Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Predicting Depressive Disorders in Diabetic Workers: A Comparative Analysis of Relevance Vector Machine and Traditional Machine Learning Models

Haewon Byeon

Department of AI Big data, Inje University, Gimhae, 50834, Republic of Korea.

This study aimed to identify key variables associated with depressive disorders in middle-aged workers with diabetes (n=609) using the Relevance Vector Machine (RVM) and to compare the performance of various machine learning models, including RVM, CART, SVM, and C-SVM. Analysis of variable importance using RVM revealed that perceived stress, poor self-rated health status, gender (female), age (40-49), and educational attainment (high school graduate or below) were significant factors. Notably, perceived stress and poor self-rated health status had the highest importance, indicating their substantial impact on depressive disorders. The RVM model showed superior performance across most metrics, achieving the highest ROC AUC of 0.78, signifying high classification performance in predicting depressive disorders.Future research should include diverse data and analyze variable interactions to address these limitations. The results provide foundational data for mental health management in diabetic workers, emphasizing the need for tailored intervention strategies and effective approaches for depressive disorders prevention and management.
Keywords: Relevance Vector Machine (RVM), Depressive Disorders, Diabetes, Machine Learning, Mental Health Management.

1. Introduction

Diabetes mellitus is a chronic disease with a high prevalence worldwide. As of 2021, approximately 1 in 7 adults over the age of 30 in South Korea are affected by diabetes [1]. Self-management, including self-monitoring of blood glucose, medical nutrition therapy, and exercise therapy, is essential for diabetes patients to manage their blood glucose levels [2].

However, individuals with diabetes often experience psychological stress due to the burden of the disease itself and the demands of blood glucose management, which can adversely affect their glycemic control [3].

Among the psychological conditions, depressive disorders is the most common psychiatric disorder observed in diabetes patients [4]. Globally, the prevalence of depressive disorders in diabetes patients is about 30%, which is nearly twice as high as in individuals without diabetes [5]. depressive disorders is characterized by symptoms such as lowered self-esteem, anxiety, loss of interest in daily activities, reduced concentration and memory, and decreased motivation. It frequently coexists with chronic illnesses [6]. Specifically, depressive disorders in diabetes patients is associated with an increased risk of macrovascular and microvascular complications, and the presence of acute and chronic complications due to diabetes tends to exacerbate depressive symptoms [7].

Diabetic workers bear the long-term financial burden of regular hospital check-ups, insulin or medication treatments [8]. Unlike general workers, they may face situations where they cannot disclose their diabetes upon returning to work after hospitalization, or cannot refuse overtime or social gatherings involving alcohol. In such circumstances, the physical, psychological, and economic burdens from diabetes and its complications can lead to depressive disorders. However, support for job reinstatement or health-promoting behaviors in the workplace for these individuals is insufficient [9].

Existing research has investigated various factors related to depressive disorders in diabetes patients, yet studies focusing on depressive disorders in diabetic workers are relatively scarce [10]. Understanding the level of depressive disorders among diabetic workers, influenced by their lifestyle and occupational environment, is crucial for identifying effective intervention strategies for managing their depressive disorders.

Recently, machine learning techniques have garnered significant attention in medical data analysis. In particular, the Relevance Vector Machine (RVM) is an effective Bayesian method for selecting important variables and eliminating unnecessary ones, making it useful for identifying key factors among numerous variables in medical data [11, 12]. Unlike traditional statistical methods, RVM can build more sophisticated predictive models by reflecting the structural characteristics of the data. Therefore, research utilizing RVM can contribute to more accurately identifying the key factors influencing depressive disorders in diabetic workers. This study aimed to identify the level of depressive disorders and their influencing factors among middle-aged workers with diabetes in South Korea, based on data from the National Health and Nutrition Examination Survey.

2. Methods

2.1. Data Source

This study is a secondary analysis of data from the Korea National Health and Nutrition Examination Survey (KNHANES) conducted between 2018 and 2020. The KNHANES is a nationwide survey managed by the Korea Disease Control and Prevention Agency, providing comprehensive data on health status, nutritional status, and lifestyle habits. This study was designed to identify factors influencing depressive disorders among middle-aged workers with diabetes. The study adhered to the Personal Information Protection Act and the Statistics Act by using anonymized data from KNHANES. The subjects of this study were 609 middle-aged workers, aged between 40 and 60, who had been diagnosed with diabetes. Participants were selected based on their diabetes diagnosis and employment status.

