IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 October 2024, accepted 22 December 2024, date of publication 25 December 2024, date of current version 6 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3522767

RESEARCH ARTICLE

Predicting Patients' Revisit Intention Based on Satisfaction Scores: Combination of Penalized Regression and Neural Networks

FARSHID ABDI[1], SHAGHAYEGH ABOLMAKAREM[1], AMIR KARBASSI YAZDI[2], PAUL LEGER[3], YONG TAN[4], AND GIULIANI COLUCCIO[2]

[1]Department of Industrial Engineering, Islamic Azad University, South Tehran Branch, Tehran 1584715414, Iran
[2]Departamento de Ingeniería Industrial y de Sistemas, Facultad de Ingeniería, Universidad de Tarapacá, Arica 10000, Chile
[3]Escuela de Ingeniería, Universidad Católica del Norte, Coquimbo 1781421, Chile
[4]School of Management, University of Bradford, BD7 1DP Bradford, U.K.

Corresponding author: Amir Karbassi Yazdi (akarbassiy@academicos.uta.cl)

ABSTRACT A company's survival in the current competitive market hinges on its ability to not only meet but exceed customer expectations, as customers are invaluable assets. Patient satisfaction is crucial in the healthcare sector, directly influencing whether patients will return to a hospital or recommend it to others. This study uses advanced data mining techniques to accurately estimate and predict patients' likelihood of returning for future appointments by assessing their satisfaction levels. In addition to feature selection models such as Random Forest, Genetic Algorithm, and Lasso Regression, the study employs various methods, including Neural Networks, Support Vector Machines, Decision Trees, k-Nearest Neighbors, Rule-based systems, and Naive Bayes algorithms. The analysis of the results indicates that while the Neural Network model shows superior prediction accuracy, the Lasso Regression method is efficient in identifying relevant features. By integrating AI approaches and thoroughly examining satisfaction ratings in the Iranian healthcare industry, this research makes a significant contribution. Moreover, the findings demonstrate that the Artificial Neural Network model best fits the predictive model and offers the highest reliability. This study aims to forecast patient satisfaction in the healthcare industry and develop a strategic roadmap for hospitals, thereby expanding the knowledge of machine learning methods for predicting customer satisfaction.

INDEX TERMS Data mining, feature selection, patients return, satisfaction.

I. INTRODUCTION

Media has extensively covered the privatization of public hospitals in recent years. Privatization entails decentralization [1]. Officials support hospital privatization and decentralization in the long term, believing that consolidation fosters competition among public hospitals, thereby enhancing patient care. Measuring patient satisfaction is crucial for both administrators and researchers, as scholars debate customer satisfaction in bureaucratic organizations. For healthcare administrators, patient happiness is paramount when planning and evaluating services [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

Surveys are instrumental in gathering feedback on hospital care. These can be conducted via paper, phone, or online methods and often include Likert scales, open-ended questions, and rating questions. Survey data helps hospitals pinpoint areas for improvement. Strategies such as staff training, process enhancements, and facility upgrades are employed to boost patient satisfaction. Patterns in patient feedback are analyzed to identify strengths and weaknesses, and satisfaction scores are used to gauge hospital performance, highlighting areas where patient expectations are unmet. Quality improvement activities in hospitals are grounded in survey results and competitor modeling. These activities may encompass training for staff communication, refining hospital processes, and upgrading facilities.

Hospitals prioritize patient preferences and needs, focusing on empathy, communication, shared decision-making, and respect for patient autonomy [7]. Technology plays a significant role in enhancing the patient experience and satisfaction, with tools such as electronic health records (EHRs), online scheduling, and patient portals contributing to improved care [8]. Feedback is gathered through digital surveys and interactive kiosks, providing real-time insights that enhance patient care and experience [9]. Combining various methods to gain patient experience insights allows hospitals to continuously monitor and improve service delivery based on feedback [10]. In the service industry, success hinges on customer retention, and many studies predict customers' return intentions [11]. Multiple models forecast consumer behavior, and machine learning techniques are effectively used to address customer-centric issues. Data mining tools ensure organizational sustainability by predicting client intentions.

Patient satisfaction boosts trust and health, enhancing the healthcare facility's growth, reputation, and patient retention. The financial performance of healthcare organizations is impacted by patient satisfaction; low satisfaction scores can harm their finances. Patient satisfaction differentiates healthcare providers and attracts new business, with online feedback driving client growth for medical facilities. Patient-centered healthcare prioritizes patient satisfaction for legitimacy and credibility, with treatment quality improved through patient involvement. Iran's healthcare system has three tiers, detecting both the largest and smallest community groups. Community Health Centers (CHCs) provide services such as posts and health houses in both rural and urban areas. Patient happiness is a significant healthcare issue, relying on meeting their expectations. Survival depends on exceptional service to ensure patient satisfaction and loyalty, which in turn boosts hospital competitiveness. Patient satisfaction is crucial for a hospital's success, as returning patients provide free advertising through referrals.

Business performance has been significantly improved by AI and ML technology. Algorithms are essential for AI systems' forecasting, recommending, and decision-making. AI systems use machine learning to generate knowledge and recognize data patterns. Machine learning, a type of AI, solves complex problems without the need for expensive human-developed algorithms, requiring machines to independently discover solutions. Initially, the validation of solutions was necessary. Solfa et al. [21] explore the use of machine learning in healthcare and patient outcome prediction. Patient outcomes and healthcare performance benefit from big data analytics in machine learning, improving patient outcomes and personalized treatments. Big data analytics helps reduce costs, improve efficiency, and allocate resources in healthcare. Kunze et al. [22] used machine learning to predict patient dissatisfaction post knee surgery. The random forest method outperformed others (c-statistic: 0.77, calibration slope: 0.74, calibration intercept: 0.087, Brier score: 0.082), accurately predicting dissatisfaction and maximizing preoperative health. Polce et al. [23] used machine learning to predict satisfaction after shoulder arthroplasty, with the support vector machine model performing best (c-statistic: 0.80). An open-access tool and the best SVM model achieved remarkable predictions for satisfaction after TSA. Five algorithms were used to forecast enhanced patient satisfaction post-hip arthroscopy using machine learning techniques, predicting satisfaction in a patient group post-hip arthroscopy. External verification of algorithm performance is necessary. Simsekler et al. [24] examined patient satisfaction factors using a random forest. Random forests analyze patient/provider characteristics and satisfaction levels during registration and consultation, highlighting significant factors for patient satisfaction. The model is useful for healthcare practitioners, managers, and academics.

Current studies predict patient intentions to return using structural equations (see [25]). Some studies did not employ machine learning for predicting patient return intentions. This study uses data mining techniques to forecast hospital readmissions by considering patients' goals. Various factors influence customer decisions regarding products or services, with user-centric aspects determining customer usage [26]. This study proposes a patient satisfaction forecasting model that incorporates patient attributes and satisfaction metrics. Effective planning relies on understanding patient satisfaction. Our research questions are: "How can we predict patients' willingness to return to the hospital?" and "How do we identify factors affecting patient willingness to return to the hospital?" The research contributions are divided into two categories: 1. Theoretical Contribution: The study combines Genetic Algorithms, Lasso regressions, and Random Forest (RF) methods. The comparison of these methods determines improvements in reliability, assessed through method comparison. An Iranian hospital examines the effectiveness of these three methods. This study aims to comprehensively evaluate patient satisfaction across all hospital types. 2. Practical Application: Exact and comprehensive input data is essential for proprietary models. Data inaccuracies reduce the credibility of predictions and conclusions. This study explores patient behaviors within the Iranian healthcare system, often disregarding interpretability in its machine learning techniques. Understanding patient satisfaction in complex models is challenging. Feature selection affects model performance and interpretation. Data analysis is used to estimate patient satisfaction and develop a hospital plan. This study does not consider other variables affecting patient satisfaction.

Ethics play a role in predicting patient behavior, with healthcare research credibility relying on ethical patient data usage. The study lacks an assessment of long-term effects regarding predictive analytics recommendations. We evaluate the model's impact on patient happiness and healthcare outcomes. The study examines patient satisfaction in the Iranian healthcare system, considering local factors. Further research is needed to explore the influence of cultural, social, and economic factors on patient satisfaction in various healthcare settings. The study focuses only on patient return, neglecting long-term medical outcomes. Satisfaction ratings overlook qualitative data and patient input, although patient feedback can enhance healthcare services. The study's limited applicability is due to its exclusive focus on the Iranian healthcare context. Additional research is required to validate predictive models in diverse healthcare settings. This paper analyzes ethical concerns regarding the use of patient data for predictive analytics in healthcare. The paper is organized into six parts as follows: 1) Introduction: Examines the contribution, research gap, and research questions; 2) Literature Review: Covers the relevant literature on the topic; 3) Research Methodology: Provides an overview of the employed methods; 4) Data Analysis and Discussion: Presents data analysis and discussion; 5) Managerial Implementation: Details the practical application of the research; 6) Conclusion: Discusses conclusions, future research, and limitations.

II. LITERATURE REVIEW ON COSTUMERS' RETURN INTENTION

Previous researchers have investigated the level of satisfaction and willingness of customers to use a service again, as well as their behavioral intentions to do so. This section presents a review of previous studies conducted in the field of customer satisfaction, loyalty, and return intentions. Kumar et al. [27] conducted a study on Artificial Intelligence (AI)-enabled Customer Relationship Management (CRM) capability in healthcare and its effect on service innovation. The researchers used a mixed-method approach to study this phenomenon. They explored how healthcare in India uses AI-enabled CRM to enhance service innovation through resource-based theory, dynamic capability theory, and productivity paradox theory. The study identified AI-enabled CRM capabilities and developed a framework for service innovation. Customer Service Flexibility (CSF) is a crucial aspect that needs attention in this relationship. The quantitative study employed Partial Least Squares Structural Equation Modeling (PLS-SEM) to estimate cause-and-effect relationships in path models with latent variables, addressing a research gap and driving performance in healthcare. Uzir et al. [28] examined the use of smartwatches in healthcare in Bangladesh during COVID-19 and the impact of artificial intelligence on user satisfaction. This study analyzed AI-powered smartwatches for health using the Stimulus-Organism-Response (S-O-R) framework. According to the theory, users' trust, satisfaction, and experience are influenced by product quality, service quality, perceived convenience, and perceived ease of use. Data from 486 smartphone users in Bangladesh were analyzed to test hypotheses through elementary analyses and PLS-SEM. The findings indicated that product quality, service quality, perceived convenience, and perceived ease of use significantly influenced user experience and trust in the product or service. User satisfaction was influenced by user experience and trust, which partially mediated the relationship between predictors and satisfaction. Gender and age moderated the relationship between experience, trust, and customer satisfaction. Simsekler et al. [24] state that patient satisfaction is a multidimensional concept that provides insight into various quality aspects of the healthcare industry. By applying the random forests algorithm, the authors estimated relationships between patient and provider-related determinants and the level of satisfaction. Van Onsem et al. [29] proposed a prediction model for patient satisfaction after total knee arthroplasty based on regression analysis. Damayanti & Kusumawardani [30] proposed predicting patients' intention to return to the clinic by examining their perception of service quality and their satisfaction level. Structural Equation Modeling (PLS-SEM) was applied to test the proposed structural relationships. According to the authors, satisfaction and perceived service quality both significantly influence patient revisit intentions, with satisfaction being cited as the most important factor.

