Deep learning for predicting non-attendance in hospital outpatient appointments

Conference or Workshop Item

Published Version

Dashtban, M. and Li, W. ORCID: https://orcid.org/0000-0003-2878-3185 (2019) Deep learning for predicting non-attendance in hospital outpatient appointments. In: 52nd Annual Hawaii International Conference on System Sciences (HICSS), pp. 3731-3740. Available at https://centaur.reading.ac.uk/79367/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://hdl.handle.net/10125/59809

www.reading.ac.uk/centaur

CentAUR

Deep Learning for Predicting Non-attendance in Hospital Outpatient Appointments

M Dashtban
Informatics Research Centre, Henley Business
School, University of Reading
m.h.dashtban@pgr.reading.ac.uk

Weizi Li
Informatics Research Centre, Henley Business
School, University of Reading
weizi.li@henley.ac.uk

Abstract

The hospital outpatient non-attendance imposes huge financial burden on hospitals every year. The non-attendance issue roots in multiple diverse reasons which makes the problem space particularly complicated and undiscovered. The aim of this research is to build an advanced predictive model for non-attendance considering whole spectrum of factors and their complexities from big hospital data. We proposed a novel non-attendance prediction model based on deep neural networks. The proposed method is based on sparse stacked denoising autoencoders (SSDAEs).

Different with exiting deep learning applications in hospital data which have separated data reconstruction and prediction phases, our model integrated both phases aiming to have higher performance than divided-classification model in predicting tasks from EPR. The proposed method is compared with some well-known machine learning classifiers and representative research works for non-attendance prediction. The evaluation results reveal that the proposed deep approach drastically outperforms other methods in practice.

1. Introduction

Missed appointments have obvious operation and financial implications for health-care systems around the world and health impact for the patient group who have unmet health needs [1]. For example, in 2014 to 2015 around 5.6 million (9% of the total) NHS outpatient appointments were missed in England [2], [3]. The non-attendance can lead to worse care for patients, inefficient use of staff, and increased waiting times. An estimate by the National Audit Office claimed that missed first outpatient appointments cost the NHS up to £225 million in 2012 to 2013 [4]. Another estimate has placed the cost of missed UK general practice (GP;

community-based family medicine) appointments at £150 million per year[5]. Recent Scottish government data suggest that each missed hospital outpatient appointment costs National Health Services (NHS) Scotland £120 [6]. Similarly, in USA, it is reported that an average no-show rate of 62 appointments per day and an estimated annual cost of $3 million in a community hospital setting [7]. It is also found that no-show and cancellation represented 31.1% of overall scheduled appointments among approximately 45,000 patients per year at a large family practice center with an estimated total annual revenue shortfall of 3 % to 14 % [8].

Understanding the complexity of factors that contribute to non-attendance and predicting patients' behaviors can develop targeted/personalized intervention to increase patient engagement and effective use of healthcare resources. Existing research on hospital non-attendance mainly focus on finding associated factors in specific patient groups such as cardiovascular and diabetes. A wider variables that incorporate social economic, patient demographic and practice factors was proposed to investigate non-attendance patterns for general practices appointment in Scotland but those variables were not analyzed with data yet [9]. Although there are digital innovations developed for secondary hospitals to engage patients through mobile text message reminders, there is no evidence about what the reminder should contain in order to minimize missed appointments [2]. The key challenge is that there is scarce knowledge in pattern recognition and risk prediction of non-attendance in secondary hospital appointment. Moreover, patient behaviour and health usage problems result from a complex interplay of several forces. It includes behaviours, social environment, surrounding physical environments, as well as health care access and quality [10]. There are very few researches studying whole spectrum of big data incorporating those factors and their complexities for non-attendance prediction.

In this research, we develop a novel method to predict non-attendance based on state-of-the-art machine learning algorithms and big data (both in-hospital and outside-hospital data). The proposed method is an end-to-end deep learning model based on sparse stacked denoising autoencoders (SSDAE) that is among the latest autoencoders introduced in the literature. We adopt the SSDAE for data reconstruction and prediction. Our model firstly learns the compact representation of data by which having missing values recovered, resulting in a better data representation. Then it uses a direct layer to predict the non- attendance event with an integrated softmax classification layer. Our approach is demonstrated to be more accurate incorporating the prediction model into hospital systems and daily practices. This research will benefit the hospitals for more targeted intervention and messages to patients and reduce non-attendance rate.

2. Related Work

Existing research on non-attendance mainly focuses on traditional quantitative and qualitative methods analysing factors and probability estimation for population groups.

