Developing a Predictive Model of Predicting Appointment No-Show by Using Machine Learning Algorithms

Abdulwahhab Alshammari[1,2,3], Raed Almalki[1,2,3], and Riyad Alshammari[4]
[1] Health Informatics Department, College of Public Health and Health Informatics, Ministry of National Guard Health Affairs, Riyadh, KSA
[2] King Saud Bin Abdulaziz University for Health Sciences (KSAU-HS), Ministry of National Guard Health Affairs, Riyadh, KSA
[3] King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard Health Affairs, Riyadh, KSA
[4] National Center for Artificial Intelligence (NCAI) - Research Labs, Riyadh 11543, Saudi Arabia
Email: raalshammari@sdaia.gov.sa

***Abstract*—Introduction: Patient no-shows are defined as patients who missed outpatient appointments, either for diagnostic or clinic tests. Identifying those patients is necessary for clinicians and healthcare settings to utilize the resources and improve healthcare efficiency appropriately. This research paper aims to develop a predictive model based on machine learning algorithms to predict patients' failure to attend scheduled appointments. A public data set was divided into training and testing data sets. Two machine learning algorithms, namely decision trees and AdaBoost, were evaluated based on Precision, Recall, True Positive Rate, False Negative Rate, F-measure, and Receiver Operating Characteristic (ROC). Results showed that the decision tree outperformed AdaBoost. The most significant predictors were age and lead time.**

***Index Terms*—no-show, decision tree, AdaBoost, machine learning**

I. Introduction

Patient no-show refers to non-attendance of scheduled appointments in healthcare settings. This recurring problem's seriousness lies in the ineffective use of human resources and increases patient waiting time. Besides, it implies a negative impact through lost time for physicians and nurses, leading to decreased productivity and quality of care. The most challenging aspect in the prediction of a no-show is to anticipate and predict human behavior patterns. However, some associated variables/attributes facilitate the prediction of patient missed appointments.

No-shows can be used as a metric of the finite resources, budget cuts, and increased costs that face any healthcare organization. It can describe the under-utilization of outpatient appointments that could affect patient satisfaction through access to care, clinic costs, and decrease clinician productivity [1].

Specifically, the predictive model of no-shows would enable a hospital to predict patients' probability of attending scheduled appointments with a high certainty level. Further, this could improve the appointment scheduling, affect provider overtime, patient waiting time, and the number of patients seen per day; the improvement may reach 30% compared to the current scheduling system in the radiotherapy center [2]. Accurately predicting no-shows could allow less waiting time, which allows hospitals to scale up their capacity during long-term strategic planning and minimize health care costs [3].

This study aims to apply and compare a Decision Tree (D.T.) and AdaBoost algorithms on clinical data to predict no-shows and choose the optimum algorithm according to various matrices. Another objective is to identify the features that consider significant predictors of no-shows. The main contributions of this paper are:

i. Show the ability of the machine learning algorithms to predict appointment no-shows with high performance.

ii. Identify minimalist attributes that are important in the prediction of appointment no-shows.

The rest of this paper is organized as follows. Section II covers the literature review, while Section III provides details on the evaluation method. Sections IV, V shows the experimental results and covers details on the solutions' analysis. Finally, Section VI presents conclusions and future work.

II. Literature Review

Several studies have widely investigated the extent of no-shows. In a systematic review, Dantas *et al.* [4] reported an average no-show rate across all studies of 23.0%. Their finding was based on 105 papers about the determination of no-shows. They used different techniques to minimize no-shows that included sending messages, patient navigation, overbooking, reminder procedure, and a high rate of no-shows persisted. In another study, Daggy *et al.* [1] reported that overbooking techniques were used to address no-shows in scheduled appointments to improve clinic efficiency.

Manuscript received January 13, 2021; revised June 24, 2021.

doi: 10.12720/jait.12.3.234-239

In an Oncology clinic, Lima *et al.* [5] reported using a patient navigation technique to build a predictive model to identify high-risk no-shows. They managed to reduce the rate by a 42% reduction at the Cancer clinic. Moreover, previous techniques' success mainly depends on a predictive model to achieve a higher decrease in no-show rates, predictions of most variables affecting no-shows, and preventing the consequences of no-shows, such as discontinuing patient treatment [6].

