Advances in Machine Learning and Explainable Artificial Intelligence for Depression Prediction

Haewon Byeon
Department of Medical Big Data, College of AI Convergence, Inje University,
Gimhae 50834, Gyeongsangnamdo, South Korea

***Abstract*—There is a growing interest in applying AI technology in the field of mental health, particularly as an alternative to complement the limitations of human analysis, judgment, and accessibility in mental health assessments and treatments. The current mental health treatment service faces a gap in which individuals who need help are not receiving it due to negative perceptions of mental health treatment, lack of professional manpower, and physical accessibility limitations. To overcome these difficulties, there is a growing need for a new approach, and AI technology is being explored as a potential solution. Explainable artificial intelligence (X-AI) with both accuracy and interpretability technology can help improve the accuracy of expert decision-making, increase the accessibility of mental health services, and solve the psychological problems of high-risk groups of depression. In this review, we examine the current use of X-AI technology in mental health assessments for depression. As a result of reviewing 6 studies that used X-AI to discriminate high-risk groups of depression, various algorithms such as SHAP (SHapley Additive exPlanations) and Local Interpretable Model-Agnostic Explanation (LIME) were used for predicting depression. In the field of psychiatry, such as predicting depression, it is crucial to ensure AI prediction justifications are clear and transparent. Therefore, ensuring interpretability of AI models will be important in future research.**

Keywords—Depression; LIME; Explainable artificial intelligence; Machine learning; SHAP

I. Introduction

Artificial Intelligence (AI) refers to computer technology that mimics human intelligence by using logical methods to reason, learn, and make decisions. With the advancement of AI technologies such as machine learning, deep learning, and natural language processing, this technology is being applied not only in professional fields but also in our daily lives. For instance, virtual assistants such as Apple's Siri, Samsung's Bixby, and Google's Assistant use natural language processing technology to provide convenience to our lives [1].

There is a growing interest in applying AI technology in the field of mental health, particularly as an alternative to complement the limitations of human analysis, judgment, and accessibility in mental health assessments and treatments [2,3]. Traditional mental health assessments rely heavily on the subjective self-reports and interviews of patients, leading to potential inaccuracies in expert decision-making [4,5]. The misdiagnosis rate of mental health conditions (e.g. bipolar disorder) can be as high as 55~76% [6,7], even among experts who have difficulty grasping the symptoms and information not reported by the patient.

The current mental health treatment service faces a gap where people who need help do not receive it due to negative perceptions of mental health treatment, lack of professional manpower and physical accessibility limitations [8,9,10]. There is a growing need for a new approach to overcome these difficulties, and AI technology is being explored as a potential solution.

In particular, explainable artificial intelligence (X-AI) with both accuracy and interpretability technology can help improve the accuracy of expert decision-making, increase the accessibility of mental health services, and solve the psychological problems of high-risk groups of depression [11]. In this mini-review, we examine the current use of X-AI technology in mental health assessments for depression. Additionally, we discuss the measures and limitations of applying AI to mental health services.

II. Materials and Methods

A. Artificial Intelligence and Machine Learning in Depressive Disorder

Depressive Disorder is a severe psychiatric disorder that results in functional impairment [12]. Currently, the diagnosis of depressive disorder relies on the identification of a minimum number of core symptoms that cause functional impairment over a certain period of time [12]. However, this symptom-based approach can lead to diagnostic discrepancies and make it challenging to interpret the results of additional studies, such as genetic studies, neuroimaging studies, and postmortem studies.

The early detection and diagnosis of subtle clinical signs in depressive disorder require highly skilled professionals working in specialized mental health services. Hence, using more objective and reliable techniques, such as neuroimaging techniques, can aid in early detection. Machine learning has the potential to make accurate diagnoses and predict the response to treatment, beyond the conventional method of comparative analysis between a patient group and a normal control group.

Artificial intelligence was first introduced at the Dartmouth Conference in 1956 by Professor John McCarthy of Dartmouth University in the US [13]. At the technological level, it refers to Narrow Artificial Intelligence (NAI), which can perform certain tasks with better-than-human capabilities [14]. Machine learning is a specific approach to implementing AI, in which a computer learns how to perform a task through an algorithm, rather than having specific decision criteria inputted directly by humans. In this process, defining appropriate features is critical to machine learning, and various algorithms, such as the Support Vector Machine, Gaussian Process Classifier, Linear Discriminant Analysis, and Decision Tree are used.

