Predicting Patient's Choices of Hospital Levels Using Deep Learning and Representation Improvements

Lichin Chen[*], Ji-Tian Sheu[†], Yuh-Jue Chuang[††]
[*]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
E-mail: li.chin@citi.sinica.edu.tw Tel: +886-2-27872391
[†]Department of Health Care Management, Chang Gung University, Taoyuan, Taiwan
E-mail: jtsheu@mail.cgu.edu.tw Tel: +886-3-2118800 #5662
[††]Department of Health Care Management, Chang Gung University, Taoyuan, Taiwan
E-mail: chuangyj@mail.cgu.edu.tw Tel: +886-3-2118800 #3420

Abstract— **In countries that enable patients to choose their own healthcare providers, a common problem is that the patients do not go to unsuitable hospital levels. This might cause problems such as overwhelming tertiary facilities with mild condition patients, and resulting in limited the treatment for acute and critical patients. Our aim is to predict patients' choices of hospital levels to support the evaluation during policy making. We proposed a deep neural network (DNN) framework, which involves an improvement of the representation for insurance data, a DNN design to make accurate predictions, and a model interpretation to further understand the decision of the model. This study used the 5-year nationwide insurance data of Taiwan. With the combination of autoencoder (AE) and DNN, the prediction results achieved an accuracy of 0.94, area under the receiver operating characteristics curve (AUC) of 0.88, sensitivity of 0.93, and specificity of 0.97 with highly imbalanced data. The result shows that changing data representation had a positive effect on the prediction model. The model interpretation results show that past medical experiences and recommendations of acquaintances are most important. Deep learning technology acts as a feasible tool that provides additional evaluation besides using traditional statistical methods.**

I. INTRODUCTION

While ensuring accessibility to care, some countries offer patients the freedom to choose health providers. Patients are to "vote with their feet" and choose healthcare providers who fit their preferences and needs [1]. This strategy empowers patients by prompting healthcare providers to compete for patients through a customer-market mechanism, improving care quality, efficiency, and wait time [2, 3]. However, such expectations have preconditions. Patients are expected to make their choices based on sufficient information and rational decision making [2, 3], which is commonly not the case. Studies have shown that patients have an inadequate ability to use comparative information during provider selection [2, 4]. Past research indicates that a patient's choice involves a complex interrelationship among the characteristics of the patients, the healthcare providers, and the incident itself [2]. The decision may differ based on the characteristics of individuals, the characteristics of the provider, and the conditions of the incident, which makes it difficult to evaluate in advance. In addition, due to the limitations of the study design, the factors are commonly defined only through pre-limited patient groups or regions. During healthcare policy making, full evaluation of the patients' choices, patient volume, and patient flow are made in order to make adequate decisions. However, in most cases, policy makers were left with insufficient information during planning for policymaking.

Researchers have developed several techniques to forecast patient choices. The gravity model, for instance, calculates the spatial interaction between a community and a hospital using the population mass of the community, capacity and service mix of hospitals, and distances (or traveling time) [5, 6]. The aggregate hospital choice model intended to model hospital choices using time-series techniques to forecast future patient volumes [7]. Forecast and simulation techniques, such as mean absolute percentage errors, autoregressive integrated moving average (ARIMA), seasonal ARIMA [8], and discrete event simulation models [9, 10], have also been used previously. However, these theories have strict preconditions and were designed to explain a particular part of the choice scenario. Recently, deep learning methods have gained popularity. They capture the underlying pattern of data, transforming them into more abstract matter, and classifying them based on the latent distribution [11, 12]. The deep learning approach has been proven effective and has shown excellent performance in a wide variety of applications [11, 13, 14], such as disease risk forecasting [15], vital sign classification into physiological symptoms [16, 17], image classification for diagnosis [18, 19], text-based medical condition recognition [20, 21], and clinical event forecasting [22].

