

The Prediction of Outpatient No-Show Visits by using Deep Neural Network from Large Data

Riyad Alshammari¹, Tahani Daghistani², Abdulwahhab Alshammari³
Health Informatics Department, College of Public Health and Health Informatics
King Saud Bin Abdulaziz University of Health Sciences
King Abdullah International Medical Research Center (KAIMRC)
Ministry of National Guard Health Affairs, Riyadh, KSA

Abstract—Patients' no-show is one of the leading causes of increasing financial burden for healthcare organizations and is an indicator of healthcare systems' quality and performance. Patients' no-show affects healthcare delivery, workflow, and resource planning. The study aims to develop a prediction model predict no-show visits using a machine learning approach. A large volume of data was extracted from electronic health records of patient visits in outpatient clinics under the umbrella of large medical cities in Saudi Arabia. The data consists of more than 33 million visits, with an 85% no-show rate. A total of 29 features were utilized based on demographic, clinical, and appointment characteristics. Nine features were an original data element, while data elements derived 20 features. This study used and compared three machine learning algorithms; Deep Neural Network (DNN), AdaBoost, and Naive Bayes (NB). Results revealed that the DNN performed better in comparison to NB and AdaBoost. DNN achieved a weighted average of 98.2% and 94.3% of precision and recall, respectively. This study shows that machine learning has the potential to improve the efficiency and effectiveness of healthcare. The results are considered promising, and the model can be an excellent candidate for implementation.

Keywords—No-show; outpatients; machine learning; prediction model introduction

I. INTRODUCTION

Reducing outpatient appointment no-show is essential to health care organizations to utilize resources, decrease the financial burden, and treat the patients who need care. Hence, accurately predicting outpatient appointment no-show can efficiently use the hospital resources, reduce the waiting list considerably, and improve patient satisfaction significantly. Thus, "improve the quality of healthcare" [1] [2] [3] [4]. The request for outpatient services in Saudi Arabia is increasing sharply, and the appointment booking system is suffering from the high outpatient's appointment no-show [5]. There is currently no useful tool in the electric health care system to identify patients of a high risk of a no-show [5] [6] [7]. The current practice is to allow overbooking and walk-in. Therefore, healthcare systems in Saudi Arabia require a useful tool to predict outpatient no-show accurately.

Existing research on clinic no-show focuses on finding factors and developing models in specific patient groups such as diabetes or particular departments such as radiology. Moreover, rely on quantitative or traditional methods to serve the prediction of no-show risk. Few kinds of research are studying high-dimensional and high-volume

big data incorporating behavioral factors for no-show prediction. Many institutions are investing heavily in technology and digitalizing the work task, especially in the healthcare sector. A massive amount of data is collected over the years by the healthcare sector. Healthcare data is forecast by the International Data Corporation (ICD) to grow to more than 163 zettabytes by 2025 [8]. Hence, Big data solution becomes an essential part of the healthcare sector to extract knowledge and insights from the vast collected digital data records. One exciting area of data science, gaining massive attention from data science and big data analytics communities is Deep learning [9] [10]. Deep learning algorithms succeed in other domains [10] [11] and becoming more popular in the medical field [12] [13]. Deep learning algorithms become a promising type of machine learning algorithm that can model complex data characteristics at a high generalization level. One of the assuring platforms that can effectively handle big data and discover insights using machine learning algorithms is Google TensorFlow [14]. Google Tensorflow is a deep learning platform that can run a complex computation on different data types. It has been created to work efficiently and quickly to find knowledge/insights from a considerable amount of data.

This paper aims to construct a data-driven approach based on a deep machine algorithm to learn from over 30 million records to predict no-show in Outpatient clinics. The goal is to seek a patient's behavioral patterns by considering the patient's history of appointments that predict no-shows' probability. The main research contributions of this paper are:

Demonstrating the ability to predict no-show using only a minimalist feature set;

Exploring the performance of training machine learning algorithm with different machine learning tools;

Exploring the power of using Big Data Machine Learning platforms to build a model from large data size.

The rest of this paper is organized as follows. Section II presents the related work, and Section III provides the methods. The results are shown in Section IV. The analysis and discussion are given in Section V. Conclusions, and future work are drawn in Section VI.

II. RELATED WORK

A. Discovering Factors Related to No-Show

There are existing research mainly focuses on discovering the factors that contribute to predicting the probability of no-show patients. Harvey et al. [15] applied statistics and logistic regression models to assess the effectiveness of using factors available in the Electronic Medical Record (EMR) such as Demographic, clinical, and health services utilization factors in predicting the absence of patients from scheduled radiology examinations. Factors that successfully predict radiology no-shows included days between scheduling and appointments, modality type (mammography, CT, PET, and MRI), and insurance type. The predictive ability was determined using the area under the receiver operator curve was 0.753.