2.1. Existing methods in analysing non-attendance in hospital appointments

Most of research in this domain studies factors contributing to non-attendance in both specific specialty and all appointments from hospital or general practice. A variety of factors were found effective on patient's attendance in pediatric urology unit [11], pulmonary rehabilitation [12], [13], psychiatric [14]–[16] and HIV [17], primary care [18], inpatient and outpatient in the hospital [19] through analyzing multiple correlation from hospital administrative database. A few studies also used survey and interviews to explore and compare the views of patient and health professionals on the reasons of non-attendance [20]–[23]. The factors relate to inaccessibility, including physical location [24], opening hours and days [25], and barriers such as language, stigma and cultural differences [26], [27] may all be important. However, the interplay between the accessibility of a service and the perceived worthiness of the attendee, or 'candidacy', competing priorities [20], [22], [28], [29] (both self-perceived and as perceived by the service provider) can also lead to differences in how likely particular groups are to 'get into, through and on' with services [30]. Morbidity differences can also affect attendance where the illness reduces the ability to navigate access to the health care system[14]. Variation in social and economic circumstances may mean certain times are inconvenient [31] and/or that the perceived importance of the appointment may vary between social groups in and of itself, or in the context of wider life complexities. Within psychiatry for example, one study found that alcohol and drug users had particularly high non-attendance rates [6], [14]. However above studies have focused single disease areas. Studies of single disease area have produced conflicting results when it comes to designing effective interventions to reduce non-attendance [32]–[35]. This may due to a reliance on small data sets and limited variables in certain specialty settings. The non-attendance in primary care [18], hospital inpatient and outpatient from all specialties [19] are studied focusing on single missed appointment. Factors reported to be associated include age, sex, transport logistics, and clinic or practitioner factors such as booking efficiency and the rapport between staff and patients [21], [22], [31], [33], [36], [37]. Williamson et al. [9] and Ellis et al. [1] focused on the patient demographic and practice factors that predict serial missed appointments in general practice. Although those studies considered multiple missing appointments as one of the factors, only a limited number of patient (age, gender, SIMD, distance, ethnicity, number of consultants per year per patient) and practice variables (SIMD, appointment delay, number of available appointments per patient, average appointment length per patient, urban/rural classification). This has led to a limited coverage of personal health, behavioral, environmental and social support information in the prediction model, lacking the capability of revealing whole spectrum of patterns at the individual level. How the whole spectrum of patterns affects patients' behavior in attendance remains unclear. Furthermore, those studies use population-based techniques rather than at an individual patient level. For example, logistic regression is mostly used to predict the probability of non-attendance by fitting numerical or categorical predictor variables in data to a logit function [1], [38]. The problem with these population-based methods is that they do not differentiate between the behaviors of individual persons and are based on small datasets therefore it will affect the effectiveness of predicting results in practice. At present, little agreement exists on what works in practice to reduce missed appointments [1]. We will use deep learning method to consider a wide range of factors and extract important features and complexities towards meaningful patterns from large dataset and more accurate at the individual level.

2.2 Deep learning in healthcare

Compared with traditional statistical methods, deep learning methods have attracted many researchers and

institutions in clinical research tasks which are difficult or even impossible to solve with traditional methods [39], [40]. They are more robust to learn knowledge from high-dimensional and high-volume data such as health, social economics and environmental information. It has proven to be competent to identify patterns and dependencies with cases superior to human experts. Therefore, deep learning methods provide great potential to present a whole picture embedded in large scale data and reveal unknown structure to better serve prediction of non-attendance risk and effective engagement to optimize the health resource usage.

Deep learning classification from EPR is initially studied to predict disease progression. For example, Choi et al [41] applied recurrent neural network (RNN) in longitudinal time stamped EPR to predict diagnoses and medications for the subsequent visit by building a generic temporal predictive model that covers observed medical conditions and medication uses, followed by the development of specific heart failure prediction model. Pham et al [42] utilize the long-short memory (LSTM) method to model disease progression and predict future risk. Recently more attention is received in using deep learning method to predict the risk of readmission. For example, Nguyen et al. [43] and Wang et al. [44] applied convolutional neural network methods to detect and combines predictive local clinical motifs to stratify the risk of readmission. Jamei et al. [45] developed an artificial neural network model to predict all cause risk of 30- day hospital readmission and Xiao et al. [46] developed a hybrid deep learning model that combines topic modelling and RNN to embed clinical concepts in short-term local context and long term global context to predict readmission. Rajkomar et al. [47] further developed a scalable deep learning model using RNN for prediction across multiple centers without site-specific data harmonization which is validated in readmission task. However, to best of our knowledge, there is no research available to predict non-attendance risk using deep learning methods.

3. Deep learning model based on sparse stacked denoising autoencoders (SSDAE)

It is well-known that in hospital systems there are large amount of missing values [48]. There are several algorithms in the literature to deal with such issues. The simplest way is to replace the missing values with the mean values, median values, or some other statistics. It is obviously fast and simple but not effective as it does not include the relations of such missing values with other known/unknown values. To this point, the SSDAE is an AI solution for reconstructing whole data instead on recovering each of which independently.

Additionally, learning highly non-linear and complicated patterns such as the relations among input features is one the prominent characteristics of SSDAE [49]. To this end, in this paper, the SSDAE was employed for recovering whole data at the first step (after data preparation from our hospital EPR system).

A denoising autoencoder (DAE) is simply a neural network with one hidden layer that should be trained to reconstruct a clean version of input X from a corrupted/current version of x'. It is accomplished by a so-called encoder that is a deterministic mapping from an input vector x into hidden representation y. X is the in-hospital and outside hospital datasets with variables to predict patient's non-attendance.

$$f_\theta(\mathbf{x}) = s(\mathbf{Wx} + \mathbf{b})$$

where the parameter θ is $(\mathbf{W}, \mathbf{b})$, $\mathbf{W}$ is a weight matrix indicating the weight of each of contributing variables of patients with non-attendance, b is an encoding bias vector.

