Neuroscience Letters 721 (2020) 134804

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research article

Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm

Kyoung-Sae Na[a], Seo-Eun Cho[a], Zong Woo Geem[b], Yong-Ku Kim[c,*]

[a] Department of Psychiatry, Gachon University College of Medicine, Gil Medical Center, Incheon, Republic of Korea
[b] Department of Energy and Information Technology, Gachon University, Seongnam-si, Republic of Korea
[c] Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Mental health
Depression
Prediction
Artificial intelligence
Machine learning

ABSTRACT

Because depression has high prevalence and cause enduring disability, it is important to predict onset of depression among community dwelling adults. In this study, we aimed to build a machine learning-based predictive model for future onset of depression. We used nationwide survey data to construct training and hold-out test set. The class imbalance was dealt with the Synthetic Minority Over-sampling Technique. A tree-based ensemble method, random forest, was used to build a predictive model. Depression was defined by 9 or more on the Center for Epidemiologic Studies – Depression Scale 11 items version. Hyperparameters were tuned throughout the 10-fold cross-validation. A total of 6,588 (6,067 of non-depression and 521 of depression) participants were included in the study. The area under receiver operating characteristics curve was 0.870. The overall accuracy, sensitivity, and specificity were 0.862, 0.730, and 0.866, respectively. Satisfactions for leisure, familial relationship, general, social relationship, and familial income had importance in building predictive model for the onset of future depression. Our study demonstrated that predicting future onset of depression by using survey data could be possible. This predictive model is expected to be used for early identification of individuals at risk for depression and secure time to intervention.

1. Introduction

Depression is defined by symptoms in emotional, cognitive, sleep, appetite, and somatic domains, resulting in functional impairments in activities of daily living [1]. The worldwide prevalence of depression reaches 10.8 % [2], which is the single largest contributor to non-fatal health loss [3]. Several studies have investigated risk factors of depression. A variety of factors may influence the prevalence of depression. Interestingly, a previous study has reported that life satisfaction is associated with depression in the Republic of Korea [4].

Regarding the treatment response of depression, the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study reported that remission for the initial treatment occurred in one third of the whole sample. This study was the largest ever conducted investigating the efficacy of antidepressants [5]. They evaluated 2876 outpatients from 41 sites. Four levels of treatment were provided to the patients when they did not respond to treatment at the previous level. As a result, remission occurred in < 30 % of patients throughout four consecutive therapeutic regimens [6]. The results of the STAR*D study suggest that overall treatment response in depression is very low despite a sufficient period of treatment. Fast and effective treatment modalities are limited; therefore, early detection and prevention should be considered as a priority. Assessing which individuals will experience depression in the near future may lead to more effective prevention programs by focusing on vulnerable individuals prior to onset.

To date, a mounting evidence has accumulated that genetic [7], neuroimaging [8], biological [9], and environmental [10] factors contribute to depression. However, the presence of such risk factors does not always indicate the presence of depression because these factors were obtained using group-level analysis. For example, a comparison of the structural MRI findings of patients with depression and healthy individuals revealed that the hippocampal volume is decreased in the depression group when compared with healthy controls. A smaller hippocampal volume has important implications for understanding the pathophysiology of MDD with respect to stress, glucocorticoid receptors, and inflammatory activities [11]. However, the use of MRI findings for the diagnosis or screening of depression may not provide the correct details because it is possible that a patient and healthy

* Corresponding author at: Department of Psychiatry, Korea University College of Medicine, Korea University Ansan Hospital, 123, Jeokgeum-ro, Danwon-gu, Ansan-si, Gyeonggi-do 15355, Republic of Korea.
E-mail address: yongku@korea.ac.kr (Y.-K. Kim).

https://doi.org/10.1016/j.neulet.2020.134804
Received 17 November 2019; Received in revised form 28 January 2020; Accepted 30 January 2020
Available online 31 January 2020
0304-3940/

individual have a larger and smaller hippocampus, respectively, than the mean hippocampal volume. Hence, an individual-level predictive model is required to assess whether individuals are at risk for developing depression.