Chua and Chow [16] applied Multiple Logistic Regression (MLR) on routinely collected administrative data to define factors associated with no-shows. Using parameter estimates from MLR, a risk-scoring model was developed to classify patients according to their risk of a no-show. The model's predictive ability was 72%, evaluated using the area under the curve (AUC). A study conducted by Dantas et al. [17] investigated each patient-related factor's influence on appointment no-show behavior in the bariatric surgery clinic. A data set of 13,230 records was used to run Logistic Regression to examine specific factors on no-show rates. As a result, predictive models were developed and perform effectively (Accuracy: 71%) with eight variables. They are later hours appointment, summer months or not, pre/post-surgery appointment, high/low lead time, higher/little no-show history, numbers of previous appointments, home distance (20 to 50 km) from the clinic, or another scheduled medical specialty than a bariatric surgeon. In contrast, gender, age, weekday, and payment were not significant factors to predict patients' no-show.

B. Using Traditional Machine Learning Algorithms to Predict No-Show

Few research studies had focused on predicting no-show of patients using different machine learning techniques, mainly logistic regression. Kurasawa et al. [18] build a logistic regression model to predict missed appointments by diabetes patients. Data were classified into two groups: the first group for clinical condition and the other for previous findings. The best predictor model was achieved in an area under the curve (AUC) = 0.958 using both groups. Precision and recall were, respectively, 0.757 and 0.659. Among all data, the appointment's day was the strongest predictor of missing the appointments (weight = 2.22).

Mohammadi et al. [19] used statistical and machine learning models to predict the next medical appointment show's chance. They applied logistic regression, artificial neural network, and Naive Bayes classifier applied on 73,811 unique appointments to identify critical features. As for finding, predictors were created, in addition to the EHR variable, as significant predictors to consider no-show

appointments. The new predictors included lead-time, prior no-show, tobacco use, cell phone ownership, and the number of days since the last appointment. Naive Bayes models had a relatively high area under the curve among three models; the model achieved 0.86.

On the other hand, Nelson et al. [20] suggested complex, high-dimensional, and non-linear predictive models based on training and evaluating a set of 22,318 sequential scheduled magnetic resonance imaging appointments. They used logistic regression, support vector machines, random forests, and AdaBoost to predict two hospitals' attendance. The results showed that Gradient Boosting models achieved the best performance with an area under the receiver operating characteristic curve of 0.852 and an average precision of 0.511.

Lenzi et al. [21] developed and validated a no-show's predictive model based on empirical data. The models were developed using Naïve and logistic regression and the Akaike Information Criteria to select the highest performance model. Fifty percent of scheduled appointments collected from a public primary care setting were used to train the model, and fifty percent were used to validate the model. Experimental results showed that an AUC of 80.9% (95% CI 80.181.7) was achieved using the two most important predictors: previous attendance and same-day appointments. Lee et al. [22] collected two years of follow up data with 25% of a no-show rate. They deployed three machine learning algorithms, namely RandomForest, Logistic regression, and decision tree. They achieved the best performance using Random Forest with an accuracy of 72.9. However, the collected dataset was for a small group of 400 patients.

C. Using Deep Learning Algorithms

On the other hand, deep learning methods have attracted many researchers and organizations in the health care field. Deep learning methods are useful with Problems, which are difficult to solve with traditional methods. They provide the optimal way to deal with high dimensional and volume data. Furthermore, present a whole picture embedded in large-scale data and disclose unknown structure. It has proven to be a superior prediction of no-show. Thus, effective optimizing of health resource usage. There is a minimal effort in using deep learning in the prediction of the patient's no-show. Only one study using deep learning had been found to predict no-show patients in outpatient clinics. Dashtban and Li [23] represented a novel prediction method for outpatients' non-attendance based on a wide range of health, environmental, and socioeconomic factors. The model was based on deep neural networks, which have integrated data reconstruction and prediction steps from in-hospital data. This integration was aiming to have higher performance than the separated classification model in predicting tasks. Comparing the proposed model with other machine learning classifiers showed that deep learning models outperform other practice methods. The model achieved (AUC (0.71), recall (0.78), accuracy (0.69). Finally, the constructed model was deployed and connected to a reminder system. To the best of our knowledge, all the previous studies were applied in small data sizes. One of this study's objectives is to

use a larger data size to predict the no-show at high accuracy. For a summary of the related work, see Table I.

