Submitted: 10 August, 2024 Accepted: 23 January, 2025 Published: 30 October, 2025 DOI:10.22514/jomh.2025.121

JOMH Journal of Men's Health

ORIGINAL RESEARCH

Identification of early myocardial infarction symptoms in adult male smokers using sparse attention mechanisms and quantile regression: results from 97,304 participants in a nationwide survey in Korea

Haewon Byeon[1,*]

[1]Worker's Care & Digital Health Lab, Department of Future Technology, Korea University of Technology and Education, 31253 Cheonan, Republic of Korea

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

Abstract

Background: Cardiovascular diseases (CVDs) are a leading cause of mortality worldwide, with myocardial infarction (MI) being a significant contributor. This study aims to identify key demographic and behavioral factors influencing the early recognition of MI symptoms among adult male smokers and non-smokers. **Methods**: Utilizing Sparse Attention Mechanisms and comparing their performance with traditional models such as Classification and Regression Tree (CART), C4.5 and Rotation Random Forest, we analyze predictors of MI symptom recognition. Data from the 2021 Community Health Survey included 97,304 participants, with 30,858 male smokers and 66,446 male non-smokers. **Results**: The findings reveal that age, marital status, residential area, education level, high-risk drinking, diabetes prevalence and occupation significantly impact MI symptom recognition. Age was the most significant predictor, with older individuals showing higher recognition rates. Marital status and residential area were also important, indicating that married individuals and those living in cities or rural areas had higher recognition rates. Higher educational attainment was associated with better recognition, emphasizing the role of health literacy. High-risk drinking and diabetes prevalence showed a trend towards significance at higher quantiles, suggesting their impact on high-risk groups. **Conclusions**: The study highlights the need for targeted educational interventions focusing on high-risk groups such as older adults, those with lower educational attainment, and individuals with high-risk drinking behaviors or diabetes. Public health strategies should address regional disparities in healthcare access and involve spouses in educational programs to enhance MI symptom recognition.

Keywords

Myocardial infarction; Symptom recognition; Sparse attention mechanisms; Multiple risk factors; Demographic factors

1. Introduction

Cardiovascular diseases (CVDs) are a leading cause of death globally, with myocardial infarction (MI) being a major contributor to this statistic. The ability to recognize MI symptoms early is crucial for seeking prompt medical intervention, which can significantly decrease mortality rates and improve patient outcomes. Smoking is a well-documented risk factor for myocardial infarction (MI), particularly among men [1]. The link between smoking and the heightened risk of acute myocardial infarction (AMI) is extensively supported by clinical research. Specifically, current smokers demonstrate a significantly higher likelihood of experiencing MI compared to their non-smoking counterparts, largely due to the damaging effects of smoking on cardiovascular health [1]. Studies [2, 3] have consistently shown that the risk of myocardial infarction nearly quadruples for smokers, underscoring the profound impact smoking has on this acute cardiovascular event. Interestingly, smoking cessation markedly reduces this risk, bringing former smokers closer to the risk levels observed in non-smokers over time.

The influence of smoking extends to other cardiovascular parameters. For instance, it is associated with myocardial damage and an exacerbated state of cardiovascular stress, particularly through increased myocardial oxygen demand coupled with restricted oxygen delivery due to vessel constriction [4, 5]. Furthermore, the so-called "smoker's paradox" has been observed in clinical settings, where smokers with AMI sometimes show paradoxically better short-term outcomes, although this should not overshadow the long-term deleterious effects of smoking on heart health [6]. Additionally, smoking has been correlated with an earlier onset of AMI incidents, illustrating its role in not only heightening risk but also precipitating cardiovascular emergencies earlier in life. These findings stress the importance of smoking cessation programs as essential interventions to reduce the incidence and severity of myocardial infarctions among men [7].

Smoking significantly increases the risk of MI, exacerbating conditions such as atherosclerosis, increasing blood coagulability and causing arterial blockages. Despite extensive research into the detrimental effects of smoking, there is still a lack of understanding regarding the specific factors that affect MI symptom recognition among smokers [8, 9]. Data from the 2021 Community Health Survey indicate that the rate of MI symptom recognition among smokers is lower than the national average, highlighting the urgent need for targeted educational interventions [10].

Historically, most research in this area has used descriptive and cross-sectional study designs, which, while valuable, have inherent limitations. These studies often struggle to establish causal relationships due to their observational nature [11, 12]. Additionally, many do not sufficiently account for the complex interactions between various demographic and behavioral factors, potentially leading to biased or incomplete conclusions [13, 14]. Traditional statistical methods commonly used in prior research, such as logistic regression models, are limited in their ability to handle high-dimensional data and complex variable interactions, further complicating the analysis [15, 16].

Attention mechanisms, especially Sparse Attention Mechanisms, present a robust solution to these limitations by enabling models to focus on the most pertinent parts of the input data. This selective focus enhances the model's capacity to identify critical factors affecting MI symptom recognition, resulting in more accurate and interpretable outcomes [17, 18]. Sparse Attention Mechanisms are particularly suitable for high-dimensional healthcare data, where many variables may be irrelevant or redundant. By concentrating on the most informative features, these mechanisms can improve model performance and provide deeper insights into the factors affecting MI symptom recognition among smokers and non-smokers [19, 20].