Previous studies used two approaches to address no-show. The first approach was mainly on developing a predictive model using a selective overbooking strategy that capable of decreasing the negative impact of no-shows for patient schedules system [1], [7]-[10]. Overbooking refers to scheduling an additional number of patients that leads to patient waiting time during outpatient visits that might affect patient satisfaction and increase providers' over time, which could affect hospital costs [3]. The overbooking strategy has succeeded in other industries, such as hospitality or transportation, which is not applicable in healthcare for several reasons [3]. The second approach was developing a prediction model. Due to machine learning effectiveness in the classification and prediction, researchers are increasingly using machine-learning algorithms to identify patients' missed appointments [11]-[14]. The second approach, the prediction model, effectiveness significantly relies on high accurate prediction associated with appointment reminder systems, usually consisting of email reminders and phone messages for those at high risk of a no-show.

Recently, the literature showed several studies that discussed the prediction of missed appointments by machine learning in healthcare. Lee *et al.* [9] aimed to predict no-shows and demonstrate how the risk scores were produced. They collected two years of follow up data with a no-show rate of 25%. Their model had 71.8% accuracy for Logistic regression and decision tree and 72.9% accuracy for the random forest. Moreover, 37 features were derived from the dataset, whereby the top features ranked by XGBoost were days since the last visit and last appointment status. Although they achieved good model performance, the model was applied for a small group of patients with a total number of 400.

Goffman *et al.* [10] developed a predictive model to identify patients at risk for missing appointments in Veterans Health while they modeled past attendance behavior by an empirical Markov model. They created 24 predictive models using logistic regression. They found that younger patients had a high probability of missing their appointments, while married patients had a low likelihood of missing an appointment. The no-show rate decreased from 51% to 13% due to multiple same-day appointments, meaning that the patient's probability of missing an appointment increases when the lead time increases. Their results were 0.762 for the training dataset and 0.713 for the test set using the receiver operator characteristic for all models. The reminder system could reduce the no-show rate by 9.9% and 15.89% if a patient received reminder calls 24 and 72 hours in advance.

A different study was published the same year, 2017, which attempts to test whether data available in Electronic Medical Records (EMR) can be effectively leveraged to predict a missed scheduled radiology examination [11]. The collected dataset consists of 54,652 patient visits with radiology examinations with a no-show rate of 6.5%. Logistic regression was developed with an area under the curve (AUC) of 0.753, focusing on the patient-specific subsets. The study has limitations to the developed model applied to a single academic medical center. Besides, the dataset has the ubiquity of errors and incompleteness in EMR design, which affects the quality of the model because it is dependent on the quality and availability of the data.

Mohammadi *et al.* proposed an experimental comparison of three different models for predicting missed appointments [8]. The predictive models were developed using logistic regression, neural network, and naïve Bayes classifier to predict missed appointments. Their study was based on creating a predictive model regarding the no-show focusing on various outpatient specialties. They compared the three models based on their resulted accuracies. They found that the Naïve Bayes classifier model outperformed the other two. Furthermore, the authors claim that their dataset's clinical specialty variable did relate to a patient's clinical characteristics. This assumption is considered a significant limitation in the study since the dataset did not have information on patients' clinical, functional, and diagnosis (e.g., heart disease, depression, pneumonia, etc.).

In this study, the authors identified variables from previous studies that affect the no-show rates' predictive models. The variables covered the date of previously scheduled visits, the patient's initial no-show rate, appointment compliance history during the last five years, new patient status, and early morning appointment. Besides, patient characteristics are important factors that contribute to the prediction of no-show; these patterns, such as younger patients, unmarried, or male, tend to miss their appointments [9]-[14]. On the other hand, Dashtban and Li [13] utilized several attributes based on clinical, socioeconomic, and environmental factors. They built a Deep Neural Network (DNN). The DNN model achieved an AUC of 0.71, recall of 0.78, and accuracy of 0.69.

III. Methods

The dataset was taken from the Kaggle database related to medical appointments, for appointments scheduled between April 29, 2016, and June 8, 2016 [14]. The datasets included (110,528) appointments were collected for all adult patients who had the appointments while pediatric patients were excluded.