TABLE I. THE SUMMARY OF RELATED WORK

The Summary of Related Work			
Authors	Algorithms	Best Algorithm	Result
Kurasawa et al. (2016)	Logistic Regression	Logistic Regression	AUC = 0.958 Precision = 0.757 Recall = 0.659
Mohammadi et al. (2018)	logistic regression artificial neural network NaïveBayes	NaïveBayes	AUC = 0.86
Nelson et al. (2019)	logistic regression support vector machines random forests AdaBoost	Gradient Boosting	AUC = 0.852 Precision = 0.511
Lenzi et al. (2019)	Naïve Bayes Logistic Regression	Naïve Bayes	AUC = 80.9
Lee et al. (2017)	RandomForest Logistic regression decision tree	Random Forest	accuracy = 72.9
Dashtban and Li (2019)	deep neural networks	deep neural networks	AUC = 0.71 Recall = 0.78 Accuracy = 0.69

III. EVALUATION METHODOLOGY

This section details the research methodology describing the datasets, features, preprocessing, machine learning algorithms, and evaluation criteria.

A. Dataset Description

Three-year datasets (Jan 2016-July, 2019) were extracted from the Medical Record Number (MRN) system at all the facilities at the central region (Riyadh) at the Ministry of National Guard Health Affairs (MNGHA) to reflect on the outpatient visits. King Abdullah International Medical Research Center (KAIMRC) review boards approved the study, and a waiver of individual consent was authorized. MNGHA has many hospitals and primary health care clinics that provide quality services for MNGHA staff and their eligible dependents. The three regions are Central (Riyadh city), Eastern (Dammam and Al-Hasa cities), and Western (Jeddah and Al-Madaina cities) and many primary health care centers all over Saudi Arabia. Approximately 77000 scheduled appointments are booked monthly on the Riyadh hospital. The central region has an on average of 77k booked appointments every month, Fig. 1.

Most outpatients live in major cities of Saudi Arabia, but many outpatients travel from across the country to Riyadh. A text reminder is sent to all outpatient three days before the appointment. The dataset includes more than 33 million (33,050,363) outpatient appointments. Table II shows a statistic description of the MNGHA

dataset according to age groups, gender, appointment type, and nationalities used in this research.

More than 33 million patients' appointments were booked in outpatients' clinics of MNGHA between January 2016 and July 2019. Of these, label classes included (85%) and (15%) for no-show class and show class, respectively. A huge no-show for the follow-up appointments (98.96%) was noticed. The noticed no-show for follow-up appointments are mainly for two reasons: I) patient need to travel a long distance to come to the hospital, and II) most of the missed appointments were related to diabetic follow up, and there is no overbook for them.

The largest group of the patients' age were between 5 and 69 years old, and the no-show rate was higher for patients over 45 years old. Most of the no-show was among national citizens (85%). The show rate for the first appointment was 76%, and more than half the patient showed up for the first visit. Of the total, (26,625,488) of the appointments were follow-up with (98.96%) no-show rate. There was a high no-show among adult patients in terms of age group, and there were no significant differences in the no-show rate among gender. Fig. 2 shows the machine learning process of the no-show prediction model.

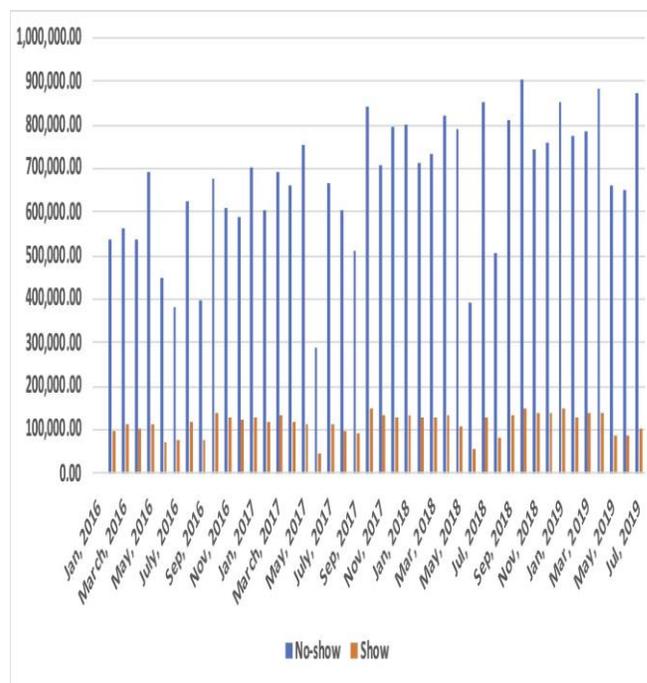


Fig. 1. Distribution of show and No-Show of Patients' Appointments from 2016 to 2019 in the Central Region.

