International Journal of Industrial Engineering and Operational Research (IJIEOR)

Contents lists available at IJIEOR
International Journal of Industrial Engineering and Operational Research
journal homepage: http://ijieor.ir
Volume 8, No. 1, 2026

Development of a Machine Learning-Based Framework for Improving Appointment Attendance Prediction in Outpatient Clinics Using Stacking Algorithm

Mohammadreza Ensafi [a], Ashkan Mozdgir [b], Nasim Ghanbar Tehrani [b], Orod Ahmadi [b]

[a] *Department of Industrial and System Engineering, Auburn University, Auburn, United States,*
[b] *Department of Industrial Engineering, Kharazmi University, Tehran, Iran.*

ARTICLE INFO

Received: 2026/04/15

Revised: 2026/05/10

Accept: 2026/06/17

Keywords:

Data Mining, Healthcare systems Stacking model, Multi-criteria decision making, TOPSIS.

ABSTRACT

Patient no-shows pose a significant challenge for healthcare providers, impacting operational efficiency and resource allocation. This study introduces an innovative machine learning framework designed to predict patient no-shows in outpatient clinics, while addressing data imbalance challenges in healthcare datasets and identifying crucial factors that influence patient attendance. The methodology encompasses data preprocessing, recursive feature elimination for feature selection, novel feature extraction techniques, and the utilization of methods such as the synthetic minority over-sampling technique and stratified k-fold cross-validation to handle class imbalance. The results highlighted the efficacy of a stacking model that integrates six distinct algorithms, achieving an impressive area under the receiver operating characteristic curve (AUC-ROC) of 0.911, an accuracy of 82%, a recall of 82%, and a precision of 85%. Finally, the technique for order of preference by similarity to ideal solution was employed based on evaluation metrics, demonstrating the superiority of the stacking model compared with other classifiers.

1 Introduction

In the context of rising demand and cost constraints, healthcare institutions face a critical need to enhance the efficiency and efficacy of their services. Despite extensive efforts, unavoidable

[a] Corresponding author email address: mze0069@auburn.edu, Mohammadreza.ensafi1998@gmail.com (Mohammadreza Ensafi).
DOI: https://doi.org/10.22034/ijieor.v8i1.211
Available online 2026/06/19

2676-3311/BGSA Ltd.

circumstances can still lead to planning and operational challenges for clinics [1]. These challenges subsequently lead to diminished productivity [2], limited access to sufficient healthcare for patients, hindered disease management practices [3], escalated costs, and underutilization of facilities and resources. An illustrative instance of such a challenge is patient no-shows, which significantly contributes to a discrepancy between anticipated supply and demand, consequently impacting service quality and system performance [4].

The phenomenon of patient no-shows, wherein patients fail to attend scheduled appointments without prior communication, poses significant challenges within healthcare facilities. These occurrences not only disrupt scheduling arrangements and impact the efficient allocation of physical space and human resources but also detrimentally affect the health of nonattending patients and cause dissatisfaction among others seeking timely appointments. The repercussions extend beyond individual care, affecting the treatment of other patients who are unable to secure timely appointments and diminishing generated revenue. Previous research has focused predominantly on understanding the adverse consequences of patient no-shows, developing strategies to mitigate them, and identifying predictive factors.

When a patient misses an appointment, healthcare access is denied to two individuals: the absent patient and someone unable to secure an appointment due to limited availability. Leveraging diverse machine learning algorithms aids in predicting the likelihood of no-shows based on patient characteristics. This predictive capability is pivotal for healthcare centers, enabling them to tailor resource allocation and scheduling strategies based on their unique patient behavior and characteristics. Ultimately, this enhances operational efficiency and patient satisfaction.

Accurate prediction outcomes can significantly enhance hospitals' decision-making processes by reducing patient no-shows, thereby improving operational efficiency and leading to economic and social advantages for healthcare institutions [5]. It has been estimated that approximately 67,000 instances of patient no-shows could incur a significant cost of nearly $7 million to the healthcare system [6]. The rate of missed appointments varies across healthcare centers globally, ranging from 10% to 50%. In North America, where the average no-show rate is reported to be 27% [7], Drabkin et al. [8] proposed implementing phone call reminders for patients, suggesting that this intervention could reduce the no-show rate from 20.99% to 7.07%.

The contributions of this study can be summarized as follows:

- A novel predictive model for missed appointments are proposed by utilizing a stacking algorithm that integrates six different machine learning algorithms.
- New features, derived from existing ones, were created to enhance model interpretability and overall performance.
- A preprocessing framework designed to handle imbalanced data effectively is introduced.
- The most significant features influencing patient attendance were identified.
- The technique for order of preference by similarity to ideal solution (TOPSIS) was implemented to rank machine learning models based on evaluation metrics.

The paper is structured as follows: Section 2 provides a comprehensive literature review, Section 3 covers the data analysis, preprocessing techniques, and prediction algorithms, Section 4 presents the results, and Section 5 concludes with a summary of the findings, limitations, and recommendations for future research directions.

2 Related works

Previous research has suggested that patient no-shows are not random events but are influenced by a variety of factors [9]. Logistic issues, inadequate scheduling systems, lack of respect for patients by healthcare providers or the healthcare system, unaffordability, inappropriate timing, patient forgetfulness, and the severity of the patient's illness have been identified as reasons for missed appointments [2, 7]. Research findings on the impact of sex on non-attendance at appointments vary; for example, some studies have shown that women have a lower rate of nonattendance compared to men [10], while contrasting findings from different studies suggest that men may exhibit higher rates of non-attendance [11]. These conflicting findings highlight the complexity of the relationship between gender and non-attendance behavior. Higher no-show rates were observed in public clinics compared to private clinics when batch appointments were implemented, highlighting the impact of group appointments on appointment attendance in pediatric settings[12].

Despite multiple interventions aimed at reducing non-attendance rates, including reminder messages and temporary account suspensions, this issue remains unresolved [13, 14]. Addressing this multifactorial problem necessitates health centers to identify key contributors, as relying solely on reminders or penalties is insufficient for a comprehensive solution. Research has shown that calls are more effective than automated reminder messages [15, 16], resulting in reduced non-attendance rates for patients [17]. One commonly used approach to address patient no-shows is

overbooking appointments, involving scheduling more appointments than the authorized capacity to optimize provider utilization. However, this strategy may lead to longer patient wait times and extended operation hours. Balancing revenue and costs while maintaining care quality and patient satisfaction presents a significant challenge. Achieving this equilibrium requires a deep understanding of facility capacity, demand patterns, and operational limits. Effective strategies can help minimize the consequences of patient no-shows and maintain operational efficiency[18-20].

Machine learning is widely used to create predictive models for various types of risks, such as hospital readmission, the identification of patient side effects [21], and the development of predictive models for in-hospital mortality [22]. Dunstan et al. [23] employed machine learning algorithms to predict pediatric patient no-shows, implementing cost-effective metrics and a reminder strategy that notably reduced overall no-show rates in a public hospital in Santiago, Chile. Lenzi et al. [24] employed mixed-effects logistic regression to identify the previous no-show rate as a potential predictor for subsequent patient appointments. Srinivas et al. [25] explored various machine learning models, including random forests, random gradient boosting, and deep neural networks, for predicting patient no-shows. Aladeemy et al. [26] employed novel wrapper-based methods using adversarial algorithms, self-compatibility, and swarm intelligence. In the context of primary care in rural areas, appointment lead time was identified as a significant predictor of patient no-shows. Logistic regression, decision trees, and tree-based group classifiers were employed to analyze this predictor [27]. Alshammari et al. [28] reported that a single model cannot achieve optimal performance in predicting patient no-shows. Dashtban et al. [29] employ sparse stacked denoising autoencoders (SDAEs) to develop an advanced predictive model for hospital outpatient non-attendance, addressing the financial burden caused by missed appointments through the integration of diverse data sources and enhanced data representation. Chen et al. [30] collected data from a Portland hospital's electronic health records to predict patient no-shows in a pediatric ophthalmology clinic. Their focus was on pediatric ophthalmology because of the preventable nature of visual loss in children and the high no-show rates for pediatric patients. They employed a combination of techniques including support vector regression, LASSO regression, XGBOOST, and random forests for their analysis. Liu et al. [31] introduced an interpretable deep learning model for absenteeism prediction, addressing the challenge of handling missing data and integrating local weather information to enhance accuracy. Their approach correctly identified 83% of absences, with a misclassification rate of less than 17%, demonstrating

the effectiveness of the model. This section concludes with Table 1, providing a concise summary of the key findings and unique features of previous studies on predicting patient no-shows.

