BMC Health Services Research

RESEARCH

Open Access

# Machine learning approach for unmet medical needs among middle-aged adults in South Korea: a cross-sectional study

Jeewuan Kim[1†], Seok-Min Ji[2†], In-Sik Kim[3], Ha-Young Jang[3], Chang-Hyun Yoo[3], Jae-Hak Kim[4] and Kyu-Min Kim[1*]

## Abstract

**Background** South Korea is reported to have higher levels of unmet medical needs (UMN) than other countries, particularly among the middle-aged adult population. Considering that this group constitutes a substantial portion of the country's productive workforce, their health requires continuous management to ensure sustained productivity. The purpose of this study is to investigate the factors associated with UMN in economically active middle-aged adults and to develop a model to predict the occurrence of UMN.

**Methods** In this study, 3,575 middle-aged adults who are economically active were selected from the 2020 Korean Health Panel Survey data. Logistic regression, Random Forest, Naïve Bayes, Gradient Boosting Method, and Neural Network were applied to create the prediction model, and tenfold cross validation was performed by checking the reliability of the analysis. The model was evaluated based on the Area Under Receiver Operating Characteristics (AUROC) as well as accuracy, precision, recall, F-1 score and MCC.

**Results** First, the prevalence of UMN in middle-aged adults was 15.6%. Second, random forest was found to be the model with the highest predictive power. It showed an AUROC of 0.831, Accuracy of 0.862, and F-1 score of 0.820. Third, the main factors influencing the occurrence of UMN were subjective stress and subjective health awareness.

**Conclusions** These findings suggest that psychological support is necessary in order to manage the occurrence of UMN among middle-aged adults, with regular stress management being especially important. However, the lower AUROC suggests that additional variables are needed to enhance the prediction model.

**Keywords** Machine learning, Unmet medical needs, Middle-aged adults, Prediction model, Korea health panel survey

[†]Jeewuan Kim and Seok-Min Ji contributed equally to this work.

*Correspondence:
Kyu-Min Kim
perves@gtec.ac.kr
[1]Department of Health Administration, Gyeonggi College of Science and Technology, Gyeonggi, South Korea
[2]Department of Cancer AI and Digital Health, Graduate School of Cancer Science and Policy, National Cancer Center, Gyeonggi, South Korea
[3]Department of Health Policy and Management, Korea University, Seoul, South Korea
[4]Department of Fitness Promotion and Rehabilitation Exercise, National Rehabilitation Center, Seoul, South Korea

## Background

The concept of unmet medical needs (UMN) refers to situations in which individuals require medical treatment or consultation but are unable to access the necessary care, serving as a significant indicator of healthcare service utilization [1]. The persistent occurrence of UMN increases the likelihood of disease onset, which can exacerbate the long-term cost and quality of healthcare costs [2]. In South Korea, all citizens are covered by the National Health Insurance (NHI), which allows them to access medical care at a relatively low cost [3]. In spite of this, South Korea exhibits a relatively high occurrence of UMN when compared with other major countries (the United States: 4.5%, Poland: 4.2%, Greece: 8.1%, and South Korea: 12.0%) [4]. Korean statistics indicate that the prevalence of UMN varies by age group (ages 19–29: 15.1%, ages 30–64: 31.6%, ages 65 and older: 30.0%) [5]. This study focuses on middle-aged adults (ages 30–64), a key demographic that plays a critical role in South Korea's economic development and stability [6]. The proportion of middle-aged adults in South Korea is relatively high compared to other countries—53.5%, compared to 38.5% in the United States, 44% in France, and 44.1% in Japan [7]. Additionally, this age group exhibits a higher rate of economic activity compared to other age groups [8]. Taking into account their importance to the economy, identifying and addressing the healthcare needs of middle-aged adults is a matter of national importance.

UMN is influenced by a complex interplay of various factors, including educational level, residential area, economic activity, chronic diseases, subjective stress, and perceived health status [5, 9-14]. A study on Korean women found that UMN was more likely to occur with lower household income, lower educational attainment, and living outside of a metropolitan area, as well as with lower perceived health status [9]. Research on lymphedema outpatients in France found that lower perceived health status, lower income, and higher out-of-pocket medical expenses were associated with higher rates of UMN [10]. Furthermore, studies conducted among adults in Greece and Iran indicated that UMN is more prevalent among those who are not economically active, have chronic diseases, lack health insurance, or do not engage in regular physical exercise [11, 12]. Specifically, numerous studies in South Korea have consistently found higher stress levels and poorer perceived health status to be related to an increased prevalence of UMN among adults [5, 13, 14].

A review of previous studies reveals that most research has focused on identifying factors influencing UMN among adults, older adults, or patients with specific diseases. Most studies were found to have a limitation in that the existing studies have only identified the factors that affect UMN in an enumerated manner. There is also a lack of research examining UMN in economically active middle-aged adults. Therefore, this study aims to develop a predictive model for UMN occurrence among economically active middle-aged adults using machine learning algorithms. Machine learning, which is a form of supervised learning, provides better accuracy in outcome prediction than traditional methods and has become widely used in healthcare to predict chronic conditions such as hypertension [15]. We intend to provide foundational data for the prevention and management of UMN among middle-aged adults through this study.

## Methods

### Study design and participants

The data for this study were obtained from the Korea Health Panel Survey (KHPS), a nationwide survey currently being conducted. The data are based on the 2005 Census of Population and Housing and include various information related to healthcare utilization, collected using two-stage probability ratio stratified cluster sampling. Moreover, it is the only dataset in the country that provides information on both unmet medical needs related to healthcare access and a variety of factors influencing healthcare expenditures [16]. For the analysis, we utilized the most current version of the data, 2020, which is publicly available from 2008 to 2020. We selected middle-aged adults between the ages of 30 and 64 who are economically active. There were 1,749 individuals with missing data. Specifically, we excluded 255 individuals who did not respond "yes or no" to the question about unmet medical needs and 1,539 individuals who did not respond to input variables such as education level, region, income, and disability. Ultimately, a total of 3,757 individuals were included in the final analysis (Fig. 1).

