



Machine Learning on Early Diagnosis of Depression

Kwang-Sig Lee¹ and Byung-Joo Ham² ✉

¹AI Center, Korea University Anam Hospital, Seoul, Republic of Korea

²Department of Mental Health, Korea University Anam Hospital, Seoul, Republic of Korea

To review the recent progress of machine learning for the early diagnosis of depression (major depressive disorder). The source of data was 32 original studies in the Web of Science. The search terms were “depression” (title) and “random forest” (abstract). The eligibility criteria were the dependent variable of depression, the interventions of machine learning (the decision tree, the naïve Bayesian, the random forest, the support vector machine and/or the artificial neural network), the outcomes of accuracy and/or the area under the receiver operating characteristic curve (AUC) for the early diagnosis of depression, the publication year of 2000 or later, the publication language of English and the publication journal of SCIE/SSCI. Different machine learning methods would be appropriate for different types of data for the early diagnosis of depression, e.g., logistic regression, the random forest, the support vector machine and/or the artificial neural network in the case of numeric data, the random forest in the case of genomic data. Their performance measures reported varied within 60.1–100.0 for accuracy and 64.0–96.0 for the AUC. Machine learning provides an effective, non-invasive decision support system for early diagnosis of depression.

Psychiatry Investig 2022;19(8):597-605

Keywords Depression; Early diagnosis; Machine learning.

INTRODUCTION

Major depressive disorder (depression hereafter) is a leading cause of disease burden in the world, affecting 300 million people on the globe.¹⁻⁶ Depression, defined as “a mood disorder that causes a persistent feeling of sadness and loss of interest,”¹ is a diagnostic category of mental disorder together with anxiety disorder.² Its global incidence registered a rapid growth of 47.86% from 172 million to 258 million during 1990 to 2017.³ It ranked third in the world for 2017 and second in Korea for 2010 in terms of years lost to disability and disability-adjusted life years, respectively.^{4,5} It is considered to have a wide range of determinants including demographic factors (age, sex), socioeconomic status (education, employment, income), neighborhood conditions (crowding, housing, pollution, violence) and health-related factors (drinking, exercise, smoking, diseases, genetics).^{6,7} Recently, on the other hand, the

terms “deep learning,” “machine learning” and “artificial intelligence” have attracted great attention all over the globe. For instance, their Google trends recorded ten-fold expansions from 10 to 100 during 2013–2018. Artificial intelligence can be defined as “the capability of a machine to imitate intelligent human behavior” (the Merriam-Webster dictionary). The definition of machine learning can be a division of artificial intelligence to “extract knowledge from large amounts of data.”⁸

Six common machine learning algorithms are the decision tree, the naïve Bayesian predictor, the random forest, the support vector machine, the artificial neural network, and the deep neural network (deep learning). A decision tree has three components: an intermediate node (a test on an independent variable), a branch (an outcome of the test) and a terminal node (a value of the dependent variable). A naïve Bayesian predictor makes an early diagnosis based on Bayes’ theorem, which states that the probability of the dependent variable given certain values of independent variables comes from the probabilities of the independent variables given a certain value of the dependent variable. A random forest is a collection of many decision trees, which make majority votes on the dependent variable (“bootstrap aggregation”). Let us take a random forest with 1,000 decision trees as an example. Let us assume that original data includes 10,000 participants. Then,

Received: March 13, 2022 **Revised:** April 19, 2022

Accepted: May 23, 2022

✉ **Correspondence:** Byung-Joo Ham, MD, PhD

Department of Mental Health, Korea University Anam Hospital, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

Tel: +82-2-2286-6843, **Fax:** +82-2-6280-5810

E-mail: byungjoo.ham@gmail.com

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

the training and test of this random forest takes two steps. Firstly, new data with 10,000 participants is created based on random sampling with replacement, and a decision tree is created based on this new data. Here, some participants in the original data would be excluded from the new data and these leftovers are called out-of-bag data. This process is repeated 1,000 times, i.e., 1,000 new data are created, 1,000 decision trees are created and 1,000 out-of-bag data are created. Secondly, the 1,000 decision trees make predictions on the dependent variable of every participant in the out-of-bag data, their majority vote is taken as their final prediction on this participant, and the out-of-bag error is calculated as the proportion of wrong votes on all participants in the out-of-bag data.