Table 1 A comprehensive summary of previous studies on predicting patient no-shows

Important Variables	AUC	Algorithms	Domain	Authors
Previous visits, average days between visits	90%	- LASSO Regression - XG-Boost - Random Forest - Support Vector Machine	Pediatric (Ophthalmology Clinic)	Chen et al. [30]
Patient registration time, visit time, appointment creation time, patient distance from the hospital	99%	- Logistic Regression - KNN - Decision Tree - Random Forest - Bagging	walk-in clinic	Fan et al. [5]
Distance between reservation day and appointment day, number of previous no-shows, number of days after the last appointment	86%	- Naïve Bayes - Neural Network - Logistic Regression	Community Health Center	Mohammadi et al. [32]
Weekday	95%	- Logistic Regression	Diabetes Outpatient Clinic	Kurasawa et al. [33]
Last appointment status (presence or absence), distance between reservation day and appointment day, number	71%	- Logistic Regression - Neural Network	Former Military Health Care Center	Goffman et al. [34]

Important Variables	AUC	Algorithms	Domain	Authors
of appointments scheduled in a day				
Age, Time interval between reservation and appointment day, Number of previous no-shows, Hospital department where the patient made the reservation	N/A	- Logistic Regression	walk-in clinic	Su et al [35]
Number of previous no-shows, Having multiple appointments in a day	81%	- Logistic Regression	Medical Center	Lenzi et al. [24]
N/A	64%	- Random Forest - Support Vector Machine - XG-Boost	Elective Urological Surgeries	Luo et al. [36]
N/A	78%	- XG-Boost - Logistic Regression - Support Vector Machine - Neural Network	Pediatrics	Liu et al. [37]
Age, Ethnicity, Insurance and healthcare services status	74%	- Logistic Regression	Anesthesia	Odonkor et al. [38]

3 Materials and methods

Fig. 1 serves as a comprehensive overview of the research methodology employed in this study, providing a deeper understanding of the overall structure of the utilized procedures. The investigation involved handling a dataset with a high degree of class imbalance, requiring meticulous preprocessing techniques to ensure reliable results. This preprocessing stage involved

several critical steps to address the imbalanced nature of the data. The initial stage of the preprocessing phase focused on eliminating undesirable features that were either irrelevant or potentially noisy for the analysis. This process involved careful consideration and domain expertise to identify and remove any features that could introduce bias or hinder the performance of the models. Simultaneously, new relevant features were incorporated into the dataset to enhance the predictive capabilities of the models. These new features were carefully selected based on their potential to provide valuable insights and improve the overall accuracy of the predictive models. The preprocessing stage involved the removal of redundant records from the dataset, ensuring the reliability and integrity of subsequent analyses.

Furthermore, to ensure that the data were on a comparable scale and to prevent any bias toward specific features, data normalization using the min-max scaler technique was employed. Given the substantial class imbalance observed in the dataset, a pivotal step in the preprocessing phase was to address this issue to prevent biased predictions, which was accomplished by employing the synthetic minority oversampling technique (SMOTE). An integral facet of the modeling process involved feature selection, performed using the recursive feature elimination (RFE) method. In our approach, we applied RFE with three classifiers (logistic regression, random forest, and gradient boosting). Following model training, we assessed feature importance scores to identify consistently important features, as determined by at least two out of the three classifiers. Additionally, to comprehensively assess the predictive performance of the models and tackle the class imbalance challenge, stratified k-fold cross-validation was employed. This approach was particularly pivotal in this research, as it prevented the models from being biased toward the majority class and enabled a robust evaluation of their performance on different subsets of the data.

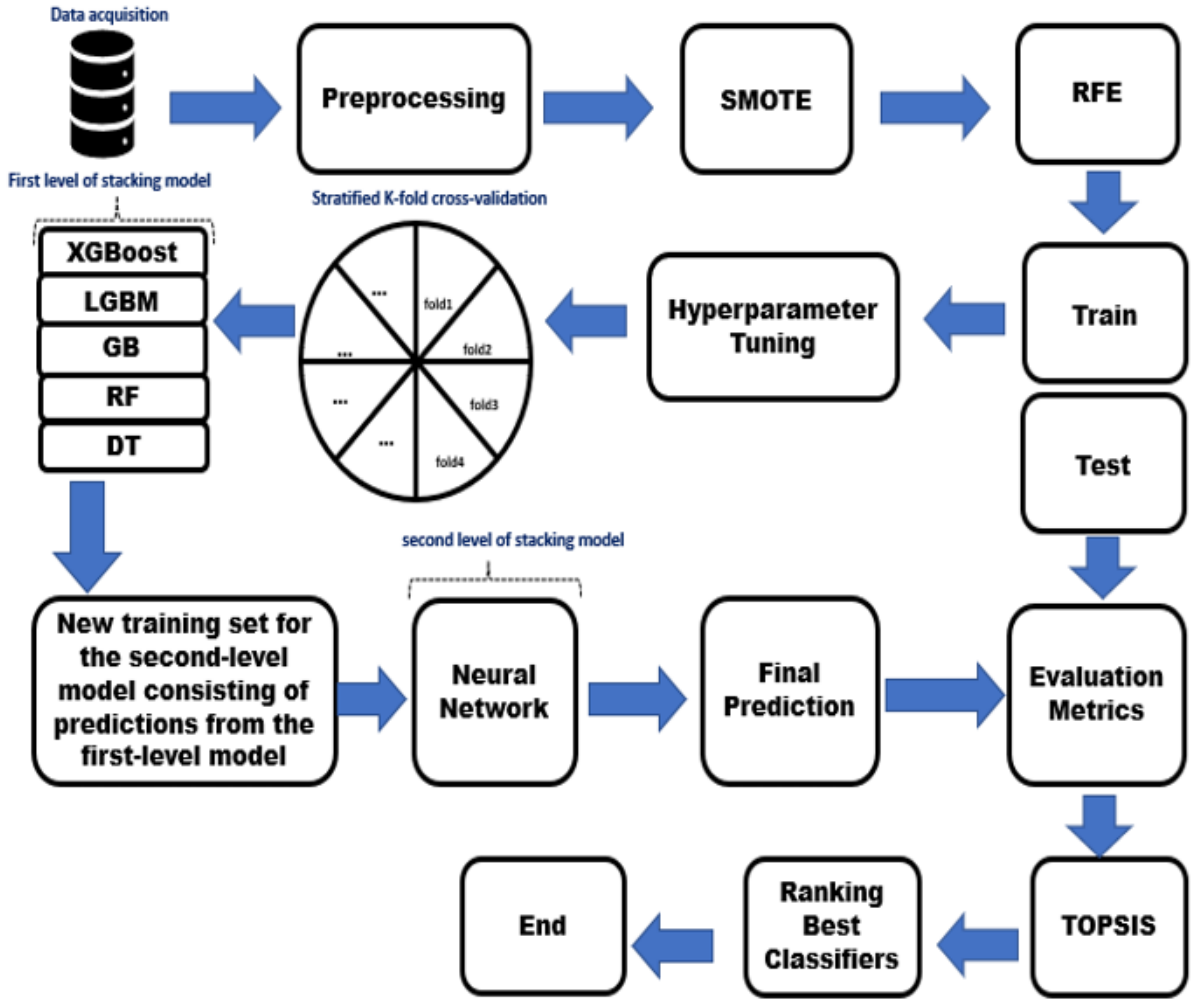

Fig. 1 Comprehensive overview of the research methodology employed in this study

Following the cross-validation phase, a range of algorithms were employed for prediction, including decision tree, random forest, gradient boosting, LightGBM, XGBoost, and neural network. Each algorithm has unique strengths and characteristics in the modeling process, allowing for a comprehensive assessment of its performance in addressing the research objectives. Table 2 represents the results obtained through grid search, which is a method used to select optimal hyperparameters for the model.

In the final step, a stacked model was constructed by combining the individual models and algorithms. The purpose of creating this stacked model was to harness the strengths and diversity of these individual models, enhancing the overall predictive performance. This approach aimed to capture the collective knowledge of the individual models and produce more precise and robust predictions. Subsequently, accuracy, precision, recall, and the area under the receiver operating

characteristic curve (AUC-ROC) were employed to assess the performance of the models. These evaluation metrics were calculated based on the weighted average To ascertain the optimal classifier, we employed the technique for order of preference by similarity to ideal solution (TOPSIS) based on diverse evaluation metrics. This method enabled us to rank the classifiers, facilitating a nuanced understanding of their relative performance and aiding in the identification of the most proficient models.

By following this comprehensive framework this research study aimed to address the challenges posed by class imbalances and enhance the predictive capabilities of the models employed. The detailed methodology described above serves as a foundation for the subsequent analysis and contributes to the robustness and validity of the research findings.

Table 2 Tuned hyperparameters for the DT, GB, RF, XGBoost, and LGBM classifiers

Model	Hyperparameters
Decision Tree	Max Depth: 6, Split Criterion: Gini
Gradient Boosting	n_estimators: 200, learning_rate: 1, max_depth: 7
Random Forest	n_estimators: 300, max_depth: 20, min_samples_split: 5
XGBoost	n_estimators: 100, max_depth: 3, learning_rate: 0.1
LGBM	n_estimators: 150, max_depth: 7, learning_rate: 0.1, num_leaves: 150

3.1 Data description

The dataset used in this research study was obtained from Kaggle ; it comprises a total of 110,528 observations and includes 14 variables. These variables, listed in Table ***3*** along with their descriptions, encompass vital appointment information, including the appointment date, scheduling time, appointment ID, and patient ID. Additionally, the dataset included demographic details, including age, gender, financial status, alcoholism status, handicap, and patient health.