Deep learning, a branch of machine learning, goes further by using given data as input [15]. This end-to-end machine learning reduces errors that can occur due to human intervention, but the quality and quantity of data provided for learning is becoming increasingly important [13]. Therefore, obtaining high quality data for AI learning is becoming more important than the algorithms used.

B. Advances in Machine Learning in Neuroimaging

Brain imaging studies can be classified into structural and functional studies. Various studies have reported the use of machine learning techniques to predict the onset of depression. Conventional structural brain imaging studies, which compare patients with major depressive disorder (MDD) and healthy controls, often use T1-weighted images, which provide high contrast between gray matter and white matter, allowing for more accurate viewing of gray matter regions that make up the cortex. However, MDD is a complex disorder with diverse symptoms, and neuroanatomical abnormalities in MDD are not limited to morphological changes in a single local area. T2-weighted imaging and diffusion tensor images are other neuroimaging techniques used to study structure. Meanwhile, functional aspects can be studied using fMRI (functional Magnetic Resonance Imaging). There are various methods used in fMRI research to predict the diagnosis of depression, such as task-related fMRI and resting fMRI. However, studies [16,17,18] on discrimination of depression using neuroimaging techniques have shown accuracy errors that vary based on sample size. Flint et al. (2021)[16] found that a study with a small sample size (n = 20) demonstrated higher accuracy than one with a medium sample size (n = 100), while a study with a large sample size (n = 1,868) showed an accuracy of only 61%. The authors emphasized the importance of considering the impact of test set size on systematic misestimation and why an overestimation effect may occur. Therefore, researchers should not disregard their models solely based on low training data, instead they should test the models on a larger set of data to assess its performance if it exhibits good results.

C. Advances in Machine Learning in Psychological Assessment

Depression is diagnosed through a structured interview, which sets it apart from many other diseases. The Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (published by the American Psychiatric Association)[19] provides diagnostic criteria that are widely used across the globe. These criteria are updated periodically by the APA. One of the main features of these criteria is that they rely solely on interviews with patients and psychological assessments. For instance, Table I displays the diagnostic criteria for major depressive disorder.

TABLE I. INSTANCE OF CRITERIA FOR MAJOR DEPRESSIVE DISORDER IN THE DSM-5

Criteria	
A	If five (or more) of the following symptoms persist for two consecutive weeks and show a change from previous functional status, at least one of the symptoms must be (1) depressed mood or (2) loss of interest or pleasure. Note that symptoms due to other apparent medical conditions should not be included
1	Depressed mood most of the day and nearly every day, subjectively reported (e.g., feeling sad, empty, or hopeless) or objectively observed (e.g., tearing); note that in children and adolescents, it may present as irritable mood.
2	Significantly diminished interest or pleasure in almost all of the usual activities nearly every day.
3	Significant weight loss (e.g., weight change of 5% or more in one month) or weight gain, or decrease or increase in appetite almost every day, without weight control; note that in children, weight gain should not exceed expectations.
4	Insomnia or hypersomnia nearly every day.
5	Psychomotor agitation or retardation nearly every day, observed objectively, not just subjective feelings of restlessness or stagnation.
6	Fatigue or loss of energy nearly every day.
7	Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) almost every day, not just remorse or guilt.
8	Decreased ability to think or concentrate, or indecisiveness nearly every day, either subjectively or objectively observable.
9	Recurrent thoughts of death (not just fear of dying), recurrent suicidal thoughts without a specific plan, or a suicide attempt or specific plan to commit suicide.

This definition of major depressive disorder makes it difficult to properly differentiate whether someone is exaggerating their symptoms using these criteria or, conversely, minimizing symptoms to avoid social stigma and prejudice as a person with a mental illness [12]. Furthermore, many risk indicators for depression have been presented through numerous studies so far, but no single risk indicator can accurately diagnose or classify depression [20,21]. This is because depression is not caused by a single factor but develops through various genetic and environmental interactions. Therefore, in order to diagnose depression clearly, it is necessary to consider the importance and influence of various risk factors in one model, which should include not only the results of face-to-face counseling but also various environmental and biological results. To overcome these limitations, several studies [22,23,24] have attempted to predict depressive disorder using machine learning.

D. Limitations of Machine Learning in Diagnosis

Studies on machine learning in the diagnosis of depressive disorder [22,23,24] have been ongoing for more than 10 years, and accuracy, sensitivity, and specificity are used to evaluate these models. Accuracy has been reported to range from the high 60% to the mid-80%, while sensitivity and specificity have been reported to be in the high 70-80% range. However, when applying machine learning theories to actual clinical practice, several problems arise that prevent its application, such as the heterogeneity of various image data, which arises from data collection, acquisition parameters, and post-processing methods. This makes it challenging to generalize the results to other data and compare.

