Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 9 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i09.47581

PAPER

Prediction of Depression Severity and Personalised Risk Factors Using Machine Learning on Multimodal Data

Adefemi Ayodele[1], Adedotun Adetunla[2,3](✉), Esther Akinlabi[4]

[1]Department of Big Data Technologies, University of East London, London, United Kingdom

[2]Department of Mechanical Engineering, University of Johannesburg, Johannesburg, South Africa

[3]Department of Mechanical & Mechatronics Engineering, Afe Babalola University, Ado, Nigeria

[4]Department of Mechanical and Construction Engineering, Northumbria University, Newcastle, United Kingdom

adetunla.adedotun@abuad.edu.ng

ABSTRACT

Depression, a prevalent global mental health disorder, often leads to a reduced quality of life and an increased risk of suicide. Despite the availability of treatments, many cases go undetected, highlighting the need for an accurate machine learning (ML) prediction model for depression severity and risk factors, particularly when dealing with multimodal datasets. Previous studies that utilized ML to predict the severity of depression encountered limitations, such as small datasets and a lack of personalization. This study proposes an optimal algorithm for predicting depression severity and personalized risk factors using ML. The potential benefits include improved accuracy in severity assessment, personalized treatment strategies, and refined risk factor identification. The random forest (RF) algorithm emerged as the most effective, exhibiting notable performance metrics on NHANES data, including a 0.93 R-squared, 0.93 explained variance score (EVS), 0.51 mean absolute error (MAE), 1.73 mean squared error (MSE), and 1.32 root mean squared error (RMSE). Notably, RF identified both general and personalized risk factors for depression severity. This model holds promise for clinical assessment, diagnosis, and intervention planning, contributing significantly to the comprehensive management of depression.

KEYWORDS

depression severity, machine learning (ML), multimodal dataset, artificial intelligence (AI), random forest (RF)

1 INTRODUCTION

Depression is a global health crisis [1]. Five of the 10 most common illnesses that globally disable people are mental illnesses, with depression ranking as the number one. Depression is a prevalent psychological condition that impacts individuals of all ages, genders, cultures, and backgrounds. Depression devastates an individual's quality of life at work, school, and home [2]. Depression results from an interplay of social, psychological, and physiological factors. According to World Health Organization (WHO) statistics, the global population suffering from depression is estimated at 3.8%,

Ayodele, A., Adetunla, A., Akinlabi, E. (2024). Prediction of Depression Severity and Personalised Risk Factors Using Machine Learning on Multimodal Data. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(9), pp. 130–143. https://doi.org/10.3991/ijoe.v20i09.47581

Article submitted 2023-12-26. Revision uploaded 2024-03-07. Final acceptance 2024-03-18.

totaling approximately 280 million individuals [3]. Early identification and assessment of depressive symptoms, along with appropriate evaluation and therapy, can significantly improve the chances of managing symptoms and the underlying disease, as well as safeguard personal, economic, and social well-being [4]. Clinical interviews and self-report tools, including the Patient Health Questionnaire-9, assess the severity of depression [5]. However, these methods can take considerable time and be subject to human error, potentially resulting in misdiagnosis or inappropriate treatment.

Machine learning (ML) techniques, which are one of the distinct areas under artificial intelligence (AI), have demonstrated great promise for improving depression diagnosis and predicting severity accurately. ML algorithms integrate various data modalities, such as neuroimaging, genetics, and behavioral data, to uncover patterns and relationships not easily seen through traditional clinical assessments [6]. Furthermore, personalized risk factor identification can help create customized treatment plans, leading to improved treatment outcomes. Numerous studies have explored the application of ML algorithms for predicting depression severity [7], [8]. A study [9] compared ML algorithms for classifying depression among senior citizens aged 60 years and older. Another study [10] developed a model to analyze textual electronic health records to predict depression diagnosis and response to treatment. [11] investigated the association between a set of biomarkers and self-reported depression using XGBoost on a Dutch dataset of 11,081 cases to enhance the diagnosis of depression. They applied various sampling techniques, such as under-sampling, over-sampling, over-under sampling, and ROSE sampling, to balance the imbalanced dataset before classification. They acknowledged that their XGBoost model had limited generalizability to other populations and contexts. Human behavioral patterns were evaluated by some studies to gain insight into possible reasons for depression [12], [13]. These studies relied on a sizeable text dataset from the public. However, the variety of its dataset limited the study, and a more comprehensive and diverse dataset is necessary. Similarly, a study was conducted by [14] to distinguish between depressed and non-depressed participants. The study involved collecting smartphone usage data, emotion elicitation data, and speech data from 102 volunteers aged 18–19 through social networks. They found that deploying features from various data modalities outperformed using a single modality, even on a benchmark dataset. The proposed approach achieved an accuracy of 86% using a support vector machine (SVM) classifier. The study's main limitation was the restricted dataset with a limited number of participants. Using large-scale, diverse datasets with clinically validated cases of depression is recommended for further exploration. A study conducted in 2022 [15] proposed a depression detection model based on quality of life scales and multimodal health and nutritional survey data. They employed an ensemble model based on the average weight of the four base algorithms (DT, ANN, SVM, and KNN) to classify depressive and non-depressive cases using the comprehensive NHANES dataset. Results indicated that the ensemble classifier model outperformed the baseline algorithms; however, mental health features that convey the symptoms of depression were not used in isolation from the prediction model. The work did not personalize risk factors; therefore, the author's recommendations were not tailored to individual circumstances.

This study investigated the prediction of depression severity and the identification of personalized risk factors using a multimodal dataset. While previous research has explored the use of ML for depression diagnosis and prediction, there is little work on utilizing multimodal data sources for personalized risk factor identification. This research aims to address this gap by creating a model that integrates various data sources, including demographic information, medical history, dietary

habits, physical activity, environmental exposures, and health outcomes, to forecast the severity of depression and pinpoint individualized risk factors.

2 METHODOLOGY

This section outlines the study's design and methods for predicting depression severity and identifying personalized risk factors using multimodal data. It describes the data collection process, statistical analyses, research approach, and data acquisition and analysis techniques. Figure 1 shows the architecture of the developed model.