Table 3 A concise overview of the variables

Variable	Description
Patient Id	Unique identification of each patient
Appointment Id	Unique identification of each appointment
Gender	Male or Female

Variable	Description
Appointment Day	The date of the actual appointment when the patient should visit the doctor
Scheduled Day	The date when patient made the appointment registration or call, which occurs prior to the appointment date
Age	The age of the patient
Neighborhood	Refers to the names of the neighborhoods where the health units are situated in the city of Vitória, in the State of Espírito Santo, Brazil
Financial aid	Assessing the patient's eligibility to obtain a governmental subsidy
Hypertension	Determining if the patient has elevated blood pressure
Diabetes	Whether the patient has diabetes
Alcoholism	Whether the patient has alcoholism
Handicap	Determining if the patient possesses a permanent physical limitation
SMS received	Whether the patient received an SMS
Appointment Status	Whether the patient attended the appointment

3.2 Data preprocessing

In this study, several critical preprocessing steps were undertaken, which included the creation of new variables, removal of irrelevant features, data transformation, normalization, and addressing class imbalances. Through these preprocessing steps, the dataset was optimized for subsequent analysis and model development.

3.2.1 New variable creation

Several new variables were created during the data preprocessing stage, including 'cumulative-appointment count ', 'cumulative no-show count,' 'no-show cumulative proportion,' 'scheduled day of the week,' 'appointment day of the week,' and 'appointment lead time.' These variables capture crucial aspects of the appointment and patient behavior. To enhance the handling of categorical data, 'scheduled day of the week' and 'appointment day of the week' were one-hot encoded. Table 4 provides a summary of these new variables, along with their descriptions.

Table 4 A summary of the new variables created during the data preprocessing phase

New Variable	Description
Cumulative-Appointment Count	Total number of appointments made by each patient
Cumulative No-Show Count	Cumulative sum of 'No-show' occurrences for each patient

New Variable	Description
No-Show Cumulative Proportion	Proportion of no-shows based on a patient's previous appointment(s)
Scheduled Day of the week	Day of the week for the scheduled appointment
Appointment Day of the week	Day of the week for the actual appointment
Appointment Lead Time	Number of days between the Scheduled Day and the actual appointment
Number of Appointments on Same Day	Number of appointments a patient has on the same day

3.2.2 Data cleaning

During the data cleaning process, several important steps were taken to enhance the quality and reliability of the dataset. First, an age range filter was applied to include only valid ages between 0 and 100, ensuring the consistency and validity of the age values. Additionally, appointments with negative waiting times, indicative of scheduling errors, were removed to maintain the chronological integrity of the data. Furthermore, Saturday appointments were excluded from the dataset, in accordance with the specification to focus solely on non-off days. These meticulous data cleaning procedures ensure that a clean and accurate dataset is provided, establishing a solid foundation for subsequent analysis and modeling.

3.2.3 Normalization

Normalization is an essential data preprocessing step, and one commonly used technique is MinMaxScaler. MinMaxScaler scales the features in the dataset to a specified range, typically between 0 and 1, by subtracting the minimum value and dividing by the range of the data. This ensures that all features are on a similar scale, preventing any single feature from dominating the learning process due to its larger magnitude. The formula used for this linear transformation is:

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

where x is the original feature, min(x) is the minimum value of the feature, and max(x) is the maximum value of the feature. The resulting scaled feature, x_{scaled}, will fall within the range of 0 to 1.

3.2.4 Handling class imbalance

To address the class imbalance in the dataset, the SMOTE was employed [39]. SMOTE is a widely used approach in machine learning that generates synthetic samples for the minority class by

interpolating between existing instances. This helps balance the class distribution and improve model performance[40].

As depicted in Fig.2, An illustration of imbalanced distribution in target variable classes, the initial dataset exhibited a notable class imbalance. The majority class comprised 88,207 instances labeled "show", contrasting with a considerably smaller count of 22,319 instances for the minority class labeled "no-show". The percentage of patients who presented was 79.81%, while the percentage of patients who did not show up was 20.19%. By applying SMOTE, the dataset was rebalanced, resulting in an equal representation of both classes with 88,207 instances each. SMOTE randomly selects minority class instances, identifies their nearest neighbors, and generates synthetic samples by interpolating between them. This process continued until an equal distribution was achieved. The application of SMOTE increased the representation of the minority class in the dataset, providing additional information for the machine learning model during training. This rebalancing significantly improved the model's ability to accurately classify instances, particularly when dealing with an under-represented minority class.

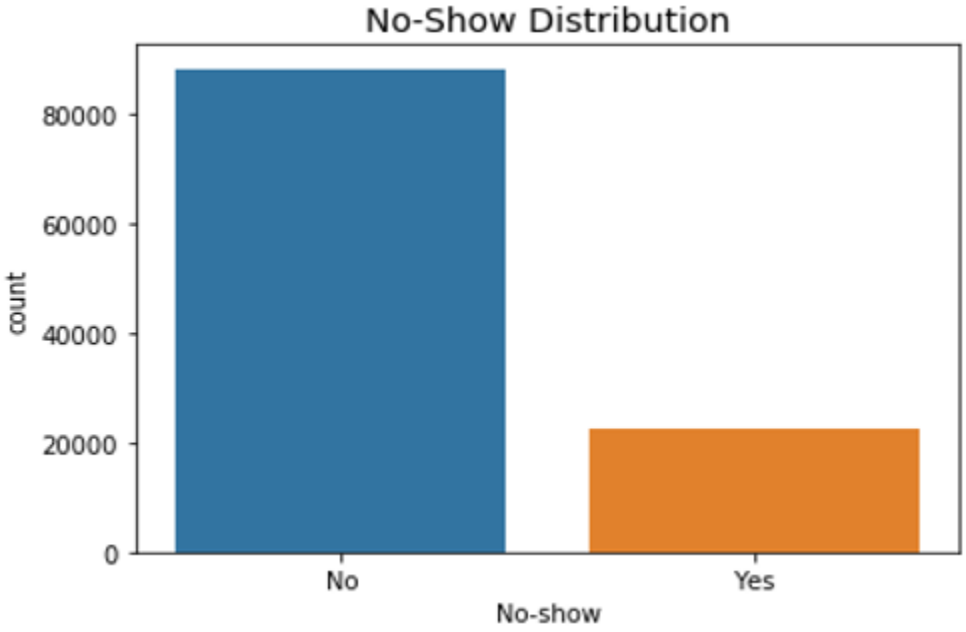

Fig.2 An illustration of the imbalanced distribution in target variable classes

3.3 Feature Selection

Feature selection is a critical aspect of machine learning, and recursive feature elimination (RFE) is a widely recognized technique [41]. RFE systematically eliminates less important features by

training a model with all features and progressively removes the least weighted features until the desired number of crucial features is attained.

3.4 Model evaluation

Model evaluation is a critical step in developing predictive models because it enables researchers to assess the performance, robustness, and generalizability of these models. This evaluation process provides valuable insights into how well a model is likely to perform on unseen data and helps make informed decisions about model selection and refinement.

3.4.1 Cross-validation

In a typical n-fold cross-validation method, a dataset is divided into n equal, non-overlapping subsets (folds). Each fold is the test set, while the remaining (n-1) folds are used for training. This process is repeated n times, providing an average accuracy estimate. Seo et al. [42] utilized cross-validation to validate their machine learning models, and achieved superior predictive accuracy, especially with extreme gradient boosting models incorporating unstructured text data, for hospitalization likelihood within 24 hours and waiting time estimation. Stratified cross-validation ensures balanced class distributions among the folds, enhancing accuracy estimation compared to regular cross-validation. Kohavi [43] found stratified cross-validation to be superior in terms of bias and variance.

