



apjtm.org



Original Article

Asian Pacific Journal of Tropical Medicine

doi: 10.4103/1995–7645.326254

5-Year Impact Factor: 2.285

Predictive modeling of 30–day readmission risk of diabetes patients by logistic regression, artificial neural network, and EasyEnsemble

Xiayu Xiang¹, Chuanyi Liu², Yanchun Zhang³, Wei Xiang^{4,5}✉, Binxing Fang⁶✉¹School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China²School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China³Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne, Australia⁴Key Laboratory of Tropical Translational Medicine of Ministry of Education, Hainan Medical University, Haikou, China⁵NHC Key Laboratory of Control of Tropical Diseases, Hainan Medical University, Haikou, China⁶Cyberspace Security Research Center, Peng Cheng Laboratory, Shenzhen, China

ABSTRACT

Objective: To determine the most influential data features and to develop machine learning approaches that best predict hospital readmissions among patients with diabetes.

Methods: In this retrospective cohort study, we surveyed patient statistics and performed feature analysis to identify the most influential data features associated with readmissions. Classification of all-cause, 30-day readmission outcomes were modeled using logistic regression, artificial neural network, and EasyEnsemble. F1 statistic, sensitivity, and positive predictive value were used to evaluate the model performance.

Results: We identified 14 most influential data features (4 numeric features and 10 categorical features) and evaluated 3 machine learning models with numerous sampling methods (oversampling, undersampling, and hybrid techniques). The deep learning model offered no improvement over traditional models (logistic regression and EasyEnsemble) for predicting readmission, whereas the other two algorithms led to much smaller differences between the training and testing datasets.

Conclusions: Machine learning approaches to record electronic health data offer a promising method for improving readmission prediction in patients with diabetes. But more work is needed to construct datasets with more clinical variables beyond the standard risk factors and to fine-tune and optimize machine learning models.

KEYWORDS: Electronic health records; Hospital readmissions; Feature analysis; Predictive models; Imbalanced learning; Diabetes

1. Introduction

Diabetes is among the most prevalent and costly chronic diseases in the world. According to the CDC 2020 National Diabetes Statistics Report[1], there are 34.2 million people diagnosed with diabetes in the United States, which accounts for 10.5% of the population. There are substantial costs associated with diabetes. In 2017, the economic burden of diabetes in the United States totaled an estimated \$327 billion. This includes \$237 billion in direct medical costs and \$90 billion in reduced productivity[2,3].

In addition, diabetes is no longer confined to developed countries or temperate zones. It has reached epidemic proportions in developing countries with tropical climates as well[4]. Tropical regions including Asia, Africa, and Latin America have more people with diabetes than other regions[5].

Immediate prospects for a cure are currently not available. Re-

Significance

We identified determinants of 30-day hospital readmission in patients with diabetes and verified models that were more suitable for 30-day readmission classification. Our findings suggest that machine learning models hold promise for integration into clinical workflow to predict readmission.

✉To whom correspondence may be addressed. E-mail: 13907533529@163.com; fangbx@bupt.edu.cn

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

©2021 Asian Pacific Journal of Tropical Medicine Produced by Wolters Kluwer-Medknow. All rights reserved.

How to cite this article: Xiang XY, Liu CY, Zhang YC, Xiang W, Fang BX. Predictive modeling of 30-day readmission risk of diabetes patients by logistic regression, artificial neural network, and EasyEnsemble. Asian Pac J Trop Med 2021; 14(9): 417-428.

Article history: Received 29 June 2021 Revision 6 August 2021

Accepted 29 August 2021 Available online 28 September 2021

hospitalizations contribute substantially to the cost and burden of diabetes care, as patients with diabetes are more likely to be hospitalized and experience readmissions. Two major challenges in the management of patients with diabetes include identifying risk factors and predicting readmissions.

Electronic health record data have the potential to be used with big data analytical approaches to predict clinical outcomes. Most previous models were developed using traditional statistical approaches, such as regression modeling. The newer alternative includes prediction models based on artificial intelligence[6]. These modern prediction algorithms are developed through machine learning techniques, and they have made the prediction of readmissions possible.

A few prior studies have identified some risk factors that influence readmissions, such as common comorbidities, multiple previous admissions, and racial and social determinants[7–9].

Despite continuous efforts, the ability of health systems to reduce hospital readmissions has been unsatisfactory. Few studies have assessed different machine learning models to predict readmissions in patients with diabetes.

Clinicians are still unable to prevent some hospital readmissions that are thought to be preventable[10]. Clinicians require more accurate methods and models to identify the highest risk patients, and to reduce the readmissions.

This study aims to identify influential factors and assess 3 machine learning models on readmission prediction in patients with diabetes. Three machine learning models and numerous data resampling methods are implemented and evaluated to predict the 30-day readmission for patients with diabetes.

2. Subjects and methods

2.1. Study population

This study used the Health Facts Database (Cerner Corporation, Kansas City, MO), a national data warehouse that collects comprehensive clinical records across hospitals throughout the United States. The Health Facts data we used were extracts representing 10 years (1999–2008) of clinical care at 130 hospitals and integrated delivery networks throughout the United States[11].