Machine learning is based on the individual-level analysis which enables to predict who will develop to depression and who will not in the future [12]. Owing to its practical utility, machine learning has received substantial amount of attention in the field of medicine, including psychiatry [13]. To date, most studies which exploited machine learning were to predict treatment response in the clinical sample [14], or classify depression cross-sectionally among community dwellers [15]. However, to the best of our knowledge, no studies built a machine learning-based model for predicting future onset of depression among community dwellers. The prediction of future depression is important because it secures time to preemptive intervention so that break the load to the depression.

In this study, we aimed to build a machine learning-based predictive model for the onset of depression using nationwide longitudinal survey data from the Korean Welfare Panel Study (KoWePS) [16]. These survey data are suitable for building predictive models because they are a nationally representative sample. In addition, survey data include many variables in different domains, such as sociodemographic, quality of life, physical health, and psychological aspects. We hypothesized that creating a predictive model with these diverse variables would have a discriminative ability from the area under the receiver operating characteristics (AUROC) of 0.8, which is considered a good level for binary classification [17]. In addition, we sought to identify which variables were important for predicting depression. Based on previous evidence [4], we expected that satisfaction for life would have importance in this predictive model.

2. Methods

2.1. Participants and data

We used data from the Korea Welfare Panel Study (KoWePS). A detailed description of the KoWePS has been reported previously [18]. Briefly, the KoWePS was developed to identify longitudinal living conditions and needs for welfare that would contribute to the formation of new policy and institutional improvements in the Republic of Korea. The KoWePS survey includes hundreds of items relating to sociodemographic characteristics, health status, status of economic activity, residence, pension, insurance, living expenses, annual family income, whole fortune, dept, living condition, lifestyle, basic living allowance, use of welfare services, subjective satisfaction, family relationship, and mental health. The sampling frame of the KoWePS corresponds to 230,000 enumeration districts that represent 90 % of the 2005 Population and Housing Census. Islands and special facilities are excluded. By applying double sampling for stratification, 7000 sample households are extracted. The first KoWePS survey began in 2006 wave 1. Data up to 2018 wave 13 is publicly available. We analyzed data from the three most recent years: 2016 wave 11, 2017 wave 12, and 2018 wave 13.

Only participants who responded to the target variable (i.e., depression) and the predictive variables were included in the study. Any participant with missing values in the target and predictive variables was excluded from this study. All participants received full explanation for the aim and protocol of the KoWePS and gave written informed consent. The KoWePS was approved by the Institutional Review Board of the Korea National Institute for Health and Social Affairs.

2.2. Depression and predictive variables

The 11-item version of the Center for Epidemiologic Studies Depression Scale (CES-D-11) was used to measure depression in the KoWePS. To measure persistent depression using the longitudinal survey data, the presence of future depression was defined by a score ≥9 on the CES-D-11 for two consecutive years, 2017 and 2018. The cut-off score 9 was determined according to previous studies [19,20]. The baseline scores on the CES-D-11 in 2016 were included as one of the predictive factors. Additionally, variables of sociodemographic (i.e., age, sex, marital status, religion, types of labor, basic living allowance, current income), quality of life (i.e., satisfaction for leisure, satisfaction for familial relationship, satisfaction for familial income, satisfaction for social relationships, satisfaction for health, satisfaction for job, satisfaction for living environment, satisfaction in general), health (i.e., health status, number of outpatient visit, private medical insurance, number of chronic diseases, presence of disability), and altruistic behaviors (i.e., volunteering activities, donation) were included as predictive variables.

Finally, data from 6588 participants non-depression, n = 6067; depression, n = 521 were used in the machine learning process.

