

## Research

# Machine Learning Prediction of Diabetic Patients Follow-Up Patterns at Family Medicine Clinic of Tertiary Health Care Provider, Riyadh, Saudi Arabia

### Aleid M

Department of Biostatistics, Epidemiology and Scientific Computing, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

### Alfantoukh L

Department of Biostatistics, Epidemiology and Scientific Computing, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

### Alhuqail N

Applications Engineering Management Department, Ministry of Health, Riyadh, Saudi Arabia

### Abomelha F

Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

### Rayes Z

Department of Family Medicine and Polyclinic, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

## ABSTRACT

**Background:** Diabetes prevalence is predicted to rise dramatically over the next 20 years, and associated spending is expected to increase threefold. We hypothesized that adding appointment data and International Classification of Diseases (ICD) codes to patient demographic data would improve predictions of follow-up appointment attendance utilizing machine learning models. Our results showed that the random forest classifier was the most accurate and sensitive, reaching 73% and 77%, respectively.

**Methods:** This study was based on retrospectively extracted patient's records of King Faisal Specialist Hospital & Research Centre patients diagnosed with diabetes mellitus type I or II who had a follow-up appointment at the Family Medicine Clinic between January 1, 2014, and December 31, 2018. We built several machine learning models, including logistic regression, decision tree, random forest, k-nearest neighbors (KNN), and support vector machine (SVM) models. We also implemented a deep learning algorithm, Deep Neural

Network (DNN).

**Results:** A total of 2,403 patients participated in the study; 3 were excluded because they had only one appointment. Of the 2,400 remaining, around 50% were female, 32% were hospital employees, and 82% were married. Non-Saudis represented around 25% of participants. A total of 19,218 appointments were analyzed, 44.33% of which were classified as "no-show." Prediction accuracy increased by an average of 7% and 10% when we added appointment data and ICD codes, respectively, to demographic data.

**Conclusions:** Our results indicate that knowing appointment-related data and ICD codes corresponding to a patient along with their demographic data is useful in predicting follow-up status.

**Keywords:** Machine learning; Pattern; Deep learning; No-show; Predictions.

## Introduction

Diabetes prevalence is predicted to rise dramatically over the next 20 years, and associated spending is expected to increase threefold [1]. Cost-effective models are needed to respond to the growing and diverse populations affected by diabetes worldwide. In the United States, chronic disease now accounts for 75% of the country's \$2 trillion annual medical costs, and adherence to prescription regimens among those with chronic conditions is estimated to range from 20% to 50% [2]. Patients' failures to keep scheduled appointments ("no-shows") lead to reduced patient-physician continuity, wasted physician time, decreased efficiency, and higher use of resources, which results

in reduced health care quality [3-4]. It has been suggested that lower follow-up rates are also associated with poorer clinical outcomes in diabetic or hypertensive patients [5-7].

In this study, the presence of digital data motivated the researchers to find electronic ways to detect patients' behaviors, such as follow-up appointment show up, utilizing data science and machine learning (ML). Both fields are effective in decision-making to improve healthcare [8]. ML, one of many artificial intelligence applications, involves designing algorithms to predict outcomes in specific contexts [9]. ML algorithms are often categorized as supervised or unsupervised. The two types of algorithms are different in terms of their output labels or

classes. While supervised algorithms use labels to predict future outcomes, unsupervised algorithms do not. A subset ML field called deep learning involves less data preprocessing by humans and offers higher accuracy when working with large amounts of data [10]. To design and implement an algorithm for any ML model, it is important to use historical data related to past events (here, past appointments) to train the model and then use the designed model to predict outcomes for future or new data [11].

In our study, we investigated and analyzed the pattern of diabetic patients' follow-up appointment attendance at family medicine clinics at five years retrospectively using different ML techniques. To the best of our knowledge, no previous study has evaluated the pattern of diabetic patients' follow-up behaviors and extracted the features that might influence such behaviors with ML techniques. Moreover, no study has analyzed the Saudi and Middle Eastern population in this context.

