Identification of Factors Correlating to Patient Appointment No-shows Using Deep & Machine Learning

Nazrinbanu Nagori
School of Professional Studies
Clark University
Worchester, MA
NNagori@clarku.edu

Kunal Malhan
School of Professional Studies
Clark University
Worchester, MA
KMalhan@clarku.edu

Emre Tokgoz
School of Professional Studies
Clark University
Worchester, MA
ETokgoz@clarku.edu

Khald Aboalayon
School of Professional Studies
Clark University
Worchester, MA
KAboalayon@clarku.edu

Hassan Musafer
School of professional studies
Clark University
Worchester, MA
HMusafer@clarku.edu

Corey Kiassat
Industrial & Systems Eng.
NC State University
Raleigh, NC
ackiassa@ncsu.edu

2025 IEEE International Conference on AI and Data Analytics (ICAD) | 979-8-3315-2472-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAD65464.2025.11114048

***Abstract*—Patient no-shows, scheduled but unattended medical appointments, pose a significant challenge in healthcare, impacting patient healthcare by disrupting the necessary treatment. Medical appointment no-shows disrupt healthcare providers' schedules that leads to suboptimal patient care. Such issues also lead to increased costs on health care, particularly in the medical field where resources are costly and in great demand. This study explores the factors influencing patient no-shows experienced by a multi-specialty clinic in the USA through the analysis of a dataset with more than 38000 medical records collected over a two-month period in 2019. Using descriptive modeling techniques, we examine key attributes such as patient demographics, appointment types, and scheduling patterns to identify trends associated with appointments missed by the patients. Our analysis reveals 22% of the appointments missed, with significant variations across different ethnic groups, appointment types, and days of the week. Notably, the Hispanic or Latino patient cluster exhibited a higher no-show rate in comparison to the others. Predictive modeling techniques used included Random Forest, Support Vector Machine (SVM), ANN, Gradient Boosting (GBM), and XGBoost to assess the feasibility of forecasting no-shows. The findings provided in this work provide an insight into appointment trends and disparities that may impact decision makers' approach to patient appointment scheduling.**

Keywords—Patient no-shows, Primary care, Missed appointments, Resource utilization, Healthcare disparities.

I. INTRODUCTION

Patient no-shows are a prevalent issue in healthcare settings, leading to significant inefficiencies and challenges in resource allocation [1,17,20]. Failed appointments result in inefficiencies and economic costs and may interrupt continuity of care [15,18,21]. Such issues cause irregularity in day-to-day healthcare operations and waste of human and space resources [2,6]. Different no-show rates were reported from 12% to 80% in various healthcare settings [2,3]. In the past decade, there have been several studies that identified the influence of gender, age, service quality, scheduling of preceding appointments and wait times on the no-show rates [3-5,7,8,11]. Another study showed that most no-shows are common among men, younger and patients belonging to lower socioeconomic profile [5,12,16]. Understanding the factors that contribute to no-show appointments is crucial for improving patient care and operational efficiency. In 2018, a study to find factors associated with no-show rates in an academic otolaryngology practice was performed and the findings suggested that the increased clinic no-show rates are associated with satellite clinics, new patient visits, younger age, and insurance types [7]. Many predictive analyses were performed to investigate factors associated with no-shows [9,13,17]. Among these, few studies have explored the influence of social determinants – such as lack of transportation, economic hardship, or limited access to childcare – on appointment adherence. For instance, research has shown that patients with limited transportation options or financial instability are more likely to miss appointments due to logistical or affordability issues [5, 22]. This study differs from such studies with a focus on other critical considerations such as appointment type, race, ethnicity, housing status, and communication consent history to investigate patients' no-show patterns. While we acknowledge the impact of broader socioeconomic factors, our analysis contributes an alternate perspective by examining trends based on available clinical and scheduling attributes. Predicted no-show appointments can be mitigated slightly by reminding patients of their upcoming appointments that can prevent them from forgetting or giving them the opportunity to cancel the appointment [14].

It's paramount to note that missed appointments cause not only logistic issues and delayed care to patients but also adversely affect healthcare revenue and day to day operations [19-21]. An analysis conducted in 2021 on the missed healthcare appointment data indicated a cost of more than 150 billion USD each year to the US health care system, and a cost of an average

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

of 200 USD per unused time slot individual to physicians [22,23]. In 2013, the cost of no-shows was studied for an outpatient endoscopy suite and the evaluated results revealed a daily loss of 725.42 USD to the health care facility that was attributed to a no-show rate of 18% with the associated costs ranging from 472.14 to 1019.29 USD over a time period based on the missed appointments' evaluations [23]. Such costs not only impact the health care provider but also cause an increased cost on the patients and impact the economy overall. Noting such considerations, and the recent significant developments in deep and machine learning techniques, we focus on descriptive and predictive modeling techniques applied on missed appointments by using health care data. The analysis presented focuses on information collected from a clinic located in the USA with the aim of analyzing patient characteristics, appointment types, and other socioeconomic profiles of a patient that may correlate to no-show rates. The descriptive and predictive modeling techniques targets to identify patterns and predict future no-shows with the following considerations:

- Descriptive modeling techniques are employed to examine key attributes such as patient demographics (age, gender, race, ethnicity, marital status, homelessness, veteran status), appointment types, and scheduling patterns to identify trends associated with missed appointments.
- Predictive analysis is utilized to develop models that can assist healthcare providers in forecasting patient behavior and mitigating the negative impact of no-shows by applying Logistic Regression, Artificial Neural Networks (ANN), and XGBoost classification.