It includes 50 features relating to demographics, diagnoses, diabetes medications, number of visits in the year preceding the encounter and payer information (Blue Cross\Blue Shield, Medicare, and self-pay), which represents patient and readmission outcomes. Information was extracted from the medical database for encounters that satisfied the following criteria:

- (1) It is an inpatient encounter (a hospital admission).
- (2) It is a diabetes encounter, that is, one during which any kind of diabetes was entered into the system as a diagnosis.
- (3) The length of hospital stay was 1–14 days.

(4) Laboratory tests were performed during the encounter.

(5) Medications were administered during the encounter.

The diabetes dataset contains encounter data, admitting physician specialty, demographics (age, sex, and race), diagnoses, and in-hospital procedures of the 101 766 inpatients documented by the Clinical Classification System International Classification of Diseases (ninth revision), Clinical Modification (ICD-9-CM) Codes, laboratory data, pharmacy data, in-hospital mortality and hospital characteristics. All data were de-identified in compliance with the Health Insurance Portability and Accountability Act of 1996 before being provided to the investigators.

2.2. Preprocessing

To obtain a clean, distinct, and transformed dataset for analysis, we performed two major preprocessing steps. The preprocessing included data cleaning and feature transformations, which resulted in 69 990 distinct records and 40 features.

2.2.1. Data cleaning

The original raw dataset contained 101 766 diabetes inpatient encounters, in which every patient had at least one-day stay at the hospital with laboratory tests and medications administered.

Our initial operations were performed on the primary dataset in order to clean duplicate records. The first encounter was defined as the index admission, obtained by the smallest encounter ID for records with the same patient ID. Other encounter records from the same patient ID were excluded ($n=29\ 353$) as duplicate encounter records. Additionally, encounters that resulted in discharge disposition to either a hospice or patient death were regarded as expired records. A total of 2 423 expired records were removed to avoid bias of the final analysis. After cleaning, reduction, and manipulation of records, 69 990 records was selected. The main outcome was the 30-day, all-cause readmission to a hospital following discharge from an index admission. Hospitalization was considered a readmission if the admission date was within 30 days after discharge.

To reduce the dimensionality of the data before training the prediction models[11], we removed 10 meaningless or low-quality data fields, including 2 patient identifiers and 8 variables with little events (<4 counts) or only a single value or high percentage (>80%) of missing values. After cleaning of the data variables, the dataset contained 40 variables.

2.2.2. Feature transformations

Among the 40 variables, there are 7 columns comprising ordinal and categorical values (*e.g.*, age, payer code, medical specialty, admission type, discharge disposition, admission source and diagnosis code) in which each variable has more than 9 subcategories. To interpret the data and better fit machine learning models, we transformed the categorical data into numerical

values. For example, in terms of the age category variable, we re-grouped the original 10-year epochs into 30-year intervals [0, 30), [30, 60), and [60, 100). This is based on a previous finding[9] that when modeling the relationship of age (grouped into intervals of 10 years) and the logistic function of the readmission rate, we can notice that there are 3 perceptible intervals ([0, 30), [30, 60), and [60, 100)) where the relationship has noticeably distinct behavior. We also re-arranged the admission type, discharge disposition, and admission source into different numeric representation collections for a better indication of various situations. The International Classification of Diseases codes grouped 3 major diagnoses (primary diagnosis, second diagnosis and third diagnosis) into 9 standard disease diagnosis groups as described previously[11].

After the abovementioned cleaning, transformation and grouping operations, there were 69 990 encounter records and 40 data variables that constituted the final harmonized data ready for patient readmission data feature analysis and prediction, in which the diabetes dataset contained 39 potential predictors and one outcome variable, namely, the 30-day readmission.

Finally, the harmonized dataset was randomly divided into training (70% records) and testing (30% records) subsets in a 7:3 ratio of encounter records[12], that is, 48 993 training records and 20 997 testing records, respectively. To avoid potential bias from the randomly split training and testing datasets, we compared the model performance between the training and the testing subsets using stratified k -fold cross-validation scores by splitting the data into k -parts ($k=10$)[13] to stratify the sampling by the class label. We trained the model on all parts except one, which was held out as a test set to evaluate the performance of the model.

2.3. Imbalanced class correction

Due to the relative infrequency (9%) of the outcome (*i.e.*, readmission) in the dataset, there was an imbalanced distribution between the minority class (readmissions) and the majority class (no readmissions). Imbalanced data are one of the obstacles for machine learning algorithms; it arises when the data are dominated by a majority class, and the minority class is infrequent[14]. As a result, the classifier performance on the minority may be insufficient when compared to the majority. For example, a dumb classifier that always predicts the majority class can achieve high accuracy.

Approaches to re-balance data, which match the number of cases to the number of controls, include oversampling (*i.e.*, adding cases to match the number of controls), undersampling (*i.e.*, reducing the number of controls to match the number of cases), and hybrid approaches. We examined several strategies for resampling, including random oversampling, random undersampling, synthetic minority over-sampling technique with Tomek link, synthetic minority over-sampling technique, and undersampling NearMiss methods.

To identify the best resampling approach for each model, every

resampling approach was applied to the training dataset, and the models with the training set were then evaluated. We also compared the resampled training and testing sets with a model using no resampling (*i.e.*, an imbalanced set) with default settings.

