scientific reports

Check for updates

OPEN

Machine learning-based predictive modelling of mental health in Rwandan Youth

Fauste Ndikumana[1,2]✉, Josias Izabayo[3], Joseph Kalisa[3], Mathieu Nemerimana[4], Emmanuel Christian Nyabyenda[1,7], Sylivain Hirwa Muzungu[1], Isaac Komezusenge[1], Melissa Uwase[5], Similien Ndagijimana[5], Celestin Twizere[6] & Vincent Sezibera[3]

Globally, mental disorders are a significant burden, particularly in low- and middle-income countries, with high prevalence in Rwanda, especially among survivors of the 1994 genocide against Tutsi. Machine learning offers promise in predicting mental health outcomes by identifying patterns missed by traditional methods. However, its application in Rwanda remains under-explored. The study aims to apply machine learning techniques to predict mental health and identify its associated risk factors among Rwandan youth. Mental health data from Rwanda Biomedical Center, collected through the recent Rwanda mental health cross-sectional study and with youth sample of 5221 was used. We used four machine learning models namely logistic regression, Support Vector Machine, Random Forest and Gradient boosting to predict mental health vulnerability among youth. The research findings indicate that the random forest model is the most effective with an accuracy of 88.8% in modeling and predicting factors contributing to mental health vulnerability and 75 % in predicting mental disorders comorbidity. Exposure to traumatic events and violence, heavy drinking and a family history of mental health emerged as the most significant risk factors contributing to the development of mental disorders. While trauma experience, violence experience, affiliation to pro-social group and family history of mental disorders are the main comorbidity drivers. These findings indicate that machine learning can provide insightful results in predicting factors associated with mental health and confirm the role of social and biological factors in mental health. Therefore, it is crucial to consider biological and social factors particularly experience of violence and exposure to traumatic events, when developing mental health interventions and policies in Rwanda. Potential initiatives should prioritize the youth who experience social hardship to strengthen intervention efforts.

Keywords Machine learning, Prediction, Mental health, Youth, Rwanda

Worldwide, one person in eight is suffering from a mental health disorder[1]. Approximately 75% of the wide burden of mental and behavioral disorders is concentrated in low and lower-middle-income countries, and a substantial majority (76–85%) remain untreated[2–4]. Mental and behavioral disorders continue to be significant contributors to the global burden, constituting 32.4% of years lived with disability and 13% of disability-adjusted life-years[2,5]. Individuals with severe mental health issues die prematurely, and this condition causes disability.

Around 20% of the world's children and teenagers have a mental health condition, with suicide the second driving cause of death among 15–29 year-olds[6]. Mental health issues' prevalence seems to be rising in Africa as the population increases, and more people, specifically young people, struggle to cope with crises to earn a living[7,8]. The recent studies in Rwanda showed that the prevalence of mental disorders is 20% and the most common are major depressive disorders (12%), panic disorders (8.1%) and post-traumatic stress disorders (3.1%) in general populations; the prevalence among genocide against Tutsi survivors was 53.3% and the most prevalent disorders are major depressive disorders (35%), post-traumatic stress disorders (27.9%), and panic disorders (26.8%)[9].

[1]African Center of Excellence in Data Sciences, University of Rwanda, Kigali, Rwanda. [2]Applied Research and Development & Foresight Incubation, National Industrial Research and Development Agency, Kigali, Rwanda. [3]Center for Mental Health, University of Rwanda, Kigali, Rwanda. [4]Maternal, Newborn, Child and Adolescent Health Program, Partners In Health/Inshuti Mu Buzima, Kigali, Rwanda. [5]College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda. [6]The Regional Centre of Excellence in Biomedical Engineering and EHealth, University of Rwanda, Kigali, Rwanda. [7]Centre for Impact, Innovation and Capacity Building for Health Information Systems and Nutrition (CIIC-HIN), Kigali, Rwanda. ✉email: nfauste@gmail.com

Given the high prevalence and significant impact of mental health conditions, particularly among vulnerable populations like youth, there is a growing need for innovative approaches to identify those at risk and provide timely interventions[10]. Novel technologies and Artificial Intelligence enhance practice-oriented research models by integrating research and practice[11,12]. In this context, leveraging data-driven approaches presents a valuable opportunity. By utilizing survey data and advanced analytical techniques, we can better understand patterns and predictors of mental health issues, enhancing the accuracy of early identification and treatment strategies[13,14].

Advancements in machine learning have shown promising potential in predicting mental health outcomes in various contexts[15,16]. Machine learning, a subset of artificial intelligence, empowers systems to learn from data and improve their performance over time, enabling a wide range of applications that enhance decision-making and automate complex tasks[17]. The use of Machine Learning (ML) in mental health gained more attention due to the growing availability of digital mental health data and the potential for ML techniques to enhance mental health care delivery[18]. ML has been applied in various areas, including the detection, diagnosis, prognosis, treatment, and support of mental health problems, as well as in public health, clinical administration, and research[15,19]. Techniques like natural language processing (NLP) have been used to predict suicide ideation and psychiatric symptoms, aiding in personalized treatment and predictive detection of mental health conditions[20,21]. Studies have demonstrated the use of ML algorithms, such as classification, regression, and deep learning, to diagnose bipolar disorder and analyze the impact of financial policies on public mental health[22,23]. The application of ML on longitudinal mobile sensing data has improved the generalizability and predictive performance for mental health symptoms, particularly depression, anxiety, and PTSD during the COVID-19 pandemic[24]. Moreover, ML models have been employed to predict mental health states across diverse populations, including students and professionals, and to assess mental healthcare service use, revealing both the potential and challenges in resource allocation[14,25]. Additionally, ML models using passive sensing signals have shown strong performance in predicting mental health risks in individuals with diabetes[26]. Therefore, ML techniques hold promise for advancing the understanding of mental illness, enhancing diagnostic accuracy, and tailoring interventions to improve outcomes for individuals with mental health conditions[27,28].

By leveraging machine-learning algorithms on a large dataset, it is possible to identify patterns and relationships that may not be easily discovered through traditional descriptive and inferential statistics[29]. ML techniques have been valuable tools for understanding psychiatric disorders and distinguishing and classifying mental health issues among patients for advanced treatments[30]. Studies identified various machine learning models and assessed their effectiveness in modeling different mental health issues in children, including Attention Deficit/Hyperactivity Disorder (ADHD), major depressive episodes, manic episodes, and anxiety, comparing their performance on different measures of accuracy[31]. Therefore, ML techniques hold promise for advancing the understanding of mental illness, enhancing diagnostic accuracy, and tailoring interventions to improve outcomes for individuals with mental health conditions[27,28].

Several studies have demonstrated the feasibility and effectiveness of machine learning approaches in modeling and predicting mental health outcomes based on behavioral data from specific populations[32–34]. For instance, Ponce et al.[35] integrated sentiment analysis, ML algorithms, and data from social network to develop a mobile application that allowed the early detection and prevention of mental health problems in Peru. This tool enabled timely interventions, contributing positively to public mental health management. Abdul et al.[36] employed machine learning techniques to predict mental wellness among university students using health behavior data. Their findings revealed that ML models could accurately predict mental wellness outcomes, outperforming traditional predictive methods. The study highlights the potential of behavior-based ML approaches to be extended and adapted for use in diverse populations beyond the university setting.

However, the use of machine learning in the prediction of mental health in Rwanda is relatively unexplored. This study aims to use machine learning to predict mental health outcomes in youth in Rwanda. Specifically, this study aims to: (a) Build a predictive machine learning model for mental health vulnerability. Design and implement a machine learning predictive model that analyzes various demographic, social, and behavioral factors to identify at-risk youth. Evaluate models based on different evaluation metrics to select the best one. (b) Determine key predictors of mental health status among Rwandan youth. Conduct feature importance analysis to identify the most significant predictors contributing to mental health issues in youth using the best model selected.

Methods

Research design and source of data

This study used recent cross-sectional records from the Rwanda Mental Health Survey (RMHS 2018) which was conducted by the Rwanda Biomedical Center under the Ministry of Health, in collaboration with partner institutions. The survey was conducted from 1st August to 31st August 2018 nationwide and employed multistage sampling methods with the targeted population being people aged 14–65 who lived in Rwanda at the time of the study. The participants of this study are youth, which means those aged between 15 and 24 according to the United Nations' definition[37]. This subsample is from across the entire country, ensuring a comprehensive and representative sample and comprised of five thousand two hundred twenty-one 5221 subjects.

