Orhan and Kurutkan *BMC Health Services Research* (2025) 25:366
https://doi.org/10.1186/s12913-025-12502-5

BMC Health Services Research

RESEARCH

Open Access

Predicting total healthcare demand using machine learning: separate and combined analysis of predisposing, enabling, and need factors

Fatih Orhan[1*] and Mehmet Nurullah Kurutkan[2]

Abstract

Objective Predicting healthcare demand is essential for effective resource allocation and planning. This study applies Andersen's Behavioral Model of Health Services Use, focusing on predisposing, enabling, and need factors, using data from the 2022 Turkey Health Survey by TUIK. Machine learning methods provide a powerful approach to analyze these factors and their combined impact on healthcare utilization, offering valuable insights for health policy.

Methods Seven different machine learning models—Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, XGBoost, and Gradient Boosting—were utilized. Feature selection was conducted to identify the most significant factors influencing healthcare demand. The models were evaluated for accuracy and generalization ability using performance metrics such as recall, precision, F1 score, and ROC AUC.

Results The study identified key features affecting healthcare demand. For predisposing factors, gender, educational level, and age group were significant. Enabling factors included treatment costs, community interest, and payment difficulties. Need factors were influenced by smoking status, chronic diseases, and overall health status. The models demonstrated high recall (approximately 0.90) and strong F1 scores (ranging from 0.87 to 0.88), indicating a balanced performance between precision and recall. Among the models, Gradient Boosting, XGBoost, and Logistic Regression consistently outperformed others, achieving the highest predictive accuracy. Random Forest and SVM also performed well, showing robust classification capability.

Conclusions The findings highlight the effectiveness of machine learning methods in predicting healthcare demand, providing valuable insights for health policy and resource allocation. Gradient Boosting, XGBoost, and Logistic Regression emerged as the most reliable models, demonstrating superior generalization and classification performance. Understanding the separate and combined effects of predisposing, enabling, and need factors on healthcare demand can contribute to more efficient and data-driven healthcare planning, facilitating strategic decision-making in resource allocation and service delivery.

*Correspondence:
Fatih Orhan
fatih.orhan@sbu.edu.tr

Full list of author information is available at the end of the article

Keywords Healthcare demand, Andersen behavioral model, Machine learning, Feature selection, Health services utilization, Predictive modeling

Introduction

The demand for healthcare services is a complex phenomenon that directly affects individuals' health status and the efficiency of health systems. Accurately predicting this demand is critically important for shaping health policies and ensuring the efficient use of health resources. Today, machine learning techniques are increasingly used to extract meaningful insights and make predictions from large datasets [1, 2]. This study aims to model healthcare demand using microdata from the 2022 Turkey Health Survey collected by the Turkish Statistical Institute.

Andersen's Behavioral Model of Health Services use forms the theoretical framework for this study [3, 4]. The model identifies three main factors influencing the use of health services: predisposing factors (demographic and social structure), enabling factors (access to health services), and perceived need factors (individuals' perception of their need for health services). This study analyzes the separate and combined effects of these three factors on healthcare demand using machine learning methods.

The analysis will be conducted using various machine learning models, and the performance of these models will be evaluated based on metrics such as accuracy, precision, and F1 score. The findings will reveal which factors are more decisive in predicting healthcare demand and will elucidate the interactions between these factors. These results will provide significant contributions to the management of health policies and resource allocation in Turkey.

Developed by Ronald M. Andersen in 1968, the Behavioral Model of Health Services Use aims to explain the factors influencing individuals' decisions to utilize health services. This model examines the use of health services across three main categories: predisposing, enabling, and need factors.

Predisposing factors include the demographic and social characteristics of individuals and influence their propensity to use health services. In a systematic review by Babitsch et al. [5], the most frequently examined predisposing factors include age, gender, educational status, and ethnicity. For example, age has a significant relationship with the use of health services; older individuals generally use more health services [5].

In terms of machine learning techniques, identifying predisposing factors can enhance the accuracy of predictions related to health service utilization. Modeling these factors is particularly important for understanding how variability in demographic data affects health service demands [6].

Enabling factors encompass the economic and organizational conditions that affect individuals' access to health services. These factors include variables such as income level, health insurance status, and the ease of accessing health services. In the study by Babitsch et al. [5], income status and health insurance are among the most frequently examined enabling factors.

Economic inequalities significantly impact the utilization of healthcare services in Turkey, particularly by contributing to unmet medical needs and difficulties in securing appointments [7]. Additionally, the rising prevalence of chronic diseases increases both perceived and clinically assessed healthcare needs, thereby leading to a greater demand for medical services [8]. In this context, the challenges faced by Turkey's healthcare system—such as limited access in rural areas, economic disparities, and the increasing prevalence of chronic illnesses [9]—can be systematically analyzed through the structured framework provided by Andersen's Behavioral Model. This approach enables a more comprehensive and country-specific examination of the dynamics influencing healthcare utilization.

The interest in machine learning (ML) techniques within the medical field has been steadily increasing [10]. As a natural reflection of this trend, the discipline of healthcare management has also been influenced, leading to a substantial number of ML-related publications. In the context of patient safety [11], artificial intelligence and ML technologies have been applied in various areas, including health technology assessment, evaluations of personalized treatments [12], and patient admission prediction in emergency departments [13]. Furthermore, the impact of ML and deep learning algorithms on patient management, healthcare policy development, and clinical decision support systems is being increasingly investigated in the domains of healthcare management and big data analytics [14].

In the context of machine learning techniques, identifying enabling factors helps in determining barriers and facilitators to accessing health services. This information is crucial for developing health policies and the efficient allocation of resources [15].

Need factors encompass elements such as health status and disease that determine individuals' need for health services. These factors are divided into evaluated health status and perceived health status. In the systematic review by Babitsch et al. [5], evaluated health status and chronic diseases are among the most frequently examined need factors.

From a machine learning perspective, accurately identifying need factors allows for better predictions of health service utilization and the provision of personalized healthcare services. Modeling these factors is crucial for the planning and improvement of health services [16]. In conclusion, Andersen's Behavioral Model of Health Services Use enables a comprehensive analysis of various factors influencing the utilization of health services. The application of this model enhances the understanding of health service demand and contributes to the improvement of the effectiveness of health systems. Machine learning techniques can further aid in accurately predicting health service utilization by identifying the factors within this model.

ML methods offer new opportunities for predicting total healthcare demand. While traditional statistical models are used to examine data and identify risk factors, they often make overly simplistic assumptions or fail to reflect complex relationships [17]. In contrast, ML algorithms can handle nonlinear correlations, account for interactions between multiple components, and adapt to changing patterns in the data.

Researchers can enhance the accuracy and validity of their predictions by combining ML techniques with existing procedures. This approach can lead to more effective preventive and intervention measures for health ministries developing policies for total healthcare demand [18]. Applying ML algorithms to predict the features affecting total healthcare demand can potentially improve the allocation of healthcare resources. Health professionals can identify high-risk areas or groups and distribute resources such as medical personnel, vaccines, and other treatments more efficiently [19]. Additionally, ML can assist in targeted intervention programs by identifying specific risk factors prominent in particular populations.

The aim of the current study is to identify the features affecting total healthcare demand based on a more comprehensive dataset that includes predisposing, enabling, and need factors, as outlined by the Andersen model. ML is a technology that enables machines to learn from data and experiences, make decisions, and generate predictions [20]. ML applications are increasingly prevalent in medical fields, with disease diagnosis and outcome prediction being two areas that benefit from ML [21]. ML algorithms are based on mathematical procedures that define relationships between variables and develop diagnoses. Unlike traditional statistical methods, ML algorithms are data-driven and independent of predefined assumptions [22]. In the healthcare domain, ML has shown improvements in disease prediction and health surveys compared to traditional methods [23, 24].

Method

Data sources and sample

The 2022 Turkey Health Survey, conducted by the Turkish Statistical Institute (TÜİK), is a comprehensive study evaluating the use of healthcare services, health behaviors, and health perceptions in Turkey. This survey encompasses a wide range of health indicators, including access to healthcare services, health statuses, the prevalence of chronic diseases, nutritional habits, and levels of physical activity across the country. The 2022 Turkey Health Survey was carried out through questionnaires administered to selected sample groups nationwide. It analyzes the health statuses, access to healthcare services, and health-related behaviors of individuals aged 15 and over.

The findings of this survey are critically important for the formulation of health policies and the planning of healthcare services. This research provides significant insights into the demand for and needs related to healthcare services in Turkey. It offers valuable information to policymakers, particularly regarding inequalities in access to healthcare, health behaviors of the population, and health perceptions[1].

Outcome variable

The outcome variable is defined as the "Total Number of Healthcare Services Utilized". This variable is not directly available in the TSA 2022 health survey. Instead, it is a composite variable calculated from various types of services received, as explained below. The TSA 2022 health survey did not include a variable specifically named "Total Number of Services Received." This variable was created to represent the number of healthcare services participants received in the past year. It was calculated based on multiple questions regarding different types of services received by participants. These questions included whether participants received inpatient treatment, outpatient day services, dental services, and other types of healthcare services, totaling twelve different service categories. Each of these questions was coded as "1" if the service was received and "0" if it was not. The calculation involved summing these responses. For instance, if a participant received inpatient services, it was coded as 1, and if not, it was coded as 0. An additional column was added to the dataset to calculate the total number of services received by summing these 12 columns.

The twelve variables used in the study are named as follows: Inpatient Services in the Last 12 Months, Outpatient Day Services in the Last 12 Months, Dental Services in the Last 12 Months, Family Doctor Services in the Last 4 Weeks, Specialist Doctor Services in the Last

[1] https://www.tuik.gov.tr/media/microdata/pdf/turkiye-saglik-arastirmasi.pdf.

12 Months, Physical Therapist Services in the Last 12 Months, Physiotherapy Specialist Services in the Last 12 Months, Kinesiotherapist Services in the Last 12 Months, Psychologist Services in the Last 12 Months, Psychotherapist Services in the Last 12 Months, Psychiatrist Services in the Last 12 Months, and Home Care Services in the Last 12 Months.

Figure 1 illustrates the distribution of the total number of healthcare services utilized within the dataset. The X-axis represents the total number of services received, while the Y-axis represents the frequency of these occurrences. Upon examining the histogram, it is evident that the majority of individuals have relatively low service utilization, with a larger proportion receiving services 2, 3, or 4 times. Most service counts are concentrated on the left side of the histogram, indicating that a significant portion of the population utilizes a low number of services. As the number of services increases, the frequency decreases, suggesting that higher service utilization is less common, and the majority of counts are clustered at lower values. The histogram displays a right-skewed distribution, as evidenced by the tail extending to the right. This skewness indicates that a small number of individuals have significantly higher service utilization, which can be considered outliers. Overall, the histogram reveals that while most individuals utilize a low number of healthcare services, a few individuals have much higher utilization, contributing to the overall right-skewed distribution.

Predictors and feature selection

Within the framework of the Andersen model, the variables classified as "Predisposing Factors" encompass fundamental demographic and socioeconomic characteristics that influence individuals' access to and utilization of healthcare services. These factors indirectly determine the need for healthcare services, as they shape individuals' access to and propensity to utilize these services.

- Gender differences are evident in healthcare utilization, with various health conditions and access to healthcare services being influenced by gender-specific disparities. For instance, women typically utilize health check-ups and preventive healthcare services more frequently than men.
- Age is a significant factor directly affecting the need for healthcare services. As age increases, the prevalence of chronic diseases rises, thereby necessitating greater healthcare services.
- The place of birth can present variations in accessible healthcare services and health habits. Rural or urban birthplaces have different dynamics in terms of access to and utilization of healthcare services.

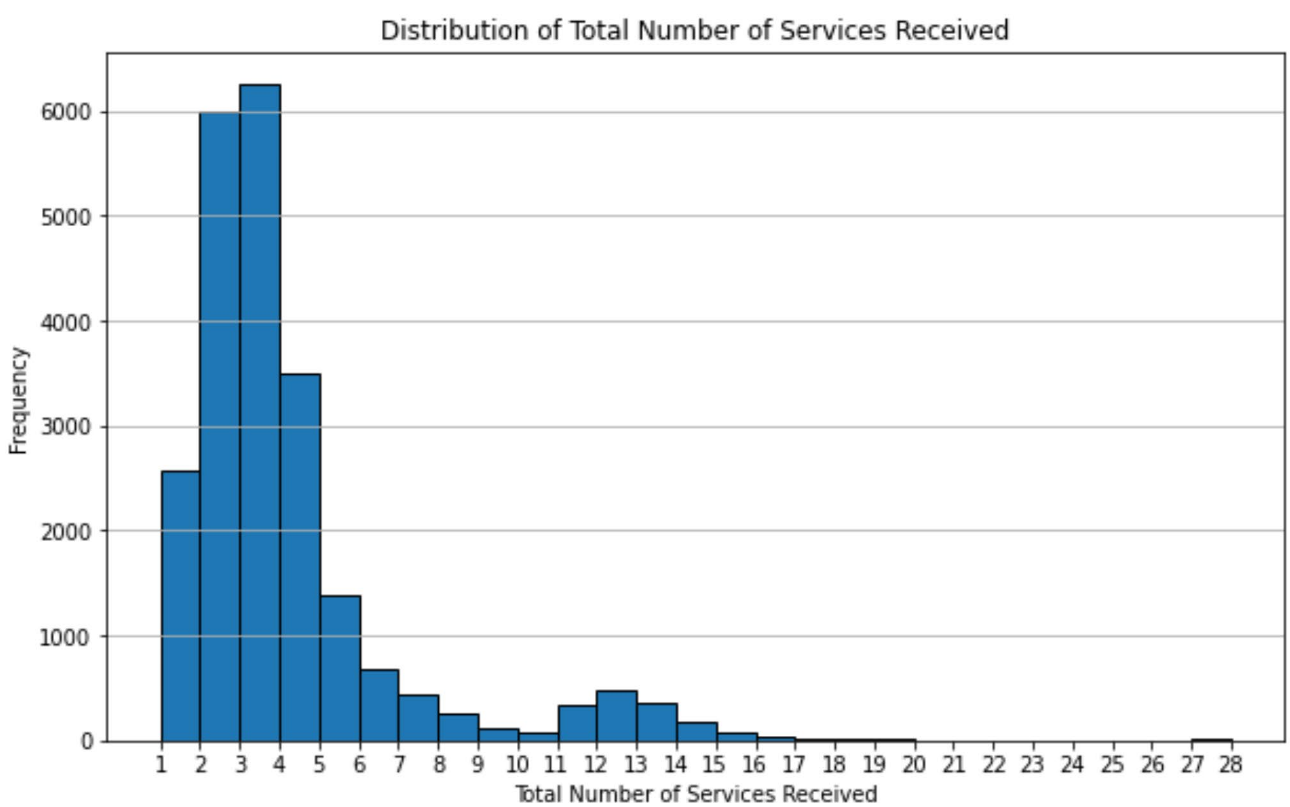


Fig. 1 Target variable: total number of services received

- Education level influences the capacity to benefit from healthcare services and the level of health-related knowledge. Higher education levels are generally associated with better access to healthcare services and a better understanding of health status.
- Married individuals often have a more stable support network compared to those living alone. This is particularly important in the management of chronic diseases and during old age.
- An individual's employment status can be a determinant of access to health insurance, income level, and consequently, access to healthcare services.
- The nature of employment, such as full-time, part-time, or self-employment, can affect access to health insurance and workplace health services.