2.4. Machine learning models

2.4.1. Logistic regression

For the data feature analysis, we derived the LR[15] model using stepwise variable selection with backward elimination, resulting in a final model with 14 variables based on the training set. We added a backward stepwise selection procedure so that the regression model started with all 39 available predictors included; then we sequentially removed non-significant predictors. At each stage, variables were considered for removal from the model if the P -value associated with the predictor was >0.05 using the Wald Test of Independence. The default procedure would remove the predictor with the largest P value among those with $P>0.05$. Variables were not allowed to re-enter the model after exclusion, and the process continued until all the remaining 14 predictors are $P<0.05$ and the confidence interval does not cross 1.

Then we assessed the association of patient risk factors with readmission risk using the readmission odds ratio for each influential data feature. For further examining the association at the factor level, we used the Fisher's exact test in the testing set to evaluate the importance of each level of the 14 risk factors.

We used logistic regression as a traditional statistical method to evaluate the baseline performance, EasyEnsemble (EE) as a typical imbalance learning model for skewed distribution, and multi-layer perceptron as a "vanilla" neural network. The models were used to simulate prediction models in the training set first and then were validated in the testing set. Sampling methods, classifier algorithms, and optimal threshold values were developed for each model from the training set and assessed on the independent testing set which played no role in model development. Our study did not perform extensive hyper-parameter optimization in the training phase in the attempt to compare the approaches as fairly as possible, especially as the neural network-based models had a very large set of parameters to choose from (such as the number of layers and dimensions of hidden vectors), which made it infeasible to explore the entire space[16].

2.4.2. Artificial neural networks

Deep learning is a naive neural network of multiple layers, which learns representations of data with various levels of abstraction. One classical architecture was the artificial neural network (ANN)[17].

In this study, a multiple-layer deep ANN classifier was first trained to distinguish between two classes (0 as not re-admitted and 1 as re-admitted) using 7 layers with 40 neurons/layer, and the sigmoid activation function was chosen for the output layer. The selected optimization algorithm was Adam, and the loss function applied was the cross-entropy loss.

2.4.3. EasyEnsemble

The EasyEnsemble[18] classifier was equipped with a built-in under-sampling algorithm, which independently sampled several subsets from negative examples. One classifier was constructed for every subset[19].

In our study, the EE model created 10 subsets of our pre-processed data, in which each subset was balanced for readmission class and no readmission class by randomly under-sampling the majority (no readmission) class. All classifiers generated from individual subsets using AdaBoost were then combined and voted on the final decision.

2.5. Evaluation

To determine whether the model generalizes well or not, we applied the following evaluation metrics to the 3 models' performance in the training and testing sets: sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value. The F1 statistic and area under the receiver operating characteristic curve (AUC ROC) of multivariate observations were drawn, which could additionally assess a model's prediction performance.

Model comparison was performed using the F1 score for classification accuracy of readmission, defined as the harmonic mean of precision (*i.e.*, positive predictive value) and recall (*i.e.*, sensitivity), with perfect precision and recall providing an F1 score of 1 and the worst precision and recall providing an F1 score of 0.

2.6. Data analysis

Statistical analyses were performed using the R programming software package with a MacBook Pro (R-64 3.6.1). Data were cleaned and transformed in a CSV spreadsheet before the start of analysis. Machine learning models were built in Python using the Scikit-learn, Tensorflow, and Keras packages. All analyses were conducted using the Sklearn version 0.21 package in Python 3.6 (Python Software Foundation). Univariate and multivariate logistic regression, multi-layer perceptron, and EE analyses were performed to examine the association of independent variables with the dependent variable (readmission) and to determine the associated factors. The factors that had $P < 0.05$ were considered statistically significant and the confidence interval does not cross 1.

3. Results

A total of 69 990 encounter records and 40 data variables were included in this study. The basic characteristics of diabetes mellitus in-patient claim data across three datasets are listed in Table 1. Diabetes medication use and diagnosis category distribution were also analyzed. A typical encounter record has 4 demographic

variables, 8 in-patient administrative claim data features, 3 laboratory measurements, and a range of 1-16 discharge diagnoses. Most hospitalized individuals had multiple coexisting conditions represented by the 3 diagnosis levels in which each level has been categorized into 8 disease comorbidity groups. Comorbidities with the first diagnosis were present in more than half of the patients, with circulatory diseases being the most common comorbidity (30.66%), followed by respiratory diseases (13.56%) and digestive diseases (9.27%). The common medication uses for diabetes include 4 categorical values ("No change", "Steady", "Up", and "Down") ("up" if the dosage was increased during the encounter, "down" if the dosage was decreased, "steady" if the dosage did not change, and "no change" if the drug was not prescribed).

Table 1. Characteristics of diabetes mellitus patients from the Health Facts Database.