However, the attention mechanism, a type of deep learning, is still limited in terms of interpretability, as it is a "black box" model, which is a feature of deep learning despite its high accuracy. To compensate for these limitations, incorporating Quantile Regression (QR) into deep learning models offers several distinct advantages. Unlike traditional mean regression models, QR allows for a more detailed analysis by estimating the conditional quantiles of the response variable [21]. This capability enables a thorough examination of the relationships between predictors and various points of the outcome distribution. In the context of MI symptom recognition, QR can elucidate how different factors influence the likelihood of early symptom recognition across varying severity levels [22]. For instance, specific demographic or behavioral factors may have a more significant impact on individuals at higher risk levels, which QR can capture [23]. Moreover, integrating QR with Sparse Attention Mechanisms enhances model interpretability by highlighting the most significant predictors for different quantiles. This combined approach offers a nuanced understanding of the factors influencing MI symptom recognition, facilitating the development of targeted educational interventions for high-risk groups [24, 25].

The objectives of this study were First; it aimed to develop a predictive model using Sparse Attention Mechanisms to identify key factors influencing early MI symptom recognition among male smokers and non-smokers. Second, it sought to compare the performance of this model with traditional models such as CART, C4.5 and Rotation Random Forest. Third, the study assessed the importance of various predictors and employed Quantile Regression to estimate the Odds Ratios and 95% confidence intervals for the top seven variables: age, marital status, residential area, education level, occupation, high-risk drinking, and diabetes prevalence. By addressing the methodological limitations of previous studies and leveraging advanced analytical techniques, this research aspired to contribute to the existing body of knowledge and support the development of effective public health interventions.

2. Method

2.1 Study design and participants

This study employed a descriptive research design to investigate the early recognition of myocardial infarction (MI) symptoms among adult male smokers and non-smokers in South Korea. Data for this study were obtained from the 2021 Korea National Health and Nutrition Examination Survey (KNHANES), an annual nationwide survey conducted by the Korea Centers for Disease Control and Prevention (KCDC). KNHANES collects comprehensive data on a wide range of health behaviors, chronic disease prevalence, and other health-related factors within the South Korean adult population.

The study population included adult male participants who completed the 2021 KNHANES. Of the initial 229,242 respondents, 18,343 individuals were excluded from the analysis due to missing data on relevant variables. This resulted in a final sample size of 210,899 participants. Subsequently, the sample was divided into two groups based on smoking status: smokers and non-smokers. Smokers were defined as individuals who reported currently smoking cigarettes, while non-smokers were defined as those who reported no current or past smoking history. The final analysis included 97,304 male participants, comprising 30,858 male smokers and 66,446 male non-smokers.

Inclusion criteria for the study were as follows: First, adult males aged 19 years or older. Second, participation in the 2021 KNHANES. Exclusion criteria included participants who: First, did not provide responses to questions regarding smoking status. Second, did not provide responses to questions regarding awareness of MI symptoms.

2.2 Data collection

This study is a secondary data study using raw data from the 2021 Community Health Survey. Data collection was carried out using structured questionnaires administered by trained interviewers in the 2021 Community Health Survey. The questionnaires included a wide range of questions related to demographic characteristics, health behaviors and recognition of MI symptoms. Specifically, the survey included items to assess the recognition of early MI symptoms, such as sudden chest pain, pain, or discomfort in the arms, back, neck, jaw or stomach, shortness of breath, nausea, or lightheadedness. Respondents were asked whether they recognized each symptom, and their responses were coded as either "yes" (1) or "no" (0).

The survey also collected detailed information on various demographic and behavioral factors that could influence the recognition of MI symptoms. These factors included age, marital status, residential area, education level, occupation, high-risk drinking, and the prevalence of diabetes. The responses to these questions were used to create the independent variables for the analysis.

2.3 Measurement

The primary outcome variable in this study is the recognition of early MI symptoms. Respondents were asked to identify whether they recognized specific symptoms associated with MI, such as sudden chest pain, pain, or discomfort in the arms, back, neck, jaw or stomach, shortness of breath, nausea or lightheadedness. The survey responses were coded to indicate whether the respondent recognized the symptoms (1) or did not recognize the symptoms (0).

The independent variables include demographic and behavioral factors such as age, marital status, residential area, education level, occupation, high-risk drinking, and diabetes prevalence. These variables were selected based on their potential influence on the recognition of MI symptoms and are detailed in the Table 1.

2.4 Model development

The Sparse Attention Mechanisms and Quantile Regression model was designed to efficiently handle large-scale datasets and provide robust predictions across different quantiles of the response variable. The key steps involved in developing this model are:

2.4.1 Sparse attention mechanisms

Sparse Attention Mechanisms aim to enhance the efficiency of attention layers in neural networks by focusing on a subset of relevant inputs rather than considering all possible inputs. This mechanism reduces computational complexity and improves scalability.

1. Attention Score Calculation: The attention scores are computed using a compatibility function, typically a dot product or scaled dot product, between the query and key vectors. Mathematically, the attention score "α_{ij}" for the (i)-th query and (j)-th key is given by: $[\alpha_{ij} = \frac{(QK^T)_{ij}}{\sqrt{d_k}}]$ where (Q) is the query matrix, (K) is the key matrix, and (d_k) is the dimensionality of the keys.