Each of these variables plays a critical role in understanding individuals' access to and utilization of healthcare services within the framework of Andersen's model. Classifying individuals based on these variables provides essential insights for more effective planning and implementation of healthcare services.

In Andersen's model, the category of "Enabling Factors" includes the conditions and resources that determine individuals' physical and economic access to healthcare services. These factors can directly impact access to healthcare, either facilitating or hindering it. Below are explanations of why each of these variables is considered an enabling factor in this context:

- Government-provided health insurance covers a significant portion of treatment costs, thereby reducing out-of-pocket expenses for individuals and facilitating access to healthcare services.
- Private health insurance typically offers more comprehensive and faster services, making it easier to access healthcare. Insured individuals can benefit from prioritized treatment and better service options.
- Personal financial resources can be a determining factor in accessing healthcare services, especially when treatments are not covered by insurance.
- A stable job and income enable individuals to maintain benefits such as health insurance and access healthcare services regularly.
- Income level is a critical factor, particularly for accessing expensive treatments. Households with higher income levels can access healthcare services more easily.
- Delays in appointment processes can hinder access to healthcare services, especially for non-emergency situations.
- Difficulty or time-consuming travel to healthcare facilities can be a barrier, particularly for those living in rural or remote areas.
- Economic hardships, especially regarding expensive medical procedures, can hinder access to healthcare services.
- Dental health services are typically expensive, and costs can be a significant barrier to access.
- Prescription medications can be costly, posing a continuous barrier to access, particularly for individuals with chronic illnesses.
- Mental health services can be particularly costly without adequate insurance coverage, restricting access to such services.
- The presence of reliable close relatives who can provide support during crises or health issues can facilitate access to healthcare services.
- Support from the community and surrounding environment can ease access to healthcare services, especially for the elderly and disabled individuals.
- Assistance from neighbors or the community can expedite access to healthcare services, particularly in emergency situations.

These variables represent the various ways and conditions that affect individuals' access to healthcare services, and therefore hold significant importance in the Andersen model.

In the Andersen model, "Need Factors" are variables that reflect individuals' objective and subjective needs for healthcare services. These factors indicate individuals' current health status and urgent needs for healthcare services. The variables listed in the provided table pertain to various chronic diseases and health conditions, directly impacting individuals' needs for access to healthcare services. Here is an explanation of each of these variables:

- Chronic Diseases (Various Chronic Conditions): Asthma, Chronic Bronchitis, Myocardial Infarction (Heart Attack), Coronary Heart Diseases, Hypertension, Stroke, Arthritis (Joint Diseases), Lower Back and Neck Issues, Diabetes (Sugar Disease), Allergies, Liver Failure, Incontinence, Kidney Diseases, Depression, High Blood Lipids, Alzheimer's Disease, Celiac Disease, Substance Abuse, Down Syndrome, Autism, and Other Chronic Diseases: These chronic conditions necessitate individuals to require continuous healthcare services. Chronic diseases typically require long-term care, regular monitoring, and treatment.
- Physical and Psychological Conditions: Bodily Pain, Depression, Restlessness: These physical and psychological conditions can influence the frequency and type of healthcare services individuals seek. For

instance, a person experiencing constant bodily pain may visit the doctor more frequently.

- Anthropometric Measurements: Height and Weight (or Body Mass Index, which considers both): These measurements reflect individuals' overall health status and potential health risks. For example, being overweight or underweight can be a risk factor for specific health issues.
- Lifestyle-Related Factors: Smoking and Alcohol Consumption: These lifestyle factors significantly impact individuals' health and can contribute to the development of chronic diseases.

Need factors are among the most critical determinants of the frequency and type of healthcare services individuals utilize. Identifying these variables aids healthcare providers in planning and implementing interventions tailored to individuals, and is also crucial for targeting and developing public health programs.

Additionally, a combined dataset (50 variables) was created by integrating these three factors, and the impact of all variables together on the total number of services received was analyzed. This study comprehensively examines the demographic and background characteristics of 22,742 participants. The research aims to analyze the distribution of variables such as gender, marital status, place of birth, age group, educational level, employment status, and body mass index (BMI) among the participants.

Since none of the variables in our dataset contain missing values, techniques for handling missing data were not employed in this study. Typically, methods such as imputation, deletion, or interpolation are applied when datasets exhibit incomplete observations. However, given the completeness of our dataset, the implementation of such techniques was deemed unnecessary. The absence of missing data enhances the robustness of our analysis, ensuring that the results are not influenced by potential biases introduced through data imputation or other corrective measures. Consequently, the findings presented in this study are derived from a dataset that remains intact, preserving the integrity and reliability of the statistical analyses conducted.

In the study, the demographic and socioeconomic characteristics of the participants are presented in Table 1. The table illustrates the distribution of various attributes, including gender, marital status, birth place, age groups, education level, employment status, and BMI group. Specifically, the gender distribution shows that 48.2% of the participants are male and 51.8% are female. The majority of the participants are married (66.2%), and 97.7% were born in Turkey. The age distribution is as follows: 7.4% are in the 0–18 age group, 27.4% are in the 19–34 age group, 29.2% are in the 35–49 age group, 22.2% are in the 50–64 age group, and 13.8% are 65 years and older. Regarding education, 10.4% have no formal education, 47.1% completed primary school, 22.3% completed high school, 18.3% have a university or college degree, and 2.0% have a master's or PhD degree. Employment status indicates that 39.9% are employed, 45.5% are unemployed, and 14.6% are retired. Lastly, the BMI group shows that 0.3% are severely underweight, 3.1% are underweight, 39.2% are normal weight, 36.6% are overweight, 15.6% are obese (Class 1 and 2), and 5.2% are severely obese.

Statistical analysis, processing of missing values and synthetic minority oversampling technique

For this analysis, Python version 3.9.7 was utilized along with several essential libraries, including pandas, matplotlib, numpy, and scikit-learn (sklearn). These libraries provided robust tools for data manipulation, visualization, numerical operations, and machine learning algorithm implementation, respectively.

A thorough examination of the dataset was conducted to address missing values, a critical step to ensure the integrity and performance of the machine learning models. Depending on the nature and context of the data, different strategies were applied to handle these missing values. For columns with sparse missing values, imputation techniques were employed. Numerical columns had missing values filled using methods such as the mean, median, or mode, while categorical variables were handled by filling with the most frequent category or a new category labeled 'Missing'. In cases where a column had a significant proportion of missing values that could not be reliably imputed, or if the column was deemed non-essential, rows or columns with missing values were removed to maintain consistency and avoid introducing biases.

To prepare the dataset for machine learning algorithms, several preprocessing steps were undertaken, particularly focusing on converting categorical variables into a format suitable for modeling. One of the key techniques used was one-hot encoding, facilitated by the `get_dummies` function from the pandas library. This process converted categorical data into a binary format where each category was represented by a separate column with values of 0 or 1, essential for algorithms that require numerical input. Additionally, although not explicitly mentioned, normalization or standardization of numerical features was considered beneficial, especially for models involving gradient descent optimization, ensuring that all features contributed equally to the model's learning process.

The preprocessing steps, including the careful handling of missing values and transformation of categorical variables, were integral to preparing the dataset for machine learning analysis. By leveraging Python 3.9.7 and powerful libraries like pandas, matplotlib, numpy, and sklearn, the data was meticulously cleaned, transformed, and

Table 1 Demographic and socioeconomic characteristics of the study population

Characteristics	Data Types	Frequency	Percentage	Mean	std
Gender	Categorical (Nominal)			0.51749	0.499705
Male	Categorical (Nominal)	10,971	48,2		
Female	Categorical (Nominal)	11,771	51,8		
Marital Status	Categorical (Nominal)				
Married	Categorical (Nominal)	15,049	66,2	0.66156	0.473189
Not Married	Categorical (Nominal)	7693	33,8	0.33843	0.473189
Birth Place	Categorical (Nominal)			0.97706	0.149691
Turkiye	Categorical (Nominal)	22,221	97,7		
Other	Categorical (Nominal)	521	2,3		
Age Groups	Categorical (Ordinal)				
0–18	Categorical (Ordinal)	1680	7,4	0.07390	0.261619
19–34	Categorical (Ordinal)	6222	27,4	0.27369	0.445862
35–49	Categorical (Ordinal)	6652	29,2	0.29270	0.455015
50–64	Categorical (Ordinal)	5044	22,2	0.22166	0.415376
65+	Categorical (Ordinal)	3141	13,8	0.13790	0.344805
Education Level	Categorical (Ordinal)				
No Formal Education	Categorical (Ordinal)	2369	10,4	0.10396	0.305223
Primary School	Categorical (Ordinal)	10,701	47,1	0.47057	0.499144
High School	Categorical (Ordinal)	5063	22,3	0.22263	0.416023
University/College	Categorical (Ordinal)	4153	18,3	0.18275	0.386474
Masters/PhD	Categorical (Ordinal)	456	2,0	0.02007	0.140247
Employement Status	Categorical (Nominal)				
Employed	Categorical (Nominal)	9069	39,9	0.39896	0.489696
Unemployed	Categorical (Nominal)	10,346	45,5	0.45486	0.497969
Retired	Categorical (Nominal)	3327	14,6	0.14617	0.353292
BMI Group	Categorical (Ordinal)				
Severely Underweight	Categorical (Ordinal)	74	0,3	0.00325	0.056980
Underweight	Categorical (Ordinal)	696	3,1	0.03063	0.172331
Normal Weight	Categorical (Ordinal)	8916	39,2	0.39200	0.488209
Overweight	Categorical (Ordinal)	8315	36,6	0.36564	0.481620
Obese (Class 1 and 2)	Categorical (Ordinal)	3549	15,6	0.15608	0.362940
Severely Obese	Categorical (Ordinal)	1192	5,2	0.05237	0.222795

made ready for robust and accurate modeling, ensuring the best possible outcomes from the machine learning algorithms applied.

Machine learning methods

In our study, we applied seven machine learning methods: logistic regression (LR), k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), decision tree (DT), and extreme gradient boosting (XGBoost). The selection of these techniques was guided by the studies of Feretzakis et al. [13, 25, 26], which served as reference points for determining the most appropriate models for our analysis. By leveraging these established methodologies, our study ensures methodological rigor and alignment with contemporary research trends in machine learning applications within the healthcare domain.

These methods were utilized to identify the factors affecting the total number of services received by individuals in Turkey and to determine the most significant variables among them.

Consequently, we compared the ability of the seven models to identify the factors influencing the total number of services received by individuals and to determine the most significant variables. The performance of these models provided valuable insights into predicting the overall demand of individuals. Each of these models offered various advantages depending on the data characteristics and structure, contributing significantly to the prediction processes.

Logistic Regression (LR)

Logistic Regression (LR) is a statistical technique used to predict the probability of a target variable belonging to one of two possible groups; for example, "yes" or "no," "sick" or "healthy." Mathematically, logistic regression estimates the probability $P(N=1)$, i.e., the probability of the target variable being 1, as a function of the independent variables M. The theory behind logistic regression involves using a logistic (sigmoid) function to constrain the predicted probabilities between 0 and 1. This function transforms any real number into the [0–1] range [27, 28]. The mathematical formula for the logistic function is as follows:

$$\sigma\ (z) = 1/1 + e^{-z}$$

In this formula, z represents the weighted sum of the independent variables. In machine learning, predictions are made using this function, and the output values of the model are probabilities ranging from 0 to 1, obtained by applying the sigmoid function. The advantage of logistic regression is that the predicted probabilities are directly interpretable. Additionally, this method is highly effective and widely used when the target variable is binary [29].

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (K-NN) is a widely used supervised learning technique in data mining and machine learning. This classification method involves a learning process based on the "similarity" information between data points. K-NN is used in both classification and regression techniques [30, 31]. In both techniques, the K nearest training examples in the dataset are used as input. The outcome is determined based on whether K-NN is applied for classification or regression. K-NN has shown to produce better results in various fields [32].

K-NN also has significant applications in the healthcare sector. For instance, a study using demographic and health survey data from India employed the K-NN algorithm for the detection and classification of diabetes. This study demonstrated that K-NN can make accurate classifications in health data [29].

Support Vector Machine (SVM)

It is a powerful classification algorithm that works by creating a hyperplane that maximizes the separation between two classes. SVM is effective even in high-dimensional datasets and aims to maximize the marginal separation. This algorithm is widely used in classification problems and has significant applications in health surveys [29, 31].

Support Vector Classifier is a definitive classifier that distinguishes between two classes by drawing a separating plane. The linear SVC draws a line between two classes, classifying all data points on one side of the line as one class and those on the other side as the second class. This line is chosen not only to separate the two classes but also to stay as far away as possible from the nearest points of both classes [33].

The various applications of SVM in health surveys demonstrate the algorithm's flexibility and accuracy. For instance, in a study conducted to predict the nutritional status of pregnant women in Bangladesh, SVM was used and found to achieve high accuracy rates in determining nutritional status [34]. Additionally, another study predicting under-vaccination status among children under five in East Africa also used SVM, yielding successful results [35]. Furthermore, research conducted to predict depression among secondhand smokers in Korea found that SVM could accurately predict depression risk by analyzing various demographic and health data [36].

Random Forest (RF)

Random Forest (RF) is an ensemble learning technique that combines the predictions of multiple decision trees. In this method, each decision tree is independently trained using randomly selected data subsets, and the overall performance of the model is determined by the majority vote or the average of these trees' predictions. RF is effectively used for both classification and regression problems, enhancing the overall model performance [29]. It creates a series of decision trees during training and outputs the class mode or average prediction of individual trees, addressing classification, regression, and other tasks. Random Forest corrects the tendency of decision trees to overfit training sets [37].

RF adds predictive power to the model by developing trees that consider the best feature from a randomly chosen subset of features, rather than searching for the best feature while splitting each node. This approach typically results in a model that performs better and generalizes more effectively. Consequently, the risk of overfitting is reduced, and the generalization capability is increased [33].