Variables	n=69 990
Demographics	
Age, year, median (Q1-Q3)	65 (50-80)
Hospital length of stay, d, median (Q1-Q3)	4.4 (3.4-5.4)
Gender	
Female, n (%)	37 239 (53.21)
Race, n (%)	
Unknown/Invalid	1 919 (2.74)
African American	12 627 (18.04)
Asian	488 (0.70)
Caucasian	52 305 (74.73)
Hispanic	1 501 (2.14)
Other	1 150 (1.64)
Hospitalization	
Diabetes medication prescribed	
Yes, n (%)	53 304 (76.16)
Diabetes medication changed	
Yes, n (%)	31 497 (45.00)
HbA1c, n (%)	
>8%	6 239 (8.91)
7%-8%	2 866 (4.09)
<7%	3 741 (5.35)
Not measured	57 144 (81.65)
Primary diagnosis ICD9, n (%)	
Circulatory	21 390 (30.56)
Respiratory	9 491 (13.56)
Digestive	6 488 (9.27)
Diabetes	5 748 (8.21)
Injury	4 696 (6.71)
Musculoskeletal	4 064 (5.81)
Genitourinary	3 441 (4.92)
Neoplasms	2 538 (3.63)
Other	12 134 (17.34)
Diabetes medication insulin, n (%)	
Dosage decreased	7 324 (10.46)
Drug not prescribed	34 268 (48.96)
No dosage change	21 621 (30.89)
Dosage increased	6 777 (9.68)
Primary outcome	
In hospital mortality, (%)	1.64
30-day readmission, n (%)	
Yes	6 285 (8.98)
No	63 705 (91.02)

Even though the record sizes were different among the 3 datasets (full, training, and testing), the frequencies of distinct record values of 40 individual data features were consistent with less than 1% difference across the 3 datasets. The primary outcome (30-day readmission) was a binary classification. The hospital readmission rates varied little among the 3 datasets (8.98% for the full set, 8.88% for the training set, and 9.21% for the testing set). The same observation is true for the diagnosis category distribution among the three datasets. For example, the first diagnosis frequencies for circulatory disease were 30.56%, 30.66%, and 30.32% in the full, training, and testing datasets, respectively.

We also found that the mean cross-validation accuracy scores and standard deviation were comparable among the full, training, and testing sets. The results suggest that the model performed equally well in general with either training or testing subsets, indicating that there was no significant bias in the random splitting process. This result is also consistent with the data attribute distribution.

3.1. Data features related to readmission

To determine whether some or all 39 variables are potential predictors for readmission, we applied the LR model for feature selection in the training set of 48 993 patient encounter records.

As can be seen in Figure 1, readmission odds based on relative influence scores were compared across 39 variables using LR, accounting for confounding patient characteristics. The 14 variables selected as the most influential data features were obtained by ranking based on relative influential scores because they were the most discriminative features with statistically significant ($P < 0.05$) and the confidence interval does not cross 1.

The 14 most influential features identified from the training set include 4 numeric fields and 10 categorical variables, which are listed in order by their odds ratios in Figure 1. The 10 categorical fields are insulin, medical specialty, diagnoses (primary, second, and third), admit source, age, diabetes medicine, repaglinide, and

discharge disposition. The 4 continuous numerical features are time in hospital, number of diagnosis, number of emergency visits, and number of in-patient visits.

The relative influential scores of the 14 most influential data features (Figure 1) ranging from 0.82 (0.75-0.89) (insulin) to 2.96 (2.86-3.06) (discharge disposition) represent the probability of readmission events. The top 5 most influential data features include discharge disposition, repaglinide, number of in-patient visits, diabetes medicine, and age. They demonstrate a high risk of readmission, evidenced by their high relative influential scores (> 1.24 , Figure 1), whereas the last 2 data features (medical specialty and insulin) have a lower probability of readmission (odds ratio < 1). For example, one-unit change in discharge disposition would make the readmission event almost 3 times as likely (odds ratio = 2.96) to occur, whereas one-unit change in insulin would lead to readmission being less likely (odds ratio = 0.82) to occur.

To further determine the data factor level that affected readmission in patients with diabetes, we included 48 993 patient encounter records in the training set for 10 categorical variables in the multivariate logistic regression model. We found that patient transfer, older age (> 60 years), primary diagnosis with circulatory diseases, or diabetes were high-risk factors associated with increased odds of readmission (Odds 1.1-4.8, $P < 0.05$, Table 2). However, primary diagnosis with respiratory disease, attending doctors with medical specialty in emergency or cardiology, and appropriate insulin usage were associated with decreased odds of readmission (Odds 0.8-0.9, $P < 0.05$, Table 2).

Except for repaglinide and HbA1c result, there was at least one factor level per categorical variable contributing to predict readmission with statistical significance ($P < 0.05$) (Table 2). Repaglinide had no significant P value at any factor level (No, Steady and Up), but it showed significant influence at the variable level (Figure 1 and Table 2). To investigate whether hemoglobin A1C (HbA1c) is related to readmission in diabetes, we included HbA1c in the LR model and Fisher's exact test.