Data collection tools

The occurrence of common mental disorders was determined using the Mini-International Neuropsychiatric Interview (MINI), version 7.0.2[38]. This diagnostic interview instrument employs a binary response format, where participants respond with either "yes" or "no".

Moreover, traumatic experiences and heavy drinking were assessed using the MINI component on traumas exposure, and section I for alcohol and section J for drugs . On the other hand, violence exposure was measured using a one dichotomous question (Yes or no) whether the person was exposed to violence, mostly domestic violence.

MINI is widely applied in the field of psychiatry. In previous investigations conducted in Rwanda, Kinyarwanda-translated versions of MINI reflecting the participants' native language were used to assess mental health conditions and showed satisfactory psychometric properties[39–41]

Description of variables

Outcome variable

In this study, the outcome variable was any mental health condition classified according to MINI. Mental health status was categorized into two groups: Individuals who screened positive for at least one mental health condition and those who did not test positive for any mental health condition.

Explanatory variables

According to the biopsychosocial model proposed by George Engel[42], human experiences are diverse and are shaped by a complex interplay of factors that influence an individual's life, including biological, psychological, and social determinants that all contribute to shaping a person's experiences, health and behavior. The explanatory variables associated with any mental health condition were the social and biological factors summarized in Table 1.

Data cleaning and pre-processing

Pre-processing of data involves preparing raw data to make it understandable and suitable for analysis. It consists of various stages, such as data cleaning, integration, transformation, and reduction[43]. In this study, data cleaning was performed using Python version 3. Key variables for predicting mental health outcomes were selected based on a biopsychosocial theoritical framework that categorizes causes into distant determinants such as social, biological, and psychological factors. The data were slightly imbalanced; therefore, no methods were applied to handle the imbalance. The selection of features was based on the filtering correlation method.

Data transformation

Data transformation involves altering the format or structure of the data to prepare it for analysis. In this study, categorical variables containing non-numeric data were converted into a numerical format through feature encoding. The dummy variables were created to analyze each variable's contribution.

Machine learning classification algorithms

Machine learning classification algorithms use input training data to predict the likelihood that subsequent data will fall into one of the predetermined categories[44]. Given n observations $\{\mathrm{x}^1, \mathrm{x}^2, \ldots, \mathrm{x}^n\}$, where x^i is an element of X, and $\{y^1, y^2, \ldots, y^n\}$, where y^i belongs to Y. Supervised learning finds a function $f : X \rightarrow Y$ or $P(Y|X)$ such that $f(\mathrm{x}) \approx y$ for all pairs $(\mathrm{x}, y) \in X \times Y$ that have the same observations[44]. For the binary classification problem, $y = \{1, 0\}$ where 0 is labeled as negative and 1 is labeled as positive observations. In this study, ML algorithms were used, including logistic regression, support vector machine, random forest, and gradient boosting.

Logistic regression

Logistic regression is a statistical method used for binary classification tasks, aiming to predict the probability that an instance belongs to a specific class[45,46]. Given a set of predictors $\mathrm{x} = (x_1, \ldots, x_p)^T$ a vector of discrete or continuous variables, a logistic regression models the likelihood that Y falls into a particular category as functions

Variables	Description	Categories
Sex	Sex of participants	0: Female, 1: Male
Age	Age of participants	$\leq$ 18: Adolescents, $>$ 18: Young adults
Marital status	Marital status of participants	0: Never married, 1: Married, 2: Separated/divorced
Education	Level of education of participants	0: No education, 1: Primary, 2: Secondary/higher
Employment status	Status of employment of participants	0: Unemployed, 1: Employed
Lifetime loss	Participant experienced lifetime loss of someone or something	0: No, 1: Yes
Affiliation to pro-social group	Whether participant is affiliated to a pro-social group	0: No, 1: Yes
Violence experience	Experiencing any kind of violence	0: No, 1: Yes
Religion	Religious beliefs of participants	0: No religion/others, 1: Muslim, 2: Christian
Medical condition	Previous experience of medical condition	0: No, 1: Yes
Heavy drinking	Heavy alcohol consumption	0: No, 1: Yes
Traumatic experience	Experiencing traumatic event	0: No, 1: Yes
Mental health history	Had personal mental health history	0: No, 1: Yes
Family history of mental health	Had history of mental health disorders in family	0: No, 1: Yes

Table 1. Explanatory variables for use in the analysis.

of the predictors. The sigmoid function is used in logistic regression to transform the linear combination of the predictor variables into a probability ranging from 0 to 1.

$$P(Y=1 \mid X)=\frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}}$$

$P(Y=1 \mid X)$ signifies the probability of the event $Y=1$ given the predictors X, and the coefficients $\beta_0, \beta_1, \ldots, \beta_p$ are estimated by the model using the training data.

The goal of the logistic regression model is to determine the optimal values of the coefficients β to minimize a loss function. The most common loss function used in logistic regression is the log loss, also known as the cross-entropy loss function, which is defined as:

$$L(Y, f(x))=-\frac{1}{N} \sum_{i=1}^{N}\left(y_i \log(p_i)+(1-y_i) \log(1-p_i)\right)$$

where N is the number of observations, y_i is the actual label for observation i, and p_i is the probability predicted that $Y=1$ for observation i.

Random forest

Random forest is an ensemble method based on decision trees, where each tree is built using a random subset of variables, with the aim of reducing variance. It is inspired by the work of Leo Breiman, Amit, and German's Lier function[47]. The random forest may be used for both categorical and continuous variables as classification and regression, respectively[48].

Given a vector of random variables $X=(x_1, \ldots, x_p)^T$ representing the input or predictor variables and Y representing the dependent variable with real values, we assume an unknown joint distribution $P_{XY}(X, Y)$. The objective is to determine a prediction function $f(X)$ to predict Y.

The prediction function is calculated using a loss function $L(Y, f(X))$, designed to minimize the expected value of the loss:

$$E_{XY}[L(Y, f(X))]$$

where E_{XY} denotes the joint distribution expectation of X and Y. The function $L(Y, f(X))$ measures how close $f(X)$ is to Y, penalizing values of $f(X)$ that deviate significantly from Y.

The loss function $L(Y, f(X))$ can be defined as:

$$L(Y, f(X))=\begin{cases} 0, & \text{if } Y=f(X) \\ 1, & \text{otherwise} \end{cases}$$

ensures that the loss is zero when the prediction matches the true value and penalizes incorrect predictions.

Support vector machine (SVM)

Support vector machine (SVM) is a robust supervised machine learning method used for classification and regression tasks[49]. Its primary objective is to find the optimal hyperplane that effectively separates the classes in the feature space. SVM achieves this by maximizing the margin between classes based on a set of training data consisting of input vectors x_i and their corresponding class labels y_i. This hyperplane is defined in the high-dimensional feature space and aims to accurately classify new data points by their proximity to this separation boundary.

The hyperplane is defined by:

$$W \cdot X+b=0$$

where W is the weight vector perpendicular to the hyperplane, and b is the bias term.

For linearly separable data, SVM determines the optimal hyperplane by solving the following optimization problem:

$$\min_{W, b} \quad \frac{1}{2}\|W\|^2$$

subject to the constraints: $y_i(W \cdot x_i) \geq 1 \quad \text{for all } i$.

Many real-world datasets are not linearly separable. To handle non-linearly separable data, kernel SVM leverages the concept of kernel functions. Kernel SVM maps the input vector x_i into a higher-dimensional space using a kernel function $\kappa(x_i, x_j)$, where the data become linearly separable. The optimization problem for kernel SVM becomes:

$$\min_{W, b} \quad \frac{1}{2} \sum_{i=1}^{N} \alpha_i y_i \kappa(x_i, x)-\sum_{i=1}^{N} \alpha_i$$

subject to the constraints: $\sum_{i=1}^{N} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C \quad \text{for all } i.$

where α_i are Lagrange multipliers and C is a regularization parameter, balancing the objective of maximizing the margin and minimizing the classification error.

