Important Features Associated with Depression Prediction and Explainable AI

Vincent Peter C. Magboo[(✉)] and Ma. Sheila A. Magboo

Department of Physical Sciences and Mathematics, University of the Philippines Manila, Manila, Philippines
{vcmagboo,mamagboo}@up.edu.ph

Abstract. Depression is a debilitating disease that leaves individuals persistently feeling sad or hopeless for more than two weeks affecting more than 300 million people globally. We applied several machine learning models with model explainability to a publicly available depression dataset. Several experiments were performed to assess the use of feature selection methods and technique to address dataset imbalance on diagnostic accuracy. The top performing model was obtained by logistic regression with excellent performance metrics (91% accuracy, 93% sensitivity, 85% specificity, 93% precision, 93% F1-score and 0.78 Matthews correlation coefficient). Feature importance was also generated for the best model. Explainable artificial intelligence method using LIME was applied to help understand the reasoning behind the model's classification of depression leading to better understanding of physicians, thus demonstrating its use in clinical practice.

Keywords: Depression prediction · Machine learning · Feature selection · Feature importance · LIME

1 Introduction

Depression, characterized by persistent sadness and a loss of interest in activities a person normally enjoys or accompanied by the inability to do usual daily activities for at least two weeks, very common nowadays affecting more than 300 million people globally [1]. It significantly affects the over-all well-being and functioning at school, family, and workplace often leading to self-harm or even suicide. With COVID-19 pandemic, depression has become even more pronounced as shown in the study by Rossi et al., indicating COVID-related stressful events to be associated with depression and anxiety symptoms in the Italian general population [2]. Depression has also been shown to be highly associated with numerous chronic diseases such as diabetes, heart disease caner, stroke, and chronic obstructive pulmonary disease [3]. Prompt recognition of the disease coupled with early professional intervention can significantly improve mental symptoms, resolve somatic problems such as gastrointestinal problems and sleeping disorders, thereby mitigating the negative implications for over-all well-being [4]. To assess depression, it is crucial to determine important contributing factors to plan the appropriate intervention. It is in this area of early diagnosis where machine learning

(ML) can be utilized, thus enhancing the whole diagnostic process leading to institution of the much-needed early intervention efforts and medical therapy.

Our objective is to predict depression using a variety of ML classification algorithms namely: Logistic Regression (LR), Naive Bayes (NB), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Adaptive Boosting (AdaBoost), and Extreme Gradient Boosting (XGBoost) evaluated on publicly available dataset. It is also our aim to determine the important features relevant to depression prediction and the logic employed by the classifiers to explain their prediction.

2   Literature Review

In the study by Grzenda et al. involving depressed 60 years and above, authors compared ML classifiers - SVM, RF, and LR on sociodemographic characteristics, baseline clinical self-reports, cognitive tests, and structural magnetic resonance imaging features to predict treatment outcomes in late-life depression [5]. RF obtained an area under receiver operating characteristic curve (AUROC) of 0.83 while SVM and LR recorded AUROC of 0.80 and 0.79, respectively. They also reported anterior and posterior cingulate volumes, depression characteristics, and self-reported health-related quality scores as the most important predictors of treatment response. Lin et al. [6], compared regression-based models (LR, lasso, ridge) and RF in depression forecasting among home-based elderly Chinese. Authors concluded that these models have good diagnostic performance in differentiating depression versus no depression. They reported life satisfaction, self-reported memory, cognitive ability, activities of daily living impairment to be the major determinants. In [7], authors applied XGBoost model to classify current depression versus no lifetime depression with a 0.86 AUROC. They further concluded that XGBoost and network analysis were useful to discover depression-related factors and their relationships and can be applied to epidemiological studies.

Sabab Zulfiker et al., applied six ML classifiers coupled with three feature selection methods and synthetic minority oversampling technique (SMOTE) to assess for presence of depression [8]. Their results showed AdaBoost with SelectKBest feature selection technique to be the best performing model with a 92.56% accuracy rate. Nemesure et al. [9], applied a novel ensemble of ML models (SVM, kNN, LR, RF, XGBoost, and neural network (NN)) to predict depression and Generalized Anxiety Disorder (GAD) with moderate predictive performance (AUROC of 0.73 for GAD and 0.60 for depression). Shapley Additive Explanations (SHAP) was used to generate feature importance.

Sousa et al. [10], determine predictors of depression and reported that sex, living status, mobility, and nutritional status appear to be the important factors to be associated with depression. They concluded that these important predictors would be crucial for prevention and for customization of interventions. In the study by Richter et al., evaluated several ML-based approaches that use behavioral data in the classification of depression and other psychiatric disorders. Authors classified these studies into laboratory-based assessments and data mining which was further divided into (a) social media usage and movement sensors data and (b) demographic and clinical information. Authors summarized the benefits and constraints and suggested future research directions to develop interventions and individually tailored treatments in the future [11].