The original data set consists of 14 attributes. New attributes were derived from the dataset. The new attribute is called "lead time," which is calculated as the duration in days between the scheduling date and the appointment date. Another attribute generated was Ante Meridiem (AM) appointment, which was derived from appointment time since morning appointments consider an essential feature according to the literature. Demographic attributes such as age were grouped into five intervals, a standard method to handle adult healthcare data. Other attributes, such as gender, appointment time, and no-show status (class), were coded as a single binary variable. The python implementation of the Recursive Feature Elimination (RFE) was used to exclude irrelevant attributes from the original dataset. Other attributes were excluded due to missing metadata of these attributes, such as address and scholarship. Table I shows the final attributes. Running the REF method is to decrease the complexity of the data and remove irrelevant attributes since including all variables may lead to high complexity in modeling. A total of 82,992 appointments was included in this study. The no-show baseline rate out of the preprocessed dataset was 19.62%. All data cleansings and preprocessing were performed using Python software, version 3.7.5.

1) Machine learning algorithms

The decision tree can be used for classification and has recently been widely used in the medical field (diagnosis of breast cancer [15], diagnosis of diabetes [16], and predicting heart disease [17]). The reason behind D.T.'s common utilization is that it can be converted into IF-THEN rules, which are commonly used in the medical domain. D.T. was used. Among them, studies in the following fields can be cited:

Unlike other machine learning algorithms, D.T. construction is based on information gain [18]. In other words, it allows us to determine which attribute in a given set of training feature vectors the most useful for discriminating between two classes. Another advantage of using D.T. is the capability of generating user-friendly rules to enable an easy understanding of the algorithm to end-user in the prediction model.

The results of D.T. can be shown in the form of a tree or a set of if-then rules. The advantage of this method algorithm is the simplicity of the result and easy to interpret. D.T. is usually considered binary splits, finding the optimal portioning by splitting each node into two sub-groups. Some node line has a resulting region complicated to describe while another node has a simple description. Each leaf represents a value of the dependent variable. Hence, DT provides a clear indication of important features.

Adaptive boosting or (AdaBoost) [19] is a boosting algorithm that is linearly combining multiple base classifiers or weak learners to correct its predecessor, focusing on the training instances that the predecessor under fitted and to produce a form of a committee whose powerful and significantly better than random. The boosting technique works by applying the base classifiers sequentially to weighted versions of each iteration. After a first base was trained and used to make predictions on the training set, the classifier decreases the weight of correctly classified instances while increasing the misclassification weight. The motivation for boosting was to choose base classifiers arbitrarily and the ability to train them in a weighted manner. In this study, a multiclass version of AdaBoost called (Stagewise Additive Modeling using a Multiclass Exponential loss function) was utilized from the scikit-learn package implemented in python.

2) Evaluation criteria

The dataset was split into two samples: 70% for the training dataset and 30% for the test dataset. Several matrices were applied to select the best model in predicting no-show: True Positive Rate, False Positive Rate, recall, precision, Receiver Operating Characteristics (ROC) f-measure. The metrics were calculated as follows:

- TPR represents the number of patients classified as high risks of non-attendance, calculated based on formula 1.

$$TPR=True\ Positive/(True\ Positive+False\ Negative) \quad (1)$$

- FPR: represents the number of patients classified as low risks of non-attendance, calculated based on formula 2.

$$FPR=False\ Positive/(False\ Positive+True\ Negative) \quad (2)$$

- Precision: represents the percentage of non-attendance patients classified as positive that were positive, which is calculated based on formula 3.

$$Precision=True\ Positive/(True\ Positive+False\ Positive) \quad (3)$$

- Recall: represents the percentage of patients who did not attend their appointments classified correctly, which is calculated based on the formula (4).

$$Recall=True\ Positive/(True\ Positive+False\ Negative) \quad (4)$$

- *F-measure*: represents the harmonic mean of precision and recall and is calculated based on the formula (5).

2 $$F\_score =2*(precision*recall)/(precision+recall) \quad (5)$$

- Receiver Operating Characteristic (ROC) curve: is a graphical way to display true positives versus false positives across a series of cut-offs and selecting the optimal cut-off.

IV. Results

A total of 82,988 appointments for 47,685 unique patients with a mean age of 46 years, 69.8% were females. Patients had scheduled appointments from April 29, 2016, to June 8, 2016, analyzed. Table I shows a list of factors considered in the model.