In stacking SSDAE as demonstrated in Figure 1, the auto-encoder layers are placed on top of each other. Each layer is trained independently ('greedily') and then is stacked on top of previous one. In denoising autoencoders, the loss function is to minimizing the reconstruction loss between a clean X and its reconstruction from Y [50]. A decoder is then used to mapped the latent representation $\mathcal{Y}$ into a reconstructed ('repaired') vector such as $z \in [0,1]^d$

$$\mathbf{z} = \mathrm{g}_{\theta'}(\mathbf{y}) = s(\mathbf{W}' \mathbf{y} + \mathbf{b}')$$

$\mathbf{W}'$ is a decoding matrix, and $\mathbf{b}'$ is decoding bias vector; The SSDAE could have several layers. For training a SSDAE, each layer is trained on top of previous one. The training process starts with per-training the first hidden layer fed the training samples as input, training the second hidden layer with the outputs flowing from the first hidden layer, and so on. This was how autoencoders stack hierarchically to form a deep SSDAE. The parameters of the model θ and θ' are optimized during the training phase to minimize the average reconstruction error,

$$\theta, \theta^{'*} = \ arg \min_{\theta, \theta^{'*}} L(\mathbf{x}, \mathbf{z}) = \arg \min_{\theta, \theta^{'*}} \frac{1}{N} \sum_{i=1}^{N} L(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}),$$

where $L(\mathbf{x}, \mathbf{z})$ is a *loss function* and N is the number of data samples in the training set. The reconstruction cross-entropy function is usually used as *loss* as depicted in the equation below:

$$L_H(\mathbf{x}, \mathbf{z}) = -\sum_{k=1}^{d} [\mathbf{x_k} log \mathbf{z_k} + (1 - \mathbf{x_k}) \log (1 - \mathbf{z_k})]$$
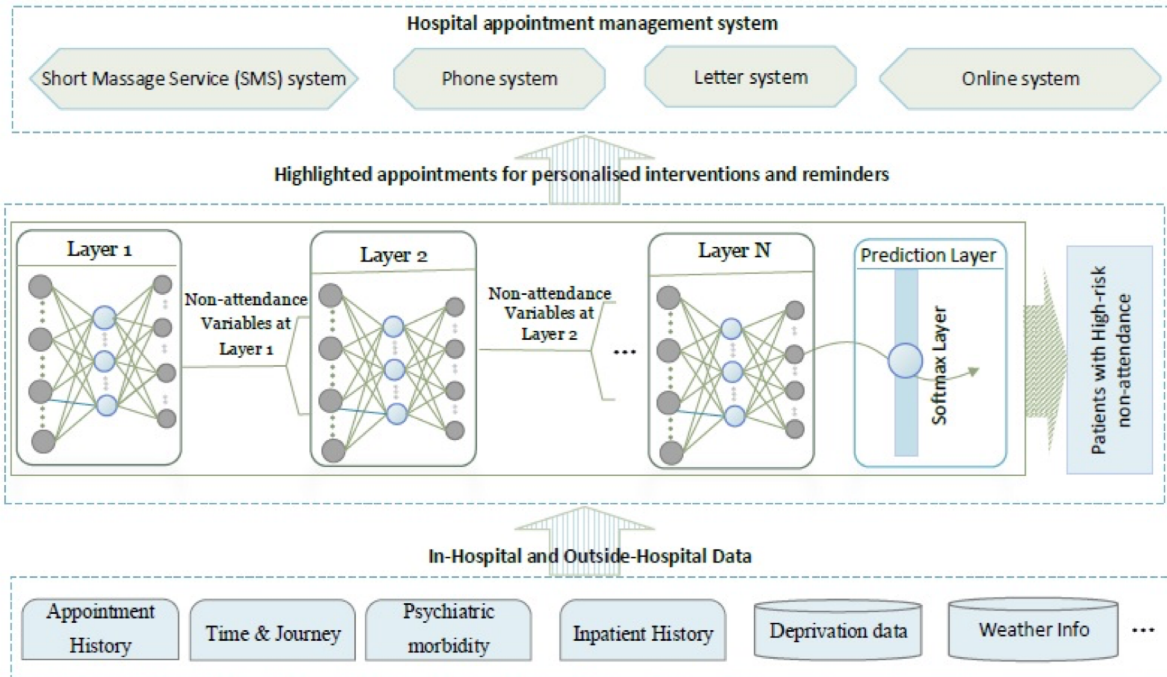
Figure 1 Non-attendance prediction model integrated with hospital appointment system

One serious issue concerning autoencoders is the size of the hidden layer that could be potentially affect the performance. If the dimensionality of the hidden unit (number of neurons) is the same as or larger than the input layer, this approach could potentially learn the identity function. it means that the model would overfit to input data instead of learning non-linear relations. Furthermore, employing larger dimensionality conduct the model to learn sparse representation of data which may result in learning more latent variables and non-linear relations. Considering merely using the denoising type may ultimately result in learning the identity function whereas, Xie et al [51] showed that sparse type of denoising autoencoders could learn other features than the denoising type. To this regard, espousing a sparsity constraint could practically solve such issues providing SSDAEs with more hidden units of larger dimensionality. The equation below depicts a sparsity constraint added to the previous equation.