We hypothesized that providing International Classification of Diseases (ICD) information in addition to appointment and demographic data would improve the prediction of patient follow-up behavior. In this study, we assessed whether adding appointment information to demographic data improved the prediction of follow-up behavior and whether adding ICD information to appointment and demographic data would also improve the prediction of a patient's attending their next follow-up appointment, thereby improving the stability of care.

## Methods

### Study design and participants :

This study was based on retrospectively extracted patient's records of King Faisal Specialist Hospital & Research Centre (KFSH&RC) patients diagnosed with diabetes mellitus type I or II who had a follow-up appointment at the KFSH&RC Family Medicine Clinic between January 1, 2014 and December 31, 2018. The study was conducted over a period of one year at KFSH&RC, Riyadh, Saudi Arabia. The Family Medicine Clinic is staffed by 26 attending family physicians, who serve the area as well as hospital employees and their families as a tertiary health care provider. The number of annual patient visits to the clinic is approximately 93,000.

The extracted data included patient demographics: age, gender, nationality, marital status, and medical history of diabetes (type I

or II). We collected subsequent follow-up appointment date and status data through the electronic family medicine scheduling system for patients who met the inclusion criteria for the study. Arrival status was categorized as either "show" or "no-show."

### Inclusion and exclusion criteria:

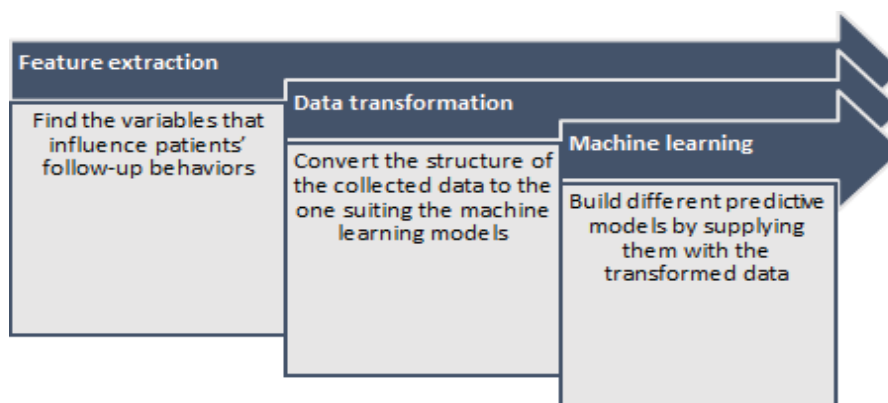
All patients aged 18 years and older who were diagnosed with diabetes type I or II and were scheduled for a follow-up appointment between January 1, 2014, and December 31, 2018, were included in the study. The inclusion criteria were chosen because our model was assumed to most strongly affect these demographics and because follow-up data could be most accurately tracked for electronic patient data.

### Experiment:

Data was an important component of this research study, as data is the main input for our proposed ML models. Therefore, before building the models, it was vital to specify the variables required for prediction, a process known as feature extraction. After obtaining the raw data corresponding to each selected feature, we processed the data by transforming them into a structure that fit the proposed ML model. Subsequently, the process of building the ML models began (**Figure 1**).

For this study, we built different ML models, including logistic regression, decision tree, random forest,  $k$ -nearest neighbors (KNN), and support vector machine (SVM), in light of the possibility of outcome variation among these techniques. We trained and evaluated the proposed models using  $k$ -fold cross-validation with  $k=10$ . Moreover, we built Deep Neural Network (DNN), a deep learning model using a feed forward neural network, which consists of a set of processing elements called "neurons". In this network, the information moves in only one direction, forward, from the input layer, through the hidden layer and to the output layer [12]. In this model, patient data flows from one layer to another without looping back. Initially, DNN produces a map of virtual neurons and appoints random numerical weights to connect inputs with the weights to be multiplied. DNN then returns an output ranging from 0 to 1. After loading the needed libraries and modules, arrays are produced for the features and the response variable, and the dataset is split into 90% training and 10% test data before modeling.