Calisir et al. [31] investigated the effects of service quality dimensions on customer satisfaction and the return intention of patients in a research study. Four hospitals were evaluated for their service quality levels using a modified SERVQUAL scale developed by the authors. Regression analysis was used to predict the patient's intention to return. Woo and Choi [32] used structural equation modeling to assess the direct effect of perceived healthcare quality on patient satisfaction and intention to return to the hospital, both in inpatients and outpatients of regional hub public hospitals. They also examined the indirect effects of healthcare quality on the intention to return via patient satisfaction. Lai et al. [33] examined the effects of perceived price and perceived service quality on patients' revisit intentions to hospitals. They considered revisit intention as the dependent variable and used the SERVQUAL model as the independent variable to examine the five-dimensional effects of service quality on patients' expectations. Structural equation modeling was employed to analyze the structural relationships between the independent and dependent variables. Mohd Isa et al. [34] used partial least squares structural equation modeling to investigate determinants of patients' intent to revisit private hospitals, including hospital image, perceived medical quality, relationship marketing, and word-of-mouth. Loureiro [35] examined the effect of hospital image on customers' emotions, perceived quality, image, and feelings of pleasure derived from the service environment it offers. Cham et al. [36] studied the relationship between word-of-mouth and social media perceptions of service quality, satisfaction, and behavioral intentions regarding hospital brands. Seow et al. [37] analyzed perceived benefits, perceived costs, and the availability of resources, intending to revisit the topic. These studies showed a significant association between hospital image and patient satisfaction and behavioral intentions. Suhail & Srinivasulu [38] analyzed differences in perception among healthcare consumers, examining the relationship between service quality, satisfaction, and behavioral intentions. They used analysis of variance (ANOVA) and a t-test to evaluate perception differences and multiple regression analysis and structural equation modeling to propose two relationship models based on the collected data. Al-Refaie [39] identified factors such as hospital performance, hospital stay, hospital facilities, interaction with patients, service quality, and patient safety culture that significantly affect patient satisfaction and Patients' Revisit Intention using structural equation modeling. Table 1 shows the previous studies.

TABLE 1. Previous study.

Author/s	Method/s
[27]	Reference examined the emergence and facets of AI-CRM capabilities in healthcare. The results indicated a linear relationship between AI-CRM capability, CSF and service innovation.
[28]	use of smartwatches in healthcare in Bangladesh during COVID-19 and the impact of artificial intelligence on user satisfaction
[38]	Analysis of variance (ANOVA), t-test, multiple regression analysi, structural equation modeling
[24]	Tree-based machine-learning algorithm, random forests
[30]	Structural relationships, structural equation modeling
[31]	Modified SERVQUAL
[32]	structural equation modeling
[33]	SERVQUAL
[34]	partial least squares structural equation modeling
[29]	Regression analysis
This research	NN, SVM,KNN, GA, DT, Rule based and NB methods

III. RESERCH METODOLOGY

In this section, the theoretical framework of the study and the proposed research framework are presented. The theoretical framework contains explanations of each algorithm used in the article. Applied feature selection methods such as LASSO regression, Genetic Algorithm (GA), and Random Forest (RF) are described separately. Additionally, a brief explanation of each algorithm used to classify patients' revisit intentions and methods for validation and evaluation of the proposed models is provided. Finally, the proposed research framework is described.

Predictor-response relationships in traditional regression models are assumed to be linear. However, real-world relationships often show complex or non-linear connections. Machine Learning (ML) models are adept at capturing complex and non-linear patterns [40]. For feature engineering, regression models often require the manual selection and manipulation of predictor variables based on domain expertise. Machine learning models can eliminate the need for manual engineering by learning features and interactions from data, identifying subtle patterns. Tree-based ML models are more robust against noise and outliers. Traditional regression models cannot handle missing values without imputation or deletion, necessitating complete datasets. ML models offer flexible preprocessing and can manage missing data more effectively. Big data is efficiently handled by machine learning techniques within distributed computing frameworks such as TensorFlow and Apache Spark. Regression models have restricted assumptions and model structures. Unlike machine learning models, which effectively capture non-linear or complex interactions, regression approaches may compromise prediction accuracy in these situations. Machine learning algorithms excel at identifying subtleties and enhancing forecasts.

A. THEORETICAL FRAMEWORK OF THE STUDY

This section provides a brief explanation of the algorithms used in this study.

1) FEATURE SUBSET SELECTION

Dataset preprocessing is crucial for preparing data for classification and improving forecast accuracy. Feature selection is a data processing technique typically applied to datasets with many variables to reduce dimensionality. If the number of features is large, selecting a subset of effective features is recommended before classification or using other appropriate techniques. This approach not only accelerates computations but also enhances classification accuracy.

Feature Subset Selection using Lasso Regression: The Lasso method is a type of linear regression used for both variable selection and regularization to enhance prediction accuracy. This method performs variable selection by shrinking the parameter coefficients in linear regression through the imposition of an L1 penalty. Regularization with L1 entails a constraint $\sum_{j=1}^{p} \left|\beta_j\right| \leq t$, where t≥0, on the objective function. This penalty facilitates automatic variable selection, allowing some coefficient values to reach zero, thereby effectively removing variables with low impact from the model. The parameter estimation solution with L1 regularization, as found by Lasso, is represented by Eq. (1) [36].

$$\arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \beta^T X_i\right)^2 + \lambda \left\|\beta\right\|_1 \tag{1}$$

where $\|\beta\|_1 = \sum_{j=1}^{d} \left|\beta_j\right|$ is the 1-norm of the vector β. The penalty $\lambda \|\beta\|_1$ is called the L_1 penalty. The LASSO method imposes a constraint on the sum of the absolute values of the model's parameters; this sum must be less than a fixed value (upper bound) for the model to be valid. A shrinking (regularization) process is applied to achieve this. In this process, the coefficients of the regression variables are penalized, and some are shrunk to zero. When selecting model features, the variables that still have non-zero coefficients after the shrinking process are chosen to be part of the model. The goal of this process is to minimize prediction error. Undoubtedly, the tuning parameter λ, which controls the strength of the penalty, plays an important role in practice. When λ is sufficiently large, the coefficients are forced to be exactly zero, thereby reducing dimensionality. As λ increases, more coefficients tend to shrink to zero. In contrast, in ordinary least squares (OLS) linear regression, there is no such constraint.

Random Forest-based Variable Importance Measures (VIMs):

Random Forest (RF) is a commonly used and highly efficient ensemble algorithm for both classification and regression problems. Let $L = \{(X_1.Y_1).(X_2.Y_2)\ldots.(X_n.Y_n)\}$ be a training set of nn i.i.d. observations of a random vector $(X.Y)$. Vector $X=(X^1.X^2.\ldots.X^n)$ contains predictor variables, and YY is either a class label or a numerical response. The Random Forest model combines a large number of decision trees constructed using several bootstrap samples from the training sample LL and randomly selecting, at each node, a subset of predictor variables XX to be used as predictors. The RF method is also used to determine the importance of variables. The most popular Variable Importance Measures (VIMs) include the Gini-based VIM (GVIM), the permutation-based VIM (PVIM), and the conditional permutation-based VIM (CPVIM). RF variable importance of X_jXj is defined as follows:

$$VI\left(X^j\right) = \frac{1}{ntree}\sum_t \left(err\widetilde{OOB}_t^j - errOOB_t\right) \qquad (2)$$

where OOB_t is considered as the associated out-of-bag sample (data not included in the bootstrap sample used to construct tree t). $errOOB_t$ denotes the error of a single tree t (Mean Squared Error for regression and misclassification rate for classification) on thisOOB_t sample.$\widetilde{OOB}_t^j$ denotes the perturbed sample obtained from randomly permuting the values of X^j in OOB_t. $err\widetilde{OOB}_t^j$ *denotes the error of predictor t*t *on the perturbed sample. The sum is over all trees t*t *of the Random Forest (RF), and n* denotes the number of trees of the RF [41], [42].

Genetic Algorithm-Based Feature Subset Selection (FSS): A Genetic Algorithm (GA) is an optimization method that employs the principles of evolutionary theory. GA is capable of efficiently searching large solution spaces and can be utilized to solve feature selection problems [43]. Typically, a genetic algorithm begins by generating a random initial population of chromosomes. Each chromosome in the population is then assigned a fitness value based on its performance in the previous step. Fitness functions are used in evolutionary processes to evaluate the quality of a solution to a problem. The initial population undergoes various genetic operators, including crossover, mutation, and selection. In crossover, two individuals are combined to form a crossover offspring. Mutation involves perturbing the genes in each chromosome by flipping bits according to the probability of gene mutation. An optimal chromosome solution is one that achieves the best fitness value after several generations. Once the termination criterion is met, the genetic algorithm terminates [44].

Convolutional Neural Networks (CNNs)-based Feature Subset Selection Feature subset selection in datasets using Convolutional Neural Networks (CNNs) is crucial for improving model performance by identifying the most relevant features for classification tasks. This process typically involves several key steps: first, features are extracted using a CNN, which is adept at automatically learning complex patterns from raw data. After extraction, various feature selection techniques are applied to evaluate and choose a subset of these features. These techniques can be categorized into three main types: filter methods, wrapper methods, and embedded methods. Filter methods assess feature relevance based on statistical measures without involving the learning algorithm, while wrapper methods evaluate subsets by training the model and measuring performance, which can be computationally expensive. Embedded methods incorporate feature selection as part of the model training process itself [45].

Feature Subset Selection using Grey Wolf Optimizer (GWO) Feature subset selection using the Grey Wolf Optimizer (GWO) is a nature-inspired algorithm that mimics the social behavior and hunting strategies of grey wolves. Initially, a population of wolves, each representing a potential subset of features, is randomly generated. The fitness of each wolf is evaluated based on a predefined criterion, typically the performance of a predictive model trained on the selected features. The wolves are ranked according to their fitness, with the best-performing wolves leading the pack. In each iteration, the positions of the wolves are updated based on the alpha, beta, and delta wolves, which represent the top three solutions. This update mechanism incorporates exploration and exploitation strategies, allowing the algorithm to search for optimal feature subsets effectively. The process continues until a stopping criterion is met, such as a maximum number of iterations or convergence of the wolves' positions. Ultimately, GWO aims to identify a compact set of features that enhances model accuracy while reducing dimensionality, thus improving computational efficiency and interpretability [46].