$$SC = L(\mathbf{X}, \mathbf{Z}) + \gamma \sum_{h=1}^{H} KL(\rho||\hat{\rho}_j)$$

where γ denotes the weight of penalty factor, H is the number of hidden units, ρ is a sparsity parameter and is typically a small value close to zero, $\hat{\rho}_j$ is the average activation value of hidden unit j over the training set, $KL(\rho||\hat{\rho}_j)$ is the Kullback–Leibler (KL) divergence as defined bellow.

$$KL(\rho||\hat{\rho}_j) = \rho \log{}^{\rho}/_{\hat{\rho}_j} + (1 - \rho) \log[(1 - \rho)/(1 - \hat{\rho}_j)]$$

The KL is principally an asymmetric measure of the distance between two given sample distributions. It provides the sparsity constraint on the coding. For instance, if two distributions are equal (e.g., $\rho = \hat{\rho}_j$), the KL would be zero. A standard backpropagation algorithm then can be used to solve this optimization problem.

Besides data recovery and construction by non-linear transformation resulting ultimately in a compact representation, the SSDAEs could include a standard predictor to do the predictions as well. This layer could be a proper function like logistic regression, max, softmax etc. In this work, we used a softmax layer as has a proven performance in most recent application. We will not only predict binary classification but also more detailed patients' attendance behaviors including attendance, non-attendance without prior notification and non-attendance with prior notification through multi-classification as the next step future research. Furthermore, using softmax will get a probability distribution which we can apply cross entropy loss function on. This layer contains a softmax function as depicted below.

$$p_{(y=j|x)} = \frac{e^{x_j}}{\sum_{k=1}^{N} e^{x_k}}$$

where x is a N dimensional vector of real numbers from the previous hidden unit and transform it into a vector of real number in range (0,1) thus it is the output probabilities for each class. As is clear in the equation, the output is always positive numbers which has been also normalized.

In brief, the training of the model comprises of two phases. At first, the model is trained using training dataset together with its associated labels. In the former phase, we try to minimize the difference between the recovered and ground truth training dataset: X vs X̂. In the later phase, purpose is to optimize the model in terms of supervised prediction performance.

It is worth mentioning that training the model using standard backpropagation algorithms usually yields poor performance. To this end, a greedy layerwise unsupervised learning algorithm is proposed by [52] to pre-train the SSDAEs layer by layer in a bottom–up way. Just afterward, fine-tuning the model's parameters in a top–down direction is applied with backpropagation to improve the performance at the same time. The training procedures of this study briefly involves the following steps drawn from the proposed algorithms of [52], [53].

Step 1: Minimize the objective function of the first autoencoder over the input data

Step 2: Minimize the second autoencoder's objective function over the output of the previous layer

Step 3: Iterates through steps 1 and 2

Step 4: Obtain the probability of no-show patient class based on the output of the last hidden layer

Step 5: Optimize whole network with backpropagation algorithms

The first three steps are literally unsupervised as is aimed to minimize the reconstruction error; whereas in the last step, where the generated labels from the last autoencoder fed to a softmax layer, all stacked layers will be optimized using backpropagation as a whole network. The optimization is performed in a supervised way based on the respective class labels.

Moreover, it is critical to consider that the number of hidden layers could potentially leverage the performance of SSDAE. Very shallow structure of SSDAE could result in poor performance whereas a very deep structure (i.e., with many hidden unit) make the constructed model very complex and diversely effect the performance as well. We pre-evaluated some architectures and found a three-layer SSDAE works better for our application. Selecting a specific architecture by testing over whole data is highly resource-intensive and not always applicable. Some shallower (one-Layer) to deeper (5-Layer) architectures were assessed using only validation set. The shallower networks resulted in poorer performance as failed to

learn proper representation while going very deeper added just complexity than any improvement. Our empirical observation was already reported in [50] as they also found stability of results (error convergence) on the three layer architecture specially for sparse types.

4. Non-attendance risk prediction using SSDAE

4.1 Data and variables

The data source is from in-hospital data (e.g. EPR) and outside hospital data (e.g. environment and social economic data). In EPR, the information of over 150,000 outpatients spanning on around 1.6 million records was gathered. The information is unevenly distributed in 6 years beginning from 2010 and going through the early 2018. Considering the period, the more records there is in the most recent years as the EPR system is more extending. Variables selected from different tables and different database can be classified into 7 categories including demographic, appointment history, inpatient history, outpatient history, deprivation index and weather information, health status. Those variables are identified through literatures and focus groups with hospital operation teams. The inpatient database contains information about in-hospital patients. Nevertheless, we used it to take the advantages of possibly available historic health data about new-coming patients. Such historic health records contain diagnostic codes which in turn could be used to draw some very informative variables such as psychiatric variables for example. If there was a patient had inpatient records for more than once, we will only use the record where there was an overlap between inpatient period and outpatient appointment time and less than 14 days gap between discharge and outpatient date. This is from focus group that patient may choose not to attend the outpatient appointment if it is within their inpatient time or it is close to their discharge date. It should be noted that some variables are particularly conditional. For instance, length of stay (LOS) is used as an input. The LOS is non-zero if and only if the patient had an immediate inpatient record in the EPR. The zero value is used in every empty element in the resulting table.