3.4.2 Evaluation metrics for predictive models

The confusion matrix summarizes a classification model's performance with true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). It facilitates the computation of key metrics such as accuracy, precision, recall, and F1-score. The matrix helps assess a classifier's strengths and weaknesses, aiding decisions on its suitability. Additionally, the area under the ROC curve is used as an evaluation metric. Various metrics, including accuracy, recall, specificity, precision, F1 score, AUC-ROC, FPR, and FNR, were used to assess model performance. Equations 2 to 7 provide the formulas for these evaluations.

$$Accuracy = \frac{TP+TN}{TN+FN+TP+FP} \tag{2}$$

$$Recall/Sensitivity/TPR = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{5}$$

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

$$AUC\text{-}ROC = \int_0^1 TPR.\, d(FPR) \quad (7)$$

3.4.2 TOPSIS

TOPSIS, which stands for technique for order of preference by similarity to ideal solution, is a multi-criteria decision-making method widely used in operations research and decision analysis. Developed to address complex decision scenarios with multiple conflicting criteria, TOPSIS helps in selecting the most suitable alternative from a set of options [44-46]. The method involves creating a matrix that represents the performance of each alternative against each criterion, followed by normalization of the matrix and determination of weighted normalized scores. Subsequently, a positive ideal solution and a negative ideal solution are identified, and the alternatives are ranked based on their proximity to the positive ideal and distance from the negative ideal. This ranking assists decision-makers in choosing the optimal solution that best balances the conflicting criteria and aligns with their objectives. In this study, TOPSIS was used to rank classifiers based on performance indicators [47].

3.5 Predictive modeling

Predictive modeling plays a crucial role in machine learning, enabling the development of models that can make accurate predictions based on historical data. In this section, we explore various techniques used for predictive modeling.

3.5.1 Decision trees

Decision trees are widely used in machine learning for classification and regression because they offer interpretable decision rules [48]. In this study, these trees were constructed by iteratively selecting features and split points, with the 'max_depth' hyperparameter set to a maximum value of 6, limiting the process. This hyperparameter controls the tree's complexity and helps prevent overfitting, enhancing generalization**.** Praveena et al. [49] applied the decision tree classifier, which achieved 89.6% accuracy, highlighting the efficacy of decision trees in predicting patient no-shows.

3.5.2 Random Forest

Random forest is an ensemble technique that combines multiple decision trees to create a precise and robust model [50]. Using random subsets of features and training data reduces overfitting and boosts accuracy. In this study, the 'n_estimators' were set to 300 to achieve the optimal balance between computational efficiency and model performance. Salazar et al. [51] addressed patient

no-shows in healthcare, developing a random forest-based predictive model with a recall rate of 0.91 and AUC of 0.969.

3.5.3 Gradient Boosting

Gradient boosting, an ensemble algorithm, combines weak models sequentially to build a robust predictive model. This algorithm minimizes errors by adjusting the parameters and minimizing the loss function, guided by the negative gradient. This process enhances the performance, reduces the bias and variance, and accommodates various data types and loss functions.

3.5.4 XGBoost

XGBoost, or extreme gradient boosting, is a fast and accurate machine learning algorithm. It builds an ensemble of decision trees, with each tree correcting prior errors. The algorithm optimizes an objective function using a loss function and regularizer to prevent overfitting, including L1 and L2 regularization. It iteratively fits new trees by calculating negative gradients, updating sample weights, and continuing until convergence, yielding an effective model.

3.5.6 LightGBM

LightGBM is a high-speed, memory-efficient gradient boosting framework for large-scale datasets [52]. It employs memory optimization techniques such as gradient-based one-sided sampling (GOSS) and exclusive feature bundling (EFB) [53]. Additionally, it uses a histogram-based algorithm for faster decision tree construction.

3.5.7 Neural networks

Neural networks, inspired by the human brain, process data through interconnected layers of nodes called neurons. These networks learn by adjusting the neuron weights and biases based on training data [54, 55]. A deep neural network (DNN) demonstrated superior performance in predicting patient no-show visits with a weighted average precision of 98.2% and recall of 94.3%, suggesting its potential to enhance healthcare efficiency and effectiveness[28].

The proposed model is a feedforward neural network with four dense layers, consisting of 128, 64, 32, and 1 neuron. The input layer size matches the input data's feature count. The first three hidden layers use ReLU activation to capture complex input-output relationships. The final layer uses sigmoid activation to provide probabilistic estimates from 0 to 1. To improve the model's performance, the Adam optimizer, a variant of stochastic gradient descent that dynamically adjusts the learning rate, is utilized. The binary cross-entropy loss function measures the difference between the predicted probabilities and actual binary labels. The model is trained for 100 epochs

with a batch size of 32. The number of neurons in each layer is determined based on the problem complexity and input data scale. ReLU activation is used for hidden layers due to its effectiveness, while sigmoid activation is chosen for the output layer to provide probabilistic estimations.

3.5.8 Stacking

Stacking is an ensemble technique in which multiple models are used to create a more accurate and robust final model. The process involves training individual models on subsets of data and using their predictions to train a meta-model. The proposed stacking model incorporates a neural network as the meta-model at the second level. In the first level, various machine learning algorithms such as XGBoost, LightGBM, gradient boosting, neural network, decision tree, and random forest, are utilized. The predictions generated by these individual models are used as inputs for the neural network at the second level. Overall, implementing the stacking algorithm with a neural network model proves to be an effective technique for improving prediction accuracy in binary classification scenarios. By combining multiple first-level models with a robust neural network meta-model, a highly accurate and resilient model is constructed.

4 Results and discussion

4.1 Descriptive analysis

provides an overview of the descriptive analysis for variables, presenting descriptive statistics such as frequency distributions with percentages for categorical variables and means/standard deviations for continuous variables. Notably, there was a greater proportion of females (64.99%) than males (35.01%), the majority of whom were not affiliated with (90.17%), and a significant proportion were not associated with hypertension (80.28%). Moreover, the average age was 37.08 years, with a standard deviation of 23.10 years, and the mean appointment lead time was 10.19 days, with a standard deviation of 15.26 days. The table also reveals that 79.81% of individuals attended their appointments, while 20.19% did not. This information is crucial because it highlights the appointment attendance rate, which can be valuable for analyzing and improving scheduling and healthcare management.

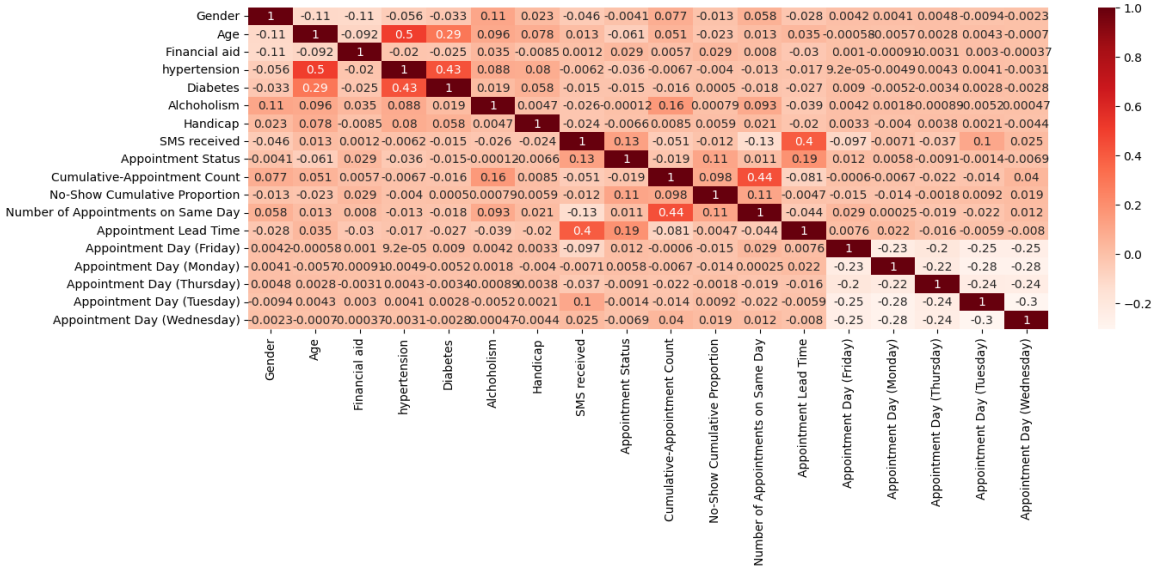

Fig. **3** presents a heatmap depicting the correlation matrix for the dataset variables. Warmer cell colors indicate higher correlations between features. The highest correlation values are as follows: 0.50 for Age and Hypertension, 0.44 for the Number of Appointments on the Same Day by each patient and Cumulative-Appointment Count by each patient, and 0.43 for Hypertension and Diabetes.