2.3. Machine learning

All machine learning process was conducted with the scikit-learn library implemented in Python 3.7. the dataset was split to the training set and test set. Given the unbalanced ratio of depression and non-depression, Synthetic Minority Over-sampling Technique (SMOTE) was used [21]. To optimize predictive algorithm with hyperparameter tuning, 10-fold cross-validation was conducted within the training set. The hold-out test set, which has never been used in the SMOTE and hyperparameter tuning process, was only used for measuring the performance of the predictive model.

For machine learning algorithm, random forest (RF) was employed. RF is a tree-based model to mitigate the overfitting problem by conducting bootstrap aggregating. The hyperparameters of the RF obtained through the cross-validation were as followed: bootstrap = True, class_weight = (0:1, 1:1), criterion = 'gini', max_depth = 2, max_features = 2, max_leaf_nodes = None, min_impurity_decrease = 0.0, min_impurity_split = None, min_samples_leaf = 1, min_samples_split = 2, min_weight_fraction_leaf = 0.0, n_estimators = 400, n_jobs = -1, oob_score = False, random_state = 1920, verbose = 0, warm_start = False.

2.4. Performance metrics

The AUROC was used as a primary performance metrics. AUROC is defined as the area under the ROC curve, which is plotted in a two-dimensional space with the x and y axes indicating the false positive (1 – specificity) and true positive rates, respectively. The AUROC estimates the classifying results relative to the true positive and false positive rate; therefore, it is recommended for measuring the performance of a binary machine learning model [22]. Generally, the AUROC of 0.8 to 0.9 was considered good and above 0.9 was considered excellent [17]. Other performance metrics such as the overall accuracy ($\frac{true\ posivie\ (TP) + true\ negative\ (TN)}{positive\ (P) + negative\ (N)}$), sensitivity ($\frac{TP}{TP+FN}$), specificity ($\frac{TN}{FP+TN}$), and precision ($\frac{TP}{TP+FP}$) were also used.

3. Results

3.1. Demographic data

A total of 6588 individuals were included in the analysis. Table 1 shows sociodemographic, economic, psychosocial, and clinical variables according to the depression. The prevalence of depression at baseline (2016), 1-year (2017), and 2-year (2018) was 7.9 % (n = 521), 7.7 % (n = 506), and 9.3 % (n = 612), respectively.

From a sociodemographic perspective, the depression group was significantly older than the non-depression group ($p < 0.001$). In addition, the there was a lower frequency of marriage and higher frequency or divorce or unmarried in the depression group when

Table 1
Sociodemographic, economic, and clinical variables by depression.