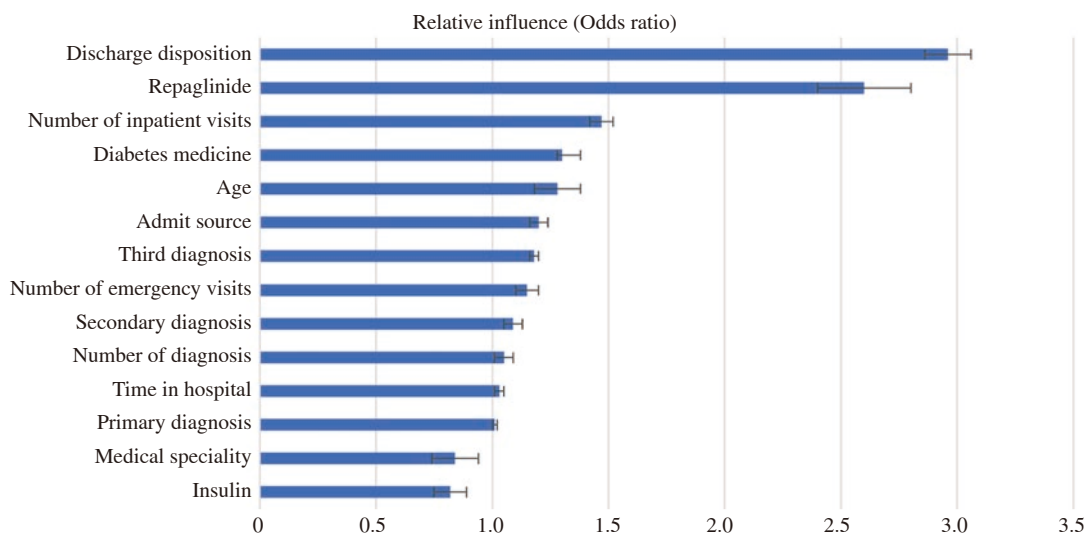


Figure 1. Identification of the most influential data features based on the logistics regression (LR) model. The bar graph compares observed relative odds ratio of readmissions with 95% confidence interval on the x-axis based on LR model predictions using 39 potential predictors in the training dataset. The 14 most influential data features were selected and plotted as blue bars on the y-axis, and the lengths of the bars represent the relative odds ratios of the individual data features for readmission prediction.

Table 2. The categorical factors of the most influential predictors of readmission in diabetes.

Factors	Coef	Odds	P value
Discharge disposition			
Transfer to special care	1.559 7	4.8	<0.001
Transfer to acute care	0.611 1	1.8	<0.001
Primary diagnosis			
Circulatory	0.110 2	1.1	0.037
Respiratory	0.146 8	0.9	0.021
Digestive	0.006 5	1.0	0.927
Diabetes	0.183 8	1.2	0.012
Injury	0.040 7	1.0	0.580
Musculoskeletal	0.131 3	0.9	0.159
Genitourinary	0.030 7	1.0	0.715
Neoplasms	0.055 9	0.9	0.583
Age			
31-60	0.112 3	1.1	0.385
>60	0.324 9	1.4	0.013
Medical specialty			
Emergency	0.202 3	0.8	0.011
Cardiology	0.233 9	0.8	0.006
Surgery	0.161 4	0.9	0.061
Orthopedics	0.185 0	0.8	0.112
Other	0.036 6	1.0	0.373
Insulin			
Drug not prescribed	0.237 8	0.8	0.005
No dosage change	0.217 0	0.8	0.001
Dosage increased	0.146 4	0.9	0.032
Admit source			
Clinic referral	0.207 1	1.2	0.010
Transfer from a hospital	0.029 8	1.0	0.774
Emergency	0.154 7	1.2	0.051
Repaglinide			
Drug not prescribed	0.512 0	1.7	0.621
No dosage change	0.780 8	2.2	0.454
Dosage increased	1.518 0	4.6	0.165
HbA1c result			
>8	0.099 9	0.9	0.312
None	0.037 4	1.0	0.649
Norm	0.038 4	1.0	0.720

Based on the results (Table 2), the addition of HbA1c did not improve the performance of the LR model, and *P* values were not significant at any factor level (Normal, None and >8) ($P>0.05$, Table 3). Therefore, HbA1c was not included in the most influential feature list (Figure 1).

The patients with higher odds ratios (>1.0) were more likely to be re-admitted, whereas those with a smaller odds ratio (<1.0) were less likely to be re-admitted. For example, readmission odds were delineated by patient age in 30-year epochs (with age 0-30 years as the reference). Age trends in the odds of readmission varied across epochs. The odds of readmission increased to 1.1 for age 31-60 years and to 1.4 for those aged above 61 years. This suggested that patients in the oldest epochs were 1.4-fold more likely to be re-admitted than younger patients (31-60). When compared to all other discharge dispositions, patients who were discharged to either special care or acute care facility had statistically significant 4.8- or 1.8-fold higher readmission likelihood ($P<0.000 1$), which indicates a strong association between patients who were discharged to either special care or acute care facilities and patients with readmission.

Interestingly, for primary diagnosis categories, the opposite odds effects of readmission were observed. Specifically, significantly high odds of readmission were associated with diabetes (1.2) and circulatory diseases (1.1). In contrast, the odds of readmission were significantly low (<1.0) with respiratory diseases (0.9), indicating that patients with primary diagnosis related to respiratory diseases would be less likely to be re-admitted.

We selected the 3 most influential data features (discharge disposition, primary diagnosis, and age) and 1 less influential data feature (HbA1c result) for further statistical testing using the test dataset.

We first stratified patient records into subgroups based on the level of data factors and then compared the differences in readmission

Table 3. Validation of selected data features in the testing dataset.

