Proceedings of 2024 IEEE 12th International Conference on Intelligent Systems (IS)

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

Mohammad Hossein Amirhosseini
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
m.h.amirhosseini@uel.ac.uk

Adefemi Lawrence Ayodele
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
u2217852@uel.ac.uk

Amin Karami
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
a.karami@uel.ac.uk

Abstract*—Depression is a widespread mental health issue with profound global impact, often leading to diminished life quality and increased suicide risk. Despite available treatments, many depression cases go unnoticed and untreated. This underscores the necessity for a precise, personalized model to predict depression severity and individual risk factors, utilizing machine learning on comprehensive, multimodal datasets. While previous efforts employing machine learning (ML) to gauge depression severity exist, their effectiveness has been curtailed by small datasets and a lack of personalization. To address this gap, we propose an advanced ML-based approach for predicting depression severity and identifying personalized risk factors. ML enhances the precision of depression severity assessments, facilitates personalized treatment strategies, and improves the identification of individual risk factors. In our study, we implemented, assessed, and compared five supervised ML algorithms—Linear Regression (LR), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Least Absolute Shrinkage and Selection Operator (LASSO)—known for their accuracy, interpretability, and computational efficiency. We utilized a multimodal dataset from the National Health and Nutrition Examination Survey (NHANES), encompassing demographic, dietary, socio-economic, lifestyle, medical, laboratory, and clinical data. The Random Forest algorithm proved to be the most effective, demonstrating an R-squared of 0.93, an explained variance score (EVS) of 0.93, a mean absolute error (MAE) of 0.51, a mean squared error (MSE) of 1.73, and a root mean squared error (RMSE) of 1.32. It effectively pinpointed both general and personalized risk factors for depression severity. Our model not only proves effective in predicting depression severity and identifying personalized risk factors but also shows promise for clinical application in assessment, diagnosis, treatment planning, and depression management.

Keywords—Depression severity, personalised risk factors, machine learning, Multimodal data

I. INTRODUCTION

Depression has been recognised as a global health crisis [1]. In fact, five of the ten most common illnesses that globally render people disabled or incapable are mental illnesses, with depression coming in as number one [2]. Depression is a prevalent psychological condition that impacts individuals across all ages, genders, cultures, and backgrounds. It devastates an individual's quality of life at work, school, and home [3]. Depression results from an interplay of social, psychological, and physiological factors [3].

The World Health Organization (WHO) statistics asserts, the global population suffering from depression is estimated at 3.8%, totalling approximately 280 million individuals [4]. Early identification and assessment of depressive symptoms and appropriate evaluation and therapy can significantly enhance the odds of controlling symptoms and the underlying disease and protect personal, economic, and social well-being [3].

Despite available treatment options for depression, it remains undertreated and undiagnosed in many cases. Akincigil and Matthews [5] highlighted that while primary care providers can accurately diagnose depression when symptoms are recognized, data indicate that depression remains undetected in about half of the cases. Health professionals face a significant challenge in understanding how to tackle depression early on and enhance treatment outcomes. There is limited research on accurately diagnosing severity of depression and identifying the personalised risk factors [5]. Therefore, developing effective and personalised approaches for predicting depression severity and identifying risk factors is essential to improve depression management and outcomes.

Clinical interviews and self-report tools, such as the Patient Health Questionnaire-9, can be utilized to determine depression severity [6]. However, these methods can take considerable time and be subject to human error, potentially resulting in misdiagnosis or inappropriate treatment.

ML algorithms have demonstrated promising results for improving depression diagnosis and multiple studies have explored their use for depression prediction [7, 8]. However, most of the previous studies have only focused on predicting depression rather than its severity. They have typically used small-sized single-modality data with depression symptom variables integrated, with no consideration for identifying personalised risk factors. This is particularly important as personalised risk factor identification can aid in creating tailored treatment plans with better treatment outcomes. Thus, more analysis of multiple-source modalities of data has been required as it can potentially facilitate the prediction of severity of depression and generate personalised risk factors.

To bridge this research gap, we explored the prediction of severity of depression and the identification of personalised risk factors, utilising ML techniques on a comprehensive, multimodal dataset. The findings presented in this paper can have significant clinical implications, such as enabling early detection and personalised treatment of depression, improving treatment outcomes, and reducing the burden of depression on individuals and society.