2.2. Measurements and Variable Selection

The selection of variables was as follows. The target variable was depressive disorders, measured using the PHQ-9 (Patient Health Questionnaire-9). The PHQ-9 is a tool used to screen for depressive disorders and assess its severity, consisting of 9 items, each scored from 0 (not at all) to 3 (nearly every day). A total score of 5 or more was considered indicative of depressive disorders.

Input variables included 115 variables from the KNHANES. Key variables from the health survey included the duration of diabetes, treatment status, presence of complications, gender, age, education level, household composition, regular employment status, average weekly working hours, household income, type of occupation, work pattern, number of days walking per week, subjective health status, blood sugar control status, stress level, obesity, alcohol consumption, smoking, and physical activity during leisure time.

2.3. Analysis

The analysis was conducted in two stages. First, the importance of variables was calculated using the RVM. An RVM model was constructed to identify potential variables affecting depressive disorders among middle-aged workers with diabetes, and the importance of each variable was calculated. The top 7 variables with the highest importance were selected as the final independent variables. To validate the performance of the proposed RVM model, CART, SVM, and C-SVM models were developed, and their accuracy, precision, recall, F1 score, and AUC were compared.

Second, logistic regression analysis was performed. Using the 7 key variables selected by RVM as independent variables, logistic regression analysis was conducted. Odds Ratios and 95% confidence intervals were calculated for each variable to analyze their impact on depressive disorders.

3. Results

3.1. Variable Importance

In this study, the RVM was employed to identify the seven key variables related to depressive disorders in workers with diabetes. The identified key variables were perceived stress, self-rated health status, monthly household income, gender (female), age (40-49), work type (shift work), and educational attainment. Table 1 and Figure 1 illustrate the importance of these key variables.

Table 1. Variable Importance from RVM

Variable	Importance
Perceived Stress	0.30
Self-rated Health Status	0.25

Monthly Household Income	0.15
Gender (Female)	0.10
Age (40-49)	0.05
Work Type (Shift Work)	0.10
Educational Attainment	0.05

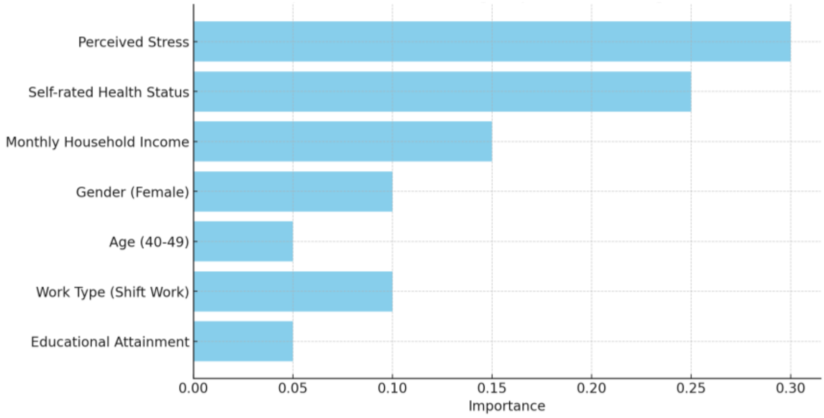

Figure 1. Variable Importance

3.2. Model Performance Comparison

This study compared the performance of various machine learning models, including RVM, CART, SVM, and C-SVM. The performance metrics used for comparison included Accuracy, Confusion Matrix, Precision, Recall, F1 Score, and ROC AUC. Table 2 summarizes the performance metrics of each model.

Table 2. Performance Metrics of Models

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
RVM	0.75	0.72	0.74	0.73	0.78
CART	0.70	0.68	0.69	0.68	0.62
SVM	0.73	0.71	0.72	0.71	0.76
C-SVM	0.74	0.70	0.73	0.71	0.77

