Submitted: 16 June, 2024 Accepted: 12 September, 2024 Published: 30 June, 2025 DOI:10.22514/jomh.2025.077

JOMH
Journal of Men's Health

ORIGINAL RESEARCH

Non-cognitive factors influencing early stroke symptom recognition among Korean working-class males with hypertension and diabetes: an integrated multi-output gradient boosting and logistic regression approach

Haewon Byeon[1,2,*]

[1]Workcare Digital Health Lab, Department of Employment Service Policy, Korea University of Technology and Education, 31253 Cheonan, Republic of Korea
[2]Department of Convergence, Korea University of Technology and Education, 31253 Cheonan, Republic of Korea

***Correspondence**
bhwpuma@naver.com
(Haewon Byeon)

Abstract

This study aims to investigate the non-cognitive factors influencing the recognition of early stroke symptoms among Korean working class males with diabetes using an integrated machine learning approach combining Multi-Output Gradient Boosting and logistic regression models. Data from the Korea National Health and Nutrition Examination Survey (KNHANES) from 2016 to 2022 were utilized, including 4125 working class males with diabetes. Participants were divided into two groups based on their recognition of early stroke symptoms. The integrated machine learning model was trained on 80% of the dataset and tested on the remaining 20%. Key predictors were identified, and logistic regression analysis provided odds ratios (OR) and 95% confidence intervals (CI) for significant factors. The study found that 72% of participants recognized early stroke symptoms, while 28% did not. Significant predictors of non-recognition included younger age ($\beta = -0.05$, OR = 0.95, $p < 0.01$), higher Body Mass Index (BMI) ($\beta = 0.12$, OR = 1.13, $p < 0.01$), hypertension ($\beta = 0.28$, OR = 1.32, $p < 0.01$), elevated cholesterol ($\beta = 0.03$, OR = 1.03, $p < 0.01$) and triglycerides ($\beta = 0.04$, OR = 1.04, $p < 0.01$), depression ($\beta = 0.25$, OR = 1.28, $p < 0.01$), stress ($\beta = 0.18$, OR = 1.20, $p < 0.01$), smoking ($\beta = 0.10$, OR = 1.11, $p < 0.01$) and alcohol consumption ($\beta = 0.08$, OR = 1.08, $p < 0.01$). Positive factors included regular physical activity ($\beta = -0.20$, OR = 0.82, $p < 0.01$) and participation in diabetes education programs ($\beta = -0.15$, OR = 0.86, $p < 0.01$). The findings highlight the multifactorial nature of stroke symptom recognition and suggest that targeted interventions focusing on both physiological and psychological factors, as well as promoting healthy lifestyle behaviors, can significantly improve symptom recognition and health outcomes in working class males with diabetes.

Keywords

Early stroke symptom recognition; Diabetic workers; Non-cognitive factors; Machine learning; Multi-output gradient boosting

1. Introduction

Stroke is one of the leading causes of death among Koreans, particularly prevalent in the elderly population. According to data from the Korean Statistical Information Service (KOSIS), the top 10 causes of death in 2019 included cancer, heart disease, pneumonia, cerebrovascular disease, suicide, diabetes, Alzheimer's disease, liver disease, chronic lower respiratory disease and hypertensive diseases [1]. Among circulatory system diseases, heart disease, cerebrovascular disease, and hypertensive diseases accounted for a significant portion of mortality, with a sharp increase in death rates observed in individuals aged 70 years and above [1]. The aging population in Korea is projected to rise substantially, with individuals aged 65 years and older expected to constitute 14.9% of the population in 2019, 20.3% by 2025 and 43.9% by 2060. This demographic shift underscores the urgent need for effective prevention and management policies for these diseases. Hypertension significantly increases the risk of complications such as stroke, angina pectoris, myocardial infarction, heart failure and renal failure [2]. Among these, stroke is the most frequently occurring complication, with hypertensive patients experiencing up to seven times higher incidence rates compared to normotensive individuals [3]. Myocardial infarction and angina pectoris also occur three times more frequently in hypertensive patients than in those without hypertension. Diabetes is another critical risk factor for stroke, with diabetic patients experiencing approximately three times higher incidence rates of stroke compared to non-diabetic individuals [2, 3]. Furthermore, the risk of mortality from ischemic heart disease is two to three times higher in diabetic patients. Therefore, preventing and managing cardiovascular and cerebrovascular diseases in patients with hypertension and diabetes is of paramount importance [1–3]. Thrombolytic therapy has proven effective in treating ischemic stroke and myocardial infarction; however, delays in pre-hospital stages—such as failure to recognize early symptoms or inadequate emergency response—can significantly impact survival and prognosis. Lack of knowledge about early symptoms and insufficient emergency response are cited as major reasons for these delays. Improving the recognition of early symptoms could serve as an effective intervention strategy to reduce pre-hospital delays. Patients diagnosed with both hypertension and diabetes are considered high-risk for cardiovascular and cerebrovascular diseases. Recognizing early symptoms of stroke and myocardial infarction and promptly transporting patients to medical facilities for reperfusion therapy can minimize sequelae, disabilities and mortality rates. Therefore, it is crucial to assess the level of recognition of early symptoms of stroke and myocardial infarction among these patients and analyze factors related to non-recognition. Previous studies have identified various factors influencing the recognition of early stroke symptoms, including gender, age, education level, income level, occupation and residential area [4]. Similarly, factors influencing the recognition of myocardial infarction symptoms include age, gender and education level [5]. However, these studies predominantly focused on the general population rather than high-risk groups such as patients with hypertension and diabetes. Additionally, factors such as education and public awareness campaigns have been highlighted as significant influences on symptom recognition [6].

According to previous studies [7–11], exposure to educational materials was significantly associated with the recognition of early stroke symptoms. Previous research [12–15] has indicated that community-based stroke education and promotion programs shortened the time from symptom onset to hospital arrival, increasing the proportion of patients arriving within the golden hour. Nonetheless, much evidence is still needed on this topic. For example, the research conducted by Azizov *et al*. [16] (2020) highlighted that structured training in hypertension and diabetes management education may positively impact blood pressure indicators in patients with Type II diabetes and arterial hypertension, but it doesn't specifically address the recognition of warning symptoms related to these conditions.

Moreover, while blood pressure and blood glucose control are fundamental in managing hypertension and diabetes, few studies have evaluated the relationship between the recognition of early cardiovascular symptoms and the awareness of blood pressure and glucose levels. This study aims to address this gap by focusing on disease management-related characteristics, such as the completion of hypertension and diabetes management education and the awareness of blood pressure and blood glucose levels, to analyze factors related to the non-recognition of early stroke symptoms.

The primary objective of this study is to analyze the factors influencing the non-recognition of early stroke symptoms in Korean working class males with diabetes using an integrated machine learning approach combining Multi-Output Gradient Boosting and logistic regression models. By identifying significant non-cognitive factors, this study aims to provide foundational data for improving the recognition of early symptoms and enhancing the efficiency of education and management programs for cardiovascular and cerebrovascular diseases. By addressing the methodological limitations of previous studies and employing advanced machine learning techniques, this research seeks to offer a comprehensive analysis and robust predictive model that can inform targeted interventions and policy development for the prevention and management of stroke in high-risk populations.

2. Materials and methods

2.1 Study design and population

This study employs a descriptive cross-sectional design to analyze the non-cognitive factors influencing the recognition of early stroke symptoms among Korean working class males with diabetes. Data from the 2016–2022 Korea National Health and Nutrition Examination Survey (KNHANES) and the Community Health Survey are utilized for this purpose. The inclusion criteria for the study population include male workers aged 40 years and above, diagnosed with both hypertension and diabetes, and who provided complete responses to the survey variables included in the analysis. A total of 4125 working class males with diabetes meeting these criteria are included in the final analysis.