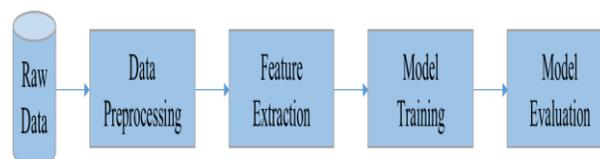


Fig. 2. Machine-Learning Process of the No-Show Prediction Model.

TABLE II. DESCRIPTIVE CHARACTERISTICS OF THE DATASET
(N=33,050,363)

Features	No-Show (N%)	Show (N%)	Total (N)
Age Group			
1-4	250,537 (65%)	134,715 (35%)	385,252
5-9	728,676 (66.9%)	360,698	1,089,374
10-14	1,004,354 (76%)	(33.1%)	1,321,454
15-19	804,462 (72%)	317,100 (30%)	1,114,550
20-24	695,384 (68.6%)	310,088 (28%)	1,014,202
25-29	729,945 (67.1%)	318,818	1,087,043
30-34	920,858 (69.2%)	(31.4%)	1,329,993
35-39	1,114,211 (73%)	357,098	1,525,420
40-44	1,274,528 (78.4%)	(32.9%)	1,626,470
45-49	1,886,169 (85.1%)	409,135	2,216,686
50-54	2,626,922 (89.9%)	(30.8%)	2,923,219
55-59	3,373,995 (92.8%)	411,209 (27%)	3,637,547
60-64	3,512,907 (94.1%)	351,942	3,732,703
65-69	3,005,732 (94.9%)	(21.6%)	3,168,725
70-74	238,1034 (94.9%)	330,517	2,509,008
75-79	1,820,760 (95.3%)	(14.9%)	1,910,993
80-84	912,299 (95%)	296,297	960,419
≥	583,117 (94.7%)	(10.1%)	615,913
		263,552 (7.2%)	
		219,796 (5.9%)	
		162,993 (5.1%)	
		127,974 (5.1%)	
		90,233 (4.7%)	
		48,120 (5%)	
		32,796 (5.3%)	
Gender			
Male	15,989,649 (84.2%)	3,005,006	18,994,655
Female	12,189,970 (86.8%)	(15.8%)	14,051,241
		186,1271	
		(13.2%)	
Appointment type			
New Patient (NP)	936,387 (23.4%)	3,060,598	3,996,985
First visit (FV)	899,145 (37%)	(76.6%)	2,427,890
Follow up (FU)	26,347,841 (98.96%)	1,528,745	26,625,488
		(63%)	
		277,647	
		(1.04%)	
Nationalities			
Saudi	27,891,825 (85%)	4,799,871	32,691,696
Egypt Sudan	63,562(88%)	(15%)	72,200
Philippines Pakistan Syria	43,753 (85%)	8,638 (12%)	51,627
Other Countries	30,696 (65%)	7,874 (15%)	46,967
	20,261 (84%)	16,271 (35%)	24,087
	7,059(67%)	3,826 (16%)	10,592
	107,617 (82%)	3,533 (33%)	131,089
		23,477 (18%)	

B. Data Preprocessing

Processing raw data needs computational powers. The data warehouse provided by the database management team was used to filter unnecessary data that is not needed in the prediction model. Data cleanings were applied to reduce noise. Categorical variables were converted into numbers, such as age and gender. For example, the age calculated based on the year and categorize by grouping every four years together.

Data cleanings methods were applied to reduce noise and remove outliers. Categorical data were transferred to numeric types such as sex and age. Data were normalized to make the data suitable for building machine learning models [24] [25].

C. Feature Extraction

New features were derived from the data. For instance, the lead-time feature was calculated, representing a

difference in days between the appointments' booking and the day of the appointments. Also, calculate the percentage of show and no-show on all the previous visits. The past behavior of patients could be a potential feature for predicting future patient's behavior. From the "Medical Treatment Reservation Type Code" feature, two categorical predictors were added to calculate the count number of walk-in and scheduled appointments. In addition to calculating the count of on foot and emergency appointments. For the "Cancellation Flag" predictor, where the patient cancelled the appointment, new predictors were added to calculate the count of Cancellation Flag Yes and No. In addition to the extracted data from the electronic medical records, the features derived were candidates for the predictive model.