Irrespective of the outstanding performance, there are still problems when using deep learning in medical fields. Structured data, such as insurance data and electronic medical records (EMRs), are originally documents on how patients are treated and the measurement of their outcomes. Such information, including diagnoses, medications, laboratory tests, and procedures, is often presented with different encodings and terminologies, which are meaningful to the human eye.

However, previous studies indicate that the existing hierarchical coding scheme of structured data is sparse, noisy, and repetitive. It does not quantify the inherited similarity between concepts and is not sufficiently representative for data modeling [23, 24]. Meanwhile, the features that enter the training process are traditionally handcrafted by experts based on domain knowledge, which requires a large amount of human effort. Another common problem is the black-box problem of deep learning technology. It merely provides the prediction results without reasons or information that indicates how the conclusion has been reached. This limited the implementation of deep learning method in the healthcare field.

The insufficient application in assisting public health has been highlighted due to the outbreak of the coronavirus (COVID-19). This research aims to use deep neural networks (DNNs) to predict patients' choices for the purpose of policy evaluation, which acts as an alternative method besides traditional statistics. Also, we attempt to address the common problems mentioned above. The proposed framework involves exploring the representation of insurance data, neural network design to make accurate predictions, and how the model makes its decision using model interpretation methods.

II. Methods

The aim of this study was to predict patients' choices of hospital level. The framework of this research consists of three sections: (1) Features were assembled into a vector that represented an incident of a hospital visit. An autoencoder (AE) was used to change the data representation. (2) DNNs were used to train the prediction model. A comparison was performed to show the best design. (3) The prediction model was further interpreted using model interpretation methods and showed the feature weights that the model made its decision on.

The data used were the insurance claims from two million clinical declaration files and the Registry for Beneficiaries files from the Taiwan National Health Insurance Research Database, dated from January 1, 2008, to December 31, 2011. The data were originally sampled to ensure their representation of the population across Taiwan. The files included the demographic information and visiting records of outpatients and emergency settings. Some publicly announced data were added to enrich the records, including "physician density" information that referred to the number of practicing physicians serving per 10,000 people in each region of Taiwan [25]. The national calendar was used to retrieve information on weekends and national holidays, and the center latitude and longitude of each district area were used to approach the distance of travel for each visit. The targeted prediction outcome of this research was the four hospital levels, namely the medical center, regional hospital, district hospital, and clinic. This research was approved by the Research Ethics Committee at National Taiwan University (No. 202004EM035), which waived informed patient consent for the data, which were already deidentified before analysis.

A. Feature Accumulation and Data Representation

Table I. Features extracted from the visiting records

Entity	Characteristics of the entity (Features)
Patients	Age, gender, low income (Yes/No), total number of visits, total number of diseases, total number of chronic diseases, usual provider of care (UPC) and least usual provider of care (LUPC), sequential continuity of care index (SECOC), and continuity of care index (COCI).
Healthcare Providers	Practicing physicians serving per 10,000 people (physician density), the most frequent provider continuity (MFPC), and the least frequent provider continuity (LFPC).
Incidents	Whether a surgery was involved (Yes/No), whether it was an emergency service (Yes/No), whether it was considered as a severe condition (Yes/No), whether the visit day was a workday (Yes/No), the disease importance rate (DIR) of the target disease during that visit, and distance of travel.

We reviewed past literature [2, 4, 26-34] and accumulated 19 features that were previously defined to affect patient choices while accessing healthcare, which could be categorized as characteristics of the patients, healthcare providers, and incidents, as summarized in Table I. The calculation of the feature was introduced in our previous work [35]. After retrieving the features, the data were assembled into a vector to represent an incident of a hospital visit.

To deal with numerical features with different scale levels, all the numerical values were normalized between 0 and 1, and the categorical features were transformed into a one-hot/dummy encoding before model training. Those indicators that were already ratio figures (values between 0 and 1) were used accordingly (including COCI, UPC, LUPC, SECOC, and DIR). Incomplete or questionable data, such as individuals without birth dates or genders (or with two genders), records without a date, birth dates later than the visit date, patients without any visiting records, patients without a primary diagnosis, incomplete information of visiting hospitals, and patients unable to determine their place of residence, were excluded.