E. Advancement of Decision Tree-based Ensemble Model Techniques for Depression Prediction and SHAP

As a result of the efforts of many researchers to create ML models with high accuracy and reproducibility over the past decade, ML models have evolved into ensemble and boosting models, as follows.

F. Random Forest

The random forest algorithm is a machine learning methodology that predicts by deriving several decision tree algorithms. The decision tree algorithm is an analysis technique that models relationships and rules of data and does not require assumptions of linearity, normality, and equal variance [25]. Random forest derives several such decision trees and synthesizes the results. Random forest randomly selects training data and independent variables when creating each decision tree to make predictions. Although individual accuracy may be low, all decision trees are aggregated and predicted. It has the advantage of increasing accuracy and stability because it performs side-by-side measurement [26]. In other words, the random forest randomly selects N independent variables and creates T decision tree algorithms that randomly select data and use the most derived value or average value as the predicted value based on the majority rule. The concept of a random forest is illustrated in Fig. 1.

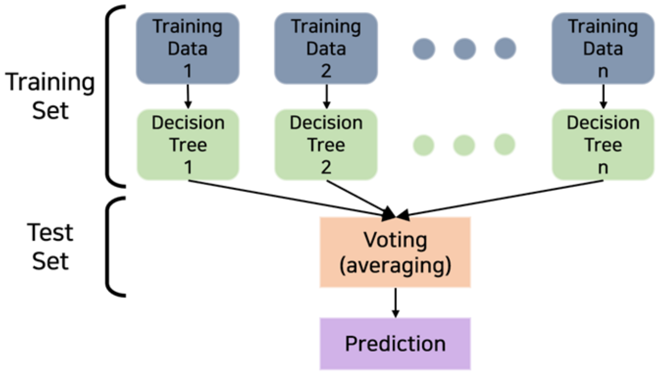

Fig. 1. The concept of a random forest.

G. Extreme Gradient Boosting: XGBoost

The boosting technique is an ensemble technique that creates a weak learner using initial sample data and iteratively adds new learners in the direction of reducing the error of the learning result. In particular, gradient boosting is an algorithm that continues to add new models that predict the residuals of previous learners [27]. However, it has the disadvantage of slow learning and overfitting. XGBoost is an algorithm that compensates for these drawbacks. The concept of XGBoost is shown in Fig. 2.

Introduced by Tianqi Chen in August 2016, XGBoost is a decision tree-based machine learning algorithm that uses a gradient boosting structure. It creates an optimized model that prevents overfitting while minimizing training loss through parallel processing, missing value processing, and regulation [28].

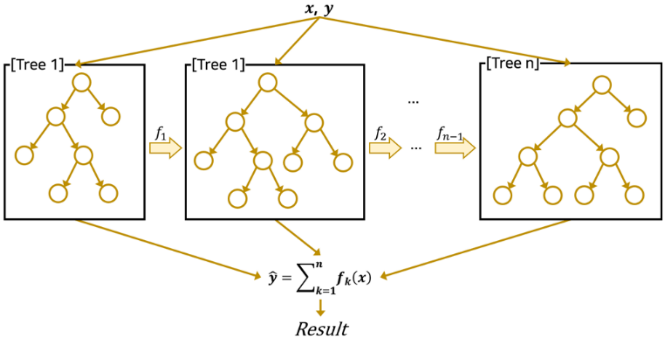

Fig. 2. The concept of XGBoost.

H. Light Gradient Boosting Machine: LightGBM

LightGBM is a fast and efficient GBDT (Gradient Boosting Decision Tree)-based algorithm designed by Microsoft MSRA (Microsoft Research Lab Asia) in 2016 [29]. Existing GBDT-based algorithms have a problem in that they do not perform well in large amounts of high-dimensional data because they have to scan all of the data to evaluate the information gain for all possible split points. Here, information gain refers to better discriminating data by selecting a certain attribute. LightGBM solved the problem by introducing two techniques, Gradient-based One-Side Sampling (GOSS) and Exclusive FeatureBundling (EFB) techniques.

In GBDT, data attributes with large gradients play a larger role in information gain. Therefore, GOSS is a technology that maintains data attributes with a large gradient and randomly removes data attributes with a small gradient with a certain probability. EFB is a technique for grouping mutually exclusive variables according to the characteristics of a sparse variable space to reduce the number of variables [30]. In other words, LightGBM uses this technology to reduce usage and achieve fast training speed. The concept of LightGBM is illustrated in Fig. 3.