II. PREVIOUS WORKS

Depression is a heterogeneous psychiatric condition characterised by persistent and pervasive negative affect, anhedonia, and reduced energy levels. The presence and

severity of core symptoms, along with additional co-morbid traits like cognitive decline, insomnia, and suicidal ideation, are used to diagnose depression. However, the diagnostic process of depression faces several challenges, such as the variability and subjectivity of symptom expression, the reliance on self-reported or clinician-rated measures, and the lack of objective and standardised criteria. These challenges can compromise the validity and reliability of depression diagnosis and lead to misdiagnosis or underdiagnosis of the disorder [9, 10].

ML is a field of artificial intelligence that enables computers to learn from data and perform predictions without the need for comprehensive programming. One of the new frontiers for ML application is in the mental health sector, especially depression diagnosis. ML can help to overcome limitations such as subjectivity and variability associated with traditional diagnostic methods by providing a more objective and data-driven approach to depression diagnosis. ML can identify new patterns and risk factors for depression that may not have been previously considered, such as genetic, environmental, or social factors [11]. ML algorithms can leverage various data types, such as clinical records, self-reports, brain imaging, speech, facial expressions, and social media posts, to identify patterns and features indicative of depression. ML can also improve the efficiency and scalability of depression diagnosis, especially in settings where resources and expertise are limited [12]. Additionally, ML algorithms can help to personalise and optimise treatment plans for patients based on their characteristics and preferences. ML algorithms are a cutting-edge tool for diagnosing depression, with the potential to increase the consistency and accuracy of diagnoses, ultimately leading to better patient outcomes.

Huang et al. [13] designed and assessed models to predict depression diagnosis, severity, and response to treatment. They utilized electronic health record (EHR) data from two sources, comprising 3,501 patients treated for depression at the Group Health Research Institute and 5,651 patients from the Palo Alto Medical Foundation, to train and evaluate regression-based models. They employed LASSO logistic regression, using features such as gender, annual visit frequency, ICD-9 codes, and specific terms for diseases and medications. They discovered that their models could predict depression diagnosis up to 12 months ahead, distinguish severe from mild depression, and identify predictors of treatment response, such as baseline severity. They also developed a model to predict baseline severity using the same features. However, they had some drawbacks with processing and labelling the data. The annotation process and gold standard labels may have introduced inaccuracies and potential biases in the model's training and testing phases.

In a different study, Priya et al. [14] utilized ML techniques to identify stress, anxiety, and depression in 348 participants who completed the Depression Anxiety Stress Scale 21 questionnaire. In their study, five ML algorithms including (1) Random Forest, (2) K-Nearest Neighbor, (3) Naive Bayes, (4) Support Vector Machine, and (5) Decision Tree were utilized. They reported that the best performing model for predicting depression was Naive Bayes, with an 85.50% accuracy. They demonstrated that ML algorithms can effectively detect anxiety, depression, and stress, with their findings contributing to early diagnosis and treatment of psychological disorders and providing insights for creating more accurate predictive algorithms. However, the study's limitation includes the imbalanced classes within the data that hinder the generalizability of the findings.

Furthermore, Haque et al. [15] utilized ML to identify depression in children and adolescents using data from Young Minds Matter (Australia's Second Child and Adolescent Survey on Mental Health and Well-being 2013–14). They integrated a Random Forest classifier with a Boruta algorithm to detect the most relevant features that could be used for detecting depression. The appropriate supervised learning algorithms were selected using Tree-based Pipeline Optimization Tool (TPOTclassifier). The depression detection model included using (1) Random Forests, (2) XGBoost, (3) Decision Tree, and (4) Gaussian Naive Bayes algorithms, with Random Forest outperforming all others models with a 95% accuracy rate. They highlighted the power of ML for predicting depression among children and adolescents using the Young Minds Matter dataset. However, the population size and the chosen data source limited the results and they needed to conduct more studies before they can generalise their findings.

Additionally, Sharma and Verbeke [16] explored the relationship between self-reported depression and various biomarkers using the XGBoost algorithm on a Dutch dataset comprising 11,081 cases to enhance depression diagnosis. To balance the imbalanced dataset before classification, they used several sampling techniques, such as ROSE sampling, over sampling, under sampling, and a combination of over-under sampling. They discovered that over-sampling and over-under sampling resulted in high-performance measures, with balanced accuracy, F1 score, recall, and precision, all above 90%. However, they acknowledged that their XGBoost model had limited generalisability to other populations and contexts.

Furthermore, Xin et al. [17] utilized the Random Forest algorithm to predict depression in a cohort of 112 Malaysian women. They used SMOTE, a random oversampling technique, to tackle the data's class imbalance and improve the overall sensitivity and accuracy of the classification. They achieved an overall accuracy rate of 90%. However, they only used a small dataset of females for their research.