Decision Tree (DT)

A Decision Tree (DT) is a modeling technique that branches out into a chain of decisions and outcomes, representing the data set. In this method, each node performs a test on a feature; the branches represent the results of the test, while the leaves show the predicted outcomes. Decision trees are widely used in classification and regression problems and have the advantage of easy visualization. They also use various measures such as the Gini index or information gain to ensure the separation of features and classes in the data set [38, 39].

When examining the recent literature on the applications of decision trees in the healthcare field, it is evident that this method has been effectively used in various studies. For instance, Kalayou et al. [40] investigated the use of machine learning methods with demographic and health survey data to predict acute respiratory infections in Ethiopian children under the age of five. This study demonstrated that decision trees are an effective tool for predicting disease risks based on various demographic characteristics. Decision trees provide high explainability in such health data and offer ease in visualizing the relationships between features.

Similarly, Zemariam et al. [41] used supervised machine learning algorithms to classify and predict anemia among young girls in Ethiopia. In this research, decision trees played a crucial role in understanding and classifying the key factors determining anemia status. The study emphasized the need to consider performance metrics such as sensitivity and specificity, especially in cases of data imbalance.

Another study by Alie and Negesse [42] applied machine learning methods to predict HIV testing services among adolescents in Ethiopia. In this research, decision trees were effectively used to analyze the impact of different socioeconomic and demographic factors. It was observed that decision trees provide valuable insights for policymakers by visualizing these complex relationships.

Finally, Kim et al. [43] developed a machine learning model using multi-faceted lifestyle data to predict the quality of life of middle-aged adults in South Korea. In this study, decision trees were effectively used to identify and visualize the interactions between multiple factors affecting quality of life. Such models provide transparency and comprehensibility in decision-making processes and play a crucial role in the development of health policies.

In conclusion, decision trees stand out as an effective tool in various health-related research, especially in identifying risk factors, predicting diseases, and planning health services. The ease of interpretation and visualization offered by this method provides significant advantages in analyzing complex health data and in the policy-making processes.

XGBoost

XGBoost (Extreme Gradient Boosting) is an ensemble learning model developed using the gradient boosting algorithm, known for its performance and speed. It delivers high accuracy rates, especially on large datasets, and includes regularization to reduce overfitting issues [44, 45]. XGBoost was developed by Chen and Guestrin in 2016 [46], based on the boosting algorithm. This method typically combines many weak learners, often decision trees, to create a more robust and accurate model.

XGBoost employs a method for calculating the importance scores of features to optimize tree structures. When creating the tree structure, features are arranged according to their importance; less important features are placed at lower levels, while more important ones are at higher levels. This arrangement contributes to a general improvement by ensuring that the trees branch out less and become more detailed. XGBoost is capable of delivering quick and accurate results even on large datasets.

XGBoost is an application of gradient-boosted decision trees designed for speed and performance. It is used for supervised learning problems where there are multiple features and a target variable in the training data. Known for its efficiency and prediction accuracy, XGBoost leverages the gradient boosting framework to enhance its performance [47].

Gradient boosting

Gradient Boosting is another ensemble machine learning technique used for classification and regression tasks. It builds the model incrementally from weak learners, usually decision trees, and generalizes by optimizing an arbitrary differentiable loss function. Due to its ability to enhance model accuracy, Gradient Boosting is effective for predictive modeling tasks [48].

This method has been found effective in evaluating malnutrition risk factors and predicting malnutrition using machine learning approaches among women in Bangladesh [49]. Additionally, a Gradient Boosting model used to improve mental health predictions in refugee camps in Sri Lanka has shown increased accuracy in predictions in this area [50]. Yin, Cao, and Sun [51] utilized this technique to examine the non-linear relationships between population density and waist-to-hip ratio, demonstrating the effectiveness of Gradient Boosting decision trees in modeling such complex relationships.

Model assessment

The performance of machine learning models was evaluated using key metrics to identify factors affecting healthcare service demand and determine the most influential variables. Seven models were compared, each offering valuable insights based on different data characteristics. These metrics are essential for assessing model accuracy and reliability.

Recall (Sensitivity) measures the proportion of true positive cases correctly identified. A high recall (≥ 0.80) in our study highlights the models' effectiveness in predicting healthcare demand. Precision reflects the proportion of true positive predictions among all positive predictions, indicating the model's ability to minimize false positives. The F1 Score balances precision and recall, representing overall performance by accounting for false positives and false negatives. High F1 scores suggest reliable and balanced models. The Area Under the ROC Curve (AUC) assesses a model's ability to distinguish between classes, with higher scores indicating superior performance [52].

Key metrics for ML (Table 2) provide critical insights for evaluating machine learning models in predicting healthcare service demand, emphasizing the importance of recall, precision, and AUC in ensuring model reliability and accuracy. The following table summarizes key metrics used to evaluate the performance of machine learning models in predicting healthcare service demand.

These formulas are the mathematical expressions of the fundamental metrics used to evaluate machine learning

Table 2 Key metrics for ML

Metric	Definition	Formula	Source
Recall (Sensitivity)	The proportion of actual positives correctly identified by the model. Indicates how well the model captures positive instances.	Recall = TP / (TP + FN)	Peretz et al. 2024 [53]
Precision	The proportion of positive predictions that are actually correct. Reflects the accuracy of positive predictions.	Precision = TP / (TP + FP)	Jin et al., 2024 [54]
F1 Score	The harmonic mean of precision and recall. Balances the trade-off between precision and recall.	F1 Score = 2 * (Precision * Recall) / (Precision + Recall)	Rachmat et al. 2024 [55]
ROC AUC	Measures the area under the Receiver Operating Characteristic curve. Evaluates the model's ability to distinguish between classes.	-	Shao et al. 2024 [56]

models' performance. Recall, precision, F1 score, and ROC AUC are critical for understanding the model's accuracy, sensitivity, and overall performance. These metrics are widely used in model evaluation and comparison processes [57].

The workflow diagram (Fig. 2) provides a comprehensive approach to building and evaluating machine learning models using data from TUIK. It begins with data collection and proceeds to data pre-processing, involving cleaning and feature engineering. The pre-processed data is then split into training (90%) and testing (10%) sets. Various machine learning models, including Decision Tree, Random Forest, SVM, KNN, Logistic Regression, XGBoost, and Gradient Boosting, are trained using the training data. These models are evaluated based on Recall, Precision, F1 Score, and ROC AUC metrics to determine their effectiveness. Additionally, feature importance analysis is conducted to identify the most influential features. This systematic process ensures a thorough evaluation and selection of the best-performing model while providing insights into the key factors driving model predictions.

Analysis steps

- Data Collection: The data underpinning this study were obtained from the 2022 Turkey Health Survey microdata provided by the Turkish Statistical Institute (TÜİK). TÜİK offers comprehensive and reliable datasets nationwide, providing valuable resources for health-related research. These data encompass various demographic, socio-economic, and health-related variables influencing healthcare demand.
- Data Pre-processing: Following data collection, the raw data undergo several pre-processing steps. Initially, data cleaning is performed to identify and appropriately address missing values, correct outliers, and resolve inconsistencies within the dataset. Data cleaning is a crucial step to enhance the accuracy and reliability of analyses. Subsequently, feature engineering is conducted, where new features (variables) are derived from the raw data, and existing features are transformed to improve the model's performance.
- Data Split: After completing the pre-processing stage, the dataset is divided into training and test sets. In this study, 90% of the dataset is allocated as training data, while the remaining 10% is designated as test data. The training data are used to train machine learning models, whereas the test data are reserved for evaluating the models' performance.
- Balancing Data Distributions in Supervised Learning with a SMOTE Application: In this study, Synthetic Minority Over-sampling Technique (SMOTE) was applied to address class imbalance before training multiple classifiers, including Decision Tree, Random Forest, SVM, KNN, Logistic Regression, XGBoost, and Gradient Boosting. The initial dataset exhibited a significant imbalance, with Class 1 (18,725 instances) vastly outnumbering the other classes, particularly Class 3 (16 instances) and Class 0 (137 instances). This imbalance could lead to biased models that favor the majority class while underperforming on minority classes. To mitigate this issue, two different SMOTE strategies were implemented. First, targeted SMOTE was applied to Classes 0, 2, and 3, increasing their instance count to 5,000 each, ensuring better representation without unnecessarily inflating the dataset. Additionally, global SMOTE was tested, which equalized all classes to match the largest class. While both approaches improved class distribution, the impact on ROC AUC scores was limited, possibly due to the extremely low initial counts of Class 3 and Class 0. By integrating SMOTE and class weighting techniques, the models became more sensitive to minority classes, leading to improvements in recall and overall model fairness. However, the presence of highly underrepresented classes in the original dataset limited the extent of improvement. This suggests that in cases of extreme imbalance, additional techniques such as data augmentation, feature engineering, or alternative modeling approaches may be necessary for further performance enhancement.

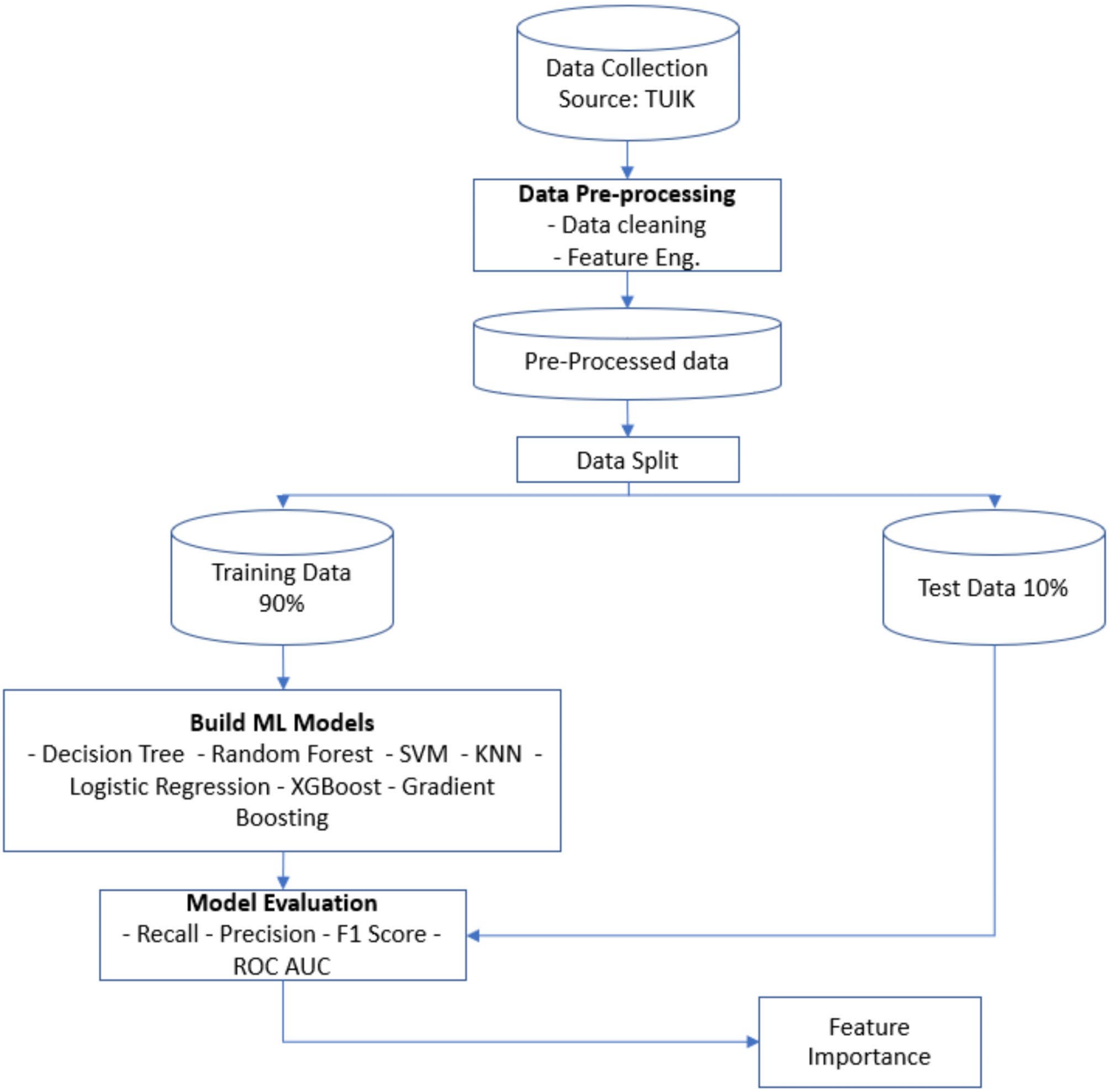


Fig. 2 Study model

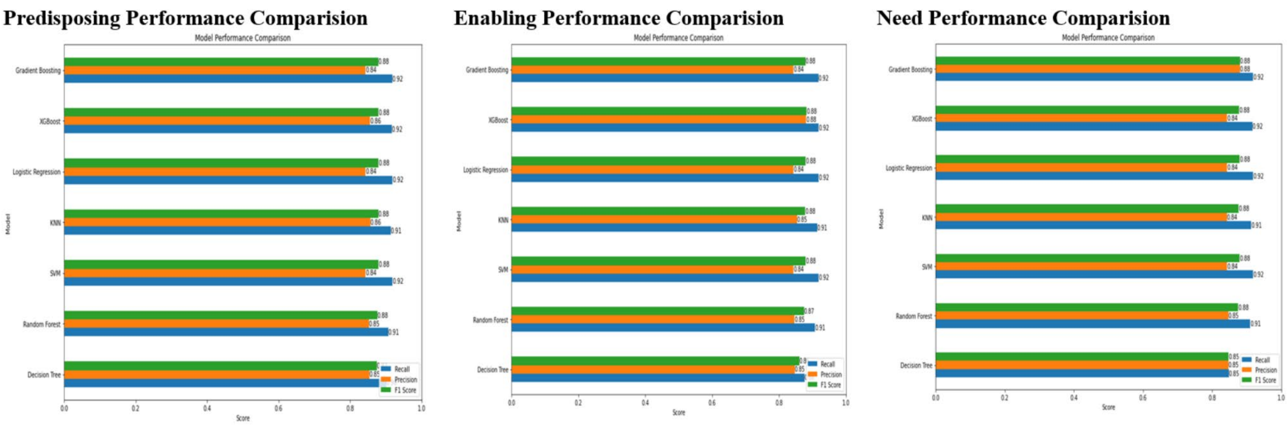
- Building ML Models: Various machine learning models are trained on the training data. This study employs Decision Trees, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression, XGBoost, and Gradient Boosting models. These models are trained and optimized to predict healthcare demand.