The list of generated features is shown in Table III. A total of (29) categorical or numerical data elements considered: (9) original data elements and (20) derived data elements. The dataset contained three categories: demographic attributes, clinical attributes, and appointment characteristics considered relevant to patient appointment history. The dataset contained imbalanced classes with 83.89% of the records for no-show class and 16.12% of the show class records. Finally, data transformation applied where all nominal values were converted to binary attributes.

D. Machine Learning Algorithms

Three machine learning algorithms were applied in this work: Deep Neural Network, AdaBoost, and Naive Bayes. Multilayer Perceptron (MLP) algorithm is a type of machine learning algorithm developed by researchers to simulate how humans function and learn. Backpropagation was added to MLP as an improvement [26]. The MLP becomes one of the golden standard algorithms in the machine learning domain because it can learn despite the lack of prior information. Deep Neural Network (DNN) Learning algorithm is an extension of the MLP algorithm with many hidden layers [27] [28]. DNN applies various activation functions, such as Rectified Linear Unit (ReLU). DNN used gradient descent optimizers and various activation functions, such as Rectified Linear Unit (ReLU). These optimizers' main task is to find the local minimum solution with the assistant of hyperparameters like the learning rate. To model the deep neural network model, the values of the hyperparameters, such as the number of hidden layers, neurons, and type of activation function, need to be chosen wisely. The hyperparameter's values have a huge effect on the training time and the model's learning capability. Having large values is going to increase the training time, while having small values will minimize the learning abilities of the model. The model was trained by choosing smaller hyperparameters and increasing the hyperparameters' values/dimensionalities depending on the model's accuracy. The final DNN model builds by using Tensorflow [14] implementing DNN. The DNN model consists of four layers: input layers (28 nodes), two hidden layers (14 nodes and 7 nodes), and one output layer with ReLU as activation function, Fig. 3.

TABLE III. DATA CATEGORIES AND PREPROCESSING FOR PATIENT HISTORICAL DATA

Data Features and Description	
Feature	Description
1) Gender	Male, Female and Unknown
2) Age Category	18 age categories
3) Nationality	The nationality of the patient
4) Address code	The primary health care clinic location (used for the patient address)
5) Medical Treatment Service Code	The code of the medical service
6) Appointment type code New Patient (NP) First visit (FV) Follow up (FU)	Type of Appointment New Patient First visit Follow up visit
7) Patient Services Department Type Cod First visit (FV) Follow up (FU)	Type of patient insurance First visit Follow up visit
8) Hospital code	The code of referral hospital /clinic
9) Medical Department code	The code of medical/clinic
10) Visit Number Count	Derived: count how many Visit Number
11) Percentage of Show visits	Derived: Calculate the percentage of a show on all the previous visits
12) Percentage of No-Show visits	Derived: Calculate the percentage of no show on all the previous visits
13) Lead Time	Derived: the time difference in days between the booking of the appointments and the day of the appointments
Cancellation Flag 14) Yes= appointment cancelled 15) No = not cancelled	Derived: count how many Check-in Cancellation was Yes as total Derived: count how many Check-in Cancellation was No as total
Medical Treatment Reservation Type Code 16) 1= Schedule 17) 2= Walk-in	Derived: count how many Scheduled reservation as total Derived: count how many Walk-in reservations as total
Medical Treatment 18) Yes= patient treated 19) N= not treated	Derived: count how many patients treated as total Derived: count how many patients not treated as total
On Foot visit 20) 1= Yes 21) 0 = No	Derived: count how many On Foot visit was Yes as total Derived: count how many On Foot visit was No as total
Emergency visit 22) 1= Yes 23) 0 = No	Derived: count how many Emergency visits was Yes as total Derived: count how many Emergency visits was No as total
24) Count of Medical Reservation Date & Time	Derived: count how much medical reservation date and time as total
25) Count of Medical Arrival Date & Time	Derived: count how many medical arrival dates and time as total
Medical Treatment Room Arrival Date & Time 26) Show 27) No-Show	Derived: count how much medical treatment room arrival date and time as total Derived: count how many times the patient came for his/her appointment Derived: count how many patients did not show for his/her appointment
Patient arrival confirmation 28) 1= Yes 29) 0 = No	Derived: count how many patient arrival confirmations was Yes as total Derived: count how many patient arrival confirmations was No as total
Class No-Show	class show=0 or No-show=1

AdaBoost is a meta-learning algorithm that is used to reduce the error of weak learning algorithms significantly. Boosting works by repeatedly running input a training set in series rounds and then combining the classifiers into a single ensemble model. The model's prediction is taken as a sum of the weighted predictions [25]. On the other hand, Naive Bayes (NB) classifiers are derived from the Bayes' theorem, which assumes each attribute is independent. Each attribute has the probability to determine the classification outcomes independently [29].