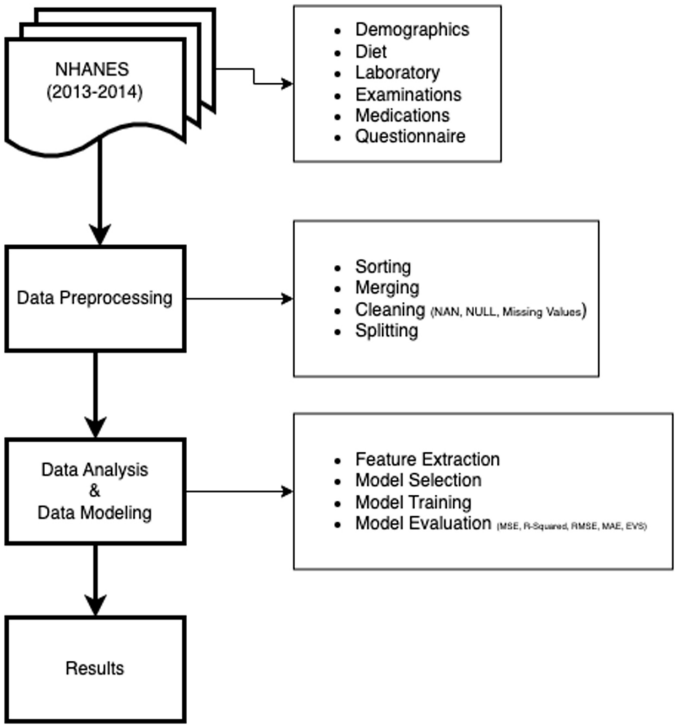

Fig. 1. Architecture of the developed model

2.1 Data collection

The National Health and Nutrition Examination Surveys (NHANES) 2013–2014 edition dataset was obtained by downloading it from Kaggle, a community for data science and ML that provides a range of tools and services for exploring, analyzing, and sharing data. The dataset was used to predict depression severity and personalized risk factors using ML algorithms. According to the Centers for Disease Control and Prevention (CDC) [16], "NHANES is a program of studies that assesses the health and nutritional states. Of adults and children in the United States, conducted by the National Center for Health Statistics, a division of the CDC. It collects data on health and nutrition through a series of health surveys conducted periodically since the early 1960s.

In the NHANES 2013–2014 edition, 14,332 individuals were sampled across 30 different survey locations in the United States. Of these, 10,175 completed the interview, and 9,813 underwent examination. NHANES uses interviews and physical examinations to collect various health and nutrition measurements from different modalities

obtained from multiple sources. The interview component included demographic, socioeconomic, dietary, and health-related questions. The examination component involved medical, dental, and physiological measurements, as well as laboratory tests administered by specialist medical personnel. This provided various forms of data, such as demographics from diverse racial backgrounds, dietary information, examination results, laboratory findings, and questionnaire responses. The study covers a wide range of areas, including health status, nutrition, risk behaviors, and environmental exposures [17]. The dataset provides a comprehensive and holistic multimodal source for developing a ML model that can predict the severity of depression and personalized risk factors.

2.2 Data processing

The collected multimodal NHANES dataset for predicting depression severity and personalized risk factors using ML was preprocessed before analysis and modeling to ensure the quality, consistency, and usability of the data. Six sub-datasets—demographics, diet, examination, laboratory, medication, and questionnaire—were sorted and merged into one dataset using the participants' unique ID, named SEQN, as the common key. The resulting dataset had 19,580 rows and 1,824 columns. Using the responses to the PHQ-9 questionnaire that each participant completed and that are contained in the dataset, the PHQ-9 score was calculated to determine the severity of depression in the dataset. The PHQ-9 score is a widely used measure of depression severity that ranges from 0 to 27, with higher scores indicating more severe depression [18]. The PHQ-9 score was added as a new column to the dataset and utilized as the target variable for predicting depression severity. The variables used to calculate the PHQ-9 score were removed from the dataset to present multicollinearity and data leakage. Additionally, rows with missing or declined information from the questionnaire were also excluded from the dataset.

The dataset was cleaned up by removing irrelevant, redundant, or erroneous rows. The missing values were replaced by imputing the median, or mean, because the variables are continuous and have a symmetric distribution. The final dataset had 19,560 rows and 1,768 columns. The numerical variables were scaled to a standard range of values between 0 and 1. This normalization process aimed to reduce the impact of outliers and account for variations in units of measurement within the dataset. Normalization could enhance the performance and convergence of certain ML algorithms [19]. The data was split into training and test sets using a stratified random sampling method. The stratification was based on the PHQ-9 score to ensure a balanced representation of different levels of depression severity in both groups. The training set contained 80% of the data, with 15,648 rows and 1,768 columns, while the test set contained 20% of the data, with 3,912 rows and 1,768 columns. The splitting was done at this stage to avoid data leakage during the feature extraction phase. Data leakage occurs when information from the test set is used or revealed in the training set, which can result in overfitting and inaccurate outcomes [19].

2.3 Feature selection

To reduce the dimensionality and complexity of the NHANES dataset and enhance the performance and interpretability of the depression severity prediction algorithms, SelectKBest, a feature selection class from the scikit-learn library, was utilized, following the approach of Zulfiker et al. (2021) [20]. According to a scoring

function, SelectKBest selects the k features with the highest scores and ranks them in order of importance for the target variable. The score function takes two arrays, X and Y, as input. Here, X represents the feature matrix, and y represents the target vector. The function returns either a pair of arrays (scores, pvalues) or a single array with scores [21].

A score function was applied to each feature to obtain the K-highest scores used for selecting the features. By calculating the statistical significance of the correlation between each feature and the target variable using the F-test, the f_regression score function assigned a score to each feature. The F-test measures how well a model fits the data by comparing the variance explained by the model with the unexplained variance. A higher F-score indicates a stronger correlation and a better fit [22]. The number of features, k, was varied from 50 to 300 in increments of 50, and the performance of each k value was evaluated using five-fold cross-validation. For each fold, the training data was split into training and test sets. The SelectKBest was fitted on the train set, and both sets were transformed using the selected features. The RandomForestRegressor was then fitted as a selector on the train set, and predictions were made on the test set. The R2 score for each fold was computed and averaged over five folds to obtain the score for each k value. k = 300 yielded the highest score, indicating that 300 features with the highest F-scores were optimal for our dataset. Thus, the predictive model used 300 features with the highest F-scores.