After digitization, a normalization procedure was applied to center the data and making them in a closed range [0,1]. The normalization considerably diminishes the inverse effect of large-scale variables to hinder the network from incorporating small-scale attribute in both the neural networks and classification models [54]. Beside the input variables, the target variable that is indeed a binary event i.e., show & no-show, should be constructed. The target vector contains either zero or

one respective to the corresponding event to each row of information. It should be mentioned that each patient may have several records from which a few might be positive event type. Hence, the model should predict the patient's behavior given a specific time point and historic information.

4.2 Model training

For training the model, we use all the information before mid 2017 as training set and validation set. The remaining records were utilized as testing data for evaluating the model performance. We tried to use a natural split as the model is going to be run over the live data, the most recent data samples were used for testing the model comprising statistically around 25% of all samples. The remaining samples were divided using stratified random sampling into of 15% validation and 85% training sets. The Stratified random sampling [55] is essential to maintain the original class distribution among both subsets.

In brief, the training model is to minimize the difference between the feeding data and recovered replicate (i.e., the output of the autoencoders) while trying to build an overall high-performance classification model with backpropagation. It is noteworthy that the pre-training the SSDAE layers are literally unsupervised as no label is being used. However, the optimization process is supervised as we exploit the target vector (i.e., prepared binary labels indicating attendance vs non-attendance). Our method was implemented and evaluated with SQL Server (for fetching data, preparing tables, some cleansing, etc.), Matlab 2018a (deep learning and machine learning packages) and Jupyter Notebook. The experiments were conducted on CPU 4Ghz, RAM 32GB, Highest Speed SSD: 1TB, and VGA Card: GTX 1080TI with 11GB of RAM having over 3600 CUDA cores.

4.3 Evaluation

The evaluation phase consists of three practical stages. In the first stage, the original test data was fed into the previously trained model. The trained model will elucidate the recovered version of the feeding test data while at the same time produce a probability of no-show event. One important advantage of this model is its flexibility about missing values or incomplete information that is widespread in real-world practical application. Another advantage is the performance of the final model is highly better than a traditional model with the same time complexity. We should notice there is two complexities in terms of time involved in building practical models. Time complexity in training and the final product. This models and similar machine learning

models in comparison with some traditional models such as logistic regression produce the result in a few milliseconds without any extra efforts. We do not need to do anything else once the model is built. One critical benefit of our proposed approach is the scalability. Scalability defines in three different ways: (1) in terms of number of variables and (2) in terms of number of samples we can use and most importantly (3) in terms of updating the model over time. We could add new variables to existing model with the same practice. New variables provide a way to incorporate more information to the model resulting in more reliable model for managers.

The proposed method was applied to the test data and its performance compared with commonly-used classifiers and representative methods. The Five well-known machine learning classifiers were used first to give us some insight about complexity of prediction. Support vector machines (SVM) (with Linear kernel), k-nearest neighbours algorithm (KNN, K=3), Decision Tree (DT), Naïve Bayes, and Random Forest based on the parameter settings suggested by [56] for imbalanced high-dimensional data. The random forest classifier was used from the widely-used 'sklearn.ensemble' library in python with all the default parameter settings but the number of tree which was set to 100 instead of its default (10) .

Considering parameter settings, it is important to note that we could not feed all data into the neural network either during the training or testing phase. Data should be fed in to the model in small parts called batch. The batch size containing 64 samples was used as adopted primarily in Adam optimizer [57] and suggested by the previous work of [45]. Other parameters such as sparsity weight has not been altered from default values. Furthermore logistic regression is the only method used to predict the probability of non-attendance so far [38]. Hence, it is also included as one of the baseline classifier to compare with our non-attendance model. Table 1 demonstrates the performance of proposed method with baseline classifiers.

In our comparison, we evaluate a similar architecture used previously in healthcare [48], as well as a representative work of [1] and logistic-regression-based method that are basically belong to statistical modelling. As it is obvious in Table 1, the proposed method markedly outperforms all the conventional methods in various aspects. It could obtain a general accuracy of almost 70%. Only the Deep Patient [48] which has a similar deep learning architecture used for disease prediction comes near in performance. However, as aforementioned in the introduction, our proposed

Table 1 Performance of different prediction methods (including existing non-attendance method and representative deep learning methods used in EPR)

Method	Measures			
	AUC-ROC	Precision	Recall	Accuracy
Our Method	**0.71**	0.69	**0.78**	**0.69**
DeepPatient-SVM [48]	0.69	**0.73**	0.67	0.67
Logistic Regression non-attendance model [1]	0.54	0.60	0.52	0.51
Logistic + Bayesian non-attendance model [38]	0.49	0.61	0.45	0.46
Random Forest	0.54	0.63	0.46	0.51
SVM	0.51	0.58	0.49	0.47
KNN	0.17	0.14	0.33	0.17
Naïve Bayes	0.33	0.25	0.76	0.32
Decision Tree	0.29	0.27	0.69	0.27

architecture is more integrative than those methods having separated the prediction phase and the construction phase.