In the study by Vincent et al. [12], they used a multilayered neural perceptron (MLP) and experimented with the backpropagation technique to assess for depression involving data collected from IT professionals. Authors reported that deep-MLP with backpropagation outperforms other machine learning-based models for effective classification of depression with a 98.8% accuracy. Jan et al., reviewed several ML algorithms for diagnosis of bipolar disorders [13]. Their survey identified 18 classification models, five regression models, two model-based clustering methods, one natural language processing, one clustering algorithms and three deep learning -based models. Magnetic resonance imaging data were mostly used for classifying bipolar patients whereas microarray expression data sets and genomic data were the least commonly used.

3   Methodology

In our research, the first step is the loading of the dataset. This is to be followed by pre-processing steps which include data cleaning, dataset normalization, feature selection techniques to select important predictors, and addressing data imbalance. We then applied various ML algorithms followed by assessment of their performance using accuracy, precision, sensitivity/recall, specificity, F1-scores, and Matthews correlation coefficient. Feature importance and AI explainability assessment were also done. The pipeline for this study is seen in Fig. 1.
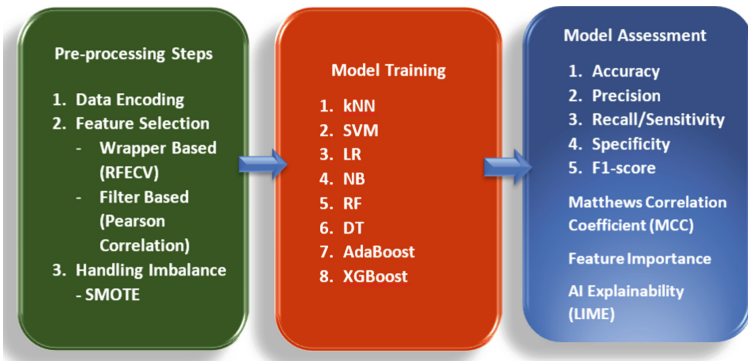

Fig. 1. Machine learning pipeline for depression prediction

3.1   Dataset Description

We used a publicly available depression dataset from github [14]. The dataset contains 604 instances involving 455:149 male–female sex ratio, with 30 predictor variables and 1 target variable (depressed or not) based on Burns Depression Checklist. The description of these attributes is shown in Table 1.

Table 1. Description of attributes of depression

Attribute	Attribute	
Age in years (AGERNG)	Physical Exercise (PHYEX)	Felt Cheated (CHEAT)
Gender	Smoker (SMOKE)	Faced threat (THREAT)
Educational Attainment (EDU)	Alcohol Drinker (DRINK)	Felt Abused (ABUSED)
Profession (PROF)	With Illness (ILLNESS)	Lost someone (LOST)
Marital Status (MARSTS)	Has Insomnia (INSOM)	Has Work/Study Pressure (WRKPRE)
Type of Residence (RESDPL)	Has Eating disorder (EATDIS)	Inferiority Complex (INFER)
Lives with Family or not (LIVWTH)	Average sleep hours (AVGSLP)	Suicidal thoughts (SUICIDE)
Satisfied with Environment or not (ENVSAT)	Taking Prescribed Med (PREMED)	In Conflict with Family or Friends (CONFLICT)
Satisfied with current position or achievements or not (POSSAT)	Ave hours in social network (TSSN)	Depressed
Financial stress (FINSTR)	Feels anxiety (ANXI)	
Had Debt (DEBT)	Feels deprived (DEPRI)	

3.2 Pre-processing Steps

Pre-processing methods were applied to the dataset in preparation for ML training. There were no missing values but there were 10 duplicate records which were promptly removed. It also shows mild data imbalance with 391 (65.82%) with depression and 203 (34.18%) without depression. We performed data encoding for the attributes and feature scaling with normalization using the StandardScaler function of scikit-learn library. All categorical predictors were dummified resulting to an increase in the number of columns. For feature selection procedure, we applied and compared a wrapper method using recursive feature elimination with cross validation (RFE-CV) and a filter method using Pearson correlation. In our study, we used a threshold correlation with the target variable of > 0.20 and a correlation between predictors of less than 0.80. As the dataset is imbalanced, we applied Synthetic Minority Over-sampling TEchnique (SMOTE). The correlation heatmap is seen in Fig. 2.

3.3 Machine Learning Models

The dummified dataset was divided into 30% testing involving 179 records and 70% training involving 415 records with tenfold cross validation. We utilized python 3.8 and its various machine learning libraries (scikit-learn, keras, tensorflow, pandas, Matplotlib, seaborn, NumPy, and LIME) in our experiment. The models tested were LR, NB, kNN, SVM, DT, RF, AdaBoost, and XGBoost. Hyperparameter tuning was performed on each

ML model. To determine the best performing model, Matthews correlation coefficient (MCC) was used.