In a more recent attempt, Uddin et al. [3] used a multimodal approach to identify symptoms of depression in text by employing one-hot encoding of robust features combined with deep learning techniques. They attained a comprehensive dataset including textual data from the public Norwegian information website ung.no. One-hot encoding of robust depression features was utilized to train a deep recurrent neural network (RNN) with 50-unit long short-term memory (LSTM) cells, effectively addressing the vanishing gradient problem and achieve high mean prediction accuracy of 98% and 99% on two datasets with 11,807 and 21,807 text data, respectively. The authors suggested that deep learning-based efficient systems should be explored further for real-time analysis and prediction of mood disorders in intelligent environments in combination with cutting-edge technologies. However, the variety of this dataset limited the study, and a more comprehensive and diverse dataset was necessary.

Thati et al. [18] also conducted a study to identify depressed and non-depressed participants by collecting smartphone usage data, emotion elicitation data, and speech data from 102 volunteers aged 18–19 through social networks. The study employed a ML method for analyzing multimodal data. To extract, choose, fuse, and classify features of these modalities, the authors used principal component analysis, pearson's correlation analysis, with different ML classifiers, including (1) naïve bayes, (2) decision tree, (3) support vector

machines, (4) Random Forests, and (5) logistic regression algorithms. They found that deploying features from various data modalities outperformed using a single modality, even on a benchmark dataset. Utilizing this approach, they attained an accuracy of 86% using an SVM classifier. The study's main limitation was the limited dataset from a limited number of participants. Using large-scale, more diverse datasets with clinically validated depression was recommended for further exploration.

In this research, the effectiveness of various ML techniques was examined to improve our understanding of depression, its severity, and risk factors. Comparing to the previous studies, we used a large-scale dataset to explore how multimodal data, which included demographic, behavioural, dietary, and clinical information, could identify the factors that induced depression in individuals and reveal the severity of depression. The results can support the prevention, early detection, and intervention of depression, improve patient outcomes, and potentially contribute to developing personalised treatment strategies.

III. METGODOLOGY

A. Data Collection

The National Health and Nutrition Examination Surveys (NHANES), 2013-2014 edition dataset was acquired in this study. According to the Centers for Disease Control and Prevention [19], the NHANES program, conducted by the National Center for Health Statistics (a division of the Centers for Disease Control and Prevention), evaluates the nutritional and health status of adults and children in the US. It collects data on health and nutrition through a series of health surveys conducted periodically since the early 1960s [19].

The NHANES 2013-2014 edition sampled 14,332 individuals across 30 different survey locations throughout the United States; 10,175 completed the interview, while they examined 9,813. NHANES gathers various health and nutrition measurements through interviews and physical examinations, utilizing data from multiple sources [19]. The interview component encompassed questions on demographics, socioeconomic status, diet, and health. The examination component included medical, dental, and physiological assessments, as well as laboratory tests performed by specialist medical personnel. This process provided diverse data types, including demographics from various racial backgrounds, dietary information, examination results, laboratory data, and questionnaire responses, covering a broad spectrum of areas such as health status, nutrition, risk behaviors, and environmental exposures [19]. The dataset offers a comprehensive and multifaceted source for developing a ML model capable of predicting severity of depression and identifying personalized risk factors.

B. Data Processing

The multimodal NHANES dataset was pre-processed before analysis and modelling to ensure the data's quality, usability and consistency. Six sub-datasets, demographics, diet, examination, laboratory, medication, and questionnaire were sorted and merged into one dataset using the participant's 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 contained in the dataset, calculating the PHQ-9 score established the depression severity in the dataset. The PHQ-9 score is a commonly used measure to evaluate the severity of depression, ranging from 0 to 27, where higher scores denote more severe levels of depression [20]. The PHQ-9 score was incorporated into the dataset as a new column and used as the target variable for predicting depression severity. The variables used to calculate the PHQ-9 score were dropped from the dataset to avoid multicollinearity and data leakage, as well as rows that contained declined or no information from the questionnaire in the dataset.

Removal of irrelevant, redundant, or erroneous rows cleaned up the dataset. The missing values were replaced by imputing the median or mean. The final dataset had 19,560 rows and 1,767 columns. The columns are representing the features. The numerical variables were scaled to a standard range of values between 0 and 1, normalising the dataset to reduce the effect of outliers and different units of measurement on the data.

The data was divided into test and training 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 sets. The training set comprised 80% of the data (15,648 rows and 1,767 columns), while the test set consisted of the remaining 20% (3,912 rows and 1,767 columns). The splitting was done at this stage to avoid data leakage during the feature extraction phase. Data leakage may happen when information from the test set is used or revealed in the training set, which can lead to overfitting and inaccurate results.

