



OPEN ACCESS

EDITED BY

Harshavardhan Sampath,
Sikkim Manipal University, India

REVIEWED BY

Seyed-Ali Sadegh-Zadeh,
Staffordshire University, United Kingdom
Delia Virga,
West University of Timișoara, Romania

*CORRESPONDENCE

Minji Gil

✉ zoemjgil@gmail.com

Eun Jeong Min

✉ ej.min@catholic.ac.kr

RECEIVED 06 April 2023

ACCEPTED 12 June 2023

PUBLISHED 12 July 2023

CITATION

Kim S-S, Gil M and Min EJ (2023) Machine learning models for predicting depression in Korean young employees.

Front. Public Health 11:1201054.

doi: 10.3389/fpubh.2023.1201054

COPYRIGHT

© 2023 Kim, Gil and Min. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning models for predicting depression in Korean young employees

Suk-Sun Kim¹, Minji Gil^{1*} and Eun Jeong Min^{2*}

¹College of Nursing, Ewha Womans University, Seoul, Republic of Korea, ²Department of Medical Life Sciences, School of Medicine, The Catholic University of Korea, Seoul, South Korea

Background: The incidence of depression among employees has gradually risen. Previous studies have focused on predicting the risk of depression, but most studies were conducted using basic statistical methods. This study used machine learning algorithms to build models that detect and identify the important factors associated with depression in the workplace.

Methods: A total of 503 employees completed an online survey that included questionnaires on general characteristics, physical health, job-related factors, psychosocial protective, and risk factors in the workplace. The dataset contained 27 predictor variables and one dependent variable which referred to the status of employees (normal or at the risk of depression). The prediction accuracy of three machine learning models using sparse logistic regression, support vector machine, and random forest was compared with the accuracy, precision, sensitivity, specificity, and AUC. Additionally, the important factors identified via sparse logistic regression and random forest.

Results: All machine learning models demonstrated similar results, with the lowest accuracy obtained from sparse logistic regression and support vector machine (86.8%) and the highest accuracy from random forest (88.7%). The important factors identified in this study were gender, physical health, job, psychosocial protective factors, and psychosocial risk and protective factors in the workplace.

Discussion: The results of this study indicated the potential of machine learning models to accurately predict the risk of depression among employees. The identified factors that influence the risk of depression can contribute to the development of intelligent mental healthcare systems that can detect early signs of depressive symptoms in the workplace.

KEYWORDS

machine learning, depression, employees, workplace, prediction

Introduction

Depression in the workplace and young adults have become a growing global concern due to greater societal costs and reduced work productivity (1). In 2019, the incremental economic burden of Korean adults with depression was an estimated a total of KRW 4.83 trillion, with 14.3% in direct costs and 85.6% in indirect costs. Among indirect costs, workplace costs accounted for the highest percentage (65.6%) including absenteeism (18%) and presenteeism (42%) (2). Employees' depression severity increased direct costs through utilization of mental health services and indirect costs attributable to overall work impairment (1, 3).

The highest prevalence of depression in young adults especially contributes to increasing an enormous economic and social burden (4). Approximately 25% of Korean younger adults aged 19–39 years experienced moderately severe depression, compared with about 18% of adults aged 40–50 years (5). However, only 7.4% of Korea employees had a diagnosis of depression by a doctor, while more than 20% had a diagnosis of depression in Canada (20.7%), United States (22.7%) and South Africa (25.6%) (6). Owing to fear of losing their job and mental health stigma, they were reluctant to disclose mental health problems and seek mental health service (6, 7).

Previous studies commonly used traditional statistical methods, such as regression analysis, to infer the relationships between depression and specific variables (8–10) which were derived from work-related theories such as the conservation of resources theory (11), self-determination theory (12), and the job demands-resources theory (13). Several studies identified factors associated with depression, including sociodemographic factors, traits, stressors, relationship stability, and cognitive processes (14, 15). In particular, work-related factors such as long working hours, workload, and burnout have been found to increase the risk of depression among employees (16, 17). However, traditional statistical methods are limited on representing real-world complexities and predicting future data due to their assumption of linearity between variables (18, 19).

Traditional statistical methods aim to test hypotheses which derive from theories, while machine learning (ML) methods focus on discovering hidden interaction in the specific data-set to make predictions (8, 9, 20). ML methods provide more accurate prediction by analyzing complex and non-linear interactions among datasets rather than separately considering the effect of one variable on an outcome of interest (21, 22). ML methods can facilitate early detection by predicting the risk of diseases (23–25). However, there were few studies on the prediction of depression using ML algorithms in young children (26), university students (27), and older adults (24).

Therefore, this study aimed to evaluate the performance the performance of different ML algorithms, such as sparse logistic regression, support vector machines (SVM), and random forest (RF), and identify the important factors influencing the risk of depression among Korean employees.

Methods

Data and sample

The target population were Millennial and Generation Z (MZ) employees in South Korea. This study included participants who were aged between 20 and 40; Millennials were born between 1983 to 1994, and Generation Zs were born between 1995 to 2004.

The Ewha womans university institutional review board approved this study (ewha-202206-0001-01). Potential participants were recruited from a website and social networking services during June 2022. Of the 505 employees completed the survey *via* the online survey platform, we excluded two participants (0.4%) who were over the age of 40. A total of 503 employees were used for data analysis.

Outcome variable

The outcome variable was the depression CES-D score (28), which consists of 20 items that are rated on a 4-point Likert scale (0–3). Possible score ranges from 0 to 60, with a higher score indicating more depression-related symptoms. The Cronbach's alpha of the CES-D was 0.85–0.90 (28) and 0.91 in the current study. Based on the CES-D cutoff score ≥ 16 (29), we divided the young employees into two groups: normal ($n = 176$) and at risk of depression ($n = 327$).

