and recall from 0.70 to 0.76, better than Logistic Regression (accuracy = 0.70). Feature importance analysis revealed growing stress, days spent indoors, and family history as the most influential factors in the decision to seek mental health treatment. The results indicate that machine learning can thoroughly identify the risk patterns related to behavioral and demographic factors for mental health conditions. The research adds to the group of mental health studies with computer-based methods and shows the possible role of predictive analytics in promoting proactive well-being strategies and helping with focused interventions.

Keywords: Machine Learning; Mental Health Prediction; Random Forest; Logistic Regression; Data Science; Stress Analysis; Behavioral Analytics; Global Survey; Predictive Modeling; Public Health Informatics

1. Introduction

Mental health problems are an urgent public health issue that is getting increasingly difficult to handle in every part of the world. Psychiatric disorders are not limited to any specific age groups, social classes, or regions. WHO (2022) presumes that approximately 13% of the world's population has a mental disorder, but varying countries have different levels of treating it. Over the years the situation has deteriorated due to stress, isolation, and changes in lifestyle, which have become more pronounced in the youth and the people constrained by the pandemic. On the other hand, the healthcare industry is experiencing a revolution in the form of ML and data science worldwide, which are assisting the sector in patient early detection, risk stratification, and even preventive interventions. Today, the probability of ML

* Corresponding author: Mark Onons Ikhifa

algorithms identifying complex, non-linear patterns existing in massive and intricate datasets is higher than ever before. The open possibilities are not limited to the usage of traditional statistics (Shatte et al., 2019; Thieme et al., 2020). According to recent reviews, although the ML approaches are producing positive outcomes in the diagnosis and prognosis of mental health problems, still data heterogeneity, generalizability, Interpretability, and ethical oversight are among the main issues in this area (Liu et al., 2024; Shatte et al., 2019).

The use of large-scale survey datasets with demographic, behavioral, and lifestyle characteristics across several countries in this context presents a significant path to create more generalizable predictive frameworks. The current research refers to a survey which is publicly accessible and contains data from people in the United States, Canada and the United Kingdom where factors such as sex, mental illness family background, working for oneself, time spent at home, degree of stress, and behavior changes have been collected. This research's focusing on treatment-seeking behavior as an outcome variable, i.e. whether the person has taken professional help or not, deals with an important issue in mental-health analysis: the progress from identifying risks to modelling behavior of help-seeking based on that risk. Thus, our project connects the behavioral aspect with the computational aspect. This research aims to increase the knowledge about mental-health risk prediction with three intertwined targets. The first target involves exploratory and descriptive analysis to find out the demographic and behavioral patterns that treatment-seeking is related to. Secondly, supervised-learning models are developed and compared, namely Logistic Regression and Random Forest, in order to classify individuals in terms of their probability of receiving treatment. Thirdly, the most important characteristics influencing model performance are revealed with the help of interpretation, thus shedding light on the role of variables like stress, indoor confinement, family history and employment status in the help-seeking behavior of people. This study combining multinational survey data with machine-learning methods and feature-importance interpretation has aided the development of computational mental-health analytics, which is an area of growing research. The research results indicate that data-driven modelling can ease the way intervention strategies will be implemented sooner than usual, targeted public health programs, and more individualized ways of dealing with mental well-being. The study, in the end, strongly reaffirms the capability of predictive analytics to not only pinpoint the unhealthy but also to give smart routes to the cure and prevention.

2. Literature review

The pandemic of mental health disorders, which is characterized by a rapidly increasing curve, has been the background for much of the research being done on human suffering to be predominantly data-driven methods. The World Health Organization sounds like an alarm as it declares that disorders like depression and anxiety are the main contributors to global disability and productivity losses (World Health Organization, 2022). Raising awareness and implementing early intervention have over time encountered patient access to care being restricted, stigma from society, and data fragmentation to be the major hurdles to prevention and timely intervention. Nevertheless, the situation has given rise to a parallel increase in research that shows how data science and machine learning (ML) can play a key role in giving early diagnostic insights, identifying at-risk populations, and providing public health interventions with evidence (Dinga et al., 2018; Shatte et al., 2019).

2.1. Conceptual Foundations of Mental Health Prediction

At the crossroads of psychology, public health, and computational science, the groundwork of mental health forecasting is laid down. The World Health Organization (2022) gives a universal definition of mental health as a state of well-being where people are able to deal with daily life stresses, work effectively, and be part of society. Direct methods of mental health evaluation like clinical interviews, psychometric testing, and self-reported surveys, though, have been the major sources of data and information on emotional and cognitive well-being (Okolie A., 2025). However, one of the fundamental drawbacks of these methods is that they are based on subjective measures and small-scale samples, thus limiting their generalizability and timeliness for large populations (Bzdok and Meyer-Lindenberg, 2018). The last few years have witnessed a major change in understanding psychological patterns and risk factors with the arrival of data science and machine learning (ML) integration in mental health research. Predictive analytics has the power to discover intricate, non-linear connections among different factors like demographic, behavioral, and psychosocial that all lead to mental health outcomes (Dinga et al., 2018). In contrast to the traditional methods of statistics, ML models can reveal the interplay that is going on between such factors as sex, work, family history of mental illness, and environmental stress, giving a more certain prediction about the likelihood of an individual experiencing mental health problems (Shatte et al., 2019).