E. Evaluation Criteria

The dataset was divided into two sets: the training set was used to develop the model while the testing set was used for validation. The training set included 23,135,254 records (70%) outpatient records, while the remaining 30% is used for testing the trained models. The evaluation metrics used to choose the best model are Precision and Recall that are defined as follows:

- Precision: shows how many of the predictions were correctly identified. Eq. 1 shows the calculation of it.

$$\text{Precision} = (\text{TruePositive}) / (\text{TruePositive} + \text{FalsePositive}) \quad (1)$$

- Recall: presents how many of the correct outcomes were identified. Eq. 2 shows the calculation of it.

$$\text{Recall} = (\text{TruePositive}) / (\text{TruePositive} + \text{FalseNegative}) \quad (2)$$

The precision and recall were used since they are influenced by the false positive and false negative. The administration is trying to minimize the costs of a no-show by reducing the false positive while physicians were interested in minimizing the overbooking by reducing the false negative. However, the trained model might be sufficient by deficit prediction using recall and precision [30].

IV. EXPERIMENTAL SETUP AND RESULTS

To show the DNN model's performance using real data, the DNN is compared with AdaBoost and NB. Weka version 3.8 is used to run AdaBoost and NB. Weka is a data mining tool developed by the University of Waikato, Hamilton, New Zealand, in 1992 [31]. On the other hand, TensorFlow [14] developed by Google is applied to build the DNN model. All the models were trained on a standalone server running Linux Centos version 6.9 with 12 CPU cores and 48GB of RAM.

Table IV shows the performance of the machine learning models per class on an unseen testing dataset. The best performing model was the Deep Neural Network (DNN). It achieved better precision and recall per class compared to AdaBoost and NB. The AdaBoost model was the second-best model, while NB was the lowest one.

Table V shows the weighted average performance of the machine learning models. DNN was the best model too. It achieved a precision and recall of 98.2% and 94.3%, respectively. The result showed that DNN performed excellently in the prediction of no-show and show based on evaluation criteria. Overall, the best machine learning models were chosen based on the highest performance using the evaluation metrics discussed.

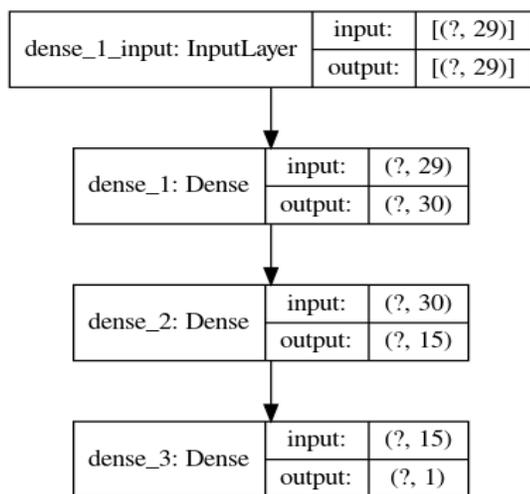


Fig. 3. The Deep Neural Network (DNN) Model Graphed by Tensorflow.

TABLE IV. THE PERFORMANCE PER CLASS FOR THE MACHINE LEARNING MODELS ON THE TEST DATASET

	DNN		AdaBoost		NB	
	S	NS	S	NS	S	NS
Precision	0.7533	0.9815	0.732	0.977	0.416	0.987
Recall	0.9068	0.9433	0.886	0.938	0.947	0.747

S=Show, NS= No-show

TABLE V. THE WEIGHTED AVERAGE PERFORMANCE OF THE MACHINE LEARNING MODELS ON THE TEST DATASET

	DNN	AdaBoost	NB
Precision	0.982	0.938	0.895
Recall	0.943	0.930	0.779

V. DISCUSSION

The accurate prediction of a patient's show is one of the most interesting and challenging healthcare providers' tasks. With the advent of new technologies in health learning, large amounts of outpatients' data have been collected and available to the health research community. Accordingly, machine-learning methods have become a popular tool for health researchers. These techniques can discover and recognize patterns and relationships using big datasets while they can effectively predict future outcomes.

In this proof-of-concept study, a vast amount of data were collected from electronic health records. With the potential to improve healthcare quality while reducing the waiting list time, these massive amounts of data, so-called 'big data' support a variety of healthcare functions, including outcome prediction, decision support, and health management. Random forests have been applied successfully using high-dimensional data (33,050,363) visits to investigate machine learning performance for the early predicting no-show in the outpatient clinic and discover potential predictors of no show.