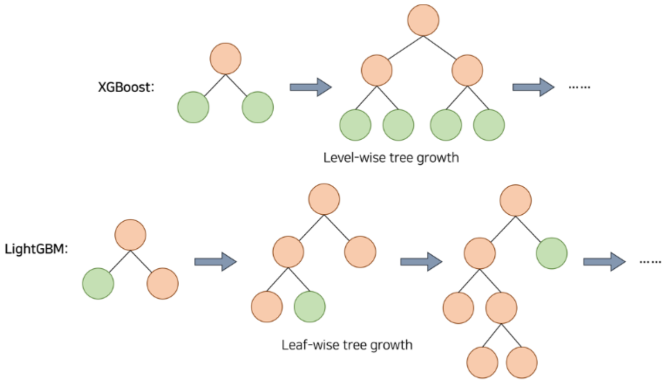

Fig. 3. The concept of LightGBM.

I. Categorical Boosting: CatBoost

The CatBoost algorithm is an ordered boosting technique that focuses on preprocessing categorical variables and solving overfitting problems [31]. Unlike conventional boosting models that sequentially learn all residual errors, ordered boosting creates a model by calculating residual errors with some data. After that, the technique calculates the residual error of the remaining data through the corresponding model.

In addition, overfitting is prevented by mixing the order of data through random permutation in sequential boosting. The CatBoost algorithm improves training speed through variable combinations that combine variables with the same information gain. Furthermore, unlike other ensemble algorithms that use Grid Search or Randomized Search to find the optimal hyperparameter, it optimizes the initial hyperparameter value, so the parameter adjustment procedure is unnecessary.

J. *Explainable Artificial Intelligence (XAI)*

Explainable artificial intelligence (XAI) refers to helping users understand the results by explaining the outcomes predicted by artificial intelligence. This makes it possible to identify the main factors influencing the result, understand the basis of the decision based on the prediction result of the machine learning model, and provide an intuitive explanation that humans can comprehend about the prediction result [32].

K. *Local Interpretable Model-Agnostic Explanation (LIME)*

LIME is a technique that uses combinations of masking or non-masking of superpixels, which are regions of interest in an image that contain important information. The goal is to create an interpretable model that checks the importance of each superpixel in the prediction of a black box model. For example, if an image is classified as a frog, LIME can help us understand why by cutting the image into explanatory units and creating multiple masked and non-masked versions of each unit. We then input these images into the black box model to determine the probability that each one is classified as a frog.

To interpret the results, we train a surrogate model that takes the number of masking cases as input values and the corresponding probabilities as output values. This model can show intuitive results and requires fewer resources than other techniques. Additionally, LIME is model-agnostic, which means it can be applied regardless of the machine learning model used.

However, LIME has some disadvantages. One is that the method used to determine the decision boundary of the model is non-deterministic, meaning that the output value may be different each time it is called. Another is that since LIME only considers one data point at a time, it may not provide a complete explanation of the entire model. The concept of LIME is shown in Fig. 4.

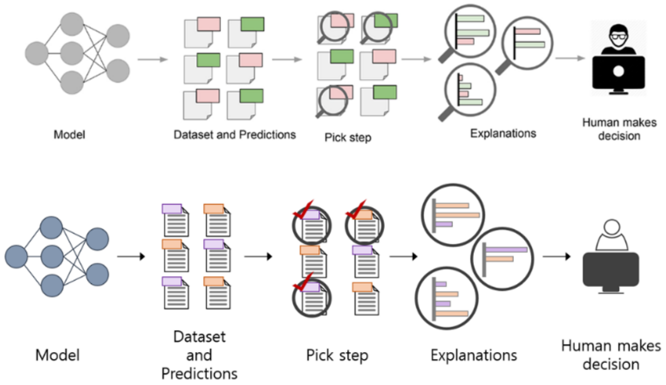

Fig. 4. The concept of LIME.

L. *Shapley Additive exPlanations (SHAP)*

SHAP is an algorithm based on Shapley Values from game theory used to describe the output of a machine learning model. The Shapley value is a value obtained through the average change according to the presence or absence of a variable after constructing a combination of several variables to determine the importance of one variable [33]. An explainable model is created based on the training data and the learned model, and the Shapley value, which expresses the influence on the prediction result in terms of direction and magnitude, is calculated for the newly input data. Through this, the technique explains the contribution that the input variable has on the output value of the learned model.