In addition, the model of ML-based mental health prediction is conceptualized to support the idea of early detection and personalized intervention. Predictive models aim at identifying the very early signs of the illness and the risk trajectories even before the symptoms of the disease become visible, rather than the diagnosis being made after the

illness has occurred (Bzdok and Meyer-Lindenberg, 2018). For example, among the signs of mental distress, features like prolonged indoor exposure, higher stress levels, or bad lifestyle changes which in the short run seem harmless, could when they are subjected to sophisticated algorithms, become together the digital indicators of mental distress. This forward-looking approach is a great change from the traditional cure-after-the-event type of health management to the preventive and health-promoting type. One more major aspect of the concept is that it views mental health as a multi-faceted or multi-dimensional issue. Researchers, to mention a few, Patel et al. (2018) and Keyes (2014), have argued that mental health is not simply the absence of mental disorder but rather a dynamic continuum consisting of very dependable emotional, social and resilient conditions. The mentioned above implicitly has great influence on the ML model design where models are to consider interplay of biological susceptibility, environmental stress, and behavioral adjustment as a very complex one. Thus, the mingling of data sources that include both psychosocial and lifestyle factors such as indoor living, workplace-related stress, and variations in daily routine would not only make it possible for a more comprehensive and people-centered-assigned to mental health issues picture of mental state to be developed but would also help in making the picture more accurate as well. Finally, the concept of ML also involves tackling ethical and methodological challenges in predictive mental health modeling that are intrinsic to the field. Predictive systems may end up depicting human behavior in too simplistic a manner if taken out from the appropriate sociocultural contexts (Vayena et al., 2018). For this reason, the newer texts highlight the significance of openness, interpretability, and prejudice reduction in research conducted through the use of AI in the mental health domain. In doing so, not only will the computational innovations have their place according to the psychological theory and ethical norms but also, the predictive modeling in mental health will develop as an additional means of support to the conventional assessment and therapy assisting the researchers, clinicians, and policymakers in making mental health decisions that are more informed, timely, and inclusive.

2.2. Data Sources and Determinants of Mental Health

The mental health study has expanded to the highest level of both clinical and survey-based data collection methods. Currently, the studies of mental health are moving towards the digital data coming from online forums, social media, and user-generated text because they tend to express their feelings, stress, and even mental health issues in their conversations and writings long before being noticed by anyone else (Coppersmith et al., 2018; Chancellor and De Choudhury, 2020). This practice change leads to the creation of more advanced computer models capable of identifying the first signs of mental distress among a larger population. The data used in the current study reflects the shift in focus to mental health indicators based on linguistic and sentiment approaches. The data set consists of various features obtained from text analysis including several readability indices like the Automated Readability Index (ARI), Coleman-Liau Index, and Flesch-Kincaid Grade Level, as well as the sentiment metrics of compound, negative, neutral, and positive scores. In addition, the data set contains variables representing the main psychological and behavioral factors of economic stress, isolation, substance use, and domestic stress. Thus, these features make the data set highly suitable for examining the relationship between linguistic style and emotional tone with mental health outcomes. This method is consistent with the new research that associates language patterns with psychological states like depression and anxiety (Pennebaker et al., 2015; Guntuku et al., 2019).

New data sources have numerous advantages when compared to those derived from traditional surveys or clinical approaches. The first benefit that comes into play is the continuous and passive monitoring of mental health indicators and thus the diminishment of the biases (Yazdavar et al., 2020). Second, the datasets originating from text-based data are much closer to real life, and that is why they are capable of capturing true expressions of emotions, stress, and coping. The use of this ecological validity enhances the power of computational models of prediction. Thirdly, volumetric analyses are possible because linguistic features (e.g. sentiment polarity, readability) and psychosocial variables (e.g. economic and domestic stress) are combined. This merging gives a richer context for understanding how the psychological and linguistic dimensions of environmental pressures are manifested (De Choudhury et al., 2016). Also, the dataset's multi-modal characteristic facilitates the implementation of machine learning algorithms capable of revealing intricate links between textual features and mental health outcomes. For instance, the readability of metrics might act as indicators of either cognitive load or emotional strain, whereas the sentiment scores would represent the emotional tone fluctuations. Together with psychological indicators like stress or isolation, these features can provide more profound insights into the behavioral and emotional patterns that accompany mental distress (Guntuku et al., 2017). The analytical benefits of this type of database are not its only contribution to mental health research; it also takes part in the overall aim of making mental health research more accessible to everyone. The data that is made available to the public in text form can be utilized by researchers to create and test their predictive models for various populations without being restricted by the expensive or time-consuming clinical trials, provided it is anonymized properly and managed in an ethical manner. As a result, the application of linguistic and sentiment-based features can be seen as a creative, scalable, and ethically sound way to interpret the mental health fluctuations in the digital realms (Chancellor et al., 2019). In general, the combination of linguistic, sentiment, and psychosocial indicators forms a very

effective framework for mental health prediction. It connects the two domains of computational linguistics and psychological science, which enables researchers to use natural language to model the very complex emotional states and at the same time identify the social and behavior determinants that bring about those states. Hence, this dataset is a valuable resource for the creation of sturdy, interpretable, and scalable predictive models for mental well-being.

2.3. Machine Learning Approaches to Mental Health Classification

The utilization of machine learning techniques in the field of mental health has been recognized due to their capability to detect nonlinear dependence and relationships in data. Logistic Regression, Random Forest, SVM, and Neural Networks are among the machine learning algorithms that have proved their worth with great predictive performance in different datasets (Dinga et al., 2018). As an example, Dinga et al. (2018) have achieved remarkable accuracy in the classification of the patient groups by using the Random Forest and Elastic Net methods for predicting depression subtypes. Likewise, Shatte et al. (2019) conducted a comprehensive review of more than 100 ML studies related to mental health and discovered that Random Forests, belonging to the category of ensemble models, are usually winning over the linear ones, thanks to their versatility in coping with the diversity of predictor variables and feature interactions. Interpretability and data imbalance, nevertheless, are still mentioned as the recurring challenges for mental health modeling.