### Measurements

#### *Target variable*

UMN occurs when an individual or physician determines that they need medical services but are unable to access those services [17]. Although a formal definition of UMN does not yet exist, previous studies primarily rely on UMN as subjectively recognized by individuals [18]. It is a direct measure based on an individual's experience with accessing medical services and, as a result, encompasses aspects related to healthcare quality. In this study, UMN were defined based on individual medical service needs. UMN was assessed using the question, 'In the past year, have you ever needed treatment or an examination at a hospital but did not receive it?' with response options of 'Yes (experienced unmet needs at least once)' and 'No (did not experience unmet needs).' Those who responded 'Yes' were classified as having UMN.

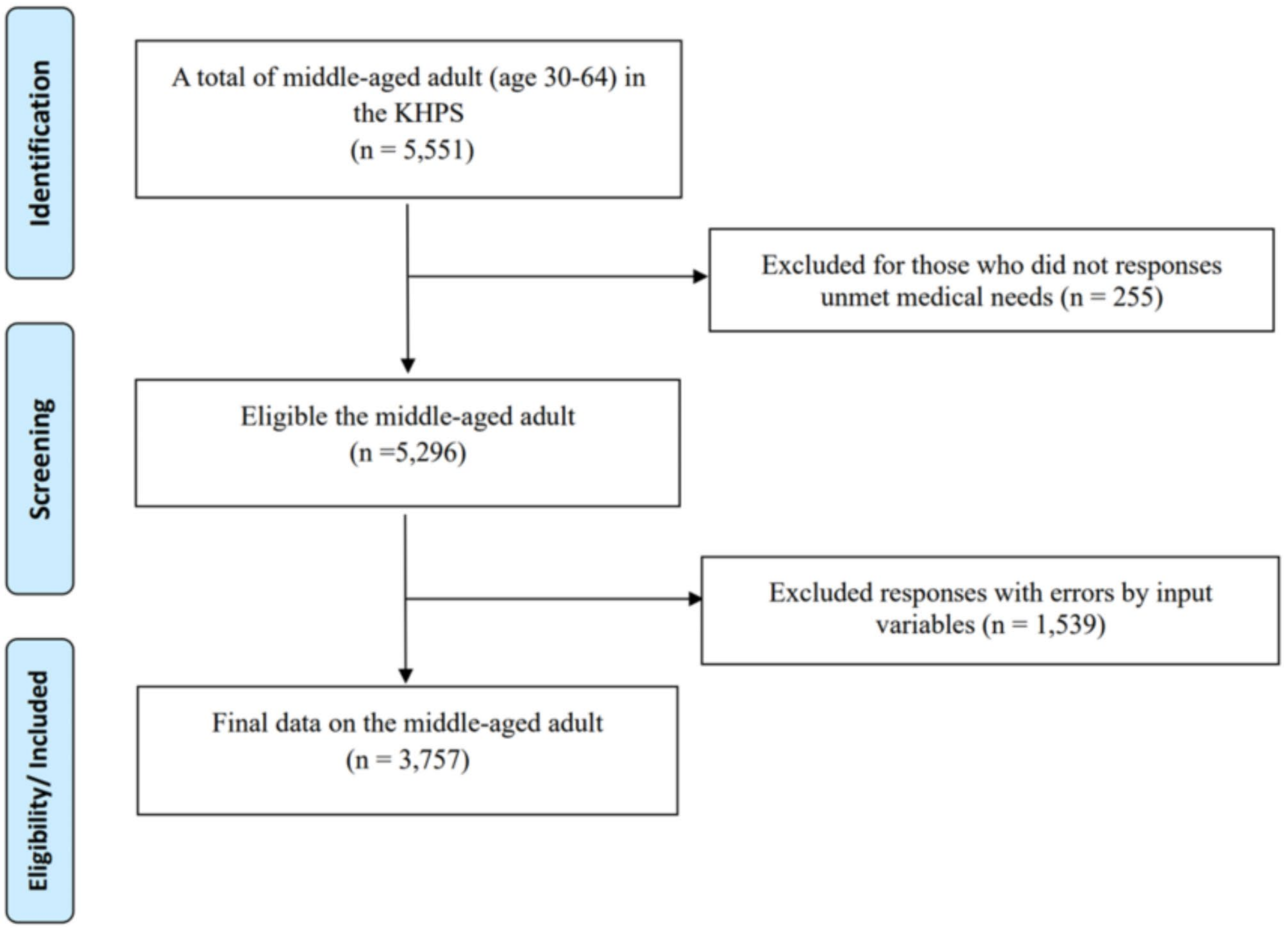


**Fig. 1** Flow diagram of the participants selection

### *Input variables*

Sex was categorized as male (0) and female (1). Age was classified as 30–49 (0), 50–64 (1). Since many previous studies have classified middle-aged adults (30–49) and late middle-aged adults (50–64) in South Korea, this study was also conducted in the same way [19, 20]. while education level was classified into four groups: elementary school or less (0), middle school (1), high school (2), and college or higher (3). Regions were grouped as metropolitan (0) for Seoul, Gyeonggi-do, and Incheon, and non-metropolitan (1) for the rest of the country. Economic conditions were categorized into employee (0) and self-employed (1) based on the 'work type in which they are currently working'. Household size was classified as one member (0), two members (1), and three or more members (2). Income was categorized into quintiles, ranging from the 1st quintile (low income) to the 5th quintile (high income), by dividing total annual household income by the square root of the actual number of household members [21]. Disability was defined as having a diagnosis from a doctor and categorized as no (0) or yes (1). Chronic disease was defined as taking medication for more than 6 months and was also categorized as no (0) or yes (1) [22]. Regular physical exercise was categorized as no (0) or yes (1) based on the question, 'In the past year, have you played sports or exercised regularly, including walking?'. Drinking status was classified as not drinking less than once a month based on 'the frequency of drinking in the last year'. Smoking status was categorized as non-smoker (0) or smoker (1), depending on whether the respondent currently smokes or not. A 5-point Likert scale is used to assess subjective health awareness, with higher scores indicating a higher level of awareness. Subjective stress awareness was assessed using a 4-point Likert scale ('I feel very much, I feel much, I feel a little, I feel nearly none'), with higher scores indicating greater levels of stress, based on responses to the question, 'How do you feel about stress in your daily life?'

### Statistical analysis

The analysis methods used in this study include frequency analysis, chi-square test, fisher test, Shapley additive explanatory analysis, and machine learning techniques. First, a frequency analysis and chi-square test were conducted in order to evaluate whether statistical differences occurred based on the demographic characteristics of the study participants and the presence of UMN. Second, Shapley additive explanatory analysis was used to evaluate the relative importance of contributing predictors in predicting UMN among middle-aged adults. Third, we used Neural networks (NNs), Logistic regression (LR), Naïve bayes (NB), Gradient boosting method (GBM), and Random Forest (RF) to develop a prediction model for the occurrence of UMN in middle-aged adults. NNs are nonlinear models and are widely used for predicting categories of target factors [23]. The LR model has the advantage of showing high predictive power when the target factor is categorical [24]. NB identifies causal relationships between independent and dependent variables by applying Bayesian statistics to represent probabilistic relationships between variables [25]. GBM has been reported to have a low probability of overfitting and a high predictive performance when combined with a variety of algorithms, such as decision trees, SVMs, and neural networks based on boosting [26]. RF has been reported to have relatively high predictive power and model stability