A support vector machine originates a line or space called a “hyperplane” (a collection of “support vectors”). The hyperplane divides data with the greatest distance between different sub-groups.⁸ An artificial neural network is a network of “neurons”, i.e., information units combined through weights. Usually, the artificial neural network has one input layer, one, two or three intermediate layers and one output layer. Neurons in a previous layer connect with “weights” in the next layer and these weights represent the strengths of connections between neurons in a previous layer and their next-layer counterparts. This process starts from the input layer, continues through intermediate layers and ends in the output layer (feed-forward operation). Then, learning happens: these weights are accommodated based on how much they contributed to the loss, a difference between the actual and predicted final outputs. This process starts from the output layer, continues through intermediate layers and ends in the input layer (back-propagation operation). The two operations are replicated until a certain expectation is met regarding the accurate diagnosis of the dependent variable. Finally, a deep neural network is an artificial neural network with a large number of intermediate layers, e.g., 5, 10 or even 1,000. The deep neural network is called “deep learning” given that learning “deepens” through numerous intermediate layers.⁹

Traditional research considers a limited scope of predictors for the early diagnosis of disease, whereas adopting logistic regression with an unrealistic assumption of *ceteris paribus*, i.e., “all the other variables staying constant.” In this context, emerging literature uses artificial intelligence for the early diagnosis of disease, e.g., arrhythmia,⁹ birth outcome,^{10–15} cancer,^{16–19} comorbidity,^{20–22} menopause²³ and temporomandibular disease.^{24,25} It does not require unrealistic assumptions of “all the other variables staying constant” while managing to analyze which predictors are more important for the early diagnosis of the dependent variable. The purpose of this study is to review the recent progress of machine learning for the early diagnosis of depression.

METHODS

Figure 1 shows the flow diagram of this study. Thirty two original studies were selected for review out of 120 original studies in the Web of Science with the search terms “depression” (title) and “random forest” (abstract). The eligibility criteria of this review were: 1) the intervention(s) of the decision tree, the naïve Bayesian predictor, the random forest, the support vector machine and/or the artificial neural network; 2) the outcome(s) of accuracy and/or the area under the receiver operating characteristic curve (AUC) for the early diagnosis of depression; 3) the publication year of 2000 or later; 4) the publication language of English; 5) the publication journal of Science Citation Index Expanded and/or Social Science Citation Index; and 6) depression being the dependent variable. The following summary measures were adopted: machine learning methods, sample size, data type, performance measures, and important predictors. Here, accuracy can be defined as the proportion of correct predictions over all observations, while the AUC can be defined as the area under the plot of the true positive rate (sensitivity) against the false positive rate (1- specificity) at various threshold settings.

RESULTS

Review summary

The summary of review is shown in Tables 1 and 2.^{26–57} The tables have five summary measures, i.e., machine learning methods, sample size, data type, performance measures, important predictors, and whether the variable importance of the random forest is reported (VI-Yes 1). Based on the results of this review, different machine learning methods would be appropriate (i.e., would show the best performance measures) for different types of data for the early diagnosis of depression: 1) logistic regression, the random forest, the support vector machine and/or the artificial neural network in the case of numeric data; 2) the random forest in the case of genomic data; 3) the random forest and/or the support vector machine in the case of radiomic data; and 4) the random forest in the case of social-network-service data. Their performance measures reported varied within 60.1–100.0 for accuracy, 68.8–95.0 for sensitivity, 76.0–94.0 for specificity, and 64.0–96.0 for the AUC (Table 1). According to the findings of this review, indeed, the following predictors would be important variables for the early diagnosis of depression: comorbid psychopathology, symptom-related disability, treatment credibility, access to therapists, time spent using certain internet-intervention modules; pain-fatigue (symptom intensity scale), comorbidity; 30 microbial markers (gut microbiota); psychological elasticity, income level; upper body move-

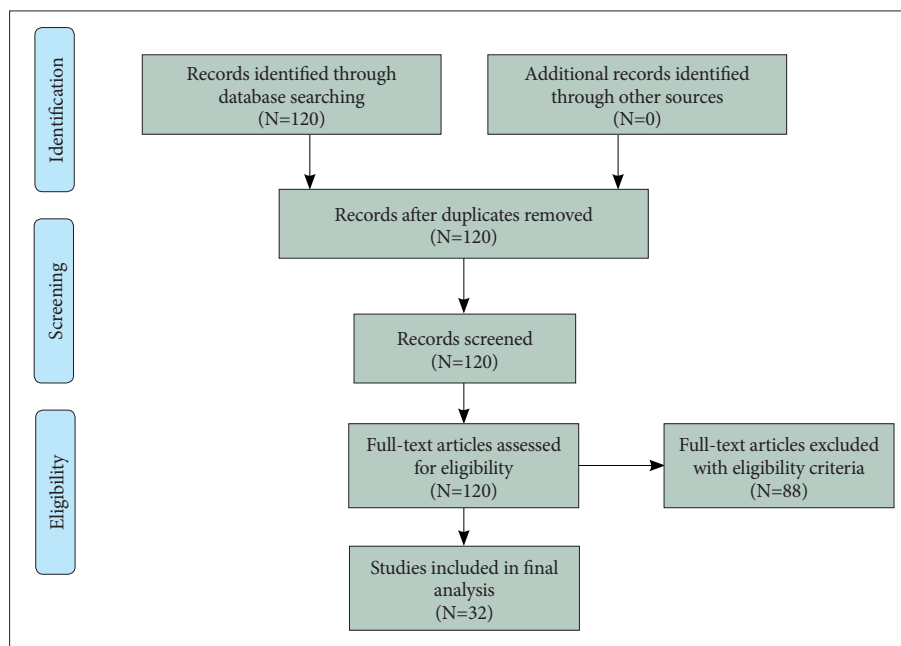


Figure 1. Flow diagram.