2.2 Study variables

The dependent variable in this study is the recognition of early stroke symptoms. The five representative stroke warning symptoms are as follows: "Sudden weakness in one side of the face, arm or leg (hemiplegia)", "Sudden slurred speech or difficulty understanding others (speech impairment)", "Sudden vision loss in one eye or half of the visual field, or seeing double (visual impairment)", "Sudden dizziness or difficulty maintaining balance (balance impairment)", and "Sudden severe headache unlike any experienced before (severe headache)". This is measured using five items from the KNHANES, with responses coded as 1 for "yes" and 0 for "no". The total score ranges from 0 to 5, with scores of 0–4 indicating non-recognition and a score of 5 indicating recognition. Independent variables include demographic factors such as age, education level, income level, marital status, occupation and residential area; physiological factors such as BMI, blood pressure and blood glucose levels; psychological factors such as depression and stress levels, measured using standardized scales; and lifestyle factors such as smoking status, alcohol consumption, physical activity, and participation in diabetes management education (Table 1).

2.3 Data preprocessing

Handling of missing values was performed using multiple imputation techniques to ensure robustness. Continuous variables such as age, BMI, blood pressure, and blood glucose levels were normalized using z-scores to standardize the data. The z-score normalization was performed using the formula:

TABLE 1. Main variables of KNHANES.

Survey Area	Survey Items
Household Survey	Gender, Age, Marital Status, Number of Household Members, Household Type, Home Ownership, Housing Type, Household Income, Health Insurance Enrollment, Private Insurance Enrollment
Education	Education Level, Graduation Status
	Parents' Education Level
Economic Activity	Employment Status, Employment Type, Current Occupation, Employment Status, Total Working Hours, Weekly Working Hours
	Permanent Employment Status, Longest Held Occupation
Morbidity	Subjective Health Status, Illness in the Last 2 Weeks, Chronic Diseases (10 for adults, 11, 13)
Medical Utilization	Unmet Medical Needs/Reasons, Outpatient Care, Inpatient Care
Health Check-up	Health Check-up Status, Cancer Screening Status
Vaccination	Influenza Vaccination Status
	Monthly Absence from Work/School/Number of Days
Activity Limitation	Activity Limitation Status/Reasons, Monthly Bedridden Experience/Number of Days, Monthly Absence from Work/School/Number of Days
	Elderly Functional Scale (LF-10)
Quality of Life	Subjective Health Perception
	Health-related Quality of Life Index (EQ-5D, EuroQol-5 Dimension: Mobility, Self-care, Usual Activities, Pain/Discomfort, Anxiety/Depression, HINT-8, Health-related Quality of Life) Instrument with 8 items: Climbing Stairs, Pain, Vitality, Work, Depression, Memory, Sleep, Happiness
Injury	Injury Experience, Number of Injuries, Time of Injury, Treatment for Injury, Bedridden Due to Injury/Number of Days/Absence from Work/School/Number of Days
Smoking	Lifetime Smoking Experience of Regular Cigarettes, Current Smoking, Age Started Smoking, Smoking Quantity
	Lifetime Smoking/Current Smoking/Past Smoking/Age Started Smoking/Daily Smoking Start Age/Smoking Quantity, Lifetime Use of Heated Tobacco Products/Current Use/Quantity/Reasons for Use, Lifetime/Monthly Use of Liquid-type E-cigarettes and Other Tobacco, Quit Smoking Duration, Quit Smoking Attempts, Quit Smoking Plans, Nicotine Dependence, Quit Smoking Methods, Indoor Secondhand Smoke Exposure at Home/Work/Public Places
Drinking	Lifetime Drinking, Age Started Drinking, Drinking Frequency, Drinking Quantity, Binge Drinking Frequency, Indirect Harm from Drinking
	Experience of Moderate Drinking Recommendation, Experience of Counseling for Drinking Problems
Physical Activity	Practice of 60 Minutes of Physical Activity a Day, Sitting Time, Strength Exercises
	International Physical Activity Questionnaire (GPAQ, Global Physical Activity Questionnaire: High/Moderate Intensity Physical Activity during Work/Leisure, Activity during Commuting, Sitting Time), Walking, Sitting Time, Strength Exercises
Sleep Health	Average Sleep Time on Weekdays/Weekends, Sleep Time and Wake Time on Weekdays/Weekends
	Obstructive Sleep Apnea Screening Tool (STOP-Bang: Snoring, Tiredness, Observed Apnea)
Mental Health	Perceived Stress, Experience of Depression, Suicidal Thoughts/Plans/Attempts, Experience of Mental Health Counseling
	Depression Screening Tool (PHQ-9, Patient Health Questionnaire-9), Perceived Sources of Stress, Generalized Anxiety Disorder Screening Tool (GAD-7)
	Use of Car Safety Equipment
Safety Awareness	Use of Front Seat in Car, Use of Bicycle Helmet
	Use of Seatbelt in Passenger Cars
	Use of Seatbelt while Driving, Use of Seatbelt in Front/Back Seat, Experience of Drunk Driving in Car/Motorcycle/Bicycle, Riding in a Vehicle with a Drunk Driver
Obesity and Weight Control	Perceived Body Shape, Effort to Control Weight, Methods of Weight Control
	Weight Change, Degree of Weight Loss, Degree of Weight Gain

TABLE 1. Continued.

Survey Area	Survey Items
Women's Health	Current Menstrual Status, Age at Menarche
	Pregnancy Experience, Childbirth Experience, Age at First/Last Childbirth
	Breastfeeding Experience/Number of Children/Duration, Age at Menopause, Causes of Artificial Menopause, Experience of Taking Oral Contraceptives, Use of Female Hormones/Age First Taken/Duration of First Use
Oral Health	Brushing Teeth, Tooth Damage, Oral Health Check-up, Dental Care, Unmet Dental Needs/Reasons
	Use of Oral Hygiene Products
	Difficulty Chewing, Difficulty Speaking

$$z = \frac{(X - \mu)}{\sigma}$$

Where (X) is the raw score, (μ) is the mean of the population and (σ) is the standard deviation of the population. Categorical variables such as education level and marital status were encoded using one-hot encoding to facilitate their inclusion in the machine learning models.

2.4 Machine learning model

The study employed an integrated machine learning framework combining Multi-Output Gradient Boosting and logistic regression models to analyze the data.

Gradient Boosting is an ensemble learning technique that builds multiple decision trees sequentially. For multi-output tasks, the algorithm is adapted to handle multiple outputs simultaneously. The model minimizes the loss function:

$$L\left(y, f\left(x\right)\right) = \sum_{i=1}^{N} l\left(y_i, f_m\left(x_i\right)\right)$$

Where (l) is the loss function, (y_i) is the observed value, ($f_m(x_i)$) is the predicted value, and (N) is the number of samples. The model iteratively adds new trees to minimize the residual errors of the previous trees:

$$f_m\left(x\right) = f_{m-1}\left(x\right) + \eta \cdot h_m\left(x\right)$$

Where (η) is the learning rate, and ($h_m(x)$) is the base learner at iteration (m). For multi-output tasks, the model optimizes the joint loss function for all outputs:

$$L\left(Y, F\left(X\right)\right) = \sum_{j=1}^{K}\sum_{i=1}^{N} l\left(y_{ij}, f_{mj}\left(x_i\right)\right)$$

Where (K) is the number of outputs.

Feature importance was derived using the feature_importances_ attribute of the GradientBoostingClassifier model from sklearn. Also, hyperparameter tuning was performed using RandomizedSearchCV.

2.5 Model training and evaluation

The dataset was divided into training (80%) and testing (20%) sets. K-fold cross-validation (k = 10) was employed to ensure model generalizability. The performance of the models was evaluated using accuracy, precision, recall and F1-Score for classification performance, and Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for regression accuracy. The formulas for MAE and RMSE are as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \left|y_i - \widehat{y_1}\right|$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(y_i - \widehat{y_1}\right)^2}$$

2.6 Statistical analysis

Descriptive statistics were used to summarize the demographic, physiological, psychological and lifestyle characteristics of the study population. Chi-Square tests and *t*-tests were conducted to identify significant differences between recognition and non-recognition groups. Multivariate logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI) for factors associated with non-recognition of early stroke symptoms. The odds ratio is calculated as:

$$OR = \frac{P(Y = 1 \mid X)}{P(Y = 0 \mid X)} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n}$$

3. Results

3.1 Descriptive statistics

The study analyzed data from 4125 working class males with diabetes extracted from the 2016–2022 KNHANES. The population was divided into two groups based on their recognition of early stroke symptoms: those who recognized the symptoms and those who did not. Of the participants, 72% (2970) recognized the symptoms, while 28% (1155) did not. Table 2 summarizes the demographic, physiological, psychological

TABLE 2. Characteristics of the sample.