when performed on data that contains a large number of input variables [27]. To develop and evaluate the predictive model, the k-fold cross validation method was applied to ensure the reliability of the results. We adopted the tenfold cross validation in which the original sample is randomly partitioned into 10 subsamples of equal size. A single subsample is retained as the validation data for testing the model, and the remaining nine subsamples are used as training data. And then the cross validation process is repeated ten times with each of the ten subsamples employed exactly once for the validation. This model was evaluated according to its Area Under the Receiver Operating Characteristic (AUROC) as well as its accuracy, precision, recall, and F- 1 score. The analysis was conducted using R and SAS Enterprise Miner.

## Results

### Sociodemographic and UMN characteristics of participants

Table 1 shows the demographic characteristics of the participants and their differences in UMN. Based on the demographic characteristics, 2,178 participants (58.0%) were categorized as male and 1,579 participants (42.0%) as female, indicating that males accounted for a greater proportion than females. Age was the most common, with 1,925 respondents (51.3%) being classified as "50–64". The most common educational background is "college or higher" with 1,922 people (51.2%), representing a relatively high educational attainment. The 'Non-metropolitan' region showed a large distribution, with 2,643 people (70.3%). Economic conditions were the most common, with 2,837 respondents (75.5%) being classified as "Employee". In terms of household size, '3 or more people' was the most common, reported by 2,510 people (66.8%), followed by '2 people' with 938 people (25.0%), and '1 person' with 309 people (8.2%). According to income levels, 1,260 people (33.5%) were in the 5 th quintile, and 1,042 people (27.7%) were in the 4 th quintile, indicating that the majority of participants had a high level of income. There were 3,659 respondents for disability (2.6%) and 2,440 respondents for chronic disease (64.9%), indicating that the majority of participants did not have a disability or chronic illness. 2,023 people (53.8%) responded 'No' to the question regarding regular physical exercise while 2,982 people (79.4%) responded 'Yes' to the question regarding regular alcohol consumption, indicating that it is the majority of participants who consume alcohol on a regular basis. In terms of smoking status, 'Non-smoker' was the most common, reported by 2,767 people (73.7%). Finally, we examined the differences between the demographic characteristics of the study subjects and the occurrence of UMN finding significant differences in the following variables: Age ($X^2$= 15.900 $p$<.001), Educational level ($X^2$= 9.259 $p$<.005), Region ($X^2$= 4.662 $p$<.005), Regular physical exercise ($X^2$= 17.133 $p$<.001), Smoking status ($X^2$= 2.863 $p$<.005).

### Continuous variables characteristic of participants

The descriptive statistics for continuous variables are shown in (Table 2). The average score on the subjective stress awareness scale, which ranges from 1 to 4, was 3.40, reflecting relatively high levels of stress. This suggests that the middle-aged adults who participated in this study generally experience elevated levels of stress. The average subjective health awareness score is 2.20, indicating relatively good health, with a standard deviation of 0.82, suggesting minimal individual variation.

### The significance of the variables impacting UMN

SHAP is used to assess the contribution of independent factors, particularly feature relative importance, to explore the various implications of independent variables on charging classes [28]. Table 3 shows the relative importance of predictors in predicting the UMN of middle-aged adults engaged in economic activities, which is the focus of this study, using SHAP values. The higher the importance ranking, the greater the influence of a predictive factor on predicting UMN. In this study, the variable with the highest importance ranking was subjective stress awareness. As such, subjective stress awareness is more significant than other factors in predicting UMN in middle-aged adults. Following subjective stress awareness, the factors of subjective health awareness, regular physical exercise, and age ranked among the most influential predictors. In contrast, educational level, sex and income were ranked lower in importance (Fig. 2).