2. Sparse Attention Selection: To induce sparsity, only the top-k attention scores are selected for each query, where (k) is a hyperparameter.

3. Weighted Sum Calculation: The output of the sparse attention mechanism is a weighted sum of the values, based on the sparse attention scores: $[Oi = \sum j\alpha_{ij}^{(sparse)} V_j]$ where (V) is the value matrix.

2.4.2 Quantile regression

Quantile Regression is employed to estimate the conditional quantiles of the response variable, providing a more comprehensive understanding of the data distribution.

TABLE 1. Description of variables used in the study.

Variable	Description
Age	Categorized into three groups: 19–44 years, 45–64 years and 65 years and older
Marital status	Single or married
Residential area	Metropolis, city or rural area
Education level	Less than high school, high school graduate, college graduate or higher
Occupation	Categorized into various occupational groups, including agriculture, professional and administrative roles
High-risk drinking	Defined as consuming more than 7 drinks per occasion for men, more than twice a week
Diabetes prevalence	Self-reported diagnosis of diabetes
Smoking cessation plan	Presence or absence of a plan to quit smoking
Physical activity	Level of physical activity, categorized as moderate or higher
Obesity	Body Mass Index (BMI) categorized as BMI <25 or BMI ≥ 25
Depression	Presence of depressive symptoms, indicated by a PHQ-9 score ≥ 10
Health examination	Participation in general health and cancer screenings within the past two years
Hypertension	Self-reported diagnosis of hypertension
Unmet medical needs	Instances where medical care was needed but not received
Subjective health status	Self-rated health categorized as good, average or bad
Socio-physical environment	Perception of the social and physical environment as good or bad

PHQ-9: Patient Health Questionnaire-9.

1. Quantile Loss Function: The quantile regression model minimizes the quantile loss function, which is defined for the (τ)-th quantile as: $[L_\tau(y, \hat{y}) = \sum_{i=1}^{n} \rho_\tau(y_i - \hat{y}i)\backslash]$ where $(\rho\{\tau\}(u) = u(\tau - \exists o\{I\}(u < 0)))$ and (I) is the indicator function.

2. Model Formulation: The linear quantile regression model is formulated as: $[Q_\tau(y \mid X) = X\beta\tau]$ where $(Q\{\tau\}(y \mid \exists b\{X\}))$ where denotes the (τ)-th quantile of (y) conditional on (X), and (β_τ) represents the quantile-specific coefficients.

2.5 Model evaluation

The models were evaluated using k-fold cross-validation (k = 10) to ensure robustness and prevent overfitting. The dataset was divided into 10 subsets, and each model was trained on 9 subsets and tested on the remaining subset. This process was repeated 10 times, and the results were averaged to obtain the final performance metrics.

The performance of the models was assessed using several metrics: accuracy, precision, recall and F1-score.

Where TP, TN, FP and FN denote true positives, true negatives, false positives, and false negatives, respectively.

- Accuracy measures the proportion of correctly classified instances: $[Accuracy = \frac{TP+TN}{TP+TN+FP+FN}]$
- Precision calculates the proportion of true positive instances among the instances classified as positive: $[Precision = \frac{TP}{TP+FP}]$
- Recall determines the proportion of true positive instances among all actual positive instances: $[Recall = \frac{TP}{TP+FN}]$
- F1-score provides the harmonic mean of precision and recall: $[F1\text{-}score = 2\times\frac{Precision\times Recall}{Precision+Recall}]$

2.6 Variable importance analysis

The importance of each variable was analyzed to understand its contribution to the prediction of the response variable. In the context of sparse attention mechanisms, attention weights were analyzed to determine the significance of each input feature. For quantile regression, the magnitude of the quantile-specific coefficients (β_τ) was used to assess the importance of each predictor.

2.7 Quantile regression coefficients and confidence intervals

For the seven most significant predictors, the quantile regression coefficients and their 95% confidence intervals (CI) were calculated to quantify the impact of these predictors across different quantiles of the response variable. The quantile regression coefficients (β_τ) represent the change in the conditional quantile of the response variable for a one-unit change in the predictor. The confidence intervals were calculated using bootstrapping techniques to provide a range within which the true quantile regression coefficient is expected to fall with 95% confidence.

3. Results

3.1 General characteristics of the subject

Table 2 presented the demographic characteristics of the study subjects. The age distribution indicated that the largest proportion of participants fell within the 45–64 age bracket, accounting for 42.02% of the total, followed by the 19–44 age group at 31.51%, and those aged 65 and above at 26.47%. Marital status revealed a predominant representation of married individuals, comprising 73.00% of the sample, while single individuals made up 27.00%. In terms of urban residence, the majority of subjects resided in cities (69.79%), with smaller proportions living in metropolises (28.32%) and villages (1.89%). Educational attainment among participants was notably high, with 45.03% having completed university education, 32.06% possessing a high school diploma, and lower percentages for middle school (9.03%) and elementary school (13.88%) education levels. The income distribution showed a significant concentration in the highest income bracket (≥400, 10,000 won), encompassing 53.94% of participants, while the remaining subjects were relatively evenly distributed across the lower income categories.