The study is conducted by using approximately 38600 patient records that belonged to approximately 13000 patients in a health care clinic in the USA during October and November of 2019. The percentage of No Show or missed appointments recorded were 22% and to assess these No Show appointments with binary classification (i.e. yes or no options), we used measures including, appointment day and date, type of appointment, scheduled and cancelled slots, race, ethnicity, languages spoken, age, sex, marital status of patient, homeless status of the patient, public housing conditions of the patient, veteran status, and patient's consent agreement for calling or texting prior to the appointment as a reminder of the appointment. This research aims to enhance understanding of appointment adherence patterns that can help decision makers on scheduling health care appointments. The findings provide a foundation for further studies that may explore ways to optimize scheduling and patient engagement strategies in healthcare settings.

The next section outlines patient demographics and cleaning of the data for preparation to the analysis. The third section covers descriptive modeling applications, The fourth section focuses on the predictive model implementations with specific models selected for practical implementations on the data set. The fifth section includes results and discussion as a broad overview of the completed work. The last section is reserved for conclusive remarks with recommendations for health care providers with potential actions that can be taken.

II. Data Set's Characteristics and Cleaning

A. Patients Demographics

Our dataset includes patients with the age range of 1 to 100 years old, with the mean age around 40 years. Gender distribution is skewed towards females, comprising 63% of the dataset, while males make up 30.63% of the data. The racial composition is diverse, with White (27.19%), Other Race (27.15%), and Black (27.14%) being the predominant categories. A smaller proportion of patients identified as American Indian, Asian, or Pacific Islander, while 6.95% declined to disclose their race. Ethnicity data reveals that 63.97% of patients are identified to be Hispanic or Latino/Spanish, while 25.57% are Not Hispanic or Latino. Majority of the patients had either Spanish (52.62%) or English (45.43%) as the language preference, with a small fraction speaking languages such as Bengali, Arabic, French, and Urdu while 0.8% of the patients without a known language preference. Most patients are classified to be Single (68.41%) that is followed by Married (17.56%) and smaller proportions identified as Separated, Divorced, Widowed, or in a Partnership. The data also contained records of socioeconomic profiles of the patients such as housing conditions; 81.3% of the patients were not in public housing, while 13.1% were residing in public housing. A notable, 5.6%, of the patients were reported to be homeless, and 0.54% were confirmed to be homeless. Veteran status indicated that 93.1% of the patients were not veterans, while 4.67% had partial veteran status. Regarding consent preferences, most of the patients consented to receive phone calls (83.91%) and text messages (89.73%) regarding appointment scheduling reminders, facilitating easier communication.

B. Dataset Cleaning

The dataset was retrieved from a database by a research assistant and preprocessed prior to the application of the modeling applied in this work therefore it was clean, and there wasn't much work required for cleaning or preprocessing the data. The dataset had a few missing values that were filled with the *Unknown* in features such as age, ethnicity, language, marital status, veteran status, and homeless status. When the data on patient consent to call or text were not available, they were replaced with False meaning no consent. Moreover, the research team merged several categories for better classification and feature selection. For instance, three race categories e.g. Black, Black or African American, and African American were all merged into one.

III. Descriptive Modelling

There were approximately 38600 medical records that belonged to approximately 13000 patients out of which 7911 patients had multiple booked appointments with the clinic. Furthermore, we will analyze every feature data that has been collected and analyze no show or missed appointments with respect to some important features.

A. Features

- The ethnicity distribution in percentage terms highlights that most patients were identified as Hispanic or Latino/Spanish, comprising a significant 63% of the

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

data. Not Hispanic or Latino accounts for 27%, indicating a notable non-Hispanic population. Among Hispanic subgroups, Dominican (4%), Latin American (1%), and Puerto Rican (0.9%) are the most represented. Around 2.9% of the patients declined to disclose their ethnicity, and 2% were rated as Unknown. Smaller groups like Mexican (0.4%), South American (0.2%), and Central American (0.1%) reflected a diverse but concentrated Hispanic representation.

- The race feature represents the racial demographics of patients in the dataset. Most patients identify as Black or African American (34%), followed by White (28%) and Other Race (27%), indicating a diverse patient population. A considerable number of individuals (7%) declined to reveal racial identity therefore they are classified as "Patient Declined", which suggests privacy concerns or reluctance to disclose racial identity. Ethnic subgroups such as Dominican (3%) and American Indian or Alaska Native (1%) provide insights into the representation of Caribbean communities. The presence of Asian (0.5%), Asian Indian (0.1%), and Other Pacific Islander (0.05%) indicates diversity in the patient base, though these groups had lower representation. The "Unknown" (1%) category suggests missing or unrecorded racial data.
- The general clinic where the data is collected has comprehensive treatment procedures. The collected dataset has records of several treatment types with maximum appointment is determined to be follow up procedures that comprised 40% of the total appointments taken and minimum number of appointments recorded was determined to be for Psychiatry treatments (1.6%). Physical therapy and New Patient appointments accounted for 16 % and 8%, respectively. The clinic also had provision for routine annual examinations that made up 2.5 % of the patient appointments while 6% of the appointments were allocated for Gynecology and other pregnancy related treatments. All other treatment types such as lab work, Injection, Vision Therapy, BP check, allergy and sick visits constituted the remaining 15% of total treatments.