2.4 Prediction algorithms and evaluation metrics

The depression severity and personalized risk factors prediction framework using ML is presented in Figure 2. The output of the feature selection process was scaled using MinMaxScaler. This process converted the values of each feature to a range, all between 0 and 1, by subtracting the minimum value and dividing by the range without distorting the differences between the values and shape of the original distribution. The scaling of the features made them comparable and prevented some from dominating others due to their large scale. The ML algorithms took the scaled features as input. Shown in Equation 1 is the scaling process [23].

$$\text{Xscaled} = \frac{\text{X} - \text{Xmin}}{\text{Xmax} - \text{Xmin}} \tag{1}$$

Where X_{min} = minimum value in X Feature, X_{max} = maximum value in X Feature.

Five ML algorithms with a literature record of accuracy, interpretability, complexity, and computational efficiency in handling regression tasks, including linear regression (LR), random forest (RF), SVM, extreme gradient boosting (XGB), and least absolute shrinkage and selection operator (LASSO), were evaluated for their performance in predicting depression severity using the preprocessed NHANES dataset. LR is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables, assuming a linear association between them. XGBoost, an advanced implementation of gradient boosting algorithms, integrates multiple weak learners to create a robust predictive model. RF is a versatile ensemble learning technique that constructs numerous decision trees and averages their predictions, making it suitable for both classification and regression tasks. SVM is a supervised learning algorithm that identifies the optimal hyperplane to separate data points into different classes by predicting the target variable according to the distance from this hyperplane. LASSO serves as an

analytical method that performs variable selection and regularization to enhance prediction accuracy and model interpretability by penalizing excessive coefficients within the statistical model. The best-performing Ml model would be utilized to generate personalized risk factors for each participant in the dataset. The importance of the feature importance would be determined to identify the most critical variables in predicting depression severity and personalized risk factors, potentially aiding in improving the understanding and management of depression.

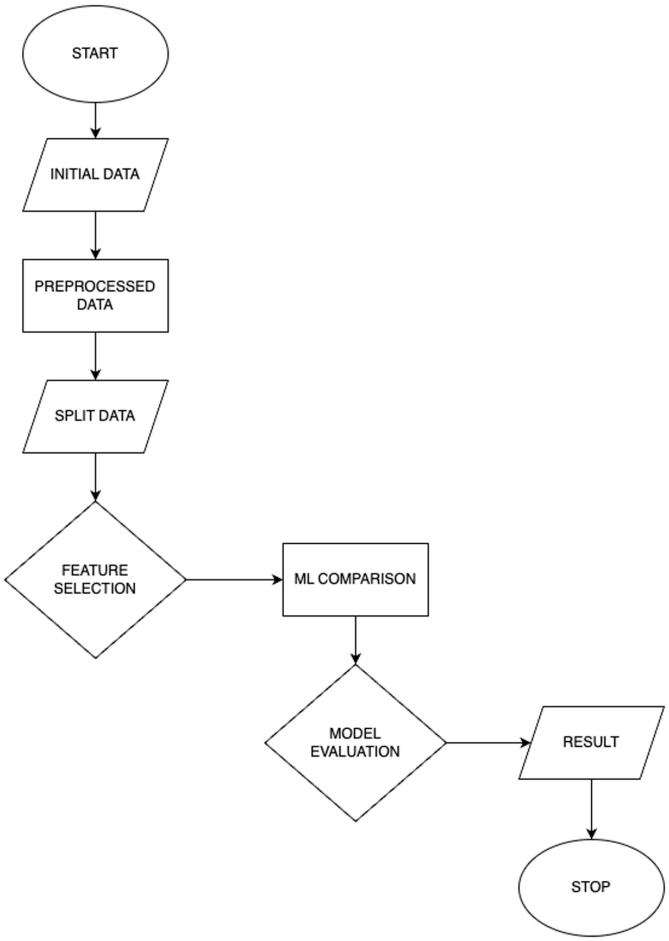

Fig. 2. Depression severity and personalized risk factors prediction framework

2.5 Personalized risk factors

Personalized risk factors are the most significant contributors to the severity of depression in each patient. The features are determined based on feature importance, which refers to the input variables that contribute most significantly to predicting depression severity in the multimodal NHANES dataset using the best-performing Ml model considered in this study. The feature importance is obtained using the "feature_importances_" attribute on the best-performing model and sorted in descending order using the NumPy function "argsort()" to identify personalized risk factors for patients in the test dataset. The important features identified in the test dataset are selected and used to calculate the corresponding feature values for each patient. Then, the features with non-zero importance scores are selected and

displayed in order of importance, helping to identify the most significant risk factors for that patient. As a result, a for loop is used to iterate over the index and patient ID of the first n patients in the test dataset. After selecting the important features using array slicing, the feature values for each patient are computed using the "iloc" attribute of the input data. Finally, the features with non-zero importance scores are selected using Boolean indexing and output in the patient's specific order of importance using the "print ()" function.

3 RESULT AND DISCUSSION

3.1 Implementation of the prediction model

Google Colaboratory Pro was used to implement the study using Python programming. This cloud-based notebook environment supports various ML tasks with enhanced system configuration capabilities. Google Colaboratory Pro provides a Google Compute Engine backend with GPU support, 25.5 GB of system RAM, and 166.8 GB of disk space. Google Colaboratory Pro was accessed using the Microsoft Edge web browser on an Apple MacBook Air M1 with 8 GB of RAM (2560 x 1600) and macOS Ventura 13.2.1, eliminating the need to install high-demand computational requirements on the local machine. This configuration facilitated the efficient and robust processing of the ML algorithms utilized in the study. The seamless and user-friendly integration of the MacBook Air M1 with Google Colaboratory Pro enabled easy implementation of study tasks. Shown in Figure 3 is the graphical user interface of Google Colab Pro.