3.2 Results using the sparse attention mechanisms model

To identify the key factors influencing the early recognition of myocardial infarction (MI) symptoms, a predictive model utilizing Sparse Attention Mechanisms was developed. This model was selected due to its proficiency in handling high-dimensional data and its capability to emphasize the most pertinent features, thereby enhancing both accuracy and interpretability. The performance of the Sparse Attention Mechanisms model was benchmarked against traditional models, including CART, C4.5 and Rotation Random Forest.

Fig. 1 presents the performance metrics of the Sparse Attention Mechanisms model in comparison to the traditional models. The Sparse Attention Mechanisms model exhibited superior performance across all evaluated metrics, including accuracy, precision, recall and F1-score.

3.3 Results of variable importance

The importance of each variable in predicting the early recognition of MI symptoms was assessed using the trained Sparse Attention Mechanisms model. The top seven predictors identified were age, marital status, residential area, education level, high-risk drinking, diabetes prevalence and occupation. Fig. 2 presents the variable importance scores for these factors.

3.4 Results of quantile regression analysis

To further analyze the relationship between the key predictors and the recognition of MI symptoms, Quantile Regression (QR) was employed. Table 3 presents the results of the Quantile Regression analysis, showing the effect of each predictor at different quantiles (0.25, 0.50, 0.75) of the outcome distribution. The coefficients, standard errors, *t*-statistics and *p*-values are reported for each quantile.

The QR analysis reveals that age, marital status, residential area, and education level are significant predictors of MI symptom recognition across all quantiles (Fig. 3). The findings

TABLE 2. General characteristics of the subject.

Variables		Total (n)	Total (%)
Age (yr)			
	19–44	30,660	31.51%
	45–64	40,886	42.02%
	≥65	25,758	26.47%
Marital status			
	Single	26,275	27.00%
	Married	71,029	73.00%
Residential area			
	Metropolis	27,554	28.32%
	City	67,909	69.79%
	Villages	1841	1.89%
Educational level			
	Elementary school	13,508	13.88%
	Middle school	8784	9.03%
	High school	31,194	32.06%
	University	43,818	45.03%
Monthly income (10,000 won)			
	<100	9300	9.56%
	100–199	11,675	12.00%
	200–299	12,493	12.84%
	300–399	11,354	11.67%
	≥400	52,482	53.94%

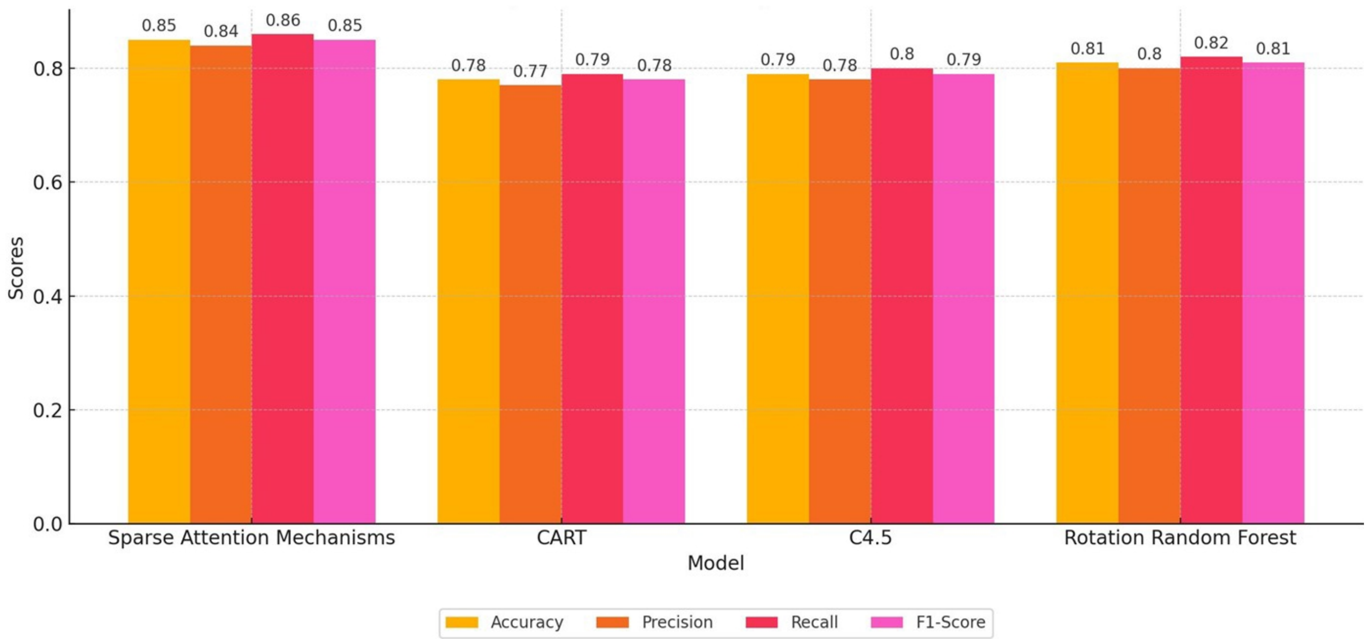


FIGURE 1. Comparison of models by metrics. CART: Classification and Regression Tree.