To address the data representation, we used AE as a preprocessing step. AE is popular for processing scarce and noisy data [23, 24, 36]. It encoded the input into a lower-dimension space Z with an encoder through deterministic mapping:

$$z = s(W + b), \tag{1}$$

where $s(\cdot)$ is a nonlinear activation function that transforms the output of the neural network, which makes it easier for the model to generalize or adapt to a variety of data. The latent representation z is then reconstructed by a decoder to generate $\tilde{x}$, as follows:

$$\tilde{x} = s(W' + b'), \tag{2}$$

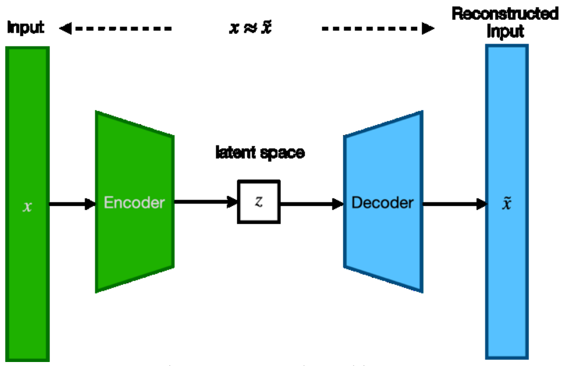

Fig. 1 Autoencoder architecture

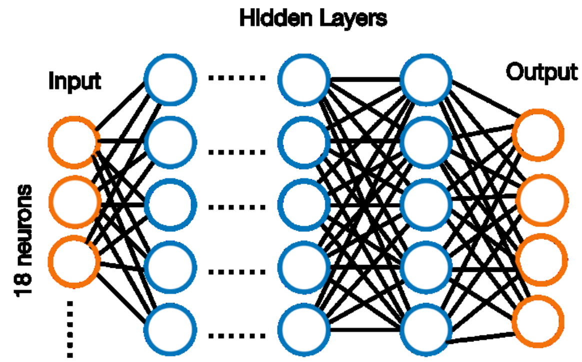

Fig. 2 Deep neural network (DNN) architecture

where the mean square error of x and $\tilde{x}$ reaches the minimization. W and W' denote the respective weights and b and b' denote the respective bias of the encoder and the decoder. The graphic architecture of AE is shown in Fig. 1. In our design, the AE consisted of five hidden layers; the neurons in those layers were 500, 250, 100, 250, and 500. The rectified linear unit (ReLU) activation function was used for the encoder (first two layers) and the decoder (last two layers) and a sigmoid activation function for the latent space conversion (the third layer). An example is demonstrated to show how AE affects the data.

B. Prediction Model Design and Training Strategy

This study used DNNs to train patients' choices of hospital levels. DNN is a complex version of an artificial neural network (ANN) that contains multiple hidden layers [37, 38], where every neuron in layer i is fully connected to every other neuron in layer $i+1$. In a multi-layer neural network, each layer of the network is trained to produce a higher level of representation of the observed pattern. Every layer produces a representation of the input pattern that is more abstract than the previous layer by composing more nonlinear operations [11, 39]. The computation of DNN is as follows:

$$\hat{y} = \sigma(\textstyle\sum_{j=1}^{d} x_j w_{ij} + b_{ij}), \tag{3}$$

where each hidden layer computes a weighted w_{ij} and bias b_{ij} of the output from the previous layer, followed by a nonlinear activation function σ that calculates the sum as outputs. The number of units in the previous layer is represented by d and the output of the previous layer by x_j. Fig. 2 demonstrates the graphic architecture of the DNN. In our design, the proposed DNN model contained 19 input nodes (based on input features) and three hidden layers with 100 neurons each. ReLU activation functions were used for each layer. Four output nodes symbolized the four hospital levels. Optimization was carried out by iterating through small subsets of the training data and modifying the parameters to minimize the loss between the prediction and the prediction target. Owing to the imbalanced distribution of hospital levels, the proposed design used a random under-sampling strategy to sample the majority label and balance the training set. The model was trained on balanced data and tested on actual distributed data [40].