This is the first study that identified big data-derived to the best of our knowledge and included complex relations

between predictors using advanced machine learning. Three key differences in the current study compared to the other studies that need to be recognized. Our work for the prediction of a no-show demonstrates several feasible advantages for outpatient clinics and provides valid prognostic information from simple features. The result was superior to traditional models (e.g., AdaBoost and NB) in predicting patients with no-show that were adopted in previous studies. The approach here classifies the clinical features and extracts rules for identifying patients at high risk of no-show for individual records. The question is how to intervene when a patient is predicted as a no-show. The management could utilize different methods, such as SMS or phone call, to remind high-risk patients of no-show about their appointment, especially in the case of a long time since the last appointment.

Thus, appointments will be managed effectively, and the negative impact they have on the operational efficiency of systems in healthcare organizations will be decreased. A patient can benefit from additional engagement through SMS reminders or investigate associated factors to avoid no-show to avoid no-shows. According to Stubbs et al. [32], SMS is the most effective way to remind patients and decrease the no-show rate. Lastly, our framework is generic and can be adopted by other outpatient clinics characterized by rates of no-shows and appointment-based patient history. Our study has two limitations. First, given that the study was limited to the central region facilities, generalizing the results across other regions is unclear. Second, our study did not account for many of the personal or environmental factors that were not available in the EMR, such as transportation and weather.

The developed model will be adopted in practice leading by the Information System and Informatics Division (ISID) under the umbrella of the MNGHA. With the model's ability to provide no-show prediction back in a meaningful way, the expectation is that the deployment will reflect the test data's results. The research team's intent and the ISID to deploy this model in the central region into production as a pilot phase, then a full roll-out to the entire organization in all regions, would follow. In addition to providing the classifications, the research team will design studies to determine the current model's effectiveness and several interventions to reduce no-shows. Additionally, the model can be improved by adding more features, e.g., medication refills, lab appointments, or special clinic orders. Besides, the model will be retrained every 6-12 months in order to maintain the model accuracy.

The deep learning model showed a high performance in predicting no-show. However, deep learning has some disadvantages, such it will not be able to provide specific recommendations to manage the risk factors that influence no-show. The DNN cannot rank the importance of the features due to the neural network's hidden layers. Furthermore, the research study has other limitations. The knowledge extracted by the machine learning algorithms were based on medical record databases. Other data based on weather conditions, cultural factors, patient education, etc. might improve model accuracy and explain the no-show rate.

VI. CONCLUSION

This work shows how machine learning can be effectively adopted in the health field to derive models that use patient data to predict an outcome of interest. Machine learning may be applied to the construction of models to predict patients at high risk of no-show for appointments using a data-driven approach, which – once evaluated and tested – may be embedded within health care systems. DNN, AdaBoost, and NB algorithms used to utilize EMR data in predicting no-show. The results show that the built model is effective and provides insightful implications for decision-making by management. The best predictive model was DNN with Precision and Recall of 0.982 and 0.943, respectively. More improvement can be achieved by adding more features, e.g., medication refills, lab appointments, or special clinic orders.

VII. FUTURE WORK

Future research should evaluate the ability of such approaches to predict no-show patients and missed by traditional algorithms then translate into better-quality clinic outcomes. Furthermore, a multistage machine learning platform is considered a future improvement where the follow-up appointments have its prediction model.

VIII. CONFLICT OF INTEREST

None of the authors has any competing interests.

IX. HUMAN SUBJECTS PROTECTIONS

No human subjects were involved in the project.