Variables	Not-depression (n = 6379)	Depression (n = 209)	*t* or χ^2	*P value*
Age	44.73 (12.06)	48.49 (12.41)	−4.43	< 0.001
Sex, male	2909 (45.6)	64 (30.6)	18.34	< 0.001
Religion, yes	2871 (45.01)	98 (46.89)	0.29	0.590
Marital status			118.93	< 0.001
Unmarried	1381 (21.65)	52 (24.88)		
Married	4358 (68.32)	89 (45.58)		
Widow	187 (2.93)	16 (7.66)		
Divorced	453 (7.10)	52 (24.88)		
Types of labor			78.38	< 0.001
Wage and salary worker	3,569 (55.95)	78 (37.32)		
Self-employed or employer	820 (12.85)	13 (6.22)		
Unpaid family worker	255 (4.00)	5 (2.39)		
Unemployed (preserved ability to work)	1,726 (27.06)	111 (53.11)		
Unemployed (inability to work)	9 (0.14)	2 (0.96)		
Receiving basic living allowance, yes	319 (5.00)	76 (36.36)	353.17	< 0.001
Current income (10,000 won)[a]	6328.71 (9166.54)	3377.02 (2905.88)	4.51	< 0.001
Chronic diseases, number			71.44	< 0.001
0	4108 (64.40)	82 (39.23)		
1	258 (4.04)	4 (1.91)		
2	137 (2.15)	5 (2.39)		
> 3	1876 (29.41)	118 (56.46)		
Volunteer activities, yes	817 (12.81)	16 (7.66)	4.86	0.027
Annual Donation (10,000 won) [a]	44.19 (1812.33)	24.43 (37.41)	0.433	0.665
Private health insurance, number, yes	1.55 (1.28)	1.00 (1.19)	6.12	< 0.001
Disability, yes	343 (5.38)	44 (21.05)	89.94	< 0.001
Health status	2.13 (0.78)	2.93 (0.92)	−14.45	< 0.001
Satisfaction for health	3.63 (0.85)	2.62 (1.04)	17.24	< 0.001
Satisfaction for familial income	3.07 (0.93)	2.29 (0.85)	11.90	< 0.001
Satisfaction for living environment	3.61 (0.79)	3.11 (0.99)	9.12	< 0.001
Satisfaction for familial relationship	4.01 (0.63)	3.31 (0.90)	15.66	< 0.001
Satisfaction for job	3.57 (0.80)	2.79 (0.88)	13.68	< 0.001
Satisfaction for social relationship	3.84 (0.60)	3.21 (0.87)	14.57	< 0.001
Satisfaction for leisure	3.40 (0.83)	2.76 (0.79)	11.16	< 0.001
Satisfaction in general	3.70 (0.62)	2.91 (0.79)	17.93	< 0.001
Familial stress	6.45 (2.24)	7.79 (2.58)	−9.09	< 0.001
Outpatients visit per year, number	9.44 (20.26)	30.81 (45.40)	−14.14	< 0.001
CES-D-11, raw scores				
2016	2.23 (3.60)	9.23 (7.25)	−26.45	< 0.001
2017	2.09 (3.12)	14.10 (4.64)	−53.80	< 0.001
2018	2.22 (3.44)	14.28 (5.03)	−49.00	< 0.001
CES-D-11, frequency of depression in 2016	410 (6.43)	111 (53.11)	605.55	< 0.001

CES-D-11: The Center for Epidemiologic Studies Depression Scale 11-item version.
[a] 1,000 won is approximately 8.3 dollar.

compared with the non-depression group ($p < 0.001$). There were no significant differences in the distribution of sex and religion between groups.

There were significant differences in the economic and health-related variables between groups. There was a higher proportion of individuals in the depression group that were receiving basic living allowance (36.36 %) when compared with the non-depression group (5.00 %) ($p < 0.001$). The mean current income of the depression group (2905.88 won) was approximately a quarter of the depression group (9166.54 won) ($p = 0.001$).

In the domain of quality of life, there was lower satisfaction for health ($p < 0.001$), familial income ($p < 0.001$), living environment ($p < 0.001$), familial relationship ($p < 0.001$), job ($p < 0.001$), social relationship ($p < 0.001$), leisure ($p < 0.001$), and general ($p < 0.001$) in the depression groups when compared with the non-depression group.

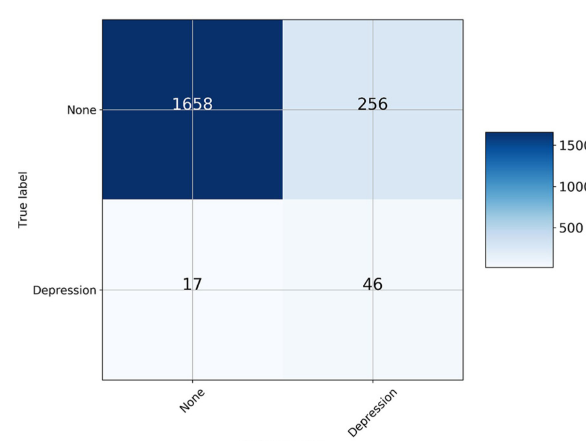

Fig. 1. Confusion matrix.