Predictor variables

The predictor variables consisted of a set of demographics, physical health-related, job-related, and study variables that were selected based on literature reviews of the risk and protective factors for depression among employees. Study variables included personality-related variables, psychosocial protective variables, psychosocial risk variables in the workplace, and psychosocial protective variables in the workplace. Cronbach's alpha test was used to determine inter-item reliability (Table 1). However, due to the potential for misleading results stemming from limited item variance (30, 31), Cronbach's alpha test was not applied to measures with fewer than three items, including the 10-item Big Five Inventory (32, 33) and relationship questions (34).

Demographic characteristics included age, gender, religion, and marital status. Age was used as a continuous variable, while the rest of variables were used as dummy variables.

Personality-related factors included physical activity per week; the amount of sleep and number of meals per day; and drinking and smoking. The frequency of physical activity per week and the number of meals per day were used as continuous variables, while the rest of the variables were used as dummy variables.

Regarding job characteristics, we examined total years of job experience, employment period at their current workplace, number of turnovers, weekly working hours, monthly salary, and income satisfaction.

As psychosocial protective factors, we considered personality, grit, attachment, satisfaction with life, and interpersonal relationships. For personality, the 10-item Big Five Inventory (32, 33) was used. Grit was measured using the short grit scale (35), and attachment was measured by the relationship questions (34). The satisfaction with life scale (36) and relationship change scale (37) were also included.

Psychosocial risk factors in the workplace included burnout and occupational stress. Burnout was measured by the burnout assessment tool (38). Occupational stress was measured by the Korean Occupational Stress Scale-Short Form (39).

Psychosocial protective factors in the workplace included occupational self-efficacy, social problem-solving style, meaning in work, work-life balance, and psychological safety. Occupational self-efficacy was assessed using the occupational self-efficacy scale (40), social problem-solving style was evaluated using the Social Problem-Solving Inventory-Revised Short Form (41), meaning in work was measured by using the Working and Meaning Inventory (42), work-life balance was assessed using the Work-life Balance Scale (43), and psychological safety was evaluated using the Team Psychological Safety Scale (44).

TABLE 1 Predictor variables.

Factors	Variables	Indicator	Cronbach' α
Demographic	Age		
	Gender:	Male = 0, Female = 1	
	Religion	Nonreligious = 0, Religious = 1	
	Marital status	Single = 0, Married = 1	
Physical health-related factors	Psychical activity per week		
	Amount of sleep	Less than 6 h = 0 More than 6 h = 1	
	Number of meals per day		
	Drinking	Nondrinker = 0, Drinker = 1	
	Smoking	Nonsmoker = 0, Smoker = 1	
Job-related factors	Total years of job experience		
	Employment period at their current workplace		
	Number of turnovers		
	Weekly working hours	Less than 40 h = 0 40 h = 1 From 40 to 52 h = 2 More than 52 h = 3	
	Monthly salary		
Psychosocial protective factors	Income satisfaction	Dissatisfied = 0 Satisfied = 1	
	Personality (10-item big five inventory)	Extraversion, Conscientiousness, Openness to experience, Neuroticism, Agreeableness	
	Grit (Short Grit scale)	Consistency of interest, Perseverance of effort	0.63
	Attachment (Relationship questions)	Secure, Fearful, Preoccupied, Dismissing attachment.	
	Satisfaction with life (Satisfaction with Life Scale)		0.84
Psychosocial risk factors in the workplace	Interpersonal relationships (Relationship Change Scale)	Satisfaction, Communication, Trust, Intimacy, Sensitivity, Openness, Understanding.	0.90
	Burnout (Burnout Assessment Tool)	Exhaustion, Mental distance, Emotional impairment, Cognitive impairment	0.88
Psychosocial protective factors in the workplace	Occupational stress (Korean Occupational Stress Scale- Short Form)	Job demands, Degree of autonomy, Job instability, Organizational system, Lack of reward, Occupational climate	0.79
	Occupational self-efficacy (Occupational Self-Efficacy Scale)		0.87
	Social Problem-Solving style (Social Problem-Solving Inventory-Revised-Short Form)	Positive problem orientation, Negative problem orientation, Rational problem solving, Impulsive/careless style, Avoidance style	0.76
	Meaning in work (Working and Meaning Inventory)	Positive meaning in work, Meaning-making through work, Greater good motivation	0.87
	Work-Life Balance (Work-Life Balance Scale)	Family, Leisure, Growth, Life	
	Psychological safety (Team Psychological Safety Scale)		0.54

With advances in data science technology, this study demonstrated the practical applicability of ML algorithms in predicting the risk of depression among MZ employees. We applied three different ML algorithms – sparse logistic regression, RF, and SVM. We found the highest accuracy of RF. Our study identified the important variables influencing the risk of depression among Korean employees such as

gender, inadequate sleep, smoking habits, occupational stress, burnout, social problem-solving styles, sense of meaning at work, attachment, interpersonal relationships, and satisfaction in life. These findings contribute to the development of intelligent mental healthcare systems for the early detection of depression. Additionally, our study can help develop target interventions designed to prevent employees'

depression and provide a situation-specific theory that predicts depression among MZ employees. However, this study focuses solely on MZ employees, and thus, careful consideration is recommended before generalizing these findings to other demographic groups.