3.4 Feature Importance and Model Explainability

For the best performing models, we generated the feature importance scores to determine the most important attributes relevant to depression prediction. To understand the local behavior of the model for a single instance of a patient with or without depresssion, we applied Local Interpretable Model-agnostic Explanations (LIME). LIME is used to explain individual prediction of a black-box machine learning model.

Fig. 2. Correlation heatmap of predictor variables for depression

4   Results and Analysis

The performance metrics of the 8 ML models for our dataset are shown in Table 2 where the effects of feature selection method are assessed. LR is the best performing model when there is no feature selection technique used as well as when Pearson correlation is used as a feature selection with accuracy rates of 91% and 89%, respectively. For Pearson correlation, a mild increase in the accuracy rate ranging from 1- 4% is seen for DT, RF, NB, kNN and SVM while a slight decrease of 2–5% is noted for LR, AdaBoost and XGBoost. Nonetheless, LR still remains to be the top model when Pearson correlation method is used as a feature selection. On the other hand, the top model for the RFE-CV feature selection is XGBoost with 85% accuracy. When RFE-CV is applied, generally there is a decrease in accuracy ranging from 3%-7% for most of the models while the

rest of the models (DT, NB, kNN) did not show any significant changes. Overall, after considering the effects of feature selection, LR with no feature selection is the best performing model as it obtained the highest MCC score at 0.78, and 91% accuracy. Hence in this dataset, it appears that all attributes seem to be important in depression prediction and no attributes need to be eliminated.

Table 2. Performance metrics for predicting depression – assessment of feature selection

	ML Model	Accuracy	Precision	Sensitivity	Specificity	F-score	MCC
No Feature	DT	0.82	0.91	0.81	0.83	0.85	0.61
Selection	LR	0.91	0.93	0.93	0.85	0.93	0.78
	RF	0.87	0.89	0.92	0.78	0.91	0.71
	NB	0.84	0.95	0.82	0.92	0.88	0.69
	kNN	0.83	0.99	0.76	0.98	0.86	0.70
	SVM	0.87	0.88	0.93	0.75	0.91	0.70
	AdaBoost	0.87	0.89	0.91	0.78	0.90	0.69
	XGBoost	0.89	0.91	0.93	0.81	0.92	0.76
Pearson	DT	0.83	0.82	0.95	0.58	0.88	0.59
Correlation	LR	0.89	0.91	0.93	0.81	0.92	0.76
	RF	0.85	0.91	0.86	0.83	0.88	0.67
	NB	0.87	0.94	0.85	0.90	0.89	0.72
	kNN	0.87	0.93	0.87	0.86	0.90	0.71
	SVM	0.88	0.90	0.93	0.80	0.91	0.73
	AdaBoost	0.83	0.89	0.85	0.80	0.87	0.63
	XGBoost	0.84	0.87	0.90	0.73	0.89	0.64
RFE-CV	DT	0.83	0.91	0.82	0.83	0.76	0.63
	LR	0.84	0.90	0.87	0.80	0.88	0.65
	RF	0.84	0.90	0.87	0.80	0.88	0.65
	NB	0.84	0.90	0.86	0.80	0.88	0.64
	kNN	0.83	0.89	0.85	0.80	0.87	0.63
	SVM	0.83	0.91	0.82	0.83	0.86	0.63
	AdaBoost	0.80	0.86	0.84	0.71	0.85	0.55
	XGBoost	0.85	0.89	0.90	0.76	0.89	0.67

To address the issue of imbalance, SMOTE was applied to the dataset. Assessment of SMOTE for the feature selection methods is seen in Table 3. The application of SMOTE to the dataset when there is no feature selection resulted to a decrease in the accuracy rate with a range of 3%-18% for most models. The only model which posted a slight increase (2%) in the accuracy is NB while there was no change for RF and AdaBoost. Nevertheless, LR obtained the highest accuracy and MCC at 84% and 0.74 respectively. The application of SMOTE to the dataset when Pearson correlation was used as a feature selection generally resulted to a very small decrease in the accuracy (1%-4%) for most