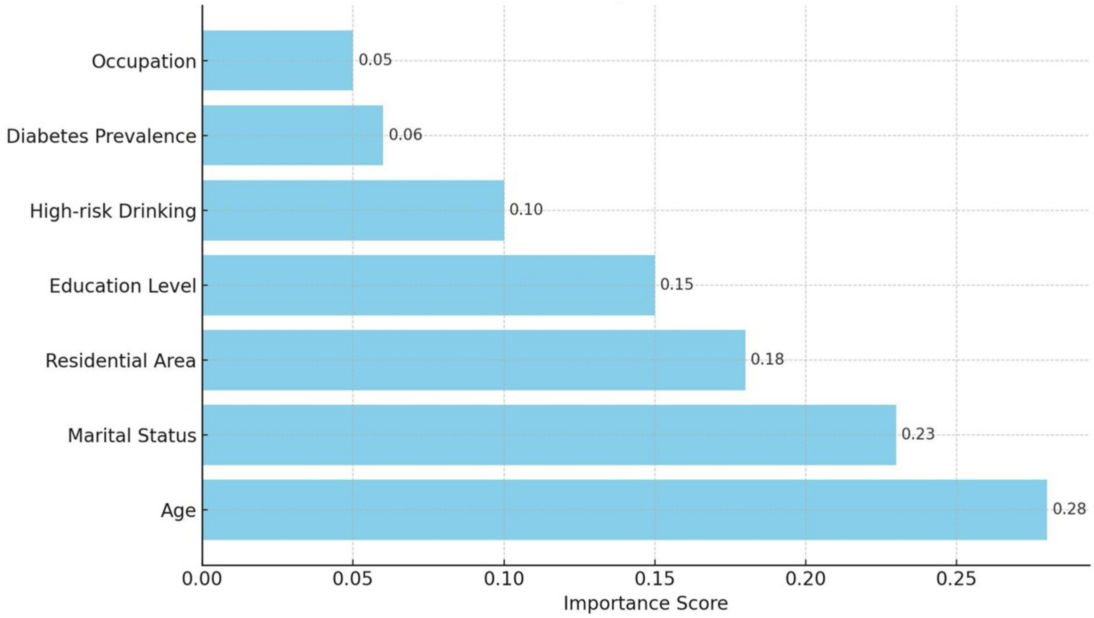
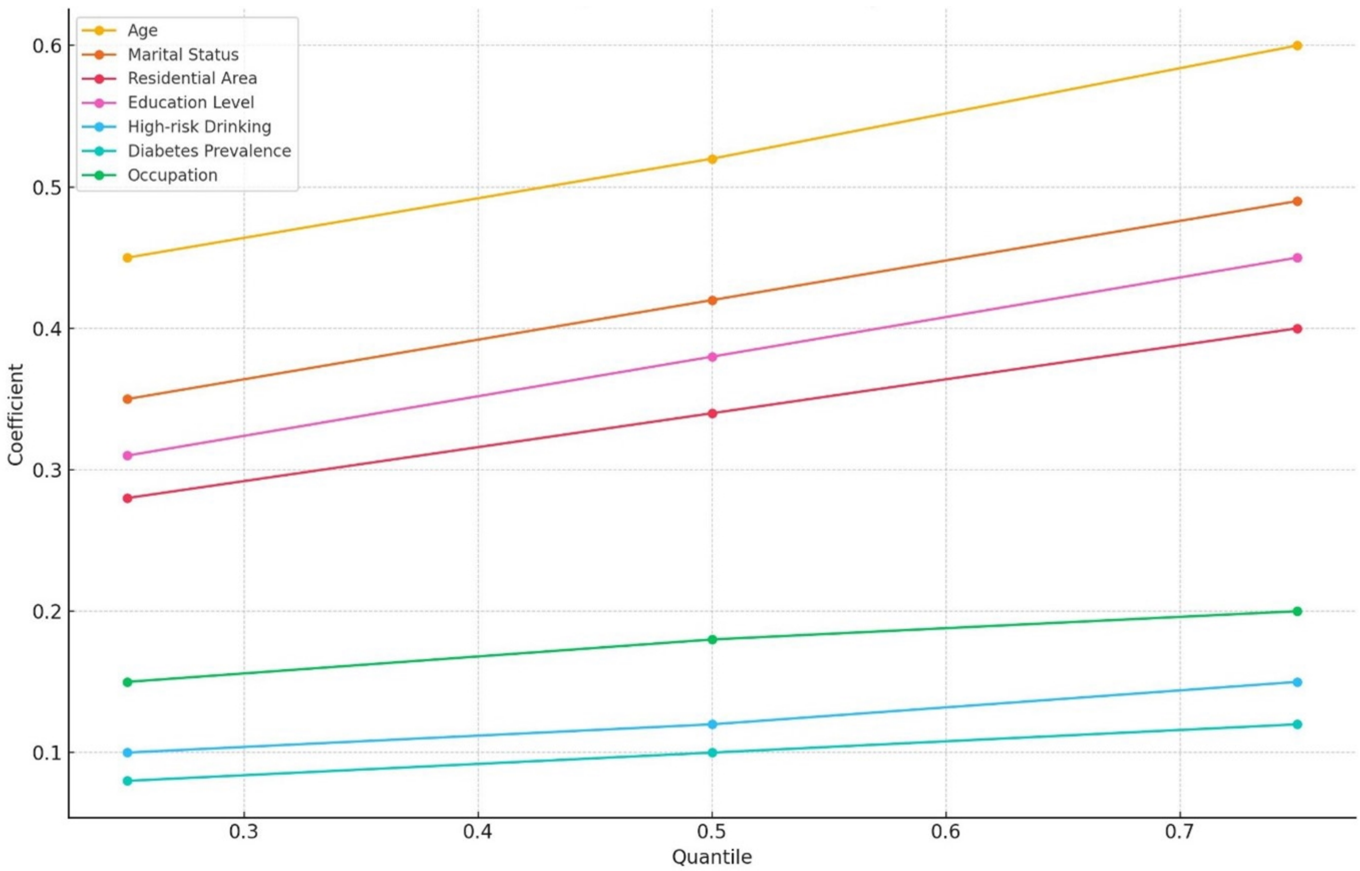


FIGURE 2. Variable importance scores.