Existing feature importance techniques use a permutation method to measure the effect of a variable on a model. This method has the advantage of high computational speed, but results may be distorted when variables are dependent on each other. Also, the negative (-) influence cannot be calculated, so the value of a specific variable may be set higher than its actual influence. On the other hand, the SHAP technique considers the possibility that variables affect each other and can calculate the negative (-) influence. Although it has the disadvantage of being slow, it can be seen as measuring the influence more accurately than the variable importance method [34]. The concept of SHAP is shown in Fig. 5.

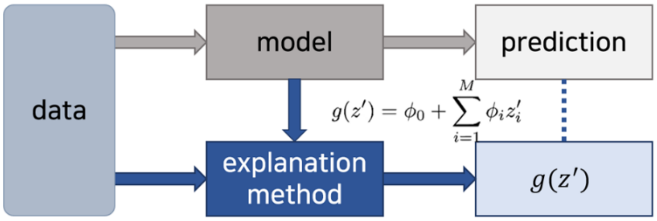

Fig. 5. The concept of SHAP.

M. *Studies of Predicting Depression based on XAI*

Explainable machine learning is a relatively new field that aims to make machine learning models and their decisions more understandable and transparent. This is especially important when making decisions that could negatively impact people's lives, such as diagnosing depression. Several studies have reported that depression was predicted using SHAP, one of the techniques of X-AI (Table II). For instance, Matthew et al. (2021)[35] propose a framework for explainable machine learning called SHAP (SHapley Additive exPlanations) that can identify depressive symptoms from social media posts. SHAP is a method that assigns importance values to each feature in a model based on how much they contribute to the prediction. The framework uses natural language processing and sentiment analysis to extract features and provide explanations for the predictions. The authors assessed the model's performance on a held-out test set and found an AUC of 0.73 (sensitivity: 0.66, specificity: 0.7) and 0.67 (sensitivity: 0.55, specificity: 0.7) for GAD and MDD, respectively. Additionally, the authors used advanced techniques such as SHAP values to illuminate which features had the greatest impact on prediction for each disease.

TABLE II. SUMMARY OF X-AI STUDIES

Article	Data	Models /Algoritms	Results

Nguyen and Byeon (2022)[32]	36,000 depression participants over 60 years old	Deep neural network, LIME	Accuracy=89.92%,
			Precision=93.55%,
			Recall=97.32%

Matthew et al. (2021)[35]	4,184 undergraduate students	Natural language processing, sentiment analysis, SHAP	GAD:
			AUC=0.73 (sensitivity: 0.66, specificity: 0.7)
			MDD:
			AUC=0.67 (sensitivity: 0.55, specificity: 0.7)

Hueniken et al. (2021)[36]	Canadian adults (aged ⩾ 18 years, N=6021) who completed web-based surveys	Random forest, gradient boosting, support vector machine, and neural network, SHAP	Average accuracy of 85% and 88% 3 most important items predicting elevated emotional distress: increased worries about finances (SHAP=0.17), worries about getting COVID-19 (SHAP=0.17), and younger age (SHAP=0.13)

Amit et al. (2021)[37]	266,544 UK women who gave birth between 2000 and 2017	Gradient tree boosting algorithm based on SHAP	Postpartum depression:
			AUC=0.805 to 0.844
			Sensitivity=0.72 to 0.76
			Specificity=0.80

Hochman et al. (2021)[38]	A nationwide longitudinal cohort that included 214,359 births between January 2008 and December 2015	Gradient-boosted decision tree algorithm	Postpartum depression
			AUC=0.712
			Sensitivity=0.349
			Specificity of 0.905

Uddin et al. (2022)[39]	Large text-based dataset from a public Norwegian information website: ung.no. (11,807 and 21,470 posts of different length)	LSTM (Long Short-Term Memory), RNN (Recurrent Neural Network), LIME	Depression Accuracy=84.2%

III. RESULTS AND DISCUSSION

Hueniken et al. (2021) [36] used machine learning methods to identify factors associated with anxiety and depression among Canadian adults during 8 months of the COVID-19 pandemic. The study analyzed data from repeated cross-sectional surveys conducted by Statistics Canada between May 2020 and December 2020, involving 6,021 respondents. Authors applied four machine learning algorithms (random forest, gradient boosting, support vector machine, and neural network) to predict anxiety and depression scores based on demographic, economic, lifestyle, and health risk variables [36]. Authors found that machine learning models performed well in predicting anxiety and depression scores, with an average accuracy of 85% and 88%, respectively [36]. Authors also identified several important predictors of anxiety and depression, including age, gender, income level, employment status, physical activity level, chronic conditions, and perceived health risk related to COVID-19 infection or vaccination.