Table 3. Comparative Performance Metrics of ML Models with and without SMOTE

	ML Model	Accuracy	Precision	Sensitivity	Specificity	F-score	MCC
No Feature Selection plus SMOTE	DT	0.64	0.97	0.47	0.97	0.64	0.44
	LR	0.88	0.92	0.91	0.83	0.91	0.74
	RF	0.87	0.90	0.91	0.80	0.90	0.71
	NB	0.86	0.93	0.86	0.86	0.89	0.70
	kNN	0.74	0.91	0.68	0.86	0.78	0.51
	SVM	0.83	0.91	0.83	0.83	0.87	0.64
	AdaBoost	0.87	0.89	0.74	0.81	0.81	0.53
	XGBoost	0.85	0.89	0.89	0.78	0.89	0.67
Pearson Correlation plus SMOTE	DT	0.83	0.89	0.85	0.78	0.87	0.62
	LR	0.89	0.94	0.90	0.88	0.92	0.77
	RF	0.86	0.92	0.87	0.85	0.89	0.70
	NB	0.85	0.94	0.83	0.90	0.88	0.70
	kNN	0.86	0.95	0.83	0.92	0.89	0.71
	SVM	0.85	0.91	0.86	0.83	0.88	0.70
	AdaBoost	0.83	0.89	0.85	0.80	0.87	0.67
	XGBoost	0.88	0.92	0.90	0.83	0.91	0.72
RFE-CV plus SMOTE	DT	0.83	0.91	0.82	0.83	0.86	0.63
	LR	0.84	0.90	0.86	0.80	0.88	0.64
	RF	0.84	0.90	0.87	0.80	0.88	0.65
	NB	0.80	0.91	0.78	0.85	0.84	0.60
	kNN	0.83	0.91	0.82	0.83	0.86	0.63
	SVM	0.84	0.90	0.86	0.80	0.88	0.64
	AdaBoost	0.83	0.91	0.82	0.83	0.86	0.63
	XGBoost	0.84	0.90	0.87	0.80	0.88	0.65

models while a small increase of 1% is seen for RF. No change in the accuracy was noted for DT, LR and AdaBoost. For the case when SMOTE was applied for RFE-CV, there were no significant changes in the accuracy rates across all models – a very slight increase of 1%-3% for SVM and AdaBoost, a decrease of 1%-4% for NB and XGBoost, while the rest of the models had no changes. Overall, LR posted the highest accuracy and MCC at 89% and 0.76, respectively for this experiment assessing the effects of SMOTE.

Table 4 highlights the confusion matrix of the best performing models for the six experiments (no feature selection (FS), with Pearson correlation, with RFE-CV, without SMOTE and with SMOTE). Comparative performance of the best models is also shown in Fig. 3. It can be deduced that the performance of the six models in the experiments seem to be similar or comparable to each other across all metrics. This suggests that for this particular dataset, we may or may not do feature selection method nor may or may not apply SMOTE to address imbalance. Nonetheless, the overall best performing model is LR without any feature selection method and with no SMOTE.

Table 4. Confusion Matrix of the best Performing ML Models in various Experiments

	LR (No FS, No SMOTE)		LR (With Pearson Correlation, No SMOTE)		XGBoost (RFE-CV, No SMOTE)		LR (No FS, With SMOTE)		LR (With Pearson Correlation, with SMOTE)		RF/XGBoost (RFE-CV, With SMOTE)	
Actual	Predicted +	-	Predicted +	-	Predicted +	-	Predicted +	-	Predicted +	-	Predicted +	-
With depression	[[50	9]	[[48	11]	[[45	14]	[[49	10]	[[52	7]	[[47	12]
No depression	[8	112]]	[8	112]]	[12	108]]	[11	109]]	[12	108]]	[16	104]]
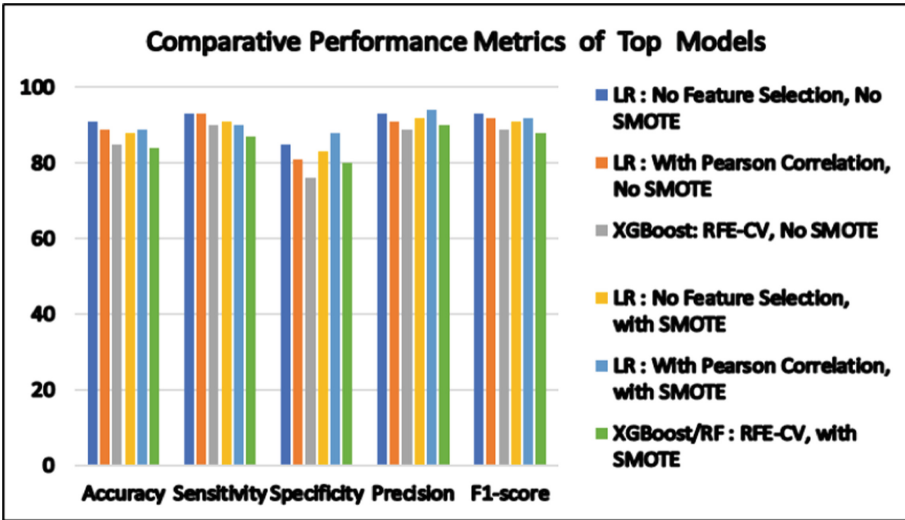
Fig. 3. Performance metrics of best models for depression prediction

The feature importance of the attributes of LR is seen in Fig. 4. The most important features relevant to depression prediction are ANXI (feels anxiety), DEPRI (feels deprived), POSSAT (satisfied or not with current position/achievement), INFER (inferiority complex) and ENVSAT (satisfied or not with environment). These features are also in consonance with clinical assessment of depression.