3.2. Predictive performance

The confusion matrix (Fig. 1) shows that 1,704 out of 1,977 individuals were correctly classified (overall accuracy = 0.862). Importantly, 1,658 out of 1,914 non-depressed individuals were correctly classified as non-depressed (specificity = 0.866). Furthermore, 256 out of 1,914 non-depressed individuals were misclassified as depressed (false positive rate = 0.134). In addition, 46 out of 63 depressed were correctly classified as depressed (sensitivity = 0.730), whereas 17 out of 63 were misclassified as non-depressed (false negative rate = 0.270). In contrast, 1,658 out of the 1,675 individuals who were predicted to be non-depressed were correctly classified (negative predictive value = 0.990). In addition, 46 out of the 302 individuals who were

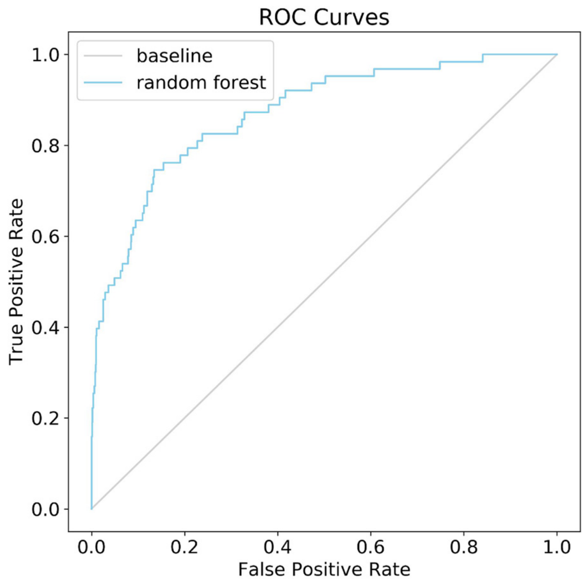

Fig. 2. Area under the receiver operating characteristics curve.

Table 2
Performance metrics of the predictive model for future depression.

Accuracy	AUROC	Sensitivity	Specificity	NPV	Precision
0.862	0.870	0.730	0.866	0.990	0.152

AUROC: area under the receiver operating characteristic curve NPV: negative predictive value.

predictive to be depressed were correctly classified (precision = 0.152).

Fig. 2 shows the AUROC, which represent the area under the receiver-operating curve. The classifier is no better than random selection when the area is 0.5 as drawn by the baseline. We calculated that the AUROC of our predictive model was 0.870. The AUROC, accuracy, specificity, sensitivity, precision, false positive rate, and false negative rates are summarized in Table 2.

3.3. Feature importance

The most predictive value was baseline satisfaction for health, which was followed by satisfaction for social relationship, health status, satisfaction for familial relationship, satisfaction for job, satisfaction in general, and satisfaction for leisure (Fig. 3).

4. Discussion

In this study, we successfully built a machine learning-based predictive model for future depression. The AUROC (0.870), overall accuracy (0.862), sensitivity (0.730), and specificity (0.866) shows that this model could be practically used for screening adults who are prone to have depression later. To the best of our knowledge, this machine learning-based model is the first to predict future onset of depression by using longitudinal nationally representative community-dwelling adults. It was not possible to directly compare our results with previous studies that used clinical samples; therefore, the level of predictability is higher than previous studies with clinical samples [23,24] in which the AUROC values ranged from 0.66 to 0.69 [23] and 0.63 to 0.76 [25]. We suggest that the superiority of our predictive model over the studies of clinical samples might have arisen from the heterogeneity and severity in the clinical sample. Indeed, clinical sample of depression consists of highly heterogeneous individuals in perspective of clinical [26], genetic variants [7], and peripheral risk factors such as inflammation [27]. Hence, once depression had occurred, it would be difficult to accurately predict prognosis of depression. That is the point of which this study targets. By focusing on the prediction of future depression among community sample, we could establish predictive model with good performance measure.