ments-postures; brain connectivity within posterior cingulate cortex, within insula, between posterior cingulate cortex and insula/hippocampus-amygdala, between insula and precuneus, between superior parietal lobule and medial prefrontal cortex; single-nucleotide polymorphisms (rs12248560, rs878567, rs17710780); cingulate isthmus asymmetry, pallidal asymmetry, ratio of the paracentral to precentral cortical thickness, ratio of lateral occipital to pericalcarine cortical thickness; self-assessed cardiac-related fear, sex, number of words to answer the first homework assignment for internet-delivered psychotherapy (Table 2). However, machine learning is a data-driven method and more study is to be done with more external data for greater external validity.

Numeric data

The review of major studies with numeric data is presented in this section. The aim of a recent study²⁹ was to adopt numeric data and machine learning for analyzing the associations of depression with participant characteristics and 8-week internet intervention (Deprexis). The data came from 283 adults in the United States and their demographic, psychopathological, environmental and intervention variables were considered. The R^2 performance of an ensemble with the elastic net and the random forest (0.25) were better than the auto-regressive model (0.17) for the prediction of post-treatment depression. Based on random forest VI, important predictors for the early diagnosis of depression were comorbid psychopathology, low symptom-related disability, treatment credibility, lower access to therapists and time spent us-

ing certain internet-intervention (Deprexis) modules. The contribution of a single predictor was small but the ensemble model with a rich collection of various predictors showed reasonable performance. Another study³⁰ employs numeric data and machine learning to identify important predictors for the early diagnosis of rheumatoid arthritis patients' depression. The source of the data was 22,131 rheumatoid arthritis patients during 1999–2008 and their demographic, socioeconomic and health-related variables were included, especially, regarding pain-fatigue (Symptom Intensity Scale) and comorbidity. Most predictors were statistically significant in logistic regression but two predictors (pain-fatigue and comorbidity) were dominant in terms of random forest VI. This finding highlights the centrality of the two predictors regarding the clinical implication.

The early diagnosis of postpartum depression is a major issue in medicine and a recent study³² builds two-stage machine learning methods to solve this issue, i.e., feature selection then depression diagnosis. The data came from a cohort of 508 women, the Edinburgh Postnatal Depression Scale within 42 days after delivery was used as the dependent variable, and their demographic, socioeconomic and health-related factors were considered as the independent variables. Four combinations of prediction models were used in this study: expert vs. random forest in feature selection; support vector machine vs. random forest in depression diagnosis. The random forest-random forest combination registered the best performance measures, AUC 78.0 and sensitivity 69.0. Psychological elasticity, depression during the third trimester and the in-

come level ranked the top 3 in terms of random forest VI. The early diagnosis of depression for university students is another great challenge in health and another study³⁴ employs the random forest and China's data on upper body movement alone

(e.g., head posture, arm/body swing) to achieve the accuracy of 91.6. In a similar vein, low adherence to internet-delivered psychotherapy for myocardial infarction patient's depression is an important research topic and a recent study⁵⁷ uses the ran-

Table 1. Summary of review: methods, sample size, data type and performance measures