B. *Patients' No-show Data Analysis*

The no-show analysis of patients by race categorization provides valuable findings into appointment adherence patterns among different racial groups as it can be seen in Fig. 1. The dataset, after filtering out races with a no-show rate below 0.5%, reveals key disparities in patient attendance rates. Among identified racial groups, the data revealed Black or African American patients to have the highest no-show rate of 23.7%. Following closely, White patients had a no-show rate of 20.8% while other Race (20.1%) and Patient Declined (20.6%) categories also show similar rates, indicating that a substantial portion of scheduled appointments are missed within these demographics. Moreover, Dominican patients show a slightly lower no-show rate of 19.6%, followed by American Indian or Alaska Native patients at 16.6%. Interestingly, the Asian patient group demonstrates the lowest no-show rate at 12.5%, suggesting better adherence to scheduled appointments.

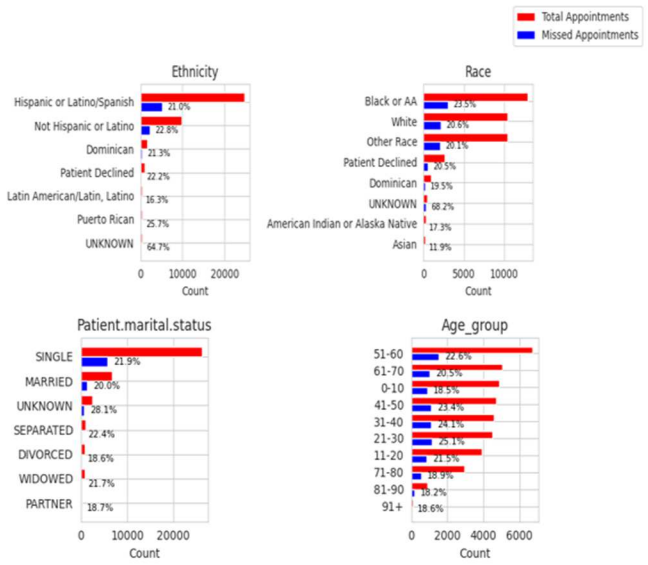


Fig. 1. Marital status, age-groups, ethnicity and race versus patients' no-show appointments.

Analyzing no-show rates across different ethnic groups reveals significant variations in appointment adherence. Among the listed ethnicities, Puerto Rican patients exhibit the highest no-show rate (25.53%). Non-Hispanic or Latino patients follow closely with a 23.04% no-show rate. Similarly, patients who declined to state their ethnicity show a high no-show rate (22.42%). Hispanic or Latino/Spanish (21.07%) and Dominican (21.35%) patients also exhibit moderate no-show rates based on our classification in this work. However, the Latin American/Latin, Latino group has the lowest no-show rate (16.17%), suggesting a relatively higher adherence to appointments compared to other ethnic groups as shown in Fig.1 although number of appointments are low for this ethnic group.

The no-show rate evaluation by focusing on the gender factor revealed no-show rates to be similar for both genders, with males (22.08%) having a slightly higher no-show rate than females (21.94%). This suggests that gender alone may not be a strong predictor of missed appointments; However, the total number of appointments for females is significantly higher than those for males, indicating that women may seek healthcare services more frequently.

The contextual analysis of the no-show rate for various age groups displayed in Fig. 1 suggests that young adults, especially between 21 to 30 years, were more prone to missing appointments at 25.1% and lowest age group missing appointments were 81 to 90 years at 18.2%. Noticeably patients in their fifties tend to book more appointments with about 22.6% of them classified to be in the no-show category. The data analysis reveals that marital status to be not a significant factor impacting appointment attendance rates. Even though the number of appointments is highest among all categories, patients of any marital status including single, married, divorced, or living with partner project No-show rate to be between 18% to 22%. Noticeably, Unknown marital status patients have the highest no-show rate at 28.12%, with a total of 2482 appointments among which 698 were missed by the patients.