After under-sampling, the data were then randomly split 80% into training data and 20% into testing data. A 5-fold cross-validation training strategy was used. The performance indicators used are the receiver operating characteristics (ROC) curve, area under the ROC curve (AUC), accuracy, sensitivity, specificity, precision, and F1 score, shown as (4)-(8), where true positive (TP) indicates the positive case that is correctly predicted as positive, true negative (TN) is the negative case that is correctly predicted as negative, false positive (FP) is the negative case that is falsely predicted as positive, and false negative (FN) is the positive case that is falsely predicted as negative.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{F1} = \frac{2 * (\text{Precision} * \text{Sensitivity})}{\text{Precision} + \text{Sensitivity}} \tag{8}$$

For a multi-class classification of hospital levels, the macro-average was used to generalize the performance index, which computed the metric independently for each class and then took the average to consider each class equally. The AUC used the one-vs-rest scheme to demonstrate the general performance. A comparison was designed to demonstrate the effect of changing representation. We compared the result of using AE processing and not using.

C. Model Interpretation Methods

This study adopted Shapley additive explanations (SHAP) [41] to interpret the prediction model [42, 43]. SHAP combines the desirable characteristics of other interpretation frameworks, including local interpretable model-agnostic explanations and deep learning important features. It interprets the models globally and locally to show the general effects of features on the whole population and individual cases. The mechanism uses all combinations of the input and the average marginal contribution of a feature value over all possible coalitions and summarizes with a feature weight, known as the SHAP value, to represent it. We interpret the DNN model and the AE+DNN model with SHAP. The mean SHAP values of the two models were demonstrated, distinguishing between positive and negative effects to show its' effect on the prediction result. This study was implemented with Python version 3.7.6, combined with PyTorch framework 1.1.0 and scikit-learn 0.22.2.

III. Results

A total of 451,317 patients and 8,039,135 visiting records were analyzed, of which 72.48% of the patients chose to go to clinics, 8.32% to the district hospital, 10.91% to the regional hospital, and 8.29% to the medical center. Tables II to IV demonstrate the demographic information of patients, healthcare providers, and incidents. The performance results are listed in Table V. AE+DNN performed better than DNN along, with AUC and accuracy reaching 0.884 and 0.937, respectively, while retaining high sensitivity and specificity (0.933 and 0.972, respectively). Observing from the difference between DNN and AE+DNN, most indicators improved when data were preprocessed with AE (beside sensitivity), which indicates that changing the data representations led to a general increase in performance. Table VI provides an example of visiting vectors before and after AE processing.

Table II. Demographic information of patients

Demographic information of patients (n = 4,511,746)	
Age, mean (SD)	49.310 (17.452)
Male, number of patients (%)	3,492,256 (43.441)
Noted low income, number of patients (%)	260,806 (3.244)
Total number of diseases, mean (SD)	17.845 (10.462)
Total number of chronic diseases, mean (SD)	17.681 (10.363)
Total number of visits per patient, mean (SD)	32.382 (24.820)
UPC, mean (SD)	0.479 (0.200)
LUPC, mean (SD)	0.084 (0.148)
COCI, mean (SD)	0.026 (0.059)
SECOC, mean (SD)	0.419 (0.217)

n = number of patients, SD = standard deviation.

Table III. Information of hospitals

Information of hospitals (n = 19,465)		
Hospital levels	Medical center, n (%)	20 (0.103)
	Regional hospital, n (%)	77 (0.396)
	District hospital, n (%)	372 (1.911)
	Clinic, n (%)	18,996 (97.591)
MFPC, mean (SD)		45.005 (168.059)
LFPC, mean (SD)		44.836 (121.646)
Physician density, mean (SD)		24.453 (21.323)

n = number of medical institutes, SD = standard deviation.