C. Feature Selection

To streamline the dataset and decrease its dimensionality and complexity, and to enhance both the performance and interpretability of the depression severity prediction algorithms, the SelectKBest feature selection method from the sci-kit-learn library was utilized, as recommended by Biswas et al. [21]. According to a scoring function, SelectKBest chooses the k features with the highest scores and ranks them in order of importance for the target variable [21, 22]. The score function accepts two input arrays, X and y, where X represents the feature matrix and y is the target vector. It returns either a pair of arrays (scores and p-values) or a single array of scores [23].

A score function was applied to each feature to obtain the K highest score used to select the features. By calculating the statistical significance of the correlation between the target variable and each feature using the F-score, the f_regression score function applied a score function to each feature. [24]. The F-score measures how well a model fits the data based on comparing the explained variance by the model and the unexplained variance. A higher F-score indicates a higher correlation and a better fit [24].

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 5-fold cross-validation. For each fold, the training data was split into test and training sets, fitted the selectKBest on the train set, transformed both sets using the selected features, fitted the Random Forest Regressor as a selector on the train set, and predicted 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.

D. Machine Learning Algorithms

The output of the feature selection process was normalized using MinMaxScaler. This method normalized the values of each feature to a range of 0 to 1 by subtracting the minimum value and dividing by the overall range, without distorting the differences between values and the original distribution's shape. Equation 1 was employed for the normalisation process, where X_{min} is the minimum value and X_{max} is the maximum value within the X Feature.

$$X_{scaled} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

The normalisation of the features made them comparable and prevented some from dominating others due to their large scale. The ML algorithms took in the normalised features as input.

Five ML algorithms with literature-record of accuracy, interpretability, complexity, and computational efficiency in handling regression tasks were implemented in this study. The models include (1) Extreme Gradient Boosting (XGBoost), (2) Support Vector Machine (SVM), (3) Linear Regression (LR), (4) Random Forest (RF), and (5) Least Absolute Shrinkage and Selection Operator (LASSO). The top-performed ML model will be used to generate personalised risk factors for each participant in the dataset, and the feature importance would be determined to identify the most critical variables in predicting depression severity and personalised risk factors and potentially aid in improving the understanding and management of depression.

E. Hyper Parameter Tuning

GridSearchCV was used for hyper parameter tuning for each model. Table 1 displays the optimal values selected for each parameter and model.

TABLE I. HYPER PARAMETER TUNING OUTCOMES

Model	Parameter	Value
LR	'fit_intercept'	True
	'normalize'	True
	'copy_X'	False
	'n_jobs'	-1
	'positive'	True
RF	'n_estimators'	100
	'max_depth'	20
	'min_samples_split'	5
	'min_samples_leaf'	2
	'bootstrap'	True
SVM	'C'	1.0
	'kernel'	rbf
	'degree'	3
	'random_state'	42
	'tol'	0.001
	'cache_size'	200
	'max_iter'	1
	'decision_function_shape'	ovr
XGBoost	'n_estimators'	200
	'max_depth'	5
	'learning_rate'	0.1
	'subsample'	1.0
LASSO	'alpha'	4
	'positive'	True
	'precompute'	True

F. Performance Evaluation

Evaluating the ML algorithms determines the most efficient predictive model for assessing severity of depression on new, unseen data. The performance of the ML algorithms was assessed using the discriminative and evaluative metrics including R-Squared, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Explained Variance Score (EVS).

R-Squared measures how well the predictive model explains the variation in the dependent variable. This coefficient ranges from 0 to 1, with higher values indicating a superior fit [25]. Moreover, MAE quantifies the average absolute difference between predicted and actual values, indicating the model's accuracy without consideration of the error direction. A lower MAE signifies smaller prediction errors and a better model fit [26].

Furthermore, MSE calculates the average squared difference between the actual values and predicted values. It represents the model's accuracy and fit, with a lower MSE indicating smaller prediction errors and a superior fit [27]. Additionally, RMSE quantifies the average difference between observed values and predicted values, presented in the same units as the target variable [26].

Finally, EVS assesses how well the model captures the variation in the observed data. It is the ratio of the variance of the predicted values to the variance of the observed values. A higher EVS indicates a better fit and less unexplained variation [28].

G. Personalised Risk Factors

Personalised risk factors are the most significant factors contributing to depression severity in each patient. They are determined based on feature importance, which refers to the input variables that most significantly contribute to predicting depression severity in the multimodal NHANES dataset, using the best-performing ML model considered in this study.