**Table 1** General characteristics of the participants and differences in UMN

| Variables | | *N* (%) | UMN (Yes) *N* (%) | $X^2$ |
|---|---|---|---|---|
| Sex | Male | 2,178 (58.0) | 335 (8.9) | 0.232 |
| | Female | 1,579 (42.0) | 252 (6.7) | |
| Age | 30–49 | 1,832(48.7) | 326(8.6) | 15.900*** |
| | 50–64 | 1,925(51.3) | 261(7.1) | |
| Educational level | Elementary school or lower | 165 (4.4) | 37 (0.9) | 9.259* |
| | Middle school | 323 (8.6) | 43 (1.1) | |
| | High school | 1,347 (35.9) | 194 (5.2) | |
| | College or higher | 1,922 (51.2) | 313 (8.4) | |
| Region | Metropolitan | 1,114 (29.7) | 196 (5.2) | 4.662* |
| | Non-metropolitan | 2,643 (70.3) | 391 (10.4) | |
| Economic conditions | Employee | 2,837(75.5) | 457(12.1) | 0.983 |
| | Self-employed | 920(24.5) | 130(3.5) | |
| Household size | 1 | 309 (8.2) | 44 (1.2) | 4.809 |
| | 2 | 938 (25.0) | 128 (3.4) | |
| | 3 or more | 2,510 66.8) | 415 (11.0) | |
| Income | 1 | 170 (4.5) | 29 (0.8) | 1.130 |
| | 2 | 468 (12.5) | 71 (1.9) | |
| | 3 | 817 (21.7) | 121 (3.2) | |
| | 4 | 1,042 (27.7) | 170 (4.5) | |
| | 5 | 1,260 (33.5) | 196 (5.2) | |
| Disability | Yes | 98 (2.6) | 18 (0.5) | 0.574 |
| | No | 3,659 (97.4) | 569 (15.1) | |
| Chronic disease | Yes | 1,317 (35.1) | 215 (5.7) | 0.755 |
| | No | 2,440 (64.9) | 372 (9.9) | |
| Regular physical exercise | Yes | 1,734 (46.2) | 225 (6.0) | 17.133*** |
| | No | 2,023 (53.8) | 362 (9.6) | |
| Drinking status | Yes | 2,982 (79.4) | 483 (12.8) | 3.601 |
| | No | 775 (20.6) | 104 (2.8) | |
| Smoking status | Smoker | 990(26.3) | 170(4.5) | 2.863* |
| | Non-smoker | 2,767(73.7) | 417(11.1) | |
| Total | | 3,757 (100.0) | 587 (15.6) | |

$P^* < 0.05$, $p^{***} < 0.001$

**Table 2** Descriptive analysis of continuous variables

| Variables | Mean | Standard deviation | Range |
|---|---|---|---|
| Subjective stress awareness | 3.40 | 0.78 | 1–4 |
| Subjective health awareness | 2.20 | 0.82 | 1–5 |

**Table 3** The significance of the variables that impact the UMN

| Variables | Ranking | Variables | Ranking |
|---|---|---|---|
| Subjective stress awareness | 1 | Drinking status | 8 |
| Subjective health awareness | 2 | Smoking status | 9 |
| Regular physical exercise | 3 | Chronic disease | 10 |
| Age | 4 | Disability | 11 |
| Region | 5 | Educational level | 12 |
| Economic conditions | 6 | Sex | 13 |
| Number of household members | 7 | Income | 14 |

### Comparison of model parameters for predicting UMN

We applied LR, GBM, NB, NN, and RF algorithms to determine which machine learning model had the highest predictive power for predicting UMN in middle-aged adults. Table 4 shows the AUROC, Accuracy, Precision, Recall, F- 1 score and Matthews correlation coefficient(MCC) for each model. Based on the AUROC values, the RF(0.831) algorithm demonstrated the highest performance in comparison with the other four algorithms. The AUROC results by algorithm were: GBM =0.728, NN =0.655, LR =0.651, and NB =0.537. Even when examining Accuracy, we find that the LR algorithm has the highest performance value (0.862). As a result, the RF algorithm provided the best predictions of UMN among South Korean middle-aged adults (Fig. 3).

## Discussion

In this study, our objective was to assess whether specific factors are associated with the occurrence of UMN among middle-aged adults and to identify the most effective model for predicting its prevalence based on these factors. Our findings revealed that the RF algorithm provided the highest accuracy in predicting UMN.