We used Microsoft Access for data transformation and RStudio (a tool based on the R programming language) to build our



**Figure 1:** The required phases for building the machine learning model.

ML models and perform data analysis and data visualization. We built the deep learning model using Keras, an open-source neural-network library written in Python.

### Dataset and Measures

Data for all patients seen or supervised by the participating attending physicians during the study period were retrospectively collected. Patient demographics, scheduled appointment provider (facility and attending department), and follow-up appointment data were extracted through KFSH&RC's electronic health record system. We categorized these data into demographic, appointment, and clinical data (**Table 1**).

We collected follow-up appointment dates and status through the hospital's electronic outpatient scheduling system, ICIS (Integrated Clinical Information Systems), for patients who met the inclusion criteria. The number of follow-ups was determined for each case based on the appointment data. Collected appointment data included scheduled appointment date, appointment provider, and arrival status. Arrival status was described as either "show" or "no-show" (i.e., whether the patient failed to appear for their appointment without canceling).

### Statistical analysis:

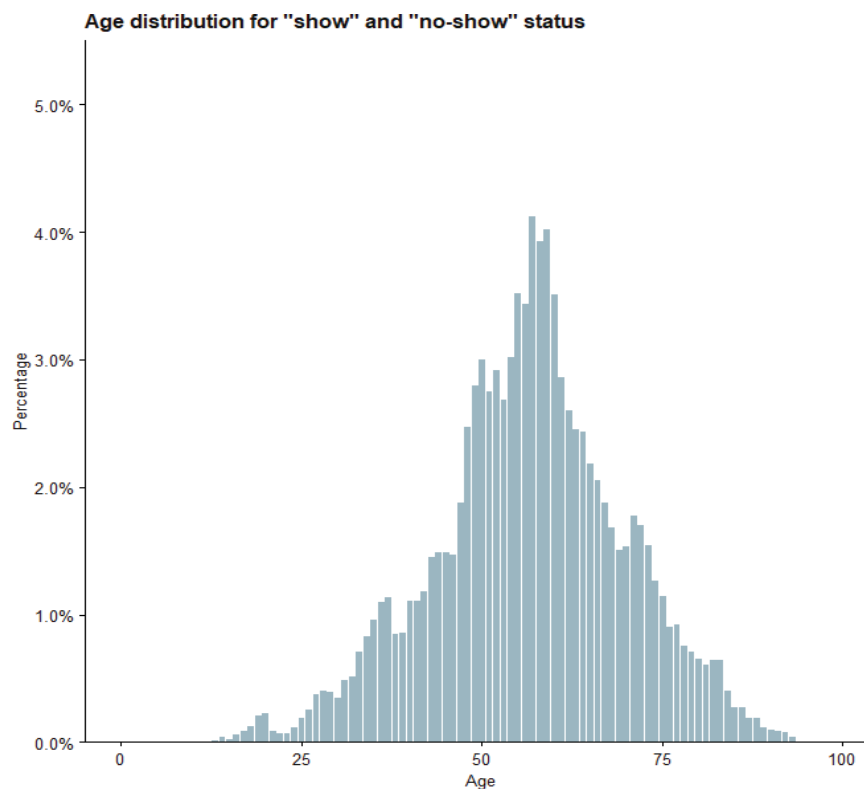
Chi-squared tests were performed to obtain  $p$  values and the standardized difference between the two groups (show and no-show). Numerical variables were expressed as means with standard deviations according to the normality of the variable distribution. Time series analysis was performed to examine the changes in no-show rates over the years. The proportions of no-show cases in relation to age group and appointment time were calculated. Multivariate analysis was conducted to examine whether there was an association between no-show status and appointment time. We then concluded with predictive analytics using various methods, including logistic regression, SVM, KNN, decision tree, random forest, and DNN.

### Results

During the study period, we studied the patient follow-up pattern, conducted exploratory data analysis, and built ML and deep learning models to predict whether a patient would appear for their follow-up appointment. Exploratory analysis was essential to understanding the shape of our data and identifying potential patterns.