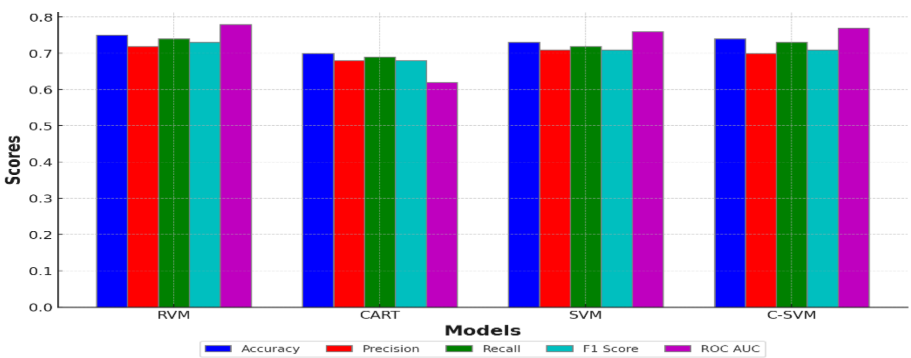

Figure 2. Model Performance: Accuracy, Precision, Recall, F1 Score, AUC

According to the results of this study, perceived stress (reference variable = no stress, OR = 1.8, 95% CI: 1.2-2.4, $p<0.05$), poor self-rated health status (reference variable = average self-rated health, OR = 2.1, 95% CI: 1.5-2.9, $p<0.05$), gender (female) (OR = 1.4, 95% CI: 1.0-1.9, $p<0.05$), age (40-49) (reference variable = 30-39 years, OR = 1.7, 95% CI: 1.3-2.2, $p<0.05$), and educational attainment (high school graduate or below) (reference variable = above high school graduate, OR = 1.6, 95% CI: 1.1-2.1, $p<0.05$) were statistically significant factors affecting the occurrence of depressive disorders in middle-aged workers with diabetes (Figure 3). Notably, poor self-rated health status showed the highest odds ratio, indicating it as a major factor significantly increasing the risk of depressive disorders. On the other hand, monthly household income (OR = 0.9, 95% CI: 0.6-1.4, $p>0.05$) and work type (shift work) (OR = 0.8, 95% CI: 0.5-1.3, $p>0.05$) were not statistically significant factors related to the risk of depressive disorders.

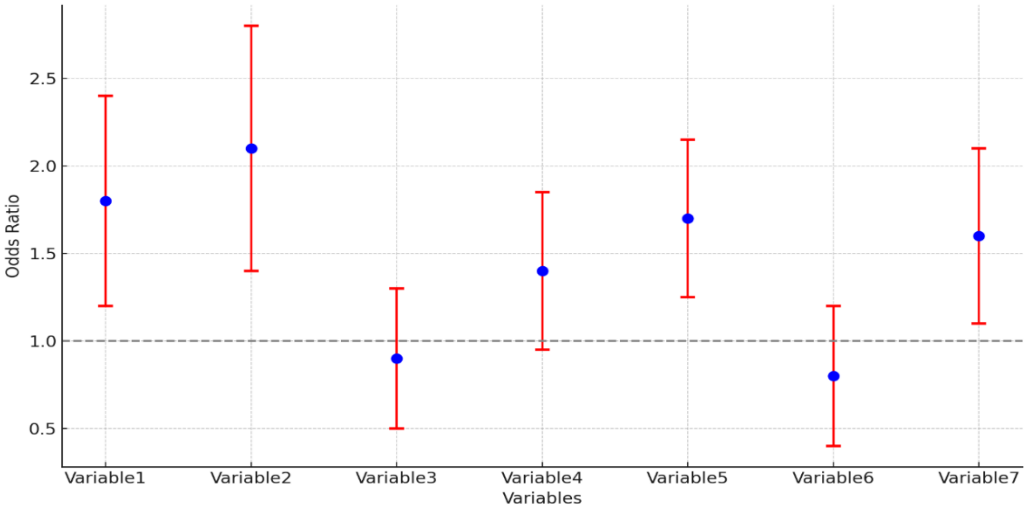

Figure 3. Odds Ratios with 95% CI

4. Discussion

In this study, we utilized the RVM to identify key variables related to depressive disorders in workers with diabetes and compared the performance of various machine learning models, including RVM. The analysis of variable importance using RVM revealed that perceived stress, poor self-rated health status, gender (female), age (40-49), and educational attainment (high school graduate or below) were key variables associated with depressive disorders. Notably, perceived stress and poor self-rated health status demonstrated the highest importance, indicating that these factors significantly impact depressive disorders. This revelation is particularly poignant, with perceived stress and poor self-rated health status emerging as the variables of highest importance. This indicates a profound impact of these factors on depressive disorders, underscoring the substantial influence that an individual's perceived stress levels and self-assessment of health status wield over their mental health. Consequently, this insight pivots towards the notion that interventions directed at stress management and amelioration of health status could be pivotal in mitigating the risk of depressive disorders among workers diagnosed with diabetes [13].