Table IV. Information of incidents

Information of incidents (n = 8,039,135)		
Hospital levels	Medical center, n (%)	666,084 (8.286)
	Regional hospital, n (%)	877,450 (10.915)
	District hospital, n (%)	668,960 (8.321)
	Clinic, n (%)	5,826,641 (72.478)
Is surgery, n (%)		198,337 (2.467)
Is ER, n (%)		149,836 (1.864)
Is severe, n (%)		272,843 (3.394)
DIR, mean (SD)		0.200 (0.185)
Workday, n (%)		6,731,605 (83.735)
Distance of travel (kilometers), mean (SD)		9.048 (30.997)

n = number of visiting records, SD = standard deviation.

Table V. Mean value of the performance indicators

	DNN	AE+DNN
AUC	0.853	0.884
Accuracy	0.910	0.937
F1 Score	0.887	0.908
Precision	0.864	0.891
Sensitivity	0.937	0.933
Specificity	0.968	0.972

DNN = Deep Neural Network. AE = autoencoder.

Table VI. An example of visiting vectors before and after AE processing

Before AE processing	After AE processing
[0.0000, 0.3738, 0.0000, 0.0833, 0.0769, 0.0219, 0.0094, 0.1429, 0.0714, 0.1538, 0.0019, 0.0077, 0.0455, 0.0000, 0.0000, 0.0000, 1.0000, 0.0714, 0.0000]	[0.00019300545741718356, 0.3745293414040272, 0.00007422724173161786, 0.08048962926685074, 0.0744074511636057, 0.017967434432186097, 0.009203413747149655, 0.14501758420491442, 0.07496150385155981, 0.15744111536258504, 0.003347708190108399, 0.007844166561940522, 0.04419625006621006, -0.000787111041561946, -0.00032199830233602665, 0.0015887867172686004, 0.9999588522411882, 0.07549837351533775, -0.00004184758831612956]

AE = autoencoder

The mean SHAP values of DNN and AE+DNN are shown in Fig. 4. MFPC, LFPC, and physician density are listed as the top three features for both models. The SHAP values of other features, even adding them together, showed a minor contribution. The AE process changed the weight ranking, shifting the physician density from the first to the third position and shifting MFPC from the third to the first position.

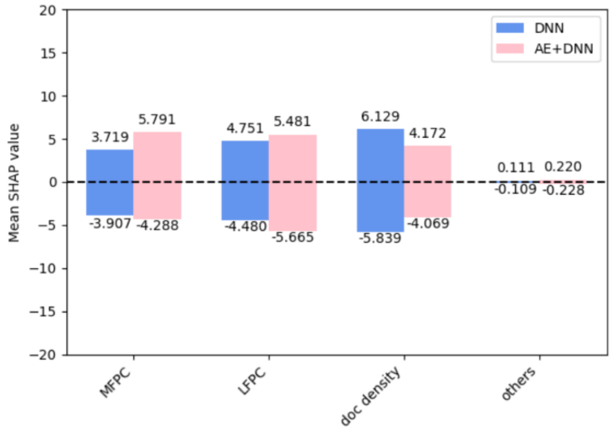

Fig. 4 Shapley additive explanations (SHAP) values of both models

IV. Discussion

Exploring patients' choices of hospital levels is the cornerstone of evaluating policies for referrals, the utilization of healthcare resources, rerouting patient volume for the hierarchization of services, and during disease outbreaks or infection control. Using deep learning technology provides an opportunity to evaluate patients' choices at hospital levels in general, which are affected by complex combinations of features and conditions. Meanwhile, the choices of patients were highly imbalanced. Not only were the choices of hospital levels maldistributed, when comparing to the general public, some conditions were not commonly seen, such as surgery (2.467%), emergency service (1.864%), or severe condition (3.394%). This is a common status for studies that focus on specific patient groups or circumstances and become scarce when applied to the general public. Despite the imbalanced distribution of the data, the deep learning model could still accurately predict with high AUC, sensitivity, and specificity and did not go through the effort of features selection.