**Table 1: Dataset categories and corresponding variables.**

Demographic	Appointment	Clinical
<ul style="list-style-type: none"> <li>· Age</li> <li>· Gender</li> <li>· Marital status</li> <li>· Nationality</li> <li>· Admission Eligibility</li> <li>· Registration type</li> <li>· City</li> </ul>	<ul style="list-style-type: none"> <li>· Facility</li> <li>· Attending department</li> <li>· Show status</li> </ul>	<ul style="list-style-type: none"> <li>· ICD code</li> </ul>



**Figure 2:** The distribution of patients ages, which is normally distributed.

### Exploratory Analysis:

A total of 2,403 patients were eligible for inclusion in the study; of these, 3 were excluded because they had only one appointment. Of the 2,400 remaining patients, around 50% were female, 32% were hospital employees, and 82% were married. Non-Saudis represented around 25% of the study population. The average age was 56.5 years (**Figure 2**). Our data included a total of 19,218 encounters (appointments), 44.33% of which were classified as “no-show.” **Table 2** provides demographic information and variable statistics for each show status.

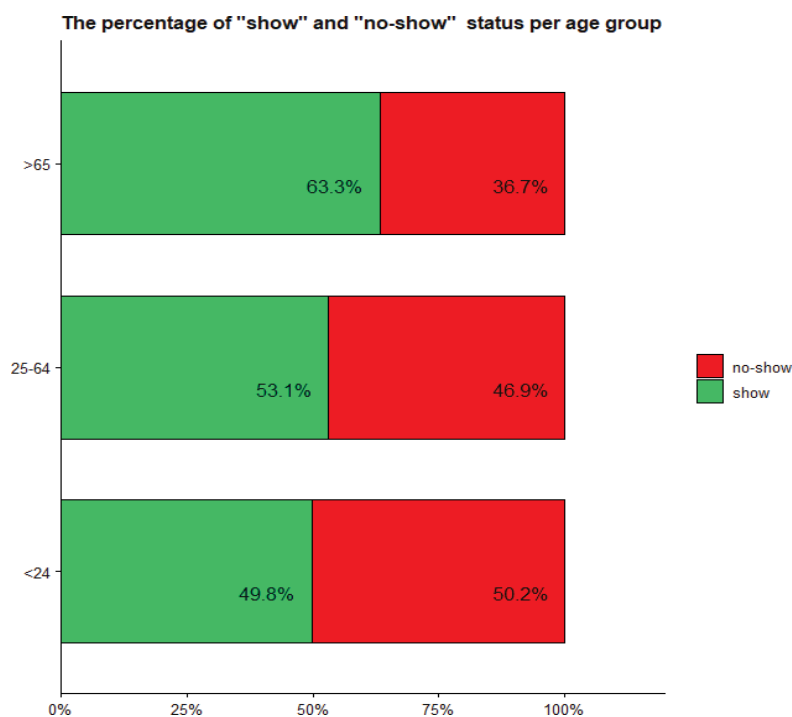
We categorized patient ages into three groups: “24 or younger,” “between 25 and 64,” and “65 and older.” (**Figure 3**) shows

that the percentage of “no-show” cases for patients in the “24 and younger” group was higher than among other groups. The patients in the “65 and older” category were the most likely to show up for their appointments compared to other groups. We also explored the changes in show status over the years and found that the percentage of “no-show” cases decreased over the years while “show” cases increased (**Figure 4**).

Each appointment has relevant associated data, such as date and time. We found a weak positive correlation equals to 0.03 between the time of the appointment and show status. As **Figure 5** shows, however, the percentage of “show” cases is similar to the percentage of “no-show” cases. Therefore, time may not be an important feature for our predictive models.