The study by Amit et al. (2021)[37] aimed to predict the risk of postpartum depression (PPD) using machine learning and electronic health records (EHR) data from primary care. PPD is a common disorder that affects mothers and their newborns. The study used data from 266,544 UK women who gave birth between 2000 and 2017 and had at least one visit to their primary care physician within a year after delivery. The machine learning algorithm used in this study was a gradient tree boosting algorithm based on SHAP. According to the findings, incorporating EHR-based forecasting with EPDS score enhanced the area under the receiver operating characteristic curve (AUC) from 0.805 to 0.844, as well as increased the sensitivity from 0.72 to 0.76 while retaining a specificity of 0.80. The study demonstrates the feasibility and value of using SHAP-based machine learning and EHR data for estimating PPD risk and improving screening and early intervention.

Hochman et al. (2021) [38] conducted a study to create and validate a model using machine learning to predict postpartum depression (PPD). The research analyzed data from a national cohort of Israeli women who gave birth between 2008 and 2015 and had a psychiatric diagnosis or prescription within a year after delivery. EHR-derived sociodemographic, clinical, and obstetric features were used with a gradient-boosted decision tree algorithm to develop the prediction model. The model's accuracy was assessed in the validation set, achieving an AUC of 0.712, with a sensitivity of 0.349 and a specificity of 0.905 at the 90th percentile risk threshold. The model identified PPDs more than three times higher than the overall set, with positive and negative predictive values of 0.074 and 0.985, respectively. The study revealed that both recognized (e.g., past depression) and less-recognized (differing patterns of blood tests) PPD risk factors were strong predictors in the model. The research demonstrated the usefulness of machine learning-based models in predicting PPD using large-scale cohort data with high accuracy.

Several studies [32, 39] have developed X-AI models to predict depression using LIME. Uddin et al. (2022) [39] were develops an interpretable machine learning model that can predict depression from multi-modal data, such as speech, text, and facial expressions. The model useed attention mechanisms and feature importance scores to provide insights into the factors influencing depression. Furthermore, as the attributes utilized by the system are grounded on the probable indications of depression, the system could produce purposeful justifications of the verdicts from machine learning algorithms through the use of an interpretable artificial intelligence technique named LIME. The accuracy of the developed depression prediction model was 84.2%.

Nguyen and Byeon (2022) [32] utilized a deep neural network (DNN) model to make predictions about depression in elderly individuals during the pandemic. They focused on social factors related to stress, health status, daily changes, and physical distancing as potential predictors. To obtain data, they used the 2020 Community Health Survey of the Republic of Korea, which included more than 97,000 participants over 60 years old. After cleansing the data, the DNN model was trained on information from over 36,000 participants and 22 variables. The researchers also integrated the DNN model with a LIME-based model to make the predictions more explainable. The study found that the model achieved an accuracy of 89.92% and had high precision (93.55%) and recall (97.32%) scores, indicating its effectiveness. The researchers highlighted the potential of this explanatory DNN model in identifying elderly patients who require early treatment due to the increased likelihood of depression caused by the pandemic.

Taken together, X-AI such as SHAP and LIME have been reported to be effective in predicting depression in several previous studies. However, the predictive performance of machine learning techniques varies across studies due to differences in data imbalance (particularly in the Y variable), the nature of the features incorporated in the model, and how the outcome variable is measured. Therefore, while some studies have shown that X-AI-based machine learning algorithms perform well, additional studies are continually needed to verify the predictive performance of each algorithm since the results cannot be generalized to all data types.

IV. Conclusion

Models that are easy to interpret often have simple structures and lower accuracy, while models that are difficult to interpret typically have more complex structures and higher accuracy. In various fields, researchers are conducting studies to apply X-AI to models to ensure interpretability while using powerful learning algorithms with excellent predictive performance. In order to introduce AI into sensitive decisions such as medical diagnoses, and to support medical professionals in their decision-making, sufficient justification for AI results needs to be established. Particularly in the field of psychiatry, such as the prediction of depression, it is crucial to ensure that the justifications for AI predictions are clear and transparent. Therefore, ensuring the interpretability of AI models will be important in future research.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2023-00237287, NRF-2021S1A5A8062526).