TABLE I. List of Factors Considered in the Model

Attributes	Number	Show	No-show	No-show rate
Gender				
Male	25142	20363	4779	(19%)
Female	57849	46340	11509	(19.89%)

Age (years old)				
18-24	9879	7410	2469	(24.98%)
25-34	14330	10991	3339	(23.3%)
35-44	14488	11374	3114	(21.49%)
45-54	15446	12619	2827	(18.3%)
55-64	14505	12181	2324	(16%)
≥65	14343	12128	2215	(15.44%)
Lead Time (Days)				
0 (same day)	27295	26286	1009	(3.69%)
1-10	30005	22760	7245	(24.15%)
11-20	10157	6973	3184	(31.35%)
21-30	7715	5288	2427	(31.46%)
31-40	3653	2450	1203	(32.93%)
41-50	1534	1052	482	(31.42%)
51-60	758	522	236	(31.13%)
61-70	1061	788	273	(25.73%)
71-80	412	281	131	(31.80%)
81-90	207	156	51	(24.64%)
>90	194	147	47	(24.23%)
SMS received				
Yes	27424	20072	7352	26.8%
No	55567	46631	8936	16.08%
Appointment Time				
AM	32171	25279	6892	21.42%
PM	50820	41424	9396	18.49%
Diabetes	7849	6443	1406	17.91%
Hypertension	21678	17943	3735	17.22%
Handicap	1850	1550	300	16.22%

The patients' largest age group was 45-54 years old (17.356%), and the no-show rate was higher for those 25-34 years old, while for patients who were older than 65 years, the no-show rate was the lowest (13.59%). The no-show rate linearly decreased as the patients getting old.

For a total of 36.19% of appointments, the patients waited for 1–10 days, and the no-show rate was the lowest when the appointments were on the same day of the scheduled day (6.19%). The majority of the appointments did not receive a text message as a reminder. Hence, a higher no-show rate (54.86%) was noticed.

The metrics, namely TPR, FNR, Recall, Precision, Receiver Operating Characteristic, and F-score, were used to compare the two prediction models' overall performance. Table II presents the results of the different evaluation matrices. In comparing the prediction results for predicting no-shows, AdaBoost achieved a higher value of TPR (0.95) than D.T., while for the remaining matrices, D.T. outperformed AdaBoost. Fig. 1 and 2 present the ROC for AdaBoost and DT.

TABLE II. PERFORMANCE COMPARISON OF DECISION TREE AND ADABOOST ALGORITHMS.

Evaluation Criteria	AdaBoost	Decision Tree
TPR	0.95	0.89
FNR	0.17	0.14
Precision	0.87	0.89
Recall	0.83	0.86
ROC	0.85	0.88
F-measure	0.84	0.87

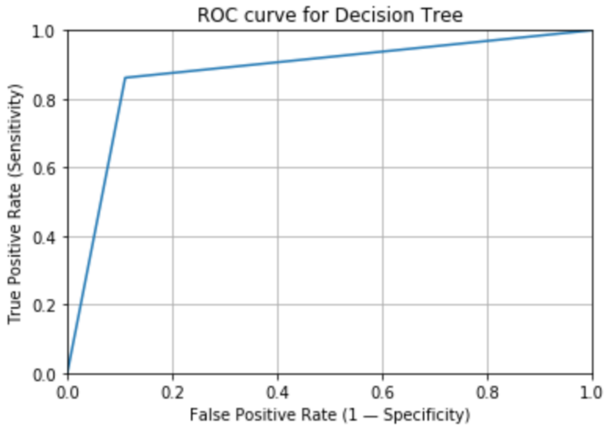

Figure 1. ROC curve for D.T.

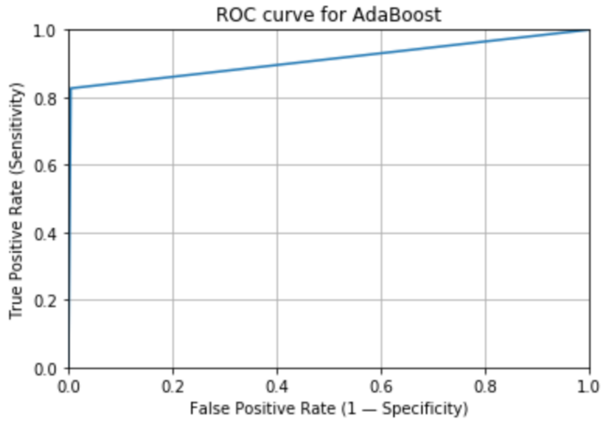

Figure 2. ROC curve for AdaBoost.