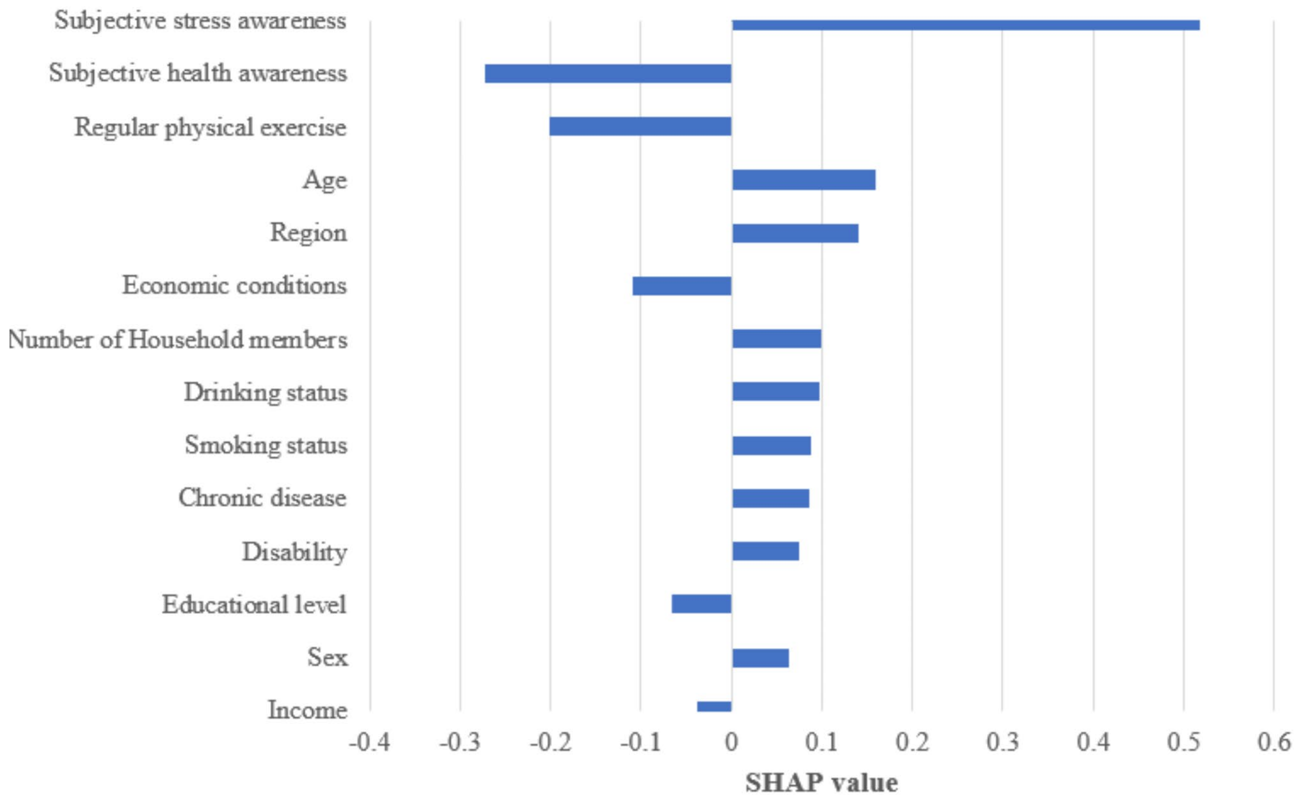


**Fig. 2** The summary plot of the SHAP values

**Table 4** Comparison of the parameters of models for prediction of UMN

| | AUROC | Accuracy | Precision | Recall | F- 1 score | MCC |
|---|---|---|---|---|---|---|
| RF | 0.831 | 0.862 | 0.860 | 0.862 | 0.820 | 0.334 |
| GBM | 0.728 | 0.851 | 0.853 | 0.851 | 0.795 | 0.236 |
| NN | 0.655 | 0.840 | 0.789 | 0.840 | 0.783 | 0.125 |
| LR | 0.651 | 0.841 | 0.866 | 0.841 | 0.768 | 0.039 |
| NB | 0.537 | 0.797 | 0.741 | 0.797 | 0.764 | 0.033 |

*AUROC* Area under the receiver operating characteristic, *MCC* Matthews correlation coefficient, *RF* Random forest, *GBM* Gradient boosting method, *NN* Neural network, *LR* Logistic regression, *NB* Navie bayes

Especially, subjective stress awareness and subjective health awareness were identified as the primary factors influencing UMN among middle-aged adults.

First, the prevalence of UMN among middle-aged adults in this study was 15.6%, which is notably higher than the 11.6% prevalence reported among general adults in Korea [29]. In comparison, other studies on South Koreans with chronic diseases found a lower UMN prevalence of 8.7% [30]. In international studies, a study conducted in Greece reported a UMN prevalence of 9.0% [11], while research on adults in Iran found an even lower rate of 4.0% [12]. These findings highlight a significant variation in UMN prevalence across different populations and health conditions. In the case of South Korea, office workers work an average of 8 h a day (9 am and 6 pm), so it is not practical to visit a hospital on weekdays. Middle-aged adults are a key segment of the economically active population and are among the most skilled workers. In order to maintain both individual well-being and economic stability, it is crucial to proactively manage and mitigate the potential occurrence of UMN in this group.

Second, we identified models of UMN occurrence in middle-aged adults and found that the RF algorithm had the highest predictive power, consistent with previous studies [30-32]. A study predicting cancer metastasis and death in breast cancer patients who received neoadjuvant chemotherapy found that a random forest model had the best predictive power [30]. Similar to this, a study predicting acute respiratory distress syndrome in patients undergoing cardiac surgery also reported a very high predictive power (AUC: 0.932) of the random forest model [31]. In addition, a study comparing decision trees, logistic regression, neural networks, and random forest models for predicting depression in older adults found that