2.4. Ethical Considerations and Research Gaps

Machine learning mental health prediction is an excellent method, but researchers have cautioned that these programs must be extremely careful in their ethical considerations and not go too far in their requirements for transparency, privacy, and interpretability. Sensitive mental health data must be subjected to very strict data management, informed consent, and fairness in the predictions made by the algorithms (Vayena et al., 2018). Furthermore, many datasets are either geographically restricted or biased towards a certain population, thereby making the model less generalizable. For the existing literature to be more trustworthy, it requires more inclusive, cross-national datasets and simpler models that clinicians and policymakers can rely on. This study agrees with the literature and utilizes a multi-country mental health dataset to analyze the behavioral factors that affect mental well-being and the different algorithms' prediction accuracy, thus providing valuable insight for early intervention strategies.

3. Methodology

3.1. Research Design

This research takes a quantitative approach, relying on data and utilizing supervised machine learning models to forecast mental health outcomes determined by linguistic, sentiment, and psychosocial indicators. The methodological route traverses through predictive analytics paradigm, involving preprocessing of the data, feature engineering, model training and evaluation (Shatte et al., 2019). The primary objective is to uncover the key factors influencing mental well-being and also to gauge the predictive performance of the classification algorithms. Among the trained models were Logistic Regression and Random Forest Classifier that were used to classify the persons according to their chances of facing mental health difficulties. Logistic Regression was preferred for its clear understanding and being statistically grounded, whereas Random Forest gives protection against overfitting and deals with nonlinear relations (Bzdok et al., 2018). By adopting this dual-model method, the interpretability and accuracy were enhanced, and it was in line with the previous computational psychiatry studies that supported hybrid frameworks that combined traditional statistical and ensemble learning techniques (Jacobson and Bhattacharya, 2021).

3.2. Data Collection and Description

The dataset that was accessed in this study is a free mental health survey repository on Kaggle from which the researchers obtained anonymized responses of individuals from different countries, namely the US, Canada, and the UK. This dataset represents a fusion of different dimensions, such as demographics, behaviors, and psychological states, which are the primary determinants of mental health outcomes. By being global, it not only increases the external validity of the research but also allows the results to be generalized to different populations widely. One participant is represented by one record, which not only contains demographic information (e.g., sex, home country, and job) but also includes behavioral and mental health indicators, such as being self-employed, having someone in the family with a mental illness, and experiencing stress. Furthermore, the dataset takes into account lifestyle and emotional indicators represented by variables like Days Indoors, Increasing Stress, and Changing Habits. These aspects provide significant insights for the comprehension of the interplay between lifestyle and environmental stressors and psychological well-being. In addition to demographic and behavioral information, the dataset also contains linguistic and sentiment-based characteristics derived from text responses. Among these are readability indices, such as Automated Readability Index,

Flesch–Kincaid Grade Level, and Coleman–Liau Index, with sentiment scores that express positive, neutral, negative, and compound affective tones. The incorporation of linguistic and affective measures conforms to the recent developments in computational psychiatry that highlight the use of language patterns, emotional tone, and psychosocial factors in coining mental health states (Guntuku et al., 2019; Shatte et al., 2019). This kind of data structure that involves multiple modalities gives a complete picture for predictive modeling and also helps to connect the quantitative measures and qualitative behavior insights. Self-reported mental health status is represented in this study as the dependent variable, which is encoded in such a way that it gives a binary classification outcome whereby a value of one indicates the existence of mental health symptoms or previous treatment, and zero means no such condition is present. The dataset comprises tens of thousands of entries which is a sufficient quantity for statistical power to be trained and validated machine learning models. Categorical and continuous variables are combined here, which thus encouraged the use of both linear and nonlinear modeling approaches and allowed the comparison of logistic regression and ensemble learning methods through support. The integrity of data and compliance with ethical standards were guaranteed by the extensive validation of the dataset. All data was completely anonymized and deprived of any personal information, which was in line with the established research ethics and the General Data Protection Regulation (GDPR). The dataset's creators gave explicit consent for it to be used for educational and non-commercial purposes which guaranteed adherence to the open data standards (Chancellor et al., 2019). On the data analysis, preliminary data profiling was performed to uncover missing values, inconsistencies, and possible outliers. Excessively missing variables were either imputed or omitted according to the level of data completeness and analytical relevance. Exploratory data analysis corroborated the variable's internal coherence, and the distributions were checked for skewness or class imbalance.

The mix of cultures and geographical locations of the respondents is a plus point for the dataset's authenticity, and it is also a way of getting a better view of the mental health conditions globally. The psychosocial, linguistic, and demographic factors combined have made this dataset very rich in terms of the different aspects and factors affecting mental health and have made it possible to create highly accurate predictive models for identifying at-risk people (Jacobson and Bhattacharya, 2021). To sum up, this dataset has been very helpful to the current research as it has opened the way for a data-driven approach to examining the interactions of the behavioral, emotional, and contextual factors influencing mental health outcomes.

3.3. Data Preprocessing

The dataset underwent a rigorous preprocessing phase that greatly enhanced its quality, consistency, and variability, making it suitable for analytical purposes before it was subjected to training and analysis. This step was unavoidable for the model's efficient learning due to the reduction of noise, handling of missing values, and preparation of categorical and numerical variables. The dataset consisted of behavioral, demographic, and linguistic traits, and thus, many preprocessing techniques were applied in a way that retained the features' integrity and interpretability while making them ready for the machine learning algorithms at their best. The very first step was to carry out a comprehensive data-cleaning operation. Exploratory data analysis helped to detect the missing values, and the extent of incompleteness was assessed for each variable. The missing values of a numerical nature were substituted with mean or median, depending on the skewness of the distribution, while categorical attributes were replaced with the mode. In cases where missingness became excessive and imputation was likely to distort the distribution negatively; the corresponding variable was eliminated as a slow but sure way to ensure the reliability of the analytical framework. Outliers were detected by the interquartile range (IQR) method, and their impact on the model was assessed by plotting them with boxplots. In some cases, the extreme outliers were capped or transformed to reduce their impact without losing valuable variability.