Variable	Recognized (n = 2970)	Not Recognized (n = 1155)	p-value
Age (yr)	55.31 ± 8.42	52.14 ± 7.90	<0.001
BMI (kg/m²)	24.65 ± 3.11	27.18 ± 4.21	<0.001
Hypertension (%)	45.52	62.81	<0.001
Cholesterol (mg/dL)	189.23 ± 35.60	202.72 ± 40.11	<0.001
Triglycerides (mg/dL)	150.11 ± 45.43	178.31 ± 60.24	<0.001
Depression (%)	18.22	35.43	<0.001
Stress (%)	25.01	42.72	<0.001
Smoking (%)	22.73	35.90	<0.001
Alcohol Consumption (%)	60.51	70.93	<0.001
Physical Activity (%)	48.64	32.43	<0.001
Diabetes Education (%)	28.21	17.76	<0.001
Occupation: Service or sales workers (%)	58.53	41.51	<0.001
Occupation: Managers/professionals/clerks (%)	61.32	38.71	<0.001
Occupation: Agricultural/forestry/fishery workers/ mechanical or manual laborers (%)	47.81	52.24	<0.001

BMI: Body Mass Index.

and lifestyle characteristics of the study population.

3.2 Machine learning model performance

The integrated Multi-Output Gradient Boosting and logistic regression model was trained using 80% of the dataset and tested on the remaining 20%. The performance metrics for the model are summarized in Table 3.

TABLE 3. Model performance metrics.

Metric	Value
Accuracy (%)	85.41
Precision (%)	82.12
Recall (%)	88.70
F1-Score (%)	85.36
Mean Absolute Error (MAE)	0.451
Root Mean Squared Error (RMSE)	0.582

The feature importance scores from the Gradient Boosting model provide insights into which factors are most influential in predicting the recognition of early stroke symptoms. Fig. 1 shows the relative importance of the top 5 features. The confusion matrix for the logistic regression model based on the Gradient Boosting predictions is shown in Fig. 2. It illustrates the number of true positives, true negatives, false positives and false negatives. The Receiver Operating Characteristic (ROC) curve for the logistic regression model is shown in Fig. 3. The area under the ROC curve (AUC) is 0.85, indicating good model performance.

3.3 Regression analysis results

The logistic regression analysis results provide the odds ratios (OR) and 95% confidence intervals (CI) for the significant factors influencing the recognition of early stroke symptoms. The results are presented in Table 4. In this study, we identified factors that negatively impact early stroke recognition (indicated by a positive odds ratio) as well as factors that positively influence early recognition (indicated by a negative odds ratio).

As a result of the analysis of the final model of this study (Figs. 4,5,6,7), First, older age was significantly associated with better recognition of early stroke symptoms. Specifically, younger male workers tended to have poorer recognition, as indicated by the negative coefficient for age ($\beta = -0.05$, OR = 0.95, $p < 0.01$). Second, higher BMI and the presence of hypertension were significantly linked to poorer recognition of early stroke symptoms. Participants with higher BMI (β = 0.12, OR = 1.13, $p < 0.01$) and those with hypertension (β = 0.28, OR = 1.32, $p < 0.01$) were less likely to recognize early stroke symptoms. Third, elevated cholesterol and triglyceride levels were found to be significant predictors of poor recognition. Higher cholesterol (β = 0.03, OR = 1.03, $p < 0.01$) and triglyceride levels (β = 0.04, OR = 1.04, $p < 0.01$) were associated with an increased likelihood of non-recognition. Fourth, psychological factors such as depression and stress were strongly associated with higher odds of non-recognition. Participants experiencing depression (β = 0.25, OR = 1.28, $p < 0.01$) and high stress levels (β = 0.18, OR = 1.20, $p < 0.01$) were less likely to recognize early stroke symptoms. Fifth, lifestyle factors played a crucial role. Smoking (β = 0.10, OR = 1.11, $p < 0.01$) and alcohol consumption (β = 0.08, OR = 1.08, $p < 0.01$) negatively impacted the recognition of early stroke symptoms. In contrast, regular physical activity ($\beta = -0.20$, OR = 0.82, $p < 0.01$) and participation in diabetes education programs ($\beta = -0.15$, OR = 0.86, $p < 0.01$) were positively

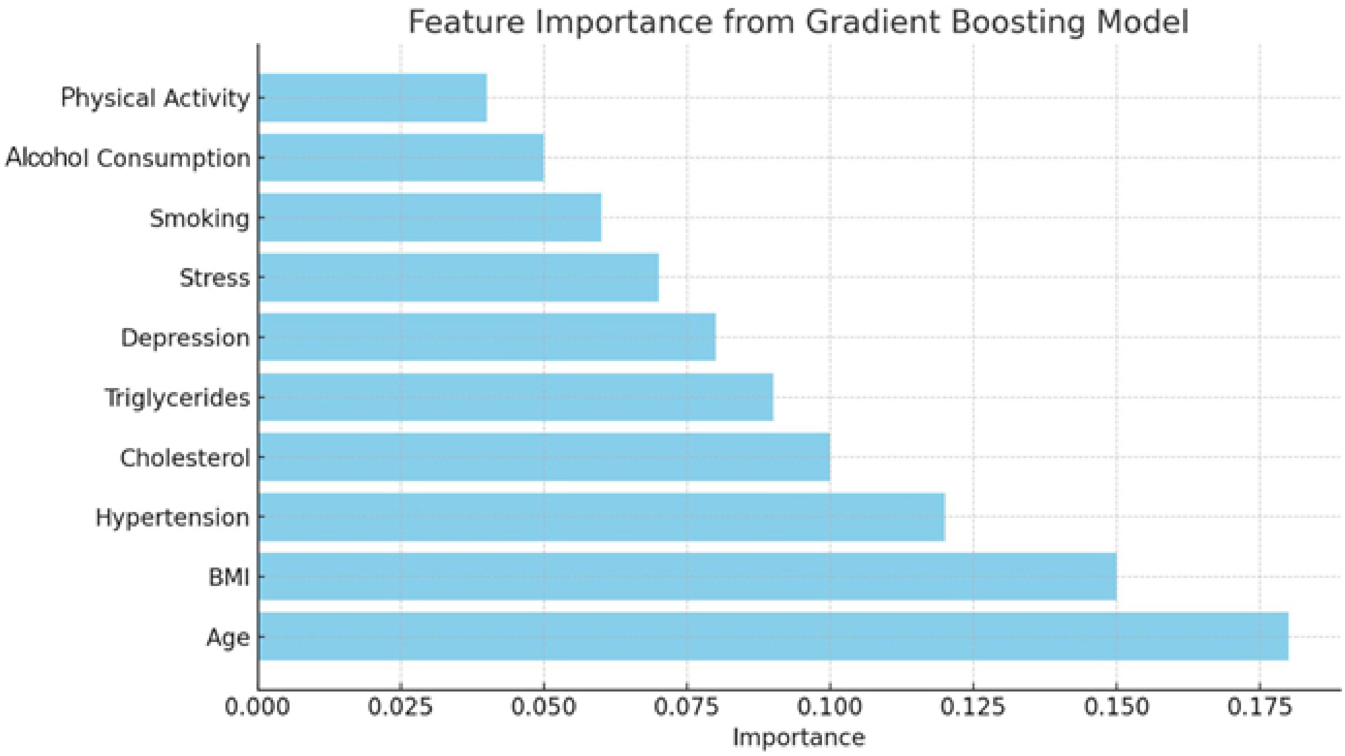

FIGURE 1. Feature importance from multi-output gradient boosting. BMI: Body Mass Index.