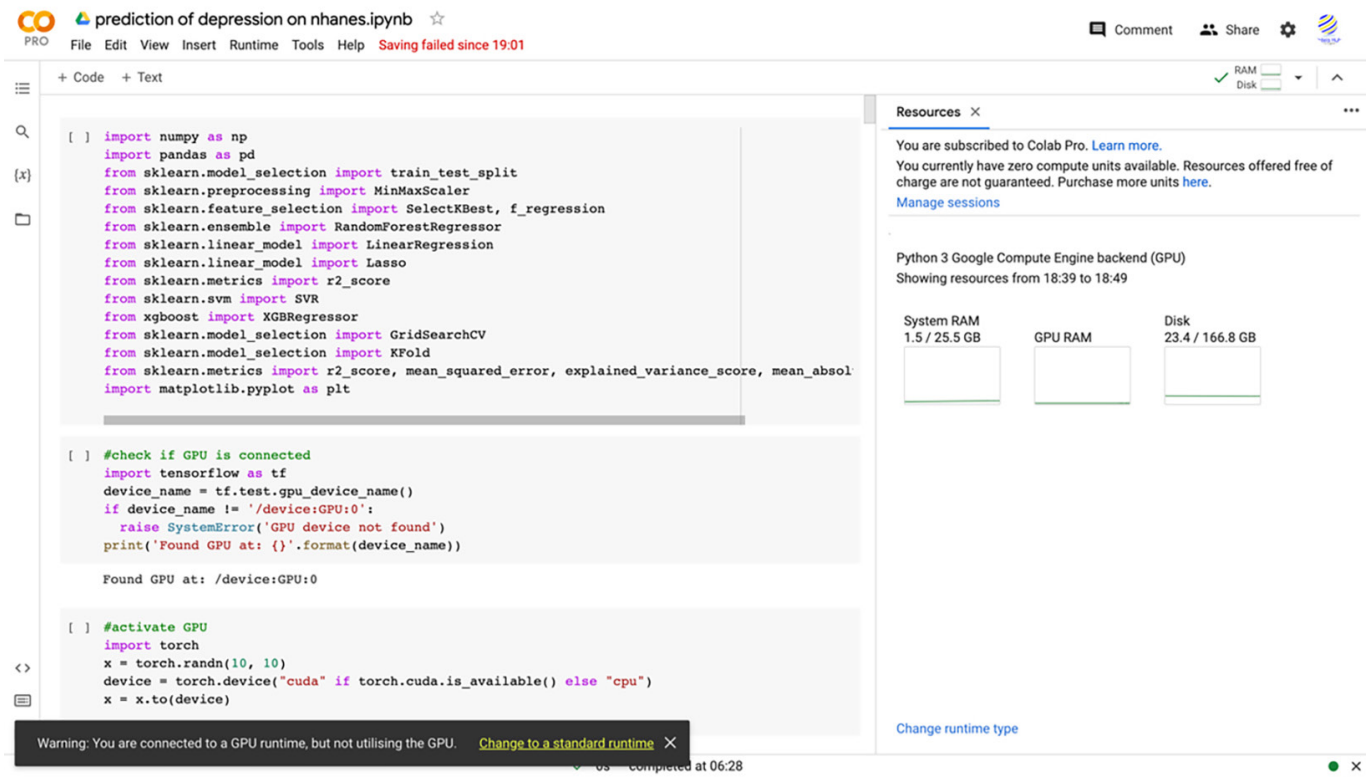

Fig. 3. Graphics user interface of google collaboratory uses the IEEE citation style

3.2 Performance evaluation of the ML algorithm

The five ML algorithms, namely LR, SVM, LASSO, XGBoost, and RF, were trained and tested to predict depression severity using the NHANES multimodal dataset in an 80:20 ratio. The performance of the ML algorithms was evaluated using evaluation metrics such as mean squared error (MSE), R-squared, root mean squared error (RMSE), mean absolute error (MAE), and explained variance score (EVS). R-squared is the coefficient of determination that indicates how well the predictive model explains the variation in the dependent variables. MAE measures the average absolute deviation between predicted and observed values. MSE calculates the average squared deviation between predicted and observed values. RMSE is a measure of the average deviation of the predicted values from the observed values, and EVS measures how well the model captures variation in the observed data. Table 1 presents the performance evaluation results of the ML algorithms: LR, SVM, LASSO, XGBoost, and RF. Upon evaluation using the specified metrics, the RF model exhibited the lowest MSE of 1.73, the lowest RMSE of 1.32, the highest R-squared value of 0.93, the smallest MAE of 0.51, and the highest EVS of 0.93. This RF outperformed the other ML algorithms considered, as illustrated in Figure 4. The RF model had the lowest MSE and RMSE, indicating the smallest average squared difference between the predicted and target values. As a result, it accurately estimates the output variables from the input features on the same scale as the target value with minimal error. Additionally, the RF model had the highest R-squared value, indicating that the input features accounted for a larger portion of the variance in the target value. The highest R-squared value indicates that the model had high explanatory power and explained most of the variation in the data.

Table 1. Evaluation result for depression severity prediction model

	LR	SVM	LASSO	XGBOOST	RF
MSE	102.03	11.70	1440.0	2.28	1.73
R-Squared	–44.10	0.52	–58.48	0.91	0.93
RMSE	33.05	3.42	37.95	1.51	1.32
MAE	3.34	1.87	2.95	0.75	0.51
EVS	–44.09	0.54	–58.46	0.91	0.93