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

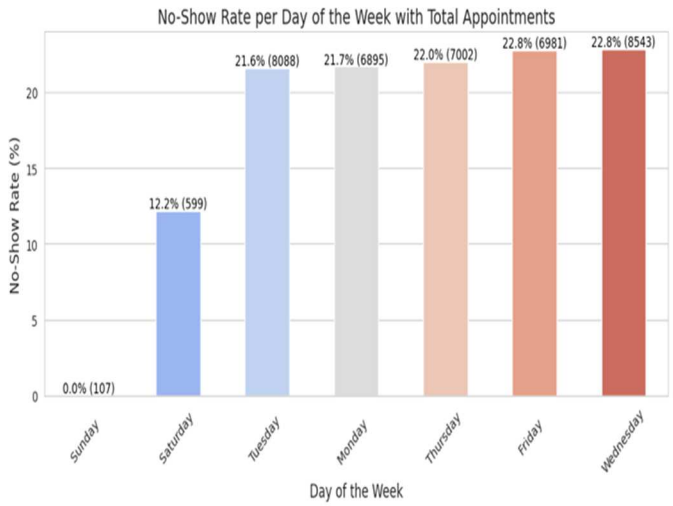


Fig. 2. Day of the Week vs No Show Appointments

Another major factor is the missed appointments during the day of the week. Fig. 2 demonstrates the No-Show Rate per Day of the Week (%) with the percentage of missed appointments for each weekday. Wednesday had the highest no-show rate at 22.85%, followed closely by Friday at 22.78% and Thursday at 22.01%. Monday (21.71%) and Tuesday (21.60%) had slightly lower but still significant no-show rates. Interestingly, Saturday has a much lower no-show rate at 12.19%, suggesting that patients are more likely to attend weekend appointments. Moreover, Sunday had no missed appointments although the number of appointments scheduled for this date was less than any other days of the week. Both observations could be due to the patient availability on non-working days.

Furthermore, our data analysis reveals important observations regarding the type of appointments as shown in Fig.3. The Annual exams, for instance, recorded 932 appointments with 282 missed visits, resulting in a relatively high no-show rate of 30.26%. This suggests that patients scheduled for annual exams may be more prone to cancellations or might face scheduling conflicts that lead to higher absenteeism. In contrast, the Follow Up category stands out with a massive volume of 14984 appointments, yet it exhibits a lower no-show rate of 22.54% despite having 3378 missed appointments. Gynecological services (GYN) show moderate engagement with 2158 total appointments and a no-show rate of 22.38% from 483 missed appointments, indicating a steady patient turnout. The New Patient group, however, demonstrates a concerning trend; with 2806 appointments and 947 missed visits, it has the highest no-show rate at 33.75%, potentially reflecting initial apprehensions or challenges in new patient onboarding. Notably, Psychiatry, with 616 appointments and only 95 out of them missed, has the lowest rate at 15.42%. These findings highlight distinct patterns of patient behavior across appointment types, providing crucial information for targeted interventions to reduce no-shows and enhance overall patient care for future research.

A total of 7911 total patients with records of no-show occurrences were determined with the least number of no-show occurrences to be zero, indicating that some patients did not miss any of their appointments. The patients that were classified into the lowest 25th and 50th (median) Percentile of the no-show category of the entire dataset was determined to be zero therefore the patients in these ranges did not miss any appointments. The 75th Percentile is determined to be the one indicating that three-quarters of patients missed at most one appointment. The highest number of missed appointments for a patient was 14, indicating the existence of patients with too many missed appointments. This analysis highlights the need for further investigation into factors influencing no-show behavior, as the distribution shows a concentration of patients with few or no missed appointments, while a small number of patients exhibit much higher rates of such missed appointments.

C. *Model and Feature Comparison*

Feature importance is analyzed through modeling on available dataset to analyze behavior of missed patient appointments against patient's demographic features, appointment type, and no-show history of patient. In this study, we used Decision Tree (DT), Random Forest (RF), Generalized Linear Model (GLM), and XGBoost (XGB) modeling techniques and determined the important features that played a role in strong modeling results.

Table 1 contains a summary of the important features that played a significant role in strong results attained for the models. Based on this table, it can be interpreted that patient with history of missing appointment has higher tendency to miss upcoming appointment. Additionally, appointment type is indicated as most influenced feature. Day of week for which appointment is booked, and patient's demographic features race, language, and age are also correlated to patient's chances of missing the appointment.

TABLE I. IMPORTANT FEATURES FOR EMPLOYED MODELS

Feature	DT	RF	GLM	XGB
Patient's history of no-show	√	√	√	√
Appointment type	√	√	√	√
Weekday of appointment			√	√
Patient's Race			√	√
Patient's Age		√		√
Patient's Language		√		
Patient's Homeless status			√	
Patient's consent to call		√		

IV. PREDICTIVE MODELING AND ANALSIS

Predictive analysis is performed on the dataset to analyze behavior of patients' missed appotinments by using patients' demographic features, appointment type, and no-show history of patient. The models used included Random Forest (Default), Random Forest - Model after SMOTE (MS), Support Vector Machines (SVM), Artifical Neural Network (ANN), Gradient Boosting Machines (GBM), and XGBoost (XGB) modeling techniques are used for predictive analysis and results.