The aim of this research was to provide an alternative way to evaluate patients' choices besides using traditional statistics. In addition to making accurate predictions, this study tried to achieve the idea of not requiring further interference after data input with representation changing. The results showed that the performance of the models can be improved by changing the data representation of the insurance data using unsupervised AE. It is straightforward that images and audio signals include disturbance and noise, and using methods to eliminate them leads to better predictions. Structured data are commonly encoded with existing encoding schemes that are meaningful to people. However, the sparsity, discreteness, and scarcity of data, which are invisible to the human eye, are difficult to notice. It can be observed from the example before and after AE processing that AE processing may be meaningful for data modeling; however, it is difficult to understand for the human eye. This study shows that changing data representations leads to improvements in the prediction results. Using features based on the underlying data pattern instead of existing knowledge may be another way to improve the prediction.

This study also tried to address the black-box problem [13, 43] using the SHAP method. According to our results, three features could interpret the majority of patients' choices of hospital levels: MFPC, LFPC, and physician density. The MFPC and LFPC represent the patients' experiences and recommendations for each institute [31, 32]. The definition allows each patient to vote for one MFPC and one LFPC, referring to their most or least visited provider. When patients were satisfied with and loyal to a specific provider, they were less likely to change healthcare providers [27]. Reports showed that patients commonly neglected the quality indicators and preferred recommendations from associates [2]. Some patients decided based on review websites [1, 44], which is another form of social approval and recommendations from others. Even professionals such as GPs attempted to decide referring to hospitals based on feedback from patients, colleagues, and their own cooperation experiences, rather than official information such as the quality of services or wait time [4]. Physician density may imply maldistributed healthcare resources across regions. A distinguishing characteristic of SHAP is that it provides an interpretation of the prediction model globally and locally. The result demonstrates the general behaviors of the people in Taiwan, which act as a reference for future policy designing.

The discipline of social economics mostly focuses on clarifying the causality and interrelationship of factors that affect patients in choosing hospitals. However, the machine learning approach attempted to seek the underlying pattern of data and accurately predict without relying on existing knowledge. The results indicated a certain underlying trend but not necessarily complete reasons and causalities. This study provides an alternative approach to observing patients' choices. The prediction was based on the trajectory of the deidentified patient-visit data that is commonly collected by insurance companies. Hence, the model is highly achievable elsewhere, as it does not involve complex information that is difficult to collect or different medical information systems in different countries. Neither was it violating patient privacy. However, because of the nature of deidentified data, distance to travel can only be approached based on projection and assumption [45] and cannot be validated accurately, which is a limitation of this research. Nevertheless, the finding of this work is responsive to existing research, which largely increases the ability to gain trust while applying it to practical use.

V. Conclusion

Deep learning technology provides an alternative way to evaluate patients' choices besides using traditional statistical methods. It overcomes the limitation of the inability to handle complex combinations of features and conditions and imbalanced datasets, which is commonly seen while facing the problem of accessibility of care. Changing the data representation of the insurance data had a positive effect on improving the performance of the prediction model. Automatically extracting features based on the underlying data pattern instead of existing knowledge can be promising. This study also attempted to propose a framework for deep learning applications in public health consisting of representation changing to extract features automatically, a prediction model making accurate predictions, and model interpretation technology to explore how the model makes its decision. Future applications using deep learning technology are promising in healthcare policy making, and further investigation is encouraged.

ACKNOWLEDGMENT

This research was funded by the Ministry of Science and Technology of Taiwan (MOST 109-2410-H-182-006) and was approved by the Research Ethics Committee at National Taiwan University (No. 202004EM035).

REFERENCES

[1] D. King *et al.*, "Identifying quality indicators used by patients to choose secondary health care providers: a mixed methods approach," *JMIR mHealth and uHealth,* vol. 3, no. 2, p. e65, 2015.

[2] A. Victoor, D. M. Delnoij, R. D. Friele, and J. J. Rademakers, "Determinants of patient choice of healthcare providers: a scoping review," *BMC health services research,* vol. 12, p. 272, 2012.