ID	Methods	Sample size	Data type	Performance measures
26	Semisupervised RF	115	Numeric	RMSE 4.50
27	RF	1,549	Numeric	Accuracy Validation 94.2 Test 93.3
28	LR DT NB RF SVM ANN	28,755	Numeric	AUC RF 88.4 SVM 86.4
29	AR EN-RF	283	Numeric	R2 EN-RF 0.25 AR 0.17
30	LR RF	22,131	Numeric	Only Coefficient P-Values Reported
31	RF	97	Genomic	AUC 81.0
32	Expert vs. RF (Feature Selection); SVM vs. RF (Classification)	508	Numeric	RF-RF AUC 78.0 Sensitivity 69.0
33	RF	126	Genomic	Accuracy 87.3
34	RF	3,669	Numeric	Accuracy 91.6
35	Lasso RF SVM	120	Radiomic	Accuracy 95.0 90.0 100.0
36	10 Models	620	Numeric	RF Accuracy 89.0/91.0 Internal/External
37	1 RF vs. 2 RFs	135	SNS	Early Risk Detection Error: 2 RFs 10.0% Lower vs. 1 RF
38	RF	201	Radiomic	Accuracy 82.4
39	RF	412	Numeric	Accuracy 76.8/81.1 Imbalanced/Balanced Data
40	LR DT NB RF SVM ANN	43	EEG	Accuracy 90.24–97.56
41	LR RF	439	Numeric	Only Coefficient P-Values Reported
42	LR RF ANN	637	Numeric	AUC 87.0–91.0
43	LR RF	150	Genomic	Only Coefficient P-Values Reported
44	LR RF	4,270	Numeric	AUC 76.0–79.0
45	RF	111	Numeric	Sensitivity 69.7, Specificity 76.8
46	RF	5,895	Numeric	Sensitivity 86.7, Specificity 91.9
47	EN (Feature Selection); RF (Classification)	41	Radiomic	Accuracy 85.4
48	RF	42	Radiomic	Accuracy 71.0–78.0
49	RF	656	Numeric	Sensitivity 95.0, Specificity 87.0
50	LR RF SVM	238	Numeric	LR AUC 93.8
51	Lasso RF SVM	62	Radiomic	Lasso SVM Accuracy 90.0
52	RF SVM	126	Numeric	Accuracy RF 60.1 SVM 59.1
53	LR RF SVM GB	170	Numeric	GB Accuracy 76.9
54	LR RF GB	573,634	Numeric	Accuracy LR 71.7 RF 72.0 GB 72.2; AUC LR 71.7 RF 72.0 GB 72.2
55	LR DT RF GB	47	Numeric	LR Accuracy 91.0 AUC 96.0 Sensitivity 92.9 Specificity 94.0
56	RF	84,317	Numeric	AUC 78.9 Sensitivity 68.8–83.9 Specificity 76.0–92.2
57	RF	90	Numeric	AUC 64.0

Different machine learning methods would be appropriate (i.e., would show the best performance measures) for different types of data for the early diagnosis of depression: 1) logistic regression, the random forest, the support vector machine and/or the artificial neural network in the case of numeric data; 2) the random forest in the case of genomic data; 3) the random forest and/or the support vector machine in the case of radiomic data; and 4) the random forest in the case of social-network-service data. Their performance measures reported varied within 60.1–100.0 for accuracy, 68.8–95.0 for sensitivity, 76.0–94.0 for specificity, and 64.0–96.0 for the AUC. ANN, artificial neural network; AR, augoregressive; AUC, area under the receiver operating characteristic curve; DT, decision tree; EEG, electroencephalogram; EN, elastic net; GB, gradient boosting; LR, logistic regression; NB, naïve bayes; RF, random forest; RMSE, root mean squared error; SNS, social network service; SNP, single nucleotide polymorphism; SVM, support vector machine

Table 2. Summary of review: important predictors and whether variable importance (VI) is reported

ID	Important predictors	VI-yes	Participants/class/predictors
26	Cognitive-behavioral features		Participants: 35 labeled 80 unlabeled
27	Patient health questionnaire-9 items		Participants: university students
28	Demographic, health-behavioral factors		Participants: pregnancy risk assessment monitoring system enrollee
29	Comorbid psychopathology, symptom-related disability, treatment credibility, access to therapists, time spent using certain internet-intervention (deprexis) modules	1	
30	Pain-fatigue (symptom intensity scale), comorbidity	1	Participants: rheumatoid arthritis patients
31	30 Microbial markers (gut microbiota)	1	Predictors: 16s-ribosomal rna gene sequences
32	Psychological elasticity, depression during the third trimester, income level	1	Participants: women with delivery
33	Blood-derived methylome and transcriptome features		
34	Upper body movements-postures	1	Participants: university students
35	19 Features of brain connectivity		Participants: parkinson's disease patients
36	Demographic, health-behavioral factors		Participants: 510/110 elders for internal/external validation
37	SNS-derived behavioral patterns		
38	Brain connectivity within posterior cingulate cortex, within insula, between posterior cingulate cortex and insula/hippocampus-amygdala, between insula and precuneus, between superior parietal lobule and medial prefrontal cortex	1	Participants: 156 advanced parkinson's disease patients and 45 normal controls (predictors: 42 brain connectivity networks)
39	Fewer contacts, fewer calls, more messages		
40	Higuchi's fractal dimension, sample entropy		
41	Fluoxetine more important than cognitive-behavioural therapy, two combined more important than one		
42	Patient-reported immune-mediated inflammatory disease measures		
43	SNPs rs12248560, rs878567, rs17710780	1	Participants: 150 depression patients on 6-month regular therapy from the psycolaus cohort (predictors: 44 snps in existing literature)
44	Psychometric properties in general health questionnaire		
45	Six cognitive-behavioral tasks		Class: anxiety, depression or mixed vs. Healthy
46	Motor activity recorded in a wearable device		
47	Prefrontal cortical activation during working memory task anticipation		Class: unipolar vs. Bipolar depression
48	Cingulate isthmus asymmetry, pallidal asymmetry, ratio of the paracentral to precentral cortical thickness, ratio of lateral occipital to pericalcarine cortical thickness	1	Class: depression relapse after electroconvulsive therapy
49	4-6 Computerized-adaptive-diagnostic-test measures		
50	Sex, age, medical insurance, marital status, education level, household income, pathological stage, psychosocial measures (social skills rating system, pittsburgh sleep quality index, european organization for research and treatment of cancer quality of life questionnaire [QLQ-C30])		Participants: non-hodgkin's lymphoma patients with chemotherapy
51	Left precuneus, left precentral gyrus, left inferior frontal cortex (pars triangularis), left cerebellum		
52	120 Behavioral patterns based on smartphone sensors including app adherence		
53	Whole body kinematic cues		
54	Age, race		Participants: women with delivery