REFERENCES

- [1] Chariatte V, Berchtold A, Akre C, Pierre-André M (2008) Missed Appointments in an Outpatient Clinic for Adolescents, an Approach to Predict the Risk of Missing. *J Adolesc Heal* 43:38–45.
- [2] Carlsen KH, Carlsen KM, Serup J (2016) Non-attendance, predictors and interventions. *Adherence in Dermatology* 29–35.
- [3] Davies M, Goffman R, May J, Monte R, Rodriguez K, Tjader Y, Vargas D (2016) Large-Scale No-Show Patterns and Distributions for Clinic Operational Research. *Healthcare* 4:15.
- [4] O. AlRowaili M, Ahmed AE, Areabi HA (2016) Factors associated with No-Shows and rescheduling MRI appointments. *BMC Health Serv Res* 16:1–7.
- [5] Almomani I, AlSarheed A (2016) Enhancing outpatient clinics management software by reducing patients waiting time. *Journal of Infection and Public Health. J Infect Public Health* 9:734–743.
- [6] Samorani M, Laganga LR (2015) Outpatient appointment scheduling given individual day-dependent no-show predictions. *Eur J Oper Res* 240:245–257.
- [7] Srinivas S, Ravindran AR (2018) Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. *Expert Syst Appl* 102:245–261.
- [8] Reinsel D, Gantz J, Rydning J (2017) Data Age 2025 : Don't Focus on Big Data; Focus on the Data That's Big. IDC White Pap 1–25.
- [9] El Fouki M, Akin N, El Kadiri KE (2017) Intelligent adapted e-learning system based on deep reinforcement learning. *ACM Int Conf Proceeding Ser* 2–5.
- [10] Zuech R, Khoshgoftaar TM, Wald R (2015) Intrusion detection and Big Heterogeneous Data: a Survey. *J Big Data*. doi: 10.1186/s40537-015-0013-4.
- [11] Hossain M, Rahman SN, Bhattacharya P, Jacks G, Saha R, Rahman M (2015) Sustainability of arsenic mitigation interventions-an evaluation of different alternative safe drinking water options provided in Matlab, an arsenic hot spot in Bangladesh. *Front Environ Sci*. doi: 10.3389/fenvs.2015.00030.
- [12] Lipton ZC, Kale DC, Elkan C, Wetzel R (2016) Learning to diagnose with LSTM recurrent neural networks. 4th Int Conf Learn Represent ICLR 2016 - Conf Track Proc 1–18.
- [13] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118.
- [14] Abadi M, Barham P, Chen J, et al (2016) TensorFlow : A System for Large-Scale Machine Learning This paper is included in the Proceedings of the TensorFlow : A system for large-scale machine learning. 12th USENIX Symp Oper Syst Des Implement (OSDI '16) 265–283.
- [15] Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, Pianykh O (2017) Predicting No-Shows in Radiology Using Regression Modeling of Data Available in the Electronic Medical Record. *J Am Coll Radiol* 14:1303–1309.
- [16] Chua SL, Chow WL (2019) Development of predictive scoring model for risk stratification of no-show at a public hospital specialist outpatient clinic. *Proc Singapore Healthc* 28:96–104.
- [17] Dantas LF, Hamacher S, Cyrino Oliveira FL, Barbosa SDJ, Viegas F (2019) Predicting Patient No-show Behavior: a Study in a Bariatric Clinic. *Obes Surg* 29:40–47.
- [18] Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, Waki K, Noguchi T, Ohe K (2016) Machine-Learning-Based Prediction of a Missed Scheduled Clinical Appointment by Patients with Diabetes. *J Diabetes Sci Technol* 10:730–736.
- [19] Mohammadi I, Wu H, Turkcan A, Toscos T, Doebbeling BN (2018) Data Analytics and Modeling for Appointment No-show in Community Health Centers. *J Prim Care Community Heal* 9:1–11.
- [20] Nelson A, Herron D, Rees G, Nachev P (2019) Predicting scheduled hospital attendance with artificial intelligence. *npj Digit Med* 2:1–7.
- [21] Lenzi H, Ben ÂJ, Stein AT (2019) Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil. *PLoS One* 14:1–14.
- [22] Lee G, Wang S, Dipuro F, Hou J, Grover P, Low LL, Liu N, Loke CY (2017) Leveraging on predictive analytics to manage clinic No-Show and improve accessibility of care. In: *Proc. - 2017 Int. Conf. Data Sci. Adv. Anal. DSAA 2017*. pp 429–438.
- [23] Dashtban M, Li W (2019) Deep Learning for Predicting Non-attendance in Hospital Outpatient Appointments. In: *Proc. 52nd Hawaii Int. Conf. Syst. Sci*. pp 3731–3740.
- [24] Biau G, Scornet E (2016) A random forest guided tour. *Test* 25:197–227.
- [25] Freund Y, Schapire RE (1996) Experiments with a New Boosting Algorithm. In: *Proc. 13th Int. Conf. Mach. Learn*. pp 148–156.
- [26] Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323:533–536.
- [27] Bengio Y (2009) Learning deep architectures for AI. doi: 10.1561/22000000006.
- [28] Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35:1798–1828.
- [29] Bricher D, Müller A (2020) A supervised machine learning approach for intelligent process automation in container logistics. *J Comput Inf Sci Eng* 20:1–9.
- [30] Ghosh S, Rana A, Kansal V (2020) A benchmarking framework using non-linear manifold detection techniques for software defect prediction. *Int J Comput Sci Eng* 21:593–614.
- [31] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software. *ACM SIGKDD Explor Newsl* 11:10–18.
- [32] Stubbs ND, Geraci SA, Stephenson PL, Jones DB, Sanders S (2012) Methods to reduce outpatient non-attendance. *Am J Med Sci* 344:211–219.