Regarding other machine learning classifiers, Decision Tree along with KNN had the lowest performance. Nonetheless, the Random Forest and SVM achieved the highest accuracy among other common methods. The RF classifier with higher running time in our experiments had slightly better performance than SVM. One reason behind that could be because the random forests usually does not require many tweaking as long as the number of estimators (the trees) is large enough. we selected the model with 100 trees which outperforms other baseline methods.

Furthermore, it is worth noting these baseline methods were utilized mostly without fine-tuning hyper parameter settings since it is highly time-consuming considering various settings and improvement steps. So, one reason of such significant different in performance of these method is because of that fact. The other reason is the complexity of hypothesis space which influence directly the performance of classifiers. For instance, decision tree which is expected usually to have a better performance particularly on high-dimensional unbalanced datasets [58], [59] obtained poorer accuracy than SVM and Naïve Bayes (although it was run with automatic hyper-parameter optimization in Matlab toolset). One reason as aforementioned is the tweaking issues and the other could be because of conflicting rules in the decision surface. In such cases, a proper rule induction method like [60] could significantly help out to build a better model that is out of this paper scope. Also, considering the AUC area of those models that are under 0.5 -which actually perform poorer than a random classifier- it is essential to note about the possibility of

having such models. Assume a case when a model learnt wrong patterns thereby generating wrong class labels a. Such issues could be because of the fix improper data-split, overfitting over training sets, missing values, class-imbalance, and other matters.

The proposed method significantly outperformed these methods in terms of various evaluation metrics. Nevertheless, the fine-tuning procedures and dealing with several free parameters are quite challenging. Perhaps in future with advancing AI technology, we would see high-scale self-adaptable algorithms. In other viewpoint, more relevant data and higher quality improves the performance of all current models. We believe the current trends for developing health-care systems in UK follow a growing consistent strategy to reduce operational costs, reduce clinical costs, and improve clinical outcomes at the same time. adopting such intelligent algorithms in health-care application with high-scale dimension could potentially contributes to this process.

5. Conclusions

In this study, we represented a novel non-attendance prediction method incorporating wide spectrum of factors relating to health, social economics and environment for improved understanding and prediction of patient behaviors. Our approach is applicable upon hospital big data from EPR systems. The proposed approach is an end-to-end deep learning model which adopted the latest architecture of sparse stacked denoising autoencoders (SSDAEs). The SSDAEs were used both for data reconstruction and classification. As for reconstruction, the stacked autoencoders were exploited to deal with the missing

values for recovering them and provide a dense representation. In prediction phase, a softmax layer that has been used in modern deep learning models was added to the network. This layer produced probability of non-attendance event based on the outputs of the last hidden unit in SSDAE.

Practically developing the proposed model required three main phases. First was the data preparation that was gathering and combining various variables from different data tables and databases in the EPR system. Collating data itself is not enough. Hence, digitization and normalization on the whole data were performed to obtain a proper input for the model. The data was separated into testing, training and evaluation sets. A target vector notating the non-attendance events containing wither zero or one was created for supervised training the model afterwards. The model was trained on the training samples and evaluated upon the testing samples. The performance of the model over the test set was compared with other classification models including logistic regression. The experiments illustrated that the proposed model significantly outperformed other models in terms of important evaluation metrics including AUC-ROC, Precision, Recall, and Accuracy.

The constructed model was finally deployed on current infrastructure to being connected to a reminder system. The limitation of this research is that only "specialty" covers clinical related information on consultant's skill and expertise. More detailed clinical data such as diagnosis, treatment specialty, attendance type (main procedure) will be included in our future work for better results. We will involve wider features from hospital database and use methods such as Recursive Feature Elimination (RFE) for feature selection to improve accuracy. Further research will also be conducted to predict more detailed patients' attendance behaviors including attendance, non-attendance without prior notification and non-attendance with prior notification through multi-classification.

6. Acknowledgement

7. References

[1] D. A. Ellis, R. McQueenie, A. McConnachie, P. Wilson, and A. E. Williamson, "Demographic and practice factors predicting repeated non-attendance in primary care: a national retrospective cohort analysis," *Lancet Public Heal.*, vol. 2, no. 12, pp. e551–e559, 2017.

[2] Department of Health. "A zero cost way to reduce missed hospital appointments.", 2016

[3] NHS England, "Quarterly Hospital Activity Data.", 2016

[4] National Audit Service, "NHS waiting times for elective care in England." 42, 2014

[5] A. George and G. Rubin, "Non-attendance in general practice: a systematic review and its implications for access to primary health care," *Fam. Pract.*, vol. 20, no. 2, pp. 178–184, 2003.

[6] K. Campbell, A. Millard, G. McCartney, and S. McCullough, "Who is least likely to attend? An analysis of outpatient appointment 'did not attend'(DNA) data in Scotland." Edinburgh: NHS Health Scotland, 2015.

[7] J. Clark, "Appointment cancellation options-a new system to help decrease no-show appointments," *IACH Inf.*, vol. 3, 2006.

[8] C. G. Moore, P. Wilson-Witherspoon, and J. C. Probst, "Time and money: effects of no-shows at a family practice residency clinic," *Fam. Med. City-*, vol. 33, no. 7, pp. 522–527, 2001.