V. DISCUSSION

The prediction of patients who might not attend an appointment is an essential challenge in the health domain. A comparison between the decision tree model and the AdaBoost model in predicting no-shows was carried in the present study. Decision Tree is often used to interpret significant factors correlated with no-shows and has been used to develop a predictive model [17]-[19]. To the best of our knowledge, this study builds and compares two models using a large-scale dataset. The D.T. and AdaBoost models were analyzed to identify the most important attributes. The most important attributes were: diabetes, hypertension, alcoholism, lead time, AM appointments, handicap, and SMS. The models' analysis indicates that those patients who missed their appointments tend to be younger, male, with morning appointments, and did not receive a text message on the phone message a reminder.

The D.T. model showed a high performance in predicting no-show. However, D.T. has some disadvantages, such it will suffer from the complexity of the tree and the robustness. Furthermore, the research study has several limitations. The knowledge extracted by the machine learning model was mainly based on the public data set. The model needs to be tested on more datasets from different kinds of patients.

Comparing with the previous study, the developed model not only showed the ability of machine learning algorithms (e.g., decision tree) to predict no-show with high accuracy but define minimalist risk factors that can shed some light to know the reasons behind missing the appointments by the patient. Further studies are needed to study the environmental, educational, and cultural factors that make patients miss appointments.

VI. Conclusion

In this study, two machine learning algorithms were compared to predict appointment no-shows. The approach involves a feature selection and building predictive models using AdaBoost and decision tree. Both algorithms were trained and tested on a public dataset. The models were evaluated according to various evaluation matrices. A decision tree can be embedded in the Electronic Health Record (EHR) system as If-then rules to early predict the possibilities of appointment no-shows and provide a list of high-risk patients. The hospital can contact the patient through phone calls or text messages to remind the high-risk patients of their appointment

In conclusion, the decision tree outperforms AdaBoost in precision, recall, ROC, and F-measure. Future work is to extend the number of attributes to improve the model's robustness and collect a larger dataset.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

R. Alshammari developed the theoretical formalism and designing of the experiment. R. Almalki carried out the experiments and contributed to interpreting the results and findings. Both R. Alshammari and A. Alshammari provided critical feedback and helped shaped the final version of the manuscript. All authors participated in reviewing the manuscript.

Acknowledgment

This study was funded by the King Abdullah International Medical Research Center (KAIMRC), National Guard, Health Affairs, Riyadh, Saudi Arabia, with research grant No. RYD-19-419812-121138.