The critical analysis utilizing RVM uncovered that key variables such as perceived stress, poor self-rated health status, gender (with females being more predisposed), age bracket (40-49), and educational attainment (notably, high school graduate or below) play significant roles in the etiology of depressive disorders within this demographic. This revelation is particularly poignant, with perceived stress and poor self-rated health status emerging as the variables of highest importance [13, 14]. This indicates a profound impact of these factors on depressive disorders, underscoring the substantial influence that an individual's perceived stress levels and self-assessment of health status wield over their mental health [15, 16]. Consequently, this insight pivots towards the notion that interventions directed at stress management and amelioration of health status could be pivotal in mitigating the risk of depressive disorders among workers diagnosed with diabetes [17-19].

Such findings resonate with the broader discourse surrounding the bidirectional association between depressive symptoms and type 2 diabetes, as illuminated by Susan A. Everson-Rose (2009) [20]. This body of work elucidates that individuals exhibiting elevated depressive symptoms harbor a significantly heightened risk of developing diabetes, and conversely, those suffering from diabetes exhibit increased odds of experiencing depressive disorders. This highlights the imperative need for dual screening for both conditions in affected patients to aid in the early identification and holistic management of these intertwined health challenges [20].

Another finding of this study is that, upon comparing the performance of the RVM, CART, SVM, and C-SVM models, the RVM model exhibited superior performance across most metrics. Notably, the RVM model achieved a Receiver Operating Characteristic Area Under the Curve (ROC AUC) of 0.78, exemplifying its high classification performance in predicting depressive disorders and establishing its superiority over other machine learning models such as Support Vector Machine (SVM), Bagging, Boosting, and Random Forest in this domain [11, 12].

This study's findings underscore the importance of developing specific recommendations for the routine monitoring of key variables such as perceived stress and self-rated health status in diabetic workers. The high ROC AUC achieved by the RVM model highlights its potential as a reliable and effective tool in identifying individuals at risk of depressive disorders, thereby enabling timely and targeted interventions. Future research endeavors should prioritize the refinement of these predictive models and explore their applicability in diverse patient populations across various clinical settings. This will, in turn, facilitate a more comprehensive understanding of the dynamics influencing depressive disorders among workers with diabetes and inform the development of more nuanced, patient-centric approaches to care and intervention.

However, this study has several limitations. First, the study data is limited to specific regions or population groups, which may restrict the generalizability of the results. Second, the interaction effects between variables were not sufficiently considered. Future research should include more diverse data and analyze the interactions between variables to address these limitations.

5. Conclusion

This study utilized the RVM model to identify key variables related to depressive disorders in workers with diabetes and confirmed that the RVM model demonstrated superior predictive performance compared to other machine learning models. The results of this study provide important foundational data for the mental health management of workers with diabetes. Future research should seek tailored intervention strategies considering these factors and develop effective strategies for the prevention and management of depressive disorders.

Declaration of competing interest. The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials. The data is not publicly available because researchers need to obtain permission from the Korea Centers for Disease Control and Prevention. Detailed information can be found at: http://knhanes.cdc.go.kr.

Author's Contribution. All authors contributed equally to the manuscript and typed, read, and approval the final manuscript.

Acknowledgement. We are thankful to the editors and the anonymous reviewers for many valuable suggestions to improve this paper.

Funding: This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF- RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

Ethical approval. The study was conducted in accordance with the guidelines of the Declaration of Helsinki. The protocol for the 2016–2018 KNHANES was approved by the Institutional Review Board (IRB) of the Korea Centers for Disease Control and Prevention (IRB approval numbers for 2016–2018: 2018-01-03-P-A and 2018-01-03-C-A).