References

[1] C. R. Yoo, S. H. Kim, and J. W. Kim, A Comparative Study of the Use of Intelligent Personal Assistant Services Experiences: Siri, Google Assistant, Bixby. Sci Emot Sensibility, vol. 23, no. 1, pp. 69-78, 2020.

[2] M. Fakhoury, Artificial intelligence in psychiatry. Adv Exp Med Biol, Vol. 1192, pp. 119-125, 2019.

[3] H. Byeon, Screening dementia and predicting high dementia risk groups using machine learning. World J Psychiatry, vol. 12, no. 2, pp. 204-211, 2022.

[4] A. De Los Reyes, E. Talbott, TJ. Power, JJ. Michel, CR. Cook, SJ. Raxz, and O. Fitzpatrick, The Needs-to-Goals Gap: How informant discrepancies in youth mental health assessments impact service delivery. Clin Psychol Rev, vol. 92, pp. 102114, 2022.

[5] T. Hansen, T. Hatling, E. Lidal, and T. Ruud, Discrepancies between patients and professionals in the assessment of patient needs: a quantitative study of Norwegian mental health care. J Adv Nurs, vol. 39, no. 6, pp. 554-562, 2002.

[6] L. Fajutrao, J. Locklear, J. Priaulx, and A. Heyes, A systematic review of the evidence of the burden of bipolar disorder in Europe. Clin Pract Epidemiol Ment Health, vol. 5, no. 3, pp.1-8, 2009.

[7] H. Shen, L. Zhang, C. Xu, J. Zhu, M. Chen, and Y. Fang, Analysis of Misdiagnosis of Bipolar Disorder in An Outpatient Setting. Shanghai Arch Psychiatry, vol. 30, no. 2, pp. 93-101, 2018.

[8] Y. Jang, DA. Chiriboga, and S. Okazaki, Attitudes toward mental health services: age-group differences in Korean American adults. Aging Ment Health, vol. 13, no. 1, pp. 127-134, 2009.

[9] C. S. Mackenzie, W. L. Gekoski, and V. J. Knox, Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. Aging Ment Health, vol. 10, no. 6, pp. 574-582, 2006.

[10] J. A. Leis, T. Mendelson, D. F. Perry, and S. D. Tandon, Perceptions of mental health services among low-income, perinatal African-American women. Womens Health Issues, vol. 21, no. 4, pp. 314-319, 2011.

[11] G. Antoniou, E. Papadakis, and G. Baryannis, Mental health diagnosis: a case for explainable artificial intelligence. Int J Artif Intell Tools, vol. 31, no. 3, pp. 2241003, 2022.

[12] C. Otte, S. M. Gold, B. W. Penninx, C. M. Pariante, A. Etkin, M. Fava, D.C. Mohr, and A. F. Schatzberg, Major depressive disorder. Nat Rev Dis Primers, vol. 2, no. 1, pp. 1-20, 2016.

[13] V. Kaul, S. Enslin, and S. A. Gross, History of artificial intelligence in medicine. Gastrointest Endosc, vol. 92, no. 4, pp. 807-812, 2020.

[14] O. Kuusi, and S. Heinonen, Scenarios From Artificial Narrow Intelligence to Artificial General Intelligence—Reviewing the Results of the International Work/Technology 2050 Study. World Futures Rev, vol. 14, no. 1, pp. 65-79, 2022.

[15] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning. Nature, vo. 521, no. 7553, pp. 436-444, 2015.

[16] C. Flint, M. Cearns, N. Opel, R. Redlich, D. M. A. Mehler, D. Emden, N. R. Winter, R. Leenings, S. B. Eickhoff, T. Kircher, A. Krug, I. Nenadic, V. Arolt, S. Clark, B. T. Baune, X. Jiang, U. Dannlowski, and T. Hahn, Systematic misestimation of machine learning performance in neuroimaging studies of depression. Neuropsychopharmacology, vol. 46, no. 8, pp. 1510-1517, 2021.

[17] B. A. Johnston, J. D. Steele, S. Tolomeo, D. Christmas, and K. Matthews, Structural MRI-Based Predictions in Patients with Treatment-Refractory Depression (TRD). PLoS One, vol. 10, no. 7, pp. e0132958, 2015.

[18] M. J. Patel, C. Andreescu, J. C. Price, K. L. Edelman, C. F. Reynolds III, and H. J. Aizenstein, Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int J Geriatr Psychiatry, vol. 30, no. 10, pp. 1056-1067, 2015.