Gradient boosting

Gradient boosting is an ensemble learning technique used for classification and regression problems, with a wide range of applications due to its accuracy and robustness against overfitting[50,51]. It combines weak learners during the gradient boosting process to obtain a strong learner. The final result, denoted as $\phi(x)$, is obtained by the addition of the results of K sequential classifier functions f_k and is calculated as follows:

$$\phi(x) = \hat{Y} = \sum_{k=1}^{K} f_k(x)$$

here, f_k represents a decision tree, and K is the total number of iterations in the boosting algorithm.

In these techniques, classification is based on the residuals of the previous iteration, with each feature's influence being sequentially assessed until a desired accuracy level is achieved. Using a loss function $L(\phi)$ that is optimized using gradient descent, the residuals are calculated. The expression for the loss function at step t is given by:

$$L(\phi)_t = \sum L(f_{t-1}, f_t) + \Omega(f_t)$$

where $\Omega(f_t)$ represents the regularization term used to compute the loss function $L(\phi)_t$ at step t.

Model training

The analysis process started by splitting the processed data set into training and testing sets with a 70–30 ratio, based on the existing literature and proven efficacy in mental health data applications. Four machine algorithms were selected for training, namely logistic regression, gradient boosting, random forest, and SVM. Each model was trained using a 10-fold cross-validation and evaluated based on the evaluation metrics from k-fold cross-validation. Moreover, we checked for overfitting and underfitting by comparing metrics on both the training and test datasets. A grid search was applied to optimize the hyperparameters and explore predefined values to determine the best-performing class-specific feature importance scores by weighting the standardized mean values of features for each class with the feature importance scores derived from the model.

Hyperparameter tuning

Following hyperparameter tuning using grid search, the SVM was optimized with the following parameters:C=10, gamma=1, kernel=ŕbf. Best parameters for the logistic regression were: C = 0.1, class weight = None, penalty = l2, solver = liblinear. Gradient boosting: learning rate=0.01, max-depth=8, n_estimators=300, subsample=0.8, best parameters for random forest=max-depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=200.

Model evaluation

Evaluating the performance of a machine learning model is key in any project; this study explores different approaches for evaluating model performance. A number of metrics used include precision, accuracy, recall, F1-score, and area under the curve, which is computed from the confusion matrix, which quantifies correct and incorrect predictions across different classes[52,53].

Data analysis

Prevalence of mental health among youth

Figure 1 illustrates the distribution of mental health status among youth, the results highlight the prevalence of mental health conditions in youth is 14.48%. This indicates that 14 individuals out of every 100 have mental health disorders.

Major depressive episode is the most prevalent mental disorder (Table 2), followed by other common conditions such as panic disorder (5.94%), obsessive-compulsive disorder (2.57%), post-traumatic disorder (1.99%), social phobia (1.32%), antisocial personality (1.28%), alcohol use disorder (1.03%), suicidal behavior disorder (0.52%), and substance use disorder (0.34%).

The majority of the individuals (67%) suffer from only one mental disorder (Fig. 2). 153 individuals (20%) have two mental disorders while 96 individuals (13%) have more than two mental disorders. The individuals, who experience severe comorbidity, may require integrated care and specialized treatment for managing multiple disorders.

Bivariate analysis

A bivariate analysis was performed using Pearson's chi-square test in a confidence interval of 95% to assess the link between the independent factors and the dependent variable. In this study, a p-value below 0.05 was considered to indicate statistical significance. Table 3 presents a summary of the participants' background characteristics. A total of 5,221 participants, with females comprising 56.6% of the sample. The data were categorized by age, with the young adults representing the most prevalent age group.

From the result in Table 3, heavy drinking, trauma experience, being divorced or separated, and violence experience are associated with any mental health vulnerability at 46.9%, 44.1%, 41.2% and 40.8%, respectively.

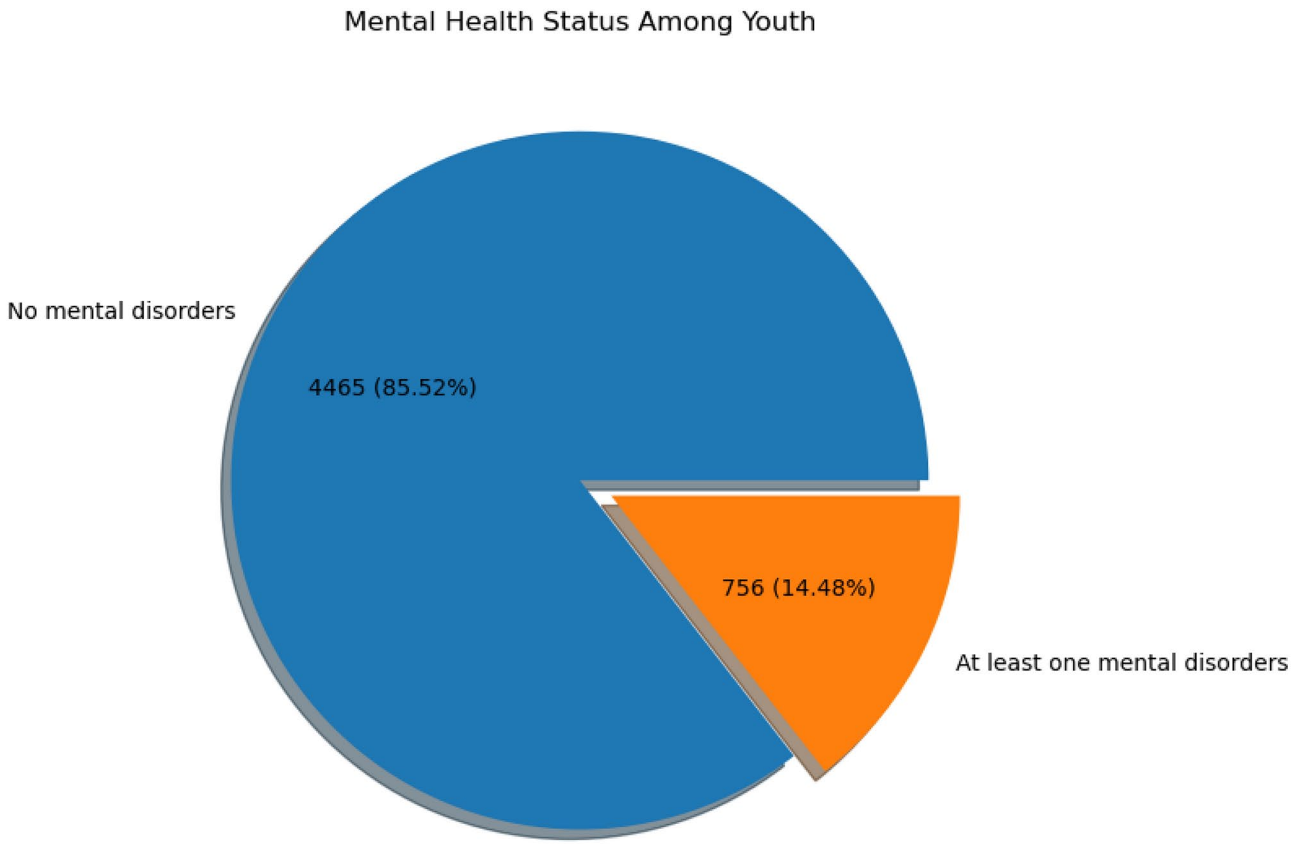


Fig. 1. Prevalence.

Mental disorder	Frequency	Prevalence (%)
Major depressive episode	347	6.65
Suicidal behavior disorder	27	0.52
Panic disorder	310	5.94
Social phobia	69	1.32
Obsessive compulsive disorder	134	2.57
Posttraumatic stress disorder	104	1.99
Manic episode	5	0.10
Hypomanic episode	4	0.08
Alcohol use disorder	54	1.03
Substance use disorder	18	0.34
Antisocial personality disorder	67	1.28
Bipolar disorder	7	0.13

Table 2. Frequency and prevalence of mental disorders.

Moreover, medical condition history, having a family history of mental health, having no religion, and being employed are associated with mental health at 28.0%, 24.2%, 23.0% and 22.3% respectively.

Generally, the bivariate analysis shows that most factors are strongly associated with the likelihood of having mental disorders. The most linked categories include sex, education, marital status, employment status, religion, mental health history, medical condition experience, heavy drinking, violence experience, trauma experience, lifetime loss, family history of mental illness, and age. On the other hand, affiliation to pro-social groups was not associated with the likelihood of developing at least one mental health disorder.