TABLE 3. Results of quantile regression analysis.

Variable	Quantile	Coefficient	Standard error	*t*-statistic	*p*-value
Age (yr)					
	0.25	0.45	0.07	6.43	<0.001
	0.50	0.52	0.06	8.67	<0.001
	0.75	0.60	0.05	12.00	<0.001
Marital status					
	0.25	0.35	0.08	4.38	<0.001
	0.50	0.42	0.07	6.00	<0.001
	0.75	0.49	0.06	8.17	<0.001
Residential area					
	0.25	0.28	0.09	3.11	0.002
	0.50	0.34	0.08	4.25	<0.001
	0.75	0.40	0.07	5.71	<0.001
Education level					
	0.25	0.31	0.10	3.10	0.002
	0.50	0.38	0.09	4.22	<0.001
	0.75	0.45	0.08	5.63	<0.001
High-risk drinking					
	0.25	0.10	0.11	0.91	0.364
	0.50	0.12	0.10	1.20	0.230
	0.75	0.15	0.09	1.67	0.096
Diabetes prevalence					
	0.25	0.08	0.12	0.67	0.505
	0.50	0.10	0.11	0.91	0.364
	0.75	0.12	0.10	1.20	0.230
Occupation					
	0.25	0.15	0.13	1.15	0.251
	0.50	0.18	0.12	1.50	0.134
	0.75	0.20	0.11	1.82	0.069



F I G U R E 3. Quantile regression coefficients by variable.

indicate that older individuals have a greater likelihood of recognizing MI symptoms compared to younger individuals, as evidenced by the progressively stronger effect of age at higher quantiles. This trend demonstrates that age is a critical factor, with the effect size increasing from the 25th to the 75th quantile, suggesting that older age groups possess heightened awareness or ability to identify MI symptoms. Marital status also shows a consistent positive association, with married individuals demonstrating higher recognition across all quantiles. The residential area is another significant predictor, with those living in cities or rural areas having higher recognition rates compared to those in metropolitan areas. The effect is more pronounced at higher quantiles, suggesting that the impact of residential areas on MI symptom recognition intensifies with increasing levels of recognition. Education level is consistently associated with higher MI symptom recognition, with individuals having higher educational attainment showing greater recognition across all quantiles. This effect is particularly strong at the 75th quantile, indicating that education plays a crucial role in the highest levels of symptom recognition. High-risk drinking and diabetes prevalence, while not significant at the 25th and 50th quantiles, show a trend towards significance at the 75th quantile. This suggests that these factors may have a more substantial impact on MI symptom recognition among individuals at higher risk levels. Occupation, although not reaching statistical significance, shows a positive trend, indicating that certain occupational groups may have higher recognition rates, particularly at the 75th quantile.

4. Discussion

The primary objective of this study was to identify the key factors influencing the early recognition of myocardial infarction (MI) symptoms among adult male smokers and non-smokers. By leveraging advanced machine learning techniques, specifically Sparse Attention Mechanisms, and comparing them with traditional models such as CART, C4.5 and Rotation Random Forest, this study provides valuable insights into the demographic and behavioral factors that significantly impact MI symptom recognition.

The study's findings highlight several important factors. Age emerged as the most significant predictor of MI symptom recognition across all quantiles. Older individuals demonstrated a higher likelihood of recognizing MI symptoms, with the effect size increasing from the 25th to the 75th quantile. This finding underscores the critical need for targeted educational interventions aimed at older adults, who are at a higher risk of MI [26–28].

One potential mechanism explaining why older individuals exhibit a higher likelihood of recognizing MI symptoms may relate to their cumulative health experiences and increased healthcare interactions over time. As people age, they are more likely to encounter health issues, either personally or through peers, which can heighten their awareness and understanding of medical conditions, including myocardial infarction. Additionally, older adults often have more frequent interactions with healthcare providers, which can lead to better education about cardiovascular risks and symptoms. This increased exposure and interaction may foster a greater familiarity with the symptoms of MI, enhancing their ability to recognize them early. Moreover, cognitive factors, such as increased vigilance or concern for health in older age, could also contribute to this heightened awareness. These insights suggest that age-specific educational strategies could be particularly effective in improving MI symptom recognition among other demographic groups, leveraging the natural advantages observed in older populations.

Marital status was consistently associated with higher MI symptom recognition. Married individuals exhibited greater recognition rates across all quantiles compared to their single counterparts. This suggests that marital support may play a crucial role in enhancing health awareness and symptom recognition. Public health initiatives should consider involving spouses in educational programs to improve MI symptom recognition [28, 29].