Various hyperparameters have been tested to improve the model's performance. Ultimately, the best scores were obtained using the hyperparameters listed below, which were therefore selected for the final model configuration. These parameters are presented in Table 3.

- Model Evaluation: The trained models are assessed using the test data, employing key performance metrics such as recall, precision, F1-score, and the area under the receiver operating characteristic curve (ROC AUC). These metrics are essential for evaluating both the predictive accuracy and the generalization capability of the models. In particular, the representation of F1-score and ROC AUC values is based on the methodology outlined in the study by

Table 3 Key hyperparmeters

Model	Key Hyperparameters
Decision Tree	max_depth = 10, min_samples_split = 5, min_samples_leaf = 2
Random Forest	n_estimators = 300, max_depth = 15, class_weight='balanced'
SVM	kernel='rbf', C = 1, gamma='scale', probability = True
KNN	n_neighbors = 7, weights='distance'
Logistic Regression	solver='liblinear', C = 1, max_iter = 2000, class_weight='balanced'
XGBoost	n_estimators = 300, learning_rate = 0.05, max_depth = 6, subsample = 0.8
Gradient Boosting	n_estimators = 300, learning_rate = 0.05, max_depth = 6, subsample = 0.8



Predisposing – Performance Metrics

	Model	Recall	Precision	F1 Score	ROC AUC
0	Decision Tree	0.901099	0.853618	0.874815	0.501052
1	Random Forest	0.906813	0.852252	0.876203	0.519270
2	SVM	0.918242	0.843168	0.879105	0.540786
3	KNN	0.913407	0.856699	0.879021	0.508306
4	Logistic Regression	0.918242	0.843168	0.879105	0.598806
5	XGBoost	0.916484	0.855535	0.879057	0.555685
6	Gradient Boosting	0.917802	0.843135	0.878886	0.612985

Enabling – Performance Metrics

	Model	Recall	Precision	F1 Score	ROC AUC
0	Decision Tree	0.876044	0.846622	0.860828	0.550719
1	Random Forest	0.907692	0.845786	0.874748	0.589547
2	SVM	0.918242	0.843168	0.879105	0.518300
3	KNN	0.913846	0.853939	0.878490	0.549409
4	Logistic Regression	0.918242	0.843168	0.879105	0.564913
5	XGBoost	0.918242	0.881007	0.882388	0.618502
6	Gradient Boosting	0.917802	0.843135	0.878886	0.613264

Need – Performance Metrics

	Model	Recall	Precision	F1 Score	ROC AUC
0	Decision Tree	0.847912	0.846589	0.847250	0.519785
1	Random Forest	0.909890	0.846524	0.875683	0.555769
2	SVM	0.918242	0.843168	0.879105	0.508706
3	KNN	0.912967	0.842770	0.876465	0.522996
4	Logistic Regression	0.918242	0.843168	0.879105	0.586715
5	XGBoost	0.916484	0.843036	0.878227	0.535873
6	Gradient Boosting	0.918242	0.879989	0.879956	0.543819

Fig. 3 Machine Learning Model Performance of Predisposing Factors, Enabling Factors, and Need Factors

Feretzakis et al. [26, 27], ensuring consistency with established research standards in machine learning-based healthcare evaluations.

- Feature Importance Determination: Finally, the feature importance of the models is analyzed. This step identifies which factors most significantly influence healthcare demand. Determining feature importance provides critical insights for shaping health policies and optimizing resource allocation.

Feature importance is a metric used in a machine learning model to measure the contribution of independent variables (features) to the model's prediction performance. It is a concept particularly prevalent in decision trees and ensemble models (a combination of multiple models). Feature importance values help determine which features are most critical in the model's decision-making process. These values indicate which features the model weights more heavily, allowing analysts and data scientists to identify key features and develop more meaningful and effective models [58].

This systematic analysis process facilitates a better understanding and prediction of healthcare service demand. The findings provide significant insights for planning and developing healthcare services and policies.

Findings

Machine learning model and ROC AUC performance of predisposing factors, enabling factors, and need factors

This section evaluates the effectiveness of different machine learning models in predicting health service demand based on three categories of factors: predisposing factors, enabling factors, and need factors. The ROC AUC scores are used to compare the models' performance in distinguishing between classes, providing insights into which models are most effective for each category of factors. This analysis highlights the strengths and weaknesses of each model and helps identify the most significant predictors of health service utilization.

The performance metrics of various machine learning models (Fig. 3) on the dataset show consistent and strong results. All models achieved a Recall of approximately 0.90, indicating effective identification of positive instances. Precision scores average around 0.85, suggesting some false positives, but overall, the models are adept at identifying true positives. The F1 Scores, ranging from 0.87 to 0.88, highlight a balanced performance between

Precision and Recall. The Decision Tree, KNN, and Gradient Boosting models stood out with F1 Scores close to 0.88. Random Forest and SVM also performed well, with F1 Scores around 0.88. Overall, these models demonstrate robust classification capabilities with high Recall and balanced F1 Scores, making them reliable choices for this dataset.

The ROC curve chart (Fig. 4) highlights the performance of various machine learning models in distinguishing between classes for Predisposing Factors, Enabling Factors, and Need Factors.

For Predisposing Factors, Gradient Boosting and Logistic Regression perform the best, with AUC scores of 0.61 and 0.60, respectively. XGBoost follows with an AUC of 0.56, while SVM, Random Forest, and KNN show moderate performance with AUCs of 0.54, 0.52, and 0.51, respectively. The Decision Tree model performs the worst with an AUC of 0.50.

In the context of Enabling Factors, XGBoost leads with an AUC of 0.62, closely followed by Gradient Boosting at 0.61. Random Forest also performs well with an AUC of 0.59. Logistic Regression and Decision Tree show moderate performance with AUCs of 0.56 and 0.55, respectively, while KNN also scores 0.55. SVM, with an AUC of 0.52, is slightly better than random guessing.

For Need Factors, Logistic Regression emerges as the best performer with an AUC of 0.59, followed by Random Forest with an AUC of 0.56. Gradient Boosting and XGBoost both have AUCs of 0.54, indicating moderate performance. KNN and Decision Tree have AUCs of 0.52, while SVM is slightly better than random guessing with an AUC of 0.51.

Overall, Gradient Boosting, XGBoost, and Logistic Regression consistently show strong performance across all factors, with XGBoost being particularly effective for Enabling Factors. In contrast, Decision Tree and KNN generally perform the worst across all factors.

Feature importance of predisposing factors, enabling factors, and need factors

Feature importance of predisposing factors

The analysis of feature importance (Fig. 5) for the different models highlights the top five most influential features for each model. In the Decision Tree model, the most important features are 'Gender,' 'Education: No Formal Education,' 'Place of Birth,' 'Working Mode 0 (not working),' and 'Age Group 19–34 (Young Adult).' The Random Forest model identifies 'Gender,' 'Place of Birth,' 'Education: No Formal Education,' 'Education: Primary School,' and 'Age Group 50–64 (Older Adult)' as the key features. The Logistic Regression model emphasizes 'Age Group 0–18 (Child and Adolescent),' 'Gender,' 'Age Group 65+ (Senior),' 'Working Mode 2 (part time),' and 'Age Group 19–34 (Young Adult).' For the XGBoost model, the top features are 'Age Group 65+ (Senior),' 'Age Group 0–18 (Child and Adolescent),' 'Employed,' 'Education: No Formal Education,' and 'Age Group 50–64 (Older Adult).' Lastly, the Gradient Boosting model highlights 'Education: No Formal Education,' 'Age Group 65+ (Senior),' 'Gender,' 'Age Group 0–18 (Child and Adolescent),' and 'Employed' as the most significant features. Each model prioritizes different aspects of the data, reflecting the varied approaches and mechanisms they use to predict the target variable effectively.

The analysis of feature importance across all models reveals several features that consistently appear at the top, highlighting their overall significance in predicting the target variable. 'Education: No Formal Education' frequently appears as a critical factor in the Decision Tree, Random Forest, Gradient Boosting, and XGBoost models, indicating that education level significantly influences the model outcomes. 'Gender' also ranks highly in the Decision Tree, Random Forest, and Gradient Boosting models, emphasizing its impact on predictions. 'Place of Birth' is another recurrent feature in the Decision Tree and Random Forest models, suggesting geographic or cultural factors play a substantial role. Additionally, 'Age Group 0–18 (Child and Adolescent)' is prominently featured in the Logistic Regression, XGBoost, and Gradient Boosting models, pointing to the importance of age demographics in the data. Lastly, 'Employed' appears consistently in the XGBoost, Logistic Regression, and Gradient Boosting models, indicating that employment status is a critical predictor across different modeling techniques. These recurring features across multiple models underscore their importance and suggest that education level, gender, place of birth, age group, and employment status are key factors influencing the predictions in this dataset.

Feature importance of enabling factors

The analysis of feature importance (Fig. 6) for the different models highlights the top five most influential features for each model. In the Decision Tree model, the most important features are 'Treatment Cost Self,' 'Community Interest,' 'Payment Difficulty Dental Care,' 'Reliable Closeness (category 3),' and 'Reliable Closeness (category 4).' The Random Forest model identifies 'Treatment Cost Self,' 'Payment Difficulty Dental Care,' 'Treatment Cost SGK,' 'Payment Difficulty Mental Health Care,' and 'Community Interest' as the key features. The Logistic Regression model emphasizes 'Appointment Delay (category 1),' 'Transportation Delay (category 1),' 'Appointment Delay (category 2),' 'Transportation Delay (category 2),' and 'Payment Difficulty Medication.' For the XGBoost model, the top features are 'Job Continuity,' 'Payment Difficulty Mental Health Care,' 'Appointment Delay (category 1),' 'Transportation Delay (category 2),'

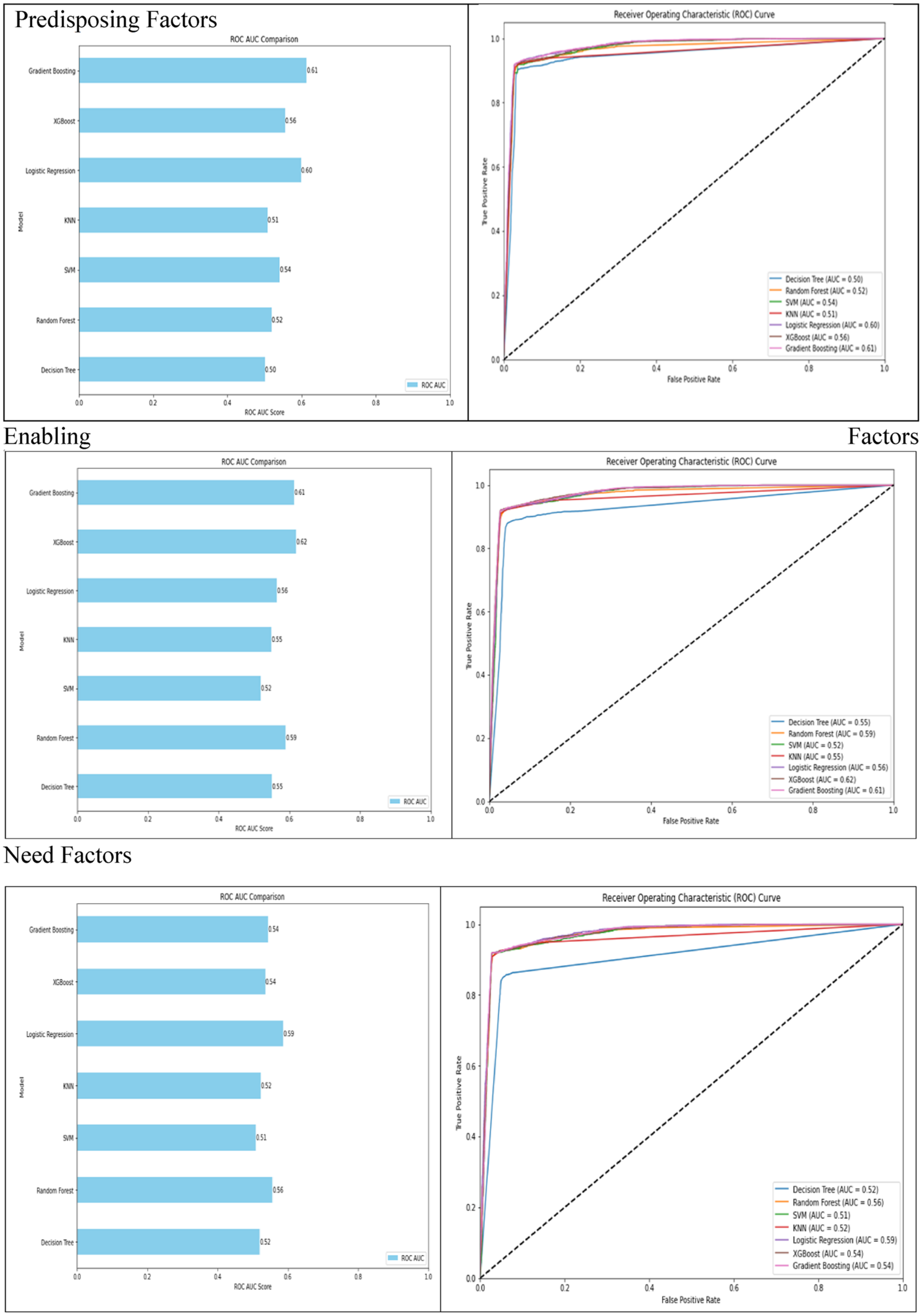


Fig. 4 ROC AUC Performance of Predisposing Factors, Enabling Factors, and Need Factors