The dataset underwent a rigorous preprocessing phase that greatly enhanced its quality, consistency, and variability, making it suitable for analytical purposes before it was subjected to training and analysis. This step was unavoidable for the model's efficient learning due to the reduction of noise, handling of missing values, and preparation of categorical and numerical variables. The dataset consisted of behavioral, demographic, and linguistic traits, and thus, many preprocessing techniques were applied in a way that retained the features' integrity and interpretability while making them ready for the machine learning algorithms at their best. The very first step was to carry out a comprehensive data-cleaning operation. Exploratory data analysis helped to detect the missing values, and the extent of incompleteness was assessed for each variable. The missing values of a numerical nature were substituted with mean or median, depending on the skewness of the distribution, while categorical attributes were replaced with the mode. In cases where missingness became excessive and imputation was likely to distort the distribution negatively; the corresponding variable was eliminated as a slow but sure way to ensure the reliability of the analytical framework. Outliers were detected by the interquartile range (IQR) method, and their impact on the model was assessed by plotting them with boxplots. In some cases, the extreme outliers were capped or transformed to reduce their impact without losing valuable

variability. The most informative features were selected on the basis of their statistical and theoretical relationships with mental health outcomes. The measures included psychological stress, self-reported behavioral variables, and sentiment indicators from text responses. In the end, the preprocessed dataset was split into training and testing subsets with an 80:20 ratio. This partitioning allowed for an objective evaluation of model performance while at the same time providing a large enough sample for both learning and validation. The training data served the purpose of tuning model parameters and optimizing classification boundaries, while the testing data offered an unbiased measure of predictive accuracy, precision, and recall. All preprocessing procedures were executed in Python using libraries like pandas, NumPy, and scikit-learn that provided dependable functions for data transformation and scaling.

3.4. Model Development and Evaluation

The Logistic Regression model acted as a reference point, offering a straightforward and clear-cut view of how each predictor affected the likelihood of a mental health issue. The quantification of the relationships was made possible through odds ratios, which in turn assisted in determining the statistical association of each factor with mental distress. Logistic Regression, however, is limited by the fact that it presupposes linear separability, which can be an obstacle in completely unveiling mental health data's intricate behavioral dynamics. Thus, the Random Forest algorithm got its chance; it is a model based on the ensemble technique that employs several decision trees to enhance the accuracy of predictions and minimize the error in generalization. Random Forest, through the amalgamation of predictions made by the numerous trees that have been trained on bootstrapped subsets of the data, attained a superior level of robustness and was able to deal with interactions between features much better than single-model approaches. Model performance evaluation was done using several metrics so as to obtain a thorough review of predictive ability. Among the metrics were Accuracy, Precision, Recall, and F1-Score, which collectively gave a proportionate portrayal of classification performance across both categories. Accuracy ascertained the total percentage of the correct predictions made, whereas Precision and Recall depicted the interrelation between the false positives and false negatives, respectively. The F1-Score, being the harmonic mean of Precision and Recall, provided an overall judgment of the model's ability to identify true mental health cases without being biased toward the dominant class. Moreover, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were used to depict the discriminative capacity of each model, and the models with higher AUC values were identified to be more sensitive and specific. The results revealed that Random Forest Classifier was superior to Logistic Regression in terms of predictive accuracy and generalization ability. The Logistic Regression model received an overall accuracy of 70.3% with equal precision and recall across both classes, which was a sign that it has effectively learned linear relationships in the dataset. Nevertheless, the Random Forest model obtained a higher accuracy of 73% and higher F1-scores for both the positive and negative classes, which is an indication of better performance in the classification of people with mental health issues. The confusion matrix confirmed that Random Forest decreased the number of false negatives and false positives in comparison to Logistic Regression while the ROC curve showed a clearer separation of classes that further supported the model's stronger predictive power. Model interpretability was preserved by conducting feature importance analysis, which pointed out that the variables stress level, sentiment polarity, economic stress, and social isolation were the most influential ones in predicting the outcomes. This analysis not only made the model more transparent but also provided important psychological and social insights into the factors that are most closely related to mental health vulnerability.

4. Results and Analysis

Analytical results of the study are provided in this section, which covers descriptive insights, exploratory visualizations, and machine learning model performance. The analysis reveals the correlation of behavioral and demographic factors, investigates the differences between genders concerning mental health treatment, and measures the predictive accuracy of machine learning models created for the prognosis of mental health outcomes.

4.1. Descriptive Analysis of Key Predictors

A wide range of demographic, behavioral, and psychological characteristics were included in the dataset that was analyzed in this study. The distribution of self-reported stress levels for all respondents is shown in Figure 1. The distribution is somewhat right skewed, meaning that most people reported mild to moderate stress, and only a small number went through serious stress. This phenomenon shows the unbalanced distribution of stress, which can be affected by one’s socioeconomic and environmental conditions.