From the analysis of feature importance in the RF algorithm, baseline depression showed the most importance. However, the importance of the baseline depression alone was less than 10 % of total feature importance. Indeed, only 40 out of 283 baseline depressive cases remained to be depressed for the consecutive two years.

Interestingly, domains in the quality of life such as satisfaction for health (1st), satisfaction for social relationship (2nd), satisfaction for familial relationship (4th), satisfaction for job (5th), satisfaction in general (6th), and satisfaction for leisure (7th) scored high importance. The satisfaction for various domains of life has become increasingly important these days. In fact, familial and social relationship have been also important in the prevention and improvement of depression in the community sample [28,29]. Leisure activity had preventive effects for depression [30]. These results suggest that the social and familial connectedness and intimacy as measured by the satisfaction for familial and social relationship seem to reflect susceptibility for depression.

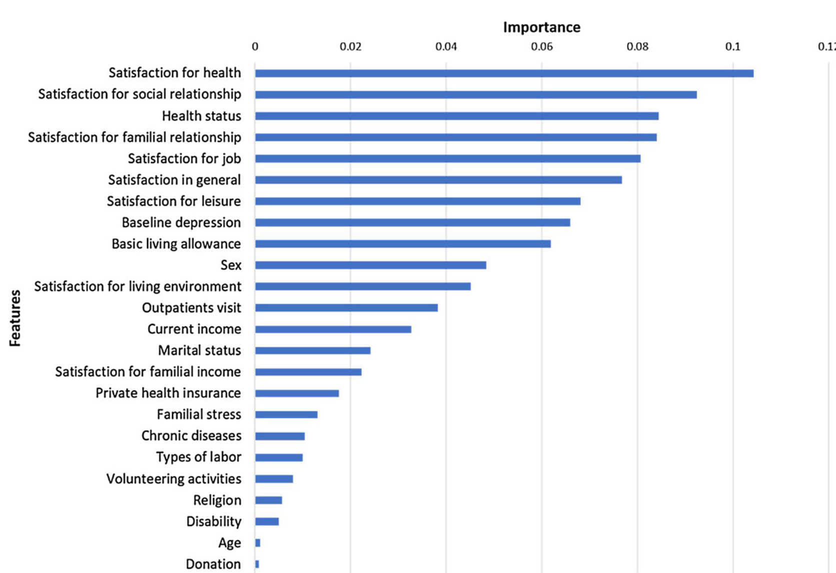

Fig. 3. Feature importance.

On the other hand, economic issue had also importance for building predictive model. Receiving basic living allowance (9^{th}) was in the top ten of feature importance. This is in line with a recent study which reported that the income of house has close association with depression [31]. These results raise a possibility that although socioeconomic status is not necessarily related with the onset of depression, poor economic conditions may limit to the access to the healthcare utilization, which in turn contribute to depression.

There are several limitations to be discussed. We did not include biomarkers in our model due to the fundamental limit of the original survey in the KoWePS. Since there have been reported that numerous biological markers contribute to the pathophysiology of depression, particularly genetic variants, [7] the absence of biological factors could limit the predictive performance in this study. However, due to reports that numerous biological markers contribute, biological factors themselves do not guarantee successful prediction. For example, a previous study revealed that single use of biomarkers for predicting prognosis of depression resulted poor performance (AUROC < 0.6) [23].

In summary, this study demonstrates machine learning-based predictive model for depression among adult community dwellers. Further studies that incorporate various types of predictive variables, such as biological, genetic, neuroimaging, and socio-environmental factors, are warranted. In addition, the optimal time and method of intervention for individuals who are at risk for depression should also be investigated in the near future.

CRediT authorship contribution statement

Kyoung-Sae Na: Conceptualization, Methodology, Validation, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Seo-Eun Cho:** Conceptualization, Methodology. **Zong Woo Geem:** Methodology. **Yong-Ku Kim:** Conceptualization, Methodology, Validation, Investigation, Data curation, Supervision.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. 2017R1C1B5073684). This paper is based on the PhD dissertation of Kyoung-Sae Na.