Results

Machine learning classification analysis

Table 4 presents the performance metrics of four machine learning models; the performance of each model is evaluated using precision score, recall score, F1 score, and accuracy.

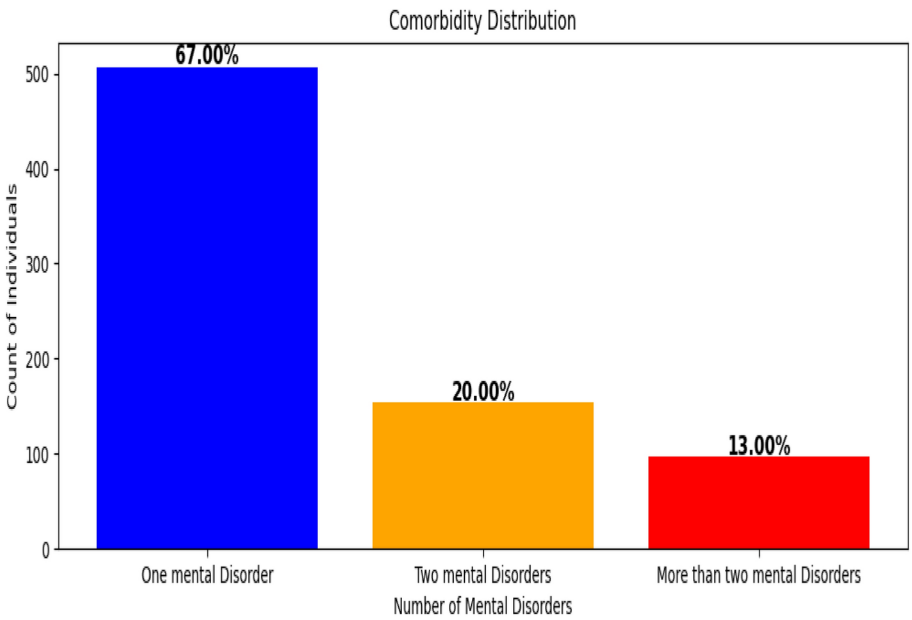


Fig. 2. Comorbidity distribution.

Variable	Category	Total (n, %)	No mental disorders (n, %)	Any mental disorders (n, %)	P-value
Sex	Female	2956 (56.6)	2475 (83.7)	481 (16.3)	0.000
	Male	2265 (43.4)	1990 (87.9)	275 (12.1)	
Education	No education	541 (10.4)	430 (79.5)	111 (20.5)	0.000
	Primary	2951 (56.5)	2510 (85.1)	441 (14.9)	
	Secondary/higher	1729 (33.1)	1525 (88.2)	204 (11.8)	
Marital status	Separated	34 (0.7)	20 (58.8)	14 (41.2)	0.000
	Married	844 (16.2)	675 (80.0)	169 (20.0)	
	Never married	4343 (83.2)	3770 (86.8)	573 (13.2)	
Employment	Employed	309 (5.9)	240 (77.7)	69 (22.3)	0.000
	Unemployed	4912 (94.1)	4225 (86.0)	687 (14.0)	
Religion	Christian	4982 (95.4)	4268 (85.7)	714 (14.3)	0.029
	Muslim	126 (2.4)	110 (87.3)	16 (12.7)	
	Other/None	113 (2.2)	87 (77.0)	26 (23.0)	
Mental health history	Yes	242 (4.6)	167 (69.0)	75 (31.0)	0.000
	No	4979 (95.4)	4298 (86.3)	681 (13.7)	
Medical condition	Yes	107 (2.0)	77 (72.0)	30 (28.0)	0.000
	No	5114 (98.0)	4388 (85.8)	726 (14.2)	
Heavy drinking	Yes	177 (3.4)	94 (53.1)	83 (46.9)	0.000
	No	5044 (96.6)	4371 (86.7)	673 (13.3)	
Violence experience	Yes	554 (10.6)	328 (59.2)	226 (40.8)	0.000
	No	4667 (89.4)	4137 (88.6)	530 (11.4)	
Trauma experience	Yes	555 (10.6)	310 (55.9)	245 (44.1)	0.000
	No	4666 (89.4)	4155 (89.0)	511 (11.0)	
Lifetime loss	Yes	2250 (43.1)	1796 (79.8)	454 (20.2)	0.000
	No	2971 (56.9)	2669 (89.8)	302 (10.2)	
Affiliation with prosocial group	Yes	1455 (27.9)	1228 (84.4)	227 (15.6)	0.165
	No	3766 (72.1)	3237 (86.0)	529 (14.0)	
Family history of mental illness	Yes	1159 (22.2)	878 (75.8)	281 (24.2)	0.000
	No	4062 (77.8)	3587 (88.3)	475 (11.7)	
Age	Adolescent	2287 (43.8)	2071 (90.6)	216 (9.4)	0.000
	Young adults	2934 (56.2)	2394 (81.6)	540 (18.4)	

Table 3. Bivariate analysis of mental health factors.

Model	Precision	Recall	F1 score	Accuracy
Logistic regression	0.870	0.880	0.857	0.880
Gradient boosting	0.859	0.872	0.844	0.872
Random forest	**0.877**	**0.888**	**0.873**	**0.888**
SVM	0.862	0.870	0.835	0.870

Table 4. Comparison of evaluation metrics for different models. Significant values are in bold.

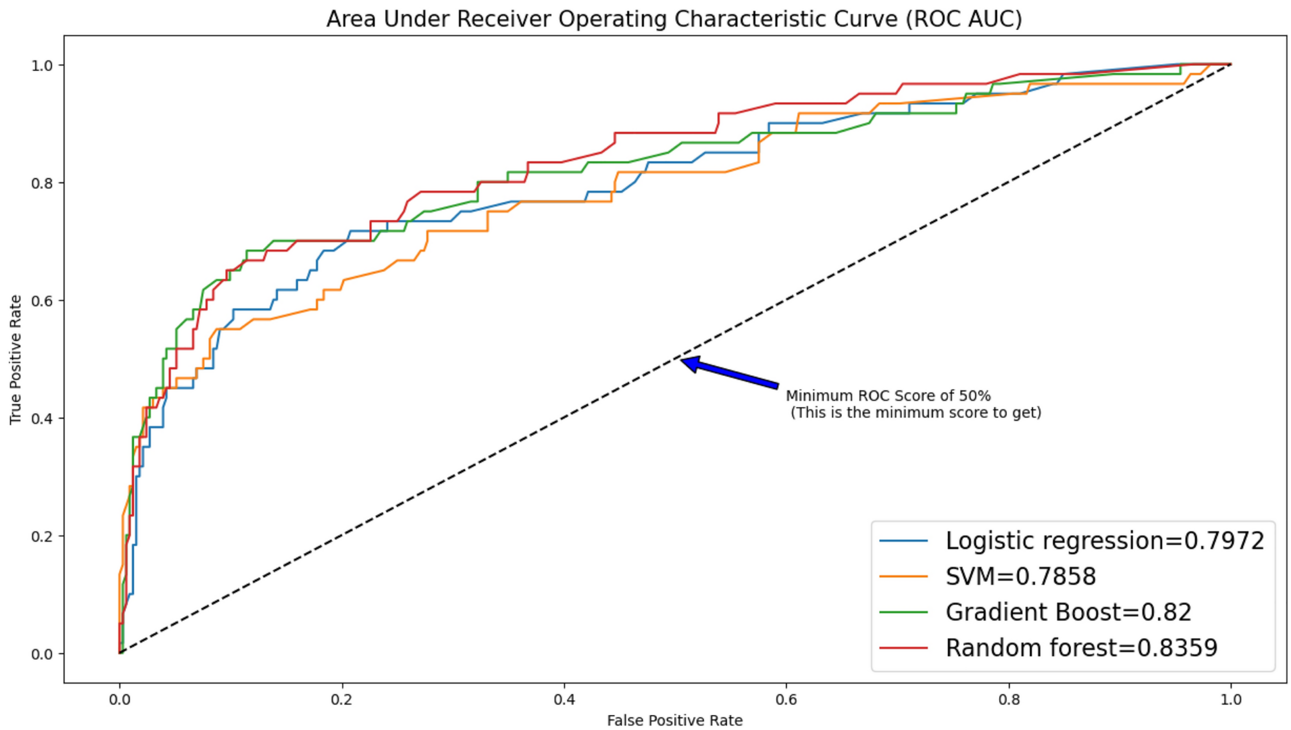
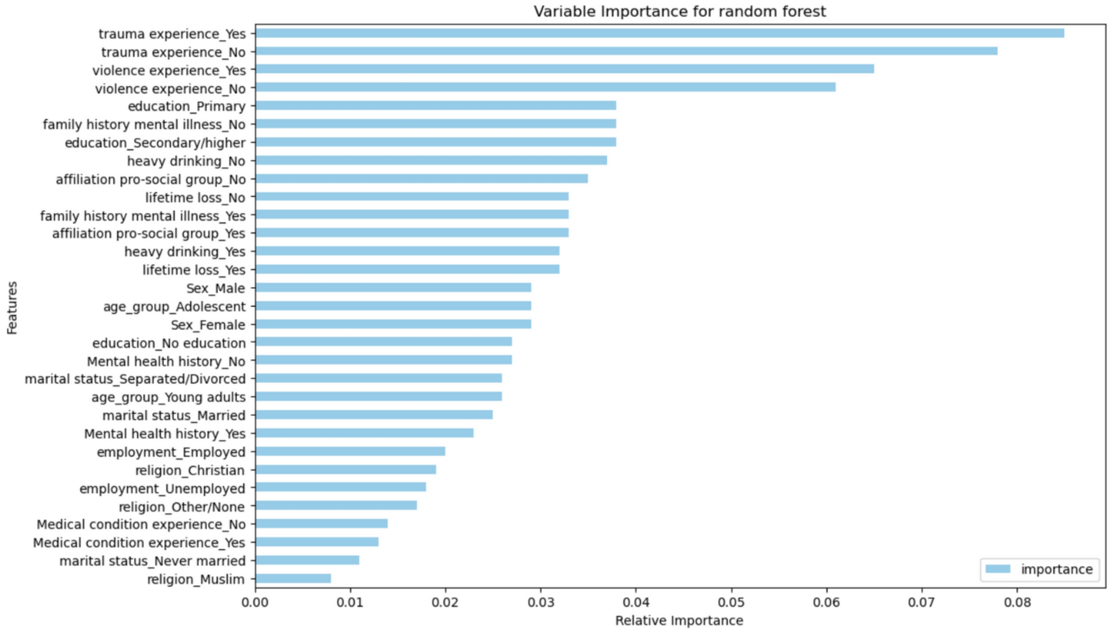