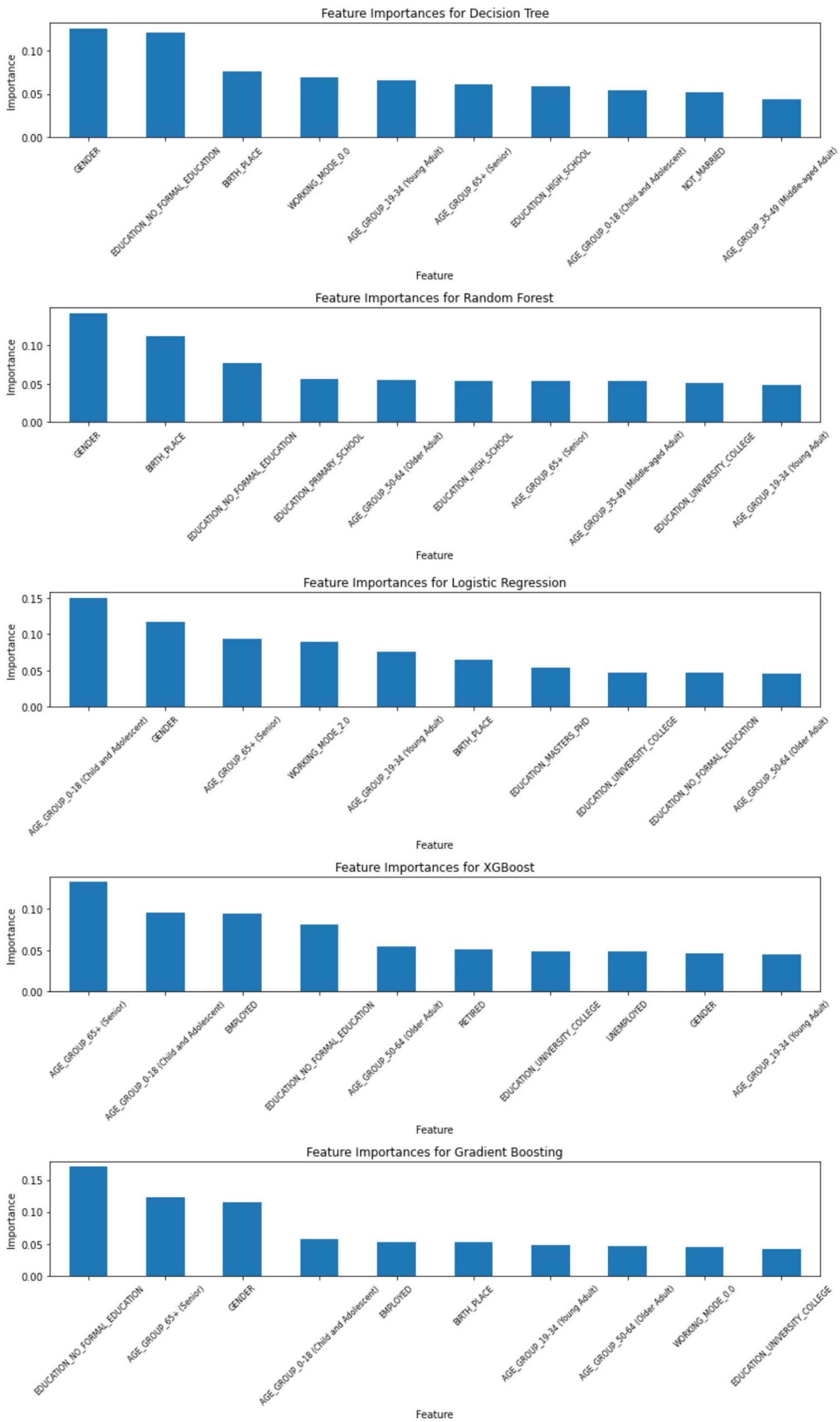


Fig. 5 Feature selection for predisposing factors

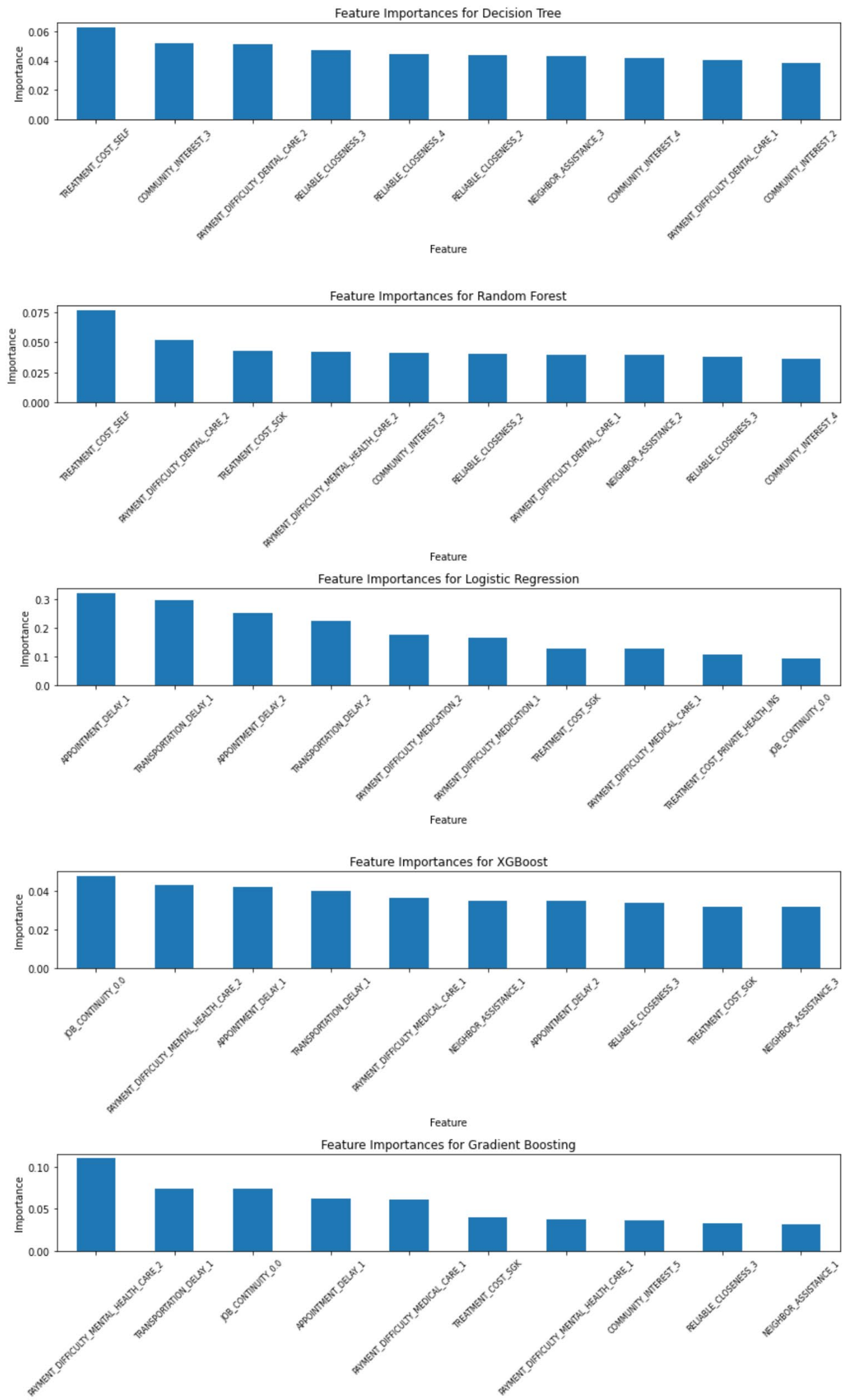


Fig. 6 Feature selection for enabling factors

and 'Payment Difficulty Medical Care.' Lastly, the Gradient Boosting model highlights 'Payment Difficulty Mental Health Care,' 'Transportation Delay (category 1),' 'Job Continuity,' 'Appointment Delay (category 1),' and 'Payment Difficulty Medical Care' as the most significant features. Each model prioritizes different aspects of the data, reflecting the varied approaches and mechanisms they use to predict the target variable effectively.

The analysis of feature importance across all models reveals several features that consistently appear at the top, highlighting their overall significance in predicting the target variable. 'Treatment Cost Self' frequently appears as a critical factor in the Decision Tree and Random Forest models, indicating that the cost of treatment borne by the individual significantly influences the model outcomes. 'Payment Difficulty Dental Care' also ranks highly in the Decision Tree and Random Forest models, emphasizing financial barriers to dental care. 'Community Interest' is another recurrent feature in the Decision Tree and Random Forest models, suggesting that community interest plays a substantial role. Additionally, 'Payment Difficulty Mental Health Care' is prominently featured in the Random Forest, XGBoost, and Gradient Boosting models, pointing to the importance of financial difficulties in accessing mental health care. Lastly, 'Appointment Delay (category 1)' appears consistently in the Logistic Regression, XGBoost, and Gradient Boosting models, indicating that delays in securing appointments are a critical predictor across different modeling techniques. These recurring features across multiple models underscore their importance and suggest that treatment costs, payment difficulties, community interest, and appointment delays are key factors influencing the predictions in this dataset.

Feature importance of need factors

The analysis of feature importance (Fig. 7) for the different models highlights the top five most influential features for each model. In the Decision Tree model, the most important features are 'Smoking Status,' 'Chronic Disease (Neck),' 'Chronic Disease (Lower Back),' 'Chronic Disease (High Blood Lipid),' and 'Chronic Disease (Allergy).' The Random Forest model identifies 'Smoking Status,' 'Chronic Disease (Allergy),' 'Chronic Disease (Lower Back),' 'Chronic Disease (Hypertension),' and 'Chronic Disease (High Blood Lipid)' as the key features. The Logistic Regression model emphasizes 'Disease Health Status,' 'Chronic Disease (Depression),' 'Chronic Disease (Neck),' 'Chronic Disease (Heart Attack),' and 'Physical Pain (category 5).' For the XGBoost model, the top features are 'Disease Health Status,' 'Restlessness (category 6),' 'Chronic Disease (Depression),' 'Physical Pain (category 6),' and 'Chronic Disease (Myocardial Infarction).' Lastly, the Gradient Boosting model highlights 'Disease Health Status,' 'Chronic Disease (Depression),' 'Restlessness (category 6),' 'Chronic Disease (Hypertension),' and 'Chronic Disease (Neck)' as the most significant features. Each model prioritizes different aspects of the data, reflecting the varied approaches and mechanisms they use to predict the target variable effectively.

The analysis of feature importance across all models reveals several features that consistently appear at the top, highlighting their overall significance in predicting the target variable. 'Smoking Status' frequently appears as a critical factor in the Decision Tree and Random Forest models, indicating that smoking habits significantly influence the model outcomes. 'Chronic Disease (Allergy)' also ranks highly in the Decision Tree and Random Forest models, emphasizing the impact of allergies on healthcare demand. 'Chronic Disease (Depression)' is another recurrent feature in the Logistic Regression, XGBoost, and Gradient Boosting models, suggesting that mental health conditions play a substantial role. Additionally, 'Disease Health Status' is prominently featured in the Logistic Regression, XGBoost, and Gradient Boosting models, pointing to the importance of overall health perception. Lastly, 'Chronic Disease (Neck)' appears consistently in the Decision Tree, Logistic Regression, and Gradient Boosting models, indicating the significant impact of neck-related chronic conditions on healthcare utilization. These recurring features across multiple models underscore their importance and suggest that smoking status, chronic diseases, mental health conditions, and overall health status are key factors influencing the predictions in this dataset.

Machine learning model and ROC AUC performance of integrated Data-Based model (Predisposing Factors + Enabling Factors + Need Factors)

The performance metrics of various machine learning models (Fig. 8) on the dataset demonstrate a range of effectiveness in handling this multiclass problem with over 100 features. In such scenarios, the F1 Score and ROC AUC are particularly critical metrics. The F1 Score balances Precision and Recall, providing a comprehensive view of the model's accuracy, while ROC AUC measures the model's ability to distinguish between classes.

The Decision Tree model shows a Recall of 0.8387, Precision of 0.8484, an F1 Score of 0.8435, and an ROC AUC of 0.4966. These metrics suggest that while the Decision Tree can identify positive instances reasonably well, it struggles with class distinction, as indicated by its lower ROC AUC. The Random Forest model, however, performs significantly better with a Recall of 0.9182, Precision of 0.8432, an F1 Score of 0.8791, and an ROC AUC of 0.6414. This indicates a strong balance between identifying true positives and distinguishing between classes.

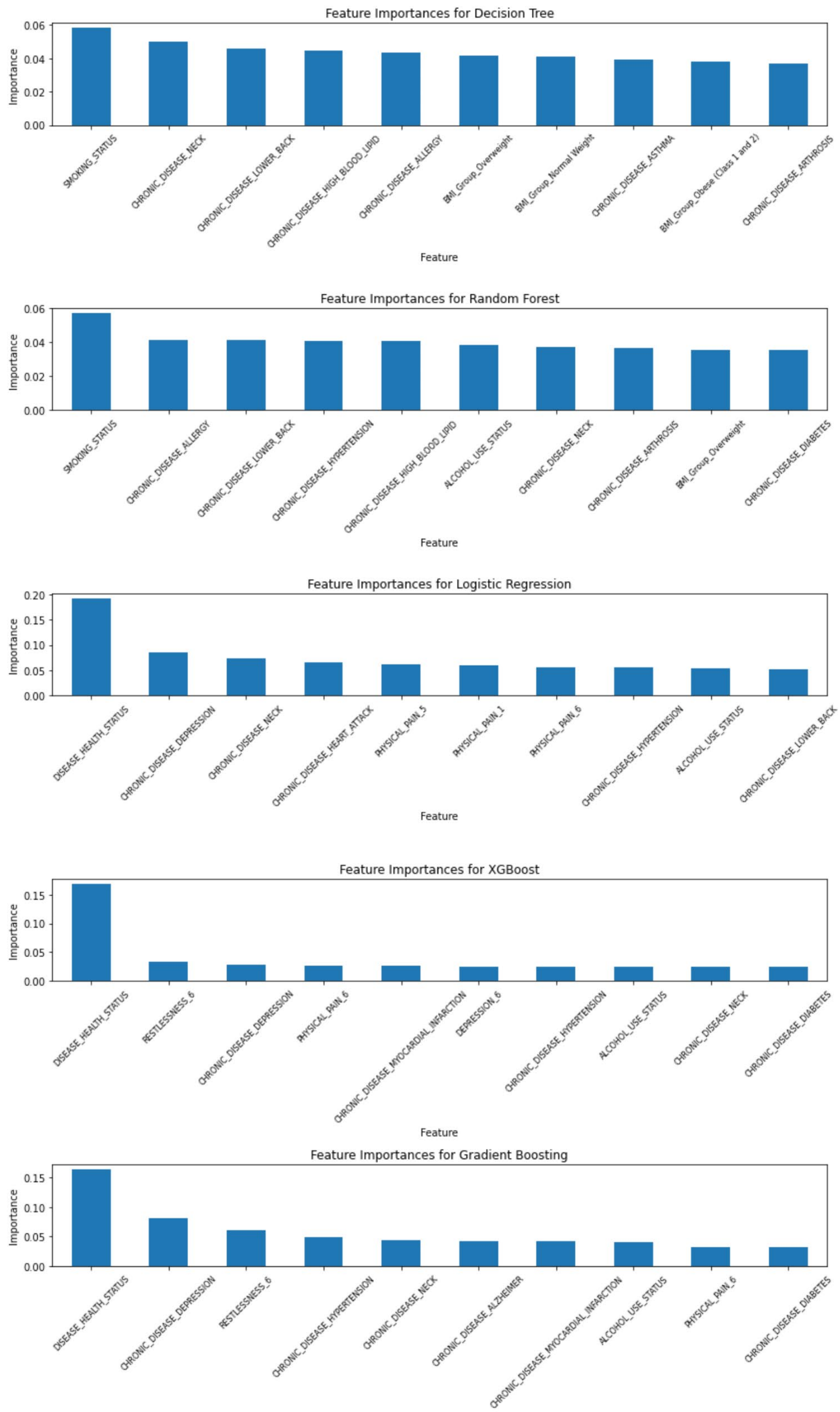
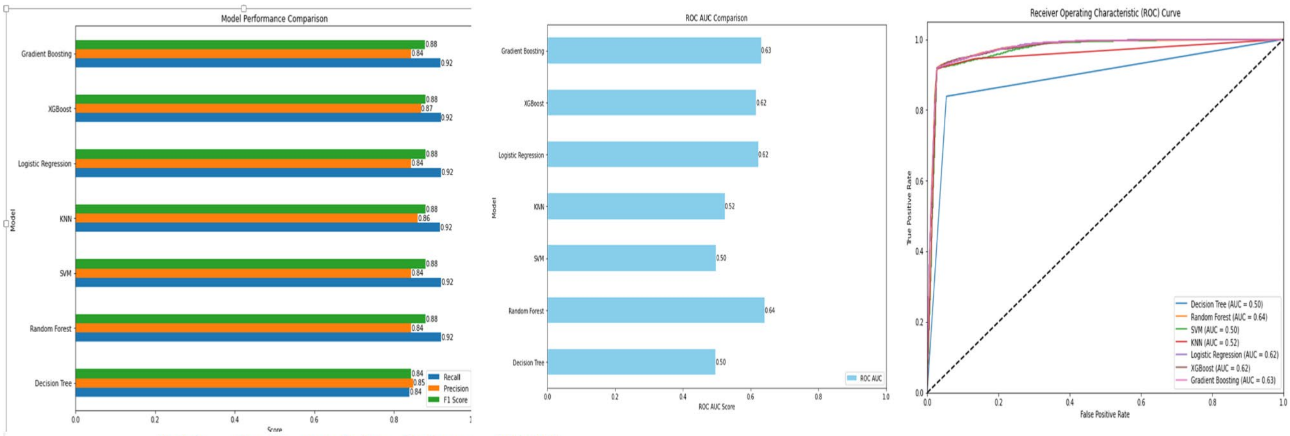