2) PATIENTS' REVISIT INTENTION CLASSIFICATION MODEL

In this study, Artificial Neural Networks (ANN), Support Vector Machines (SVM), N-Nearest Neighbors (KNN), Decision Trees (DT), Rule-Based (RB), and Naïve Bayes (NB) methods have been utilized for modeling. Below is a brief description of each of these classification algorithms:

Artificial Neural Networks (ANN): Artificial neural networks (ANN) are one of the most commonly used methods for both classification and regression problems. In classification scenarios, the output variable is categorical (often binary), and the input vector or pattern is trained to select a class from a list of possibilities. Figure 1 depicts a typical neural network structure for a classification problem. The network consists of three types of layers: an input layer, hidden layers, and an output layer. Each input variable has a neuron in the input layer. The hidden layer represents the nonlinear relationship between variables and completes the data processing. There can be any number of neurons in each hidden layer. The output layer produces the final result, either producing the categorical class label or predicting continuous target indices. Input data typically undergo normalization. The two most common normalization methods are MinMaxScaler and StandardScaler. MinMaxScaler scales the data to a fixed range, usually between 0 and 1, while StandardScaler rescales the data to have a mean of 0 and a standard deviation of 1, resulting in a distribution with unit variance and a zero mean value. Each hidden layer receives input from the previous layer and provides output to another hidden layer or the output layer. Various nonlinear activation functions such as ReLU and Sigmoid can be used in the hidden layers. For classification problems, softmax and sigmoid activation functions are commonly used in the output layer [47].

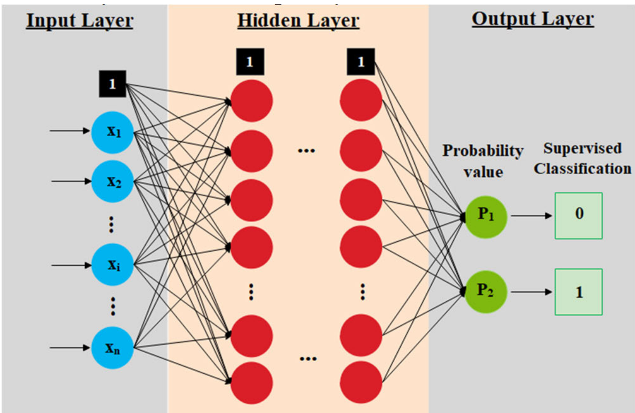


FIGURE 1. Structure of ANN for classification.

Support Vector Machines (SVM): For a given dataset $D = \{x_i, y_i)\}_{i=1}^{n}$, where x_i denotes the input sets and $y_i \in \{-1, +1\}$ for $i = 1, 2, \ldots, n$ are the matching output values. Each data record belongs to one of the two classes separated by a linear classifier with a hyperplane. It is necessary to choose the hyperplane with the maximum margin between the two classes for the best classification. The hyperplane that separates the data is given by (3).

$$f(x) = w^T x + b = \sum\nolimits_{j-1}^{n} w_i x_j + b = 0 \tag{3}$$

where w is the J-dimensional vector and b is the scalar used to define the position of the separating hyperplane. Thus, the width of the margin is $\frac{2}{|w|}$, maximizing $\frac{2}{|w|}$, which means minimizing $\frac{1}{2}\|w\|^2$. The best hyperplane can be obtained by minimizing (4).

$$\frac{1}{2}w^2 + c\sum\nolimits_{i-1}^{J} \xi_i$$
$$\text{Subject to,}$$
$$\{y_i((w^T x + b) \geq 1 - \xi_i)\}(i = 1, 2, \ldots, JM) \tag{4}$$

where ξ_i denotes the slack variable, and c is the penalty for error [42].

K-Nearest Neighbors (KNN): The K-Nearest Neighbor (KNN) algorithm is one of the widely used methods for classification. KNN could be described as learning by similarity. Test records are described using n features, and each record is actually a point in an N-dimensional space. The class of an unknown and unlabeled record is determined according to the classification of the nearest K neighbors. Most often, more than one neighbor is taken into consideration. KNN usually applies Euclidean distance to determine the similarity between the training records and the test record [43].

Decision Trees (DT): The decision tree (DT) is one of the most powerful and common tools for classification and prediction. A DT generally has a structure like a flowchart, where each internal node contains a test about one particular attribute, and every branch shows the outcomes of the test; each terminal leaf is labeled with one of the possible classes. The highest node in a tree is called the root. Leaf nodes represent class distributions. Classification for a given sample record, such as X, whose label is completely unknown and whose attribute values are tested in the decision tree, is done using the decision tree in such a way that, given the attribute values of X, a path is traced from the root to a leaf node that holds the class prediction for the given sample record [48] [49].

Rule-Based (RB): Rule induction algorithms often apply the sequential covering strategy and learn a list of rules from the training data sequentially, or one by one. When the algorithm learns a new rule, it eliminates the corresponding training records that it covers, that is, those training examples that satisfy the rule antecedent. This learning process is repeated until rules can cover the whole training data or no new rule can be learned from the remaining training data [49].

Naïve Bayes (NB): A Naive Bayes algorithm is a probabilistic classifier based on Bayes' theorem. The NB model is easy to construct, without complicated iterative parameter estimation, making it useful in several fields. This algorithm calculates the conditional probabilities between the input variables and the target and determines which input properties are most likely to play a role in predicting the target variable. The basis of the Bayesian classification algorithm is Bayes' theorem. Mathematically, Bayes' theorem is stated as:

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)} \tag{5}$$

where A and B are two independent events. The probability of event A is denoted by $P(A)$, and the probability of event B is denoted by $P(B)$, respectively. $P(A \mid B)$ denotes the conditional probability of A given B, while $P(B|A)$ is the probability of event B with respect to event A [43].

3) VALIDATION AND EVALUATION OF RESULTS

Validation of Results: The k-fold method is applied to validate the model used. In this method, the entire dataset is divided into k batches, with one batch used as the test set while the remaining k-1 batches are used as the training set. This process is repeated k times, with each batch serving as the test set once. Finally, the average difference of the k obtained results is declared as the error. In this research, k is considered equal to 10.

Evaluation of Results: In this section, modeling should be examined using model evaluation metrics. For instance, in this research, evaluation metrics such as Accuracy, Precision, and Recall have been utilized to determine which classification method is more accurate compared to others. A matrix called the confusion matrix is used to evaluate the classification algorithm, as defined in Table 2. Accuracy, Precision, and Recall metrics are also calculated using equations 6, 7, and 8, respectively.

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

The research methodology is discussed in this section.

TABLE 2. Confusion matrix.

		Prediction	
		Positive Class	Negative Class
Real value	Positive Class	True Negative (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

B. PROPOSED FRAMEWORK

1) DATASET DESCRIPTION

This study utilized a dataset containing information about 340 patients referred to a private hospital in Iran for medical treatment. The data was collected over a period of 3 months. At the studied hospital, upon discharge, patients are asked to complete a questionnaire to assess their satisfaction with the services provided. Each question in the questionnaire was considered a research variable, as listed in Table 3. To assess the quality of hospital service, two sets of variables were collected based on the demographic characteristics of the patients and their ratings of the hospital. The first set includes demographic information such as age, education, gender, insurance status, type of admission, and discharge, while the second set pertains to various aspects of service provision. These variables include patient satisfaction with factors such as the speed of admission, availability of nursing staff, quality of physician care, nutrition, room ventilation, ward comfort, hospital amenities, and the performance of paraclinical units.To mitigate response bias, we designed our survey to encourage honest feedback from all participants, minimizing the influence of extreme experiences on overall satisfaction scores. Additionally, we employed stratified sampling to capture a diverse demographic representation, thereby reducing sampling bias. We conducted pilot testing with a small group of participants to refine the survey questions and enhance clarity and relevance. Furthermore, participants were blinded to the specific hypotheses and objectives of the study, which helped reduce expectations and promote more candid responses. This approach prevented participants' expectations from influencing their feedback.

The distribution of satisfaction scores from 340 patients across ten variables (V7-V16) reflects a range of perceptions regarding hospital service quality. For most variables, a significant number of patients rated their satisfaction as 5 (Very High Satisfaction), indicating strong approval of services such as "Speed of File Formation," "Transfer to the Ward," "Performance of the Treating Physician," and "Nutrition Status." However, some variables, like "Timely Presence of Nursing Staff" and "Welfare Facilities," showed more varied distributions, with a notable percentage of patients expressing dissatisfaction through lower scores, highlighting areas requiring attention and improvement.

Overall, the descriptive analysis of satisfaction score distributions provides valuable insights into the strengths and weaknesses of hospital services. High scores in areas like "Sanitary Status" and "Room Ventilation Status" suggest patients feel comfortable and well cared for, while lower scores in aspects such as "Calmness in the Ward" and "Providing the Necessary Training" indicate areas for potential improvement. By examining these distributions, healthcare providers can identify specific areas for enhancement, ultimately aiming to improve patient satisfaction and the overall quality of care delivered.

Table 3 outlines 17 variables, of which 16 are independent or predictive variables, and "Intent to revisit" represents the target variable. In this context, "Intent to revisit" refers to a patient's willingness to return to the same hospital for future healthcare needs. This research aims to create a binary variable for analysis purposes. Variables 7-16 pertain to indicators measuring patient satisfaction, with patients rating each aspect on a scale from 1 to 5, indicating very low, low, medium, high, and very high satisfaction levels, respectively. The "Age" variable is numerical but has also been converted into a categorical variable for analysis.

Every individual in Iran has a one-of-a-kind identification number that must be shown whenever they visit a location or seek any service from any type of establishment. Additionally, the individual or company must register this identification number in their systems to complete any kind of procedure, and service providers are required to provide their identification to work. In light of the previous arguments, this procedure guarantees that only patient responses will be provided to this specific inquiry.

In the following, a brief description of each of the satisfaction indicators used in this research is presented:

1. Speed of File Formation and Transfer to the Ward: Efficient administrative processes, including file formation and transfer, contribute to seamless patient care delivery. Timely completion of paperwork and smooth transitions between departments enhance the patient experience and satisfaction.
2. Timely Presence of Nursing Staff: Patients rely on nursing staff for assistance, medication administration, and emotional support. A timely presence ensures a prompt response to patient needs, promotes safety, and enhances trust and satisfaction with the healthcare facility.

TABLE 3. Variable used in the research.