Fig. 3. Area under the curve (AUC).

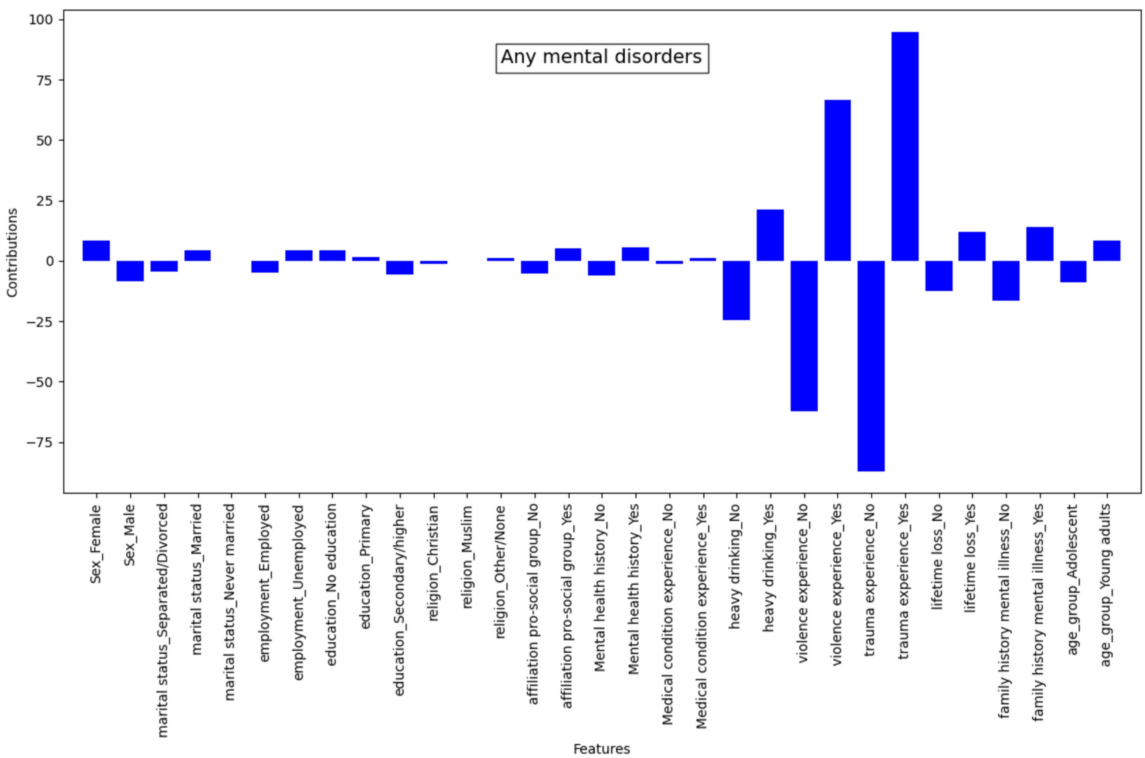
Random forest has the highest precision, recall, and accuracy, resulting in the best F1 score among the models. Logistic regression shows a high balance between precision and recall, leading to a solid F1 score; it also has high accuracy. Gradient boosting performs well with a slightly lower precision compared to other models. SVM shows the lowest F1 score and recall, resulting in low accuracy. Random forest has the best overall performance.

The performance of the model shown in Fig. 3 is measured using the ROC-AUC curve. As the graph indicates, the random forest classifier outperforms other models with an AUC of 83.59%. Random forest is quite good and has good discrimination capability, which means it is good at differentiating between those with no mental disorders and those screened with any mental disorders. Feature importance is defined as a method that assigns a score to independent variables based on their predictive utility for a target variable. The achieved scores illustrate which feature may be more important to the objective. The plot created for the feature importance is the output of the random forest classifier.

The Fig. 4 displays the feature importance and contributions as determined by the random forest classifier trained on the Rwanda mental health survey youth dataset. The Fig. 4a represents the feature importance where each bar represents a feature from the dataset and its relative importance in predicting those who are screened for at least one mental health disorder. The experiencing traumatic events feature has a high value of relative importance, indicating that trauma experience has a strong influence on the model's predictions. It is followed by violence experience, heavy drinking, and family history of mental health, as it is illustrated in Fig. 4a . Moreover, we evaluated the contributions of each feature on the target variable. The Fig. 4b highlights experiencing traumatic events and violence has high feature importance scores in predicting any mental health conditions. Other key predictors include heavy drinking, family history of mental illness, and a lifetime loss of a loved one or something significant. Additionally, being female and a young adult are associated with a high likelihood of mental disorders. Conversely, not experiencing these factors significantly reduces the likelihood of having mental health condition.



(a)



(b)

Fig. 4. Feature importance and their contributions.

Analysis of factors contributing to the comorbidity

Beyond binary classification, which identifies individuals as positive if they had at least one mental disorder, a multiclass classification approach using the Random Forest model, which outperformed others in the binary, was applied to predict and determine factors associated with comorbidity. Comorbidity was categorised into three levels namely: single mental disorder, two disorders, and multiple ($\geq$ 3) disorders.

The multiclass modeling using random forest demonstrated an overall solid performance in predicting mental disorder comorbidity among youth, with an accuracy of 75 % and an AUC score of 0.738, indicating a reasonable level of class separability. The precision (0.758) and recall (0.750) reveal that the model makes relatively accurate predictions and successfully identifies a significant proportion of true cases. However, the Fig. 5 and the F1-score point to a slight imbalance in performance across classes.

Figure 6 highlights the most influential variables in predicting the level of mental disorder comorbidity among youth. The top contributors include trauma experience, violence experience, affiliation to pro-social group and family history of mental disorders which all show relatively high importance score. These variables show the role of social support systems and adverse life experiences are critical in determining mental health outcomes.

Other notable predictor are lifetime loss, heavy drinking and mental health history, emphasizing the role of environmental and behavioral factors. Variables like being divorced/separated, employment status and religion exhibit relatively low importance, which means they may play a lesser role in distinguish levels of comorbidity in this specific dataset.