The feature importance was obtained using the "feature_importances_" attribute on the best-performing model and sorted in descending order using the NumPy function "argsort()" to identify personalised risk factors for patients in the test dataset. The identified important features in the test dataset were selected and used to compute each patient's corresponding feature values. Then, the features with non-zero importance scores were selected in order of importance, helping to identify the most significant risk factors for that patient.

As a result, using a for loop, the index and patient ID of the first n patients in the test dataset was iterated over. Following the important features' selection using array slicing, the feature values for each patient were computed using the "iloc" attribute of the input data. Finally, the features with non-zero importance scores were selected using Boolean indexing and output in the patient's specific order of importance.

IV. RESULTS

The five implemented ML models were trained and tested in predicting depression severity using the NHANES multimodal dataset in an 80:20 ratio. The performance of the ML models was examined using various evaluation metrics

including (1) R-Squared, (2) MSE, (3) MAE, (4) RMSE, and (5) EVS. Table 2 shows the results of the performance evaluation.

TABLE II. PERFORMANCE EVALUATION RESULTS

	Machine Learning Models				
	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

In addition, Table 3 shows the identified general depression severity risk factors in order of importance, with matching variable description.

TABLE III. IDENTIFIED GENERAL DEPRESSION SEVIRETY RISK FACTORS IN ORDER OF IMPORTANCE

	Variable	**Variable Description**
1	**DR1TP204**	PFA 4:20 Eicosatetraenoic (gm)
2	**MGXH1T2E**	Whether the patient exerted a maximal or questionable effort during the test 2 on hand 1, as assessed by the technician
3	**DRD370EQ**	Number of time cod was eaten in the past 30 days
4	**BMXBMI**	Body Max Index (kg/m**2)
5	**BMDAVSAD**	Average Sagittal Abdominal Diameter (cm)
6	**MGQ100**	Have you had any pain, aching or stiffness in your left hand in the last 7 days?
7	**RIDSTATR**	Interview and Examination Status of the Sample Person
8	**BMXHEAD**	Head Circumference (cm)
9	**DMDHSEDU**	What is the highest degree you have/ SPOUSE has received
10	**DRD350F**	Mussels eaten during past 30 days
11	**DMDFMSIZ**	Total number of people in the family
12	**MIALANG**	Language of the MEC CAPI interview
13	**FIAPROXY**	Was a proxy respondent used in the conducting family interview?
14	**BPXDI2**	Diastolic: Blood pressure (second reading) mm Hg
15	**OHDEXSTS**	Overall Oral Health Exam Status
16	**OHX06TC**	Tooth Count: Upper right cuspid
17	**DR1TFF**	Food folate (mcg)
18	**DRD370NQ**	# of times sardines eaten past 30 days
19	**WTMEC2YR**	Full sample 2 years MEC exam weight
20	**DR1TSUGR**	Total sugars (gm)
21	**FIAINTRP**	Was an interpreter used to conduct the family interview?
22	**DR1TVK**	Vitamin K (mcg)
23	**BPXDI4**	Diastolic: Blood pressure (fourth reading if necessary) mm Hg
24	**DMQADFC**	Did {you/SP} ever serve in a foreign country during a time of armed conflict or on a humanitarian or peas-keeping mission?
25	**INDHHIN2**	Total household income (Dollars)
26	**DR1TCRYP**	Beta-cryptoxanthin (mcg)
27	**DR1TFA**	Folic Acid (mcg)
28	**DR1TFDFE**	Folate, DFE (mcg)
29	**DR1TM201**	MFA 20:1 (Eicosenoic) (gm)
30	**DR1TCARB**	Carbohydrate (gm)
31	**MIAINTRP**	Was an interpreter used to conduct the MEC CAPI interview?
32	**DR1TIRON**	Iron (mg)
33	**DR1TCHL**	Total choline (mg)
34	**DRD370OQ**	# of times sea bass eaten in past 30 days
35	**DRD350FQ**	# of times oyster eaten in past 30 days
36	**DR1TFOLA**	Total Folate (mcg)
37	**WTINT2YR**	Interviewed participants
38	**DRD370BQ**	# of time tuna eaten in past 30 days
39	**DR1TWS**	Tap water source
40	**SIALANG**	Language of the Person Interview
41	**DR1TB12A**	Added vitamin B12 (mcg)
42	**OHX12TC**	Tooth Count: Upper left 1st bicuspid / 1st Primary moral (1B)
43	**DR1TLZ**	Lutein + Zeaxanthin (mcg)
44	**MIAPROXY**	Was a proxy respondent used in conducting the MEC CAPI interview?
45	**MGQ110**	Is the pain, aching or stiffness in your left hand caused by arthritis, tendonitis, or carpal tunnel syndrome?
46	**DR1TVC**	Vitamin C (mg)
47	**DMDHREDU**	HH reference person's education level
48	**DMDHHSZE**	Number of adults aged 60 years or older in the household
49	**MBIARML**	Upper Arm Length Comment
50	**DMDHRMAR**	HH reference person's marital status