[3] A. Victoor, R. D. Friele, D. M. Delnoij, and J. J. Rademakers, "Free choice of healthcare providers in the Netherlands is both a goal in itself and a precondition: modelling the policy assumptions underlying the promotion of patient choice through documentary analysis and interviews," *BMC health services research,* vol. 12, p. 441, 2012.

[4] H. O. Birk and L. O. Henriksen, "Which factors decided general practitioners' choice of hospital on behalf of their patients in an area with free choice of public hospital? A questionnaire study," *BMC health services research,* vol. 12, no. 126, 2012.

[5] F. W. Porell and E. K. Adams, "Hospital choice models: a review and assessment of their utility for policy impact analysis," *Medical Care Research and Review,* vol. 52, no. 2, pp. 158-195, 1995.

[6] D. Fabbri and S. Robone, "The geography of hospital admission in a national health service with patient choice," *Health Economics,* vol. 19, no. 9, pp. 1029-1047, 2010.

[7] F. Kadri, F. Harrou, S. Chaabane, and C. Tahon, "Time series modelling and forecasting of emergency department overcrowding," *Journal of medical systems,* vol. 38, no. 9, p. 107, 2014.

[8] K. Kim, C. Lee, K. O'Leary, S. Rosenauer, and S. Mehrotra, "Predicting patient volumes in hospital medicine: A comparative study of different time series forecasting methods," *Northwestern University, Illinois, USA, Scientific Report,* 2014.

[9] B. Mielczarek and J. Uziałko-Mydlikowska, "Using simulation to forecast the demand for hospital emergency services at the regional level," in *Proceedings of the winter simulation conference*, 2012, pp. 1-2.

[10] N. R. Hoot *et al.*, "Forecasting emergency department crowding: a discrete event simulation," *Annals of emergency medicine,* vol. 52, no. 2, pp. 116-125, 2008.

[11] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics,* vol. 19, no. 6, pp. 1236-1246, 2018.

[12] S. Chauhan, L. Vig, M. De Filippo De Grazia, M. Corbetta, S. Ahmad, and M. Zorzi, "A Comparison of Shallow and Deep Learning Methods for Predicting Cognitive Performance of Stroke Patients From MRI Lesion Images," (in English), *Frontiers in Neuroinformatics,* Original Research vol. 13, no. 53, 2019-July-31 2019, doi: 10.3389/fninf.2019.00053.

[13] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review," *Journal of the American Medical Informatics Association,* vol. 25, no. 10, pp. 1419-1428, 2018.

[14] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nature medicine,* vol. 25, pp. 24-29, 2019.

[15] H.-H. Wang, Y.-H. Wang, C.-W. Liang, and Y.-C. Li, "Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer," *JAMA dermatology,* vol. 155, no. 11, pp. 1277-1283, 2019.

[16] M. Alhussein and G. Muhammad, "Voice pathology detection using deep learning on mobile healthcare framework," *IEEE Access,* vol. 6, pp. 41034-41041, 2018.

[17] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Computer methods and programs in biomedicine,* vol. 161, pp. 1-13, 2018.

[18] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, pp. 115-118, 2017.

[19] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature medicine,* vol. 25, pp. 954-961, 2019.

[20] S. Gehrmann *et al.*, "Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives," *PloS one,* vol. 13, no. 2, p. e0192360, 2018.

[21] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," in *Proceedings of the international conference on machine learning*, 2013, vol. 28: ACM New York, USA.

[22] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," in *Proceedings of the 1st Machine Learning for Healthcare Conference*, 2016, vol. 56, pp. 301-318.

[23] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE journal of biomedical and health informatics,* vol. 22, no. 5, pp. 1589-1604, 2018.

[24] 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," *Scientific reports,* vol. 6, no. 26094, 2016.

[25] *Statistics Yearbook of Practicing Physicians and Health Care Organizations in Taiwan*, 2018.