Base control group			Testing group			Statistics test	
Name	Size	REA rate	Name	Size	REA rate	OR	P value
Discharge disposition							
Other locations	16 371	7.70%	Transfer to special care facility	430	27.67%	4.6 (3.6-5.6)	<0.001
Other locations	16 371	7.70%	Transfer to acute care facility	4 196	13.18%	1.8 (1.6-2.0)	<0.001
Age							
0-30 years	537	6.33%	31-60 years	6 603	7.27%	1.2 (0.8-1.6)	0.487
0-30 years	537	6.33%	61-100 years	13 857	10.24%	1.7 (1.2-2.2)	0.002
Primary diagnosis							
Other diagnosis	3 648	9.70%	Circulatory diseases	6 367	10.04%	1.0 (0.9-1.1)	0.602
Other diagnosis	3 648	9.70%	Respiratory diseases	2 904	6.99%	0.7 (0.6-0.8)	<0.001
HbA1c							
None	17 108	9.35%	Norm	1 164	9.02%	1.0 (0.8-1.2)	0.755
None	17 108	9.35%	>7	826	8.11%	0.9 (0.7-1.1)	0.244
None	17 108	9.35%	>8	1 899	8.53%	0.9 (0.8-1.0)	0.260

REA rate: Readmission rate; 95% confidence interval for odds ratio is also presented.

Table 4. Evaluation of readmission prediction models and sampling algorithms.

Sampling algorithms	Model	Threshold	Data type	F1 score	Sensitivity	PPV
No	LR	0.500 000	Testing	0.011	0.006	0.524
No	ANN	0.500 000	Testing	0.066	0.040	0.193
No	LR	0.079 885	Testing	0.213	0.591	0.130
No	ANN	0.111 484	Testing	0.207	0.460	0.133
No	EE	0.499 497	Testing	0.216	0.570	0.133
Random oversampling and undersampling	LR	0.312 963	Testing	0.213	0.572	0.131
Random oversampling	LR	0.474 926	Testing	0.212	0.570	0.130
Condensed nearest neighbor rule undersampling	LR	0.254 011	Testing	0.211	0.534	0.131
Random undersampling	LR	0.309 724	Testing	0.208	0.579	0.127
Random oversampling and undersampling	ANN	0.348 134	Testing	0.183	0.424	0.117
Undersampling	ANN	0.393 258	Testing	0.178	0.406	0.114
Undersampling nearMiss	LR	0.510 219	Testing	0.172	0.645	0.645
Undersampling nearMiss	ANN	0.558 734	Testing	0.171	0.629	0.099
Oversampling	ANN	0.394 007	Testing	0.161	0.248	0.119
SMOTETomek	ANN	0.549 018	Testing	0.149	0.155	0.143
Oversampling SMOTE	ANN	0.437 848	Testing	0.134	0.139	0.129
Oversampling SMOTE	LR	0.570 423	Testing	0.054	0.036	0.108
SMOTETomek	LR	0.580 079	Testing	0.053	0.035	0.107
Average performance values in the testing set				0.165	0.396	0.160
Undersampling nearMiss	ANN	0.558 734	Training	0.965	0.965	0.966
Oversampling SMOTE	ANN	0.437 848	Training	0.943	0.934	0.952
SMOTETomek	ANN	0.549 018	Training	0.942	0.932	0.953
Random oversampling and undersampling	ANN	0.348 134	Training	0.856	0.890	0.825
Random undersampling	ANN	0.393 258	Training	0.856	0.875	0.837
Random oversampling	ANN	0.394 007	Training	0.907	0.933	0.883
Oversampling SMOTE	LR	0.570 423	Training	0.923	0.881	0.969
SMOTETomek	LR	0.580 079	Training	0.923	0.879	0.970
No	ANN	0.111 484	Training	0.298	0.687	0.190
No	ANN	0.500 000	Training	0.242	0.147	0.700
Undersampling nearMiss	LR	0.510 219	Training	0.665	0.654	0.676
No	EE	0.499 497	Training	0.225	0.613	0.138
Random oversampling	LR	0.474 926	Training	0.613	0.612	0.615
Random oversampling and undersampling	LR	0.312 963	Training	0.514	0.609	0.445
No	LR	0.079 885	Training	0.219	0.628	0.133
No	LR	0.500 000	Training	0.010	0.005	0.382
Random undersampling	LR	0.309 724	Training	0.517	0.625	0.441
Condensed nearest neighbor rule undersampling	LR	0.254 011	Training	0.437	0.575	0.353
Average performance values in the training set				0.614	0.691	0.635

SMOTE: Synthetic minority over-sampling technique; SMOTETomek: Synthetic minority over-sampling technique with Tomek link.

rate between paired subgroups using Fisher’s exact test.

The statistical test results from the test dataset’s paired factor comparisons for each data feature are listed in Table 3. In general, the Fisher’s exact test results using the test set (Table 3) confirmed the LR model *Chi*-square statistics results using the training set in Table 2. There was no statistically significant difference among readmission rates among the 3 patient groups partitioned by hemoglobin HbA1c results (normal, >7, and >8) (Table 3). We observed that the readmission rates and odds values from the 2 discharge disposition subgroups and the 61-100 years age group were significantly higher. The patients with primary respiratory disease diagnosis had lower readmission rates than those with circulatory diseases or other diagnoses (Table 3).