Fig. 7 Feature selection for need factors



	Model	Recall	Precision	F1 Score	ROC AUC
0	Decision Tree	0.838681	0.848359	0.843453	0.496588
1	Random Forest	0.918242	0.843168	0.879105	0.641355
2	SVM	0.918242	0.843168	0.879105	0.497405
3	KNN	0.916044	0.859895	0.879638	0.523633
4	Logistic Regression	0.918242	0.843168	0.879105	0.622297
5	XGBoost	0.917802	0.867795	0.879731	0.616376
6	Gradient Boosting	0.916923	0.843069	0.878446	0.630897

Fig. 8 Integrated data-based model

The SVM model achieves similar high Recall and F1 Score values to Random Forest, both at 0.9182 and 0.8791 respectively, but its ROC AUC of 0.4974 suggests limitations in class distinction, highlighting an area for potential improvement. KNN, with a Recall of 0.9160, Precision of 0.8599, an F1 Score of 0.8796, and an ROC AUC of 0.5236, shows strong Precision and F1 Score, suggesting it is adept at identifying true positives, though its ROC AUC indicates moderate performance in distinguishing classes.

Logistic Regression maintains high performance with a Recall of 0.9182, Precision of 0.8432, an F1 Score of 0.8791, and an ROC AUC of 0.6223, demonstrating effectiveness in both identifying positives and distinguishing between classes. XGBoost stands out with a Recall of 0.9178, a notably high Precision of 0.8678, an F1 Score of 0.8797, and an ROC AUC of 0.6164, indicating a strong overall performance, particularly in maintaining a balance between Precision and Recall.

Finally, the Gradient Boosting model also shows robust performance with a Recall of 0.9169, Precision of 0.8431, an F1 Score of 0.8784, and an ROC AUC of 0.6309. This indicates reliable classification capabilities and effective distinction between classes. Overall, these metrics underscore the importance of the F1 Score and ROC AUC in evaluating model performance in multiclass problems. Random Forest, Logistic Regression, XGBoost, and Gradient Boosting emerge as the most effective models for this dataset, each demonstrating a strong balance of Recall, Precision, F1 Score, and ROC AUC, making them reliable choices for handling complex, feature-rich datasets.

Feature importance of integrated Data-Based model (Predisposing Factors + Enabling Factors + Need Factors)

The analysis of feature importance (Fig. 9) for the different models highlights the top ten most influential features for each model. In the Decision Tree model, the most important features are 'Gender,' 'Smoking Status,' 'Community Interest (category 3.0),' 'BMI Group (Normal Weight),' 'Treatment Cost (Self),' 'BMI Group (Overweight),' 'Physical Pain (category 1.0),' 'Physical Pain (category 6.0),' 'BMI Group (Obese 1 and 2),' and 'Age Group (50–64, Older Adult).' The Random Forest model identifies 'Smoking Status,' 'Gender,' 'Community Interest (category 3.0),' 'BMI Group (Normal Weight),' 'BMI Group (Overweight),' 'Education (Primary School),' 'Neighbor Assistance (category 2.0),' 'Reliable Closeness (category 2.0),' 'Treatment Cost (Self),' and 'Reliable Closeness (category 3.0)' as the key features. The Logistic Regression model emphasizes 'Transportation Delay (category 1.0),' 'Appointment Delay (category 1.0),' 'Appointment Delay (category 2.0),' 'Transportation Delay (category 2.0),' 'Working Mode (category 2.0),' 'Disease Health Status,' 'Treatment Cost (Private Health Insurance),' 'Patient Difficulty (Medication, category 1.0),' 'Patient Difficulty (Medical Care, category 1.0),' and 'Patient Difficulty (Medical, category 2.0).' For the XGBoost model, the top features are 'Disease Health Status,' 'Restlessness (category 6.0),' 'Patient Difficulty (Mental Health Care, category 2.0),' 'Chronic Disease (Depression),' 'Physical Pain (category 6.0),' 'Employment Status (Employed),' 'Chronic Disease (Myocardial Infarction),' 'Depression (category 6.0),' 'Reliable Closeness (category 3.0),' and 'Chronic Disease (Hypertension).' Lastly, the Gradient Boosting model highlights 'Disease Health

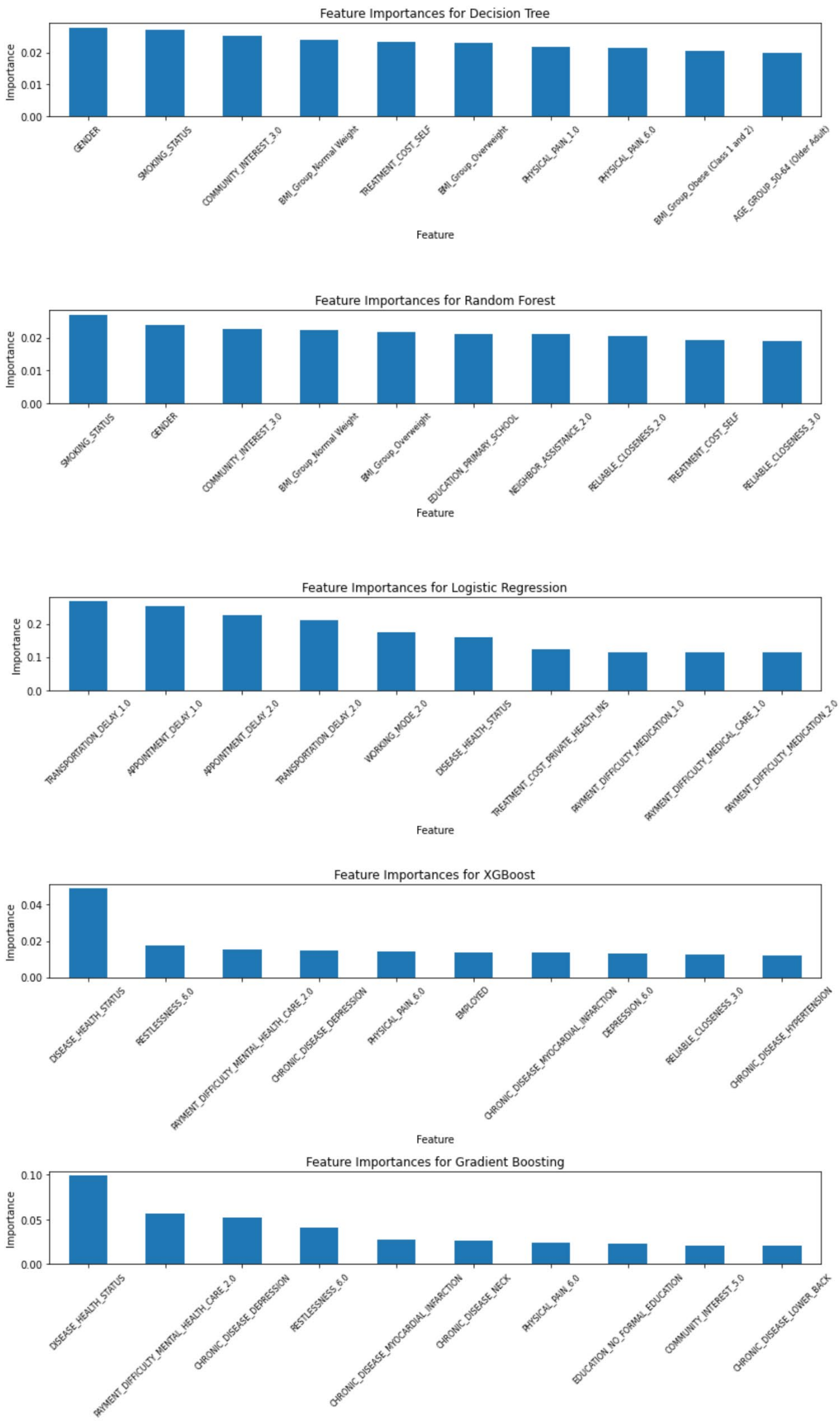


Fig. 9 Feature Selection for the Integrated Data-Based Model

Table 4 Evaluation of factors affecting health service utilization

No	Feature	Description	Factor Group	Evaluation
1	Smoking status	Smoking habits of the individual	Need	Higher predictions for individuals who smoke, indicating a significant impact of smoking on the number of services received. Positive
2	Gender	The gender of the individual	Predisposing	Shows substantial influence with higher predictions based on gender-specific health service utilization. Positive
3	Community interest	Level of community interest and involvement (category 3.0)	Enabling	Higher predictions for individuals with moderate community interest, highlighting the role of social involvement in receiving more services. Positive
4	BMI Group (Normal weight)	Body Mass Index category indicating normal weight	Need	Higher predictions for individuals with normal weight, suggesting active engagement in preventive health services. Positive
5	BMI Group (Overweight)	Body Mass Index category indicating overweight	Need	Higher predictions for overweight individuals, reflecting increased healthcare service use due to related health issues. Positive
6	Education (Primary school)	Education level	Predisposing	Higher predictions for individuals with primary school education, indicating that education level impacts the number of services received. Positive
7	Neighbor Assistance	Level of neighbor assistance (category 2.0)	Enabling	Higher predictions for individuals receiving moderate neighbor assistance, suggesting that social support correlates with more services received. Positive
8	Reliable closeness	Reliability of close relationships (category 2.0)	Enabling	Higher predictions for individuals with moderately reliable close relationships, highlighting the positive impact of social support on service use. Positive
9	Treatment Cost (Self)	Cost of treatment borne by the individual	Enabling	Higher predictions for individuals bearing their treatment costs, reflecting that financial contribution is associated with increased service utilization. Positive
10	Reliable closeness	Reliability of close relationships (category 3.0)	Enabling	Higher predictions for individuals with reliable close relationships, emphasizing the importance of strong social support networks in healthcare service use. Positive

Status,' 'Patient Difficulty (Mental Health Care, category 2.0),' 'Chronic Disease (Depression),' 'Restlessness (category 6.0),' 'Chronic Disease (Myocardial Infarction),' 'Chronic Disease (Neck),' 'Physical Pain (category 6.0),' 'Education (No Formal Education),' 'Community Interest (category 5.0),' and 'Chronic Disease (Lower Back)' as the most significant features. Each model prioritizes different aspects of the data, reflecting the varied approaches and mechanisms they use to predict the target variable effectively.

The analysis of feature importance across all models reveals several features that consistently appear at the top, highlighting their overall significance in predicting the target variable. 'Smoking Status' frequently appears as a critical factor in the Decision Tree and Random Forest models, indicating that smoking habits significantly influence the model outcomes. 'Gender' also ranks highly in the Decision Tree and Random Forest models, emphasizing the impact of gender on healthcare demand. 'Community Interest (category 3.0)' is another recurrent feature in the Decision Tree and Random Forest models, suggesting that social involvement plays a substantial role. Additionally, 'Disease Health Status' is prominently featured in the Logistic Regression, XGBoost, and Gradient Boosting models, pointing to the importance of overall health perception. Lastly, 'Patient Difficulty (Mental Health Care, category 2.0)' appears consistently in the XGBoost and Gradient Boosting models, indicating the significant impact of financial and logistic challenges in accessing mental health care. These recurring features across multiple models underscore their importance and suggest that smoking status, gender, community interest, health status, and patient difficulties are key factors influencing the predictions in this dataset.

Based on the highest recall, precision, F1 score, and ROC AUC values, the feature importance of the Random Forest model was examined. The analysis of feature importance in the Random Forest model shows that several key factors significantly influence the number of healthcare services received (Table 4). Smoking status and gender are critical predisposing and need factors, reflecting lifestyle and demographic influences. Community interest and reliable closeness, categorized under enabling factors, highlight the importance of social involvement and support networks in accessing healthcare services. BMI categories indicate that individuals with normal and overweight statuses have higher healthcare service use, aligning with preventive and health-related needs. Education level and self-borne treatment costs emphasize the impact of socioeconomic status on healthcare service utilization. These findings underscore the multifaceted nature of healthcare service use, driven by a combination of lifestyle, demographic, social, and economic factors.