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

TABLE II. FEATURES AND CORRESPONDING CATEGORIES SELECTED FOR PREDICTIVE MODELING

Feature	Original Category	Grouped Category
Race	American Indian, Mexican American Indian, Spanish American Indian, Standing Rock Sioux	American Indian/Alaska Native
	Asian, Asian Indian	Asian
	Black, Black or African American, African American, Jamaican	Black/African American
	White, European	White
	Dominican, Other Pacific Islander, Native Hawaiian or Other Pacific Islander, Patient Declined, UNKNOWN	Other
Ethnicity	Hispanic or Latino/Spanish, Dominican, Mexican, Cuban, Spaniard	Hispanic or Latino /Spanish
	Not Hispanic or Latino	Not Hispanic or Latino
	Latin American/Latin, Latino, South American, Central American, Honduran, Ecuadorian	Latin Amarican/Latin/Latino
	Puerto Rican, Patient Declined, UNKNOWN	Puerto Rican or Patient Declined
Language	English	English
	Spanish	Spanish
	Albanian, Arabic, Armenian, Bengali, Chinese, Danish, French, Italian, Portuguese, Serbian, Tagalog, Urdu, Vietnamese, Patient Declined, UNKNOWN	Other

A. Feature Selection and Data Preparation

We followed and performed the following data preparation steps prior to modeling noting the necessity for data to be in a specific format for the application of machine learning models:

- All walk-in records have been removed from datasets.
- Missing call and text consent data is assumed for consent to be not given.
- Any missing observation was removed.
- Race was categorized into American Indian/Alaska Native, Asian, Black/African American, White, or Other by using the original categorization of 17 levels.
- Ethnicity was categorized into Hispanic or Latino /Spanish, Not Hispanic or Latino, Latin Amarican/Latin/Latino, Puerto Rican or Patient Declined by using original categorization of 14 levels.
- Language was classified as English, Spanish, or Other by using the original categorization of 16 levels.
- A calculated field is added to indicate the Patient's history of no-show that calculates the number of instances in the history patients didn't show-up after having a scheduled appointment.

B. Parameter Selection for Predictive Modeling

A successful modeling approach requires the selection of appropriate parameters for the success of the models. We applied a few feature engineering techniques applied to create age groups from age feature as described in previous section.

Categorical parameters were encoded and converted to numerical features. Various classification models were trained which are Random Forest, SVM, GBM, XGBoost and simple feed forward Neural network (ANN). Using the grid search technique with 5-fold cross validation, best parameter values of *n_estimators* and *max_depth* for random forest classifier found to be 50 and 10 respectively. ANN was trained with relu activation functions and sigmoid in last layer which was used with threshold=0.5 for classification of no-show category(y/n).

Performances of various machine learning models are compared for a binary classification task, evaluating accuracy, precision (P), recall (R), F1-score, and the confusion matrix. Precision (P) and recall (R) represent the model's ability to correctly identify positive samples and its ability to detect all relevant positive samples, respectively. ANN (Deep Learning), XGBoost, and Random Forest (MS) models all have an accuracy of 0.87 with similar F1-scores while XGBoost resulting in the highest precision for class 1 (0.98). The RF (Default) model performs poorly with an accuracy of 0.82, making ANN, XGBoost, and Random Forest (MS) models more reliable. XGBoost is the best performing model due to its balance of precision and recall for class 1. Table 3 contains a summary of the results attained for the models selected models. Additionally, we identified the appointment type to be the most influential feature among the features.

TABLE III. MODEL PERFORMANCE METIRCS

Model	Accuracy	P(0) /R(0)	F1	P(1) /R(1)	F1	ROC-AUC
RF (Default)	0.82	0.89/0.89	0.89	0.59/0.59	0.59	0.87
RF (MS)	0.87	0.93/0.81	0.86	0.83/0.94	0.88	0.51
SVM	0.86	1.00/0.73	0.84	0.78/1.00	0.88	0.89
ANN	0.87	1.00/0.74	0.85	0.80/1.00	0.89	0.90
GBM	0.87	0.97/0.77	0.86	**0.81/0.98**	0.89	0.92
XGBoost	0.87	0.97/0.77	0.86	**0.81/0.98**	0.89	0.92

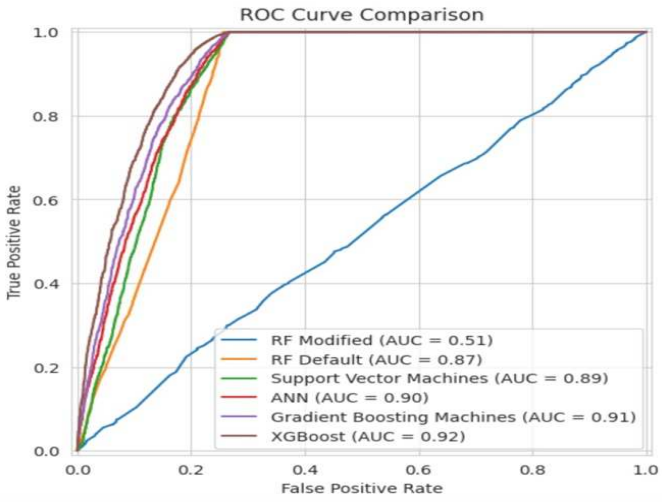


Fig. 3. ROC-AUC curves of classification models

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

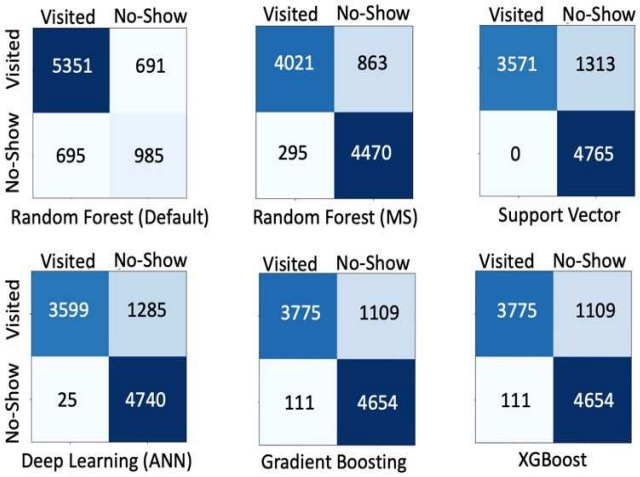


Fig. 4. Confusion Matrices for classification models.