[9] A. E. Williamson, D. A. Ellis, P. Wilson, R. McQueenie, and A. McConnachie, "Understanding repeated non-attendance in health services: a pilot analysis of administrative data and full study protocol for a national retrospective cohort," *BMJ Open*, vol. 7, no. 2, Feb. 2017.

[10] U.-G. Gerdtham and M. Johannesson, "The relationship between happiness, health, and socio-economic factors: results based on Swedish microdata," *J. Socio. Econ.*, vol. 30, no. 6, pp. 553–557, 2001.

[11] R. Bush, V. Vemulakonda, S. Corbett, and G. Chiang, "Can We Predict A National Profile Of Non-Attendance Pediatric Urology Patients: A Multi-Institutional Electronic Health Record Study," *Inform. Prim. Care*, vol. 21, no. 3, p. 132, 2014.

[12] R. Sabit *et al.*, "Predictors of poor attendance at an outpatient pulmonary rehabilitation programme," *Respir. Med.*, vol. 102, no. 6, pp. 819–824, 2008.

[13] C. Hayton *et al.*, "Barriers to pulmonary rehabilitation: characteristics that predict patient attendance and adherence," *Respir. Med.*, vol. 107, no. 3, pp. 401–407, 2013.

[14] A. J. Mitchell and T. Selmes, "A comparative survey of missed initial and follow-up appointments to psychiatric specialties in the United Kingdom," *Psychiatr. Serv.*, vol. 58, no. 6, pp. 868–871, 2007.

[15] A. J. Mitchell and T. Selmes, "Why don't patients attend their appointments? Maintaining engagement with psychiatric services," *Adv. Psychiatr. Treat.*, vol. 13, no. 6, pp. 423–434, 2007.

[16] H. Killaspy, S. Banerjee, M. King, and M. Lloyd, "Prospective controlled study of psychiatric out-patient non-attendance: Characteristics and outcome," *Br. J. Psychiatry*, vol. 176, no. 2, pp. 160–165, 2000.

[17] S. L. Catz, J. B. McClure, G. N. Jones, and P. J. Brantley, "Predictors of outpatient medical appointment attendance among persons with HIV," *AIDS Care*, vol. 11, no. 3, pp. 361–373, 1999.

[18] D. Giunta, A. Briatore, A. Baum, D. Luna, G. Waisman, and F. G. B. de Quiros, "Factors associated with nonattendance at clinical medicine scheduled outpatient appointments in a university general hospital," *Patient Prefer. Adherence*, vol. 7, pp. 1163–1170, 2013.

[19] P. Kheirkhah, Q. Feng, L. M. Travis, S. Tavakoli-Tabasi, and A. Sharafkhaneh, "Prevalence, predictors and economic consequences of no-shows," *BMC Health Serv. Res.*, vol. 16, no. 1, p. 13, 2016.

[20] E. Harte *et al.*, "Reasons why people do not attend NHS Health Checks: a systematic review and qualitative synthesis," *Br J Gen Pr.*, vol. 68, no. 666, pp. e28–e35, 2018.

[21] V. L. Lawson, P. A. Lyne, J. N. Harvey, and C. E. Bundy, "Understanding why people with type 1 diabetes do not attend for specialist advice: A qualitative analysis of the views of people with insulin-dependent diabetes who do not attend diabetes clinic," *J. Health Psychol.*, vol. 10, no. 3, pp. 409–423, 2005.

[22] C. Martin, T. Perfect, and G. Mantle, "Non-attendance in primary care: The views of patients and practices on its causes, impact and solutions," *Fam. Pract.*, vol. 22, no. 6, pp. 638–643, 2005.

[23] M. Husain-Gambles, R. D. Neal, O. Dempsey, D. A. Lawlor, and J. Hodgson, "Missed appointments in primary care: questionnaire and focus group study of health professionals.," *Br J Gen Pr.*, vol. 54, no. 499, pp. 108–113, 2004.

[24] K. E. Lasser, I. L. Mintzer, A. Lambert, H. Cabral, and D. H. Bor, "Missed appointment rates in primary care: the importance of site of care," *J. Health Care Poor Underserved*, vol. 16, no. 3, pp. 475–486, 2005.

[25] V. Chariatte, P. Michaud, A. Berchtold, C. Akré, and J. Suris, "Missed appointments in an adolescent outpatient clinic: descriptive analyses of consultations over eight years," *Swiss Med. Wkly.*, vol. 137, no. 47/48, p. 677, 2007.

[26] W. Franks, N. Gawn, and G. Bowden, "Barriers to access to mental health services for migrant workers, refugees and asylum seekers," *J. Public Ment. Health*, vol. 6, no. 1, pp. 33–41, 2007.

[27] F. M. Burns, J. Y. Imrie, J. Nazroo, A. M. Johnson, and K. A. Fenton, "Why the (y) wait? Key informant understandings of factors contributing to late presentation and poor utilization of HIV health and social care services by African migrants in Britain," *AIDS Care*, vol. 19, no. 1, pp. 102–108, 2007.

[28] M. D. Woods *et al.*, "Vulnerable groups and access to health care: a critical interpretive review," *Natl. Coord. Cent. NHS Serv. Deliv. Organ RD Retrieved May*, vol. 27, p. 2012, 2005.