References

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th edition, American Psychiatric Publishing, Washington, USA, 2013 (DSM–5).
[2] G.Y. Lim, W.W. Tam, Y. Lu, C.S. Ho, M.W. Zhang, R.C. Ho, Prevalence of depression in the community from 30 countries between 1994 and 2014, Sci. Rep. 8 (2018) 2861.
[3] World Health Organization, Depression and Other Common Mental Disorders: Global Health Estimates, World Health Organization, Geneva, Switzerland, 2017.
[4] E.H. Seo, S.G. Kim, S.H. Kim, J.H. Kim, J.H. Park, H.J. Yoon, Life satisfaction and happiness associated with depressive symptoms among university students: a cross-sectional study in Korea, Ann. Gen. Psychiatry 17 (2018) 52.
[5] National Institute of Mental Health, Questions and Answers About the NIMH Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Study — All Medication Levels, (2016).
[6] National Institute of Mental Health, Questions and Answers About the NIMH Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Study — All Medication Levels Vol. 2019 (2006).
[7] D.M. Howard, M.J. Adams, T.K. Clarke, J.D. Hafferty, J. Gibson, M. Shirali, J.R.I. Coleman, S.P. Hagenaars, J. Ward, E.M. Wigmore, C. Alloza, X. Shen, M.C. Barbu, E.Y. Xu, H.C. Whalley, R.E. Marioni, D.J. Porteous, G. Davies, I.J. Deary, G. Hemani, K. Berger, H. Teismann, R. Rawal, V. Arolt, B.T. Baune, U. Dannlowski, K. Domschke, C. Tian, D.A. Hinds, T. Me Research, C. Major Depressive Disorder Working Group of the Psychiatric Genomics, M. Trzaskowski, E.M. Byrne, S. Ripke, D.J. Smith, P.F. Sullivan, N.R. Wray, G. Breen, C.M. Lewis, A.M. McIntosh, Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions, Nat. Neurosci. 22 (2019) 343–352.
[8] L.P. Pereira, C.A. Kohler, B. Stubbs, K.W. Miskowiak, G. Morris, B.P. de Freitas, T. Thompson, B.S. Fernandes, A.R. Brunoni, M. Maes, D.A. Pizzagalli, A.F. Carvalho, Imaging genetics paradigms in depression research: systematic review and meta-analysis, Prog. Neuropsychopharmacol. Biol. Psychiatry 86 (2018) 102–113.
[9] A.F. Carvalho, C.A. Kohler, R.S. McIntyre, C. Knochel, A.R. Brunoni, M.E. Thase, J. Quevedo, B.S. Fernandes, M. Berk, Peripheral vascular endothelial growth factor as a novel depression biomarker: a meta-analysis, Psychoneuroendocrinology 62 (2015) 18–26.
[10] S.A. Stuart, J.K. Hinchcliffe, E.S.J. Robinson, Evidence that neuropsychological deficits following early life adversity may underlie vulnerability to depression, Neuropsychopharmacology 44 (2019) 1623–1630.
[11] Y.K. Kim, K.S. Na, A.M. Myint, B.E. Leonard, The role of pro-inflammatory cytokines in neuroinflammation, neurogenesis and the neuroendocrine system in major depression, Prog. Neuropsychopharmacol. Biol. Psychiatry 64 (2016) 277–284.
[12] R.J. Janssen, J. Mourao-Miranda, H.G. Schnack, Making individual prognoses in psychiatry using neuroimaging and machine learning, Biol. Psychiatry Cogn. Neurosci. Neuroimaging 3 (2018) 798–808.
[13] D.B. Dwyer, P. Falkai, N. Koutsouleris, Machine learning approaches for clinical psychology and psychiatry, Annu. Rev. Clin. Psychol. 14 (2018) 91–118.
[14] Z. Nie, S. Vairavan, V.A. Narayan, J. Ye, Q.S. Li, Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study, PLoS One 13 (2018) e0197268.