For the explainable AI part of this research, we used LIME which is a technique that approximates any black box machine learning model with a local, interpretable model to explain each individual prediction. LIME is model agnostic hence can give explanations for any supervised machine learning model. To illustrate how LIME works, we randomly selected two patients the first one without depression while the second one has depression. Let us take the case of our first patient diagnosed as having no depression and was correctly classified by LR to be 0 or "not depressed" as illustrated in Fig. 5. The LIME output in Fig. 5 consists of 3 parts: left, center, and right. The left shows the classification predicted by LR which in this case is 0 or "not depressed" and with a confidence of 90%. The center shows the features that influenced the classification. For

this patient, LIME was able to generate the important features for LR to arrive at the classification "no depression" and these are: patient has no anxiety (ANXI_Yes = 0), has no inferiority complex (INFER_Yes = 0), has no suicidal thoughts (SUICIDE_Yes = 0), has not recently lost someone close to him (LOST_YES = 0), was not in conflict with family or friends (CONFLICT_Yes = 0), was not physically, sexually, or emotionally abused (ABUSED_YES = 0), never felt cheated by someone recently (CHEAT_YES = 0), and average sleep of was not 8 h (AVGSLP_8 = 0). Note that there are also feature values that are leaning towards "depression" for this particular case which are: not satisfied with his current position or achievements (POSSAT_YES = 0) and felt deprived of something he deserves (DEPRI_YES = 1). However, the effects of these two features are not enough to oppose the effects of the other features contributing to a "no depression" classification. The rightmost part of the LIME output shows the actual values of the first 10 most important features for this patient. LIME can be an effective tool to explain the logic by the model to arrive at the prediction.

For the second patient who was diagnosed as "depressed", the LIME output in Fig. 6 shows the classification predicted by LR which in this case was 1 or "depressed" and with a confidence of 100%. The center shows the features that influenced the "depressed" classification which are: patient is not satisfied with current position or achievements (POSSAT_True = 0), has anxiety (ANXI_Yes = 1), felt deprived of something he deserves (DEPRI_YES = 1), has inferiority complex (INFER_Yes = 1), and is undergoing financial stress (FINSTR_True = 1). Note that there are also values contributing to "no depression" for this particular case which are: did not lost someone (LOST_Yes = 0), nor felt abused (ABUSED_Yes = 0), nor cheated (CHEAT_Yes = 0), is not in conflict with family or friends (CONFLICT_Yes = 0) nor threatened (THREAT_Yes = 0). However, the effects of these features are not enough to oppose the effects of the features contributing to a "depression" classification. The top features that influenced the "depression" classification for this patient is in agreement with the top features selected by LR as most influential to a "depression" classification as seen in Fig. 3. The explainability feature of LIME can help health professionals understand and interpret classifier's prediction leading to increased trust in the use of these methods.

In our study, we applied a filter method using Pearson correlation with the target variable (presence of depression) and among predictor variables. Feature selection aims to remove redundant features which can be expressed by other attributes and irrelevant features which do not contribute to the performance of the model in predicting depression [15]. RFECV reduces model complexity by removing attributes one at a time until it automatically finds an optimal number of features based on the cross-validation score of the model [16, 17]. It is a commonly used due to its ease of use. Using the associated feature weights, those attributes with small feature weights close to zero contribute very little to predicting depression. But we must take note that removing a single attribute would also lead to a change in the feature weights, which suggest that elimination of the features should be done in a stepwise fashion. On the other hand, pairwise correlation identifies highly correlated features and keeps only one of them to achieve predictive power using few features as possible since highly correlated features bring no new information to the dataset. These highly correlated features only increase model complexity, increase the chance of overfitting, and require more computations [18, 19].

SMOTE is an oversampling method that creates artificial minority data points within the cluster of minority class samples in a balanced way which render it to be an effective method in reducing negative effects of imbalance leading to increased performance [8, 20–24]. It works by utilizing a kNN algorithm to create synthetic data by first selecting a random data from the minority (no depression) class and then kNN from the data are set. That is, synthetic data is created between the random data and the randomly selected kNN. As such, there is not only an increase in the number of datapoints but an increase in its variety. However, SMOTE has its disadvantages such as sample overlapping, noise interference and blindness of neighbor selection as well as their suitability for clinical datasets [22, 25, 26].