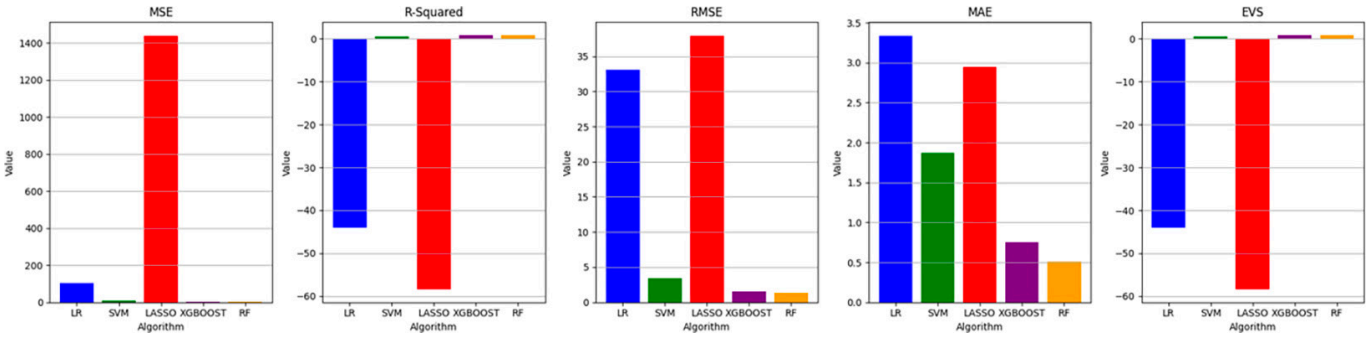

Fig. 4. Performance evaluation of depression severity of machine learning

3.3 Identifying depression severity risk factors

Algorithms cite

The best-performing depression severity prediction model, the RF model, was used to identify 50 important features that are considered general risk factors. These features are presented in Figure 5, with matching variable names in Figure 6. These generalized risk factors are contributing factors that can inform the development of preventive and early intervention strategies. In addition, the RF model recorded the lowest MAE, indicating that the model has a minor average absolute difference between the predicted and target values. This demonstrates low error and the ability to precisely estimate the output variables from the input features without being influenced by outliers. With the highest EVS, which represents the ratio of variance in the target value explained by the predicted value, the RF model demonstrated a high level of consistency and the ability to make accurate predictions close to the target value. Negative values of the R-squared and EVS imply that both the LR and LASSO regression models have not effectively captured the underlying relationships between independent and dependent variables. Such negative values indicate a significant prediction error, with the models capturing noise or random fluctuations in the training data rather than discernible patterns that can be generalized.

```
Index(['DR1TP204', 'MGXH1T2E', 'DRD370EQ', 'BMXBMI', 'BMDAVSAD', 'MGQ100',
       'RIDSTATR', 'BMXHEAD', 'DMDHSEDU', 'DRD350F', 'DMDFMSIZ', 'MIALANG',
       'FIAPROXY', 'BPXDI2', 'OHDEXSTS', 'OHX06TC', 'DR1TFF', 'DRD370NQ',
       'WTMEC2YR', 'DR1TSUGR', 'FIAINTRP', 'DR1TVK', 'BPXDI4', 'DMQADFC',
       'INDHHIN2', 'DR1TCRYP', 'DR1TFA', 'DR1TFDFE', 'DR1TM201', 'DR1TCARB',
       'MIAINTRP', 'DR1TIRON', 'DR1TCHL', 'DRD370OQ', 'DRD350FQ', 'DR1TFOLA',
       'WTINT2YR', 'DRD370BQ', 'DR1TWS', 'SIALANG', 'DR1TB12A', 'OHX12TC',
       'DR1TLZ', 'MIAPROXY', 'MGQ110', 'DR1TVC', 'DMDHREDU', 'DMDHHSZE',
       'BMIARML', 'DMDHRMAR'],
      dtype='object')
```