Fig. 3 and 4 contain ROC-AUC and confusion matrix metrics of the developed models respectively. Using the AUC summary displayed in Fig. 4, the least significant model with AUC performance is determined to be 0.57 for RF (modified) while the best AUC performance value of 0.92 is attained for the XGBoost model. The other models consisting of RF Default (0.87), SVM (0.89), ANN (0.9), and GBM (0.91) had relatively similar AUC values that indicated similarities in the model performances for this metric. Fig. 5 is a summary of confusion matrices of the models based on the visits completed and missed appointments. GBM and XGBoost were determined to have the same exact performance in classification while the other models had varying performances as it can be seen based on the numbers attained in this figure.

V. Results and Discussion

In this work, descriptive and predictive modeling is conducted on a data set that contains approximately 38600 medical records. The data is attained from the database of a general clinic located in the USA that contained significant patterns of patients that missed their health care appointments. The rate of missed appointments among the appointments scheduled in 2019 is determined to be 22%. The demographic analysis showed no-show rates varying across different racial and ethnic groups, with Black or African American patients having the highest no-show rate at 23.7%, followed by White (20.8%) and Other Race (20.1%). Puerto Rican and Non-Hispanic or Latino exhibit a higher no-show rate compared to non-Hispanic individuals. Gender-wise, there was no significant difference in the likelihood of missing appointments between males and females.

Appointment types also played a role in adherence. Follow-up appointments, which constituted the majority (40%), had a relatively lower no-show rate, whereas new patient visits, annual exams and physical therapy appointments had higher rates of missed appointments. Socioeconomic factors further influenced attendance. Patients who were homeless or resided in public housing had higher no-show rates, indicating a possible barrier to healthcare access. Language also appeared to be a

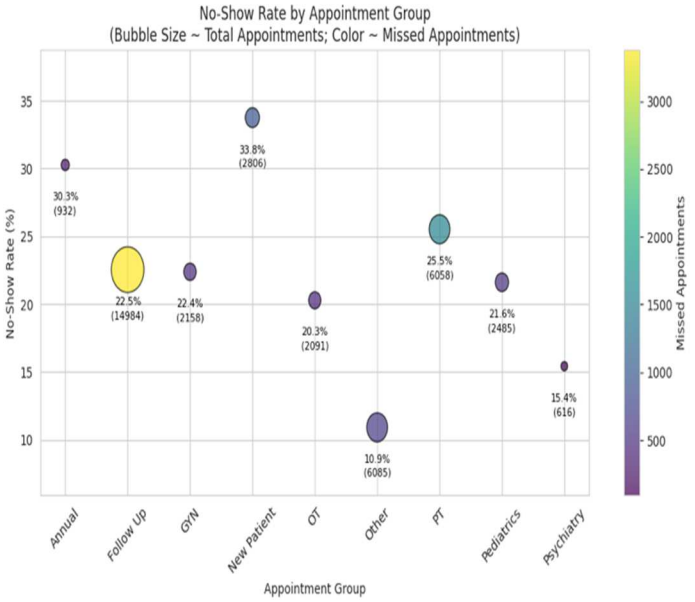


Fig. 5. Appointment Type versus No Show Appointments

contributing factor, with non-English speakers, particularly those who spoke Spanish, exhibiting slightly higher no-show rates. Moreover, patients who did not consent to phone or text reminder were determined to be more likely to miss their appointments.

Fig. 5 is a bubble chart summary of the no-show rates based on the appointment groupings determined by using missed appointments. The color coding pallete on the right side of the figure displays the intensity of the missed appointments within these groups while the bubble size demonstrates the intensity of the missed appointments. As it is easy to see, the missed follow-up appointments has the most significant impact on the missed appointments. Other category which constitute 15% of total treatments e.g. lab work, Injection, Vision Therapy, BP check, allergy and sick visits display lowest no-show rate of 10%.

Our predictive modeling approach, particularly by utilizing ANN, GBM, and XGBoost, demonstrated the feasibility of forecasting patient no-shows with reasonable accuracy. While predictive insights can aid in proactive scheduling interventions, this study primarily focused on descriptive analysis to understand the underlying trends affecting appointment adherence.

There are several important aspects that decision makers need to keep in mind that relate to the analysis of the data:

- The weekdays are determined to experience the highest rates of no-shows, that may be due to the significant number of appointments scheduled for these days. Due to the "domino effect" of patient treatment timing, patients may not necessarily wait for their appointments that may show as "no-shows", but the patients might had arrived at the location and preferred to not wait due to the wait-time in the clinic.
- Automated systems of appointment cancellation and reminders can reduce appointment costs. Such systems

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

may work to reschedule appointments and shuffle appointments based on availability of the health care unit and the patient.