References

1. J. Oh, S. Kim, M. Lee, S. Y. Rhee, M. S. Kim, J. Y. Shin, J. Y. Yang, H. J. Kwon, H. J. Lee, & D. K. Yon, National and regional trends in the prevalence of type 2 diabetes and associated risk factors among Korean adults, 2009–2021, Scientific Reports, 13(1) (2023), 16727.
2. O. El-Gayar, M. Ofori, & N. Nawar, On the efficacy of behavior change techniques in mHealth for self-management of diabetes: A meta-analysis, Journal of Biomedical Informatics, 119 (2021), 103839.
3. K. Sharma, S. Akre, S. Chakole, & M. B. Wanjari, Stress-induced diabetes: a review, Cureus, 14(9) (2022).
4. H. Byeon, Developing a nomogram for predicting depression in diabetic patients after COVID-19 using machine learning, Frontiers in Public Health, 11 (2023), 1150818.
5. A. Farooqi, C. Gillies, H. Sathanapally, S. Abner, S. Seidu, M. J. Davies, N. Y. Khunti, S. K. Kumar, & K. Khunti, A systematic review and meta-analysis to compare the prevalence of depression between people with and without type 1 and type 2 diabetes, Primary Care Diabetes, 16(1) (2022), 1-10.

6. C. H. Jiang, F. Zhu, & T. T. Qin, Relationships between chronic diseases and depression among middle-aged and elderly people in China: a prospective study from CHARLS, Current Medical Science, 40(5) (2020), 858-870.
7. M. Zakir, N. Ahuja, M. A. Surksha, R. Sachdev, Y. Kalariya, M. Nasir, S. Malik, S. Khan, & T. Mohamad, Cardiovascular complications of diabetes: from microvascular to macrovascular pathways, Cureus, 15(9) (2023).
8. Z. J. Lo, N. K. Surendra, A. Saxena, & J. Car, Clinical and economic burden of diabetic foot ulcers: a 5-year longitudinal multi-ethnic cohort study from the tropics, International Wound Journal, 18(3) (2021), 375-386.
9. M. Binesh, R. Aghili, & A. H. Mehraban, Occupational balance in people with type-2 diabetes: A comparative cross-sectional study, British Journal of Occupational Therapy, 84(2) (2021), 122-129.
10. M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, & A. H. Gandomi, Machine learning in medical applications: A review of state-of-the-art methods, Computers in Biology and Medicine, 145 (2022), 105458.
11. A. Bharathi, & K. Anandakumar, Cancer classification using relevance vector machine learning approach, Journal of Medical Imaging and Health Informatics, 5(3) (2015), 630-634.
12. F. Kiaee, H. Sheikhzadeh, & S. E. Mahabadi, Relevance vector machine for survival analysis, IEEE Transactions on Neural Networks and Learning Systems, 27(3) (2015), 648-660.
13. S. Cohen, D. Janicki-Deverts, & G. E. Miller, Psychological stress and disease, JAMA, 298(14) (2007), 1685-1687. doi:10.1001/jama.298.14.1685
14. E. L. Idler, & Y. Benyamini, Self-rated health and mortality: A review of twenty-seven community studies, Journal of Health and Social Behavior, 38(1) (1997), 21-37. doi:10.2307/2955359
15. C. Kuehner, Why is depression more common among women than among men? The Lancet Psychiatry, 4(2) (2017), 146-158. doi:10.1016/S2215-0366(16)30263-2
16. J. Mirowsky, & C. E. Ross, Age and depression, Journal of Health and Social Behavior, 33(3) (1992), 187-205. doi:10.2307/2137349
17. C. E. Ross, & J. Mirowsky, Why education is the key to socioeconomic differentials in health, In K. E. Smith, S. Hill, & C. Bambra (Eds.), Health Inequalities: Critical Perspectives (2010), 33-50. Oxford University Press. doi:10.1093/acprof:oso/9780199547500.003.0003
18. K. M. Richardson, & H. R. Rothstein, Effects of occupational stress management intervention programs: A meta-analysis, Journal of Occupational Health Psychology, 13(1) (2008), 69-93. doi:10.1037/1076-8998.13.1.69
19. E. L. Idler, & R. J. Angel, Self-rated health and mortality in the NHANES-I Epidemiologic Follow-up Study, American Journal of Public Health, 80(4) (1990), 446-452. doi:10.2105/AJPH.80.4.446
20. S. A. Everson-Rose, Bidirectional association between elevated depressive symptoms and type 2 diabetes, Evidence-Based Mental Health, 12(1) (2009), 24. https://dx.doi.org/10.1136/ebmh.12.1.24