Furthermore, Table 4 shows examples of personalised risk factors which have been identified for two patients in the dataset.

TABLE IV. IDENTIFIED PERSONALISED RISK FACTORS FOR TWO PATIENTS WITHIN THE DATASET

Patient ID	**Personalised Risk Factors in Order of Importance**		
2558	1- DR1TP204 2- DRD370EQ 3- BMXBMI 4- BMDAVSAD 5- MGQ100 6- RIDSTATR 7- BMXHEAD 8- DMDHSEDU 9- DRD350F 10- DMDFMSIZ 11- MIALANG 12- FIAPROXY 13- BPXDI2 14- OHX06TC 15- DR1TFF	16- DRD370NQ 17- WTMEC2YR 18- DR1TSUGR 19- DR1TVK 20- BPXDI4 21- DMQADFC 22- INDHHIN2 23- DR1TCRYP 24- DR1TFA 25- DRITFDFE 26- DRITH201 27- DRITCARB 28- DRITIRON 29- DRITCHL 30- DRD3700Q	31- DRD3SOPO 32- DRITPOLA 33- WTINT2YR 34- DRD37OBQ 35- DRITWS 36- SIALANG 37- DRITB12A 38- DRITLZ 39- MIAPROXY 40- MGQ110 41- DRITVC 42- DHDHREDU 43- DMDHHS2E 44- BMIARML 45- DMDHRMAR
2559	1- MGXH1T2E 2- DRD370EQ 3- BMXBMI 4- MGQ100 5- RIDSTATR 6- BMXHEAD 7- DMDHSEDU 8- DRD350F 9- DMDFMSIZ 10- MIALANG 11- FIAPROXY 12- BPXDI2 13- OHX06TC 14- DR1TFF 15- WTMEC2YR	16- DR1TSUGR 17- FIAINTRP 18- DR1TVK 19- BPXDI4 20- DMQADFC 21- INDHHIN2 22- DR1TCRYP 23- DR1TFA 24- DR1TFDFE 25- DR1TM201 26- DR1TCARB 27- DR1TIRON 28- DR1TCHL 29- DRD370OQ 30- DR1TFOLA	31- WT1NT2YR 32- DR1TWS 33- SIALANG 34- DR1TB12A 35- OHX12TC 36- DR1TLZ 37- MGQ110 38- DR1TVC 39- DMDHREDU 40- DMDHHSZE 41- DMDHRMAR

V. DISCUSSION

On evaluation using the considered metrics, the RF model outperformed the other considered ML models. This model achieved the lowest MSE of 1.73 and the lowest RMSE of 1.32, indicating the lowest average squared difference between the actual values and predicted values. As a result, it estimates the output variables accurately from the input features on the same scale as the target value with little error. Moreover, the RF model had the highest R-squared of 0.93, implying that the input features explained a more significant portion of the variance in the target value. The highest R-

squared feat showed that the model had high explanatory power and accounted for most of the variation in the data.

Furthermore, the Random Forest (RF) model achieved the lowest MAE of 0.51, indicating a minimal average absolute difference between the actual values and predicted values. This demonstrates the model's low error rate and its ability to accurately estimate the output variables from the input features without being affected by outliers. In addition, with the highest Explained Variance Score (EVS) of 0.93, which is 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 closely aligned with the actual target values.

As a result, the best-performing depression severity prediction model, the RF model, was used to identify fifty important features that is considered the general risk factors, presented in table 3. These generalized risk factors are contributing factors that can inform the development of preventive and early intervention strategies. Moreover, the RF model was utilized to identify personalised risk factors for each patient in the dataset.

Table 4 shows that the model identified 45 personalised risk factors for patient 2558, and 41 personalised risk factors for patient 2559. The reason for difference between the number of identified personalized risk factors for each patient is that only the features with non-zero importance scores were selected and displayed. In other words, the features with zero importance score were ignored and they were not displayed. In addition, figure 1 clearly shows that the order of identified personalized risk factors for each patient is different based on their depression severity. This approach provides a more individualized and targeted treatment and clinical support approach, which can lead to improved patient outcomes.