Row	Variable	Name	Type of Variable	Role of Variable
1	V1	Age	Numeric	Predictor
2	V2	Gender	Binomial	Predictor
3	V3	Education	Categorical	Predictor
4	V4	Insurance	Binomial	Predictor
5	V5	Admission type	Categorical	Predictor
6	V6	Demission type	Binomial	Predictor
7	V7	Speed of file formation and transfer to the ward	Categorical	Predictor
8	V8	Timely presence of nursing staff	Categorical	Predictor
9	V9	Providing the necessary training	Categorical	Predictor
10	V10	Performance of the treating physician	Categorical	Predictor
11	V11	Nutrition status	Categorical	Predictor
12	V12	Room ventilation status	Categorical	Predictor
13	V13	Calmness in the ward	Categorical	Predictor
14	V14	Welfare facilities	Categorical	Predictor
15	V15	Sanitary Status	Categorical	Predictor
16	V16	Para-clinical units	Categorical	Predictor
17	V17	Y: Intent to revisit	Binomial	Target

3. Providing Necessary Training: Educating patients and their families about their condition, treatment plan, and self-care measures empowers them to actively participate in their own healthcare. It can lead to better treatment adherence, reduced readmission rates, and increased patient satisfaction.
4. Performance of the Attending Physician: This indicator typically refers to patients' perceptions of how well their doctor communicated, listened, and provided care during their treatment. It reflects patient satisfaction with the doctor's expertise, empathy, and overall quality of care. This indicator is crucial because satisfied patients are more likely to comply with treatment plans, have better health outcomes, and positively impact the hospital's reputation. Additionally, it helps healthcare providers identify areas for improvement and enhance the patient experience.
5. Nutrition Status: Proper nutrition is vital for patient recovery and overall well-being. Monitoring and ensuring adequate nutrition can speed up recovery, prevent complications, and enhance patient satisfaction.
6. Room Ventilation Status: Proper ventilation in hospital rooms is crucial for infection control and maintaining a comfortable environment for patients. Good ventilation reduces the risk of airborne pathogens, odors, and discomfort, leading to higher patient satisfaction and safety.
7. Calmness in the Ward: A calm and peaceful environment in hospital wards promotes healing and comfort for patients. It reduces stress, anxiety, and the perception of pain, contributing to overall satisfaction with the hospital stay.
8. Welfare Facilities in Hospitals: Encompassing aspects like comfortable accommodations for staff, access to amenities, and support services, these indicators are vital as they directly impact the well-being and morale of healthcare professionals. By providing adequate welfare facilities, hospitals can enhance staff satisfaction, reduce burnout, improve retention rates, and ultimately contribute to a positive work environment conducive to delivering quality patient care.
9. Sanitary Status: This indicator assesses patients' perceptions of the cleanliness and hygiene levels within the hospital environment, including patient rooms, bathrooms, waiting areas, and common areas. It reflects how well the hospital maintains cleanliness standards to prevent the spread of infections and ensure a safe environment for patients, visitors, and staff. This indicator is crucial for several reasons, such as infection control, patient safety, perception of care quality, and employee and patient morale.
10. Performance of Para-clinical Units: Efficient functioning of para-clinical units such as laboratories, radiology, and pharmacy is essential for timely diagnosis, treatment, and medication provision. Delays or errors in these units can negatively impact patient outcomes and satisfaction.

The range of patient perceptions regarding the quality of hospital services is reflected in the distribution of satisfaction scores, which were based on responses from 340 patients across ten variables (V7-V16). A considerable proportion of patients rated their satisfaction as 5 (Very High Satisfaction) for most variables, indicating a strong endorsement of services such as "Speed of File Formation," "Transfer to the Ward," "Performance of the Treating Physician," and "Nutrition Status." Nevertheless, certain variables such as "Timely Presence of Nursing Staff" and "Welfare Facilities" exhibited a diverse distribution, with a considerable proportion of patients expressing dissatisfaction through lower scores. This highlights the necessity for focus and enhancement in these domains.

In general, the descriptive analysis of satisfaction score distributions offers valuable insights into the strengths and weaknesses of hospital services. The presence of high scores in categories such as "Sanitary Status" and "Room Ventilation Status" indicates a sense of comfort and quality care for patients, while lower scores in areas like "Calmness in the Ward" and "Providing the Necessary Training" highlight potential concerns that may affect the patient experience. Through the analysis of these distributions, healthcare providers have the ability to identify precise areas for enhancement, ultimately with the goal of enhancing patient satisfaction and the overall quality of care provided.

Overall, these satisfaction indicators play a critical role in ensuring patient-centered care, improving outcomes, and building trust between patients and healthcare providers.

Monitoring and addressing these indicators can help healthcare facilities continuously improve their services and meet the needs and expectations of their patients.

2) THE PROPOSED MODEL

Figure 2 presents the flowchart of the research. As we can see, the proposed model is a three-stage approach: 1) Data Gathering: This is the first step in the data collection process; 2) Feature Subset Selection: This second step involves implementing feature subset selection. As part of the study, several different variable selection methods have been used, including Lasso regression, Random Forest, and Genetic Algorithms, to select a subset of variables that are indicative of patients' willingness to return; 3) Patients' Revisit Intention Classification Model: In the third stage, building a model of the patient's willingness to return is necessary. During this stage, the selected features are used as input for commonly used machine learning algorithms, namely Neural Networks (NN), Support Vector Machines (SVMs), K-Nearest Neighbors algorithm (KNN), Decision Tree algorithm (DT), Rule-based algorithm (RB), and Naive Bayes algorithm (NB). At the end of the process, evaluation metrics such as accuracy, precision, and recall are used to compare the performance of the models, and the best one is selected based on its performance. Figure 2 illustrates the process of this research.

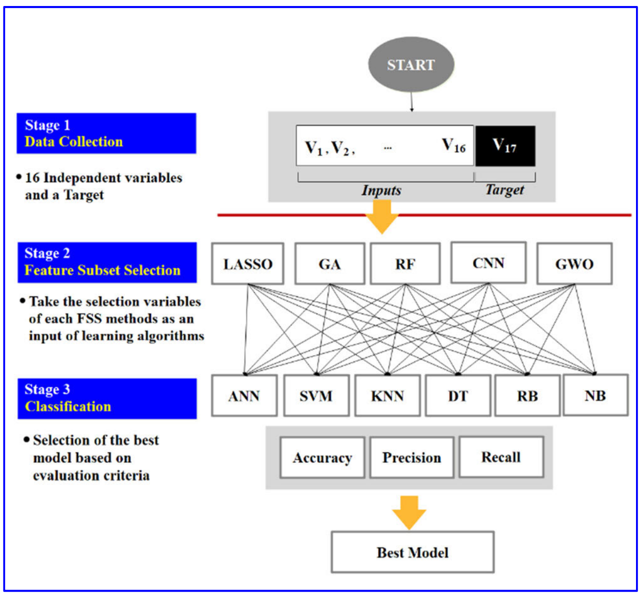


FIGURE 2. Flowchart of the research.

3) BENEFITS OF THE TECHNIQUES

a: FEATURE SELECTION METHODS

LASSO (Least Absolute Shrinkage and Selection Operator) Feature Selection Techniques:

LASSO penalizes large coefficients to prevent overfitting and create interpretable models. It produces sparse solutions with improved generalization performance. LASSO induces sparsity, focusing on significant features for easier interpretation.

Random Forest:

Random forests combine decision trees to improve prediction accuracy and reduce overfitting. They are less prone to overfitting compared to individual decision trees and provide a measure of feature importance, indicating each feature's contribution to the model's predictive ability. Random forests can handle missing data without requiring imputation.

Genetic Algorithms:

Genetic algorithms are useful for solving complex global optimization problems by exploring various potential solutions and avoiding local optima. They are applicable to a wide range of optimization issues, including feature selection and model optimization. Genetic algorithms can be parallelized to explore multiple solutions simultaneously, thereby speeding up convergence. They are capable of optimizing problems without the need for gradient information.

b: BENEFITS OF THE MODELING TECHNIQUES

Artificial Neural Networks (ANNs) are ideal for capturing complex data patterns due to their ability to capture non-linear correlations. They are versatile and can be used for pattern recognition, regression, and classification tasks. ANNs handle noisy data, including missing values, due to their distributed nature and ability to generalize from examples. Deep learning architectures, a subclass of ANNs, learn hierarchical data representations without requiring human feature engineering.

Support Vector Machines (SVMs): SVMs excel in high-dimensional feature spaces and utilize kernel functions to manage non-linear decision boundaries, recognizing intricate patterns in data. They optimize class margins, reducing overfitting in high-dimensional fields, and prioritize the global optimum over local optima, resulting in more resilient models.

K-Nearest Neighbors (KNN): KNN is a simple and suitable approach for classification and regression tasks. KNN is effective for online learning and dynamic situations due to its elimination of the training phase. It is resistant to outliers and noisy data due to nearest neighbor averaging or majority voting.

Decision Trees: Decision trees generate clear and simple models for feature importance and decision-making. They can handle missing and categorical/numerical data directly and capture non-linear correlations in complicated datasets. Decision trees can be integrated into ensemble techniques such as Gradient Boosting Machines and Random Forests, which often perform better than individual decision trees.

Rule-based Approaches: Rule-based approaches create clear and understandable rules that describe decision logic in a human-readable manner. They simplify interpretation and validation by encoding subject knowledge explicitly. Rule-based systems are efficient and flexible, allowing them to handle big datasets and complex decision-making tasks. They also handle noisy data and missing values by considering uncertainty and unpredictability.

Naive Bayes Classifiers: Naive Bayes classifiers are efficient and require minimal memory and training time, making them suitable for big datasets and real-time applications. They are easy to use and require simple probability calculations. Naive Bayes classifiers excel with noisy or incomplete data due to their assumption of feature independence and perform well with small datasets by providing accurate class probability estimates with limited training samples.

IV. RESULTS AND DISCUSSIONS

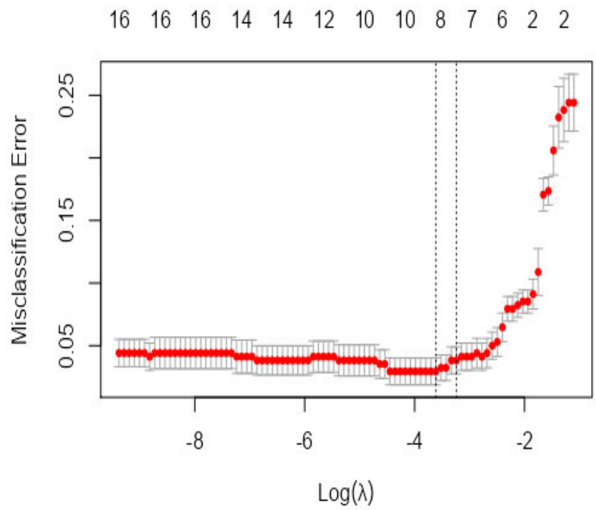
A. FEATURE SUBSET SELECTION

To ensure a more reliable and accurate classification, it is highly efficient to identify the set of appropriate and most influential input variables. We utilized LASSO regression, Genetic Algorithms, and Random Forest methods to determine the importance of available features concerning patients' likelihood to return to the hospital in the future when they require medical services again. The aim of feature selection in this research is to choose a concise subset of features that maximizes the predictive capability of learning algorithms while also avoiding model complexity.