Table 5 Overview of descriptive analysis of the variables

Column	Frequency	Percentage
Gender		
0 (Female)	71,802	64.99%
1 (Male)	38,675	35.01%
Scholarship		
0 (No)	99,619	90.17%
1 (Yes)	10,858	9.83%
Hypertension		
0 (No)	88,696	80.28%
1 (Yes)	21,781	19.72%
Diabetes		
0 (No)	102,541	92.82%
1 (Yes)	7,936	7.18%
Alcoholism		

Column	Frequency	Percentage
0 (No)	107,119	96.96%
1 (Yes)	3,358	3.04%
Handicap		
0 (No handicap)	108,243	97.98%
1 (Not significant))	2,035	1.84%
2 (Mild handicap)	183	0.17%
3 (Severe handicap)	13	0.01%
4 (Extreme handicap)	3	0.00%
SMS_received		
0 (No)	75,009	67.90%
1 (Yes)	35,468	32.10%
Appointment Status		
0 (show)	88,175	79.81%
1 (no-show)	22,302	20.19%
Appointment Day of the week		
Wednesday	25,866	23.41%
Tuesday	25,638	23.21%
Monday	22,711	20.56%
Friday	19,018	17.21%
Thursday	17,244	15.61%
	Mean	**Standard Deviation**
Age	37.08	23.10
Appointment Lead Time	10.19	15.26
No-Show Cumulative Proportion	0.09	0.25
Cumulative No-Show Count	0.41	0.80
Cumulative-Appointment Count	2.27	3.91

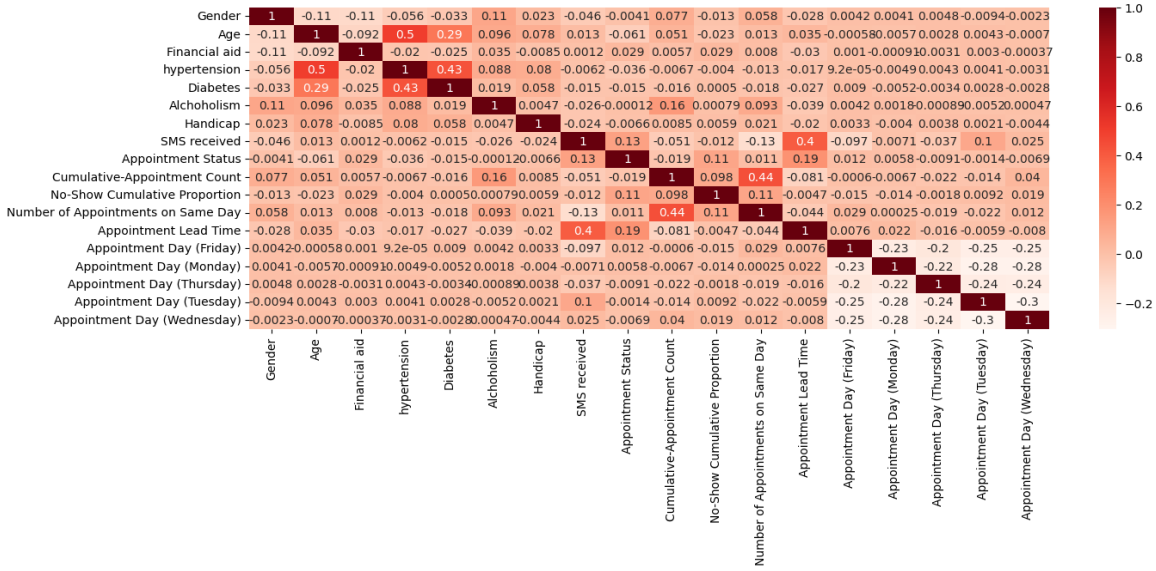

Fig. 3 Heatmap of the correlation matrix for the dataset variables

4.2 Performance Comparison of Different Algorithms

In evaluating the performance of various machine learning models, we considered key metrics including AUC, accuracy, recall, precision, and F1-score, and the results are presented in Table 6. The stacking model emerged as the top performer with an AUC of 0.91, demonstrating its robust discriminative ability. This model also achieved an accuracy of 0.82, along with balanced recall (0.82), precision (0.85), and F1-score (0.83), reflecting a harmonious trade-off between identifying positive instances and minimizing false positives.

Comparatively, the neural network exhibited lower performance, with an AUC of 0.76 and an accuracy of 0.68. Despite its lower scores, it demonstrated consistency in recall (0.68), precision (0.72), and F1-score (0.69). The decision tree, gradient boosting, random forest, Xgboost, and Lgbm models exhibited varying degrees of performance, with gradient boosting showcasing strong AUC (0.89), F1-score (0.82), and Lgbm demonstrating high precision (0.79). These results provide valuable insights into the comparative strengths and weaknesses of the models, aiding in the selection of the most suitable approach for the given task. Fig. 4 displays the AUC-ROC curves for various algorithms, offering a comprehensive performance overview across classification thresholds. This curve illustrates the sensitivity-specificity trade-off, facilitating algorithm comparisons.

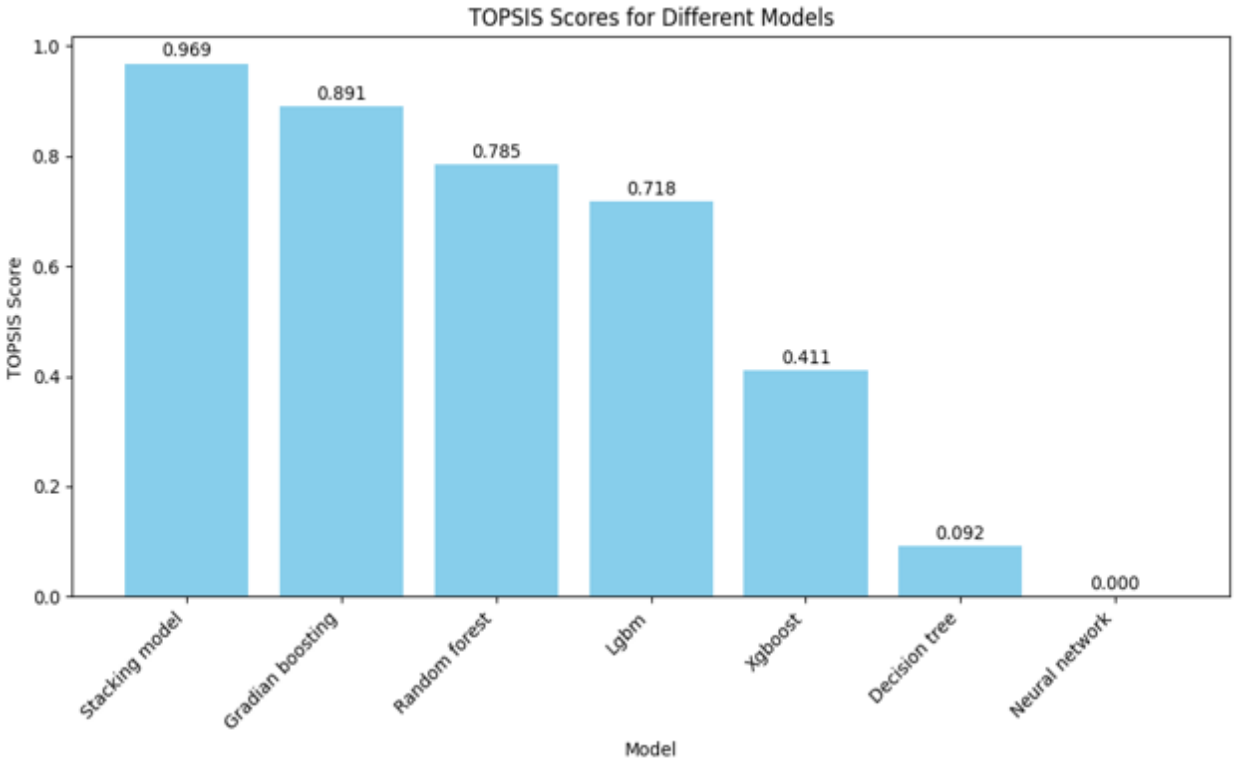

Fig. **5** displays the TOPSIS scores, emphasizing the superior performance of the stacking algorithm. This approach streamlines decision-making, aiding in the selection of classifiers aligned with specific evaluation criteria.

Table 6 Performance of the Machine Learning Models on the Test Dataset

Model	AUC	Accuracy	Recall	Precision	F1-score
Stacking model	0.91	0.82	0.82	0.85	0.83
Neural network	0.76	0.68	0.68	0. 72	0.69
Decision tree	0.76	0.680	0.69	0.74	0.71
Gradian boosting	0.89	0.81	0.82	0.82	0.82
Random forest	0.88	0.79	0.80	0.80	0.80
Xgboost	0.83	0.73	0.74	0.75	0.74
Lgbm	0.88	0.78	0.78	0.79	0.79