Statistical analysis

Prediction models

Our goal is to predict the class of a sample given set of predictor variable values. Three different models are considered for the risk prediction: sparse logistic regression, support vector machine, and random forest. Logistic regression is one of the most widely used statistical prediction model for binary classification problem. When the response variable is binary, logistic regression predicts the probability to be classified to one of two groups given a set of covariate values. The model has the nice property that the estimated coefficients is log odds ratio. However, it is hard to interpret the results when the number of variables gets bigger. To overcome this problem, sparse logistic regression that employs least absolute shrinkage and selection operator (LASSO) in the model is considered in our problem. This model conducts feature selection and the estimation simultaneously, which enables interpretation with few selected important predictors.

Support vector machine is a famous machine learning technique for the binary classification problem. SVM seeks a decision boundary that well separate the data into two groups. It is well known that SVM performs well when data exhibit high-dimensionality while its computational cost is relatively cheap compared to other machine learning methods (45).

Random forest is the representative classification method of ensemble models, which consists of many decision trees. Ensemble method is an approach to combines prediction results from numerous algorithms to improve prediction power by avoiding overfitting. Random forest aggregates the prediction results from many decision trees to make the final decision.

To evaluate the prediction performance of above three estimated models, we used 70% of dataset for the estimation of models and last 30% of data were used for the test. To select the optimal hyperparameters in each model, 5-fold cross validation (CV) were used. For sparse logistic regression, we re-fit logistic regression using only selected features from sparse logistic regression with CV to avoid possible bias of results. Various measures including accuracy, precision, sensitivity, specificity, F1, and AUC were calculated to compare prediction performance. All statistical analyses were performed using R version 4.1.13 statistical package (R Project for Statistical Computing).

Results

Table 2 shows the participants' characteristics by their level of depressive symptoms. The *p*-values were computed using the t-test or chi-square test depending on the type of each variable. The results demonstrated that gender and marital status were significantly different between the normal and depression-risk groups.

Performance measures calculated from analysis results of three methods are shown in Table 3. While all three models show relatively comparable performance, random forest shows little higher

performance on every performance measure except AUC. This implies that random forest predicts the risk of depression more accurately. Also, random forest ranks the importance of features by comparing the reduction of average impurity on prediction for each feature. Figure 1 shows the results including first five features selected as important variables: Cognitive impairment, Negative problem orientation, emotional impairment, satisfaction with life, and employment period at their current workplace.

It is worth to note that sparse logistic regression shows better performance on AUC compared to random forest. These results imply that sparse logistic regression shows better performance in terms of arrangement of samples from low depression probability to high probability. In addition to that, sparse logistic regression yields interpretable results with the estimated effect size of selected features. Table 4 includes the estimated result of logistic regression model using selected features from the sparse logistic regression. Gender, amount of sleep hours, smoking status, year of job experience, weekly working hours, interpersonal relationships, occupational stress, and social problem-solving style were shown to be important to predict the risk of depression for workers in Korea. Several features including negative problem orientation, emotional impairment is commonly shown as important features for the risk prediction in both random forest and sparse logistic regression models. Figure 1 shows significant RF variables. Burnout, social problem-solving style, and satisfaction with life were selected as important features in the RF.

Discussion

This study provides evidence that the ML algorithms can help reduce bias and accurately predict the likelihood of depression among MZ employees. The main strength of this study was the use of the MZ employees' dataset to predict and identify personal and work-related factors of depression using ML techniques. MZ employees exhibited unique depression symptoms, such as narcissistic tendencies, a feeling of victimization from supervisors, difficulty accepting criticism, and an inferiority complex (46). Our study applied to advanced ML techniques to improve intelligent mental healthcare systems which will be used to detect early depressive symptoms and increase access to mental health services for MZ employees in Korea.

The primary objective of this study is to develop ML algorithms to predict the risk of depression among MZ employees. This objective differs from hypothesis testing, which seeks statistically effective variables in relationship with response variables. In this case, too small sample size leads to low power of the test, which means inefficient use of resources including data and time. To overcome the problem, often optimal sample size calculation for getting enough power based on the expected type I and II error of hypothesis test can be conducted (47). This approach can be wildly found in randomized control trials (RCT), especially clinical trials.

On the other hands, our study is focused on developing the predictive model, and we assess their performance using several prediction performance measures such as sensitivity and specificity. To assess the generalizability of their prediction power, ML typically uses partial portion of the data called the training dataset, and test their prediction performance using the data unused for the model construction, called the test dataset. The test set is used to get the accuracy of the ML algorithms and assess model performance

TABLE 2 Socio-demographic characteristics (N=503).

		Young employees' depression		<i>p</i> -Values
		Normal range (<i>n</i> =176, 35.0%)	Risk of depression (<i>n</i> =327, 65.0%).	
Age	Median [1st, 3rd quartile]	32.00 [29.00, 3,500]	32.00 [30.00, 36.00]	0.204
Gender				
Male	<i>n</i> (%)	81 (46.0%)	112 (34.3%)	0.010
Female	<i>n</i> (%)	95 (54.0%)	215 (65.7%)	
Religion				
Nonreligious	<i>n</i> (%)	115 (65.3%)	220 (67.3%)	0.661
Religious	<i>n</i> (%)	61 (34.7%)	107 (32.7%)	
Marital status				
Single	<i>n</i> (%)	109 (61.9%)	245 (74.9%)	0.003
Married	<i>n</i> (%)	67 (38.1%)	82 (25.1%)	
Work related characteristics				
Total years of experience	Median [1st, 3rd quartile]	6.00 [3.00, 10.00]	5.00 [3.00, 9.00]	0.482
Employment period at their current workplace	Median [1st, 3rd quartile]	3.00 [1.00, 5.00]	4.00 [2.00, 5.00]	0.001
Number of turnovers	Median [1st, 3rd quartile]	1.00 [0.00, 2.00]	0.00 [0.00, 2.00]	0.371
Monthly salary (mln won)	Median [1st, 3rd quartile]	2.80 [2.50, 3.50]	3.00 [2.50, 3.20]	0.453

TABLE 3 Performance of machine learning algorithms.