[29] M. Mackenzie, E. Conway, A. Hastings, M. Munro, and C. O'Donnell, "Is 'candidacy'a useful concept for understanding journeys through public services? A critical interpretive literature synthesis," *Soc. Policy Adm.*, vol. 47, no. 7, pp. 806–825, 2013.

[30] A. Hirst, J. Delvaux, S. Rinne, C. Short, and A. McGregor, "Multiple and complex needs initiative: programme evaluation report," 2009.

[31] R. D. Neal, M. Hussain-Gambles, V. L. Allgar, D. A. Lawlor, and O. Dempsey, "Reasons for and consequences of missed appointments in general practice in the UK: questionnaire survey and prospective review of medical records," *BMC Fam. Pract.*, vol. 6, no. 1, p. 47, 2005.

[32] T. N. O. Lehmann, A. Aebi, D. Lehmann, M. B. Olivet, and H. Stalder, "Missed appointments at a Swiss university outpatient clinic," *Public Health*, vol. 121, no. 10, pp. 790–799, 2007.

[33] K. M. Nielsen, O. Faergeman, A. Foldspang, and M. L. Larsen, "Cardiac rehabilitation: health characteristics and socio-economic status among those who do not attend," *Eur. J. Public Health*, vol. 18, no. 5, pp. 479–483, 2008.

[34] S. B. Cashman, J. A. Savageau, C. A. Lemay, and W. Ferguson, "Patient health status and appointment keeping in an urban community health center," *J. Health Care Poor Underserved*, vol. 15, no. 3, pp. 474–488, 2004.

[35] Y. Masuda *et al.*, "Personal features and dropout from diabetic care," *Environ. Health Prev. Med.*, vol. 11, no. 3, pp. 115–119, 2006.

[36] J. Waller and P. Hodgkin, "Defaulters in general practice: Who are they and what can be done about them?," *Fam. Pract.*, vol. 17, no. 3, pp. 252–253, 2000.

[37] A. Murdock, C. Rodgers, H. Lindsay, and T. C. K. Tham, "Why do patients not keep their appointments?

Prospective study in a gastroenterology outpatient clinic," *J. R. Soc. Med.*, vol. 95, no. 6, pp. 284–286, 2002.

[38] A. Alaeddini, K. Yang, C. Reddy, and S. Yu, "A probabilistic model for predicting the probability of no-show in hospital appointments," *Health Care Manag. Sci.*, vol. 14, no. 2, pp. 146–157, 2011.

[39] J. Wu, J. Roy, and W. F. Stewart, "Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches," *Med. Care*, vol. 48, no. 6, pp. S106–S113, 2010.

[40] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Heal. Inf. Sci. Syst.*, vol. 2, no. 1, p. 3, 2014.

[41] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," in *Machine Learning for Healthcare Conference*, 2016, pp. 301–318.

[42] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Deepcare: A deep dynamic memory model for predictive medicine," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2016, pp. 30–41.

[43] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "Deepr: A Convolutional Net for Medical Records," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 22–30, 2017.

[44] H. Wang, Z. Cui, Y. Chen, M. Avidan, A. Ben Abdallah, and A. Kronzer, "Cost-sensitive Deep Learning for Early Readmission Prediction at A Major Hospital," *Canada Proc. BIOKDD*, no. 17, 2017.

[45] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, and E. Liu, "Predicting all-cause risk of 30-day hospital readmission using artificial neural networks," *PLoS One*, vol. 12, no. 7, 2017.

[46] C. Xiao, T. Ma, A. B. Dieng, D. M. Blei, and F. Wang, "Readmission prediction via deep contextual embedding of clinical concepts," *PLoS One*, vol. 13, no. 4, 2018.

[47] A. Rajkomar, E. Oren, K. Chen, A. Dai, … N. H. preprint arXiv, and U. 2018, "Scalable and accurate deep learning for electronic health records," *arxiv.org*.

[48] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, p. 26094, 2016.

[49] H.-I. Suk, S.-W. Lee, D. Shen, and A. D. N. Initiative, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Funct.*, vol. 220, no. 2, pp. 841–859, 2015.

[50] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[51] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in neural information processing systems*, 2012, pp. 341–349.

[52] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[53] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in neural information processing systems*, 2007, pp. 153–160.

[54] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[55] A. I. Marqués, V. García, and J. S. Sánchez, "On the suitability of resampling techniques for the class imbalance problem in credit scoring," *J. Oper. Res. Soc.*, vol. 64, no. 7, pp. 1060–1070, 2013.

[56] M. Dashtban and M. Balafar, "Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts," *Genomics*, vol. 109, no. 2, pp. 91–107, 2017.

[57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.

[58] L. Rokach and O. Z. Maimon, *Data mining with decision trees: theory and applications*, vol. 69. World scientific, 2008.

[59] M. Dashtban, M. Balafar, and P. Suravajhala, "Gene selection for tumor classification using a novel bio-inspired multi-objective approach," *Genomics*, vol. 110, no. 1, pp. 10–17, 2018.

[60] G. M. Weiss and F. Provost, "Learning when training data are costly: The effect of class distribution on tree induction," *J. Artif. Intell. Res.*, vol. 19, pp. 315–354, 2003.