**Table 2: Demographic comparison results.**

Variable	Value	No-show	Show	<i>p</i>	SMD
Age <i>n</i> (%)	> 65 years	1,823 (21.40)	3,142 (29.37)	< 0.01	0.185
	25–64 years	6,591 (77.36)	7,451 (69.65)		
	18–24 years	106 (1.24)	105 (0.98)		
Gender <i>n</i> (%)	Female	4,679 (54.92)	5,293 (49.48)	< 0.01	0.109
	Male	3,841 (45.08)	5,405 (50.52)		
Nationality <i>n</i> (%)	Non-Saudi	2,280 (26.76)	2,676 (25.01)	0.006	0.040
	Saudi	6,240 (73.24)	8,022 (74.99)		
City <i>n</i> (%)	Riyadh	4,548 (53.38)	6,617 (61.85)	< 0.01	0.172
	Other	3,972 (46.62)	4,081 (38.15)		
Marital Status <i>n</i> (%)	Divorced	203 (2.38)	182 (1.70)	< 0.01	0.087
	Married	6,824 (80.09)	8,824 (82.48)		
	Single	1,185 (13.91)	1,284 (12.00)		
	Unknown	36 (0.42)	23 (0.21)		
	Widowed	272 (3.19)	385 (3.60)		



**Figure 3:** Percentage of show status by age group.

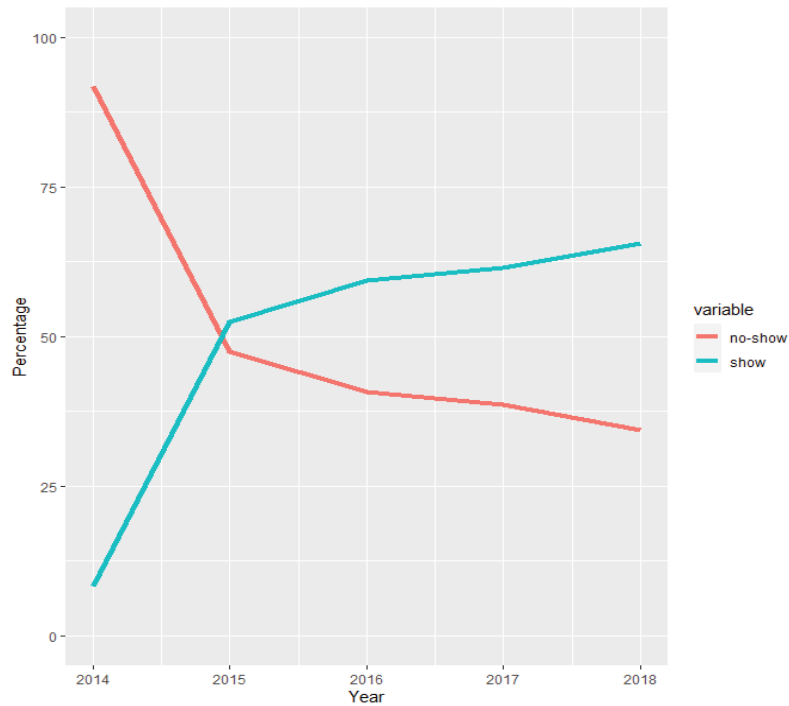


Figure 4: Changes in show status over time (by year).