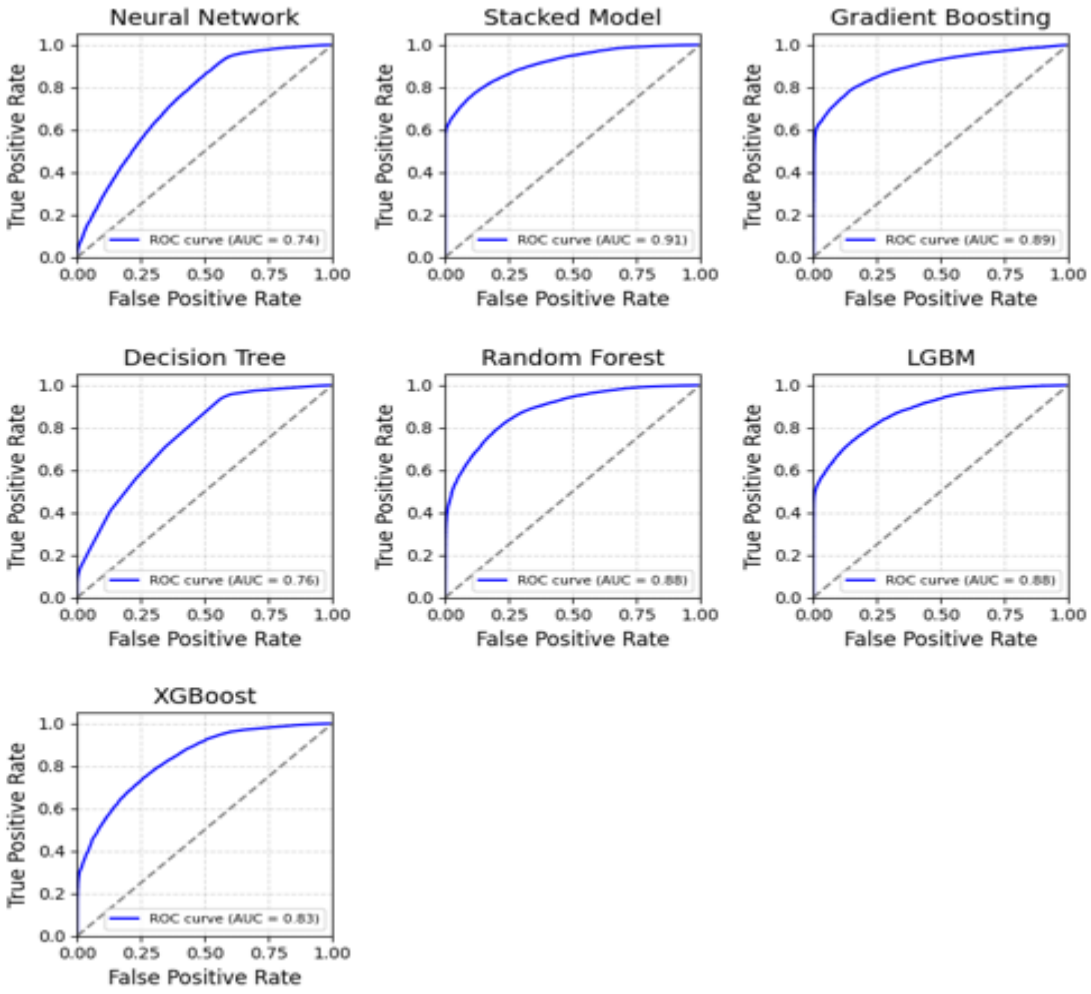

Fig. 4 ROC curves of the Machine Learning Models

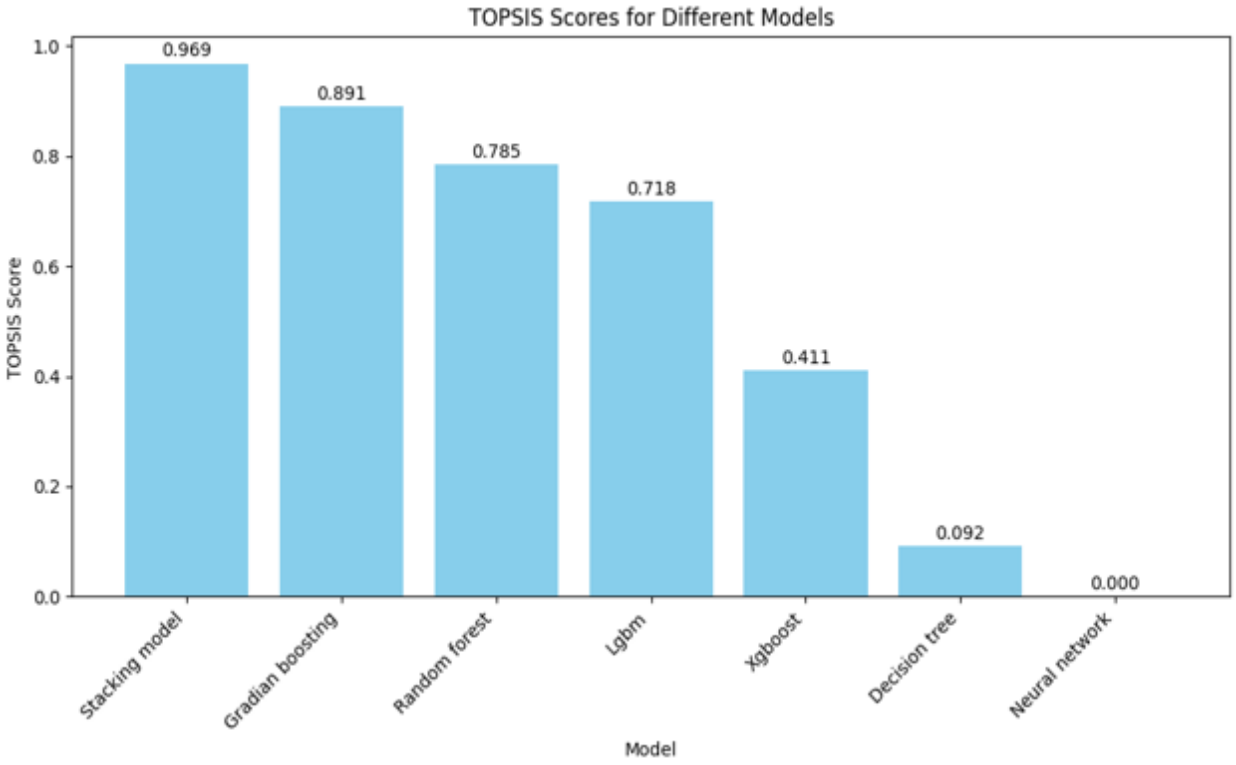

Fig. 5 Comparative Ranking of Models Using TOPSIS Scores

4.3 Important predictor variables

In the feature selection process, RFE was employed with three distinct classifiers: logistic regression, random forest, and gradient boosting. The objective was to identify robust features that consistently contribute to predictive performance across multiple algorithms. Fig. 6 displays the ranking of each feature according to the three classifiers. The numbers represent the ranking assigned by each algorithm, where a lower number indicates a higher ranking. Features that are consistently highly ranked across classifiers are likely to be more influential in predictive models. Based on the rankings, our aim was to identify features that were consistently deemed important by at least two out of the three classifiers.

Several variables have been identified as essential for predicting patient appointment no-shows. Notable factors, including 'Age,' 'Cumulative-Appointment,' 'Appointment Lead Time,' 'No-Show Cumulative Proportion,' 'Financial Aid,' 'Alcoholism,' 'SMS Received,' and 'Number of Appointments on the Same Day,' were selected for further examination due to their significant impact. Analyzing these variables provides valuable insights into the factors affecting patient attendance and the effectiveness of appointment management strategies.