Overall, the random forest (RF) model outperformed the other algorithms in predicting depression severity in the NHANES multimodal dataset. It demonstrated the lowest error, highest precision, and greatest accuracy, showcasing its ability to detect complex relationships between the output variables and the input features. It also identified both generalised and personalised 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.

VI. Conclusion

In this study, a depression severity and personalised risk factors predictive model was developed using a ML algorithm on a multimodal dataset. Five ML algorithms, including (1) Extreme Gradient Boosting (XGBoost), (2) Support Vector Machine (SVM), (3) Linear Regression (LR), (4) Random Forest (RF), and (5) Least Absolute Shrinkage and Selection Operator (LASSO) were implemented and evaluated in this work. The NHANES multimodal dataset, which contains various types of information related to demographics, diet, socio-economic status, medical history, and clinical measurements was used in this research.

The results indicated that the RF model outperformed the other models in predicting depression severity with the least MSE of 1.73, least RMSE of 1.32, highest R-squared of 0.93, least MAE of 0.51, and highest EVS of 0.93. The study further revealed the important features in the NHANES multimodal dataset that were contributive and informative for predicting depression severity and personalised risk factors. These features could help developing preventive and early intervention approaches based on generalised risk factors and more individualised and targeted treatment strategies based on the personalised risk factors, resulting in improved patient outcomes.

The novelty of this work stems from its comprehensive approach to understanding depression, which goes beyond merely predicting its occurrence to assess its severity and identify personalized risk factors. Unlike previous studies that relied on small, single-modality datasets focusing solely on depression symptoms, our study utilises a comprehensive multimodal dataset. The innovative approach used in this research is pivotal in developing customized treatment plans, thereby enhancing the effectiveness of interventions and improving patient outcomes. By integrating multiple data sources, our research offers a more holistic and detailed understanding of depression, paving the way for early detection and individualized treatment strategies. Consequently, this can significantly lessen the overall impact of depression on individuals and society, marking a substantial advancement in mental health research and clinical practice.

References

[1] Samareh, A.,Jin, Y., Wang, Z., Chang, X., and Huang, S. (2018) Detect depression from communication: how computer vision, signal processing, and sentiment analysis join forces, IISE Transactions on Healthcare Systems Engineering, 8:3, 196-208, DOI: 10.1080/24725579.2018.1496494

[2] Ding, Y., Chen, X., Fu, Q. and Zhong, S. (2020). A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm. *IEEE Access*, 8, pp.75616–75629. doi:https://doi.org/10.1109/access.2020.2987523.

[3] Uddin, M.Z., Dysthe, K.K., Følstad, A. and Brandtzaeg, P.B. (2021). Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*. doi:https://doi.org/10.1007/s00521-021-06426-4.

[4] World Health Organization (2023). *Depressive Disorder (Depression)*. [online] World Health Organization. Available at: https://www.who.int/news-room/fact-sheets/detail/depression

[5] Akincigil, A. and Matthews, E.B. (2017). National Rates and Patterns of Depression Screening in Primary Care: Results From 2012 and 2013. *Psychiatric Services*, 68(7), pp.660–666. doi:https://doi.org/10.1176/appi.ps.201600096.

[6] Cheung, R.Y.M. (2023). Patient Health Questionnaire-9 (PHQ-9). *Handbook of Assessment in Mindfulness Research*, pp.1–11. doi:https://doi.org/10.1007/978-3-030-77644-2_63-1.

[7] Lennon, M.J. and Harmer, C. (2023). Machine learning prediction will be part of future treatment of depression. *Australian and New Zealand Journal of Psychiatry*, p.000486742311582-000486742311582. doi:https://doi.org/10.1177/00048674231158267.

[8] Chung, J. and Teo, J. (2022). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Applied Computational Intelligence and Soft Computing*, pp. 9970363. doi:https://doi.org/10.1155/2022/9970363.

[9] Kroenke, K., Alford, D.P., Argoff, C., Canlas, B., Covington, E., Frank, J.W., Haake, K.J., Hanling, S., Hooten, W.M., Kertesz, S.G., Kravitz, R.L., Krebs, E.E., Stanos, S.P. and Sullivan, M. (2019). Challenges with Implementing the Centers for Disease Control and Prevention Opioid Guideline: A Consensus Panel Report. *Pain Medicine*, 20(4), pp.724–735. doi:https://doi.org/10.1093/pm/pny307.