Feature importance allows us to detect which features in our depression dataset have predictive power by assigning a score to each feature based on its ability to improve predictions and allow us to rank these features. The increase in the model prediction error after permuting the values of that feature determines its feature importance. An increase in the model error also increases the importance of that feature for predicting depression, while if the accuracy of the model remains the same or slightly decreases, then the feature is deemed unimportant for depression prediction [27–29]. However, this method has also some disadvantages such as prohibitive computational cost and cannot be used as a substitute for statistical inference [30].
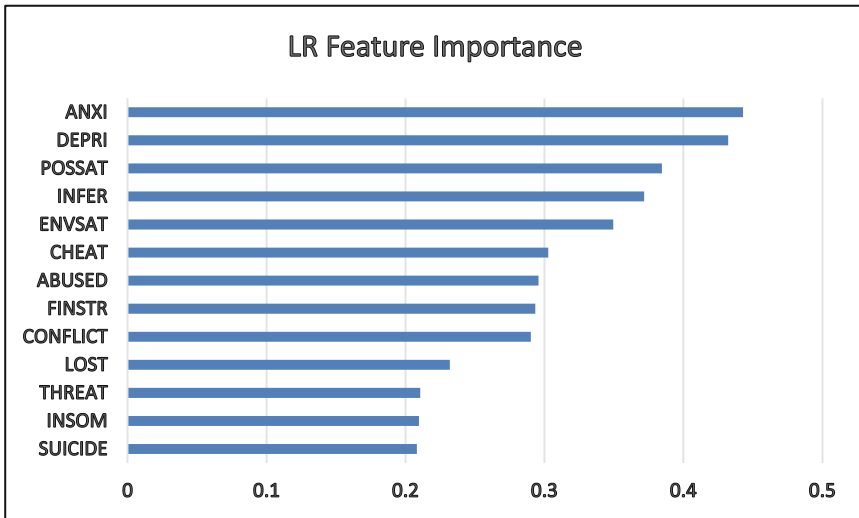

Fig. 4. Feature importance of LR model for depression prediction

Our results are comparable with the other studies [5, 6, 9, 13] in the literature with respect to depression prediction. Our top performing models have very good sensitivity and specificity rates allowing the mental health professionals to use these models as a screening tool for depression in their clinical practice. Additionally, we highlighted the importance of utilizing LIME as an XAI tool in depression prediction. In [31], authors have validated the use of their XAI-ASD in improving diagnostic performance
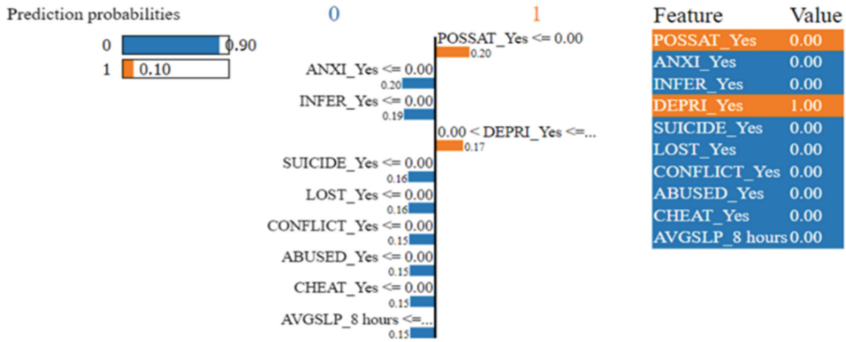
Fig. 5. A sample of feature explainability for a correctly classified patient without depression by logistic regression
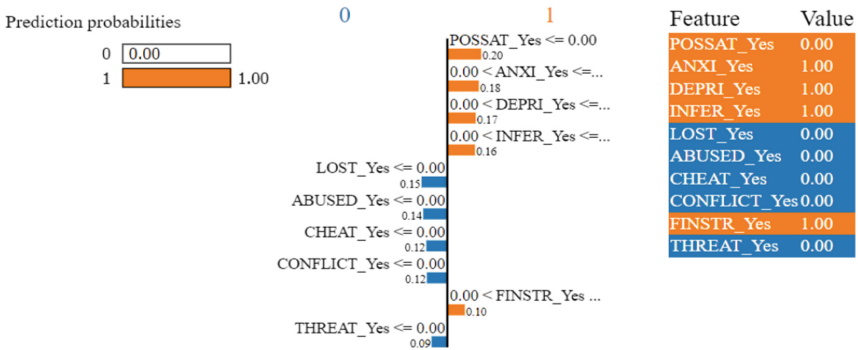

Fig. 6. A sample of model explainability for a correctly classified patient with depression by logistic regression

in predicting presence of depression and reported that explainability allows humans to appropriately understand and trust the emerging AI phenomenon. It has brought the machines closer to humans because of its capability to explain the logic behind the diagnosis. It needs to be emphasized that insufficient explainability and transparency in most existing AI systems seem to be a major reason for unsuccessful implementation and integration of AI tools in the routine clinical practice. Our findings suggest the utility of XAI models to make a diagnosis of depression with acceptable results. The clinical relevance of our experiment is even more highlighted with XAI models that can provide faster and with high reliability to help physicians in the screening of patients for depression. An early accurate diagnosis leading to prompt intervention efforts is very crucial to improve patient's quality of life, diminished risk for developing chronic diseases, improve productivity and prevention of suicide cases [5, 8, 24, 32]. This research, thus, provided useful insights in the development of an automated models that can assist healthcare workers in the assessment of depressive disorders.