Table 2. Summary of review: important predictors and whether variable importance (VI) is reported (continued)

ID	Important predictors	VI-yes	Participants/class/predictors
55	Physical activity and light exposure measured by a wearable device, sleep efficiency measured in a survey		
56	Demographic, health-behavioral factors		
57	Self-assessed cardiac-related fear, sex, number of words to answer the first homework assignment	1	Class: adherence to internet-delivered psychotherapy for myocardial infarction patients' anxiety and depression

The following predictors would be important variables for the early diagnosis of depression: comorbid psychopathology, symptom-related disability, treatment credibility, access to therapists, time spent using certain internet-intervention modules; pain-fatigue (symptom intensity scale), comorbidity; 30 microbial markers (gut microbiota); psychological elasticity, income level; upper body movements-postures; brain connectivity within posterior cingulate cortex, within insula, between posterior cingulate cortex and insula/hippocampus-amygdala, between insula and precuneus, between superior parietal lobule and medial prefrontal cortex; single-nucleotide polymorphisms (rs12248560, rs878567, rs17710780); cingulate isthmus asymmetry, pallidal asymmetry, ratio of the paracentral to precentral cortical thickness, ratio of lateral occipital to pericalcarine cortical thickness; self-assessed cardiac-related fear, sex, number of words to answer the first homework assignment for internet-delivered psychotherapy. ANN, artificial neural network; AR, autoregressive; AUC, area under the receiver operating characteristic curve; DT, decision tree; EEG, electroencephalogram; EN, elastic net; GB, gradient boosting; LR, logistic regression; NB, naïve bayes; RF, random forest; RMSE, root mean squared error; SNS, social network service; SNP, single nucleotide polymorphism; SVM, support vector machine

dom forest to identify important factors for the adherence. The source of the data was 90 myocardial infarction patients participated in Uppsala University Psychosocial Care Programme Heart study in Sweden. Adherence was defined as completing more than two homework assignments during 14-week therapy. The top-3 factors for the adherence were self-assessed cardiac-related fear, sex and the number of words to answer the first homework assignment. Examining a causal mechanism between linguistic factors and internet-delivered psychotherapy is expected to make a great contribution in this direction.

Genomic and radiomic data

The review of important studies with genomic and radiomic data is reported in this section. The aim of a recent study³¹ is to adopt genomic data and machine learning for analyzing the effect of quetiapine treatment on microbiota and investigating the utility of microbiota as a biomarker for the diagnosis and treatment of bipolar depression (BD). Based on the results of univariate analysis for 16S-ribosomal RNA gene sequences, the composition of gut microbiota is significantly different between BD participants and their normal counterparts. Bacteroidetes and Firmicutes were dominant in BD patients and their normal counterparts, respectively. Quetiapine treatment for BD participants altered the composition of their gut microbiota. According to the findings of the random forest, 30 microbial markers were effective predictors of BD participants compared to their normal counterparts (AUC 81.0). This study concludes that quetiapine treatment would change the composition of gut microbiota and microbial markers would be effective for the diagnosis and treatment of BD. Likewise, another study⁴³ employed genomic data and the random forest to examine the effects of generic variants on major depressive dis-

order (depression hereafter) during 6-month regular therapy. The source of the data was 150 depression patients on 6-month regular therapy from the population-based PsyCoLaus cohort in Switzerland. The independent variables were 44 single nucleotide polymorphisms in existing literature. Among the 44 predictors, rs12248560, rs878567, and rs17710780 ranked the top 3 in terms of random forest VI. This study was a rare attempt to demonstrate that combining different types of data would break new ground for this area.