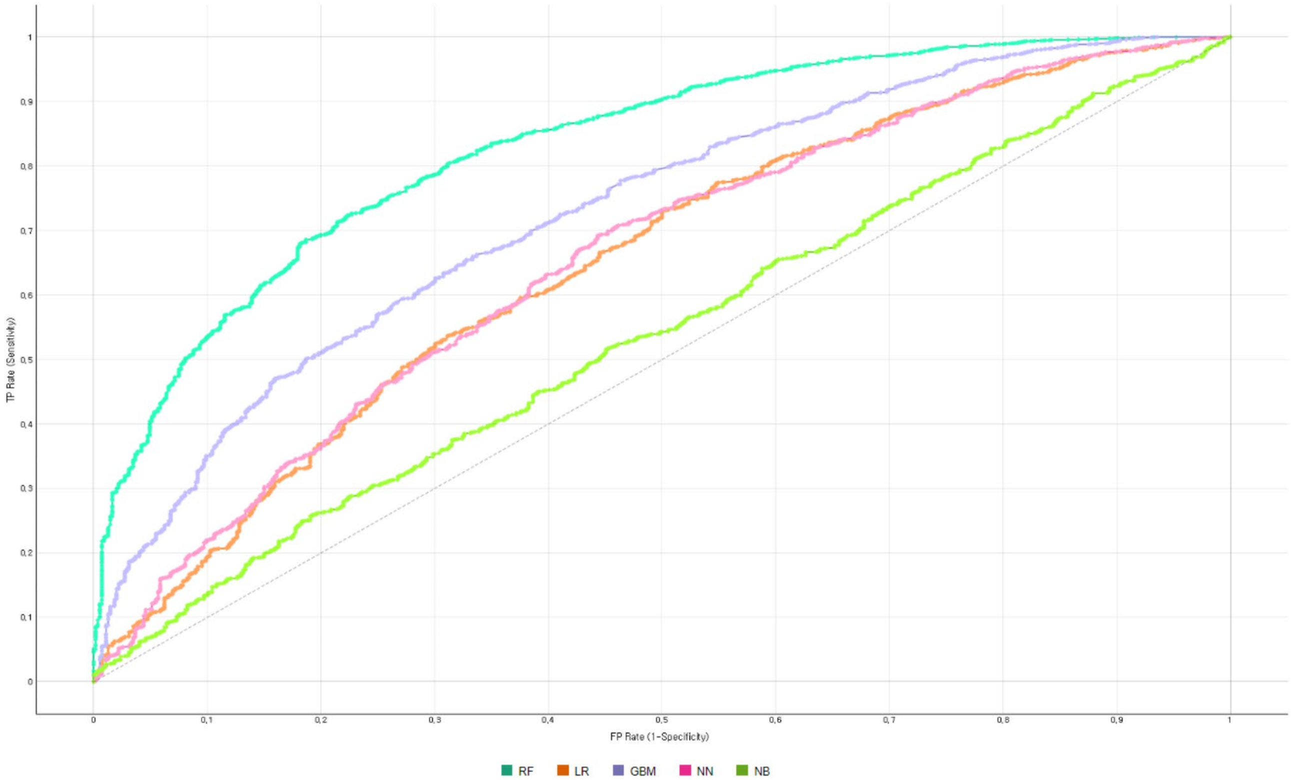


**Fig. 3** The AUROC of UMN prediction models using various machine learning algorithms

decision trees and random forests were the top predictors [32]. Based on the results of these studies, RF is particularly effective at dealing with complex relationships between variables, and it is widely used in healthcare [33]. However, the AUROC value was judged to be relatively low. The low AUROC indicates that variables in the current study were insufficient to adequately predict the UMN. Various factors, such as job position, total work period, commute time, sleep hours and interpersonal relationships, can also affect UMN [34, 35].

Third, factors influencing the occurrence of UMN were assessed, with subjective stress awareness and subjective health awareness being the most significant predictors. Specifically, individuals with higher subjective stress awareness and lower subjective health awareness were more likely to experience UMN. These results are consistent with previous studies [9, 11, 13, 34]. As one example, a study of female baby boomers (born between 1955 and 1963) concluded that high levels of stress are associated with an increased risk of developing UMN [13]. Research on older women in South Korea discovered that lower subjective health awareness, lower levels of education, and lower household income were related to an increased risk for UMN [9]. Furthermore, a study conducted on Korean adults with chronic diseases found that higher stress levels and lower subjective health awareness were both associated with an increased incidence of UMN [34]. The key factors can be considered important intervention targets in order to decrease the prevalence of UMN. The results shown in (Table 2) indicate that participants' subjective levels of stress awareness were very high, with a mean score of 3.40 (range 1–4). These findings highlight the need for managing stress on a regular basis. A potential approach is to integrate stress management into existing public health programs. Representatively, the National Cancer Screening Program implemented by the country can be used. Stress levels can be checked regularly by including stress measurement in the examination items. And the identified high-risk stress group is secondarily linked. In the second case, it is classified and systematically managed by the presence or absence of the Employee Assistance Program in the company where the person is attending. If there is an Employee Assistance Program (EAP), the company provides support for self-management by linking high-risk groups of stress with EAP. In addition, the status of them should be reported to the nearby public health center to continuously monitor quality control.

If there is no EAP, a social network is established so that it can be continuously managed in connection with community resources through accurate diagnosis in connection with primary health care institutions close to the residence. Typically, it will be possible to manage regularly if an additional team is formed for middle-aged adults within a local program operated for the health care of the Vulnerable group (older adults, low-income households). In the long term, support is needed to operate EAP programs in all workplaces, and it will be necessary to organize a system so that quality management can be maintained through public health centers. It has been demonstrated that workplace wellness programs are effective in increasing employees' subjective perceptions of health and reducing stress [35]. It has also been shown that workplace stress reduction programs are effective in reducing stress and depression [36].

Finally, there are several limitations to this study. First, we did not examine various factors that contribute to UMN occurrence among middle-aged adults. A number of factors have been considered in previous studies, however due to data limitations, we were not able to include all of them. Second, we were not able to differentiate between occupational groups. The purpose of this study was to investigate the likelihood of developing UMN among middle-aged adults who are economically active, and it is reported that the likelihood differs depending on the occupation, such as white-collar, blue-collar, and self-employed. There is a need for further investigation into the different occupational groups. Third, there was no identification of the specific causes of UMN in this study. There are three general types of causes for UMNs: availability, accessibility, and acceptability [37]. A future study should identify the specific causes and compare the impact of UMNs based on these causes. Fourth, the AUROC in this study was lower than that in previous studies, which suggests that the features utilized in this study may be inadequate to sufficiently interpret the reasons for UMN. Also, A relatively disproportionate number of UMNs may be the cause of the low AUROC. Fifth, this study was conducted on the South Korean population, so it is difficult to generalize the results to all populations. Nevertheless, this study is significant in that it examined the relationship between various factors affecting the occurrence of UMN in economically active middle-aged adults and developed a model for predicting UMN risk characteristics for the study population.