5 Conclusion

Depression is debilitating disease that leaves individuals persistently feeling sad or hopeless for more than two weeks affecting more than 300 million people globally. We applied several machine learning models with model explainability to a publicly available depression dataset. After a series of experiments to assess the effects of the use of feature selection methods and the technique to address dataset imbalance, the best performing model was logistic regression (LR) with a 91% accuracy, 93% sensitivity, 85% specificity, 93% recall, 93% F1-score and 0.78 Matthews correlation coefficient. Feature importance identified the most important attributes necessary to make a depression classification are also in consonance with clinical assessment of depression. LIME method provided tools to visualize the reasoning behind the classification of depression by the machine learning model for better understanding of physicians. Incorporation of XAI tools in clinical practice can further enhance the diagnostic acumen of health professionals. The primary limitation of our research is the use of small datasets due unavailability of large and open-source depression datasets.

Future enhancement of this study should focus on inclusion of other tools for feature importance as well as techniques in XAI such as SHAP for better understanding of the models by healthcare providers. Moreover, mixed type of datasets combining symptoms with neuroimaging features seen in functional magnetic resonance imaging can also be explored to generate more superior diagnostic accuracy. Our findings are promising and have generated useful insights in the development of automated models that are faster and with high reliability which can be of use to physicians in predicting depression. Nonetheless, early intervention efforts and treatment for depression ensure the best quality of care for our patients.

References

1. World Health Organization: Mental Health and Substance Abuse (2021). http://www.emro. who.int/mnh/what-you-can-do/index.html#accordionpan4 Last accessed 10 Jan 2022
2. Rossi, R., Jannini, T.B., Socci, V., Pacitti, F., Lorenzo, G.D.: Stressful life events and resilience during the COVID-19 lockdown measures in italy: association with mental health outcomes and age. Frontiers in Psychiatry **12**, 635832 (2021). https://doi.org/10.3389/fpsyt.2021. 635832
3. Li, H., Ge, S., Greene, B., Dunbar-Jacob, J.: Depression in the Context of Chronic Disease in the United States and China. Int. J. Nurs. Sci. **6**(1), 117–122 (2019). https://doi.org/10.1016/ j.ijnss.2018.11.007
4. Uddin, M.Z., Dysthe, K.K., Følstad, A., Brandtzaeg, P.B.: Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Comp. Appl. **34**, 721–744 (2022). https://doi.org/10.1007/s00521-021-06426-4
5. Grzenda, A., Speier, W., Siddarth, P., Pant, A., Krause-Sorio, B., Narr, K., Lavretsky, H.: Machine learning prediction of treatment outcome in late-life depression. Frontiers in Psychiatry **12** (2021). https://doi.org/10.3389/fpsyt.2021.738494
6. Lin, S., Wu, Y., Fang, Y.: Comparison of regression and machine learning methods in depression forecasting among home-based elderly chinese: a community based study. Frontiers in psychiatry **12**, 764806 (2022). https://doi.org/10.3389/fpsyt.2021.764806

7. Nam, S.M., Peterson, T.A., Seo, K.Y., Han, H.W., Kang, J.I.: Discovery of depression-associated factors from a nationwide population-based survey: epidemiological study using machine learning and network analysis. J. Medi. Intern. Res. **23**(6), e27344 (2021). https://doi.org/10.2196/27344

8. Sabab Zulfiker, M., Kabir, N., Biswas, A.A., Nazneen, T., Shorif Uddin, M.: An in-depth analysis of machine learning approaches to predict depression. Curr. Res. Behavi. Sci. **2**, 100044 (2021). https://doi.org/10.1016/j.crbeha.2021.100044

9. Nemesure, M.D., Heinz, M.V., Huang, R., Jacobson, N.: Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Scientific Reports **11**, 1980 (2021). https://doi.org/10.1038/s41598-021-81368-4

10. Sousa, S., Paúl, C., Teixeira, L.: Predictors of major depressive disorder in older people. Int. J. Environm. Res. Pub. Health **18**, 11894 (2021). https://doi.org/10.3390/ijerph182211894

11. Richter, T., Fishbain, B., Richter-Levin, G., Okon-Singer, H.: Machine Learning-Based Behavioral Diagnostic Tools for Depression: Advances, Challenges, and Future Directions. J. Personal. Medi. **11**, 957 (2021). https://doi.org/10.3390/jpm11100957

12. Vincent, P., Mahendran, N., Nebhen, J., Deepa, N., Srinivasan, K., Hu, Y.C.: Performance assessment of certain machine learning models for predicting the major depressive disorder among IT professionals during pandemic times. Computational intelligence and neuroscience **2021**, 9950332 (2021). https://doi.org/10.1155/2021/9950332