	Accuracy	Precision	Sensitivity	Specificity	AUC
Sparse logistic regression	0.8675	0.8700	0.9255	0.7719	0.9171
SVM	0.8675	0.8491	0.9574	0.7192	0.8384
Random Forest	0.8874	0.8889	0.9362	0.8070	0.8716

regardless of sample size (10, 48). This approach enables us to estimate a prediction error (generalizability) on new data. In our study, the dataset was divided into a training set (70%) and a testing set (30%). Also, cross-validation method the we employed for ML model training is one of the representative devised method to overcome the situation with not enough sample size situation.

Machine learning techniques, sparse logistic regression, SVM, and RF, were applied to develop models for predicting the risk of depression among MZ employees. In this study, sparse logistic regression, SVM, and RF techniques yielded very close accuracies, with RF being slightly higher. Feature selection performed using sparse logistic regression and RF showed similar variables as the important factors of the risk of depression. Sparse logistic Regression provided interpretable results *via* feature selection procedure with the estimated effect size and *p*-values for testing its significance, while RF gives a comprehensive view of variable importance through impurity reduction (49, 50). SVM also showed comparable prediction performance, but also limitation of the method is clear that it does not yields any results regarding importance of each feature for prediction. The important variables identified by Sparse logistic Regression and RF could be useful as a selection tool for mental health professionals to identify employees at risk of depression.

We found that female employees were more likely to suffer from depression than male employees. This result in consistent with

previous studies which reported higher levels of depression among female employees (51–53). Although the mechanisms that underlie this gender difference remain unclear, one possible explanation is related to sex-specific factors. A reduction in estrogen levels may contribute to an increased risk of depression among women (54). Moreover, female employees in East Asia face heavier domestic workloads, including housework and childcare than males, which may contribute to their depression (2, 55). These results highlight the need to identify the mechanisms underlying depression among female employees and develop tailored interventions to address their needs.

Our finding indicates that employees who sleep less than six hours per night were at a higher risk of depression than those who sleep for more than six hours. This is consistent with previous studies which reported that short sleep was associated with a performance of works and depression (56, 57). Our result suggests that adequate sleep is crucial for preventing depression among employees. Additionally, smokers were more likely to suffer from depression than non-smokers in this study. This result was consistent with previous studies that depression was associated with current smoking (58, 59). Chronic nicotine exposure can affect neurotransmitters such as dopamine and 5-HT, leading to depression (60, 61). To successfully prevent the risk of young employees' depression, smoking cessation and preventive interventions must be developed.