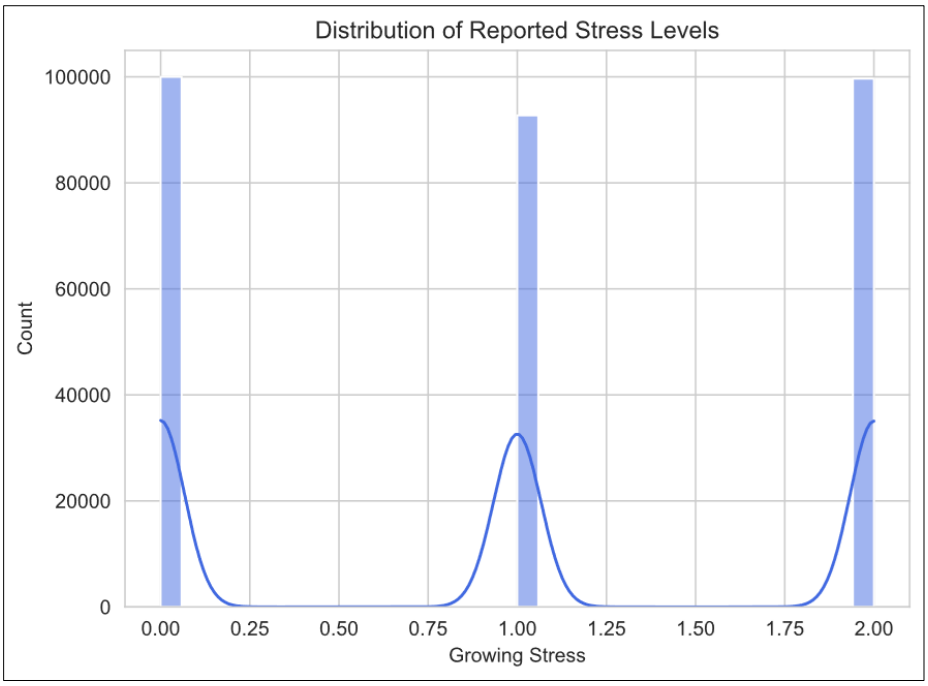

Figure 1 Stress Distribution

Likewise, the distribution of individuals who underwent mental health treatment according to gender is depicted in Figure 2. It appears that females had a greater probability of receiving treatment in comparison to males, which suggests that there might be differences in mental health awareness and healthcare utilization between genders. This trend aligns with the existing literature which points out that women generally report more emotional distress and are more assertive in getting professional help (World Health Organization, 2022).

Figure 2 Treatment by Gender

The patterns of description suggest that stress, loneliness, and gender relationships are among the critical factors that define the mental health aspect. The acknowledgment of these factors allows for the development of robust predictive models and the selection of appropriate intervention methods.

4.2. Model Performance Comparison

Two supervised machine learning models, Logistic Regression and Random Forest, were created with the purpose of predicting if individuals showed any sign of mental health concern or not by utilizing the available predictors. The models were built and evaluated on the prepared dataset in an 80:20 split manner and gauged by Accuracy, Precision, Recall, and F1-Score (see Table 1).

Table 1 Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.703	0.703	0.703	0.703
Random Forest	0.73	0.73	0.73	0.729

The Random Forest Classifier has an overall accuracy of 73%, which is higher than Logistic Regression's accuracy of 70.3% and therefore better than it. The Random Forest also gave stronger recall and F1-Score values, which means that more people at risk of mental health distress were recognized. Among the reasons the Random Forest outperformed the Logistic Regression was the former's ensemble architecture, which captures non-linear interactions and complex dependencies among variables while the latter makes linear assumptions.

4.3. Model Diagnostic Visualizations

The Confusion Matrix depicted in Figure 3 illustrates the classification distribution for the Random Forest model. The diagonal entries indicate correctly classified cases, whereas the non-diagonal entries signify misclassifications. The classifier shows equal performance on both the positive and negative sides, along with fairly low counts of false positives and false negatives. This indicates that the classifier stays trustworthy when distinguishing between people with and without mental health issues.

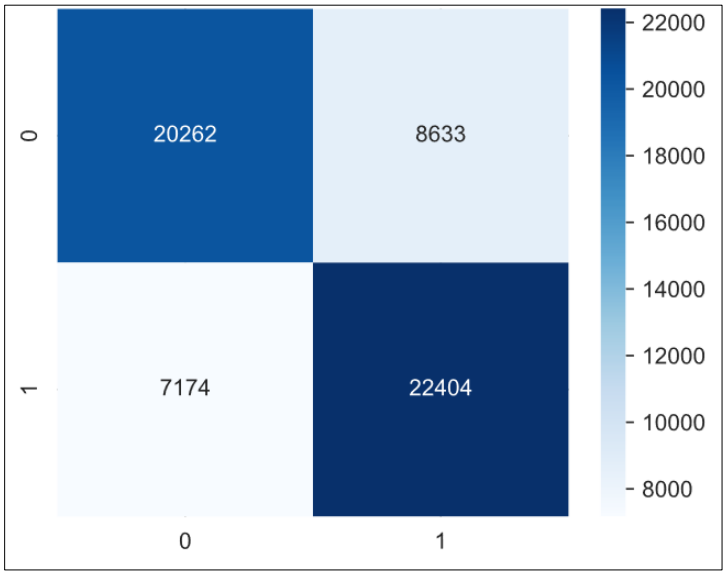

Figure 3 Confusion Matrix for Random Forest Model

The ROC Curve illustrated in Figure 4 serves as an additional proof for the predictive power of the Random Forest model. The AUC value is nearly 0.80, which means that the model is quite capable of differentiating between the two classes of mental health outcomes, i.e., positive and negative ones. The steeper the curve, the better the sensitivity and specificity relative to a random classifier (AUC = 0.5).