- Additional data collection and analysis on specific health care units of cancelled appointments can provide insights about the reasons for cancelling specific health care units when high rates of cancellations occur.
- Identification of high-rate no-show patients that we presented in this work can provide "chronic" issues of no-show patients and such history can be utilized to ensure that the appropriate actions are taken during the appointment of such patients.

VI. Conclusion

In conclusion, patient no-shows remain to be a critical issue, influenced by demographic, socioeconomic, and appointment-related factors in health care. Our findings provide an insight on the missed health care appointments by patients based on certain features selected by using descriptive and predictive modeling, and such information can serve as foundational results for future research aimed at optimizing scheduling strategies and improving patient engagement to reduce missed appointments.

The predictive models developed in this study can support healthcare providers in making more informed scheduling decisions. One concrete implementation is integrating these models into Electronic Health Record (EHR) systems to automatically flag patients at high risk of no-shows. These flags can trigger targeted outreach, such as live reminder calls and flexible rescheduling. Clinics may also adopt dynamic overbooking strategies, especially in appointment types with historically high no-show rates like annual exams or new patient visits, to reduce idle time without overburdening staff.

In terms of workflow, front-desk systems could prioritize confirmation calls or SMS reminders based on model output. For instance, patients with prior no-shows, lacking consent to reminders, or those in vulnerable housing conditions could be routed to staff for pre-visit check-ins. These simple adjustments may reduce waste and increase attendance rates without requiring large-scale technology overhauls.

Reducing no-shows directly translates to better resource utilization and potentially lower overall healthcare costs. As missed appointments contribute to an estimated $150 billion annual loss in the U.S. healthcare system [22], even modest improvements could yield significant economic benefit. On the other hand, over-reliance on predictive models may unintentionally introduce bias or result in denial of appointment slots for patients deemed "high risk." Therefore, these tools must be used to augment—not replace—human judgment and patient engagement strategies.

Future work can focus on designing and piloting specific interventions derived from model insights, such as adaptive scheduling windows or real-time appointment reshuffling. Continuous model retraining with more data and feedback loops will be essential to maintain performance and equity. Additionally, scaling the system to other healthcare settings such as pediatric care, mental health services, or rural clinics may require feature adjustments to reflect population differences. Integrating these tools into widely used scheduling platforms or patient portals would enhance adoption and allow longitudinal monitoring of attendance behavior trends over time.

References

[1] Glowacka KJ, Henry RM, May JH. A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. J Oper Res Soc. 2009;60(8):1056–1068. doi:10.1057/jors.2008.177

[2] Lee, V.J., Earnest, A., Chen, M.I. et al. Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases. BMC Health Serv Res 5, 51 (2005). https://doi.org/10.1186/1472-6963-5-51

[3] Barron WM. Failed appointments. Who misses them, why they are missed, and what can be done. *Prim Care*. 1980;7(4):563–574. [PubMed] [Google Scholar]

[4] Reid MW, May FP, Martinez B, Cohen S, Wang H, Williams DL Jr, Spiegel BM. Preventing Endoscopy Clinic No-Shows: Prospective Validation of a Predictive Overbooking Model. Am J Gastroenterol. 2016 Sep;111(9):1267-73. doi: 10.1038/ajg.2016.269. Epub 2016 Jul 5. PMID: 27377518.

[5] Barghash, M., & Saleet, H. (2018). Enhancing outpatient appointment scheduling system performance when patient no-show percent and lateness rates are high. International journal of health care quality assurance, 31(4), 309–326. https://doi.org/10.1108/IJHCQA-06-2015-0072

[6] Marbouh, D., Khaleel, I., Al Shanqiti, K., Al Tamimi, M., Simsekler, M. C. E., Ellahham, S., Alibazoglu, D., & Alibazoglu, H. (2020). Evaluating the Impact of Patient No-Shows on Service Quality. Risk management and healthcare policy, 13, 509–517. https://doi.org/10.2147/RMHP.S232114

[7] Fiorillo, C. E., Hughes, A. L., I-Chen, C., Westgate, P. M., Gal, T. J., Bush, M. L., & Comer, B. T. (2018). Factors associated with patient no-show rates in an academic otolaryngology practice. The Laryngoscope, 128(3), 626–631. https://doi.org/10.1002/lary.26816

[8] Shaw, T., Metras, J., Ting, Z. A. L., Courtney, E., Li, S. T., & Ngeow, J. (2018). Impact of Appointment Waiting Time on Attendance Rates at a Clinical Cancer Genetics Service. Journal of genetic counseling, 27(6), 1473–1481. https://doi.org/10.1007/s10897-018-0259-z

[9] Sowter, N., King, L., Calderbank, A., & Eccles, F. J. R. (2022). Factors predicting first appointment attendance at a traumatic brain injury clinical neuropsychology outpatient clinic: a logistic regression analysis. Disability and rehabilitation, 44(22), 6861–6866. https://doi.org/10.1080/09638288.2021.1970254