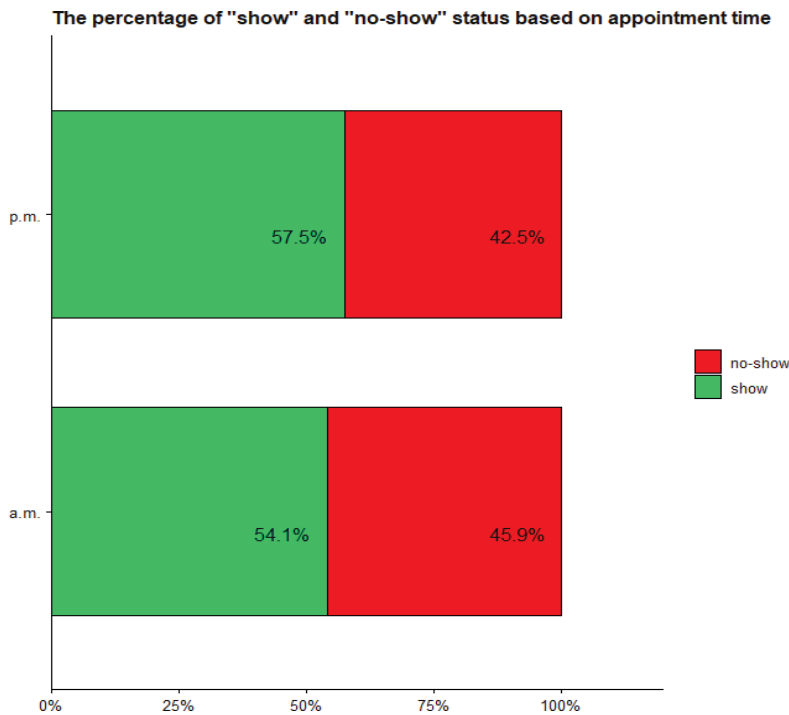


Figure 5: Percentage of show status based on appointment time (am or pm).

**Predictions:**

To run the ML models, we needed to perform feature selection. Therefore, we defined three groups of datasets based on feature type:

- Demographic data
- Demographic and appointment data
- Demographic, appointment, and ICD code data

Let  $ML$  be a function that represents the ML model. To produce  $outcome_i$  (show or no-show), we used the following formula:

$$outcome_i = ML(group_i)$$

$ML$  can be logistic regression, SVM, KNN, decision tree, random forest, or DNN. Represents the group type, and represents all features belonging to that group. Running the ML function produces  $outcome_i$ , which is a list of predicted values that represent follow-up status. From that  $outcome_i$  list, we can compute accuracy, sensitivity, and specificity as follows:

where  $TP$  represents the number of “show” cases classified correctly,  $TN$  represents the number of “no- show” cases classified correctly,  $FP$  represents the number of “no-show”

**Table 3: Prediction accuracy results for different models.**

Features	Logistic Regression	SVM	KNN	Decision Tree	Random Forest	DNN
$group_A$	59%	59%	56%	60%	62%	58%
$group_B$	61%	66%	65%	68%	69%	67%
$group_C$	61%	69%	71%	72%	<b>73%</b>	70%

**Abbreviations:** SVM: Support Vector Machine; KNN: K-nearest neighbors; DNN: Deep Neural Network

**Table 4: Prediction sensitivity results for different models.**

Features	Logistic Regression	SVM	KNN	Decision Tree	Random Forest	DNN
$group_A$	56%	58%	51%	57%	61%	62%
$group_B$	61%	67%	61%	68%	70%	62%
$group_C$	61%	70%	71%	73%	<b>77%</b>	71%

**Abbreviations:** SVM: Support Vector Machine; KNN: K-nearest neighbors; DNN: Deep Neural Network

**Table 5: Prediction specificity results for different models.**

Features	Logistic Regression	SVM	KNN	Decision Tree	Random Forest	DNN
$group_A$	60%	60%	61%	62%	63%	59%
$group_B$	61%	65%	68%	68%	68%	67%
$group_C$	61%	68%	71%	<b>72%</b>	71%	69%

**Abbreviations:** SVM: Support Vector Machine; KNN: K-nearest neighbors; DNN: Deep Neural Network

cases classified incorrectly as “show,” and  $FN$  represents the number of “show” cases classified incorrectly as “no-show.” Accuracy indicates the extent to which all cases were classified correctly, sensitivity describes the extent to which “show” cases were classified correctly, and specificity denotes the extent to which “no-show” cases were classified correctly (**Tables 3-5**).

We computed the results from the ML and deep learning models. Our findings indicated that  $group_C$  had a higher performance than  $group_A$ . Letting *average difference* represent performance variation, we find the differences between the two values for each group and take the average as the following equation:

$$average\ difference = \frac{\sum_{k=1}^n acc_{C_k} - acc_{A_k}}{N}$$

where  $acc_{A_k}$  and  $acc_{C_k}$  are the accuracy values of groups  $A$  and  $C$  for model  $k$  and  $N$  is the number of models. The *average difference* for accuracy is 10%. We applied the same formula to compute the average difference between groups  $A$  and  $B$ , with a result of 7%.