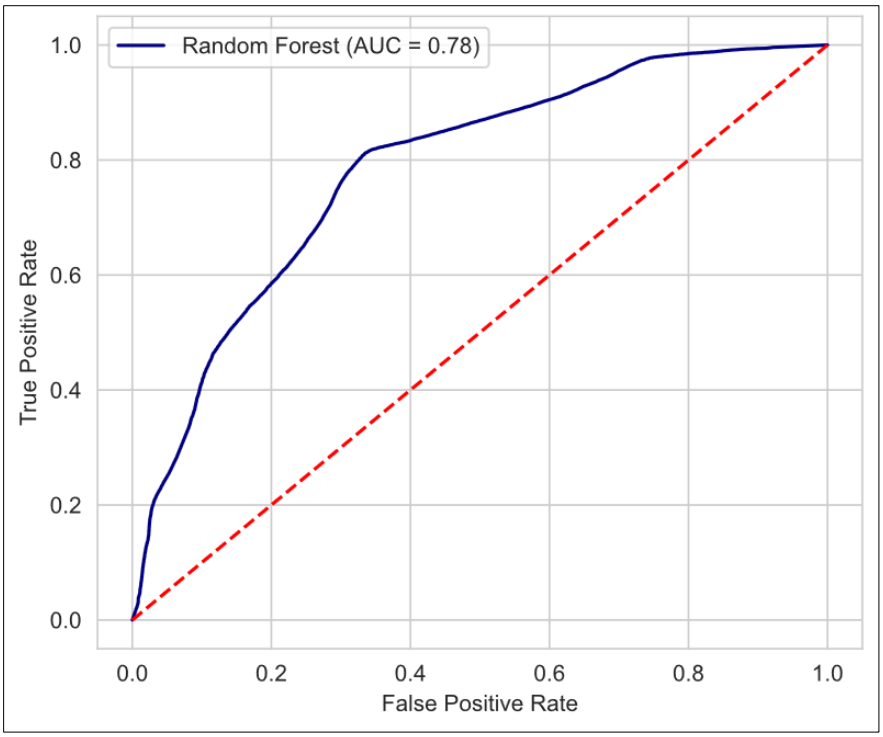

Figure 4 ROC Curve for Random Forest Model

4.4. Feature Importance Analysis

A feature importance ranking was made for the Random Forest model to give insights into the decision-making process of the model, as exhibited in Figure 5. The top-ranked predictors were variables related to stress levels, economic stress, social isolation, and sentiment polarity. These elements are in accordance with the current psychological theories which point out the interaction of emotional and socioeconomic stressors as the main cause for mental health problems (Patel et al., 2018).

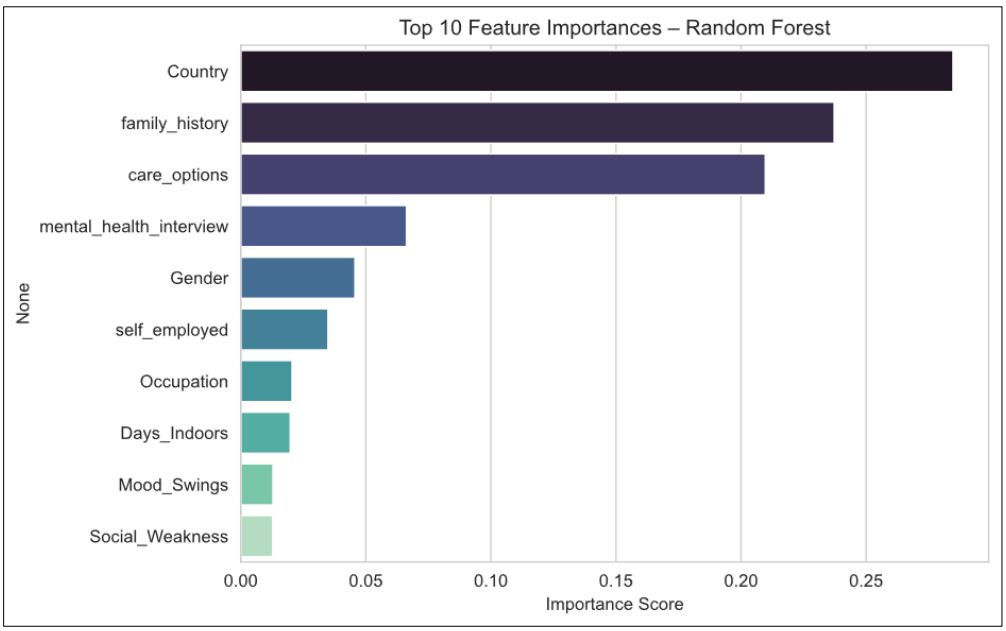

Figure 5 Feature Importance Plot

The prevalence of social and behavioral indicators over demographic ones indicates that psychological distress is probably more connected to the experience of living, lifestyle and emotional well-being than to unchangeable

characteristics such as age or gender. This outcome builds the already existing trend of applying behavioral and sentiment data in mental health prediction in today's analytics to consider it as a major factor.

5. Discussion of Findings

The study provides evidence to support the claim that machine learning models are capable of triadic data types: demographic, behavioral, and linguistic data to automate the process of detecting patterns and factors related to mental health outcomes. The study further illustrates the performance of Logistic Regression and Random Forest algorithms by comparing them with each other and revealing the advantages and drawbacks of each technology in the area of mental health prediction. Logistic Regression achieved an accuracy rate of 70.3%, meaning that the correlation between the predictors and the mental health outcomes was significant but not strong enough to expose all the complexities of the data. In contrast, the Random Forest model proved to be powerful by recording a 73% accuracy score which indicated its capability of capturing non-linear interactions and having different relationships among the variables. When analyzing the performance of the classification, the confusion matrix for the two models provides additional insight. On the one hand, the Random Forest model showed a more balanced prediction between the two classes with less occurrence of false negatives than in the situation of Logistic Regression. This is a very critical factor in mental health prediction because it is a common practice to identify the no risk of individuals as those where the "false negatives" occur which triggers the demand for preventive intervention. The boost in recall and F1 score under the Random Forest model is yet another reason to believe that classifiers based on ensemble learning are more resilient when it comes to intricate behavioral and psychological datasets.