The residential area was another significant predictor, with individuals living in cities or rural areas having higher recognition rates compared to those in metropolitan areas. The effect was more pronounced at higher quantiles, indicating that geographic location and associated factors, such as access to healthcare and health education, play a critical role in symptom recognition. Efforts to improve MI symptom recognition should address regional disparities in healthcare access and education [30, 31].

Higher educational attainment was consistently associated with greater MI symptom recognition. This effect was particularly strong at the 75th quantile, highlighting the importance of education in health literacy and awareness. Educational interventions that focus on improving health literacy among individuals with lower educational attainment could significantly enhance MI symptom recognition [32, 33].

While high-risk drinking and diabetes prevalence were not significant predictors at the 25th and 50th quantiles, they showed a trend towards significance at the 75th quantile. This suggests that these factors may have a more substantial impact on MI symptom recognition among individuals at higher risk levels. Public health strategies should include targeted interventions for individuals with high-risk drinking behaviors and diabetes to improve their symptom recognition [34–36].

Although occupation did not reach statistical significance, it exhibited a positive trend, indicating that certain occupational groups may have higher recognition rates, particularly at the 75th quantile. Further research is needed to explore specific occupational categories that contribute to higher MI symptom recognition rates. Occupational health programs could be tailored to address the unique needs of different job categories [36]. These models could be utilized by healthcare providers such as cardiologists and primary care physicians for early detection and intervention, potentially reducing the impact of myocardial infarction on healthcare systems. Furthermore, we have considered the generalizability of our results to other populations, including women, and the application of similar models in countries with diverse healthcare systems. These considerations provide a more comprehensive understanding of the study's impact and future applications.

The findings of this study have significant implications for public health strategies aimed at improving MI symptom recognition. By identifying the most important predictors, targeted educational interventions can be developed to address the specific needs of high-risk groups. For instance, older adults, individuals with lower educational attainment, and those living in metropolitan areas could benefit from tailored educational programs that enhance their ability to recognize MI symptoms early.

Furthermore, the study underscores the importance of considering demographic and behavioral factors in public health initiatives. Interventions that involve spouses, address regional disparities, and focus on high-risk behaviors and chronic conditions can significantly improve MI symptom recognition and ultimately reduce mortality rates associated with cardiovascular diseases.

While this study offers significant insights, it is crucial to recognize its limitations. First, the study relies on self-reported data, which may be subject to recall bias and social desirability bias. Second, the cross-sectional design of the study limits the ability to establish causal relationships between the predictors and MI symptom recognition. Future research should explore longitudinal designs to establish causality and investigate the impact of targeted interventions on MI symptom recognition. Moreover, qualitative studies could provide deeper insights into the barriers and facilitators of MI symptom recognition among different demographic groups. Third, it is important to consider potential uncontrolled factors that may have influenced the results. For instance, the sample size and data collection methods could introduce variability that impacts the generalizability of the findings. Ensuring a representative sample and employing rigorous data collection procedures in future studies will be essential to validate and extend the applicability of these results.

5. Conclusions

This study identified key demographic and behavioral factors influencing the early recognition of myocardial infarction (MI) symptoms among adult male smokers and non-smokers. By leveraging advanced machine learning techniques, the study provides a comprehensive understanding of the variability in predictor effects across different levels of symptom recognition. These findings underscore the significance of targeted educational interventions that address the unique needs of high-risk groups, which will ultimately contribute to improved health outcomes and reduced mortality rates associated with cardiovascular diseases. Future research should explore the application of Sparse Attention Mechanisms Models on tabular datasets to further refine predictive capabilities and enhance the generalizability of findings across diverse populations.

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. The author contributed to editorial changes in the manuscript. The author 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). All study participants provided written informed consent.

ACKNOWLEDGMENT

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

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) and a new professor research program of KOREATECH in 2025.

CONFLICT OF INTEREST

The author declare no conflict of interest. HB is serving as one of the Guest editors of this journal. We declare that HB 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 WYCW.

REFERENCES

[1] Sia CH, Ko J, Zheng H, Ho AFW, Foo D, Foo LL, *et al.* Association between smoking status and outcomes in myocardial infarction patients undergoing percutaneous coronary intervention. Scientific Reports. 2021; 11: 6466.

[2] Duncan MS, Greevy RA, Tindle HA, Vasan RS, Lipworth L, Aldrich MC, *et al.* Inclusion of smoking data in cardiovascular disease risk estimation. JAMA Cardiology. 2022; 7: 195–203.

[3] Biery DW, Berman AN, Singh A, Divakaran S, DeFilippis EM, Collins BL, *et al.* Association of smoking cessation and survival among young adults with myocardial infarction in the partners YOUNG-MI registry. JAMA Network Open. 2020; 3: e209649.

[4] Yaman B, Akpinar O, Cerit L, Kemal HS, Usalp S, Yüksek Ü, *et al.* Effects of chronic cigarette smoking on myocardial deformation parameters by two-dimensional speckle tracking echocardiography. Echocardiography. 2019; 36: 2122–2128.

[5] Klein LW. Systemic and coronary hemodynamic effects of tobacco products on the cardiovascular system and potential pathophysiologic mechanisms. Cardiology in Review. 2022; 30: 188–196.