Discussion

The primary objective of this research was to employ machine learning approaches to predict mental health vulnerability and identify potential contributing factors among youth, using insights derived from cross-sectional survey data. The models were trained on a designated training set, and their performance was evaluated on unseen test data using several metrics, including classification accuracy, confusion matrix, recall, precision, and area under the curve (AUC).

The results revealed that the Random Forest algorithm outperformed other techniques, achieving the highest performance metrics with an accuracy of 88.8% and an AUC of 83.59% in predicting mental health vulnerability. It also showed strong performance in predicting comorbidity, with an accuracy of 75% and an AUC of 0.738. The most influential predictors of mental health vulnerability included traumatic experiences, exposure to violence, heavy drinking, family history of mental illness, lifetime loss, and being female or a young adult. In the context of mental health comorbidity, key contributing factors included traumatic experiences, exposure to violence, affiliation with pro-social groups, family history of mental disorders, experiences of lifetime loss, heavy drinking, and prior mental health conditions.

These findings align with existing literature, where Random Forest has consistently demonstrated high predictive power in mental health studies. For instance, a study focused on international students in the United Kingdom utilized Random Forest to predict depression, identifying important factors such as loan status,

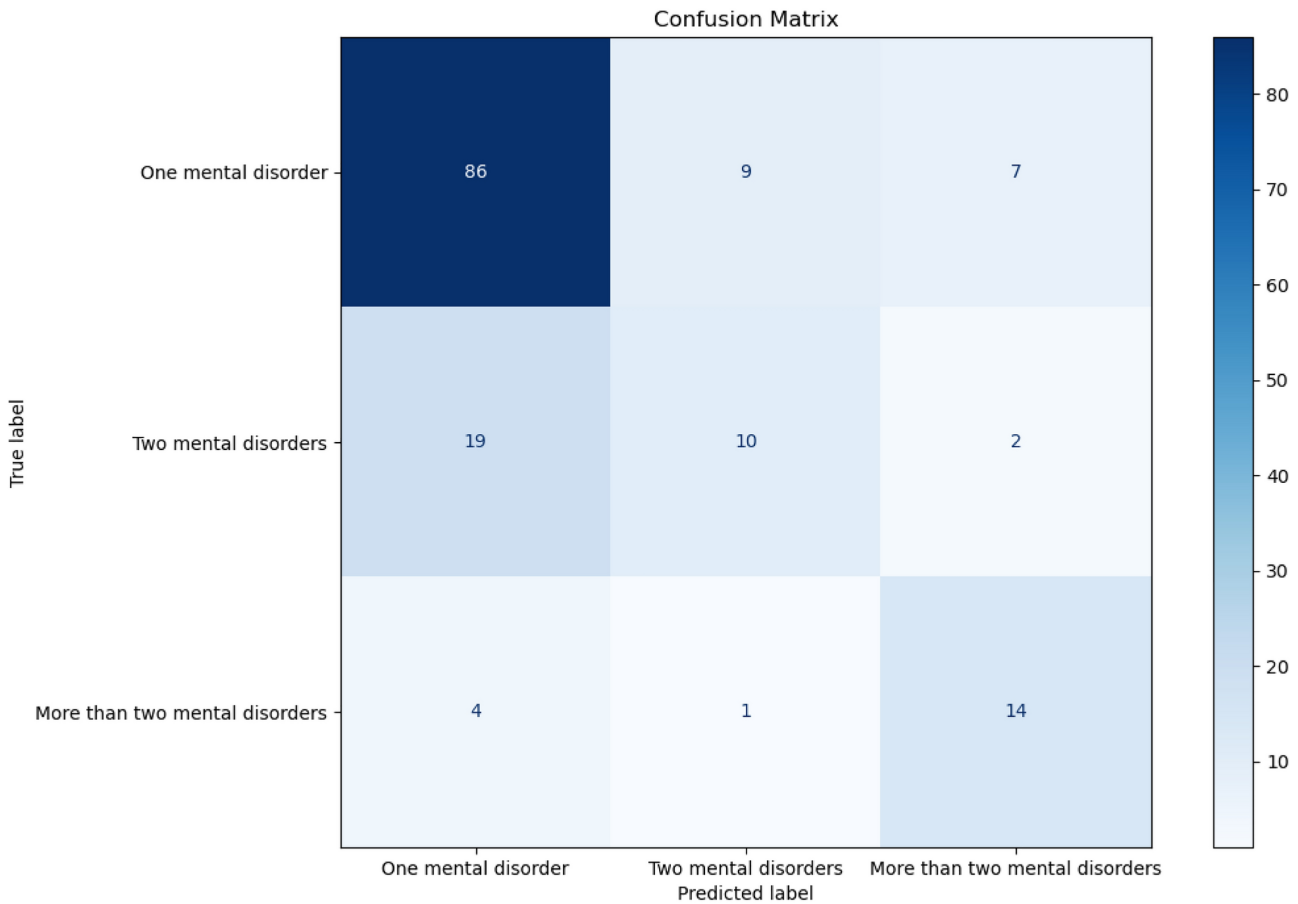


Fig. 5. Confusion matrix.

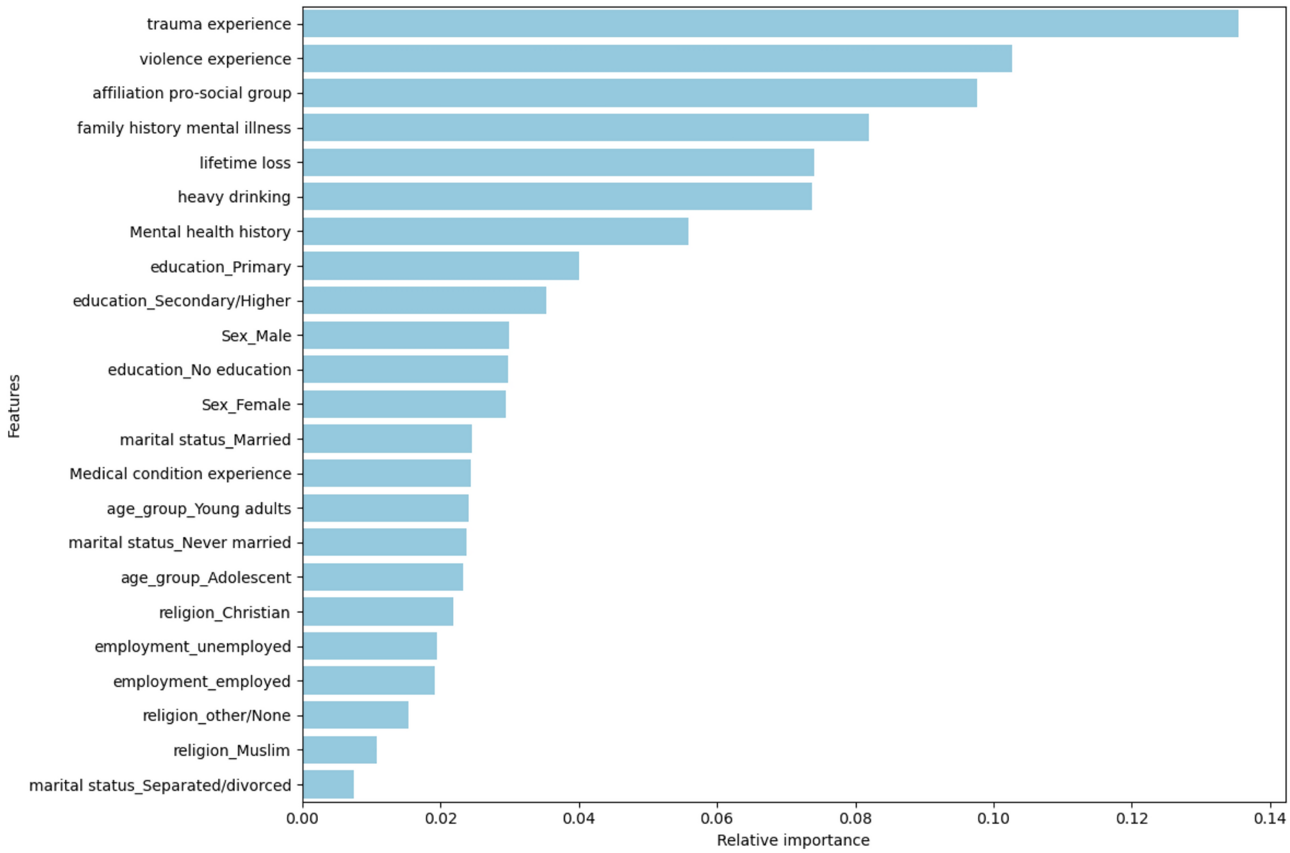


Fig. 6. Features importances.