Discussion and limitations

This study provides a comprehensive evaluation of healthcare demand prediction by employing machine learning models based on Andersen's Behavioral Model of Health Services Use. By integrating predisposing, enabling, and need factors, the analysis offers valuable insights into the factors influencing healthcare utilization in Turkey. The results demonstrate the effectiveness of machine learning techniques in identifying critical predictors of healthcare service demand. Among the seven models tested, Random Forest, Logistic Regression, XGBoost, and Gradient Boosting consistently outperformed other models in terms of recall, precision, F1 score, and ROC AUC. This highlights the robustness and adaptability of ensemble methods and logistic regression in handling complex, nonlinear relationships within healthcare data.

Demographic characteristics such as gender, age, and educational level emerged as significant predictors of healthcare utilization. Consistent with global trends, women tend to use healthcare services more frequently than men. For instance, a study analyzing gender differences in outpatient service utilization among adults aged 20 years or older found that females made much more use of services overall. [59, 60]. Additionally, older individuals and those with lower educational attainment exhibited higher healthcare demand, emphasizing the impact of age-related health needs and the influence of health literacy on service utilization. Research indicates that age and gender profiles of healthcare agency populations clarify service use patterns, identifying high proportions of women in health and social care populations, particularly in older care populations [61, 62].

In Turkey, women tend to utilize healthcare services more frequently compared to men. This phenomenon can be attributed, on one hand, to higher health awareness among women and their widespread participation in regular medical check-ups. Additionally, it is influenced by gender-focused approaches in healthcare practices. Notably, policies such as pregnancy monitoring, postnatal care, and family planning services play a crucial role in enhancing women's access to healthcare services.

The community interest variable reflects how an individual's living environment affects their access to healthcare services. In Turkey, neighborhood relationships, family support, and social solidarity networks are key determinants of healthcare utilization. Furthermore, economic factors play a critical role, particularly in determining access to healthcare services for individuals with lower socioeconomic status. For instance, social assistance mechanisms and public health insurance systems are activated to facilitate healthcare access for individuals facing financial difficulties. These findings highlight the importance of considering gender and community interest factors in the development of healthcare policies.

Financial and social resources play a critical role in facilitating access to healthcare services. The analysis revealed that individuals who bore their treatment costs or received moderate levels of neighbor assistance utilized healthcare services more frequently. This underscores the importance of financial and social support in accessing care. A study on the economics of healthcare access found a severe economic burden across health conditions, wealth quintiles, and study types, highlighting the financial barriers to healthcare utilization [63]. Furthermore, providing social support services, including homemaker services, assessment by a social worker, transportation, and emotional support, can positively impact healthcare utilization and reduce costs [64].

Health-related behaviors and conditions, such as smoking status, chronic diseases, and perceived health status, significantly influenced healthcare utilization. The prominence of smoking status as a predictor aligns with existing literature on the health risks associated with smoking and its impact on long-term healthcare needs. A study comparing healthcare utilization among different smoking groups found that current smokers had higher healthcare utilization and costs compared to former and never smokers [65, 66].

Chronic diseases, including hypertension and depression, were consistently highlighted as critical drivers of healthcare demand across multiple models. Research indicates that the impact of smoking on healthcare utilization and medical costs varies among patients with chronic conditions, suggesting that tobacco control might be more effective at reducing the burden of disease for patients with chronic obstructive pulmonary disease (COPD) and coronary heart disease (CHD) than for patients with diabetes [67].

The superior performance of ensemble models, particularly Random Forest and Gradient Boosting, can be attributed to their ability to capture complex interactions between features. These models demonstrated high recall rates (≥ 0.90), indicating their effectiveness in identifying true positive cases. An ensemble learning approach for diabetes prediction using boosting techniques achieved high accuracy, demonstrating the effectiveness of ensemble methods in healthcare predictions [68].

The relatively lower ROC AUC scores for certain models (e.g., Decision Tree) suggest potential challenges in distinguishing between borderline cases, highlighting the need for further model tuning and feature engineering. The use of Logistic Regression, which also performed well, underscores the continued relevance of traditional statistical methods in healthcare prediction [69]. Despite the advent of more complex algorithms, logistic regression remains a reliable tool, particularly for interpretable and transparent model outputs [70]. A comprehensive review of logistic regression techniques in predicting health outcomes and trends emphasizes its robustness and applicability in health research [71].

Additionally, a study comparing machine learning methods with logistic regression found that logistic regression had similar performance to optimized machine learning algorithms in certain clinical settings, highlighting its effectiveness [72].

Despite the application of SMOTE to address class imbalance, the improvement in ROC AUC scores remained limited across all classifiers. One of the primary reasons for this was the extremely low instance counts of Class 3 (16 instances) and Class 0 (137 instances) in the original dataset, which posed challenges for the models in learning meaningful decision boundaries, even after synthetic samples were introduced. While targeted SMOTE successfully increased minority class representation to 5,000 instances without inflating the dataset unnecessarily, and global SMOTE ensured full class balance by matching all categories to the largest class, neither approach led to a substantial increase in model performance. The high disparity in original class distributions may have caused certain models, particularly tree-based classifiers like Random Forest, XGBoost, and Gradient Boosting, to exhibit overfitting tendencies when trained on synthetic data. Additionally, class weighting techniques, although integrated to enhance model sensitivity to underrepresented classes, did not fully compensate for the inherent imbalance in the dataset. These findings highlight that SMOTE alone may not always be sufficient to improve model performance in cases of extreme class imbalance. Future research could explore alternative solutions, such as data augmentation, advanced ensemble methods, or more sophisticated feature engineering approaches, to better address this issue and enhance classification performance.

The integration of machine learning in healthcare and security applications has demonstrated significant potential for optimizing system performance and resource allocation. In the healthcare domain, Korkmaz et al. [73] explored the role of machine learning algorithms in non-intrusive room occupancy detection, highlighting the effectiveness of Random Forest, k-NN, and Decision Tree models in predicting patient presence based on environmental sensor data. Their study emphasized how feature importance analysis, particularly with light and CO_2 sensors, contributed to accurate predictions, reinforcing the role of machine learning in enhancing healthcare facility management and patient care efficiency. Similarly, Korkmaz et al. [74] applied Artificial Neural Networks (ANNs) with data balancing techniques, including SMOTE and BorderlineSMOTE, to improve firewall packet classification in network security. Their findings underscored the critical importance of addressing class imbalance in machine learning models, as applying oversampling techniques led to substantial improvements in classification accuracy, particularly for underrepresented classes.

Building upon these insights, our study leverages machine learning techniques to predict healthcare demand, integrating feature selection and classification models such as Decision Tree, Random Forest, SVM, KNN, Logistic Regression, XGBoost, and Gradient Boosting. By identifying key factors influencing healthcare utilization, our approach aligns with previous research demonstrating the value of machine learning in predictive modeling. Moreover, similar to the necessity of data balancing in security applications [73, 74], our study also addressed class imbalance in healthcare demand prediction, ensuring more reliable classification outcomes. The combined application of machine learning, feature selection, and predictive analytics in our study further supports the growing role of data-driven models in optimizing both healthcare management and system performance.

The findings have important implications for health policy and resource allocation in Turkey. By identifying the key factors driving healthcare demand, policymakers can tailor interventions to address specific population needs. For example, targeted smoking cessation programs or chronic disease management initiatives could significantly reduce healthcare demand in the long term. Additionally, enhancing social support systems and financial assistance programs may alleviate barriers to accessing care, particularly for vulnerable populations.

Despite its strengths, this study has certain limitations. The reliance on self-reported survey data introduces the potential for recall bias and underreporting of healthcare service utilization. Additionally, the exclusion of certain variables due to missing data may limit the generalizability of the findings.

Moreover, in self-reported data, biases such as recall bias, which arises from participants' uncertainty when recalling past experiences, and social desirability bias, stemming from the tendency to provide responses aligned with societal expectations, constitute significant limitations [75]. It should be acknowledged that the results are based on subjective assessments and rely on health conditions determined by a health authority. Despite these limitations, survey data remains one of the most valuable sources of information for studies based on real-world data.

In our study, we utilized data from the Turkey Health Survey, conducted biennially by the Turkish Statistical Institute (TÜİK). The exclusion of participants under the age of 15 represents a notable demographic limitation in terms of the generalizability of the findings. However, upon examining the dataset, it is evident that many variables included in our model do not contain data for individuals under 15 years of age. For instance, in the case of chronic disease records and critical factors such as general health status, TÜİK did not collect responses from this demographic group during the survey. Therefore, it is essential to emphasize this specific limitation regarding children and adolescent populations.

Additionally, it is possible to conduct international comparisons using a similar survey, such as the European Health Examination Survey, which has been widely utilized in Continental European countries [76]. This allows for meaningful cross-national analyses in healthcare research.

Future research should explore the integration of additional data sources, such as electronic health records (EHRs) or administrative healthcare databases, to enhance model accuracy and comprehensiveness. Moreover, expanding the analysis to include temporal trends and regional variations in healthcare demand could provide deeper insights into the evolving healthcare landscape.

The use of machine learning in healthcare demand prediction raises concerns about transparency, fairness, and bias. Ensuring interpretability and accountability in AI-driven models is essential for ethical decision-making. Bias in training data may lead to disparities in healthcare predictions, disproportionately affecting vulnerable populations. To mitigate this, fairness-aware algorithms and continuous model evaluation should be prioritized. Additionally, AI should support, not replace, human expertise in healthcare planning. Establishing regulatory frameworks and ethical guidelines is crucial to ensuring equitable and responsible AI integration in healthcare policy [77, 78].

Conclusion

This study aims to predict the total demand for healthcare services using machine learning methods, utilizing microdata from the Turkish Statistical Institute's (TÜİK) 2022 Turkey Health Survey. The analyses are based on Andersen's Behavioral Model of Health Services Use, encompassing predisposing factors, enabling factors, and perceived need factors. Seven different machine learning models were employed in this study: Decision Tree, Random Forest, SVM, KNN, Logistic Regression, XGBoost, and Gradient Boosting.

As a result of the analyses, the most important features affecting the demand for healthcare services were identified, and the performance of the models was evaluated in terms of prediction accuracy and generalization capability. Among the predisposing factors, gender, education level, and age group stood out, while treatment costs, community interest, and payment difficulty were significant among the enabling factors. Need factors were determined by variables such as smoking status, chronic diseases, and overall health status.

The performance evaluation of the machine learning models was conducted using F1 Score and ROC AUC metrics. Random Forest, Logistic Regression, XGBoost, and Gradient Boosting models demonstrated the best performance with high recall and precision values. Notably, the Random Forest model stood out with its high accuracy and generalization capability.

The findings of this study demonstrate the effectiveness of machine learning methods in predicting the demand for healthcare services. The use of such analyses contributes significantly to the more effective planning of health policies and resource allocation. Understanding the separate and combined effects of predisposing, enabling, and need factors on healthcare service demand will enable more efficient and effective planning of healthcare services. This study provides valuable insights into the factors affecting access to and use of healthcare services and supports the development of strategic interventions accordingly.

The key predictors identified in our study provide significant insights for the more effective formulation of healthcare policies. Variables such as financial difficulties in accessing dental care and chronic diseases, particularly depression, highlight structural inequalities in healthcare accessibility. In this context, several policy recommendations are proposed. Chronic mental illnesses, such as depression, have been identified as fundamental determinants of healthcare demand. To enhance access to mental health services in Turkey, it is essential to expand psychiatric and psychotherapy services, establish mental health counseling units within family health centers, and develop free or low-cost therapy programs, particularly in disadvantaged regions. To mitigate the effects of chronic diseases such as depression, early diagnosis and regular monitoring systems should be established. The formation of multidisciplinary chronic disease management teams in hospitals and the enhancement of training programs on chronic disease management within primary healthcare services may optimize healthcare demand and contribute to a more efficient allocation of healthcare resources.

The study also indicates that individuals facing financial difficulties in accessing dental care are unable to fully utilize healthcare services. This issue may negatively impact overall health outcomes in the long term. Expanding public dental health programs, extending insurance coverage for low-income individuals, and introducing state-supported mobile dental clinics in rural areas could contribute to addressing this problem.

Furthermore, the study demonstrates that smoking is a key determinant of healthcare demand. The habit of smoking not only affects individual health outcomes but also significantly increases the overall demand for healthcare services. Smoking is a primary risk factor for numerous chronic diseases, including respiratory disorders (e.g., COPD, asthma, chronic bronchitis), cardiovascular diseases (e.g., coronary artery disease, stroke), and certain types of cancer (e.g., lung cancer). The inability to prevent or intervene at an early stage in these diseases increases the demand for both primary and advanced healthcare services, such as hospital admissions and intensive care needs. Particularly, the high cost of treating smoking-related health conditions imposes a substantial financial burden on the healthcare system in the long run. Given the prevalence of smoking in Turkey and its impact on the healthcare system, the following aspects become critical for healthcare policy:

Increasing public awareness campaigns aimed at reducing smoking, implementing regulations to prevent smoking initiation at a young age, and raising tobacco taxes could contribute to decreasing long-term healthcare demand. Expanding smoking cessation programs within primary healthcare services, implementing regular pulmonary function tests, and promoting screening programs for the early detection of smoking-related diseases are essential measures. Additionally, ensuring greater accessibility to pharmacological and behavioral interventions for smoking addiction, particularly in socioeconomically disadvantaged regions where smoking rates are high, is crucial for effectively managing healthcare demand. Strategic steps should also be taken concerning other key predictors identified in the study.

Authors' contributions

Dr. Fatih Orhan was primarily responsible for the conception and design of the study, as well as the preparation and writing of the manuscript. Dr. Mehmet Nurullah Kurutkan conducted the data analysis and contributed significantly to the interpretation of the results. Both authors reviewed and approved the final version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The datasets generated and analyzed during this study are not publicly available but are held by the authors. Data can be made available upon reasonable request. Please contact the corresponding author for access.