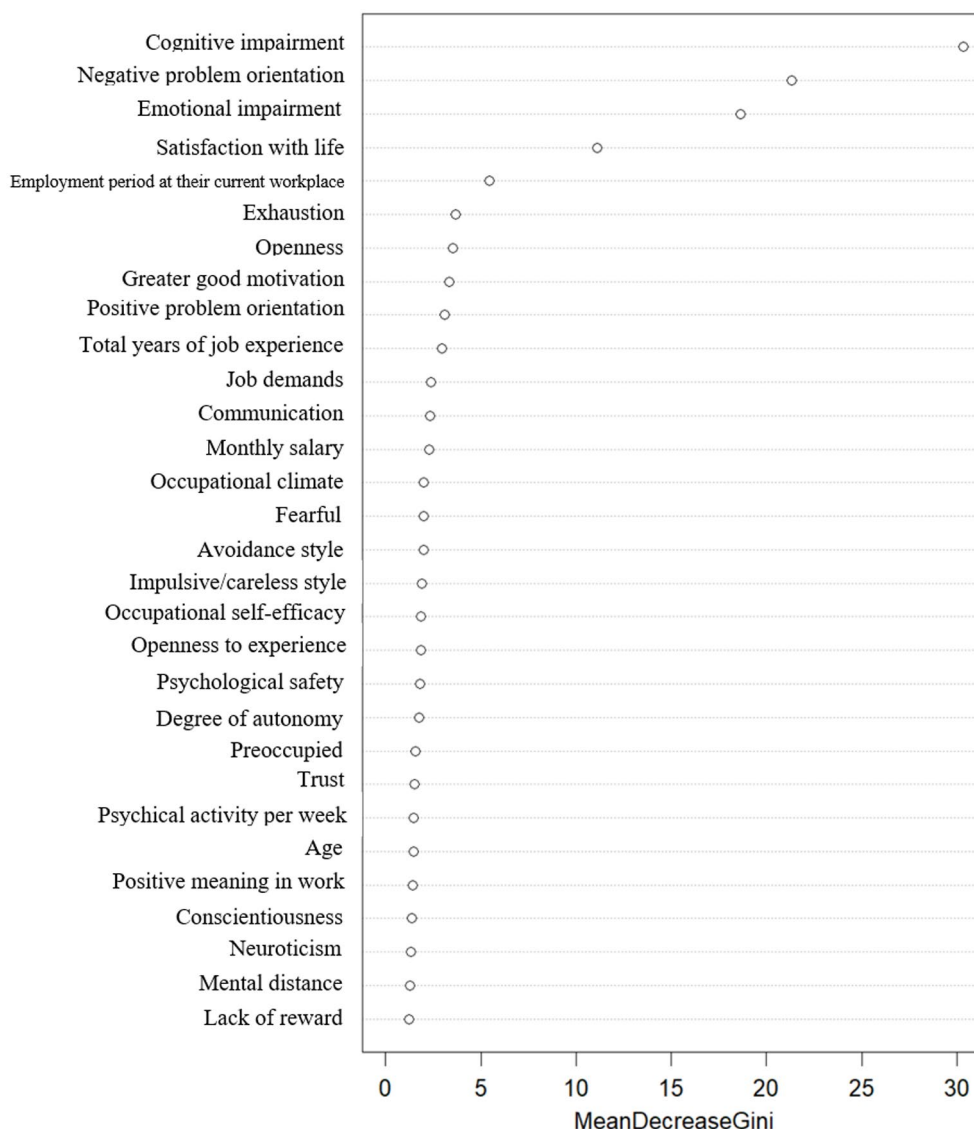


FIGURE 1
Random forest analysis results.

In our study, occupational stress and burnout were identified as psychosocial risk factors in the workplace pivotal in predicting depression risk in young generation employees. These findings were in similar with the results from prior studies, which reported that work-related stress was associated with higher depression in employees (4, 62). A logistic regression analysis also showed that higher occupational stress from highly demanding jobs were more likely to lead to depression. However, employees whose years of job experience were longer and weekly work hours were between 40 to 52 h showed a lower risk of depression. A potential explanation is that employees who face high job demands such as high workloads, time pressure, and long working hours may tend to develop the risk of depression by feeling hopelessness and powerlessness at work when they have little or no control over their work (63–66).

Moreover, higher burnout from emotional impairment was a predictor in both Logistic Regression and RF approach. This

supports the results of previous studies showing that individuals with major depressive disorder struggle with regulating emotions due to a dysfunction of their emotional brain systems (67, 68). This result can be explained that employees who experience difficulty in understanding, recognizing, and controlling emotions may undergo decreased contextual information and memory processing that subsequently impairs the autonomic nervous system and brain structure, which results in depression (67, 69, 70). Consistent with this notion, an emotional regulation program such as mindfulness and Emotional Focused Therapy (EFT) is needed to prevent young employees' depression.

We discovered a social problem-solving style and meaning in work as significant psychosocial protective factors in the workplace for predicting employees' depression. According to earlier studies, the deterioration of social problem-solving was associated with depression (71). Negative interpretations of the

TABLE 4 Logistic regression analysis results.

	Estimate	Std error	Exp coef	p-Value
Gender				
Male	–	–	–	–
Female	1.216	0.544	3.374	0.025
Marital status				
Single	–	–	–	–
Married	–0.836	0.579	0.434	0.149
Sleep hours				
Less than 6 h.	–	–	–	–
More than 6 h.	–1.069	0.536	0.343	0.046
Smoking status				
Non-smoker	–	–	–	–
Currently smoker	4.530	1.263	92.776	0.000
Total years of job experience	–0.227	0.081	0.797	0.005
Weekly working hours				
Less than 40h	–	–	–	–
40 h	–1.578	0.850	0.206	0.063
From 40 to 52 h	–2.379	1.012	0.093	0.019
More than 52h	0.690	1.701	1.994	0.685
Income satisfaction	–0.664	0.697	0.515	0.341
Personality				
Conscientiousness	–0.042	0.423	0.959	0.921
Openness to experience	0.376	0.329	1.457	0.253
Grit				
Consistency of interest	–0.624	0.439	0.536	0.155
Attachment				
Fearful	0.418	0.180	1.520	0.020
Preoccupied	0.192	0.197	1.211	0.330
Satisfaction with life	–0.115	0.062	0.891	0.065
Interpersonal relationships				
Openness	–0.211	0.098	0.809	0.031
Burnout				
Exhaustion	0.294	0.359	1.342	0.413
Emotional impairment	1.493	0.486	4.451	0.002
Cognitive impairment	0.368	0.418	1.446	0.378
Work and meaning				
Greater good motivation	–0.256	0.141	0.774	0.071
Occupational stress				
Job demands	0.034	0.017	1.035	0.042
Social Problem-Solving style				
Negative problem orientation	0.240	0.087	1.271	0.006
Work-Life Balance				
Work-Family Balance	–0.026	0.227	0.975	0.909
Work-Growth Balance	–0.296	0.222	0.744	0.182

work environment and events tend to be likely to increase depressive rumination as individuals recall more negative past experiences. In contrast, a greater sense of meaning in work predicted lower depression and higher psychological well-being. Employees who have a desire to help others and contribute to society experienced fewer symptoms of depression (72, 73). To reduce depression, intervention can help MZ employees view their work as meaningful, understand challenges, and develop abilities to deal with stress and difficulties at work.

Three psychosocial protective factors, including attachment, interpersonal relationships, and satisfaction in life, were identified in predicting depression in young employees. Young employees with fearful attachments were more likely to become depressed, while young employees with interpersonal openness were less likely to become depressed. Employees with fearful attachment likely have a negative view of themselves and others causing social isolation and loneliness, but those who are confident in self-expression and have higher self-esteem feel socially connected to others which results in preventing depression (74, 75).

The current study has some limitations. First, cross-sectional data restricts the interpretation of causal relationships. Secondly, the sample was limited to young Korean young employees, making it difficult to generalize to older employees, those at different career stages, and individuals from other cultural backgrounds. Finally, the RF approach identified several important predictive factors, but the direction of effect is unclear.