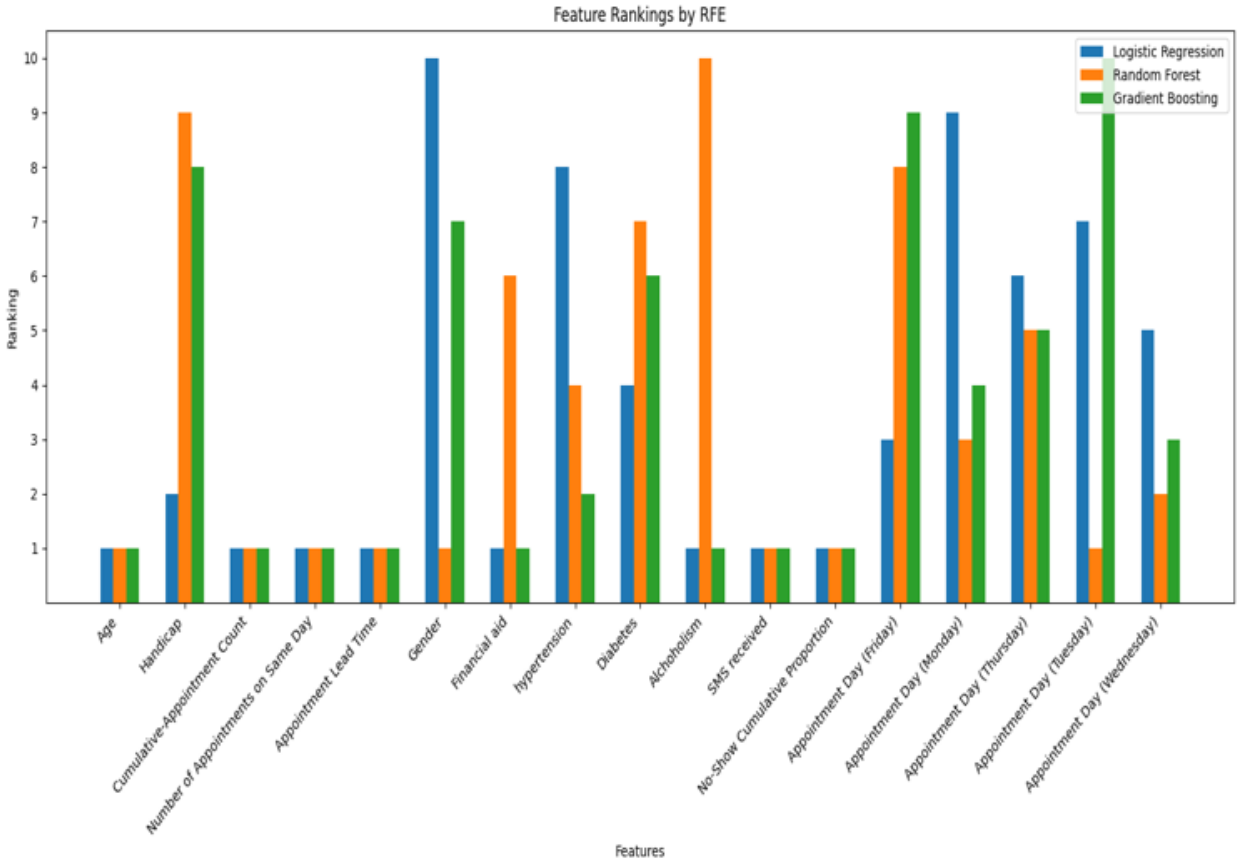

Fig. 6 Illustration of the ranking of each feature obtained through Recursive Feature Elimination using three distinct classifiers: logistic regression, random forest, and gradient boosting

In addition to the factors mentioned above, patient characteristics such as income, marital status, and race can influence non-attendance behavior [2, 56]. Moreover, it has been observed that non-attendance rates tend to be higher on Mondays and for afternoon appointments, possibly because the face-to-face system is typically closed on weekends [13]. Several studies have shown that weather conditions on the day of patient visits to face-to-face clinics are an important predictor of non-attendance behavior [13, 57]. Su et al. [35] identified key factors associated with non-attendance, which included age, the time interval between appointment scheduling, prior missed appointments, and clinical referral. Notably, their research revealed no significant relationship between non-attendance and gender, day of the week, or appointment cost.

In a rural population with limited healthcare access, several predictive factors for missed appointments were identified. These factors include prior no-show rates, scheduling-to-appointment waiting time, season, day of the week, type of healthcare service, age, gender, and

language proficiency [58]. These findings provide valuable insights into the intricate determinants of missed appointments within underserved rural communities.

Other factors that have been shown to influence missed appointments in other studies include younger age [2, 59], distance to the clinic [2, 60, 61], and insurance company [62]. As mentioned by [63], it has been asserted that younger patients generally display a lower rate of no-shows for appointments. Nevertheless, Dantas et al. [61] suggested that there is no significant effect of age on the likelihood of missed appointments. The waiting period for an appointment significantly influences the likelihood of no-shows. Waiting for more than six months is associated with significantly greater no-show rates than a one-week wait [64].

5 Conclusions, limitations, and prospects for future research

In this research study, an innovative machine learning framework was proposed to address the challenge of patient no-show occurrences in healthcare institutions. Through rigorous preprocessing techniques and advanced modeling approaches, the aim was to improve the predictive capabilities of the models and enhance operational efficiency in outpatient clinics. The methodology involved various steps, including data preprocessing, feature selection, oversampling, cross-validation, algorithm selection, hyperparameter tuning, and model stacking. These steps were carefully designed to ensure the models' accuracy and reliability in predicting patient no-shows. The effectiveness of the proposed framework in predicting patient no-shows was demonstrated by the results. Significant improvements in predictive accuracy and performance metrics were achieved by utilizing diverse machine learning algorithms and addressing class imbalance through SMOTE and stratified k-fold cross-validation. Adding new variables to the dataset and removing irrelevant features improved the models' ability to capture meaningful patterns and relationships within the dataset.

The findings of this study have practical implications for healthcare institutions. Accurate prediction of patient no-shows can help clinics optimize resource allocation, reduce costs, and enhance operational efficiency. By leveraging the insights gained from the predictive models, clinics can develop targeted strategies to mitigate the occurrence of no-shows and improve patient attendance rates.

The limitations of this research must be acknowledged. First, the study was conducted using a specific dataset sourced from Kaggle, which may not fully capture the complexity and variability of patient no-show occurrences in different healthcare settings. Further research should aim to

validate the findings on larger and more diverse datasets. Additionally, the generalizability of the predictive models should be tested in real-world healthcare environments to assess their performance in practical settings.

The proposed machine learning framework demonstrated its potential to improve operational efficiency and resource allocation in outpatient clinics. By addressing these limitations, conducting further validation studies, and exploring advanced modeling techniques, future research can build upon these findings and advance the field of patient attendance prediction in healthcare. Ultimately, successfully implementing predictive models can lead to better patient outcomes, reduced costs, and enhanced healthcare services.

Furthermore, while the proposed framework achieved notable improvements in predictive accuracy, there is still room for future work to enhance the models' performance. For instance, exploring ensemble techniques and leveraging more advanced deep learning architectures could yield even better results. Additionally, investigating the temporal dynamics and seasonal patterns of patient no-show occurrences could provide valuable insights for designing targeted interventions. Incorporating additional data sources, such as weather, traffic, and socio-economic factors, can improve predictive models by providing a more comprehensive understanding of patient attendance. Real-time prediction models and proactive interventions such as personalized reminders and rescheduling options can reduce patient no-shows, enhancing resource allocation and patient satisfaction. Comparative analysis across healthcare institutions, regions, or countries can reveal attendance variations and the effectiveness of mitigation strategies, guiding the development of tailored solutions. Ethical considerations, including fairness, bias, and privacy, must be explored to establish responsible guidelines for predictive model usage. Long-term impact assessment through longitudinal studies is crucial to understanding the sustained benefits and challenges associated with implementing predictive models and interventions.

References

[1] Gupta, D. and B. Denton, *Appointment scheduling in health care: Challenges and opportunities.* IIE Transactions, 2008. **40**(9): p. 800-819.

[2] Daggy, J., et al., *Using no-show modeling to improve clinic performance.* Health Informatics J, 2010. **16**(4): p. 246-59.

[3] Nguyen, D.L., R.S. Dejesus, and M.L. Wieland, *Missed appointments in resident continuity clinic: patient characteristics and health care outcomes.* J Grad Med Educ, 2011. **3**(3): p. 350-5.

[4] Glowacka, K.J., R.M. Henry, and J.H. May, *A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling.* Journal of the Operational Research Society, 2009. **60**(8): p. 1056-1068.

[5] Fan, G., et al., *Machine learning-based prediction models for patients no-show in online outpatient appointments.* Data Science and Management, 2021. **2**: p. 45-52.
[6] Berg, B.P., et al., *Estimating the Cost of No-Shows and Evaluating the Effects of Mitigation Strategies.* Medical Decision Making, 2013. **33**(8): p. 976-985.
[7] Turkcan, A., et al. *No-Show Modeling for Adult Ambulatory Clinics*. 2013.
[8] Drabkin, M.J., et al., *Telephone reminders reduce no-shows: A quality initiative at a breast imaging center.* Clin Imaging, 2019. **54**: p. 108-111.
[9] Dantas, L.F., et al., *No-shows in appointment scheduling – a systematic literature review.* Health Policy, 2018. **122**(4): p. 412-421.
[10] Liu, N., *Optimal Choice for Appointment Scheduling Window under Patient No-Show Behavior.* Production and Operations Management, 2016. **25**(1): p. 128-142.
[11] Kheirkhah, P., et al., *Prevalence, predictors and economic consequences of no-shows.* BMC Health Services Research, 2016. **16**(1): p. 13.
[12] Abdus-Salaam, H. and L.B. Davis, *Impact of batch appointments on no-show rates: a public vs private clinic perspective.* Health Systems, 2012. **1**(1): p. 36-45.
[13] Peng, Y., et al., *Large-scale assessment of missed opportunity risks in a complex hospital setting.* Inform Health Soc Care, 2016. **41**(2): p. 112-27.
[14] Kopach, R., et al., *Effects of clinical characteristics on successful open access scheduling.* Health Care Manag Sci, 2007. **10**(2): p. 111-24.
[15] Parikh, A., et al., *The effectiveness of outpatient appointment reminder systems in reducing no-show rates.* Am J Med, 2010. **123**(6): p. 542-8.
[16] Hasvold, P.E. and R. Wootton, *Use of telephone and SMS reminders to improve attendance at hospital appointments: a systematic review.* J Telemed Telecare, 2011. **17**(7): p. 358-64.
[17] Liu, C., et al., *Text Message Reminders Reduce Outpatient Radiology No-Shows But Do Not Improve Arrival Punctuality.* J Am Coll Radiol, 2017. **14**(8): p. 1049-1054.
[18] Alaeddini, A., et al., *A probabilistic model for predicting the probability of no-show in hospital appointments.* Health Care Management Science, 2011. **14**(2): p. 146-157.
[19] Samorani, M. and L.R. LaGanga, *Outpatient appointment scheduling given individual day-dependent no-show predictions.* European Journal of Operational Research, 2015. **240**(1): p. 245-257.
[20] Abdus-Salaam, H. and L.B. Davis, *Tactical allocation and acceptance of multiple patient classes in the presence of no-shows.* Health Systems, 2015. **4**(2): p. 93-103.
[21] Rochefort, C.M., et al., *A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data.* Journal of the American Medical Informatics Association, 2015. **22**(1): p. 155-165.
[22] Tabak, Y.P., et al., *Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS).* Journal of the American Medical Informatics Association, 2014. **21**(3): p. 455-463.
[23] Dunstan, J., et al., *Predicting no-show appointments in a pediatric hospital in Chile using machine learning.* Health Care Management Science, 2023. **26**(2): p. 313-329.
[24] Lenzi, H., Â.J. Ben, and A.T. Stein, *Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil.* PLOS ONE, 2019. **14**(4): p. e0214869.
[25] Srinivas, S. and H. Salah, *Consultation length and no-show prediction for improving appointment scheduling efficiency at a cardiology clinic: A data analytics approach.* International Journal of Medical Informatics, 2021. **145**: p. 104290.
[26] Aladeemy, M., et al., *New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows.* Applied Soft Computing, 2020. **86**: p. 105866.
[27] Lekham, L.A., *Two-Step Predictive Model for Missed Appointments at Outpatient Primary Care Settings Serving Rural Areas*. 2020, State University of New York at Binghamton.