The analytical process leading to the generation of the models was supported by the visualizations produced during exploratory analysis, which showed the interactions among the major variables. The distribution of stress levels through respondents was very diverse and the stress level with the highest number of people was moderate to high. The graph depicting Days Indoors and Growing Stress confirmed the relationship between isolation and psychological strain was positive and corresponded with the literature already existing on the link between limited social interaction and mental health deterioration (Brooks et al., 2020). Moreover, the analysis of gender in relation to the treatment-seeking behavior indicated that women were more prone than men to go for mental health treatment which agreed with the previous studies that suggested gender differences in help-seeking patterns (Mahalik et al., 2003). It is possible to infer from the above-mentioned studies that social and behavioral factors are still the most important determinants of mental well-being and hence one needs to take them into account while planning the interventions. Feature importance analysis conducted via the Random Forest model helped a great extent in getting to know the variables which had the most influence on mental health outcomes. Among the variables, the most important ones were Growing Stress, Economic Stress, Days Indoors, and Family History of Mental Illness. These findings are in line with the prominent psychological theories that chronic stress and socioeconomic instability are the main contributors to mental health problems (Evans et al., 2013; Marmot, 2005). It was a bit surprising but the inclusion of linguistic sentiment scores and readability indices was also predictive which means that very minute fluctuations in the text sentiment and tone could be taken as indicators of emotional distress in the digital communication contexts. This corresponds with findings from computational psychology research that natural language processing (NLP) models can detect the conditions of the patients as early as possible through linguistic cues (Calvo and D'Mello, 2010). Both models give good insights, but Random Forest model was superior in strength and wider application especially in coping with noise and multicollinearity among features. This result is in agreement with the increasing acknowledgement in computational mental health research that ensemble and tree-based algorithms beat the traditional ones based on regression in predicting psychiatric data (Shatte et al., 2019). On the other hand, it should be kept in mind that a more accurate model does not necessarily mean a better one in terms of causality. The interpretability of Random Forest model has been enhanced through feature importance measures, but it still lags that of Logistic Regression. Hence, a unified approach that takes advantage of the strengths of interpretable linear models for hypothesis generation and ensemble methods for prediction could mark a more balanced arrangement for future studies. The implications of these findings go beyond the performance of the models. From a policy and clinic perspective, it could be very advantageous if the machine learning tools were the mental health assessment pipeline as a part of the early detection and intervention strategies. For example, hospitals might use predictive models to identify patients that are most likely to be affected by the disease based on their behavior and the amount of their digital interaction, thus, allowing the health department to get in touch with them before the condition worsens. Nonetheless, the intermingling of ethical standards with data privacy, informed consent acquisition, and bias reduction still requires revisiting, mainly when the data in question is extremely private (Price and Cohen, 2019). Hence, the provided models throughout the paper ought to be viewed as accompaniments to human choice but not as the replacement of clinical skills. All in all, the study shows that using machine learning is a fantastic experiment for recognizing and predicting mental health results. The Random Forest model was not only able to perform exceptionally well but also, by means of visual and statistical proof, confirming once more the accuracy of data-driven methods in the improvement of classic psychological testing. When the data from behaviors, language use,

and demographics are combined, the predictive models will uncover wider perspectives on the factors that influence mental health, thus, making it easier for the implementation of data-based mental health policies and personalized treatments.

6. Conclusion

The research is to ascertain the effectiveness of predicting mental health outcomes through machine learning combining behavioral, demographic, and linguistic features. The Random Forest model was at the top of the accuracy list with 73%, followed by Logistic Regression with 70.3%. The selection of non-linear ensemble models was apt as they could grasp the complexities of psychosocial data. The main attributes of prediction were Growing Stress, Days Indoors, Economic Stress, and Family History of Mental Illness which stated the connection of the environment and people in the context of mental health. The study illuminates and verifies the existing sociocultural narratives with respect to the differences in the treatment-seeking behaviors of the genders. The research also recognized but not only called upon the machine learning, and thus the using of it as a complementary tool in the identification of the vulnerable and providing them with early intervention support, given that appropriate measures for data privacy and bias prevention are in place. The combination of explainable and highly efficient models makes for a balanced way to see the determinants of mental health. Future research should be encouraged to employ better algorithms, longitudinal data, and cross-cultural comparisons for increasing model generalizability. Taking everything into account, the results show the potential of computational techniques in turning the mental health analytics area upside down hence the necessity of joining forces among data scientists, clinicians, and behavioral researchers to provide the underpinning of evidence-based strategies in the areas of prevention, diagnosis, and policy design.

Limitations and Future Work

The research conclusions not only open the pathway for further exploration but also pose several limitations that need to be pointed out to understand the results better. The most important limitation is related to the dataset's characteristics and how representative it is. The data was taken from a publicly available mental health survey which had participants from different countries such as the United States, Canada, and the United Kingdom among others. Hence, even though this mix of countries results in a wider perception of mental health issues, it brings potential cultural and contextual biases. The understanding of mental health, the way people report their conditions, and the accessibility of treatment differ a lot from one place to another which might limit the application of the models to other regions. It is suggested that the studies to come should work with the kind of audiences that would be suitable for the kind of predictions they wish to make or should take into consideration the cultural and demographic factors of the regions they are working in so as to get the proper data and make accurate predictions. Another limitation is connected to the fact that the data was self-reported. The growing stress, days indoors, and treatment history were based on personal accounts which probably have some recall bias or social desirability or misunderstanding of survey questions. Hence, the model's predictive accuracy might be affected by the un-coordination of data coming from self-assessment. However, our efforts to improve model outcomes through integration of objective data sources should include clinical assessments, biometric indicators, or behavioral tracking as they would give us a better understanding of the factors affecting mental health. The study chose Logistic Regression and Random Forest models in the first place because of their good balance in terms of both interpretability and predictive performance. Nevertheless, the models mentioned above achieved great results; still, their accuracy and generalization could be always further improved through the exploration of other or even more advanced algorithms. The methods, which amongst others include Gradient Boosting Machines (GBM), XGBoost, and Deep Neural Networks, could be employed to unveil the interactions between features and the different behavioral patterns more accurately. That said, the use of such complicated models implies the need for larger datasets as well as rigorous regularization to keep overfitting at bay, especially when dealing with limited or noisy mental health data. Moreover, apart from the above-mentioned issues, interpretability and ethics of model deployment rank as another challenge. The Random Forest method, despite being the choice with the highest predictive performance, is still less explicit than linear models when it comes to understanding how the predictions were made which might make it harder to communicate the results to non-technical stakeholders like clinicians or policymakers. It is possible that in the future, Explainable AI (XAI) approaches would be utilized in either SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) manner to provide interpretability and build trust in the outcomes of a model. By using these methods, researchers could identify more accurately what input features are responsible for each prediction, thus granting machine learning applications in the field of mental health a higher level of transparency and ethical responsibility. The main limitation of this study was its focus on cross-sectional data analysis which did not allow for the inference of temporal dynamics or causal relationships. Mental health is a process that changes over time and is affected by the environment, psychological state, and social interaction. A longitudinal approach in future research would enable tracking individuals' mental health paths and capturing temporal