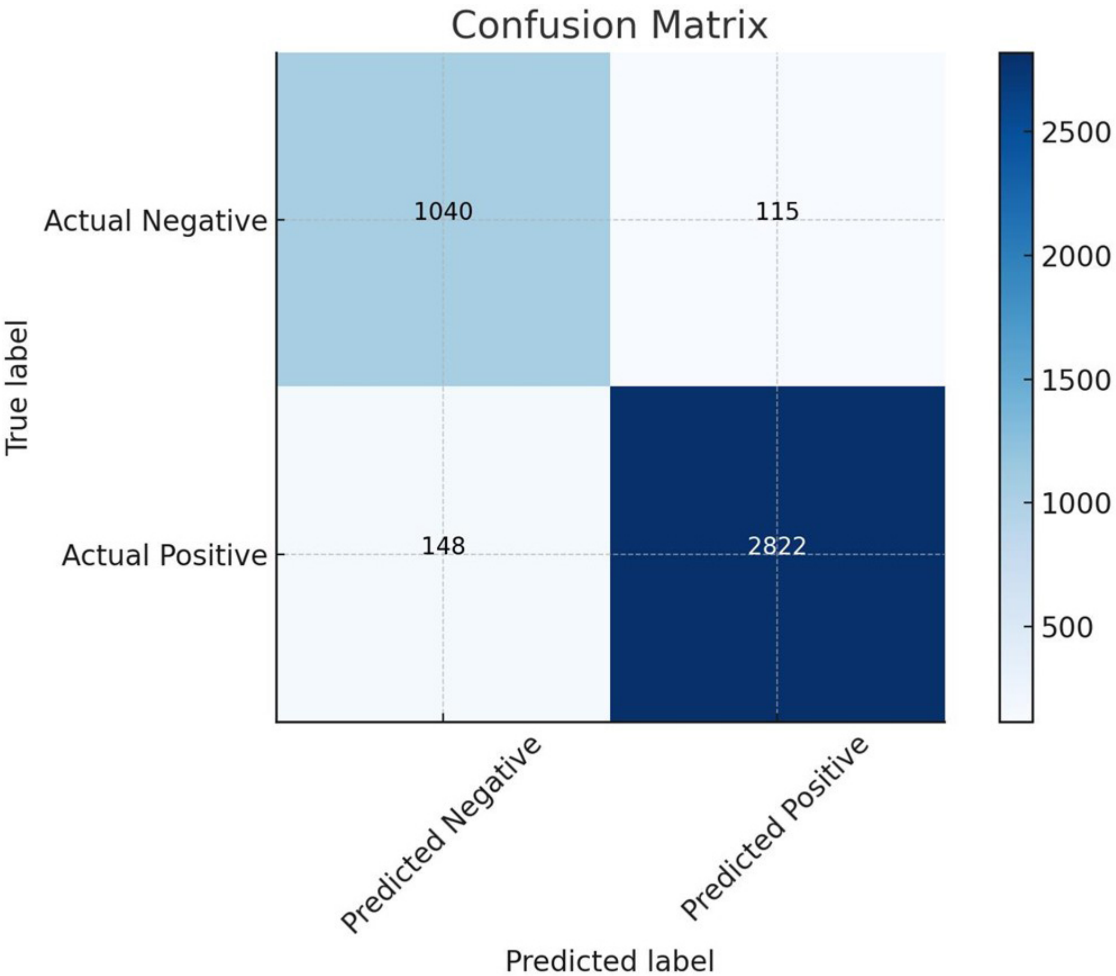

FIGURE 2. Confusion matrix of model.

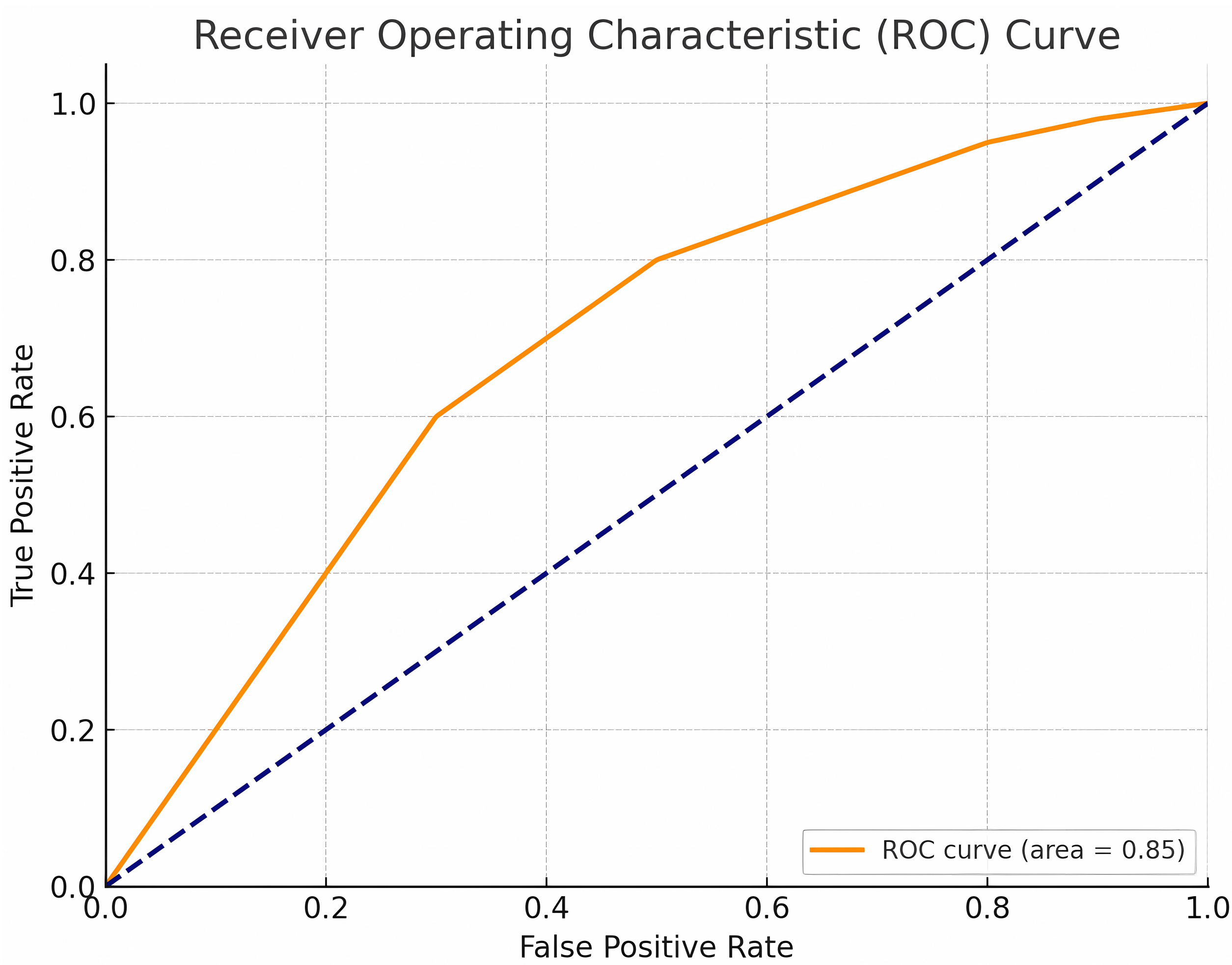

FIGURE 3. ROC curve.

TABLE 4. Logistic regression coefficients: multivariate results for factors associated with non-recognition of early stroke symptoms.