## Discussion

Several methods have been investigated to decrease no-show rates, including mail, telephone, and short message service (SMS) reminder systems; advance or same-day scheduling systems, which promote patient-driven scheduling [13-15]. Of these, SMS reminder systems have increasingly been used due to their ease of applicability and financial feasibility. However, more effort is still needed to minimize no-shows given their high therapeutic, educational, and financial burdens [16].

To our knowledge, this is the first study to evaluate the association between strategies to improve compliance and follow-up appointment attendance in primary care settings [17]. We searched the relevant literature from the inception of the database through 1999 using terms such as ML, follow up,

prediction, appointment and schedules, appointment compliance or adherence, and no-show. The searched databases included PubMed, Ovid Medline, and Scopus as well as open databases such as Google Scholar and Microsoft Academic Search.

Various challenges exist in building a ML model, such as feature selection and the choice of model [18]. There are several studies in the area of predicting patients' attendance of follow-up appointments. Harrell et al., used the random forest classifier and neural networks for prediction, while Elvira et al., used a gradient boosting algorithm [19]. To predict missing appointments among patients with diabetes, Choi et al., used logistic regression, linear discriminant analysis, quadratic discriminant analysis, and KNN algorithms [20]. Kurasawa et al., also used logistic regression in their study of patients with diabetes [21]. The features selected in these studies were related to demographic, clinical, and appointment data. Almuhaideb et al., 2019, used artificial intelligence to build a model that predicted no-shows for individual outpatient clinic appointments and produced an accuracy of 76.44% and 77.13% [22].

In our study, we used artificial intelligence to build a model that predicted no-shows for individual family medicine clinic appointments. Our algorithms produced accuracy and sensitivity with (among others) the random forest model, which achieved 73% and 77%, respectively. The decision tree model had the best performance in terms of specificity at 72%. This study contributes to the body of literature in this area by showing that demographic data alone is not sufficient to improve prediction performance in terms of accuracy. Adding appointment data is useful in increasing accuracy, as it increased the value by 7%. The same is true of ICD code data, which increased accuracy by 10%. The deep learning model had a similar performance to other traditional ML algorithms and did not have the best performance, as it requires large amounts of data to improve performance [23].



This study has several noticeable limitations. First, the study was not randomized. This could have affected the patterns of the data and may have reduced the chance of observing positive effects. Second, we could not collect patient demographics such as ethnicity, smoking status, diet, employment status, or socioeconomic status, which may influence the pattern of the appointments. Third, we could not assess whether patients attended clinics outside our family medicine clinic. These occurrences may invalidate the effectiveness of the pattern of the study. Fourth, the study findings may not be generalizable to other practice settings with different population characteristics. Despite these limitations, however, this study provided new evidence that demographic data is not sufficient to improve prediction performance in terms of accuracy. Considering the therapeutic, educational, and financial burdens of no-shows in family medicine clinics, further studies are needed to improve prediction performance.

### Conclusion

In this study, we examined the pattern of diabetic patients' follow-up behavior at KFSH&RC. We utilized ML and deep learning to predict whether a patient would come to a follow-up appointment using demographic, appointment, and ICD data. Our results indicate that knowing appointment-related data and ICD codes corresponding to a patient along with demographic data is useful in predicting follow-up status. In fact, having this information increased prediction accuracy by 7% and 10% on average, which is superior to using demographic data exclusively. We applied different ML models and found that the random forest algorithm had the best performance, reaching 73% and 77% accuracy and sensitivity, respectively. In future work, we plan to include more variables, such as patient data and medications, to improve prediction performance.

### Declaration of Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This study was approved by the KFSH&RC Research Ethics Committee (RAC#2191170).