Conclusion

With advances in data science technology, this study demonstrated the practical applicability of ML algorithms in predicting the risk of depression among MZ employees. We applied three different ML algorithms – sparse logistic regression, RF, and SVM. We found the highest accuracy of RF. Our study identified the important variables influencing the risk of depression among Korean employees such as gender, inadequate sleep, smoking habits, occupational stress, burnout, social problem-solving styles, sense of meaning at work, attachment, interpersonal relationships, and satisfaction in life. These findings contribute to the development of intelligent mental healthcare systems for the early detection of depression. Additionally, our study can help develop target interventions designed to prevent employees' depression and provide a situation-specific theory that predict depression among MZ employees. However, this study focuses solely on MZ employees, and thus, careful consideration is recommended before generalizing these findings to other demographic groups.

References

1. Chow W, Doane MJ, Sheehan J, Alphas L, Le H. Economic burden among patients with major depressive disorder: an analysis of healthcare resource use, work productivity, and direct and indirect costs by depression severity. *Am J Manag Care*. (2019):e188–96.
2. Lee S, Song HY. Analysis of the convergence factors affecting depression of married working women: the 7th Korean longitudinal survey of women and families. *J Korea Converg Soc*. (2021) 12:475–85. doi: 10.15207/JKCS.2021.12.11.475
3. Eßl-Maurer R, Flamm M, Hösl K, Osterbrink J, van der Zee-Neuen A. Absenteeism and associated labour costs according to depressive symptom severity in the German

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ewha Womans University Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

S-SK, MG, and EM performed the study and manuscript conceptualization and contributed to methods, results, and discussion. S-SK and MG contributed to the background. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. NRF 2022R1A2C2004867, 2022R1A6A3A01086554, and 2021R1F1A1058613).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

general population: why preventive strategies matter. *Int Arch Occup Environ Health*. (2022) 95:409–18. doi: 10.1007/s00420-021-01788-7

4. Law PCF, Too LS, Butterworth P, Witt K, Reavley N, Milner AJ. A systematic review on the effect of work-related stressors on mental health of young workers. *Int Arch Occup Environ Health*. (2020) 93:611–22. doi: 10.1007/s00420-020-01516-7

5. Lim HN. (2021). Available at: <https://url.kr/xv9q2s> (accessed March 12, 2023)

6. Evans-Lacko S, Knapp M. Global patterns of workplace productivity for people with depression: absenteeism and presenteeism costs across eight diverse countries.