Variable	Coefficient (β)	Standard Error	Odds Ratio (OR)*	95% CI	*p*-value
Age	−0.05	0.01	0.95	0.93–0.97	<0.001
BMI	0.12	0.02	1.13	1.09–1.17	<0.001
Hypertension	0.28	0.05	1.32	1.19–1.46	<0.001
Cholesterol	0.03	0.01	1.03	1.01–1.04	<0.001
Triglycerides	0.04	0.01	1.04	1.03–1.06	<0.001
Depression	0.25	0.04	1.28	1.17–1.40	<0.001
Stress	0.18	0.03	1.20	1.11–1.30	<0.001
Smoking	0.10	0.02	1.11	1.06–1.16	<0.001
Alcohol Consumption	0.08	0.03	1.08	1.02–1.15	<0.001
Physical Activity	−0.20	0.05	0.82	0.74–0.91	<0.001
Diabetes Education	−0.15	0.04	0.86	0.79–0.94	<0.001
Occupation—Managers/Professionals/Clerks	0.25	0.22	1.28	0.75–1.73	0.280
Occupation—Agricultural/Forestry/Fishery Workers/Mechanical or Manual Laborers	0.35	0.14	1.42	1.08–1.53	0.015

**The regression model used in this study incorporates all input variables as confounding factors. CI: confidence intervals; BMI: Body Mass Index.*

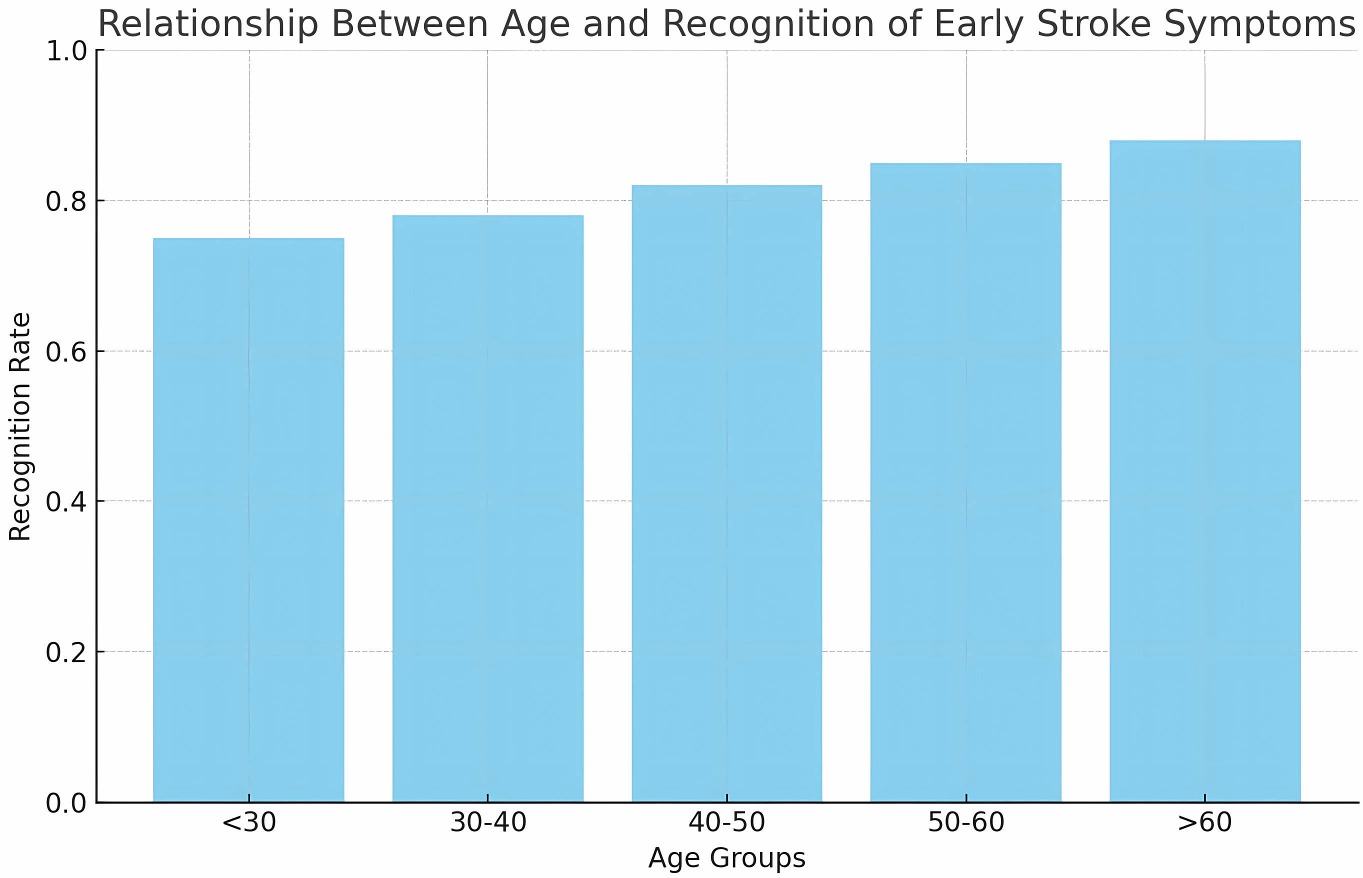

FIGURE 4. Relationship between age and recognition of early stroke symptoms.

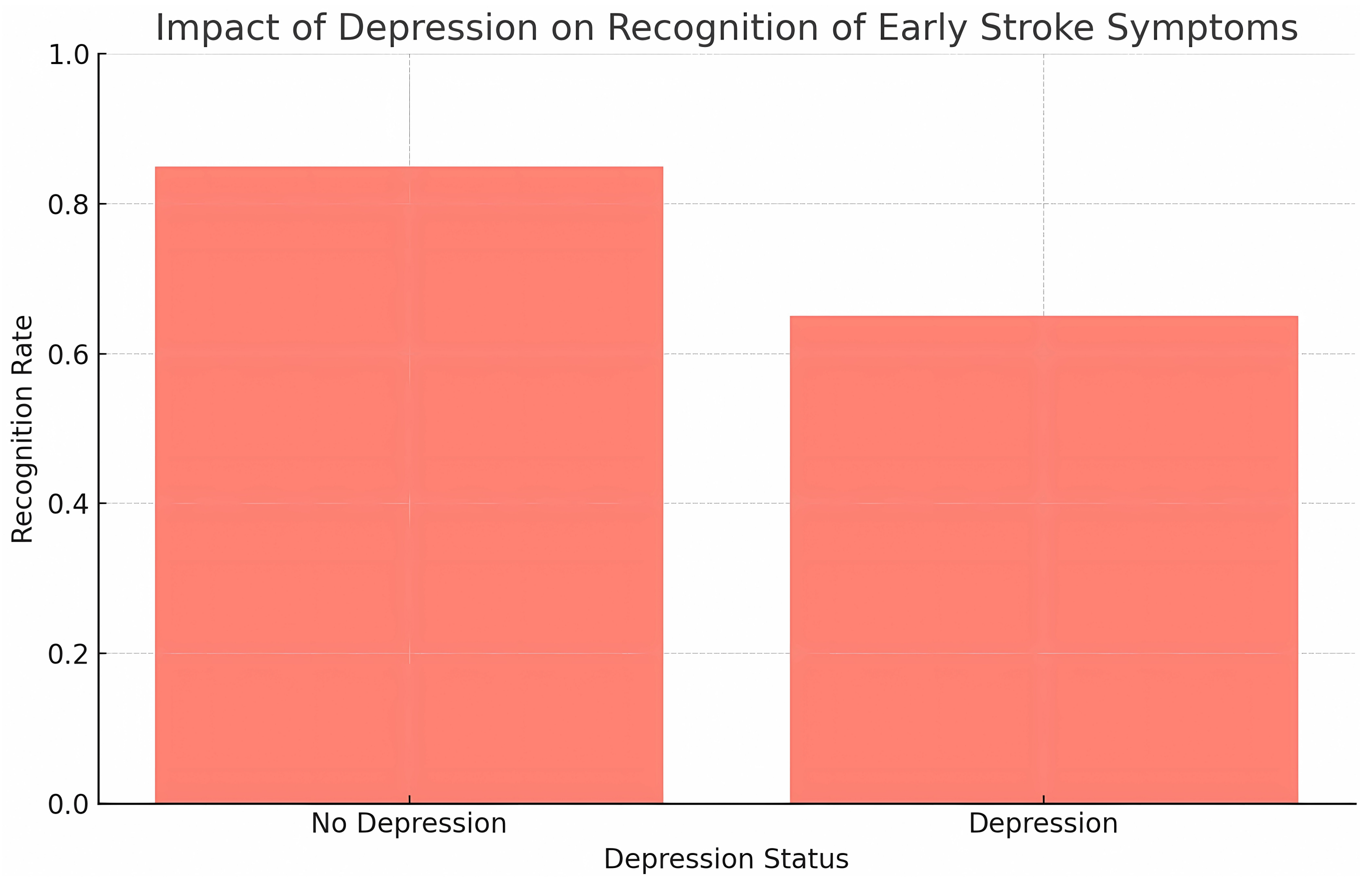

FIGURE 5. Impact of depression on recognition of early stroke symptoms.