[28] Alshammari, R., T. Daghistani, and A. Alshammari, *The Prediction of Outpatient No-Show Visits by using Deep Neural Network from Large Data*. International Journal of Advanced Computer Science and Applications, 2020. **11**(10).
[29] Dashtban, M. and W. Li, *Predicting non-attendance in hospital outpatient appointments using deep learning approach.* Health Systems, 2022. **11**(3): p. 189-210.
[30] Chen, J., et al., *Application of Machine Learning to Predict Patient No-Shows in an Academic Pediatric Ophthalmology Clinic.* AMIA Annu Symp Proc, 2020. **2020**: p. 293-302.
[31] Liu, D., et al., *Machine learning approaches to predicting no-shows in pediatric medical appointment.* npj Digital Medicine, 2022. **5**(1): p. 50.
[32] Mohammadi, I., et al., *Data Analytics and Modeling for Appointment No-show in Community Health Centers.* J Prim Care Community Health, 2018. **9**: p. 2150132718811692.
[33] Kurasawa, H., et al., *Machine-Learning-Based Prediction of a Missed Scheduled Clinical Appointment by Patients With Diabetes.* J Diabetes Sci Technol, 2016. **10**(3): p. 730-6.
[34] Goffman, R.M., et al., *Modeling Patient No-Show History and Predicting Future Outpatient Appointment Behavior in the Veterans Health Administration.* Mil Med, 2017. **182**(5): p. e1708-e1714.
[35] Su, W., et al., *Who Misses Appointments Made Online? Retrospective Analysis of the Outpatient Department of a General Hospital in Jinan, Shandong Province, China.* Risk Manag Healthc Policy, 2020. **13**: p. 2773-2781.
[36] Luo, L., et al., *Machine learning for identification of surgeries with high risks of cancellation.* Health Informatics J, 2020. **26**(1): p. 141-155.
[37] Liu, L., et al., *Mining patient-specific and contextual data with machine learning technologies to predict cancellation of children's surgery.* International Journal of Medical Informatics, 2019. **129**: p. 234-241.
[38] Odonkor, C.A., et al., *Factors Associated With Missed Appointments at an Academic Pain Treatment Center: A Prospective Year-Long Longitudinal Study.* Anesthesia & Analgesia, 2017. **125**(2).
[39] Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique.* Journal of artificial intelligence research, 2002. **16**: p. 321-357.
[40] Devi, D., S.K. Biswas, and B. Purkayastha, *Correlation-based Oversampling aided Cost Sensitive Ensemble learning technique for Treatment of Class Imbalance.* Journal of Experimental & Theoretical Artificial Intelligence, 2022. **34**(1): p. 143-174.
[41] Chen, X.-w. and J.C. Jeong. *Enhanced recursive feature elimination*. in *Sixth international conference on machine learning and applications (ICMLA 2007)*. 2007. IEEE.
[42] Seo, H., et al., *Prediction of hospitalization and waiting time within 24 hours of emergency department patients with unstructured text data.* Health Care Management Science, 2023.
[43] Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. in *Ijcai*. 1995. Montreal, Canada.
[44] Shafaei, R. and A. Mozdgir, *Master surgical scheduling problem with multiple criteria and robust estimation.* Scientia Iranica, 2019. **26**(1): p. 486-502.
[45] Mozdgir, A. and R. Shafaei, *Case Mix Planning using The Technique for Order of Preference by Similarity to Ideal Solution and Robust Estimation: a Case Study*. International Journal of Engineering, 2017. **30**(9): p. 1362-1371.
[46] Afsordegan, A., et al., *Absolute order-of-magnitude reasoning applied to a social multi-criteria evaluation framework.* Journal of Experimental & Theoretical Artificial Intelligence, 2016. **28**(1-2): p. 261-274.
[47] Tashakkori, R., et al., *The prediction of NICU admission and identifying influential factors in four different categories leveraging machine learning approaches.* Biomedical Signal Processing and Control, 2024. **90**: p. 105844.
[48] Yaesoubi, R., et al., *Generating simple classification rules to predict local surges in COVID-19 hospitalizations.* Health Care Management Science, 2023. **26**(2): p. 301-312.

[49] Praveena, M.D.A., J.S. Krupa, and S. SaiPreethi. *Statistical Analysis Of Medical Appointments Using Decision Tree*. in *2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*. 2019.
[50] Shahid Ansari, M., et al., *Identification of predictors and model for predicting prolonged length of stay in dengue patients.* Health Care Management Science, 2021. **24**(4): p. 786-798.
[51] Salazar, L.H.A., et al., *Application of Machine Learning Techniques to Predict a Patient’s No-Show in the Healthcare Sector.* Future Internet, 2022. **14**(1): p. 3.
[52] Ke, G., et al., *Lightgbm: A highly efficient gradient boosting decision tree.* Advances in neural information processing systems, 2017. **30**.
[53] Haghbayan, M., et al., *Increasing the Efficacy of Umbilical Cord Blood Banking Using Machine Learning Algorithms: A Case Study from Royan Cord Blood Bank.* Scientia Iranica, 2023: p. -.
[54] Mousavi, H., S.A. Darestani, and P. Azimi, *An artificial neural network based mathematical model for a stochastic health care facility location problem.* Health Care Management Science, 2021. **24**(3): p. 499-514.
[55] Zarrin, M., J. Schoenfelder, and J.O. Brunner, *Homogeneity and Best Practice Analyses in Hospital Performance Management: An Analytical Framework.* Health Care Management Science, 2022. **25**(3): p. 406-425.
[56] Lee, V.J., et al., *Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases.* BMC Health Services Research, 2005. **5**(1): p. 51.
[57] DeFife, J.A., et al., *Psychotherapy appointment no-shows: rates and reasons.* Psychotherapy (Chic), 2010. **47**(3): p. 413-417.
[58] Torres, O., et al., *Risk factor model to predict a missed clinic appointment in an urban, academic, and underserved setting.* Popul Health Manag, 2015. **18**(2): p. 131-6.
[59] Fiorillo, C.E., et al., *Factors associated with patient no-show rates in an academic otolaryngology practice.* Laryngoscope, 2018. **128**(3): p. 626-631.
[60] Guzek, L.M., W.F. Fadel, and M.R. Golomb, *A Pilot Study of Reasons and Risk Factors for "No-Shows" in a Pediatric Neurology Clinic.* J Child Neurol, 2015. **30**(10): p. 1295-9.
[61] Dantas, L.F., et al., *Predicting Patient No-show Behavior: a Study in a Bariatric Clinic.* Obesity Surgery, 2019. **29**(1): p. 40-47.
[62] Rosenthal, B.D., et al., *The effect of payer type on clinical outcomes in total knee arthroplasty.* J Arthroplasty, 2014. **29**(2): p. 295-8.
[63] Zhou, Y., D. Dong, and W. Jiang, *Influence Factors of Patient No Show in a Outpatient Department.* IOP Conference Series: Materials Science and Engineering, 2018. **439**: p. 032047.
[64] Rosenbaum, J.I., et al., *Understanding Why Patients No-Show: Observations of 2.9 Million Outpatient Imaging Visits Over 16 Years.* J Am Coll Radiol, 2018. **15**(7): p. 944-950.