gender, age, marital status, financial difficulties, academic stress, homesickness, and loneliness. The model achieved an accuracy of 80%, emphasizing the need for targeted interventions that address the multidimensional aspects of students' lives[54]. Similarly, another study evaluated the effectiveness of machine learning in predicting treatment outcomes in a digital mental health intervention targeting depression and anxiety. The Random Forest model achieved high accuracy and highlighted key predictors such as pre-treatment symptoms, self-reported motivation, referral type, and work productivity[55]. In Southeast Asia, research involving university students developed machine learning classifiers, with Random Forest attaining high accuracy in predicting mental well-being. Key predictors included body mass index, weekly physical activity, GPA, sedentary behavior, and age[56]. Furthermore, a study on bipolar disorder patients in Colombia demonstrated the potential of ML in predicting hospital admissions and re-admissions using electronic health records. The Random Forest model achieved an accuracy of 95.1% and an AUC of 98%, identifying psychiatric emergency visits, outpatient follow-up appointments, and age as primary predictors[57].

Feature importance analysis further confirmed that exposure to traumatic events and experiences of violence were the most significant contributors to screening positive for at least one mental health condition, as illustrated in Fig. 4a. These findings are consistent with existing literature. Traumatic experiences increase vulnerability to mental health disorders, often leading individuals to cope through harmful behaviors such as substance use or self-harm[58]. Likewise, individuals who have experienced violence are at a significantly higher risk of developing mental health conditions[59]. Therefore, the present findings support prior evidence indicating that violence and trauma are key drivers of mental health vulnerability.

Limitations

This study offers valuable insights into the use of machine learning for predicting mental health outcomes and serves as a significant resource for exploring factors associated with both mental health conditions and comorbidity among youth. However, several limitations and concerns regarding the generalizability of the findings must be acknowledged. Firstly, the study utilized cross-sectional data, capturing information at a single time point. This approach may not adequately reflect the temporal dynamics and variability of mental health conditions, thereby limiting the model's capacity to generalize across different real-world scenarios when applied in practice.

Secondly, the analysis was restricted to social and biological factors due to the nature of the dataset. Key psychological and cognitive variables; such as resilience, coping mechanisms, and self-esteem along with other influential factors like family functioning, stigma, discrimination, help-seeking behaviors, digital environment, and cyberbullying, were not included. The omission of these potentially significant variables may introduce selection bias and reduce the comprehensiveness and generalizability of the results.

Thirdly, the external validity of the model may be constrained. Its performance might not translate well to different populations or settings that vary in terms of demographics, resources, social support systems, behavioral norms, or care practices. Moreover, the model's temporal validity may be affected as health care practices, risk factor prevalence, and population characteristics can evolve over time, further influencing model applicability.

Lastly, although the multiclass classification approach employed in this study provides a more refined analysis than binary classification, it does not account for the exact combinations or co-occurrence patterns of specific mental disorders. Additionally, it does not incorporate the severity or intensity of the conditions, which are critical in understanding the full spectrum of mental health comorbidity. This limitation underscores the need for more detailed clinical data and longitudinal approaches in future research to better capture the complexity of mental health conditions.

Conclusion

This study represents an innovative effort to leverage machine learning techniques to predict mental health outcomes and identify potential risk factors associated with mental health comorbidity among youth, using data from the Rwanda Mental Health Cross-Sectional Survey. The study demonstrated the potential of machine learning in predicting mental health vulnerability. The findings indicate that the random forest model outperformed other models in identifying factors contributing to mental health vulnerability, achieving an accuracy of 87.8% and an AUC of 83.59%. Exposure to traumatic events, violence, heavy drinking, and a family history of mental health issues emerged as the most significant risk factors associated with the development of at least one mental disorder. The random forest model also showed solid performance in predicting mental health comorbidity, with an accuracy of 75% and an AUC of 0.738. However, the F1-score of 71.1% indicates a modest limitation in detecting comorbidity cases, particularly in the context of highly imbalanced data.

Key factors such as traumatic experiences, exposure to violence, affiliation with pro-social groups, family history of mental disorders, experiences of lifetime loss, heavy drinking, and prior mental health conditions were identified as significant predictors of mental health comorbidity. These findings are consistent with existing literature and further underscore the importance of addressing these risk factors in youth mental health interventions.

This study contributes to the existing body of knowledge by demonstrating the applicability of machine learning models in mental health research within the Rwandan context. Moreover, it highlights the importance of identifying and addressing high-risk factors in mental health interventions and policies. The findings emphasize the need for targeted support and tailored strategies for individuals who are at an increased risk of developing mental health conditions.

Data availability

The used and/or analysed datasets during this study are available from the corresponding author upon reasonable request.

Code availability

The source code used in this study is available at https://github.com/nfauste/Mental-health-prediction. The code repository includes all scripts necessary to reproduce the analyses and results presented in this work.

Received: 28 October 2024; Accepted: 29 April 2025
Published online: 08 May 2025

References

1. Khune, A. A., Rathod, H. K., Deshmukh, S. P. & Chede, S. B. Mental health, depressive disorder, and its management: A review. *GSC Biol. Pharm. Sci.* **25**(2), 1–13 (2023).
2. Collaborators, G. M. D. et al. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry.* **9**(2), 137–150 (2022).
3. Mudiyanselage, K. W. W. et al. The effectiveness of mental health interventions involving non-specialists and digital technology in low-and middle-income countries-a systematic review. *J. Psychosomatic Res.* **169**, 111320 (2023).
4. Purgato, M. et al. Promotion, prevention and treatment interventions for mental health in low-and middle-income countries through a task-shifting approach. *Epidemiol. Psychiatr. Sci.* **29**, e150 (2020).
5. Consortium WWMHS, *et al.* Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Jama.* **291**(21), 2581–2590 (2004).
6. Kovacevic, R. Mental health: Lessons learned in 2020 for 2021 and forward. *World Bank Blogs Retrieved November.* **7**, 2022 (2021).
7. Sankoh, O., Sevalie, S. & Weston, M. Mental health in Africa. *Lancet Glob. Health.* **6**(9), e954–e955 (2018).
8. McNab, S. E. et al. The silent burden: A landscape analysis of common perinatal mental disorders in low-and middle-income countries. *BMC Pregnancy Childbirth.* **22**(1), 342 (2022).
9. Kayiteshonga, Y., Sezibera, V., Mugabo, L. & Iyamuremye, J. D. Prevalence of mental disorders, associated co-morbidities, health care knowledge and service utilization in Rwanda-towards a blueprint for promoting mental health care services in low-and middle-income countries?. *BMC Public Health.* **22**(1), 1–13 (2022).
10. Patel, V. et al. The Lancet Commission on global mental health and sustainable development. *Lancet.* **392**(10157), 1553–1598 (2018).
11. Atzil-Slonim, D., Penedo, J. M. G. & Lutz, W. Leveraging novel technologies and artificial intelligence to advance practice-oriented research. *Administration Policy Mental Health Mental Health Services Res.* **51**(3), 306–317 (2024).
12. Kazdin, A.E. Nonprecsion (standard) psychosocial interventions for the treatment of mental disorders. *Int. J. Mental Health Promotion.* **24**(4), (2022)
13. Katiyar, K. AI-based predictive analytics for patients' psychological disorder. in *Predictive Analytics of Psychological Disorders in Healthcare: Data Analytics on Psychological Disorders*, 37–53. (Springer, 2022)