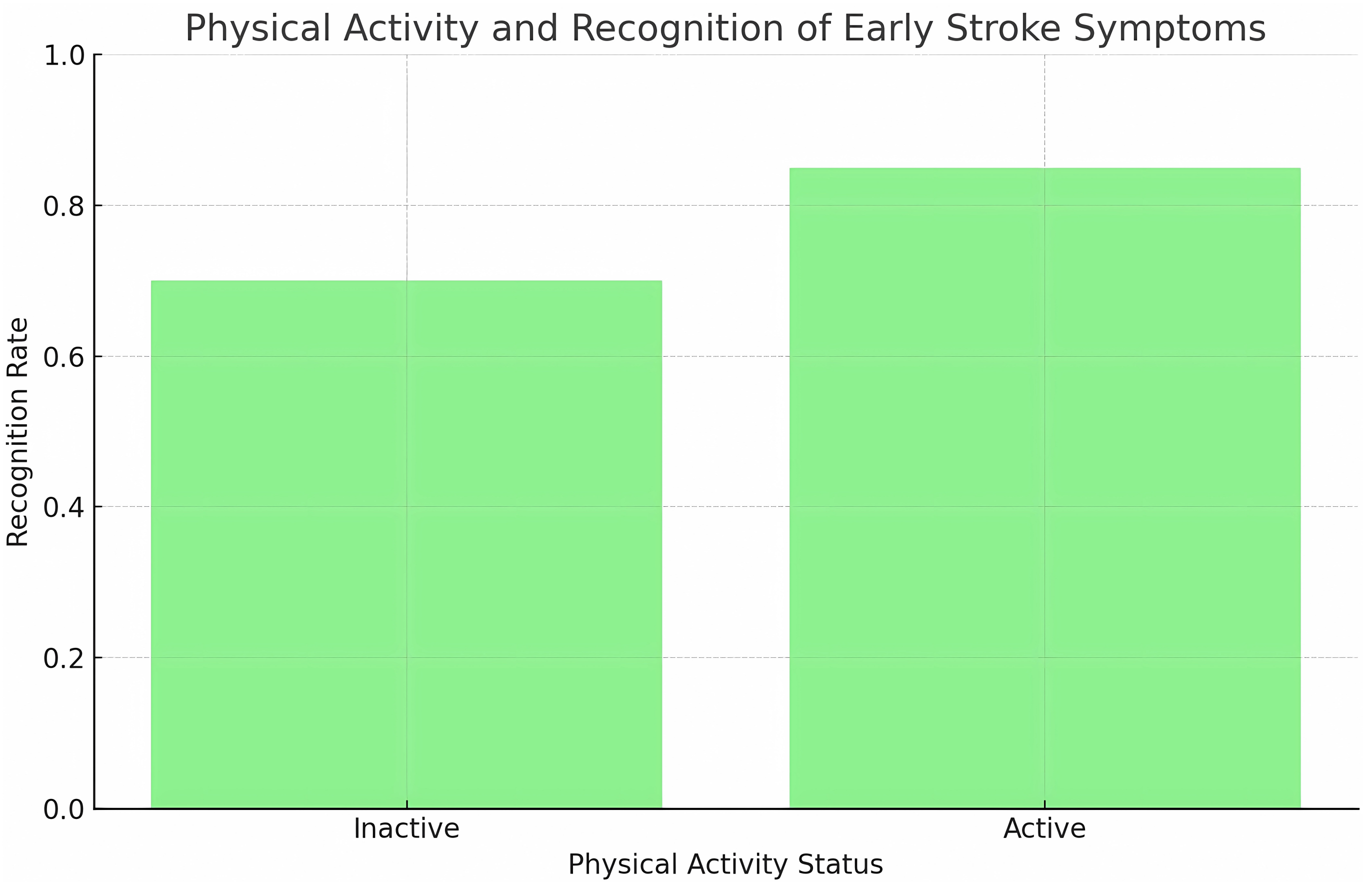

FIGURE 6. Physical activity and recognition of early stroke symptoms.

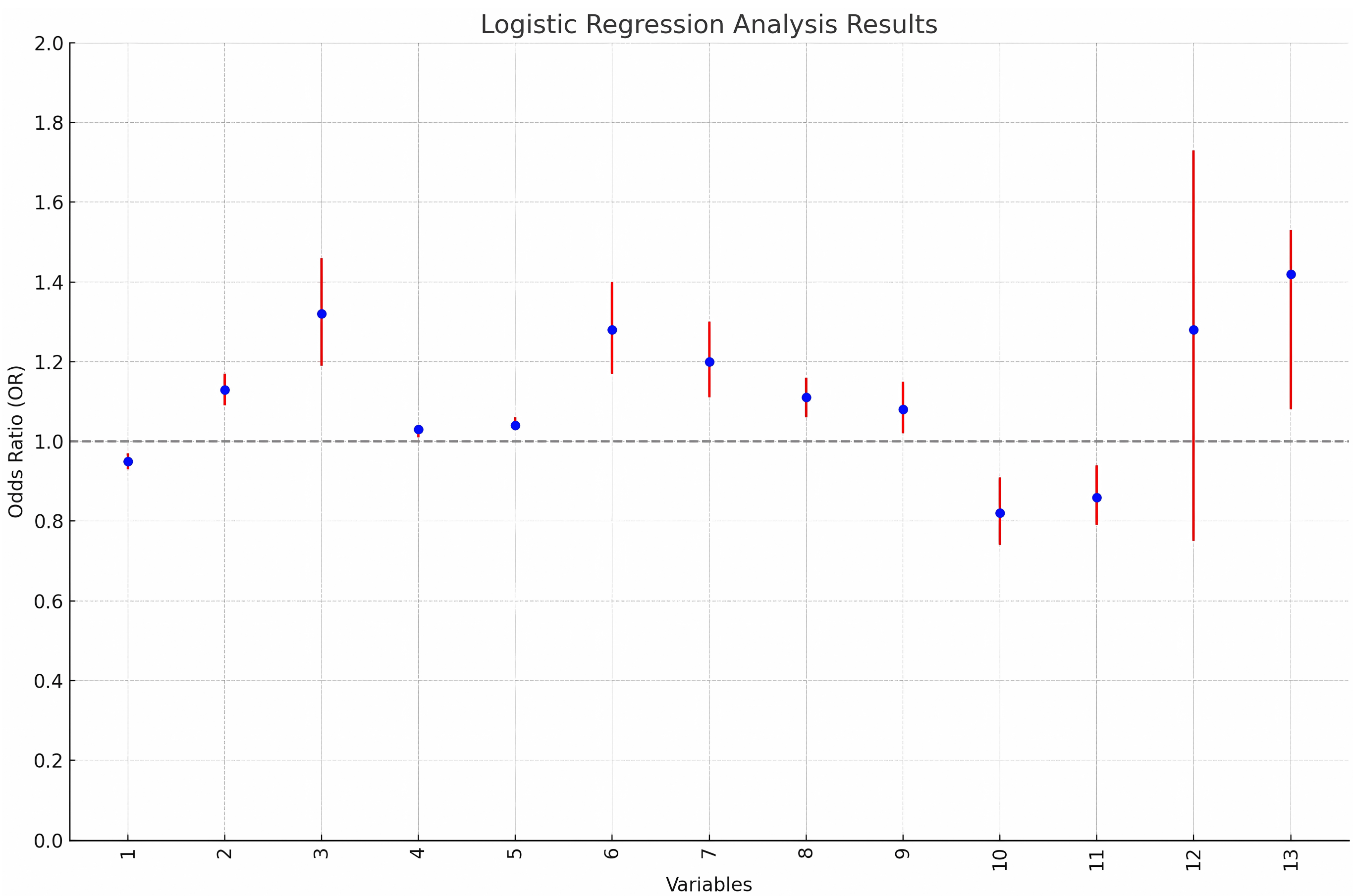

FIGURE 7. Results of multi-output gradient boosting and logistic Regression Model.

associated with better recognition. Sixth, occupation was an important factor. Compared to service or sales workers, managers/professionals/clerks did not show a significant difference in the recognition of early stroke symptoms (OR = 1.28, 95% CI = 0.75–1.73, p = 0.280). However, agricultural/forestry/fishery workers and mechanical or manual laborers were significantly less likely to recognize early stroke symptoms (OR = 1.42, 95% CI = 1.08–1.53, p = 0.015).