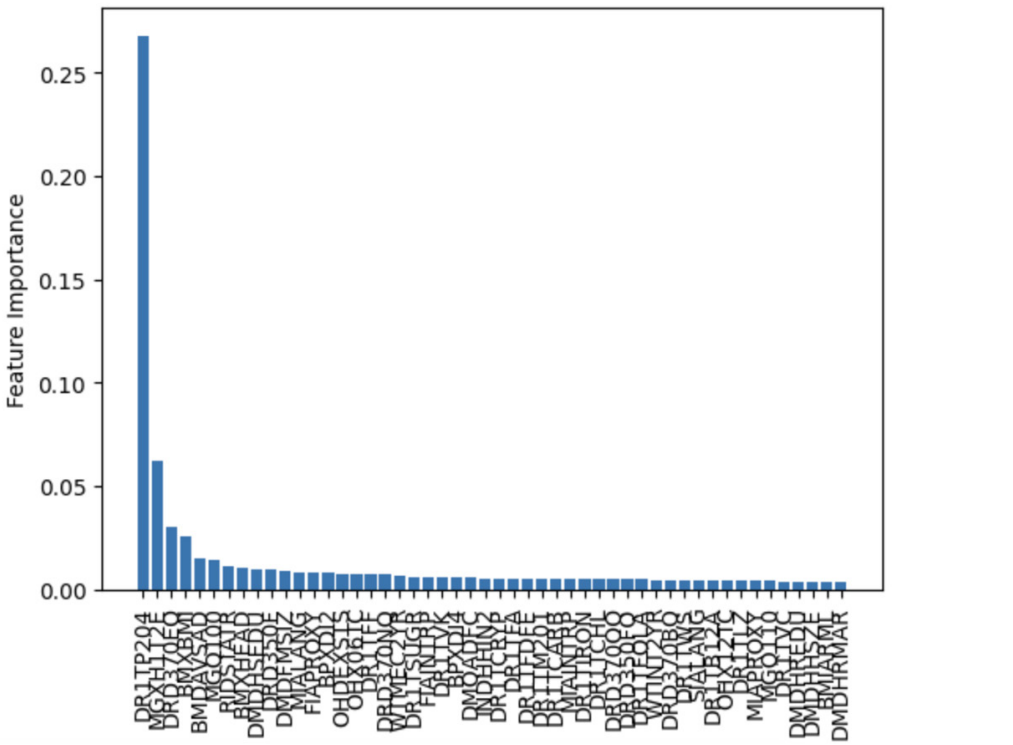

Fig. 5. Identified depression severity risk factors

Variable	Variable Names
DR1TP204	PFA 4:20 (Eicosatetraenoic) (gm)
MGXH1T2E	Whether the participant exerted a maximal or questionable effort during the test 2 on hand 1, as assessed by the technician.
DRD370EQ	Number of times cod was eaten in the past 30 days
BMXBMI	Body Mass Index (kg/m**2)
BMDAVSAD	Average Sagittal Abdominal Diameter (cm)
MGQ100	Have you had any pain, aching or stiffness in your left hand in the past 7 days?
RIDSTATR	Interview and Examination Status of the Sample Person.
BMXHEAD	Head Circumference (cm)
DMDHSEDU	What is the highest grade or level of school {you have/NON-SP SPOUSE has completed or the highest degree {you have/he/she has} received
DRD350F	Mussels eaten during past 30 days
DMDFMSIZ	Total number of people in the Family
MIALANG	Language of the CAPI MEC Interview Instrument
FIAPROXY	Was a Proxy respondent used in conducting the Family Interview?
BPXDI2	Diastolic: Blood pressure (second reading) mm Hg
OHDEXSTS	Overall Oral Health Exam Status
OHX06TC	Tooth Count: Upper right cuspid ©
DR1TFF	Food folate (mcg)
DRD370NQ	# of times sardines eaten past 30 days
WTMEC2YR	Full sample 2 year MEC exam weight
DR1TSUGR	Total sugars (gm)
FIAINTRP	Was an interpreter used to conduct the Family interview?
DR1TVK	Vitamin K (mcg)
BPXDI4	Diastolic: Blood pressure (fourth reading if necessary) mm Hg
DMQADFC	Did {you/SP} ever serve in a foreign country during a time of armed conflict or on a humanitarian or peace-keeping mission? (This would include National Guard or reserve or active duty monitoring or conducting peace keeping operations in Bosnia and Kosovo, in the Sinai between Egypt and Israel, or in response to the 2004 tsunami or Haiti in 2010.)
INDHHIN2	Total household income (reported as a range value in dollars)
DR1TCRYP	Beta-cryptoxanthin (mcg)
DR1TFA	Folic acid (mcg)
DR1TFDFE	Folate, DFE (mcg)
DR1TM201	MFA 20:1 (Eicosenoic) (gm)
DR1TCARB	Carbohydrate (gm)
MIAINTRP	Was an interpreter used to conduct the MEC CAPI interview?
DR1TIRON	Iron (mg)
DR1TCHL	Total choline (mg)
DRD370OQ	# of times sea bass eaten past 30 days
DRD350FQ	# of times oysters eaten in past 30 days
DR1TFOLA	Total folate (mcg)
WTINT2YR	Interviewed participants
DRD370BQ	# of times tuna eaten in past 30 days
DR1TWS	Tap water source
SIALANG	Language of the Sample Person Interview Instrument
DR1TB12A	Added vitamin B12 (mcg)
OHX12TC	Tooth Count: Upper left 1st bicuspid/1st primary molar (1B)
DR1TLZ	Lutein + zeaxanthin (mcg)
MIAPROXY	Was a Proxy respondent used in conducting the MEC CAPI Interview?
MGQ110	Is the pain, aching or stiffness in your left hand caused by arthritis, tendonitis, or carpal tunnel syndrome?
DR1TVC	Vitamin C (mg)
DMDHREDU	HH reference person's education level
DMDHHSZE	Number of adults aged 60 years or older in the household
BMIARML	Upper Arm Length Comment
DMDHRMAR	HH reference person's marital status

Fig. 6. Depression severity risk factors

The RF model was utilized to identify personalized risk factors for each patient in the dataset, as illustrated in Figure 7. This approach provides more individualized and targeted treatment and clinical support, which can lead to improved patient outcomes. Overall, the RF model outperformed the other algorithms in predicting depression severity in the NHANES multimodal dataset. It had the lowest error rate, highest precision, and highest accuracy, showcasing its capability to detect complex relationships between the input features and the output variables. It also identified both generalized