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### References

- Huang ES, Basu A, O'Grady M, Capretta JC. Projecting the future diabetes population size and related costs for the US. *Diabetes Care* 2009; 32: 2225-2229.
- Zhang P, Zhang X, Brown J, Vistisen D, Sicree R, et al. Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes Res Clin* 2010; 87: 293-301.
- Johnson BJ, Mold JW, Pontious JM. Reduction and management of no-shows by family medicine residency practice exemplars. *Ann Fam Med* 2007; 5: 534-539.
- Hixon AL, Chapman RW, Nuovo J. Failure to keep clinic appointments: implications for residency education and productivity. *Fam Med* 1999; 31: 627-630.
- Glynn LG, Murphy AW, Smith SM, Schroeder K, Fahey T. Interventions used to improve control of blood pressure in patients with hypertension. *Cochrane Database Syst Rev* 2010; 3: 1-7.
- Nuti LA, Lawley M, Turkcan A, Tian Z, Zhang L, et al. No-shows to primary care appointments: subsequent acute care utilization among diabetic patients. *BMC Health Serv Res* 2012; 12: 304.
- Schectman J, Schorling J, Voss J. Appointment adherence and disparities in outcomes among patients with diabetes. *J Gen Intern Med* 2008; 23: 1685-1687.
- <https://link.springer.com/content/pdf/10.1007/978-3-642-40017-9.pdf>
- Das S, Dey A, Pal A, Roy N. Applications of artificial intelligence in machine learning: Review and prospect. *Int J Comput Appl* 2015; 115: 31-41.
- <https://link.springer.com/book/10.1007/978-1-4842-2845-6#about>
- Kotsiantis SB. Supervised machine learning: A review of classification techniques. *Informatica* 2007; 31: 249-268.
- Wu H, Zhou Y, Luo Q, Basset MA. Training Feedforward Neural Networks Using Symbiotic Organisms Search Algorithm. *Comput Intel Neurosc* 2016; 16: 1-15.
- Guy R, Hocking J, Wand H, Stott S, Ali H, et al. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. *Health Serv Res* 2012; 4: 614-632.
- Macharia WM, Leon G, Rowe BH, Stephenson BJ, Haynes RB. An overview of interventions to improve compliance with appointment keeping for medical services. *JAMA* 1992; 267: 1813-1817.
- Parikh A, Gupta K, Wilson AC, Fields K, Cosgrove NM, et al. The effectiveness of outpatient appointment reminder systems in reducing no-show rates. *Am J Med* 2010; 123: 542-548.
- Kaplan-Lewis E, Percac-Lima S. No-show to primary care appointments: Why patients do not come. *J Prim Care Community Health* 2013; 4: 251-255.
- Sorita A, Funakoshi T, Kashan G, Young ER, Park J. Impact of prescription patterns on compliance with follow-up visits at an urban teaching primary care continuity clinic. *J Prim Care Comm Heal* 2014; 5: 188-193.
- <https://ieeexplore.ieee.org/document/8031199>
- Elvira C, Ochoa A, González JC, Mochón F. Machine-learning-based no show prediction in outpatient visits. *Int J Interact Multi Art Intel* 2018; 4: 29-34.
- Choi BG, Rha SW, Kim SW, Kang JH, Park JY, et al. Machine learning for the prediction of new-onset diabetes mellitus during 5-year follow-up in non-diabetic patients with cardiovascular risks. *Yonsei Med J* 2019; 60: 191-199.
- Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, et al. Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *J Diabetes Sci Technol*, 2019; 10: 730-736.

22. Almuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. *Ann Saudi Med* 2019; 39: 373-381.
23. Reiss A, Indlekofer I, Schmidt P, Van Laerhoven K. Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors* 2019; 19: 3079.

**ADDRESS FOR CORRESPONDENCE:**

Aleid M, Department of Biostatistics, Epidemiology and Scientific Computing, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia, E-mail: mahaeid@kfshrc.edu.sa, Tel: +966505216367

Submitted: June 25, 2020; Accepted: July 16, 2020; Published: July 24, 2020