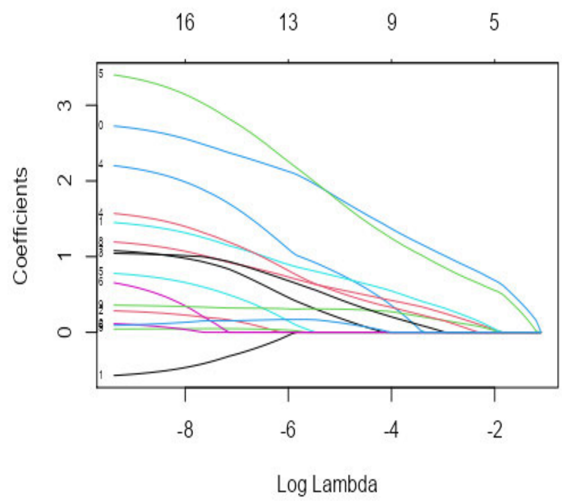
1) LASSO REGRESSION

Ten-fold cross-validation was performed to tune the optimal value of lambda λ that yields the minimum mean cross-validated error. Out of 16 variables, 7 features were selected based on non-zero coefficients. Figure 3 displays the results of the Lasso regression.

In Fig. (3), (a) the tuning parameter λ (lambda) selection in the LASSO model used 10-fold cross-validation based on the minimum criteria. The misclassification error was plotted against log(λ). Dotted vertical lines were drawn at the optimal values determined by the minimum criteria. A λ value of 0.03918369 with log(λ) of -3.239495 was selected through 10-fold cross-validation. (b) LASSO coefficient profiles of the 16 features were generated. A coefficient profile plot was created against the log(λ) sequence. A vertical line was drawn at the value selected by 10-fold cross-validation, resulting in the optimal λ with 7 non-zero coefficients. The variables selected using LASSO regression are presented in Table 7.



(a) Tuning parameter λ



(b) LASSO coefficients

FIGURE 3. Feature selection using LASSO binary logistic regression model.

2) GENETIC ALGORITHM BASES FSS

To implement genetic algorithm-based feature selection, the optimized selection operator (evolutionary operator) provided by RapidMiner Machine Learning Tools is utilized. As shown in Table 4, there are parameters associated with this operator. Depending on the population size, you can specify how many individuals will be generated for each generation. The fitness function used is the classification accuracy of the K-nearest neighbor algorithm. The parents of the next generation are selected through a tournament selection process. Initially, it is assumed that an attribute will be switched on with a probability of 0.5. The mutation rate, set at 0.03, determines the likelihood that a feature will be changed during reproduction. Uniform crossover is employed during reproduction, selecting individuals with a probability of 0.5 for crossover.

The selected variables using Genetic Algorithm are presented in Table 7. As observed, V_1, V_2, V_5, V_7, V_8, V_{10}, V_{12}, V_{13}, V_{14}, *and* V_{15} have non-zero weights.

3) RANDOM FOREST VIMS

To implement random forest-based Variable Importance Measures (VIMs), the R package "randomForest" is utilized. We consider the mean decrease in accuracy over all classes and the mean decrease in Gini index as variables

TABLE 4. The parameter setting of genetic algorithm.

Parameter	Value
Population size	16
Fitness function	K-nearest neighbor
Maximum number of generation	30
Selection scheme	Tournament
p initialize	0.5
p mutation	0.03
p crossover	0.5
Crossover type	Uniform

TABLE 5. The parameter setting of CNN.

Parameter	Value
Learning Rate:	0.001
Batch Size	32
Epochs	5, 10, 15, 20, 30, 40, 50, 60
Early Stopping Patience	5
Permutation Importance Repeats	10, 20, 30, 40, 50

of importance in classification, and the mean decrease in accuracy and the mean decrease in Mean Squared Error (MSE) in regression as variables of importance in regression analysis [50] [51].

The parameter mtry (the number of variables to randomly sample as candidates at each split) specifies the number of input variables randomly chosen at each split, and the number of trees in the forest is determined by the ntree parameter. These parameters are set to their default values (ntree=500 and $mtry = \sqrt{p}$)) for classification, where p is the number of predictor variables in the dataset. The average weights assigned to input variables by Random Forest are presented in Table 7. As observed, variables $V_{10}, V_{15}, and V_{11}$ have the highest weights, respectively.

4) CNN-BASED FEATURE SUBSET SELECTION (FSS)

To implement CNN, the input data is reshaped to fit the expected input format of the CNN. A CNN model is then defined and compiled using the Adam optimizer and a binary cross-entropy loss function. The model is trained with early stopping to prevent overfitting, monitoring validation loss. After training, the model's performance is evaluated using accuracy, precision, and recall metrics. To assess feature importance, permutation importance is calculated on the test set, which measures the impact of each feature on the model's predictions by permuting feature values and observing changes in performance. The CNN parameter settings are presented in Table 5, while the selected variables using CNN are shown in Table 7. As observed, variables $V_{10}, V_{11}, V_9, V_8, V_6, V_{12}, V_7$, and$V_3$ have non-zero weights.

TABLE 6. The parameter setting of GWO algorithm.

Parameter	Value
No. of wolves	5,10,15,20,25
No. of iterations	10,20,30,40,50
Fitness Function:	Accuracy score from the Naive Bayes mode

5) GREY WOLF OPTIMIZER (GWO)

The feature selection process using GWO begins by initializing a population of wolves, each representing a potential subset of features. The fitness of each wolf is evaluated based on the accuracy of a Naive Bayes model trained on the selected features. The wolves are ranked by their fitness values, and their positions are updated according to the positions of the alpha, beta, and delta wolves (the top three solutions). This update mechanism incorporates exploration and exploitation strategies, enabling the algorithm to search for optimal feature subsets effectively. The process continues for a specified number of iterations, and the best-performing wolf (the one with the highest fitness value) is selected as the optimal feature subset. The parameter settings for GWO are provided in Table 6, and the selected variables using CNN are shown in Table 7. As observed,$V_{11}, V_{16}, V_{12}, and V_1$ have non-zero weights.

Fig. 4 shows the importance of each independent variable in predicting patients' willingness to return to the hospital using different feature selection methods, including LASSO, GA, and RF. When the LASSO method is used for feature selection, variables $V_8, V_9, V_{10}, V_{11}, V_{13}, V_{14}, \ and\ V_{15}$ have non-zero coefficients, with V_{10}having the highest weight. V_{15} achieved the next highest rank, with the LASSO method assigning a weight of 0.95137 and the Random Forest method assigning a weight of 23.190336 to this variable. In weighting using the RF method, the same seven variables that have a non-zero weight in the LASSO method are the top seven features with the highest weights. Variables $V_{10}, V_{15}, \ and\ V_{11}$ have the highest weights using the Random Forest method. The next variables we can find $V_1, V_2, V_5, V_7, V_8, V_{10}, V_{12}, V_{13}, V_{14}, and V_{15}$have non-zero weights and were selected using GA methods. Therefore, the variables related to measuring customer satisfaction (patients) in the hospital have a greater impact on the return of patients than demographic variables, among which the performance of the treating physician and the sanitary conditions are the most important factors.

B. PARENTS' REVISIT INTENTION CLASSIFICATION MODEL

This section presents the results of modeling with the selected variables. The variables chosen in the previous stage are assumed to be inputs for Artificial Neural Networks (ANN), Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbors (KNN) algorithms. The k-fold method

TABLE 7. Weights Assigned to Variables based on LASSO, RF, GA, CNN, and GWO methods.

Id	Variable	LASSO	RF	CNN	GA	GWO
V10	Performance of the treating physician	1.101188	31.4651726	0.057190	1	0
V15	Sanitary Status	0.95137	23.190336	0	1	0
V11	Nutrition status	0.374223	13.4859998	0.022549	0	1
V9	Providing the necessary training	0.186321	12.4445506	0.040850	0	0
V8	Timely presence of nursing staff	0.303948	11.2288313	0.033007	1	0
V14	Welfare facilities	0.197817	8.9068204	0	1	0
V13	Calmness in the ward	0.05676	4.1218699	0	1	0
V16	Para-clinical units	0	4.1044337	0	0	1
V6	Demission type	0	3.8207376	0.008497	0	0
V12	Room ventilation status	0	3.1429437	0.007190	1	1
V1	Age	0	2.7349224	0	1	1
V7	Speed of file formation and transfer to the ward	0	2.474263	0.020915	1	0
V3	Education	0	1.9811921	0.001634	0	0
V5	Admission type	0	1.0004453	0	1	0
V4	Insurance	0	0.7631545	0	0	0
V2	Gender	0	0.4745147	0	1	0

has been employed to validate the models. Additionally, accuracy, precision, and recall indices have been used to evaluate the models. The parameter settings for all applied methods are presented in Table 8.

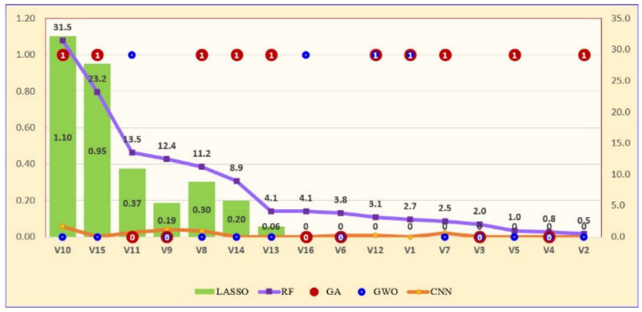


FIGURE 4. Weights Assigned to Variables based on LASSO, RF, GA CNN, and GWO methods.

TABLE 8. The parameter setting for the applied models.

Model	Parameters
ANN	Learning rate= 0.3, momentum= 0.2, hidden layer= 3 (neurons= 5,10,15)
SVM	Kernel= sigmoid, gamma= 0.05
KNN	K= 5,10,15,20,25,30
DT	Criterion= Gain Ratio, maximal depth= 20, confidence= 0.25, minimal gain= 0.1, minimal leaf size= 2, minimal size for split= 4, number of prepruning alternatives= 3
Rule based	Criterion= Information gain, pureness= 0.9, minimal prune benefit= 0.25
NB	NB has no parameters to set.

To demonstrate the positive effect of feature selection on the performance of classification algorithms, modeling has been conducted with all variables included in the model and features selected using LASSO, RF, and GA methods. As mentioned, the LASSO method assigned a non-zero coefficient to seven variables. The seven variables selected using the LASSO method were the same seven variables with the highest weight assigned by the Random Forest method. The GA method assigned non-zero weights to ten variables. The Random Forest method did not assign zero weight to any of the variables and did not eliminate any of the variables. Therefore, modeling with each of the used classification algorithms has been conducted in four modes as follows: 1) The first mode involved entering all variables into the selected classification algorithms; 2) Entering the seven selected variables by the LASSO method, which are the same seven variables with the highest weight assigned by the Random Forest method, into the selected classification algorithms; 3) Entering the ten variables with the highest weight using the Random Forest method into the selected classification algorithms; 4) Entering the ten selected variables by the genetic algorithm method into the selected classification algorithms. Table 9 presents the results of modeling by classification algorithms including ANN, SVM, KNN, DT, RB, and NB, and several FSS methods such as LASSO, RF, and GA.

TABLE 9. Classification results for selected classifier.