14. Srividya, M., Mohanavalli, S. & Bhalaji, N. Behavioral modeling for mental health using machine learning algorithms. *J. Med. Syst.* **42**, 1–12 (2018).
15. Shatte, A. B., Hutchinson, D. M. & Teague, S. J. Machine learning in mental health: A scoping review of methods and applications. *Psychol. Med.* **49**(9), 1426–1448 (2019).
16. Tutun, S. et al. An AI-based decision support system for predicting mental health disorders. *Inform. Syst. Front.* **25**(3), 1261–1276 (2023).
17. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science.* **349**(6245), 255–260 (2015).
18. Jiang, T., Gradus, J. L. & Rosellini, A. J. Supervised machine learning: A brief primer. *Behav, Therapy.* **51**(5), 675–687 (2020).
19. Tiffin, P. A. & Paton, L. W. Rise of the machines? Machine learning approaches and mental health: Opportunities and challenges. *Br. J. Psychiatry.* **213**(3), 509–510 (2018).
20. Cook, B. L., Progovac, A. M., Chen, P., Mullin, B., Hou, S., Baca-Garcia, E, et al. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Computational and mathematical methods in medicine. 2016;2016.
21. D'Alfonso, S. AI in mental health. *Curr. Opin. Psychol.* **36**, 112–117 (2020).
22. Jan, Z. et al. The role of machine learning in diagnosing bipolar disorder: Scoping review. *J. Med. Internet Res.* **23**(11), e29749 (2021).
23. Alanazi, S. A. et al. Publicâ€™s mental health monitoring via sentimental analysis of financial text using machine learning techniques. *Int. J. Environ. Res. Public Health.* **19**(15), 9695 (2022).
24. Samuelson, K. W. et al. Mental health and resilience during the coronavirus pandemic: A machine learning approach. *J. Clin. Psychol.* **78**(5), 821–846 (2022).
25. Van Mens, K. et al. Predicting future service use in Dutch mental healthcare: A machine learning approach. *Administration Policy Mental Health Mental Health Services Res.* **49**(1), 116–124 (2022).
26. Yu, J. et al. A machine learning approach to passively informed prediction of mental health risk in people with diabetes: Retrospective case-control analysis. *J. Med. Internet Res.* **23**(8), e27709 (2021).
27. Graham, S. et al. Artificial intelligence for mental health and mental illnesses: An overview. *Curr. Psychiatry Rep.* **21**, 1–18 (2019).
28. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **14**, 91–118 (2018).
29. Dipnall, J. F. et al. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS One.* **11**(2), e0148195 (2016).
30. Chung, J. & Teo, J. Mental health prediction using machine learning: Taxonomy, applications, and challenges. *Appl. Comput. Intell. Soft Comput.* **2022**, 1–19 (2022).
31. Sumathi, M., Poorna, B. Prediction of mental health problems among children using machine learning techniques. *Int. J. Adv. Comput. Sci. Appl.* **7**(1). (2016)
32. Dehghan, P., Alashwal, H. & Moustafa, A. A. Applications of machine learning to behavioral sciences focus on categorical data. *Discover Psychol.* **2**(1), 22 (2022).
33. Kim, J. et al. Machine learning for mental health in social media Bibliometric study. *J. Med. Internet Res.* **23**(3), e24870 (2021).
34. Nilsen, P. et al. Accelerating the impact of artificial intelligence in mental healthcare through implementation science. *Implement. Res. Practice.* **3**, 26334895221112030 (2022).
35. Ponce, E. K., Cruz, M.F., & Andrade-Arenas, L. Machine learning applied to prevention and mental health care in Peru. *Int. J. Adv. Comput. Sci. Appl.* **13**(1). (2022)
36. Abdul Rahman, H., Kwicklis, M., Ottom, M., Amornsriwatanakul, A., H Abdul-Mumin, K., Rosenberg, M, et al. Machine LearningBased prediction of mental WellBeing using health behavior data from University Students. *Bioengineering.* **10**(5), 575. (2023)
37. Clark-Kazak, C. R. Towards a working definition and application of social age in international development studies. *J. Develop. Stud.* **45**(8), 1307–1324 (2009).
38. Sheehan, D. V. et al. The MiniInternational Neuropsychiatric Interview MINI: The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry.* **59**(20), 22–33 (1998).
39. Betancourt, T. et al. Validating the center for epidemiological studies depression scale for children in Rwanda. *J. Am. Acad. Child Adolesc. Psychiatry.* **51**(12), 1284–1292 (2012).
40. Munyandamutsa, N., Mahoro Nkubamugisha, P., Gex-Fabry, M. & Eytan, A. Mental and physical health in Rwanda 14 years after the genocide. *Social Psychiatry Psychiatric Epidemiol.* **47**, 1753–1761 (2012).
41. Muwonge, J., Umubyeyi, A., Rugema, L. & Krantz, G. Suicidal behaviour and clinical correlates in young adults in Rwanda: A population-based, cross-sectional study. *J. Glob. Health Rep.* **3**, e2019080 (2019).
42. Engel, G. L. The clinical application of the biopsychosocial model. *J. Med. Philos.* **6**(2), 101–124 (1981).
43. Maharana, K., Mondal, S. & Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transitions Proc.* **3**(1), 91–99 (2022).
44. Azencott, C. A. Introduction au Machine Learning-2e éd. Dunod; (2022).
45. Agresti, A. *Categorical Data Analysis*, Vol. 792. (Wiley, 2012).
46. Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., & Friedman, J. Unsupervised learning. The elements of statistical learning: Data mining, inference, and prediction. p. 485–585, (2009)
47. Biau, G. & Scornet, E. A random forest guided tour. *Test.* **25**, 197–227 (2016).
48. Breiman, L. Random forests. *Machine learning.* **45**, 5–32 (2001).
49. Vapnik, V. *The Nature of Statistical Learning Theory*. (Springer2Verlag, 1995)
50. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **54**, 1937–1967 (2021).
51. Chen, T., & Guestrin, C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; p. 785–794. (2016)
52. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **17**(1), 168–192 (2020).
53. Wardhani, N. W. S. et al. international conference on computer, control, informatics and its applications (IC3INA). *IEEE.* **2019**, 14–18 (2019).
54. Rahman, M. A. & Kohli, T. Mental health analysis of international students using machine learning techniques. *Plos One.* **19**(6), e0304132 (2024).
55. Hornstein, S., Forman-Hoffman, V., Nazander, A., Ranta, K. & Hilbert, K. Predicting therapy outcome in a digital mental health intervention for depression and anxiety A machine learning approach. *Digital Health.* **7**, 20552076211060660 (2021).
56. Rahman, H. A., Kwicklis, M., Ottom, M., Amornsriwatanakul, A., AbdulMumin, K. H., Rosenberg, M., et al. Prediction Modeling of Mental Well-Being Using Health Behavior Data of College Students. Research Square. (2022)
57. Palacios-Ariza, M. A. et al. Prediction of patient admission and readmission in adults from a Colombian cohort with bipolar disorder using artificial intelligence. *Front. Psychiatry.* **14**, 1266548 (2023).
58. Butler, L. D., Critelli, F. M. & Rinfrette, E. S. Trauma-informed care and mental health. *Directions Psychiatry.* **31**(3), 197–212 (2011).
59. Patel, D. Violence and Mental Health: Opportunities for Prevention and Early Detection: Proceedings of a Workshop. National Academies Press (2018)

Acknowledgements
We gratefully acknowledge the contribution of the Rwanda Biomedical Center in Rwanda for providing access to data used in this study. We also extend our sincere appreciation to the University of Rwanda's Centre for mental health and African Centre of Excellence in Data Science for providing the required resources to execute this project.

Author contributions
This study was conceptualized and designed by FN, VS and CT. FN, VS, CT, JK and JI made contributions to the acquisition of the dataset. FN carried out data analysis, VS, CT, JI, JK, MN, SN contributed to the review of data analysis and the interpretation of the results. MN, IK, MU, SHM, CEN and CT provided critical feedback and helped in revising the manuscript. The overall project was supervised by VS. All authors reviewed and approved the final manuscript.

Funding
We gratefully acknowledge the funding provided by Research training in Data Science for Health in Rwanda under NIH Grant: U2RTW012122-04 in collaboration with the University of Rwanda, The Regional Centre of Excellence in Biomedical Engineering and EHealth, African Institute for Mathematical Sciences and Washington University in Saint Louis. The funder had no involvement in the study design, data acquisition and analysis, decision to publish, or the preparation of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
Prior to starting the survey, ethical approval was acquired from the Rwanda National Ethics Committee (RNEC) under reference number 0061/RNEC/2018, dated February 15, 2018 and every participant granted informed consent by signing an informed consent form before participating in the survey. All methods were performed in accordance with the relevant guidelines and regulations.

Additional information
Correspondence and requests for materials should be addressed to F.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025