and personalized risk factors for depression, which could support clinical diagnosis, decision-making, and intervention planning. The RF model thus demonstrated its potential utility for the practical assessment and management of depression.

```
For patient 75468:
Depression severity: Minimal or none
For patient 75469:
Depression severity: Minimal or none
For patient 75469:
Depression severity: Minimal or none
For patient 75469:
Depression severity: Minimal or none
For patient 75470:
Depression severity: Mild
For patient 75471:
Depression severity: Severe
For patient 75472:
Depression severity: Minimal or none
For patient 75472:
Depression severity: Minimal or none
For patient 75472:
Depression severity: Minimal or none
For patient 73557:
The personalised risk factors in order of importance are: Index(['DR1TP204', 'MGXH1T2E', 'DRD370EQ', 'BMXBMI', 'MGQ100', 'RIDSTATR',
       'BMXHEAD', 'DMDHSEDU', 'DRD350F', 'DMDFMSIZ', 'MIALANG', 'FIAPROXY',
       'BPXDI2', 'OHX06TC', 'DR1TFF', 'DRD370NQ', 'WTMEC2YR', 'DR1TSUGR',
       'FIAINTRP', 'DR1TVK', 'BPXDI4', 'DMQADFC', 'INDHHIN2', 'DR1TCRYP',
       'DR1TFA', 'DR1TFDFE', 'DR1TM201', 'DR1TCARB', 'MIAINTRP', 'DR1TCHL',
       'DRD370OQ', 'DRD350FQ', 'DR1TFOLA', 'WTINT2YR', 'DRD370BQ', 'DR1TWS',
       'SIALANG', 'DR1TB12A', 'OHX12TC', 'DR1TLZ', 'MIAPROXY', 'MGQ110',
       'DR1TVC', 'DMDHREDU', 'DMDHHSZE', 'DMDHRMAR'],
      dtype='object')
For patient 73557:
The personalised risk factors in order of importance are: Index(['DR1TP204', 'DRD370EQ', 'BMXBMI', 'BMDAVSAD', 'MGQ100', 'RIDSTATR',
       'BMXHEAD', 'DMDHSEDU', 'DRD350F', 'DMDFMSIZ', 'MIALANG', 'FIAPROXY',
       'BPXDI2', 'OHX06TC', 'DR1TFF', 'DRD370NQ', 'WTMEC2YR', 'DR1TSUGR',
       'DR1TVK', 'BPXDI4', 'DMQADFC', 'INDHHIN2', 'DR1TCRYP', 'DR1TFA',
       'DR1TFDFE', 'DR1TM201', 'DR1TCARB', 'DR1TIRON', 'DR1TCHL', 'DRD370OQ',
       'DRD350FQ', 'DR1TFOLA', 'WTINT2YR', 'DRD370BQ', 'DR1TWS', 'SIALANG',
       'DR1TB12A', 'OHX12TC', 'DR1TLZ', 'MIAPROXY', 'MGQ110', 'DR1TVC',
       'DMDHREDU', 'DMDHHSZE', 'BMIARML', 'DMDHRMAR'],
      dtype='object')
For patient 73558:
The personalised risk factors in order of importance are: Index(['MGXH1T2E', 'DRD370EQ', 'BMXBMI', 'MGQ100', 'RIDSTATR', 'BMXHEAD',
       'DMDHSEDU', 'DRD350F', 'DMDFMSIZ', 'MIALANG', 'FIAPROXY', 'BPXDI2',
       'OHX06TC', 'DR1TFF', 'WTMEC2YR', 'DR1TSUGR', 'FIAINTRP', 'DR1TVK',
       'BPXDI4', 'DMQADFC', 'INDHHIN2', 'DR1TCRYP', 'DR1TFA', 'DR1TFDFE',
       'DR1TM201', 'DR1TCARB', 'DR1TIRON', 'DR1TCHL', 'DRD370OQ', 'DR1TFOLA',
       'WTINT2YR', 'DR1TWS', 'SIALANG', 'DR1TB12A', 'OHX12TC', 'DR1TLZ',
       'MGQ110', 'DR1TVC', 'DMDHREDU', 'DMDHHSZE', 'DMDHRMAR'],
      dtype='object')
For patient 73558:
The personalised risk factors in order of importance are: Index(['DR1TP204', 'MGXH1T2E', 'DRD370EQ', 'BMXBMI', 'BMDAVSAD', 'MGQ100',
       'RIDSTATR', 'BMXHEAD', 'DMDHSEDU', 'DRD350F', 'DMDFMSIZ', 'MIALANG',
       'FIAPROXY', 'OHX06TC', 'DR1TFF', 'DRD370NQ', 'WTMEC2YR', 'DR1TSUGR',
       'FIAINTRP', 'DR1TVK', 'BPXDI4', 'DMQADFC', 'INDHHIN2', 'DR1TCRYP',
       'DR1TFA', 'DR1TFDFE', 'DR1TM201', 'DR1TCARB', 'DR1TIRON', 'DR1TCHL',
       'DRD370OQ', 'DRD350FQ', 'DR1TFOLA', 'WTINT2YR', 'DRD370BQ', 'DR1TWS',
       'SIALANG', 'DR1TB12A', 'OHX12TC', 'DR1TLZ', 'MIAPROXY', 'MGQ110',
       'DR1TVC', 'DMDHREDU', 'DMDHHSZE', 'BMIARML', 'DMDHRMAR'],
      dtype='object')
```

Fig. 7. Identified personalized risk factors of some patients

4 CONCLUSION

A depression severity and personalized risk factor prediction model was developed using ML algorithms on a multimodal dataset. Five ML algorithms, namely LR, SVM, LASSO, XGBoost, and RF, were applied to the NHANES multimodal dataset, which contains various types of information related to demographics, diet, socio-economic status, medical history, and clinical measurements. The results show that the RF algorithm outperformed the other algorithms in predicting depression severity using the NHANES multimodal data. The RF algorithm achieved the lowest

error rate, the highest precision, and the highest accuracy. The study further revealed important features in the NHANES multimodal dataset that were contributory and informative for predicting depression severity and personalized risk factors. These features could help develop preventive and early intervention approaches based on generalized risk factors and more individualized and targeted treatment strategies based on personalized risk factors, resulting in improved patient outcomes.

This study has proven the effectiveness of utilizing ML algorithms to predict outcomes and assist in the treatment of depression. This makes it a valuable, feasible, and dependable tool for clinical diagnosis, decision-making, intervention planning, and addressing this significant public health challenge.

5 REFERENCES

[1] M. L. Wainberg *et al.*, "Challenges and opportunities in global mental health: A research-to-practice perspective," *Current Psycchiatry Reports/Current Psychiatry Reports*, vol. 19, no. 5, p. 28, 2017. https://doi.org/10.1007/s11920-017-0780-z

[2] C. Gupta, P. Chandrashekar, T. Jin, C. He, S. Khullar, and Q. Chang, "Bringing machine learning to research on intellectual and developmental disabilities: Taking inspiration from neurological diseases," *J. Neurodev. Disord.*, vol. 14, no. 28, pp. 1–22, 2022. https://doi.org/10.1186/s11689-022-09438-w

[3] E. Olubunmi, A. Adetunla, S. Olufemi, A. Adeoye, and M. Glucksberg, "Heliyon an archetypal model of a breathable air-circuit in an electro-pneumatic ventilator device," *Heliyon*, vol. 8, no. 5, p. e09378, 2022. https://doi.org/10.1016/j.heliyon.2022.e09378

[4] Z. Uddin, "Deep learning for prediction of depressive symptoms in a large textual dataset," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 721–744, 2022. https://doi.org/10.1007/s00521-021-06426-4

[5] A. Nadeem, M. Naveed, M. I. Satti, H. Afzal, T. Ahmad, and K. Kim, "Depression detection based on hybrid deep learning SSCL framework using self-attention mechanism: An application to social networking data," *Sensors*, vol. 22, no. 24, p. 9775, 2022. https://doi.org/10.3390/s22249775

[6] V. Gabbay *et al.*, "A preliminary study of cytokines in suicidal and nonsuicidal adolescents with major depression," *Journal of Child and Adolescent Psychopharmacology*, vol. 19, no. 4, pp. 423–430, 2009. https://doi.org/10.1089/cap.2008.0140