Model	Feature Selection	Classification Algorithms	Accuracy	Precision	Recall
Model 1	All Variables (16)	ANN	96.18% +/- 2.96%	97.75% +/- 3.01%	97.26 % +/- 2.49%
Model 2	LASSO (7) - RF(7)	ANN	96.76% +/- 2.44%	97.43% +/- 3.24%	98.45 % +/- 1.90%
Model 3	RF(10)	ANN	96.47% +/- 1.76%	97.05% +/- 2.73%	98.45 % +/- 1.90%
Model 4	GA (10)	ANN	94.41% +/- 3.59%	96.02% +/- 4.53%	96.89 % +/- 2.89%
Model 5	CNN (8)	ANN	95.00% +/- 2.96%	96.28% +/- 3.23%	97.29 % +/- 2.47%
Model 6	GWO(4)	ANN	86.18% +/- 4.93%	89.13% +/- 4.61%	93.37 % +/- 3.53%
Model 7	All Variables (16)	SVM	75.59% +/- 1.35%	75.59% +/- 1.35%	100.00 % +/- 0.00%
Model 8	LASSO (7) - RF(7)	SVM	96.18% +/- 3.49%	95.95% +/- 3.41%	99.23 % +/- 1.54%
Model 9	RF(10)	SVM	95.00% +/- 4.17%	94.56% +/- 3.90%	99.23 % +/- 1.54%
Model 10	GA (10)	SVM	94.71% +/- 5.06%	94.68% +/- 4.98%	98.85 % +/- 2.46%
Model 11	CNN (8)	SVM	95.59% +/- 4.00%	96.00% +/- 4.04%	98.45 % +/- 1.90%
Model 12	GWO(4)	SVM	86.18% +/- 4.37%	88.97% +/- 3.47%	93.40 % +/- 3.45%
Model 13	All Variables (16)	KNN	78.24% +/- 5.29%	81.07% +/- 2.21%	93.03 % +/- 8.20%
Model 14	LASSO (7) - RF(7)	KNN	94.41% +/- 3.82%	94.16% +/- 3.60%	98.85 % +/- 1.76%
Model 15	RF(10)	KNN	93.24% +/- 3.96%	92.51% +/- 3.92%	99.23 % +/- 1.54%
Model 16	GA (10)	KNN	93.24% +/- 4.37%	92.27% +/- 4.54%	99.62 % +/- 1.15%
Model 17	CNN (8)	KNN	94.12% +/- 3.48%	93.50% +/- 3.42%	99.23 % +/- 1.54%
Model 18	GWO(4)	KNN	85.00% +/- 5.95%	87.39% +/- 4.29%	93.74 % +/- 4.42%
Model 19	All Variables (16)	DT	90.88% +/- 5.49%	94.43% +/- 4.84%	93.77 % +/- 5.04%
Model 20	LASSO (7) - RF(7)	DT	94.12% +/- 3.22%	95.92% +/- 3.31%	96.51 % +/- 4.05%
Model 21	RF(10)	DT	92.06% +/- 5.59%	94.42% +/- 4.34%	95.32 % +/- 5.16%
Model 22	GA (10)	DT	92.65% +/- 4.41%	93.90% +/- 4.91%	96.86 % +/- 3.88%
Model 23	CNN (8)	DT	92.94% +/- 5.29%	96.98% +/- 4.33%	93.75 % +/- 4.73%
Model 24	GWO(4)	DT	86.47% +/- 3.77%	89.06% +/- 3.62%	93.78 % +/- 2.54%
Model 25	All Variables (16)	RB	92.65% +/- 4.79%	96.87% +/- 2.94%	93.40 % +/- 6.00%
Model 26	LASSO (7) - RF(7)	RB	95.00% +/- 3.49%	96.92% +/- 2.92%	96.48 % +/- 3.24%
Model 27	RF(10)	RB	93.53% +/- 4.12%	95.92% +/- 3.86%	95.71 % +/- 4.45%
Model 28	GA (10)	RB	94.41% +/- 4.04%	94.25% +/- 4.47%	98.82 % +/- 1.81%
Model 29	CNN (8)	RB	93.24% +/- 3.49%	95.19% +/- 3.96%	96.11 % +/- 3.48%
Model 30	GWO(4)	RB	85.88% +/- 4.52%	90.68% +/- 4.38%	91.02 % +/- 5.93%
Model 31	All Variables (16)	NB	92.35% +/- 4.20%	98.33% +/- 2.04%	91.45 % +/- 4.55%
Model 32	LASSO (7) - RF(7)	NB	92.65% +/- 4.97%	97.58% +/- 3.27%	92.63 % +/- 5.10%

According to Table 9, Model 2 exhibits higher accuracy when using the LASSO FS method. The accuracy, precision, and recall values are 96.76%, 97.43%, and 98.45%, respectively. The selected number of features in this model is 7. Based on Table 9, across all classification algorithms,

TABLE 9. ***(Continued.)*** **Classification results for selected classifier.**

Model 33	RF(10)	92.65% +/- 4.60%	98.33% +/- 2.05%	91.82 % +/- 5.13%
Model 34	GA (10)	92.94% +/- 3.77%	98.77% +/- 1.89%	91.85 % +/- 4.70%
Model 35	CNN (8)	92.94% +/- 4.40%	96.92% +/- 4.10%	93.80 % +/- 2.56%
Model 36	GWO(4)	81.47% +/- 6.71%	93.73% +/- 4.44%	80.94 % +/- 6.09%

the LASSO method achieves higher accuracy compared to the other three modes: "all variables", "RF(10)", and "GA(10)". Even though only 7 characteristics out of 16 variables are included in the model, the accuracy obtained is higher than other models. Therefore, LASSO feature selection brings us closer to the goal of feature selection, as selecting a compact feature subset leads to maximal predictive capability.

The SVM and KNN algorithms did not perform well when considering all variables as input. Model 7 is the model that exhibits the worst classification performance, with the lowest accuracy measure. In this model, all variables are taken as input for the SVM algorithm. As observed, the use of feature selection has resulted in an increase in the accuracy of SVM and KNN algorithms. Figure 5 illustrates the comparison of classification results based on the FS technique and the number of selected features for each of the applied models.

Table 10 represents the performance of Model 2 in the form of a confusion matrix.

As indicated in the table, out of 257 individuals who are actually willing to return to the hospital if necessary, 253 individuals are predicted as recurrent patients. Furthermore, 97.31% of the actual recurrent patients are correctly categorized.

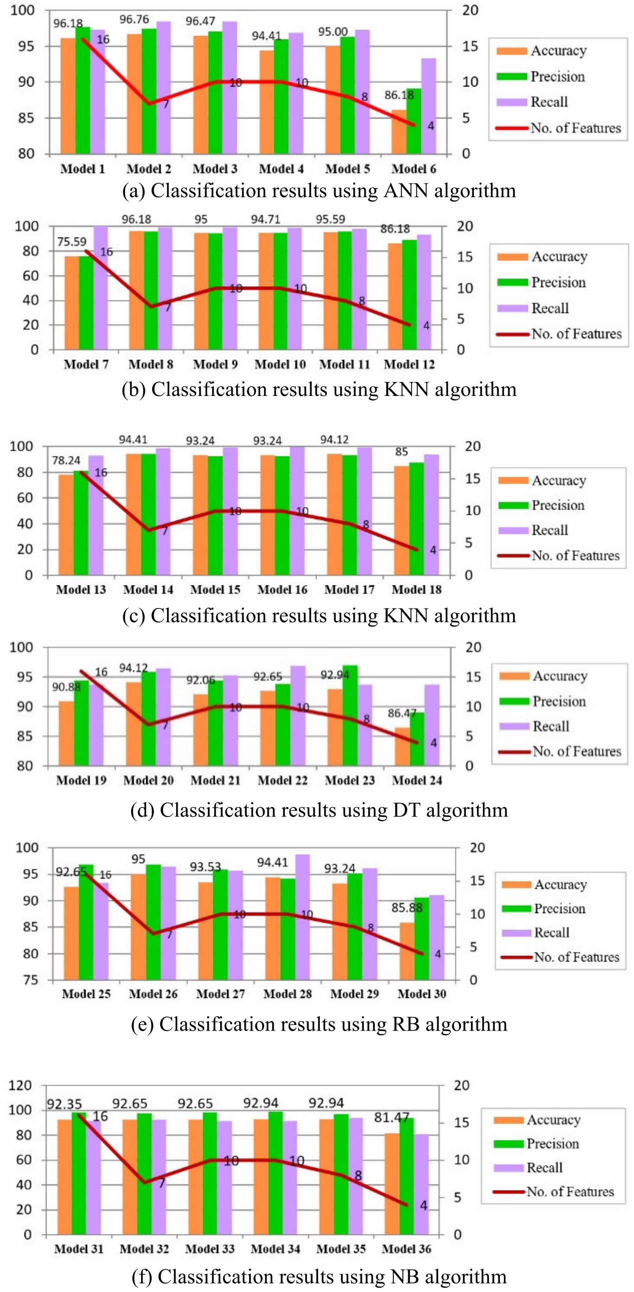


FIGURE 5. **Classification results based on FS technique and the number of selected features for each of the applied models.**

C. EXTERNAL VALIDATION OF THE PROPOSED MODEL

Once the model is trained and tested, to ensure its generalizability, we utilized a new dataset comprising 290 records to validate the presented model. The new dataset, without labels, was input into the proposed model. The results revealed that the model correctly predicted 278 out of 290 records, yielding an accuracy of approximately 96% based on the new data.

Following the external evaluation of the model and confirmation of its predictive quality, data pertaining to 95 new patients, collected over a period of approximately one month, were obtained. For these patients, the label variable indicating their intention to return to the hospital was unknown. To predict the patients' future behavior, this data was input into the prediction model. Essentially, the proposed model was applied to the data of new patients to predict whether they desire to return to the hospital.

The results indicated that approximately 25% (24 out of 95) of these patients have a high likelihood of revisiting the hospital. Individuals with higher predicted revisit probabilities were identified as those most likely to revisit the hospital. These individuals can then be targeted for proactive interventions, such as personalized follow-up care or targeted communication strategies, aimed at enhancing their overall experience and potentially reducing the likelihood of revisits.

TABLE 10. Classification results for selected classifier.

	true 0	true 1	class precision
pred. 0	76	4	95.00%
pred. 1	7	253	97.31%
class recall	91.57%	98.44%	

D. PRACTICAL APPLICATION OF THE PROPOSED MODEL

In this paper, a model is proposed that combines penalized regression with neural networks. The aim of the proposed model is to predict patients' revisit intentions based on their satisfaction scores.