## Conclusions

This study focused on identifying factors related to the UMN occurrence in economically active middle-aged adults, as highlighted in previous research, and developed a predictive model based on these factors. The LR model was found to be the most accurate, with subjective stress awareness and subjective health awareness emerging as key risk factors. In order to address UMN occurrence in middle-aged adults, supporting policies emphasizing regular stress management are crucial. Moreover, as noted in the study's limitations, the factors influencing the occurrence of UMNs may differ according to occupation. To better understand how UMN occurrence differs according to occupation and its influencing factors, comparative analyses that separate occupational groups are necessary.

### Abbreviations

| | |
|---|---|
| UMN | Unmet medical needs |
| NHI | National health insurance |
| KHPS | Korea health panel survey |
| NN | Neural network |
| LR | Logistic regression |
| NB | Naïve bayes |
| GBM | Gradient boosting method |
| RF | Random forest |
| AUROC | Area under the receiver operating characteristics |
| MCC | Matthews correlation coefficient |
| SHAP values | Shapley additive explanation analysis |


### Acknowledgements

Not applicable.


### Authors' contributions

KMK had full access to all the data used in this study and takes responsibility for the integrity of the data and accuracy of the data analysis. Study concept and design: JWK, SMJ, KMK, Data acquisition: ISK, KMK, Statistical analysis: JWK, SMJ, ISK, KMK, Interpretation of the results: JWK, SMJ, ISK, HYK, CHY, JHK, KMK, Manuscript drafting: JWK, ISK, KMK. All the authors have read and approved the final version of the manuscript.


### Funding

Not applicable.


### Data availability

The data will be made available for special purposes only upon request to the corresponding authors.

## Declarations

### Ethics approval and consent to participate

This study was approved by the Korea national institute for bioethics policy Institutional Review Board (IRB No. IRB- 202410 -01 -041). The IRB of Korea University waived informed consent since this study was retrospective and blinding of the personal information in the data was performed. This data is publicly accessible and written informed consent is obtained from all the participants before participating in the survey. Respondents' information was completely anonymized for use for research purposes and unidentified prior to analysis. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 14 February 2025 / Accepted: 15 April 2025
Published online: 13 May 2025