[10] Rosenbaum, J. I., Mieloszyk, R. J., Hall, C. S., Hippe, D. S., Gunn, M. L., & Bhargava, P. (2018). Understanding Why Patients No-Show: Observations of 2.9 Million Outpatient Imaging Visits Over 16 Years. Journal of the American College of Radiology : JACR, 15(7), 944–950. https://doi.org/10.1016/j.jacr.2018.03.053

[11] Samuels, R. C., Ward, V. L., Melvin, P., Macht-Greenberg, M., Wenren, L. M., Yi, J., Massey, G., & Cox, J. E. (2015). Missed Appointments: Factors Contributing to High No-Show Rates in an Urban Pediatrics Primary Care Clinic. Clinical pediatrics, 54(10), 976–982. https://doi.org/10.1177/0009922815570613

[12] Mohammadi, I., Wu, H., Turkcan, A., Toscos, T., & Doebbeling, B. N. (2018). Data Analytics and Modeling for Appointment No-show in Community Health Centers. Journal of primary care & community health, 9,2150132718811692. https://doi.org/10.1177/2150132718811692

[13] Harvey, H. B., Liu, C., Ai, J., Jaworsky, C., Guerrier, C. E., Flores, E., & Pianykh, O. (2017). Predicting No-Shows in Radiology Using Regression Modeling of Data Available in the Electronic Medical Record. Journal of the American College of Radiology : JACR, 14(10), 1303–1309. https://doi.org/10.1016/j.jacr.2017.05.007

[14] Drabkin, M. J., Lobel, S., Kanth, N., Martynov, A., Hunt, H. W., 3rd, Guerrero, D., Fogel, J., Grechanik, A., Mancuso, C. D., & Lev, S. (2019). Telephone reminders reduce no-shows: A quality initiative at a breast imaging center. Clinical imaging, 54, 108–111. https://doi.org/10.1016/j.clinimag.2018.12.007

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.

[15] Marbouh, D., Khaleel, I., Al Shanqiti, K., Al Tamimi, M., Simsekler, M. C. E., Ellahham, S., Alibazoglu, D., & Alibazoglu, H. (2020). Evaluating the Impact of Patient No-Shows on Service Quality. Risk management and healthcare policy, 13, 509–517. https://doi.org/10.2147/RMHP.S232114

[16] Yuhan, B. T., Yasuda, M. A., Joshi, R., Charous, S., & Hurtuk, A. (2024). No-Show Rates in an Academic Otolaryngology Practice Before and During the COVID-19 Pandemic. Cureus, 16(2), e54015. https://doi.org/10.7759/cureus.54015

[17] Understanding Why Patients No-Show: Observations of 2.9 Million Outpatient Imaging Visits Over 16 Years Rosenbaum, Joshua I. et al. Journal of the American College of Radiology, Volume 15, Issue 7, 944 – 950

[18] Chen A. M. (2023). Socioeconomic and demographic factors predictive of missed appointments in outpatient radiation oncology: an evaluation of access. Frontiers in health services, 3, 1288329. https://doi.org/10.3389/frhs.2023.1288329

[19] Philpott-Morgan S, Thakrar DB, Symons J, Ray D, Ashrafian H, Darzi A, et al. Characterising the nationwide burden and predictors of unkept outpatient appointments in the national health service in England: a cohort study using a machine learning approach 2021. PLoS Med. (2021) 18:e1003783. 10.1371/journal.pmed.1003783 [DOI] [PMC free article] [PubMed] [Google Scholar]

[20] Kheirkhah, P., Feng, Q., Travis, L.M. et al. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 16, 13 (2015) https://doi.org/10.1186/s12913-015-1243-z

[21] Marbouh D, Khaleel I, Al Shanqiti K, Al Tamimi M, Simsekler MCE, Ellahham S, et al. Evaluating the impact of patient no-shows on service quality. Risk Manag Healthc Policy. (2020) 13:509–17. 10.2147/RMHP.S232114 [DOI] [PMC free article] [PubMed] [Google Scholar]

[22] Gier J. Missed appointments cost the U.S. Healthcare system $150B each year. Nashville, TN, United States: Healthcare Innovation; (2017). Available at: https://www.hcinnovationgroup.com/clinical-it/article/13008175/missed-appointments-cost-the-us-healthcare-system-150b-each-year (Accessed January 1, 2023). [Google Scholar]

[23] Berg BP, Murr M, Chermak D, et al. Estimating the Cost of No-Shows and Evaluating the Effects of Mitigation Strategies. Medical Decision Making. 2013;33(8):976-985. doi:10.1177/0272989X13478194

[24] Tarabichi Y, Higginbotham J, Riley N, Kaelber DC, Watts B. Reducing Disparities in No Show Rates Using Predictive Model-Driven Live Appointment Reminders for At-Risk Patients: a Randomized Controlled Quality Improvement Initiative. J Gen Intern Med. 2023 Oct;38(13):2921-2927. doi: 10.1007/s11606-023-08209-0. Epub 2023 May 1. PMID: 37126125; PMCID: PMC101506

979-8-3315-2472-2/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: Istanbul Bilgi University. Downloaded on July 31,2026 at 23:20:33 UTC from IEEE Xplore. Restrictions apply.