- Soc Psychiatry Psychiatr Epidemiol. (2016) 51:1525–37. doi: 10.1007/s00127-016-1278-4
7. Shin D, Lee KJ, Adeluwa T, Hur J. Machine learning-based predictive modeling of postpartum depression. *J Clin Med*. (2020) 9:2899. doi: 10.3390/jcm9092899
8. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. (2018) 15:233–4. doi: 10.1038/nmeth.4642
9. Grimmer J, Roberts ME, Stewart BM. Machine learning for social science: an agnostic approach. Annual review of political science. *Annu Rev Pol Sci*. (2021) 24:395–419. doi: 10.1146/annurev-polisci-053119-015921
10. Stewart M. The actual difference between statistics and machine learning. Towards data science. (2019). Available at: <https://medium.com/towards-data-science/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3> (accessed March 29, 2023).
11. Jin W, Miao J, Zhan Y. Be called and be healthier: how does calling influence employees' anxiety and depression in the workplace. *Int J Ment Health Promot*. (2022) 24:1–12. doi: 10.32604/ijmh.2022.018624
12. Kranabetter C, Niessen C. Appreciation and depressive symptoms: the moderating role of need satisfaction. *J Occup Health Psychol*. (2019) 24:629–40. doi: 10.1037/ocp0000153
13. Bakker AB, de Vries JD. Job demands–resources theory and self-regulation: new explanations and remedies for job burnout. *Anxiety Stress Coping*. (2021) 34:1–21. doi: 10.1080/10615806.2020.1797695
14. Hammen C. Risk factors for depression: an autobiographical review. *Annu Rev Clin Psychol*. (2018) 14:1–28. doi: 10.1146/annurev-clinpsy-050817-084811
15. Ramón-Arбуés E, Gea-Caballero V, Granada-López JM, Juárez-Vela R, Pellicer-García B, Antón-Solanas I. The prevalence of depression, anxiety and stress and their associated factors in college students. *Int J Environ Res Public Health*. (2020) 17:7001. doi: 10.3390/ijerph17197001
16. Golonka K, Mojsa-Kaja J, Blukacz M, Gawłowska M, Marek T. Occupational burnout and its overlapping effect with depression and anxiety. *Int J Occup Med Environ Health*. (2019) 32:229–44. doi: 10.13075/ijom.1896.01323
17. Li Z, Dai J, Wu N, Jia Y, Gao J, Fu H. Effect of long working hours on depression and mental well-being among employees in Shanghai: the role of having leisure hobbies. *Int J Environ Res Public Health*. (2019) 16:4980. doi: 10.3390/ijerph16244980
18. Ray S. A quick review of machine learning algorithms. In: *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)*. Manhattan: IEEE (2019) 35. doi: 10.1109/COMITCon.2019.8862451
19. Ryo M, Rillig MC. Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere*. (2017) 8:e01976. doi: 10.3390/polym14153065
20. Li JJ, Tong X. Statistical hypothesis testing versus machine learning binary classification: distinctions and guidelines. *Patterns*. (2020) 1:100115. doi: 10.1016/j.patter.2020.100115
21. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. (2016) 3:243–50. doi: 10.1016/S2215-0366(15)00471-X
22. Handing EP, Strobl C, Jiao Y, Feliciano L, Aichele S. Predictors of depression among middle-aged and older men and women in Europe: a machine learning approach. *Lancet Reg Health Eur*. (2022) 18:100391. doi: 10.1016/j.lanpe.2022.100391
23. Lee Y, Ragugett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, et al. Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *J Affect Disord*. (2018) 241:519–32. doi: 10.1016/j.jad.2018.08.073
24. Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S, Tiffin PA. Predicting persistent depressive symptoms in older adults: a machine learning approach to personalized mental healthcare. *J Affect Disord*. (2019) 246:857–60. doi: 10.1016/j.jad.2018.12.095
25. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med*. (2019) 49:1426–48. doi: 10.1017/S0033291719000151
26. Haque UM, Kabir E, Khanam R. Detection of child depression using machine learning methods. *PLoS One*. (2021) 16:e0261131. doi: 10.1371/journal.pone.0261131
27. Gil M, Kim S-S, Min EJ. Machine learning models for predicting risk of depression in Korean college students: identifying family and individual factors. *Front Public Health*. (2022) 10:1023010. doi: 10.3389/fpubh.2022.1023010
28. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. (1977) 1:385–401. doi: 10.1177/014662167700100306
29. Cho MJ, Kim KH. Diagnostic validity of the CES-D (Korea version) in the assessment of DSM-III-R major depression. *J Korean Neuropsychiatric Assoc*. (1993) 32:381–99.
30. Hinkin TR, Tracey JB, Enz CA. Scale construction: developing reliable and valid measurement instruments. *J Hosp Tour Res*. (1997) 21:100–20. doi: 10.1177/109634809702100108
31. Vaske JJ, Beaman J, Sponarski CC. Rethinking internal consistency in cronbach's alpha. *Leis Sci*. (2017) 39:163–73. doi: 10.1080/01490400.2015.1127189
32. Gosling SD, Rentfrow PJ, Swann WB Jr. A very brief measure of the big-five personality domains. *J Res Pers*. (2003) 37:504–28. doi: 10.1016/S0092-6566(03)00046-1
33. Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German. *J Res Pers*. (2007) 41:203–12. doi: 10.1016/j.jrp.2006.02.001
34. Bartholomew K, Horowitz LM. Attachment styles among young adults: a test of a four-category model. *J Pers Soc Psychol*. (1991) 61:226–44. doi: 10.1037/0022-3514.61.2.226
35. Duckworth AL, Quinn PD. Development and validation of the short grit scale (GRIT-S). *J Pers Assess*. (2009) 91:166–74. doi: 10.1080/00223890802634290
36. Diener ED, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess*. (1985) 49:71–5. doi: 10.1207/s15327752jpa4901_13
37. Schlein A, Guernsey BG, Stover L. *Relationship enhancement*. San Francisco: Josey-Bass (1971).
38. Schaufeli WB, Desart S, De Witte H. Burnout assessment tool (BAT)—development, validity, and reliability. *Int J Environ Health Res*. (2020) 17:9495. doi: 10.3390/ijerph17249495
39. Chang SJ, Koh SB, Kang DM, Kim SA, Kang MG, Lee CG, et al. Developing an occupational stress scale for Korean employees. *Korean J Occup Environ Med*. (2005) 17:297–317. doi: 10.35371/kjoem.2005.17.4.297
40. Schyns B, Von Collani G. A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *Eur J Work Organ Psychol*. (2002) 11:219–41. doi: 10.1080/13594320244000148
41. D'zurilla TJ, Nezu AM. Development and preliminary evaluation of the social problem-solving inventory. *Psychol Assess*. (1990) 2:156–63. doi: 10.1037/1040-3590.2.2.156
42. Steger MF, Dik BJ, Duffy RD. Measuring meaningful work: the work and meaning inventory (WAMI). *J Career Assess*. (2012) 20:322–37. doi: 10.1177/1069072711436160
43. Kim JW, Park JG. A study on the development of the 'Work-life Balance' scale. *J Leisure Stud*. (2008) 5:53–69. doi: 10.22879/slos.2008.5.3.53
44. Edmondson A. Psychological safety and learning behavior in work teams. *Adm Sci Q*. (1999) 44:350–83. doi: 10.2307/26669
45. Chauhan VK, Dahiya K, Sharma A. Problem formulations and solvers in linear SVM: a review. *Artif Intell Rev*. (2019) 52:803–55. doi: 10.1007/s10462-018-9614-6
46. Kawanishi Y, Shin-Gata utsu (a new type of depression) in Japan: social and clinical implications of the emerging phenomenon. *Int J Ment Health*. (2017) 46:52–64. doi: 10.1080/00207411.2016.1264041
47. Maxwell SE. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol Methods*. (2004) 9:147–63. doi: 10.1037/1082-989X.9.2.147
48. Vabalas A, Gowen E, Poliakkoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One*. (2019) 14:e0224365. doi: 10.1371/journal.pone.0224365
49. Lantz B. *Machine learning with R: Expert techniques for predictive modeling*. Birmingham: Packt Publishing (2019).
50. Kirašich K, Smith T, Sadler B. Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Sci Rev* (2018) 1: 9. Available at: <https://scholar.smu.edu/datasciencereview/vol1/iss3/9> (accessed April 5, 2023)
51. Kim HR, Kim SM, Hong JS, Han DH, Yoo SK, Min KJ, et al. Character strengths as protective factors against depression and suicidality among male and female employees. *BMC Public Health*. (2018) 18:1084–11. doi: 10.1186/s12889-018-5997-1
52. Labonté B, Engmann O, Purushothaman I, Menard C, Wang J, Tan C, et al. Sex-specific transcriptional signatures in human depression. *Nat Med*. (2017) 23:1102–11. doi: 10.1038/nm.4386
53. Plaisier I, de Bruijn JG, de Graaf R, ten Have M, Beekman AT, Penninx BW. The contribution of working conditions and social support to the onset of depressive and anxiety disorders among male and female employees. *Soc Sci Med*. (2007) 64:401–10. doi: 10.1016/j.socscimed.2006.09.008
54. Albert PR. Why is depression more prevalent in women? *J Psychiatry Neurosci*. (2015) 40:219–21. doi: 10.1503/jpn.150205
55. Koura U, Sekine M, Yamada M, Tatsuse T. Work, family, and personal characteristics explain occupational and gender differences in work–family conflict among Japanese civil servants. *Public Health*. (2017) 153:78–90. doi: 10.1016/j.puhe.2017.08.010
56. Barnes CM, Watson NF. Why healthy sleep is good for business. *Sleep Med Rev*. (2019) 47:112–8. doi: 10.1016/j.smrv.2019.07.005
57. Watson NF, Badr MS, Belenky G, Bliwise D, Buxton OM, Buysse D, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of sleep medicine and Sleep Research Society. *J Clin Sleep Med*. (2015) 38:843–4. doi: 10.5665/sleep.4716
58. Park JN, Han MA, Park J, Ryu SY. Prevalence of depressive symptoms and related factors in Korean employees: the third Korean working conditions survey (2011). *Int J Environ Health Res*. (2016) 13:424. doi: 10.3390/ijerph13040424