[15] J. Oh, K. Yun, U. Maoz, T.S. Kim, J.H. Chae, Identifying depression in the National Health and Nutrition Examination Survey data using a deep learning algorithm, J. Affect. Disord. 257 (2019) 623–631.
[16] Korea Institute for Health and Social Affairs, Social Welfare Research Center of Seoul National University, Korean Welfare Panel Study User Guidelines, Seoul (2017).
[17] D.W. Hosmer Jr, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression Vol. 398 John Wiley & Sons, NY, 2000.
[18] Korea Institute for Health and Social Affairs, Korea Welfare Panel Study Vol. 2019 (2014) Sejong.
[19] E. Torres, Psychometric properties of the center for epidemiologic studies depression scale in african american and Black Caribbean US adults, Issues Ment. Health Nurs. 33 (2012) 687–696.
[20] M. Hoe, Exploring latent trajectory classes of change in depression measured using CES-D, Korean J. Soc. Welfare 66 (2014) 307–331.
[21] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002).
[22] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recogn. 30 (1997) 1145–1159.
[23] R. Dinga, A.F. Marquand, D.J. Veltman, A.T.F. Beekman, R.A. Schoevers, A.M. van Hemert, B. Penninx, L. Schmaal, Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach, Transl. Psychiatry 8 (2018) 241.
[24] Y. Lee, R.M. Ragguett, R.B. Mansur, J.J. Boutilier, J.D. Rosenblat, A. Trevizol, E. Brietzke, K. Lin, Z. Pan, M. Subramaniapillai, T.C.Y. Chan, D. Fus, C. Park, N. Musial, H. Zuckerman, V.C. Chen, R. Ho, C. Rong, R.S. McIntyre, Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review, J. Affect. Disord. 241 (2018) 519–532.
[25] R.C. Kessler, H.M. van Loo, K.J. Wardenaar, R.M. Bossarte, L.A. Brenner, T. Cai, D.D. Ebert, I. Hwang, J. Li, P. de Jonge, A.A. Nierenberg, M.V. Petukhova, A.J. Rosellini, N.A. Sampson, R.A. Schoevers, M.A. Wilcox, A.M. Zaslavsky, Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports, Mol. Psychiatry 21 (2016) 1366–1371.
[26] M.M. Weissman, K.R. Merikangas, P. Wickramaratne, K.K. Kidd, B.A. Prusoff, J.F. Leckman, D.L. Pauls, Understanding the clinical heterogeneity of major depression using family data, Arch. Gen. Psychiatry 43 (1986) 430–434.
[27] K.S. Na, K.J. Lee, J.S. Lee, Y.S. Cho, H.Y. Jung, Efficacy of adjunctive celecoxib treatment for patients with major depressive disorder: a meta-analysis, Prog. Neuropsychopharmacol. Biol. Psychiatry 48 (2014) 79–85.
[28] A.R. Teo, H. Choi, M. Valenstein, Social relationships and depression: ten-year follow-up from a nationally representative study, PLoS One 8 (2013) e62396.
[29] M. Gilligan, J.J. Suitor, S. Nam, B. Routh, M. Rurka, G. Con, Family networks and psychological well-being in Midlife, Soc. Sci. 6 (2017) 94.
[30] H.Y. Lee, C.P. Yu, C.D. Wu, W.C. Pan, The effect of leisure activity diversity and exercise time on the prevention of depression in the middle-aged and elderly residents of Taiwan, Int. J. Environ. Res. Public Health 15 (2018).
[31] M. Economou, L.E. Peppou, K. Souliotis, G. Konstantakopoulos, T. Papaslanis, K. Kontoangelos, S. Nikolaidi, N. Stefanis, An association of economic hardship with depression and suicidality in times of recession in Greece, Psychiatry Res. 279 (2019) 172–179.