The process of building the predictive model starts by collecting historical data on patients' demographic information, satisfaction scores, and revisit outcomes. LASSO regression is then applied to the dataset to select the most relevant features that contribute to predicting revisit intentions. Subsequently, an ANN model is designed and trained using the selected features from LASSO. The k-fold method is applied to validate the model. The ANN model is trained on the training set, and validation is performed using the test set to ensure it generalizes well to unseen data. Once the model is trained and validated, it is used in practice by inputting new satisfaction scores of patients into the model. The ANN processes this information through its layers, incorporating the learned relationships from the training data to predict the likelihood of patients revisiting the hospital. Individuals with higher predicted revisit probabilities are identified as those most likely to revisit the hospital.

The proposed model can be practical for the hospital in several aspects:

1. Planning to Create a Good Experience for Patients: Identify patients at high risk of revisiting the hospital based on the model's predictions. Since the model predicts the willingness of patients to return, it can estimate the proportion of referring patients in the future. This allows the hospital to make the necessary plans to provide optimal services and create a superior experience for these patients.
2. Resource Allocation: Allocate resources more efficiently by prioritizing patients with a higher likelihood of revisit. This could involve assigning care coordinators or case managers to patients identified as high-risk to ensure they receive appropriate support and follow-up care.
3. Personalized Care Plans: Develop personalized care plans for patients based on their individual risk factors identified by the model. Tailoring interventions to address specific needs can improve patient outcomes and satisfaction while reducing unnecessary healthcare utilization.
4. Quality Improvement Initiatives: Use insights from the model to inform quality improvement initiatives aimed at reducing hospital readmissions. Hospitals can target areas for improvement based on the factors driving revisit intentions identified by the model.
5. Patient Engagement: Engage patients in their own care by sharing the model's predictions and discussing strategies to reduce their risk of revisiting the hospital. This proactive approach can empower patients to take an active role in managing their health and preventing future admissions.
6. Performance Monitoring: Continuously monitor the performance of the predictive model in real-time clinical settings. Hospitals can refine the model over time based on feedback and new data to ensure its accuracy and relevance in guiding patient care decisions.

Integrating the predictive model into clinical workflows enables hospitals to adopt a proactive approach to patient care, improving outcomes, reducing healthcare costs, and enhancing the overall patient experience. The results of applying this model can significantly impact our understanding of patient behaviors and provide valuable insights. By accurately predicting whether patients are likely to return, hospitals can anticipate patients' needs and adjust their services accordingly, leading to increased patient satisfaction. Also, understanding patient behavior through predictive modeling can lead to a more effective allocation of resources and improvement in the overall quality of patient care. By using this model, healthcare providers can gain insights into the factors influencing patients' decisions to return to the hospital. They can personalize care plans and recommendations for patients. Using this model, the hospital can identify areas for improving the quality of its services and provide the necessary training to its employees to deliver better experiences for patients. By predicting which patients are more likely to return to the hospital, targeted interventions aimed at reducing the rate of return visits can be implemented. This could include improved discharge planning or better coordination of post-discharge care. Using this model helps the hospital build stronger relationships with their patients. By more effectively addressing patient needs and preferences, hospitals can increase patient satisfaction and loyalty. Satisfied patients are more likely to return for follow-up care and recommend the hospital to others. By identifying patients who are less likely to be readmitted, potential areas where interventions can improve patient satisfaction and encourage repeat visits are highlighted. This proactive approach can prevent dissatisfaction and resolve issues before they escalate. By using this model, the hospital can communicate more effectively with patients. For example, they can provide reminders to encourage patients to return for necessary appointments or treatments. Since the model presented in this research uses feature selection algorithms to identify the factors affecting patients' return, it enables the hospital to identify areas for improving the quality of their services. This could include enhancing certain services, streamlining processes, or addressing any factors that contribute to patient dissatisfaction.

There are several ethical considerations when developing predictive models for hospitals, particularly when using sensitive data like patient satisfaction scores. Key considerations include ensuring patient privacy and confidentiality, obtaining informed consent from participants, minimizing bias in data collection and analysis, and more. Protecting patients' personal information is paramount. Researchers must ensure that data are anonymized or de-identified before analysis to prevent the possibility of re-identification. Obtaining informed consent from participants is essential. Participants should be fully informed about the nature of the study, including its purpose, potential risks and benefits, and how their data will be used. Another important ethical principle is transparency in methods, findings, and any conflicts of interest. This includes clearly documenting data sources and preprocessing steps, openly sharing data where possible, and accurately reporting limitations and uncertainties in research findings.

Furthermore, resource computation cost is an important consideration in machine learning and data science projects, as it directly affects the time, effort, and infrastructure required to perform various tasks. The resource computation cost for the proposed Lasso-ANN predictive modeling in this research can vary depending on factors such as the complexity of the neural network architecture, the size of the dataset, and the computational resources available. For the dataset used in this research with 17 variables and 340 records, training a neural network with Lasso regularization might take longer compared to traditional linear models due to the iterative optimization process involved in training neural networks. Techniques like hyperparameter optimization involve running multiple experiments with different configurations to find the optimal settings. Each experiment adds to the computation cost, especially when using techniques like grid search or random search across a wide range of hyperparameters. Training machine learning models involves iterative processes of updating model parameters based on input data, which can be computationally intensive.

V. CONCLUSION

Hospital marketing faces several challenges due to the growth of private hospitals and increasing competition. As part of their competitive strategy, hospitals should strive to improve patient referrals to gain an edge in the market. Patient satisfaction is a critical concept in today's healthcare system and is considered a priority for health systems. The experience of illness and the treatment process highlight the importance of patient satisfaction because patients are vulnerable and require comprehensive support. Factors affecting satisfaction include patients' willingness to return to the hospital if necessary. Understanding these factors is crucial for preparation. Data mining methods have been used in this research to address this need.

Feature selection techniques revealed that the performance of the treating physician and the patient's sanitary conditions are major factors in determining their willingness to return to the clinic. Several models were used for prediction, including an Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), rule-based model, and net-based model. Compared with other models, the ANN model had the best predictive performance, while the Lasso model excelled in feature selection. Lasso aids in feature selection by shrinking coefficients of less important variables, while the ANN model is adept at capturing nonlinear relationships between predictors and outcomes.

In this model, users can input patient satisfaction scores, which the model processes by considering both linear and nonlinear relationships to predict the likelihood of patients revisiting the hospital. Based on the predictions generated by the model, individuals with higher revisit probabilities are identified as those most likely to return. This information can be used to prioritize follow-up care or interventions for these individuals, potentially improving patient satisfaction and healthcare outcomes. By leveraging the strengths of both Lasso and ANN, this approach provides a robust framework for identifying individuals most likely to revisit the hospital based on satisfaction scores, enabling healthcare providers to allocate resources more effectively and improve patient outcomes.

By applying the proposed model, hospitals can develop outreach programs targeting patients identified as high-risk for revisits. These programs can include proactive follow-up calls, home visits, or personalized communication to ensure patients receive appropriate care and support after discharge. Integrating predictive analytics tools like the ANN model into clinical decision-making processes can help healthcare providers identify patients at high risk of hospital revisits, improving care planning and resource allocation for more targeted and effective interventions.

LIMITATIONS OF THIS RESEARCH

The accuracy and comprehensiveness of input data are critical to the success of predictive models. Incomplete, erroneous, or biased data may lead to unreliable findings and forecasts. The study focuses on the healthcare system in Iran, which may have unique characteristics and patient behaviors, limiting the generalizability of the conclusions and models to other healthcare environments. While machine learning methods are known for their accuracy, they sometimes lack interpretability, making it difficult to fully understand patient satisfaction criteria and model predictions. Additionally, the study does not assess the long-term effects of implementing the recommendations from predictive analytics; instead, it focuses on patient satisfaction and developing a hospital plan. To evaluate the efficacy of the prediction model, it is essential to examine the impact of these interventions on healthcare outcomes and long-term patient satisfaction.

ACKNOWLEDGMENT

(Farshid Abdi, Shaghayegh Abolmakarem, Amir Karbassi Yazdi, Paul Leger, Yong Tan, and Giuliani Coluccio contributed equally to this work.)

REFERENCES

[1] J. E. Hasemann Lara, "Health sector reform in Honduras: Privatization as institutional bad faith," *Med. Anthropology*, vol. 42, no. 1, pp. 62–75, Jan. 2023.

[2] C. Ferrario, R. Levaggi, and M. Piacenza, "Efficiency and equity effects of healthcare decentralization: Evidence from Italy," in *Handbook on the Political Economy of Health Systems*. Cheltenham, U.K.: Edward Elgar, Jun. 2023, pp. 85–111.

[3] A. R. Komijan, A. K. Yazdi, Y. Tan, L. Ocampo, and F. Nasrollahpourniazi, "Spherical fuzzy multicriteria decision making for evaluating healthcare service quality of hospitals during the global pandemic," *Int. J. Comput. Intell. Syst.*, vol. 17, no. 1, p. 105, Apr. 2024.

[4] A. K. Yazdi, F. M. Muneeb, P. F. Wanke, T. Hanne, and A. Ali, "How, when, & where temporary hospitals fit in turbulent times: A hybrid MADM optimization in the middle east," *Comput. Ind. Eng.*, vol. 175, p. 108761, Jan. 2023.

[5] A. M. Burgener, "Enhancing communication to improve patient safety and to increase patient satisfaction," *Health Care Manager*, vol. 39, no. 3, pp. 128–132, 2020.

[6] P. Thomas, C. Baldwin, L. Beach, B. Bissett, I. Boden, S. M. Cruz, R. Gosselink, C. L. Granger, C. Hodgson, A. E. Holland, A. Y. Jones, M. E. Kho, L. van der Lee, R. Moses, G. Ntoumenopoulos, S. M. Parry, and S. Patman, "Physiotherapy management for COVID-19 in the acute hospital setting and beyond: An update to clinical practice recommendations," *J. Physiotherapy*, vol. 68, no. 1, pp. 8–25, Jan. 2022.

[7] R. Bohmer, *Managing Care: How Clinicians Can Lead Change and Transform Healthcare*. Oakland, CA, USA: Berrett-Koehler, 2021.

[8] O. Fennelly, C. Cunningham, L. Grogan, H. Cronin, C. O'Shea, M. Roche, F. Lawlor, and N. O'Hare, "Successfully implementing a national electronic health record: A rapid umbrella review," *Int. J. Med. Informat.*, vol. 144, Dec. 2020, Art. no. 104281.

[9] I. J. Udo and M. E. Ekpenyong, "Improving emergency healthcare response using real-time collaborative technology," in *Proc. 4th Int. Conf. Med. Health Informat.*, Aug. 2020, pp. 165–173.

[10] P. Chauhan, A. Bali, and S. Kaur, "Breaking barriers for accessible health programs: The role of telemedicine in a global healthcare transformation," in *Transformative Approaches To Patient Literacy and Healthcare Innovation*. Hershey, PA, USA: IGI Global, 2024, pp. 283–307.

[11] S. Hwang, J. Kim, E. Park, and S. J. Kwon, "Who will be your next customer: A machine learning approach to customer return visits in airline services," *J. Bus. Res.*, vol. 121, pp. 121–126, Dec. 2020.