59. Stubbs B, Vancampfort D, Firth J, Solmi M, Siddiqi N, Smith L, et al. Association between depression and smoking: a global perspective from 48 low-and middle-income countries. *J Psychiatr Res.* (2018) 103:142–9. doi: 10.1016/j.jpsychires.2018.05.018
60. Bertrand D. The possible contribution of neuronal nicotinic acetylcholine receptors in depression. *Dialogues Clin Neurosci.* (2022) 7:207–16. doi: 10.31887/DCNS.2005.7.3/dbertrand
61. Dongelmans M, Durand-de Cuttoli R, Nguyen C, Come M, Duranté EK, Lemoine D, et al. Chronic nicotine increases midbrain dopamine neuron activity and biases individual strategies towards reduced exploration in mice. *Nat Commun.* (2021) 12:6945. doi: 10.1038/s41467-021-27268-7
62. Pereira H, Feher G, Tibold A, Costa V, Monteiro S, Esgalhado G. Mediating effect of burnout on the association between work-related quality of life and mental health symptoms. *Brain Sci.* (2021) 11:813. doi: 10.3390/brainsci11060813
63. Almroth M, Hemmingsson T, Kjellberg K, Sörberg Wallin A, Andersson T, van der Westhuizen A, et al. Job control, job demands and job strain and suicidal behaviour among three million workers in Sweden. *Occup Environ Med.* (2022) 79:681–9. doi: 10.1136/oemed-2022-108268
64. Almroth M, Hemmingsson T, Wallin AS, Kjellberg K, Burström B, Falkstedt D. Psychosocial working conditions and the risk of diagnosed depression: a Swedish register-based study. *Psychol Med.* (2022) 52:3730–8. doi: 10.1017/S003329172100060X
65. Bonde JPE. Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence. *Occup Environ Med.* (2008) 65:438–45.
66. Theorell T, Hammarström A, Aronsson G, Träskman Bendz L, Grape T, Hogstedt C, et al. A systematic review including meta-analysis of work environment and depressive symptoms. *BMC Public Health.* (2015) 15:738–14. doi: 10.1186/s12889-015-1954-4
67. Park C, Rosenblat JD, Lee Y, Pan Z, Cao B, Iacobucci M, et al. The neural systems of emotion regulation and abnormalities in major depressive disorder. *Behav Brain Res.* (2019) 367:181–8. doi: 10.1016/j.bbr.2019.04.002
68. Visted E, Vøllestad J, Nielsen MB, Schanche E. Emotion regulation in current and remitted depression: a systematic review and meta-analysis. *Front Psychol.* (2018) 9:756. doi: 10.3389/fpsyg.2018.00756
69. Bayes A, Tavella G, Parker G. The biology of burnout: causes and consequences. *World J Biol Psychiatry.* (2021) 22:686–98. doi: 10.1080/15622975.2021.1907713
70. LeMoult J, Gotlib IH. Depression: a cognitive perspective. *Clin Psychol Rev.* (2019) 69:51–66. doi: 10.1016/j.cpr.2018.06.008
71. Hasegawa A, Yoshida T, Hattori Y, Nishimura H, Morimoto H, Tanno Y. Depressive rumination and social problem solving in Japanese university students. *J Cogn Psychother.* (2015) 29:134–52. doi: 10.1891/0889-8391.29.2.134
72. Allan BA, Dexter C, Kinsey R, Parker S. Meaningful work and mental health: job satisfaction as a moderator. *J Ment Health.* (2018) 27:38–44. doi: 10.1080/09638237.2016.1244718
73. Wallace JE. Meaningful work and well-being: a study of the positive side of veterinary work. *Vet Rec.* (2019) 185:571. doi: 10.1136/vr.105146
74. Chinvararak C, Kirdchok P, Lueboonthavatchai P. The association between attachment pattern and depression severity in Thai depressed patients. *PLoS One.* (2021) 16:e0255995. doi: 10.1371/journal.pone.0255995
75. Gittins CB, Hunt C. Self-criticism and self-esteem in early adolescence: do they predict depression? *PLoS One.* (2020) 15:e0244182. doi: 10.1371/journal.pone.0244182