It is noteworthy to address recent innovations based on the combination of radiomic data and machine learning for the early diagnosis of depression. Depression is reported to be under-diagnosed in Parkinson's disease patients, given that the two diseases are overlapped in their symptoms and it is very challenging to take accurate measures in old patients with Parkinson's disease. In this context, a recent study³⁸ uses radiomic data and the random forest to highlight brain connectivity networks as effective predictors of depression among Parkinson's disease patients. The data came from 156 advanced Parkinson's disease patients and 45 normal controls. The independent variables were their 42 brain connectivity networks. Among the 42 predictors, the following networks ranked the top 6 in terms of random forest VI: 1) within posterior cingulate cortex; 2) within insula; 3/4) between posterior cingulate cortex and insula/hippocampus-amygdala; 5) between insula and precuneus; and 6) between superior parietal lobule and medial prefrontal cortex. The accuracy of the random forest was 82.4. This study concludes that brain connectivity networks would be useful predictors of depression among Parkinson's disease patients based on radiomic data and machine learning. In a similar context, the focus of another study⁴⁸ was to improve the prediction of depression relapse after 6-month electroconvulsive therapy (which is reported to have a relapse

rate higher than 50%). The source of the data was 42 depression patients with 6-month electroconvulsive therapy in the United States. Top predictors in terms of random forest VI were cingulate isthmus asymmetry, pallidal asymmetry, the ratio of the paracentral to precentral cortical thickness and the ratio of lateral occipital to pericalcarine cortical thickness (accuracy 71.0–78.0). Structural imaging features are expected to have great utility for the prediction of the prediction of depression relapse after 6-month electroconvulsive therapy.

DISCUSSION

This study presented one of the most comprehensive reviews regarding the recent progress of machine learning for the early diagnosis of depression. This study reviewed thirty two original studies out of 120 original studies in the Web of Science. Also, this study covered a wide range of summary measures, i.e., machine learning methods, sample size, data type, performance measures, important predictors, and whether the VI of the random forest is reported (VI-Yes 1). Current studies on the early diagnosis of depression based on machine learning has the following limitations. Firstly, many studies adopted cross-sectional data and employing longitudinal data would strengthen the performance of machine learning. Secondly, many studies used data with small sizes in single centers. Using big data (e.g., national health insurance claims data) would make valuable contributions for this area. Thirdly, most studies did not consider possible mediating effects among predictors. Fourthly, some studies reported accuracy or the AUC below 70.0 and these results would not be appropriate as diagnostic tests. Likewise, one study reported the accuracy of 100.0 and there could be overfitting in that study. Fifthly, binary categories (no, yes) are popular now but they can be refined to multiple categories with more clinical insights. Sixthly, combining different types of machine learning approaches for different types of depression data would bring new innovations in many aspects. Seventhly, this study compared the performance measures of the six machine learning methods only for different data types. How other data characteristics affect the performance measures of machine learning approaches would be an important topic for future research. In conclusion, however, this study demonstrates that machine learning provides an effective, non-invasive decision support system for early diagnosis of depression.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Kwang-Sig Lee, Byung-Joo Ham. Data curation: Kwang-Sig Lee, Byung-Joo Ham. Formal analysis: Kwang-Sig Lee, Byung-Joo Ham. Funding acquisition: Byung-Joo Ham. Investigation: Kwang-Sig Lee, Byung-Joo Ham. Methodology: Kwang-Sig Lee, Byung-Joo Ham. Project administration: Kwang-Sig Lee, Byung-Joo Ham. Resources: Byung-Joo Ham. Software: Kwang-Sig Lee, Byung-Joo Ham. Supervision: Kwang-Sig Lee, Byung-Joo Ham. Validation: Kwang-Sig Lee, Byung-Joo Ham. Visualization: Kwang-Sig Lee, Byung-Joo Ham. Writing—original draft: Kwang-Sig Lee, Byung-Joo Ham. Writing—review & editing: Kwang-Sig Lee, Byung-Joo Ham.

ORCID iDs

Kwang-Sig Lee <https://orcid.org/0000-0002-0576-0098>
Byung-Joo Ham <https://orcid.org/0000-0002-0108-2058>

Funding Statement

This study was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2020M3E5D9080792).

REFERENCES

1. Mayo Clinic. Depression (major depressive disorder). Available at: <https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007>. Accessed January 28, 2022.
2. World Health Organization. Depression and other common mental disorders: global health estimates. Geneva: World Health Organization; 2017.
3. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: findings from the global burden of disease study. *J Psychiatr Res* 2020;126:134-140.
4. Institute for Health Metrics and Evaluation. Findings from the Global Burden of Disease Study 2017. Seattle: Institute for Health Metrics and Evaluation; 2018.
5. Lee KS, Park JH. Burden of disease in Korea during 2000-10. *J Public Health (Oxf)* 2014;36:225-234.
6. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. *Lancet* 2016;387:1672-1685.
7. Won E, Ham BJ. Imaging genetics studies on monoaminergic genes in major depressive disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 2016;64:311-319.
8. Lee KS, Ahn KH. Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth. *Diagnostics (Basel)* 2020;10:733.
9. Park J, Kim JK, Jung S, Gil Y, Choi JI, Son HS. ECG-signal multi-classification model based on squeeze-and-excitation residual neural networks. *Appl Sci* 2020;10:6495.
10. Goodwin LK, Iannacchione MA, Hammond WE, Crockett P, Maher S, Schlitz K. Data mining methods find demographic predictors of preterm birth. *Nurs Res* 2001;50:340-345.
11. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Inf Sci Syst* 2020;8:14.
12. Fergus P, Cheung P, Hussain A, Al-Jumeily D, Dobbins C, Iram S. Prediction of preterm deliveries from EHG signals using machine learning. *PLoS One* 2013;8:e77154.
13. Gao C, Osmundson S, Velez Edwards DR, Jackson GP, Malin BA, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. *J Biomed Inform* 2019;100:103334.
14. Grigorescu I, Cordero-Grande L, Edwards AD, Hajnal J, Modat M, Deprez M. Interpretable convolutional neural networks for preterm birth classification. *arXiv [Preprint]*. 2019 [Accessed March 1, 2022].