3.2. Model selection

To develop models and algorithms for readmission prediction,

we applied various data sampling algorithms to achieve a balanced class distribution in the training set. In table 4, we carried out 36 evaluation tests with the 3 predictive models (LR, EE, and ANN) and two input datasets (training and testing).

When we compared 18 training set results with their corresponding testing results (Table 4), we noticed that the training data performed better than the testing dataset (*e.g.*, average F1 score from the training set was 0.614 compared with 0.165 from the testing set). Specifically, training results from the deep machine learning ANN model were much better than those from the corresponding testing set results (ANN F1 score in the training set >0.85 *vs.* ANN F1 score in the testing set <0.21), indicating that the ANN model could be trained to achieve a much better performance in the training set compared to the results from the testing set.

However, traditional machine learning LR and EE models resulted in much smaller differences in performance between the training set and the testing data, suggesting that LR and EE model

Table 5. Readmission prediction with selected data features.

Model	Data type	F1 score	Sensitivity	PPV	
LR	Full training	0.613	0.612	0.615	
	Influential training	0.607	0.599	0.616	
	Less influential training	0.555	0.560	0.550	
	Full testing	0.212	0.570	0.130	
	Influential testing	0.213	0.561	0.132	
	Less influential testing	0.177	0.539	0.106	
	ANN	Full training	0.907	0.933	0.883
		Influential training	0.750	0.793	0.711
		Less influential training	0.660	0.675	0.646
Full testing		0.161	0.248	0.119	
Influential testing		0.188	0.460	0.118	
Less influential testing		0.168	0.442	0.104	
EE		Full training	0.225	0.613	0.138
		Influential training	0.218	0.625	0.132
		Less influential training	0.182	0.567	0.108
	Full testing	0.216	0.570	0.133	
	Influential testing	0.213	0.593	0.130	
	Less influential testing	0.181	0.547	0.108	

results from the testing dataset are more consistent than those from the ANN model. Therefore, the LR and EE models were relatively superior to the ANN model because of stable and balanced performance between the training and testing sets. Furthermore, the LR and EE models exhibited better prediction ability than the ANN model in two of the three metrics (F1 scores: 0.216, 0.213 *vs.* 0.207; sensitivity: 0.57, 0.591 *vs.* 0.46). Of the 3 analyzed models, the LR and EE models are more suitable for the readmission classification problem in patients with diabetes. The deep machine learning ANN model did not outperform the traditional LR and EE models in this study.

Overall, the average F1 score of all models and sampling algorithms in the testing set was 0.156 (0.011–0.216), and the best performance in the testing dataset was derived from the EE model with an F1 score of 0.216 (Table 4), a sensitivity of 0.57, and a PPV of 0.133. The poor performance results (F1 scores < 0.1) were from the LR and ANN models with the default method (*i.e.*, no sampling algorithm and default threshold of 0.5) and the LR model with oversampling SMOTE and SMOTETomek.

The highest sensitivity (0.645) was from the LR model with Undersampling Edited Nearest Neighbors algorithm, and comparable results (0.629, 0.57 and 0.57) were also observed from the ANN models with Undersampling Edited Nearest Neighbors algorithm, LR model with random oversampling, and EE model. The lowest sensitivity (0.006) was from the LR model with the default threshold of 0.5 and no sampling algorithm.

We also evaluated the AUC ROC of the 36 tests. When comparing 18 training set results with their corresponding testing results, we observed that the training data also performed better than the testing dataset (*e.g.*, average AUC score for the training set was 0.812 *vs.* 0.589 for the testing set). We reached an identical conclusion, that is, the ANN model tends to overfit the training

data, whereas the other two algorithms led to much smaller differences between the training and testing datasets.

3.3. Readmission prediction with selected data features

To further validate the most influential data features identified in Figure 1, we partitioned the diabetes datasets (training and testing) into 3 sets, including the full-size dataset with 39 independent variables, the influential subset with 14 most influential data features in Figure 1, and the less influential subset with the remaining less influential 25 data features.

We evaluated the performance of the EE, LR, and ANN models in the feature split training and testing subsets with the oversampling method and optimal threshold values. We examined whether the selected 14 data features could improve readmission prediction performance by 18 prediction evaluation tests (Table 5). For a given model, there are two panels of results (training and testing), in which each panel includes results from the three datasets (full, influential, and less influential).

We found that a given model's training set performances were better than those from the corresponding testing datasets (Table 5). The performance values from the ANN model were significantly higher in the 3 training sets (*e.g.*, F1 scores: 0.907, 0.75, 0.66) than those from the 3 testing sets, which were 0.161, 0.188, and 0.168, respectively. This suggests an overfitting problem in the ANN model. In this regard, the EE model produced the smallest differences between the 3 training and testing sets. From the results in Table 5, it can be concluded that the EE model performed better and was more stable than the LR and ANN models in both full and influential datasets, with the F1 score of 0.225 and 0.218 for the training full and influential sets and 0.216 and 0.213 for the testing full and influential sets, respectively.