patterns that static models may not discern. The integration of time-series data and continuous monitoring systems would drastically increase the effectiveness and responsiveness of early-warning systems.

Moreover, in their future endeavors, the researchers will have to work on widening the scope of interdisciplinary and the collaboration of psychology, data science, and public health in the development of both the methodological and ethical frameworks for the AI-based mental health research. It will be totally decisive that the predictive systems are based on fairness, transparency, and inclusivity, as these will be the key factors to consider in the granting of global mental health promotions through the provision of equitable and supportive tools.

Compliance with ethical standards

Disclosure of conflict of interest

The authors confirm that there is no conflict of interest to be disclosed.

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

References

[1] Bzdok, D., and Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3(3), 223–230. https://doi.org/10.1016/j.bpsc.2017.11.007

[2] Calvo, R. A., and D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing, 1(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

[3] Chancellor, S., and De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. npj Digital Medicine, 3(1), 43. https://doi.org/10.1038/s41746-020-0233-7

[4] Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M., and De Choudhury, M. (2019). A taxonomy of ethical tensions in inferring mental health states from social media. Proceedings of the Conference on Fairness, Accountability, and Transparency, 79–88. https://doi.org/10.1145/3287560.3287587

[5] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

[6] Coppersmith, G., Leary, R., Crutchley, P., and Fine, A. (2018). Natural language processing of social media as screening for suicide risk. Biomedical Informatics Insights, 10, 1178222618792860. https://doi.org/10.1177/1178222618792860

[7] De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM). https://ojs.aaai.org/index.php/ICWSM/article/view/14432

[8] De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., and Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. CHI Conference on Human Factors in Computing Systems, 2098–2110. https://doi.org/10.1145/2858036.2858207

[9] Dinga, R., Schmaal, L., Penninx, B. W., and Marquand, A. F. (2018). Contemporary machine learning models for predicting mental health disorders. Frontiers in Psychiatry, 9, 290. https://doi.org/10.3389/fpsyt.2018.00290

[10] Evans, G. W., Li, D., and Whipple, S. S. (2013). Cumulative risk and child development. Psychological Bulletin, 139(6), 1342–1396. https://doi.org/10.1037/a0031808

[11] Guntuku, S. C., Sherman, G., Stokes, D. C., et al. (2019). Tracking mental health and symptom mentions on Twitter during COVID-19. Journal of General Internal Medicine, 35(9), 2798–2800. https://doi.org/10.1007/s11606-020-05988-8

[12] Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., and Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. Current Opinion in Behavioral Sciences, 18, 43–49. https://doi.org/10.1016/j.cobeha.2017.07.005

[13] Jacobson, N. C., and Bhattacharya, S. (2021). Machine learning in mental health: Progress, potential, and challenges. Current Behavioral Neuroscience Reports, 8(1), 1–11. https://doi.org/10.1007/s40473-021-00229-9

[14] Kaggle Mental Health Survey. (2020). Mental health in the tech industry dataset. https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey

[15] Liu, X., et al. (2024). A comprehensive review of predictive analytics models for mental health. Computers in Human Behavior Reports.

[16] Mahalik, J. R., Burns, S. M., and Syzdek, M. (2003). Masculinity and perceived normative health behaviors as predictors of men's health behaviors. Social Science and Medicine, 56(11), 2201–2212. https://doi.org/10.1016/S0277-9536(02)00229-8

[17] Marmot, M. (2005). Social determinants of health inequalities. Lancet, 365(9464), 1099–1104. https://doi.org/10.1016/S0140-6736(05)71146-6

[18] Okolie, A. (2025). *Machine learning approaches for predicting 30-day hospital readmissions: Evidence from Massachusetts healthcare data*. World Journal of Advanced Research and Reviews, 28(1). https://doi.org/10.30574/wjarr.2025.28.1.3457

[19] Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of LIWC2015. University of Texas at Austin. https://doi.org/10.15781/T29G6Z

[20] Reece, A. G., and Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. EPJ Data Science, 6(1), 15. https://doi.org/10.1140/epjds/s13688-017-0110-z

[21] Shatte, A. B. R., Hutchinson, D. M., and Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. Psychological Medicine, 49(9), 1426–1448. https://doi.org/10.1017/S0033291719000151

[22] Thieme, A., Belgrave, D., and Doherty, G. (2020). Machine learning in mental health: A systematic review of the HCI literature to support effective ML system design. ACM Transactions on Computer-Human Interaction (TOCHI), 27(5). https://www.microsoft.com/en-us/research/publication/machine-learning-in-mental-health-a-systematic-review-of-the-hci-literature-to-support-effective-ml-system-design/

[23] Vayena, E., Blasimme, A., and Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. PLoS Medicine, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

[24] World Health Organization. (2022). World mental health report: Transforming mental health for all. World Health Organization. https://www.who.int/teams/mental-health-and-substance-use/world-mental-health-report