[7] J. Masih and W. Verbeke, "Immune system function and its relation to depression: How exercise can alter the immune system-depression dynamics," *J. Depress. Anxiety*, vol. 7, no. 4, 2019. https://doi.org/10.4172/2167-1044.1000325

[8] O. Guerra and E. Eboreime, "The impact of economic recessions on depression, anxiety, and trauma-related disorders and illness outcomes – a scoping review," *Behav. Sci.*, vol. 11, no. 9, p. 119, 2021. https://doi.org/10.3390/bs11090119

[9] M. Amitai *et al.*, "The relationship between plasma cytokine levels and response to selective serotonin reuptake inhibitor treatment in children and adolescents with depression and/or anxiety disorders," *Journal of Child and Adolescent Psychopharmacology*, vol. 26, no. 8, pp. 727–732, 2016. https://doi.org/10.1089/cap.2015.0147

[10] K. Opoku *et al.*, "Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis," *Pervasive Mob. Comput.*, vol. 83, p. 101621, 2022. https://doi.org/10.1016/j.pmcj.2022.101621

[11] A. Sharma and W. J. M. I. Verbeke, "Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers dutch dataset (n = 11,081)," *Frontiers in Big Data*, vol. 3, pp. 1–11, 2020. https://doi.org/10.3389/fdata.2020.00015

[12] H. E. Villavicencio, J. Gamboa-Cruzado, J. López-Goycochea, and L. S. Soto, "The role of artificial intelligence in the diagnosis of neoplastic diseases: A systematic and bibliometric review," *International Journal of Online and Biomedical Engineering (iJOE),* vol. 20, no. 4, pp. 43–68, 2024. https://doi.org/10.3991/ijoe.v20i04.45429

[13] D. O. Bhende, G. Sakarkar, P. Khandar, S. S. Uparkar, and A. Bhave, "Enhancing classification performance through feature boost thyro: A comparative study of machine learning algorithms and feature selection," *International Journal of Online and Biomedical Engineering (iJOE),* vol. 20, no. 4, pp. 29–42, 2024. https://doi.org/10.3991/ijoe.v20i04.45413

[14] R. P. Thati, A. Singh, D. Praveen, and K. Sainaba, "A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms," *Multimedia Tools and Applications*, *Multimed. Tools and Appl.,* vol. 82, no. 4, pp. 4787–4820, 2023. https://doi.org/10.1007/s11042-022-12315-2

[15] Y. Li, K. Konstabel, and R. Mõttus, "Temporal associations between objectively measured physical activity and depressive symptoms: An experience sampling study," *Front. Psychiatry*, vol. 13, pp. 1–9, 2022. https://doi.org/10.3389/fpsyt.2022.920580

[16] F. Xia, Q. Li, X. Luo, and J. Wu, "Machine learning model for depression based on heavy metals among aging people: A study with national health and nutrition examination Survey 2017–2018," *Frontiers in Public Health*, vol. 10, 2022. https://doi.org/10.3389/fpubh.2022.939758

[17] C. L. Id and H. Kim, "Machine learning-based predictive modeling of depression in hypertensive populations," *PLoS ONE,* vol. 17, no. 7, pp. 1–17, 2022. https://doi.org/10.1371/journal.pone.0272330

[18] Y. Sun, Z. Fu, Q. Bo, Z. Mao, X. Ma, and C. Wang, "The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital," *BMC Psychiatry*, vol. 20, no. 474, pp. 1–7, 2020. https://doi.org/10.1186/s12888-020-02885-6

[19] N. P. A. Zuithoff *et al.*, "The patient health questionnaire-9 for detection of major depressive disorder in primary care: Consequences of current thresholds in a crosssectional study," *BMC Fam., Pract.*, vol. 11, no. 98, 2010. https://doi.org/10.1186/1471-2296-11-98

[20] S. Zulfiker, N. Kabir, A. Amin, and T. Nazneen, "An in-depth analysis of machine learning approaches to predict depression current research in behavioral sciences: An in-depth analysis of machine learning approaches to predict depression," *Curr. Res. Behav. Sci.*, vol. 2, p. 100044, 2021. https://doi.org/10.1016/j.crbeha.2021.100044

[21] I. C. Obagbuwa, S. Danster, and O. C. Chibaya, "Supervised machine learning models for depression sentiment analysis," *Front. Artif. Intell.,* vol. 6, 2022. https://doi.org/10.3389/frai.2023.1230649

[22] K. Vayadande, A. Bodhankar, A. Mahajan, and D. Prasad, "Classification of depression on social media using distant supervision," *ITM Web of Conferences,* vol. 50, no. 01005, p. 16, 2022. https://doi.org/10.1051/itmconf/20225001005

[23] A. Kuber, S. Kulthe, and P. Kosamkar, "Detecting depression in tweets using natural language processing and deep learning," *Information and communication technology for competitive strategies (ICTCS 2021),* vol. 400, pp. 453–461, 2023. https://doi.org/10.1007/978-981-19-0095-2_43

6 AUTHORS

Adefemi Ayodele holds Master's degrees in Computer Science and Big Data Technologies from the University of East London, United Kingdom and Ladoke Akintola University of Technology, Nigeria, respectively. He is currently employed as a Data Scientist at Generate First, United Kingdom. His research interests include Machine Learning, AI and Big Data.

Dr. Adedotun Adetunla holds a Ph.D. in Mechanical Engineering Science from the University of Johannesburg, South Africa. He is currently a Research Associate at the University of Johannesburg and a Senior Lecturer in the Department of Mechanical & Mechatronics Engineering at Afe Babalola University, Ado, Nigeria. He is interested in Robotics, Materials Science, Additive Manufacturing, and Machine Learning (E-mail: adetunla.adedotun@abuad.edu.ng).

Prof. Esther Akinlabi is a Professor of Mechanical Engineering, currently serving as the Deputy Faculty Pro Vice Chancellor for Research and Knowledge Exchange, Faculty of Engineering and Environment, Northumbria University, Newcastle, United Kingdom. Her research interests are in modern and advance manufacturing processes which include Friction Stir Welding and Additive Manufacturing.