[6] Muhyieddeen A, Sharif Z, Rasheed O, Ranasinghe S, Chang J, Saleh K. Investigating the smoker's paradox: acute myocardial infarction outcomes in smokers. American Heart Journal. 2024; 250: 97–101.

[7] Luo JG, Yang M, Han L, Gao K, Chen X, Chen LW. A 5-year follow up study on smoking and current smoking cessation status in patients with acute myocardial infarction from a hospital in Xicheng district, Beijing. Chinese Journal of Prevention and Control of Chronic Diseases. 2011; 19: 231–234.

[8] Yusuf S, Hawken S, Ounpuu S, Bautista L, Franzosi MG, Commerford P, *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. The Lancet. 2004; 364: 937–952.

[9] Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. The BMJ. 2004; 328: 1519.

[10] Ford ES, Caspersen CJ. Cardiovascular disease and diabetes: a combination of epidemics. American Journal of Preventive Medicine. 2006; 31: 345–352.

[11] Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, *et al.* Heart disease and stroke statistics—2019 update: a report from the American Heart Association. Circulation. 2019; 139: e56–e528.

[12] Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, *et al.* Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation. 2015; 131: e29–e322.

[13] Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, *et al.* Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. JAMA. 2003; 290: 891–897.

[14] Kannel WB, McGee DL. Diabetes and cardiovascular disease: the Framingham study. JAMA. 1979; 241: 2035–2038.

[15] Huxley R, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. The Lancet. 2011; 378: 1297–1305.

[16] Peters SA, Huxley RR, Woodward M. Smoking as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 81 cohorts, including 3,980,359 individuals and 42,401 strokes. Stroke. 2014; 45: 1024–1031.

[17] Pérez J, Barceló P, Marinkovic J. Attention is Turing-complete. Journal of Machine Learning Research. 2021; 22: 1–35.

[18] Su X, Zhao X, Ren J, Li Y, Rätsch M. Pre-training neural machine translation with alignment information via optimal transport. Multimedia Tools and Applications. 2024; 83: 48377–48397.

[19] Gardazi NM, Daud A, Malik MK, Bukhari A, Alsahfi T, Alshemaimri B. BERT applications in natural language processing: a review. Artificial Intelligence Review. 2025; 58: 166.

[20] Tamura M, Saito K, Yamasaki T. Design and analysis of efficient attention in transformers for social group activity recognition. International Journal of Computer Vision. 2024; 132: 4269–4288.

[21] Koenker R, Bassett Jr G. Regression quantiles. Econometrica. 1978; 46: 33–50.

[22] Koenker R. Quantile regression. 1st edn. Cambridge University Press: United Kingdom. 2005.

[23] Yu HF, Huang FL, Lin CJ. Dual coordinate descent methods for logistic regression and maximum entropy models. Machine Learning. 2011; 85: 41–75.

[24] Zhu H, Hastie T. Kernel logistic regression and the import vector machine. Journal of Computational and Graphical Statistics. 2005; 14: 185–205.

[25] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B. 1996; 58: 267–288.

[26] Bibbins-Domingo K, Pletcher MJ, Lin F, Vittinghoff E, Gardin JM, Arynchyn A, *et al.* Racial differences in incident heart failure among young adults. New England Journal of Medicine. 2009; 360: 1179–1190.

[27] Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, *et al.* Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic impact goal through 2020 and beyond. Circulation. 2010; 121: 586–613.

[28] Dupre ME, Beck AN, Meadows SO. Marital trajectories and mortality

among US adults. American Journal of Epidemiology. 2009; 170: 546–555.

[29] Rohrmann S, Overvad K, Bueno-de-Mesquita HB, Jakobsen MU, Egeberg R, Tjønneland A, *et al*. Meat consumption and mortality—results from the European Prospective Investigation into Cancer and Nutrition. BMC Medicine. 2013; 11: 63.

[30] Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, *et al*. Neighborhood of residence and incidence of coronary heart disease. New England Journal of Medicine. 2001; 345: 99–106.

[31] Rundle A, Neckerman KM, Freeman L, Lovasi GS, Purciel M, Quinn J, *et al*. Neighborhood food environment and walkability predict obesity in New York City. Environmental Health Perspectives. 2009; 117: 442–447.

[32] Ross CE, Wu CL. The links between education and health. American Sociological Review. 1995; 60: 719–745.

[33] Cutler DM, Lleras-Muney A. Education and health: evaluating theories and evidence. 2006. Available at: https://doi.org/10.3386/w12352 (Accessed: 15 July 2024).

[34] Mukamal KJ, Jensen MK, Grønbæk M, Stampfer MJ, Manson JE, Pischon T, *et al*. Drinking frequency, mediating biomarkers, and risk of myocardial infarction in women and men. Circulation. 2005; 112: 1406–1413.

[35] Huxley R, Barzi F, Woodward M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. The BMJ. 2006; 332: 73–78.

[36] Kivimäki M, Nyberg ST, Batty GD, Fransson EI, Heikkilä K, Alfredsson L, *et al*. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. The Lancet. 2012; 380: 1491–1497.

How to cite this article: Haewon Byeon. Identification of early myocardial infarction symptoms in adult male smokers using sparse attention mechanisms and quantile regression: results from 97,304 participants in a nationwide survey in Korea. Journal of Men's Health. 2025; 21(10): 1-10. doi: 10.22514/jomh.2025.121.