[26] N. Gutacker, L. Siciliani, G. Moscelli, and H. Gravelle, "Choice of hospital: Which type of quality matters?," *Journal of health economics,* vol. 50, pp. 230-246, 2016.

[27] A. Shabbir and S. A. Malik, "Measuring patients' healthcare service quality perceptions, satisfaction, and loyalty in public and private sector hospitals in Pakistan," *International Journal of Quality & Reliability Management,* vol. 33, no. 5, pp. 538-557, 2016.

[28] C. E. Pollack, P. S. Hussey, R. S. Rudin, D. S. Fox, J. Lai, and E. C. Schneider, "Measuring care continuity: a comparison of claims-based methods," *Medical care,* vol. 54, no. 5, pp. e30-e34, 2016.

[29] C.-L. Chan, H.-J. You, H.-T. Huang, and H.-W. Ting, "Using an integrated COC index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions," *BMC health services research,* vol. 12, no. 405, 2012.

[30] MCHP. "Concept: Measuring Majority of Care." Manitoba Centre for Health Policy.

[31] A. Higgins, T. Zeddies, and S. D. Pearson, "Measuring the performance of individual physicians by collecting data from multiple health plans: the results of a two-state test," *Health affairs,* vol. 30, no. 4, pp. 673-681, 2011.

[32] C. Saint-Pierre, F. Prieto, V. Herskovic, and M. Sepúlveda, "Relationship between Continuity of Care in the Multidisciplinary Treatment of Patients with Diabetes and Their Clinical Results," *Applied Sciences,* vol. 9, no. 2, p. 268, 2019.

[33] T. C. Rosenthal and C. Fox, "Access to health care for the rural elderly," *JAMA,* vol. 284, no. 16, pp. 2034-2036, 2000.

[34] C. Y. Roh and M. J. Moon, "Nearby, but not wanted? The bypassing of rural hospitals and policy implications for rural health care systems," *Policy Studies Journal,* vol. 33, no. 3, pp. 377-394, 2005.

[35] L. Chen, Y. Tsao, and J.-T. Sheu, "Using Deep Learning and Explainable Artificial Intelligence in Patients' Choices of Hospital Levels," *arXiv preprint arXiv:2006.13427,* 2020.

[36] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Interspeech*, 2013, vol. 2013, pp. 436-440.

[37] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing,* vol. 234, pp. 11-26, 2017.

[38] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE,* vol. 105, no. 12, pp. 2295-2329, 2017.

[39] A. M. Bur, M. Shew, and J. New, "Artificial intelligence for the otolaryngologist: a state of the art review," *Otolaryngology–Head and Neck Surgery,* vol. 160, no. 4, pp. 603-611, 2019.

[40] Q. Wei and R. L. Dunbrack Jr, "The role of balanced training and testing data sets for binary classifiers in bioinformatics," *PloS one,* vol. 8, no. 7, p. e67863, 2013.

[41] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, 2017, pp. 4765-4774.

[42] F. Arcadu, F. Benmansour, A. Maunz, J. Willis, Z. Haskova, and M. Prunotto, "Deep learning algorithm predicts diabetic retinopathy progression in individual patients," *NPJ digital medicine,* vol. 2, no. 92, 2019.

[43] M. V. García and J. L. Aznarte, "Shapley additive explanations for NO2 forecasting," *Ecological Informatics,* vol. 56, p. 101039, 2020.

[44] S. Li and A. Hubner, "The Impact of Web-Based Ratings on Patient Choice of a Primary Care Physician Versus a Specialist: Randomized Controlled Experiment," *Journal of medical Internet research,* vol. 21, no. 6, p. e11188, 2019.

[45] L. Min-Hau, Y. Ann-Chi, and W. Tzai-Hung, "Using Regional Differences and Demographic Characteristics to Evaluate the Principles of Estimation of the Residence of the Population in National Health Insurance Research Databases (NHIRD)," *Taiwan Journal of Public Health,* vol. 30, no. 4, 2011.