These findings suggest that targeted interventions focusing on both physiological and psychological factors, as well as promoting healthy lifestyle behaviors, can significantly improve the recognition of early stroke symptoms and overall health outcomes in working class males with diabetes.

4. Discussion

The results of this study provide valuable insights into the multifactorial nature of early stroke symptom recognition in working class males with diabetes. The findings emphasize the significant impact of demographic, physiological, psychological and lifestyle factors on the recognition of early stroke symptoms, suggesting the need for targeted interventions to improve health outcomes in this population.

A notable discovery is the relation between older age and better recognition of early stroke symptoms [17]. This aligns with the research by Reed Mszar *et al*. [18] (2020), which demonstrated the importance of sociodemographic variables such as ethnicity, immigration status, and education level in stroke symptom awareness among young adults in the United States. Likewise, a study examining stroke symptom awareness among Iranian adults identified age, education level, and prior knowledge about stroke as key predictors of stroke awareness [19].

Also, physiological factors, such as higher BMI and hypertension, were identified as significant predictors of poorer recognition of early stroke symptoms. This finding is consistent with the study by Z. Jarou *et al*. [20] (2013), which reported that individuals with a higher number of cumulative stroke risk factors had a significantly reduced ability to identify stroke symptoms. Additionally, the observation that women generally have better knowledge of major stroke symptoms than men, as documented in a study by T. Truelsen and L.-H. Krarup (2010) in Denmark [21], highlights the intricate relationship between gender and stroke symptom recognition. Further supporting the need for targeted interventions, N. Ojike *et al*. [22] (2016) highlighted racial and ethnic disparities in stroke awareness, emphasizing the importance of nuanced public health strategies. Moreover, C. Ellis and L. Egede (2009) [23] found that individuals with a history of stroke are better at recognizing symptoms, suggesting that experiential learning could play a significant role in enhancing stroke awareness.

In this study, psychological factors such as depression and stress were strongly associated with higher odds of not recognizing early stroke symptoms. This finding aligns with existing literature that shows the adverse impact of mental health issues on health awareness and self-care behaviors [24–29]. Addressing mental health through counseling and stress management programs could substantially improve symptom recognition and overall health outcomes for diabetic workers.

Lifestyle factors also played a crucial role in the recognition of early stroke symptoms. Smoking and alcohol consumption were negatively associated with symptom recognition, while regular physical activity and participation in diabetes education programs had a positive association [30–33]. These findings underscore the importance of promoting healthy lifestyle behaviors and providing continuous education to enhance health awareness and symptom recognition.

Occupational factors were also significant in influencing the recognition of early stroke symptoms. While managers, professionals, and clerks did not show a significant difference in recognition compared to service or sales workers, agricultural, forestry, fishery workers, and mechanical or manual laborers were significantly less likely to recognize symptoms [31, 34, 35]. This suggests that occupational health programs should be tailored to address the specific needs and challenges faced by different occupational groups.

The findings from this study suggest several important implications for clinical practice and public health interventions aimed at improving the recognition of early stroke symptoms among working class males with diabetes. First, targeted interventions are crucial. Younger male workers, who were found to have poorer recognition of early stroke symptoms, may benefit significantly from tailored educational programs and support systems. These programs should focus on enhancing their understanding of stroke symptoms and providing practical strategies to improve symptom recognition in the context of their work environment. Second, comprehensive care that addresses both physiological and psychological factors is essential for effective health management. The strong association between mental health issues, such as depression and stress, and poor symptom recognition highlights the need for integrating mental health support into routine diabetes care. Healthcare providers should screen for psychological distress and offer appropriate interventions, such as counseling or stress management programs, to help patients manage their mental health alongside their diabetes. Third, promoting healthy lifestyle behaviors is fundamental to improving health awareness and symptom recognition. The study found that lifestyle factors such as smoking, alcohol consumption, and physical activity significantly impacted the recognition of early stroke symptoms. Healthcare providers should encourage diabetic patients to adopt healthier behaviors, such as regular physical activity and smoking cessation, and provide resources and support to help them make these changes. Additionally, reducing alcohol consumption should be emphasized as part of a comprehensive health management plan.

Finally, continuous education is vital for empowering patients to manage their condition effectively and recognize early stroke symptoms. Participation in diabetes education programs was associated with better symptom recognition, underscoring the importance of ongoing education. Healthcare providers should offer regular educational sessions that cover various aspects of diabetes management, including the recognition of early stroke symptoms, and ensure these sessions are accessible and engaging for male workers. By implementing these evidence-based strategies, healthcare providers can help working class males with diabetes achieve better recognition of early stroke symptoms, improve their overall health outcomes and enhance their quality of life.

Despite its valuable contributions, this study has several limitations that should be considered in future research. First, the cross-sectional design restricts the ability to draw causal inferences between the factors identified and the recognition of early stroke symptoms. To establish causal relationships, longitudinal studies are required, which can track changes over time and provide a clearer picture of how these factors influence symptom recognition. Second, the study's reliance on self-reported data may introduce reporting bias. Participants might inaccurately report their recognition of stroke symptoms or other relevant variables. Future research should aim to incorporate objective measures, such as clinical assessments or validated instruments, to improve data accuracy. Third, the study focused solely on male workers, limiting the generalizability of the findings to other groups, such as female workers or non-working individuals. Future studies should include a broader demographic to assess whether the observed associations are consistent across different populations and employment statuses. Fourth, there may be discrepancies between feature importance plots and the odds ratios (OR) from logistic regression. For example, age may have a small coefficient in logistic regression but be a significant factor in the Gradient Boosting model. Future research should aim to integrate both methods to enhance interpretability and predictive accuracy. Fifth, while the integrated machine learning model offered valuable insights, further validation using external datasets is essential to confirm the model's robustness and generalizability. Additionally, examining the interactions between different factors and their combined effects on symptom recognition could yield more comprehensive insights into the underlying mechanisms. Future research should refine the model and test it in diverse settings to ensure its applicability and reliability across various contexts.

5. Conclusions

This study emphasizes the complex nature of early stroke symptom recognition in working-class males with diabetes, and highlights the critical need for targeted interventions to enhance health outcomes. By addressing both physiological and psychological factors and promoting healthy lifestyle behaviors, healthcare providers can improve the recognition of early stroke symptoms and mitigate the risk of adverse health events in this high-risk group. Future research should prioritize longitudinal studies, involve more diverse populations, and validate predictive models to deepen our understanding and improve the management of stroke symptom recognition. Additionally, subsequent research should explore factors related to stroke knowledge among very old adults and teenagers, as understanding these demographics could provide valuable insights for tailored educational and preventive strategies.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are provided at the request of the corresponding author. 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 CONTRIBUTIONS

HB—conceptualization; software; methodology; validation; investigation; writing-original draft preparation; formal analysis; writing-review and editing; visualization; supervision; project administration; funding acquisition; contributed to editorial changes in the manuscript; read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Before conducting the survey, written informed consent was acquired from all participants. This study employed only pre-existing, anonymized data. It adhered to the principles outlined in the Declaration of Helsinki. The protocol for the Panel Study of Worker's Compensation Insurance received approval from the Institutional Review Board (IRB) of the KNHANES (IRB approval numbers: 2018-01-03-5C-A).

ACKNOWLEDGMENT

The author would like to thank the Korea Centers for Disease Control and Prevention for providing the raw data.

FUNDING

This research was 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).

CONFLICT OF INTEREST

The authors declare no conflict of interest. Haewon Byeon is serving as one of the Editorial Board members/Guest Editor of this journal. We declare that Haewon Byeon had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to WYS.

REFERENCES

[1] Hong KS, Bang OY, Kang DW, Yu KH, Bae HJ, Lee JS, *et al*. Stroke statistics in Korea: part I. Epidemiology and risk factors: a report from the Korean stroke society and clinical research center for stroke. Journal of Stroke. 2013; 15: 2–20.