- Available at: <https://doi.org/10.48550/arXiv.1910.00071>.
15. Lee KS, Kim ES, Kim DY, Song IS, Ahn KH. Association of gastroesophageal reflux disease with preterm birth: machine learning analysis. *J Korean Med Sci* 2021;36:e282.
 16. Zhu SL, Dong J, Zhang C, Huang YB, Pan W. Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics. *PLoS One* 2020;15:e0244869.
 17. Cysouw MCF, Jansen BHE, van de Brug T, Oprea-Lager DE, Pfaehler E, de Vries BM, et al. Machine learning-based analysis of [18F]DCFPyL PET radiomics for risk stratification in primary prostate cancer. *Eur J Nucl Med Mol Imaging* 2021;48:340-349.
 18. Park YR, Kim YJ, Ju W, Nam K, Kim S, Kim KG. Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images. *Sci Rep* 2021;11:16143.
 19. Lee KS, Jang JY, Yu YD, Heo JS, Han HS, Yoon YS, et al. Usefulness of artificial intelligence for predicting recurrence following surgery for pancreatic cancer: retrospective cohort study. *Int J Surg* 2021;93:106050.
 20. Lee KS, Park KW. Social determinants of association among diabetes mellitus, visual impairment and hearing loss in a middle-aged or old population: artificial-neural-network analysis of the Korean Longitudinal Study of Aging (2014–2016). *Geriatrics (Basel)* 2019;4:30.
 21. Lee KS, Park KW. Social determinants of the association among cerebrovascular disease, hearing loss and cognitive impairment in a middle-aged or older population: recurrent neural network analysis of the Korean Longitudinal Study of Aging (2014-2016). *Geriatr Gerontol Int* 2019;19:711-716.
 22. Lee KS, Park KW. Artificial intelligence approaches to social determinants of cognitive impairment and its associated conditions. *Dement Neurocogn Disord* 2020;19:114-123.
 23. Ryu KJ, Yi KW, Kim YJ, Shin JH, Hur JY, Kim T, et al. Machine learning approaches to identify factors associated with women's vasomotor symptoms using general hospital data. *J Korean Med Sci* 2021;36:e122.
 24. Lee KS, Kwak HJ, Oh JM, Jha N, Kim YJ, Kim W, et al. Automated detection of TMJ osteoarthritis based on artificial intelligence. *J Dent Res* 2020;99:1363-1367.
 25. Lee KS, Jha N, Kim YJ. Risk factor assessments of temporomandibular disorders via machine learning. *Sci Rep* 2021;11:19802.
 26. Lu S, Shi X, Li M, Jiao J, Feng L, Wang G. Semi-supervised random forest regression model based on co-training and grouping with information entropy for evaluation of depression symptoms severity. *Math Biosci Eng* 2021;18:4586-4602.
 27. Sawangreerak S, Thanathamathee P. Random forest with sampling techniques for handling imbalanced prediction of university student depression. *Information* 2020;11:519.
 28. Shin D, Lee KJ, Adeluwa T, Hur J. Machine learning-based predictive modeling of postpartum depression. *J Clin Med* 2020;9:2899.
 29. Pearson R, Pisner D, Meyer B, Shumake J, Beevers CG. A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychol Med* 2019;49:2330-2341.
 30. Wolfe F, Michaud K. Predicting depression in rheumatoid arthritis: the signal importance of pain extent and fatigue, and comorbidity. *Arthritis Rheum* 2009;61:667-673.
 31. Hu S, Li A, Huang T, Lai J, Li J, Sublette ME, et al. Gut microbiota changes in patients with bipolar depression. *Adv Sci (Weinh)* 2019;6:1900752.
 32. Zhang W, Liu H, Silenzio VMB, Qiu P, Gong W. Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study. *JMIR Med Inform* 2020;8:e15516.
 33. Bhak Y, Jeong HO, Cho YS, Jeon S, Cho J, Gim JA, et al. Depression and suicide risk prediction models using blood-derived multi-omics data. *Transl Psychiatry* 2019;9:262.
 34. Fang J, Wang T, Li C, Hu X, Ngai E, Seet BC, et al. Depression prevalence in postgraduate students and its association with gait abnormality. *IEEE Access* 2019;7:174425-174437.
 35. Zhang X, Cao X, Xue C, Zheng J, Zhang S, Huang Q, et al. Aberrant functional connectivity and activity in Parkinson's disease and comorbidity with depression based on radiomic analysis. *Brain Behav* 2021;11:e02103.
 36. Sau A, Bhakta I. Predicting anxiety and depression in elderly patients using machine learning technology. *Healthc Technol Lett* 2017;4:238-243.