References

[1] J. Daggy, *et al*., "Using no-show modeling to improve clinic performance," *Health Informatics J.*, vol. 16 pp. 246-259, 2010.
[2] D. Bentayeb, N. Lahrichi, and L. M. Rousseau, "Patient scheduling based on a service-time prediction model: A data-driven study for a radiotherapy center," *Health Care Manag. Sci.*, vol. 22, pp. 768-782, 2019.
[3] Z. Huang, M. Ashraf, H. Gordish-Dressman, and P. Mudd, "The financial impact of clinic no-show rates in an academic pediatric otolaryngology practice," Am. J. Otolaryngol., vol. 38, pp. 127-129, 2017.
[4] L. F. Dantas, J. L. Fleck, F. L. C. Oliveira, and S. Hamacher, "No-shows in appointment scheduling – A systematic literature review," *Health Policy*, vol. 122, pp. 412-421, 2018.
[5] S. Percac-Lima, P. R. Cronin, D. P. Ryan, B. A. Chabner, E. A. Daly, and A. B. Kimball, "Patient navigation based on predictive modeling decreases no-show rates in cancer care," *Cancer*, vol. 121, pp. 1662-1670, 2015.
[6] H. Kurasawa, *et al.*, "Machine-Learning-Based Prediction of a missed scheduled clinical appointment by patients with diabetes," *J. Diabetes Sci. Technol.*, vol. 10, pp. 730-736, 2016.
[7] A. Alaeddini, K. Yang, P. Reeves, and C. K. Reddy, "A hybrid prediction model for no-shows and cancellations of outpatient appointments," *IIE Trans. Healthc. Syst. Eng.*, vol. 5, pp. 14-32, 2015.
[8] I. Mohammadi, H. Wu, A. Turkcan, T. Toscos, and B. N. Doebbeling, "Data analytics and modeling for appointment no-show in community health centers," *J. Prim. Care Community Heal.*, vol. 9, pp. 1-11, 2018.
[9] G. Lee, *et al.*, "Leveraging on predictive analytics to manage clinic no-show and improve accessibility of care," in *Proc. Int. Conf. Data Sci. Adv. Anal.*, 2017, pp. 429-438.
[10] R. M. Goffman, *et al.*, "Modeling patient no-show history and predicting future outpatient appointment behavior in the veterans health administration," *Mil. Med.*, vol. 182, pp. e1708-e1714, 2017.
[11] H. B. Harvey, *et al.*, "Predicting no-shows in radiology using regression modeling of data available in the electronic medical record," *J. Am. Coll. Radiol.*, vol. 14, pp. 1303-1309, 2017.
[12] H. Lenzi, Â. J. Ben, and A. T. Stein, "Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil," *PLoS One*, vol. 14, pp. 1-14, 2019.
[13] M. Dashtban and W. Li, "Deep learning for predicting non-attendance in hospital outpatient appointments," in *Proc. 52nd Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 3731-3740.
[14] V.2 K. (2016). Medical appointment no shows. Version 3. In Kaggle. [Online]. Available: https://www.kaggle.com/joniarroba/noshowappointments/version/3
[15] A. M. Elsayad and H. A. Elsalamony, "Diagnosis of breast cancer using decision tree models and SVM," *Int. J. Comput. Appl.*, vol. 83, pp. 19-29, 2013.
[16] A. Iyer, S. Jeyalatha, and R. Sumbaly, "Diagnosis of diabetes using classification mining techniques," *Int. J. Data Min. Knowl. Manag. Process.*, vol. 5, pp. 1-14, 2015.
[17] M. Abdar, "Using decision trees in data mining for predicting factors influencing of heart disease," *Carpathian J. Electron. Comput. Eng.*, vol. 2, pp. 31-36, 2015.
[18] E. Alpaydin, *Introduction to Machine Learning*, third edition, MIT Press, 2014.
[19] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Mach. Learn.*, 1996, pp. 148-156.

Dr. Riyad Alshammari is an Associate Professor in Computer Science, senior Data Scientist and Artificial Intelligence consultant in the National Canter for Artificial Intelligence, Riyadh, Saudi Arabia. Dr. R. Alshammari studied at Dalhousie University, Halifax, N.S., Canada, and graduated with a Bachelor of Computer Science in 2005, Master of Computer Science in 2008, and Doctor of Philosophy in Computer Science in 2012. Dr. R. Alshammari is specialized in big data, data modeling, artificial network, and machine learning. Dr. R. Alshammari's research interests include but are not limited to the artificial network, Machine Learning, Clinical Informatics, e-Health, and data analytics. He has published his research in leading peer-reviewed journals and international conferences and has attended many conferences and workshops in data science and artificial intelligence. Dr. R. Alshammari is a member of many IEEE societies and review committees of several IEEE international conferences and journals.

Mr. Raed Almalki received a Master's degree in health informatics from the Department of Health Informatics, College of Public Health and Health Informatics, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia. Mr. Almalki maintained an "A" GPA during his Master's degree. His involvement and interaction during classes and after graduation influenced his research interest. His main research interest focuses on Machine Learning and health informatics. In particular, his research interest focuses on how can Machine Learning affect patients, physicians, and hospitals, as it plays a crucial role in clinical decision support and early detection of chronic diseases to improve healthcare outcomes.

Dr. Abdulwahhab Alshammari is an Assistant Professor at the Department of Health Informatics, College of Public Health and Health Informatics, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia. Dr. A. Alshammari received a Ph.D. from the University of Washington, Seattle, Washington, U.S.A., in Biomedical and Health Informatics, University of Washington ranked one of the top 10 universities globally based on the U.S. News & World Report's Rankings. Dr. A. Alshammari graduated with a Master's degree in Health Informatics and a Bachelor's degree in Computer Science from Dalhousie University, Halifax, N.S, Canada. Dr. A. Alshammari has many publications, various skills, deep understanding of the Health Informatics field challenges and future needs, which shaped his abilities and skills to meet healthcare stakeholders' expectations and needs. His research interest focuses on Clinical Informatics, Consumer Health Informatics, User-centered Design and Evaluation, Infectious and Communicable Conditions reporting, and Usability. He is a member of many health informatics societies and organizations.