[10] Nanni, V., Uher, R. and Danese, A. (2012). Childhood Maltreatment Predicts Unfavorable Course of Illness and Treatment Outcome in Depression: A Meta-Analysis. *American Journal of Psychiatry*, 169(2), pp.141–151. doi:https://doi.org/10.1176/appi.ajp.2011.11020335.

[11] Wang, H., Yang, G., Wang, S., Zheng, X., Zhang, W. and Li, Y. (2018). The Most Commonly Treated Acupuncture Indications in the United States: A Cross-Sectional Study. *The American Journal of Chinese Medicine*, pp.1–33. doi:https://doi.org/10.1142/S0192415X18500738.

[12] Shatte, A.B.R., Hutchinson, D.M. and Teague, S.J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, 49(09), pp.1426–1448. doi:https://doi.org/10.1017/s0033291719000151.

[13] Huang, S.H., LePendu, P., Iyer, S.V., Tai-Seale, M., Carrell, D. and Shah, N.H. (2014). Toward personalizing treatment for depression: predicting diagnosis and severity. *Journal of the American Medical Informatics Association*, 21(6), pp.1069–1075. doi:https://doi.org/10.1136/amiajnl-2014-002733.

[14] Priya, A., Garg, S. and Tigga, N.P. (2020). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Computer Science*, 167, pp.1258–1267. doi:https://doi.org/10.1016/j.procs.2020.03.442.

[15] Haque, U.M., Kabir, E. and Khanam, R. (2021). Detection of child depression using machine learning methods. *PloS One*, [online] 16(12), p.e0261131. doi:https://doi.org/10.1371/journal.pone.0261131.

[16] Sharma, A. and Verbeke, W.J.M.I. (2020). Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Frontiers in Big Data*, 3. doi:https://doi.org/10.3389/fdata.2020.00015.

[17] Xin, L.K. and Rashid, N. binti A. (2021). Prediction of Depression among Women Using Random Oversampling and Random Forest. *2021 International Conference of Women in Data Science at Taif University*. doi:https://doi.org/10.1109/widstaif52235.2021.9430215.

[18] Thati, R.P., Dhadwal, A.S., Kumar, P. and P, S. (2022). A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms. *Multimedia Tools and Applications*. doi:https://doi.org/10.1007/s11042-022-12315-2.

[19] Centers for Diseas Control and Prevention (2013) National Health and Nutrition Examination Survey. Available at: https://wwwn.cdc.gov/Nchs/Nhanes/ContinuousNhanes/Overview.aspx?BeginYear=2013

[20] Kroenke, K., Spitzer, R.L. and Williams, J.B.W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), pp.606–613. doi:https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

[21] Biswas, S., Islam, Md.Mozahidul., Sarker, U., Hridoy, R.H. and Habib, Md.T. (2022). Machine Learning-Based Depression Detection. *Computer Networks and Inventive Communication Technologies*, pp.809–821. doi:https://doi.org/10.1007/978-981-19-3035-5_60.

[22] Tislenko, M.D., Gaidel, A.V. and Kupriyanov, A.V. (2022). Comparison of feature selection algorithms for Data classification problems. *2022 VIII International Conference on Information Technology and Nanotechnology (ITNT)*. doi:https://doi.org/10.1109/itnt55410.2022.9848765.

[23] Vishwakarma, A.K. (2021). Network Traffic Based Botnet Detection Using Machine Learning. doi:https://doi.org/10.31979/etd.4nd6-m6hp.

[24] Pirbazari, A.M., Chakravorty, A. and Rong, C. (2019). Evaluating Feature Selection Methods for Short-Term Load Forecasting. *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. doi:https://doi.org/10.1109/bigcomp.2019.8679188.

[25] Li, Z., Du, C., Tang, Y. and Li, X. (2020). Experimental and Statistical Investigation of Reservoir Properties with the Effect of Waterflooding Treatment. *ACS Omega*, 5(33), pp.20922–20931. doi:https://doi.org/10.1021/acsomega.0c02374.

[26] Saigal, S. and Mehrotra, D. (2012) Performance Comparision of Time Series Data Using Predictive Data Mining Techniques, *Advances in Information Mining*, 4(1), pp.57-66.

[27] Hossin, M. and Sulaiman, M.N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, [online] 5(2), pp.01-11. doi:https://doi.org/10.5121/ijdkp.2015.5201.

[28] Demaris, A. (2002). Explained Variance in Logistic Regression A Monte Carlo Study of Proposed Measures. *Sociological Methods & Research*, 31, pp.27–74. doi:https://doi.org/10.1177/0049124102031001002.