**References**

1. Donabedian A. Aspects of medical care administration: specifying requirements for health care. Cambridge: Harvard University Press; 1973.
2. Aday LA, Andersen R. A framework for the study of access to medical care. Health Serv Res. 1974;9(3):208–20.
3. Kim YS, Lee J, Moon Y, Kim KJ, Lee K, Choi J, Han SH. Unmet healthcare needs of elderly people in Korea. BMC Geriatr. 2018;18(1):98. https://doi.org/10.1186/s12877-018-0786-3.
4. OECD library. Health immediately. 2021. https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2021_13aff239-en;2021. Accessed 25 June 2024.
5. Korea Health Panel Survey. Unmet medical needs. https://www.khp.re.kr:444/web/research/board/surveyresult_view.do?&bbsid=60&seq=11;2021. Accessed 25 June 2024.
6. Venkataramani AS, O'Brien R, Whitehorn GL, Tsai AC. Economic influences on population health in the united States: toward policymaking driven by data and evidence. PLoS Med. 2020;17(9):e1003319. https://doi.org/10.1371/journal.pmed.1003319.
7. OECD library. Perspectives on Decentralisation and Rural-Urban Linkages in Korea. https://www.oecd-ilibrary.org/urban-rural-and-regional-development/population-pyramid-by-age-gender-and-place-of-residency-korea-2017_92b5020c-en;2021. Accessed 25 June 2024.
8. Statistics Korea. Economically Active Population Survey. hppts://www.kostat.go.kr/board.es?mid=a10301030200&bid=210;2024. Accessed 25 June 2024.
9. Choi JA, Kim OS, Factors Influencing Unmet Healthcare Needs among Older Korean Women. Int J Environ Res Public Health. 2021;18(13):6862. https://doi.org/10.3390/ijerph18136862.
10. Mercier G, Pastor J, Clément V, Rodts U, Moffat C, Quéré I. Out-of-pocket payments, vertical equity and unmet medical needs in France: A National multicenter prospective study on lymphedema. PLoS ONE. 2019;14(5):e0216386. https://doi.org/10.1371/journal.pone.0216386.
11. Pierrakos G, Goula A, Latsou D. Predictors of unmet healthcare needs during economic and health crisis in Greece. Int J Environ Res Public Health. 2023;20(19):6840. https://doi.org/10.3390/ijerph20196840.
12. Connolly S, Wren MA. Unmet healthcare needs in Ireland: analysis using the EU-SILC survey. Health policy. (Amsterdam Netherlands). 2017;121(4):434–41. https://doi.org/10.1016/j.healthpol.2017.02.009.
13. Park MJ, Chung MY. Factors affecting unmet healthcare needs in female baby boomers: Andersen model application in Korea. PloS one. 2023;18(6). https://doi.org/10.1371/journal.pone.0286425.
14. Jung B, Ha IH. Determining the reasons for unmet healthcare needs in South Korea: a secondary data analysis. Health Qual Life Outcomes. 2021;19(1):99. https://doi.org/10.1186/s12955-021-01737-5.
15. El-Hasnony IM, Elzeki OM, Alshehri A, Salem H. Multi-Label active learning-Based machine learning model for heart disease prediction. Sensors. 2022;22(3):1184. https://doi.org/10.3390/s22031184.
16. Kim KM, Kim CY. The association between types of usual source of care and unmet health care needs. Health Soc Sci. 2020;53(1):105–28.
17. Ha R, Jung-Choi K, Kim CY. Employment status and Self-Reported unmet healthcare needs among South Korean employees. Int J Environ Res Public Health. 2018;16(1):9. https://doi.org/10.3390/ijerph16010009.
18. Allin S, Grignon M, Le Grand J. Subjective unmet need and utilization of health care services in Canada: what are the equity implications? Soc Sci Med. 2010;70(3):465–72. https://doi.org/10.1016/j.socscimed.2009.10.027.
19. Cha KW. Factors affecting purchase decision and premiums of private health insurance. Financial Plann Rev. 2022;15(1):1–32.
20. Jia RH, Choi YG. The impact of foreign workers on native employment across different age groups in the Korean labor market. Int Area Stud Rev. 2024;28(4):99–118.
21. Dudel C, Garbuszus JM, Schmied J. Assessing differences in household needs: a comparison of approaches for the Estimation of equivalence scales using German expenditure data. Emp Econ. 2021;60(4):1629–59.
22. Boscart V, Crutchlow LE, Taucar S, Johnson L, Heyer K, Davey M, Costa M, Heckman AP. Chronic disease management models in nursing homes: a scoping review. BMJ Open. 2020;10(2):e032316. https://doi.org/10.1136/bmjopen-2019-032316.
23. Du C, Liu PX, Zheng M. Classification of Imbalanced Electrocardiosignal Data using Convolutional Neural Network. Comput Methods Prog Biomed. 2022;214:106483. https://doi.org/10.1016/j.cmpb.2021.106483.
24. Monti K, Bachi K, Gray M, Mahajan V, Sweeney G, Oprescu AM, Munjal KG, Hurd YL, Lim S. Data mining-based clinical profiles of substance use-related emergency department utilizers. Am J Emerg Med. 2022;53:104–11. https://doi.org/10.1016/j.ajem.2021.12.059.
25. Nopour R, Kazemi-Arpanahi H, Shanbehzadeh M, Azizifar A. Performance analysis of data mining algorithms for diagnosing COVID-19. J Educ Health Promotion. 2021;10:405. https://doi.org/10.4103/jehp.jehp_138_21.
26. Li W, Huang G, Tang N, Lu P, Jiang L, Lv J, Qin Y, Lin Y, Xu F, Lei D. Effects of heavy metal exposure on hypertension: A machine learning modeling approach. Chemosphere. 2023;337:139435. https://doi.org/10.1016/j.chemosphere.2023.139435.
27. Wallace ML, Mentch L, Wheeler BJ, Tapia AL, Richards M, Zhou S, Yi L, Redline S, Buysse DJ. Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction. BMC Med Res Methodol. 2023;23(1):144. https://doi.org/10.1186/s12874-023-01965-x.
28. Ullah I, Liu K, Yamamoto T, Zahid M, Jamal A. Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction. Travel Behav Soc. 2023;31:78–92.
29. Yoon YS, Jung B, Kim D, Ha IH. Factors underlying unmet medical needs: a cross-sectional study. Int J Environ Res Public Health. 2019;16(13):2391.
30. Jin Y, Lan A, Dai Y, Jiang L, Liu S. Development and testing of a random forest-based machine learning model for predicting events among breast cancer patients with a poor response to neoadjuvant chemotherapy. Eur J Med Res. 2023;30(1):394. https://doi.org/10.1186/s40001-023-01361-7.
31. Wang W, Li L, Gu H, Chen Y, Zhen Y, Dong Z. Random Forest-Based prediction of acute respiratory distress syndrome in patients undergoing cardiac surgery. Heart Surg Forum. 2022;30(6):E854–9. https://doi.org/10.1532/hsf.5113.
32. Kim KM, Kim JH, Rhee HS, Youn BY. Development of a prediction model for the depression level of the elderly in low-income households: using decision trees, logistic regression, neural networks, and random forest. Sci Rep 2023'16;13(1):11473. https://doi.org/10.1038/s41598-023-38742-1
33. Kim S, Kosorok MR, Arbeeva L, Schwartz TA, Callahan LF, Golightly YM, Nelson AE, Allen KD. Precision Medicine-Based machine learning analyses to explore optimal exercise therapies for individuals with knee osteoarthritis: random Forest-Informed Tree-Based learning. J Rheumatol. 2023;50(10):1341–5. https://doi.org/10.3899/jrheum.2022-1039.
34. Kim JY, Yang Y. Factors affecting unmet medical needs of patients with diabetes: A population-based study. Nurs Open. 2023;10(10):6845–55. https://doi.org/10.1002/nop2.1933.
35. Reif J, Chan D, Jones D, Payne L, Molitor D. Effects of a workplace wellness program on employee health, health beliefs, and medical use: A randomized clinical trial. JAMA Intern Med. 2020;180(7):952–60. https://doi.org/10.1001/jamainternmed.2020.1321.
36. Lacerda SS, Little SW, Kozasa EH. A stress reduction program adapted for the work environment: A randomized controlled trial with a Follow-Up. Front Psychol. 2018;9:668. https://doi.org/10.3389/fpsyg.2018.00668.
37. Detollenaere J, Hanssens L, Vyncke V, De Maeseneer J, Willems S. Do we reap what we sow?? Exploring the association between the strength of European primary healthcare systems and inequity in unmet need. PLoS ONE. 2017;12(1):e0169274. https://doi.org/10.1371/journal.pone.0169274.

## Publisher's Note