[2] Han MK, Huh Y, Lee SB, Park JH, Lee JJ, Choi EA, *et al*.; Korean consortium for health and aging research. Prevalence of stroke and transient ischemic attack in Korean elders: findings from the Korean Longitudinal Study on Health and Aging (KLoSHA). Stroke. 2009; 40: 966–969.

[3] Lorber D. Importance of cardiovascular disease risk management in patients with type 2 diabetes mellitus. Diabetes, Metabolic Syndrome and Obesity. 2014; 7: 169–183.

[4] Pearlman DN, Affleck P, Goldman D. Disparities in awareness of the warning signs and symptoms of a heart attack and stroke among Rhode Island adults. Medicine & Health Rhode Island. 2011; 94: 183–185.

[5] Ma J, Zhang N, Xiao JY, Wang JX, Li X, Wang J, *et al*. Public awareness of acute myocardial infarction symptoms and emergency response in community residents. European Heart Journal. 2021; 42: ehab724.3133.

[6] Alharbi MF, Alanazi G, Almesned B, Alenazi Y, Alruwaili R, Alanazi W, *et al*. Public awareness of early symptoms of acute myocardial infarction in Arar, Saudi Arabia. International Journal of Medicine in Developing Countries. 2024; 8: 151–157.

[7] Ghanem N, Elshal A, Darwish SY, Emam M, ElWalili M, Mehanna R. Role of awareness activities in mega-events. European Heart Journal. 2021; 23: suab069.019.

[8] Uhland D. Evaluating the impact of a standardized education class on a person diagnosed with chronic kidney disease, stage IV. Himmelfarb Health Sciences Library. 2018; 1–37.

[9] Morrison S, Subrahmanian KN, Ali S, Osterberg L. Health outreach program to educate (HOPE) palo alto. Journal of Community Health. 2016; 41: 1047–1055.

[10] Myke-Mbata B, Dioka C, Meludu S, Adebisi SA, Oghagbon EK. Preponderance of increased glycosylated hemoglobin (HBA1c) and chronic disease in Nigerian adults. Journal of Diabetes Research. 2018; 2018: 1–8.

[11] Harini K, Rekha KKS. Evaluating performance of identifying at-risk students and learning achievement model using accuracy and f-measure by comparing logistic regression, generalized linear model and gradient boost machine algorithm. International Conference for Advancement in Technology. 2022; 4: 1–7.

[12] Balaji C, Chakkaravarthy A. Automatic identification of fake news circulation in social media using logistic regression over naïve bayes and XGBoost algorithm to improve accuracy. PURE Network Journals. 2022; 13: 654–660.

[13] Vivek R, R M. Analyze the lack of accuracy in loan prediction using logistic regression compared with random forest to improve accuracy. Eighth International Conference on Science Technology Engineering and Mathematics. 2023; 174: 1–5.

[14] Vasu VN, R S, S SM, Nelson M. Prediction of defective products using logistic regression algorithm against linear regression algorithm for better accuracy. International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies. 2022; 161–166.

[15] Vishnuvardhan T, Rama A. Comparison of accuracy rate in prediction of cardiovascular disease using random forest with logistic regression. Cardiometry. 2022; 25: 1526–1531.

[16] Azizov V, Mirzazade VA, Garibova C. The effect of structured training on blood pressure indicators in Type II diabetes mellitus patients having arterial hypertension. Vestnik Sankt-Peterburgskogo Universiteta—Meditsina. 2020; 13: 7–12. (In Russian)

[17] Harmawati H, Etriyanti E, Hardini S. Deteksi dini gejala awal stroke. Journal of Acute Stroke. 2021; 3: 186–189. (In Indonesian)

[18] Mszar R, Mahajan S, Valero-Elizondo J, Yahya T, Sharma R, Grandhi GR, *et al*. Association between sociodemographic determinants and disparities in stroke symptom awareness among us young adults. Stroke. 2020; 51: 3552–3561.

[19] Hosseininezhad M, Ebrahimi H, Seyedsaadat S, Bakhshayesh B, Asadi M, Ghayeghran A. Awareness toward stroke in a population-based sample of Iranian adults. Iranian Journal of Neurology. 2017; 16: 7–14.

[20] Jarou Z, Harris N, Gill L, Azizi M, Gabasha S, LaBril R. Public stroke knowledge: those most at risk, least able to identify symptoms. Medical Student Research Journal. 2013; 3: 003–008.

[21] Truelsen T, Krarup L. Stroke awareness in Denmark. Neuroepidemiology. 2010; 35: 165–170.

[22] Ojike N, Ravenell J, Seixas A, Masters-Israilov A, Rogers A, Jean-Louis G, *et al*. Racial disparity in stroke awareness in the US: an analysis of the 2014 national health interview survey. Journal of Neurology and Neurophysiology. 2016; 7: 365.

[23] Ellis C, Egede L. Stroke recognition among individuals with stroke risk factors. The American Journal of the Medical Sciences. 2009; 337: 5–10.

[24] Swartz RH, Bayley M, Lanctôt KL, Murray BJ, Cayley ML, Lien K, *et al*. Post-stroke depression, obstructive sleep apnea, and cognitive impairment: rationale for, and barriers to, routine screening. International Journal of Stroke. 2016; 11: 509–518.

[25] Lawrence M, Booth J, Mercer S, Crawford E. A systematic review of the benefits of mindfulness-based interventions following transient ischemic attack and stroke. International Journal of Stroke. 2013; 8: 465–474.

[26] Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, *et al*. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2016; 47: e98–e169.

[27] Fonarow GC, Calitz C, Arena R, Baase C, Isaac FW, Lloyd-Jones D, *et al*. Workplace wellness recognition for optimizing workplace health: a presidential advisory from the American Heart Association. Circulation. 2015; 131: e480–e497.

[28] Yang X, Li Z, Sun J. Effects of cognitive behavioral therapy-based intervention on improving glycaemic, psychological, and physiological outcomes in adult patients with diabetes mellitus: a meta-analysis of randomized controlled trials. Frontiers in Psychiatry. 2020; 11: 711.

[29] Hinwood M, Ilicic M, Gyawali P, Coupland K, Kluge MG, Smith A, *et al*. Psychological stress management and stress reduction strategies for stroke survivors: a scoping review. Annals of Behavioral Medicine. 2023; 57: 111–130.

[30] Duque AS, Fernandes L, Correia AF, Calvinho I, Pinto M, Freitas P, *et al*. Awareness of stroke risk factors and warning signs and attitude to acute stroke. International Archives of Medicine. 2015; 8: 1–18.

[31] Oh GJ, Lee K, Kim K, Lee YH. Differences in the awareness of stroke symptoms and emergency response by occupation in the Korean general population. PLOS ONE. 2019; 14: e0218608.

[32] Wang C, Huang X, Xiang S, Lian XG, Yuan C, Jiang X, *et al*. The effect of an overall healthy lifestyle on early-onset stroke: a cross-sectional study. Annals of Palliative Medicine. 2020; 9: 2623–2630.

[33] Douma MJ, Aves T, Allan KS, Bendall JC, Berry DC, Chang WT, *et al*. First aid cooling techniques for heat stroke and exertional hyperthermia: a systematic review and meta-analysis. Resuscitation. 2020; 148: 173–190.

[34] Boltaevich UM, Umarovich AT. First AID for stroke. European Journal of Innovation in Nonformal Education. 2024; 4: 266–270.

[35] Landry KK, Alexander KS, Zakai N, Judd S, Kleindorfer D, Howard V, *et al*. Association of stroke risk biomarkers with stroke symptoms: the Reasons for Geographic and Racial Differences in Stroke cohort. Journal of Thrombosis and Haemostasis. 2017; 15: 21–27.

How to cite this article: Haewon Byeon. Non-cognitive factors influencing early stroke symptom recognition among Korean working-class males with hypertension and diabetes: an integrated multi-output gradient boosting and logistic regression approach. Journal of Men's Health. 2025; 21(6): 21-32. doi: 10.22514/jomh.2025.077.