Declarations

Ethics approval and consent to participate

The data utilized in this study are secondary in nature, sourced from the Turkish Health Survey micro datasets provided by the Turkish Statistical Institute (TÜİK). These datasets are made accessible to researchers upon request and payment of a specified fee. As the data are anonymized, no personally identifiable information is disclosed, ensuring participant confidentiality. Consequently, this study was conducted using an already publicly accessible secondary dataset, and no additional consent or ethical approval was required. This approach is consistent with prior studies using older versions of the same dataset [79], where no additional research permission or ethical committee approval was necessary for generating scientific outputs.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]University of Health Sciences, Gülhane Vocational School of Health, Ankara, Turkey
[2]Department of Health Management, Duzce University, Duzce, Turkey

Received: 12 January 2025 / Accepted: 28 February 2025
Published online: 12 March 2025

References

1. Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, Ting DSW. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Rep Med 2024;5(2).
2. Giddings R, Joseph A, Callender T, Janes SM, Van der Schaar M, Sheringham J, Navani N. Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. Lancet Digit Health. 2024;6(2):e131–44.
3. Lee HY, Yoon YJ, Choi YJ, Ham YH. Factors associated with Korean American women's Health-Related internet use: findings from Andersen's behavioral model. J Immigr Minor Health. 2024;26(1):124–32.
4. Mubeen Z, Fatmi Z, Hameed W, Asim M. Barriers and facilitators to accessing adolescents' mental health services in Karachi: users and providers perspectives. BMC Health Serv Res. 2024;24(1):157.
5. Babitsch B, Gohl D, Von Lengerke T. Re-revisiting Andersen's behavioral model of health services use: a systematic review of studies from 1998–2011. GMS Psycho Soc Med. 2012;9:1-15.
6. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995;1–10.
7. Terzi M, Kurutkan MN, Şahin D, Kara O. Unmet medical care needs due to payment difficulty. Sağlık Akademisyenleri Dergisi. 2020;9(4):309–16.
8. Genç G, Kurutkan MN. Eşitsizlik Bağlamında Karşılanmayan Sağlık İhtiyacı: Türkiye Sağlık Araştırması Verilerinden Kanıtlar. SDÜ Sağlık Yönetimi Dergisi. 2021;3(1):34–51.
9. Alkan Ö, Bayrakçeken E, Yarali S. Prevalence and predictors of tobacco use in older adults: a case in Turkey. J Subst Use. 2024;29(4):562–7.
10. Tsoni R, Kaldis V, Kapogianni I, Sakagianni A, Feretzakis G, Verykios VS. A machine learning pipeline using KNIME to predict hospital admission in the MIMIC-IV database. In: 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA). IEEE. 2023;1–6.
11. Arjmandnia F, Alimohammadi E. The value of machine learning technology and artificial intelligence to enhance patient safety in spine surgery: a review. Patient Saf Surg. 2024;18(1):11.
12. Zhang Y, Kreif N, Gc VS, Manca A. Machine learning methods to estimate individualized treatment effects for use in health technology assessment. Med Decis Making. 2024;44(7):756–69.
13. Feretzakis G, Sakagianni A, Anastasiou A, Kapogianni I, Tsoni R, Koufopoulou C, Verykios VS. Machine learning in medical triage: A predictive model for emergency department disposition. Appl Sci. 2024;14(15):6623.
14. Vajjhala NR, Eappen P. An exploratory review of machine learning and deep learning applications in healthcare management. In: International conference on machine learning and big data analytics. Cham: Springer. 2025;315–324.
15. Andersen RM, Davidson PL. Improving access to care in America: individual and contextual indicators. In: Andersen RM, Rice TH, Kominski EF, editors. Changing the U.S. Health care system: key issues in Health services, policy, and management. Jossey-Bass; 2001. pp. 3–30.
16. Aday LA, Andersen R. A framework for the study of access to medical care. Health Serv Res. 1974;9(3):208.
17. Moulaei K, Shanbehzadeh M, Mohammadi-Taghiabad Z, Kazemi-Arpanahi H. Comparing machine learning algorithms for predicting COVID-19 mortality. BMC Med Inf Decis Mak 2022;22(2).
18. Samuel O, Zewotir T, North D. Application of machine learning methods for predicting under-five mortality: analysis of Nigerian demographic health survey 2018 dataset. BMC Med Inf Decis Mak. 2024;24(1):86.
19. Del Solfa G, F., Simonato FR. Big data analytics in healthcare: exploring the role of machine learning in predicting patient outcomes and improving healthcare delivery. International Journal of Computations, Information and Manufacturing (IJCIM); 2023. p. 3.
20. Géron A. Hands-on machine learning with Scikit-Learn, Keras, and tensorflow. O'Reilly Media, Inc.; 2022.
21. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19:1–18.

22. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina. 2020;56(9):455.
23. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. Knee Surg Sports Traumatol Arthrosc. 2022;30(3):753–7.
24. Tesfaye SH, Seboka BT, Sisay D. (2024). Application of machine learning methods for predicting childhood anaemia: analysis of Ethiopian demographic health survey of 2016. PLoS ONE 19(4):e0300172.
25. Feretzakis G, Sakagianni A, Loupelis E, Karlis G, Kalles D, Tzelves L, Kaldis V. Predicting hospital admission for emergency department patients: A machine learning approach. Informatics and technology in clinical care and public health. IOS; 2022;297–300.
26. Feretzakis G, Karlis G, Loupelis E, Kalles D, Chatzikyriakou R, Trakas N, Kaldis V. Using machine learning techniques to predict hospital admission at the emergency department. J Crit Care Med. 2022;8(2):107–16.
27. Islam MM, Ferdousi R, Rahman S, Bushra HY. Likelihood prediction of diabetes at early stage using data mining techniques. In: Computer vision and machine intelligence in medical image analysis. Singapore: Springer; 2020. p. 113–25.
28. Nusinovici S, Tham YC, Yan MYC, Ting DSW, Li J, Sabanayagam C, Cheng CY. Logistic regression was as good as machine learning for predicting major chronic diseases. J Clin Epidemiol. 2020;122:56–69.
29. Thotad PN, Bharamagoudar GR, Anami BS. Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods. Diabetes Metabolic Syndrome: Clin Res Reviews. 2023;17(1):102690.
30. Zhang S. Challenges in KNN classification. IEEE Trans Knowl Data Eng. 2021;34(10):4663–75.
31. Bansal M, Goyal A, Choudhary A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decis Analytics J. 2022;3:100071.
32. Wang H, Xu P, Zhao J. Improved KNN algorithm based on preprocessing of center in smart cities. Complexity. 2021;20211:5524388.
33. Sneha N, Gangil T. Analysis of diabetes mellitus for early prediction using optimal features selection. J Big Data. 2019;6(1):1–19.
34. Begum N, Rahman MM, Omar Faruk M. (2024). Machine learning prediction of nutritional status among pregnant women in Bangladesh: evidence from Bangladesh demographic and health survey 2017–18. PLoS ONE, 19(5):e0304389.
35. Tadese ZB, Nigatu AM, Yehuala TZ, Sebastian Y. Prediction of incomplete immunization among under-five children in East Africa from recent demographic and health surveys: a machine learning approach. Sci Rep. 2024;14(1):11529.
36. Kim NH, Kim M, Han JS, Sohn H, Oh B, Lee JW, Ahn S. Machine-learning model for predicting depression in second-hand smokers in cross-sectional data using the Korea National health and nutrition examination survey. Digit Health. 2024a;10:20552076241257046.
37. Palimkar P, Shaw RN, Ghosh A. Machine learning technique to prognosis diabetes disease: Random forest classifier approach. In: Advanced computing and intelligent technologies: proceedings of ICACIT 2021. Singapore: Springer. 2022;219–244
38. Miglio R, Soffritti G. The comparison between classification trees through proximity measures. Comput Stat Data Anal. 2004;45(3):577–93.
39. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. J Appl Sci Technol Trends. 2021;2(01):20–8.
40. Kalayou MH, Kassaw AAK, Shiferaw KB. Empowering child health: Harnessing machine learning to predict acute respiratory infections in Ethiopian under-fives using demographic and health survey insights. BMC Infect Dis. 2024;24(1):338.
41. Zemariam AB, Yimer, Abebe GK, Wondie W T, Abate BB, Alamaw AW… Ngusie H S. Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia. Sci Rep 2024;14(1):9080.
42. Alie MS, Negesse Y. Machine learning prediction of adolescent HIV testing services in Ethiopia. Front Public Health. 2024;12:1341279.
43. Kim J, Jeong K, Lee S, Baek Y. Machine-learning model predicting quality of life using multifaceted lifestyles in middle-aged South Korean adults: a cross-sectional study. BMC Public Health. 2024b;24(1):159.
44. Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM. Extreme gradient boosting as a method for quantitative Structure–Activity relationships. J Chem Inf Model. 2016;56(12):2353–60.
45. Carmona P, Climent F, Momparler A. Predicting failure in the U.S. Banking sector: an extreme gradient boosting approach. Int Rev Econ Finance. 2019;61:304–23.
46. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016;785–794.
47. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Environ. 2023;31(6):3360–79.
48. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev. 2021;54:1937–67.
49. Turjo EA, Rahman MH. Assessing risk factors for malnutrition among women in Bangladesh and forecasting malnutrition using machine learning approaches. BMC Nutr. 2024;10(1):22.
50. Sahoo I, Amona E, Kuttikat M, Chan D. Enhancing mental health predictions: A gradient boosted model for Sri Lankan camp refugees. Social Sci. 2024;13(5):255.
51. Yin C, Cao J, Sun B. Examining non-linear associations between population density and waist-hip ratio: an application of gradient boosting decision trees. Cities. 2020;107:102899.
52. Bowers AJ, Zhou X. Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. J Educ Students Placed Risk (JESPAR). 2019;24(1):20–46.
53. Peretz O, Koren M, Koren O. Naive Bayes classifier–An ensemble procedure for recall and precision enrichment. Eng Appl Artif Intell. 2024;136:108972.
54. Jin C, Jia C, Hu W, Xu H, Shen Y, Yue M. Predicting antimicrobial resistance in E. coli with discriminative position fused deep learning classifier. Comput Struct Biotechnol J. 2024;23:559–65.
55. Rachmat R, Yusuf M, Basmar MF, Suherwin S. Analysis of algorithms and data processing efficiency in movie recommendation systems based on machine learning. Jurnal Mandiri IT. 2024;13(2):273–9.
56. Shao H, Xu Q, Yang Z, Wen P, Peifeng G, Huang Q. Weighted Roc curve in cost space: extending Auc to cost-sensitive learning. Adv Neural Inf Process Syst 2024;36.
57. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
58. Balachandran PV, Xue D, Theiler J, Hogden J, Gubernatis JE, Lookman T. Importance of feature selection in machine learning and adaptive design for materials. Materials discovery and design: by means of data science and optimal learning. Cham: Springer International Publishing; 2018. pp. 59–79.
59. Noh J-W, Kim K-B, Park H, Kwon YD. Gender differences in outpatient utilization: A pooled analysis of data from the Korea health panel. J Womens Health. 2017;26(2):178–85. https://doi.org/10.1089/jwh.2016.5771.
60. Kammer-Kerwick M, Cox K, Purohit I, Watkins SC. (2024). The role of social determinants of health in mental health: an examination of the moderating effects of race, ethnicity, and gender on depression through the all of Us research program dataset. PLOS Mental Health, 1(3), e0000015.
61. Keene J, Li X. Age and gender differences in health service utilization. J Public Health. 2005;27(1):74–9.
62. Placiszewski K, Wierzba W, Ostrowski J, Pinkas J, Jankowski M. (2023). Factors associated with the use of public eHealth services in Poland-a 2022 nationwide cross-sectional survey. Ann Agric Environ Med 30(1).
63. de Siqueira Filha NT, Li J, Phillips-Howard PA, Quayyum Z, Kibuchi E,Mithu MIH… Elsey H. The economics of healthcare access: a scoping review on the economic impact of healthcare access for vulnerable urban populations in low-and middle-income countries. Int J Equity Health 2022;21(1):191.
64. Shier G, Ginsburg M, Howell J, Volland P, Golden R. Strong social support services, such as transportation and help for caregivers, can lead to lower health care use and costs. Health Aff. 2013;32(3):544–51.
65. Wacker M, Holle R, Heinrich J, Ladwig KH, Peters A, Leidl R, Menn P. The association of smoking status with healthcare utilisation, productivity loss and resulting costs: results from the population-based KORA F4 study. BMC Health Serv Res. 2013;13:1–9.
66. Mayer J, Brandstetter S, Tischer C, Seelbach-Göbel B, Malfertheiner SF, Melter M… Apfelbacher C. Utilisation of supplementary prenatal screening and diagnostics in Germany: cross-sectional study using data from the KUNO Kids Health Study. BMC Pregnancy Childbirth 2022;22(1):436.
67. Ye BZ, Wang XY, Wang YF, Liu NN, Xie M, Gao X, Liang Y. Impact of tobacco smoking on health care utilization and medical costs in chronic obstructive pulmonary disease, coronary heart disease and diabetes. Curr Med Sci. 2022;42(2):304–16.

68. Ganie SM, Pramanik PKD, Malik B, Mallik M, S., Qin H. An ensemble learning approach for diabetes prediction using boosting techniques. Front Genet. 2023;14:1252159.
69. Fang J. Why logistic regression analyses are more reliable than multiple regression analyses. J Bus Econ. 2013;4(7):620–33.
70. Yang Y, Wu M. Explainable machine learning for improving logistic regression models. In: 2021 IEEE 19th International Conference on Industrial Informatics (INDIN). IEEE. 2021;1–6.
71. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Disease. 2019;11(Suppl 4):S574.
72. Sievering AW, Wohlmuth P, Geßler N, Gunawardene MA, Herrlinger K, Bein B… Stang A. Comparison of machine learning methods with logistic regression analysis in creating predictive models for risk of critical in-hospital events in COVID-19 patients on hospital admission. BMC Med Inform Decis Making 2022;22(1):309.
73. Korkmaz A, Koşunalp S, Korkmaz AÇ. Enhancing Healthcare Facility Management through Non-intrusive Occupancy Detection: A Comparative Study of Machine Learning Algorithms. In 2024 5th International Conference on Communications, Information, Electronic and Energy Systems (CIEES). IEEE. 2024;1–9.
74. Korkmaz A, Bulut S, Talan T, Kosunalp S, Iliev T. Enhancing firewall packet classification through artificial neural networks and synthetic minority Over-Sampling technique: an innovative approach with evaluative comparison. Appl Sci. 2024;14(16):7426.
75. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. J Appl Psychol. 2003;88(5):879–903.
76. Donfrancesco C, Buttari B, Marcozzi B, Sieri S, Di Lonardo A, Lo Noce C… Palmieri L. Nutrition, physical activity and smoking habit in the Italian general adult population: CUORE project health examination survey 2018–2019. Healthcare 2024;12(4):475. MDPI.
77. Kasula BY. Ethical implications and future prospects of artificial intelligence in healthcare: A research synthesis. Int Meridian J. 2024;6(6):1–7.
78. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R,… Naganawa S. Fairness of artificial intelligence in healthcare: review and recommendations. Japanese J Radiol 2024;42(1):3–15.
79. Yetim B, Celik Y. What predicts the quality of life in Turkey: the results of a country representative study. Health Soc Care Commun. 2020;28(2):431–8.

Publisher's Note