[12] O. Demirdöğen, H. Erdal, and A. İ. Akbaba, "Comparing various machine learning methods for prediction of patient revisit intention: A case study," *Selcuk Univ. J. Eng.,Science Technol.*, vol. 5, no. 4, pp. 386–401, Dec. 2017.

[13] I. Ouedraogo, B. M. J. Some, R. Benedikter, and G. Diallo, "Mobile technology as a health literacy enabler in African rural areas: A literature review," 2021.

[14] D. Solnet and J. Kandampully, "How some service firms have become part of 'service excellence' folklore: An exploratory study," *Manag. Service Quality: Int. J.*, vol. 18, no. 2, pp. 179–193, Mar. 2008.

[15] L. Bosetti, "Health and safety," in *Encyclopedia of Sustainable Management*. Springer, 2023, pp. 1850–1861.

[16] K. H. Mead, S. Raskin, A. Willis, H. Arem, S. Murtaza, L. Charney, and M. Pratt-Chapman, "Identifying patients' priorities for quality survivorship: Conceptualizing a patient-centered approach to survivorship care," *J. Cancer Survivorship*, vol. 14, no. 6, pp. 939–958, Dec. 2020.

[17] R. Rezapour, A. A. Dorosti, M. Farahbakhsh, S. Azami-Aghdash, and I. Iranzad, "The impact of the covid-19 pandemic on primary health care utilization: An experience from Iran," *BMC Health Services Res.*, vol. 22, no. 1, p. 404, Mar. 2022.

[18] S. Islam and N. Muhamad, "Patient-centered communication: An extension of the HCAHPS survey," *Benchmarking: Int. J.*, vol. 28, no. 6, pp. 2047–2074, Jun. 2021.

[19] K. Z. Htoo, "Marketing mix on patient satisfaction and loyalty at ma Har myaing hospital," M.S. thesis, MERAL Portal, 2023.

[20] E. Alpaydin, *Introduction To Machine Learning*. Cambridge, MA, USA: MIT Press, 2020.

[21] F. Del Giorgio Solfa and F. R. Simonato, "Big data analytics in healthcare: Exploring the role of machine learning in predicting patient outcomes and improving healthcare delivery," *Int. J. Computations, Inf. Manuf. (IJCIM)*, vol. 3, no. 1, pp. 1–9, Jun. 2023.

[22] K. N. Kunze, E. M. Polce, A. J. Sadauskas, and B. R. Levine, "Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty," *J. Arthroplasty*, vol. 35, no. 11, pp. 3117–3122, Nov. 2020.

[23] E. M. Polce, K. N. Kunze, M. C. Fu, G. E. Garrigues, B. Forsythe, G. P. Nicholson, B. J. Cole, and N. N. Verma, "Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty," *J. Shoulder Elbow Surgery*, vol. 30, no. 6, pp. e290–e299, Jun. 2021.

[24] M. C. E. Simsekler, N. H. Alhashmi, E. Azar, N. King, R. A. M. A. Luqman, and A. Al Mulla, "Exploring drivers of patient satisfaction using a random forest algorithm," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, p. 157, Dec. 2021.

[25] Y. Ma, N. Cui, X. Li, J. Li, Y. Zhang, Y. Liu, and J. Jin, "Predicting critical care Nurses' intention to use physical restraints in intubated patients: A structural equation model," *J. Nursing Manage.*, vol. 2023, pp. 1–9, Jul. 2023.

[26] K.-J. Kwon, L.-W. Mai, and N. Peng, "Determinants of consumers' intentions to share knowledge and intentions to purchase on s-commerce sites: Incorporating attitudes toward persuasion attempts into a social exchange model," *Eurasian Bus. Rev.*, vol. 10, no. 1, pp. 157–183, Mar. 2020.

[27] P. Kumar, S. K. Sharma, and V. Dutot, "Artificial intelligence (AI)-enabled CRM capability in healthcare: The impact on service innovation," *Int. J. Inf. Manage.*, vol. 69, Apr. 2023, Art. no. 102598.

[28] M. U. H. Uzir, H. Al Halbusi, R. Lim, I. Jerin, A. B. Abdul Hamid, T. Ramayah, and A. Haque, "Applied artificial intelligence and user satisfaction: Smartwatch usage for healthcare in Bangladesh during COVID-19," *Technol. Soc.*, vol. 67, Nov. 2021, Art. no. 101780.

[29] S. Van Onsem, C. Van Der Straeten, N. Arnout, P. Deprez, G. Van Damme, and J. Victor, "A new prediction model for patient satisfaction after total knee arthroplasty," *J. Arthroplasty*, vol. 31, no. 12, pp. 2660–2667, Dec. 2016.

[30] A. A. M. P. Damayanti and K. A. Kusumawardani, "Predicting the revisit intention at private hospitals mediated by perceived service quality and satisfaction," in *Proc. 1st Int. Congr. Regional Econ. Develop., Inf. Technol., Sustain. Bus.*, 2020, pp. 105–122.

[31] F. Calisir, A. E. Bayraktaroglu, C. A. Gumussoy, and B. Kaya, "Effects of service quality dimensions including usability on perceived overall quality, customer satisfaction, and return intention in different hospital types," *Int. J. Adv. Operations Manage.*, vol. 6, no. 4, p. 309, 2014.

[32] S. Woo and M. Choi, "Medical service quality, patient satisfaction and intent to revisit: Case study of public hub hospitals in the republic of Korea," *PLoS ONE*, vol. 16, no. 6, Jun. 2021, Art. no. e0252241.

[33] K. P. Lai, Y. Yee Yen, and C. Siong Choy, "The effects of service quality and perceived price on revisit intention of patients: The Malaysian context," *Int. J. Quality Service Sci.*, vol. 12, no. 4, pp. 541–558, Nov. 2020.

[34] S. Mohd Isa, G. S. S. Lim, and P. N. Chin, "Patients' intent to revisit with trust as the mediating role: Lessons from Penang Malaysia," *Int. J. Pharmaceutical Healthcare Marketing*, vol. 13, no. 2, pp. 140–159, Jun. 2019.

[35] S. M. C. Loureiro, "Medical tourists' emotional and cognitive response to credibility and servicescape," *Current Issues Tourism*, vol. 20, no. 15, pp. 1633–1652, Nov. 2017.

[36] T. H. Cham, Y. M. Lim, N. C. Aik, and A. G. M. Tay, "Antecedents of hospital brand image and the relationships with medical tourists' behavioral intention," *Int. J. Pharmaceutical Healthcare Marketing*, vol. 10, no. 4, pp. 412–431, Nov. 2016.

[37] A. N. Seow, Y. O. Choong, K. Moorthy, and L. M. Chan, "Intention to visit Malaysia for medical tourism using the antecedents of theory of planned behaviour: A predictive model," *Int. J. Tourism Res.*, vol. 19, no. 3, pp. 383–393, May 2017.

[38] P. Suhail and Y. Srinivasulu, "Perception of service quality, satisfaction, and behavioral intentions in ayurveda healthcare," *J. Ayurveda Integrative Med.*, vol. 12, no. 1, pp. 93–101, Jan. 2021.

[39] A. Al-Refaie, "A structural model to investigate factors affect patient satisfaction and revisit intention in Jordanian hospitals," in *Investigations Into Living Systems, Artificial Life, and Real-World Solutions*. Hershey, PA, USA: IGI Global, 2013, pp. 136–147.

[40] J. Shuja, M. A. Humayun, W. Alasmary, H. Sinky, E. Alanazi, and M. K. Khan, "Resource efficient geo-textual hierarchical clustering framework for social IoT applications," *IEEE Sensors J.*, vol. 21, no. 22, pp. 25114–25122, Nov. 2021.

[41] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010.
[42] C. Aldrich and L. Auret, "Fault detection and diagnosis with random forest feature extraction and variable importance methods," *IFAC Proc. Volumes*, vol. 43, no. 9, pp. 79–86, 2010.
[43] S. Oreski and G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2052–2064, Mar. 2014.
[44] R. A. Welikala, M. M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T. H. Williamson, and S. A. Barman, "Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy," *Computerized Med. Imag. Graph.*, vol. 43, pp. 64–77, Jul. 2015.
[45] B. A. Alabsi, M. Anbar, and S. D. A. Rihan, "CNN-CNN: Dual convolutional neural network approach for feature selection and attack detection on Internet of Things networks," *Sensors*, vol. 23, no. 14, p. 6507, 2023, doi: 10.3390/s23146507.
[46] J. Zhang, Y. Dai, and Q. Shi, "An improved grey wolf optimization algorithm based on scale-free network topology," *Heliyon*, vol. 10, no. 16, 2024, Art. no. e35958.
[47] S. Heo and J. H. Lee, "Fault detection and classification using artificial neural networks," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 470–475, 2018.
[48] F. Moslehi, A. Haeri, and M. R. Gholamian, "Investigation of effective factors in expanding electronic payment in Iran using datamining techniques," *J. Ind. Syst. Eng.*, vol. 12, no. 2, pp. 61–94, Mar. 2019.
[49] P. Aggarwal and S. K. Sharma, "Analysis of KDD dataset attributes-class wise for intrusion detection," *Proc. Comput. Sci.*, vol. 57, pp. 842–851, Jun. 2015.
[50] L. Breiman and A. Cutler, *Random Forests*, Berkeley, CA, USA, 2005.
[51] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

FARSHID ABDI received the M.Sc. and Ph.D. degrees in industrial management. He is currently an Assistant Professor of industrial engineering with Islamic Azad University, South Tehran Branch, Tehran, Iran. He has published more than 100 papers in high-quality peer-reviewed international journals and conferences. His research interests include soft computing, machine learning, and customer relationship management.

SHAGHAYEGH ABOLMAKAREM received the Ph.D. degree in industrial engineering. She has published more than ten papers in high-quality peer-reviewed international journals and conferences. Her research interests include soft computing, machine learning, data mining, and customer relationship management.

AMIR KARBASSI YAZDI received the Ph.D. degree in industrial management. He is currently an Assistant Professor at Universidad de Tarapacá, Arica, Chile. He has published more than 40 articles in high-quality peer-reviewed international journals. His research interests include operations research, MCDM, SCM, and fuzzy sets.

PAUL LEGER received the Ph.D. degree in computer science from the University of Chile. He is currently an Associate Professor at Universidad Católica del Norte, Chile. His research interests include programming languages, software engineering, and different programming approaches.

YONG TAN is currently an Associate Professor in economics and finance at the University of Bradford. He has published his research in a number of highly influential academic journals. His research interests include financial economics and production economics. His research is widely recognized internationally, he has been featured in the list of the Top 2% scientists in the world published by Stanford University.

GIULIANI COLUCCIO received the degree in industrial engineering from Universidad de Tarapacá, in 2012, and the Ph.D. degree in management from Universidad Adolfo Ibáñez, in 2019. He is currently an Assistant Professor with the Industrial and Systems Department, Universidad de Tarapacá. His primary research interests include business administration, organizational behavior, and organizational dynamics.