13. Jan, Z., et al.: The role of machine learning in diagnosing bipolar disorder: scoping review. J. Medi. Intern. Res. **23**(11), e29749 (2021). https://doi.org/10.2196/29749

14. Sabab31/Depression-Repository: https://github.com/Sabab31/Depression-Repository.git last accessed 10 Nov 2021

15. Demircioğlu, A.: Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. Insights Imaging **12**, 172 (2021). https://doi.org/10.1186/s13244-021-01115-1

16. Chang, W., Ji, X., Wang, L., Liu, H., Zhang, Y., Chen, B., Zhou, S.: A machine-learning method of predicting vital capacity plateau value for ventilatory Pump failure based on data mining. Healthcare **9**, 1306 (2021). https://doi.org/10.3390/healthcare9101306

17. Li, D., et al.: Application of machine learning classifier to candida auris drug resistance analysis. Frontiers in Cellular and Infection Microbiology **11** (2021). https://doi.org/10.3389/fcimb.2021.742062

18. Rieta, J.J., Senan, E.M., Abunadi, I., Jadhav, M., Fati, S.M.: Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms. Computational and Mathematical Methods in Medicine **2021**, article 8500314 (2021). https://doi.org/10.1155/2021/8500314

19. Magboo, V.P.C., Magboo, M.S.A.: Machine learning classifiers on breast cancer recurrences. Procedia Computer Science **192**, 2742–2752 (2021). https://doi.org/10.1016/j.procs.2021.09.044

20. Çakır, H., İncereis, N., Akgün, B.T., Taştemir, A.S.Y.: Comparison of sampling methods using machine learning and deep learning algorithms with an imbalanced data set for the prevention of violence against physicians. In: 2021 15th Turkish National Software Engineering Symposium (UYMS), pp. 1–7 (2021). https://doi.org/10.1109/UYMS54260.2021.9659758

21. Huang, C.Y., Dai, H.L.: Learning from class-imbalanced data: review of data driven methods and algorithm driven methods. Data Sci. Fina. Econo. **1**(1), 21–36 (2021). https://doi.org/10.3934/DSFE.2021002

22. Wang, S., Dai, Y., Shen, J., Xuan, J.: Research on expansion and classification of imbalanced data based on SMOTE algorithm. Scientific Reports **11**, 24039 (2021). https://doi.org/10.1038/s41598-021-03430-5

23. Risi, M., Wang, J.B., Zou, C.A., Fu, G.H.: AWSMOTE: An SVM-based adaptive weighted SMOTE for class-imbalance learning. scientific programming **2021**, article 9947621 (2021). https://doi.org/10.1155/2021/9947621

24. Magboo, V.P.C., Magboo, M.S.A.: Imputation techniques and recursive feature elimination in machine learning applied to type II diabetes classification. In: 2021 4th Artificial Intelligence and Cloud Computing Conference (AICCC '21), pp. 201-207. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3508259.3508288

25. Jiang, Z., Pan, T., Zhang, C., Yang, J.: A new oversampling method based on the classification contribution degree. Symmetry **13**, 194 (2021). https://doi.org/10.3390/sym13020194

26. Beinecke, J., Heider, D.: Gaussian noise up-sampling is better suited than SMOTE and ADASYN for clinical decision making. BioData Mining **14**, 49 (2021). https://doi.org/10.1186/s13040-021-00283-6

27. Ljubobratovic, D., Vukovic, M., Brkic Bakaric, M., Jemric, T., Matetic, M.: Utilization of explainable machine learning algorithms for determination of important features in 'Suncrest' peach maturity prediction. Electronics **10**, 3115 (2021). https://doi.org/10.3390/electronics10243115

28. Mi, X., Zou, B., Zou, F., Hu, J.: Permutation-based identification of important biomarkers for complex diseases via machine learning models. Nature Communications **12**, 3008 (2021). https://doi.org/10.1038/s41467-021-22756-2

29. Inglis, A., Parnell, A., Hurley, C.: Visualizing variable importance and variable interaction effects in machine learning models. J. Compu. Graphi. Statis. https://doi.org/10.1080/10618600.2021.2007935

30. Oh, S.: Predictive case-based feature importance and interaction. Information Sciences **593**, 155–176 (2022). https://doi.org/10.1016/j.ins.2022.02.003

31. Uddin, M.Z., et al.: Deep Learning for prediction of depressive symptoms in a large textual dataset. Neural Comp. Appl. **34**, 721–744 (2022). https://doi.org/10.1007/s00521-021-06426-4

32. Magboo, V.P.C., Abu, P.A.R.: Deep neural network for diagnosis of bone metastasis. In: 2022 The 5th International Conference on Software Engineering and Information Management (ICSIM) (ICSIM 2022), pp. 144–151. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3520084.3520107