[19] D. A. Regier, E. A. Kuhl, and D. J. Kupfer, The DSM-5: Classification and criteria changes. World Psychiatry, vol. 12, no. 2, pp. 92-98, 2013.

[20] K. S. Kendler, and M. B. First, Alternative futures for the DSM revision process: iteration v. paradigm shift. Br J Psychiatry, vol. 197, no. 4, pp. 263-265, 2010.

[21] H. D. Schmidt, R. C. Shelton, and R. S. Duman, Functional biomarkers of depression: diagnosis, treatment, and pathophysiology. Neuropsychopharmacology, vol. 36, no. 12, pp. 2375-2394, 2011.

[22] M. Sajjadian, R. W. Lam, R. Milev, S. Rotzinger, B. N. Frey, C. N. Soares, S. V. Parikh, J. A. Foster, G. Turecki, D. J. Müller, S. C. Strother, F. Farzan. S. H. Kennedy, and R. Uher, Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. Psychol Med, vol. 51, no. 16, pp. 2742-2751, 2021.

[23] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, Predicting women with depressive symptoms postpartum with machine learning methods. Sci Rep, vol. 11, no. 1, pp. 1-15, 2021.

[24] Y. Park, J. Hu, M. Singh, I. Sylla, I. Dankwa-Mullan, E. Koski, and A. K. Das, Comparison of Methods to Reduce Bias From Clinical Prediction Models of Postpartum Depression. JAMA Netw Open, vol. 4, no.4, pp. e213909, 2021.

[25] M. Park, S. Choi, A. M. Shin, and C. H. Koo, Analysis of the characteristics of the older adults with depression using data mining decision tree analysis. J Korean Acad Nurs, vol. 43, no. 1, pp. 1-10, 2013.

[26] H. Byeon, Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's Disease with Normal Cognition?. Int J Environ Res Public Health, vol. 17, no. 7, pp. 2594, 2020.

[27] A. Ogunleye, and Q. G. Wang, XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Trans Comput Biol Bioinform, vol. 17, no. 6, pp. 2131-2140, 2020.

[28] H. J. Hwang, S. H. Kim, and G. W. Song, Xgboost model to identify potential factors improving and deteriortating elderly cognition. Korean Inst Next Gener Comput, vol. 14, no. 14, pp. 16-24, 2018.

[29] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electron Commer Res Appl, vol. 31, pp. 24-39, 2018.

[30] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst, vol. 30, pp. 3146-3154, 2017.

[31] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, CatBoost: unbiased boosting with categorical features. arXiv preprint, 2017.

[32] H. V. Nguyen, and H. Byeon, Explainable Deep-Learning-Based Depression Modeling of Elderly Community after COVID-19 Pandemic. Mathematics, vol. 10, no. 23, pp. 4408, 2022.

[33] S. M. Lundberg, and S. I. Lee, A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst, vol. 30, pp. 4766-4775, 2017.

[34] H. Byeon, Predicting South Korean adolescents vulnerable to obesity after the COVID-19 pandemic using categorical boosting and shapley additive explanation values: A population-based cross-sectional survey. Front Pediatr, vol. 10, pp. 955339, 2022.

[35] M. D. Nemesure, M. V. Heinz, R. Huang, and N. C. Jacobson, Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep, vol. 11, pp. 1980, 2021.

[36] K. Hueniken, N. H. Somé, M. Abdelhack, G. Taylor, T. E. Marshall, C. M. Wickens, H. A. Hamilton, S. Wells, and D. Felsky, Machine Learning-Based Predictive Modeling of Anxiety and Depressive Symptoms During 8 Months of the COVID-19 Global Pandemic: Repeated Cross-sectional Survey Study. JMIR Ment Health, vol. 8, no. 11, pp. e32876, 2021.

[37] G. Amit, I. Girshovitz, K. Marcus, Y. Zhang, J. Pathak, V. Bar, and P. Akiva, Estimation of postpartum depression risk from electronic health records using machine learning. BMC Pregnancy Childbirth, vol. 21, pp. 630, 2021.

[38] E. Hochman, B. Feldman, A. Weizman, A. Krivoy, S. Gur, E. Barzilay, H. Gabay, J. Levy, O. Levinkron, and G. Lawrence, Development and validation of a machine learning-based postpartum depression prediction model: A nationwide cohort study. Depress Anxiety, vol. 38, no. 4, pp. 400-411, 2021.

[39] M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Comput Appl, vol. 34, no. 1, pp. 721-744, 2022.