 37. CACHEDA F, FERNANDEZ D, NOVOA FJ, CARNEIRO V. Early detection of depression: social network analysis and random forest techniques. *J Med Internet Res* 2019;21:e12554.
 38. Lin H, Cai X, Zhang D, Liu J, Na P, Li W. Functional connectivity markers of depression in advanced Parkinson's disease. *Neuroimage Clin* 2020;25:102130.
 39. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. *J Am Med Inform Assoc* 2020;27:522-530.
 40. Čukić M, Stokić M, Simić S, Pokrajac D. The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. *Cogn Neurodyn* 2020;14:443-455.
 41. Foster S, Mohler-Kuo M, Tay L, Hothorn T, Seibold H. Estimating patient-specific treatment advantages in the 'Treatment for Adolescents with Depression Study'. *J Psychiatr Res* 2019;112:61-70.
 42. Tennenhouse LG, Marrie RA, Bernstein CN, Lix LM; CIHR Team in Defining the Burden and Managing the Effects of Psychiatric Comorbidity in Chronic Immunoinflammatory Disease. Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. *J Psychosom Res* 2020;134:110126.
 43. Kanders SH, Pisanu C, Bandstein M, Jonsson J, Castela E, Pistis G, et al. A pharmacogenetic risk score for the evaluation of major depression severity under treatment with antidepressants. *Drug Dev Res* 2020;81:102-113.
 44. Elovainio M, Hakulinen C, Pulkki-Råback L, Aalto AM, Virtanen M, Partonen T, et al. General Health Questionnaire (GHQ-12), Beck Depression Inventory (BDI-6), and Mental Health Index (MHI-5): psychometric and predictive properties in a Finnish population-based sample. *Psychiatry Res* 2020;289:112973.
 45. Richter T, Fishbain B, Fruchter E, Richter-Levin G, Okon-Singer H. Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *J Psychiatr Res* 2021;141:199-205.
 46. Zanello-Calzada LA, Galván-Tejada CE, Chávez-Lamas NM, Gracia-Cortés MDC, Magallanes-Quintanar R, Celaya-Padilla JM, et al. Feature extraction in motor activity signal: towards a depression episodes detection in unipolar and bipolar patients. *Diagnostics (Basel)* 2019;9:8.
 47. Manelis A, Iyengar S, Swartz HA, Phillips ML. Prefrontal cortical activation during working memory task anticipation contributes to discrimination between bipolar and unipolar depression. *Neuropsychopharmacology* 2020;45:956-963.
 48. Wade BSC, Sui J, Hellemann G, Leaver AM, Espinoza RT, Woods RP, et al. Inter and intra-hemispheric structural imaging markers predict depression relapse after electroconvulsive therapy: a multisite study. *Transl Psychiatry* 2017;7:1270.
 49. Gibbons RD, Hooker G, Finkelman MD, Weiss DJ, Pilkonis PA, Frank E, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. *J Clin Psychiatry* 2013;74:669-674.
 50. Hu C, Li Q, Shou J, Zhang FX, Li X, Wu M, et al. Constructing a predictive model of depression in chemotherapy patients with non-Hodgkin's lymphoma to improve medical staffs' psychiatric care. *Biomed Res Int* 2021;2021:9201235.
 51. Shimizu Y, Yoshimoto J, Toki S, Takamura M, Yoshimura S, Okamoto Y, et al. Toward probabilistic diagnosis and understanding of depression based on functional MRI data analysis with logistic group LASSO. *PLoS One* 2015;10:e0123524.
 52. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing

- and support for people with depression: a pilot trial in the wild. *JMIR Mhealth Uhealth* 2016;4:e111.
53. Li W, Wang Q, Liu X, Yu Y. Simple action for depression detection: using kinect-recorded human kinematic skeletal data. *BMC Psychiatry* 2021;21:205.
 54. Park Y, Hu J, Singh M, Sylla I, Dankwa-Mullan I, Koski E, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. *JAMA Netw Open* 2021;4:e213909.
 55. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. *JMIR Mhealth Uhealth* 2019;7:e14149.
 56. Kasthurirathne SN, Biondich PG, Grannis SJ, Purkayastha S, Vest JR, Jones JF. Identification of patients in need of advanced care for depression using data extracted from a statewide health information exchange: a machine learning approach. *J Med Internet Res* 2019;21:e13809.
 57. Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, et al. Predicting adherence to internet-delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: machine learning insights from the U-CARE heart randomized controlled trial. *J Med Internet Res* 2018;20:e10754.