The performance values from the full dataset and the most influential set achieved a comparable level of prediction in all panels (Table 5). For example, the logistic model F1 score in the full training set and the most influential set was 0.613 and 0.607, respectively. In contrast, a smaller F1 score value was observed from the less influential sets in all panels. The results suggest that the 14 most influential features could replace the full feature list to predict readmission, and the 25 less influential features were weaker predictors with poorer performance regardless of the predictive models, which is consistent with the findings presented in Table 2 and 3.

In summary, using the LR model, we identified the 14 most influential features (Figure 1, Table 2, 3 and 5) from the patients with diabetes in the training set and validated them in the testing set. They include 4 numeric variables (time in hospital, number of diagnosis, number of emergency visits, and number of in-patient visits) and 10 categorical data variables [insulin, medical specialty, diagnoses (first, second, third), admit source, age, diabetes medicine, repaglinide, and discharge disposition].

The performances of predictions with the 14 selected influential predictors were better than those with the 25 less influential predictors, regardless of the dataset types (training or testing) or prediction models (Table 5). Our result is supported by a study using the random forest algorithm with the same data source, in which 11 out of 15 data features overlapped with those in Figure 1, namely, discharge disposition, number of in-patient visits, age, time in hospital, number of emergency visits, number of diagnosis, primary diagnosis, secondary diagnosis, third diagnosis, diabetes medicine, and insulin[20]. Robbins *et al.*[21] collated 76 distinct risk factors for readmission of people with diabetes from published literature. The identified risk factors included co-morbidity burden, age, race, and insurance type, in which two out of the four risk factors (co-morbidity and age) are consistent with our results (Figure 1).

Moreover, we found that there are 2 factor levels with opposite influential directions on readmissions. For instance, patient transfer, older age (>60 years), and primary diagnosis with circulatory diseases or diabetes were high-risk factors associated with increased odds of readmission (OR 1.1-4.8, $P < 0.05$, Table 2). The result suggests that these patients were more vulnerable to readmission. However, primary diagnosis with respiratory disease, or attending doctors with a medical specialty in emergency or cardiology, or insulin usages were factors associated with decreased odds of readmission (OR 0.8-0.9, $P < 0.05$, Table 2). Besides the known risk factors of diabetes (co-morbidity burden and age), our results might identify quality markers, such as service providers in respiratory disease. Emergency or cardiology practitioners may provide better services than other teams. The results might also suggest that the types of administrations in those teams are easier and more straightforward to be managed better in clinical settings.

The diabetes dataset in the current study had an imbalanced binary classification problem for readmission prediction, where

91% were in the majority class without readmission and 9% were in the minority class with readmission (Table 1). When we ran LR or ANN models with default settings (*i.e.*, no sampling algorithm and classification threshold at 0.5), the predicted performance of both LR and ANN was very poor (F1 scores: 0.011 and 0.066, respectively, Table 4).

Hence, we conducted a systematic approach to identify optimal thresholds and the best resampling and classification algorithms for readmission. We found that either an optimal threshold value or random oversampling algorithm could improve the performance of a classifier. The threshold that achieved the best evaluation metric in the training set was then adopted for the model. We made predictions on the testing dataset based on the model that achieved a big rise in F1 scores (20-fold increase for the LR model and 3-fold increase for the ANN model) (Table 4).

4. Discussion

Diabetes is related to oxidant stress[22], not only significantly affects the quality of life and imparts disease-related morbidity and mortality but also greatly increases medical expenses due to its chronic nature and hospital readmissions[23]. Hospital readmission for patients with diabetes serves as a quality indicator that is subject to scrutiny from commercial payers and consumers alike. Hence, there is a great need to identify readmission-related data features and to predict hospital readmissions for patients with diabetes. This study aims to identify the most influential data features related to readmission and establish a predictive model to help reduce readmission rate. We found that machine learning is an effective approach for data feature selection and readmission prediction.

Feature selection plays a significant role in machine learning because irrelevant attributes may lead to low accuracy, and model over-fitting results in classification errors[24]. Identifying predictive variables of readmission is essential because this information can be translated to care interventions to further reduce readmissions[25]. Unfortunately, very few related readmission predictors for diabetes have been reported in the past literature.

By using a logistic regression model, we identified 14 most influential features and evaluated their performance of predictions. The selected influential predictors were better than those with less influential ones.

However, contrary to popular belief[26], the HbA1c result is one of the less influential data variables presented in the current study. We found that neither HbA1c nor any of its 3 factor levels (>8, None and Norm) demonstrated any significant association with readmission, suggesting that HbA1c result might not be a reliable predictor of readmission in patients with diabetes. It is well-known that for patients with diabetes, HbA1c measures the average amount of glucose attached to hemoglobin over the past 3 months.

The higher the levels, the greater the risk of developing diabetes complications. While they are not necessarily related to 30-day readmission outcomes. There was no difference in the average HbA1c values between the group that was re-admitted and the one that was not (7.63% vs. 7.55%, $P=0.470$)^[27]. Several other studies have also suggested that the elevated preoperative HbA1c level was not associated with increased readmission^[28,29], or HbA1c was only a less-common risk for readmission of in-patients with diabetes^[21].

The machine learning algorithms (LR or ANN) can predict a probability of class membership that is interpreted into a meaningful class label. When the training dataset is imbalanced, the resulting model is biased towards the majority class, which makes it inherently difficult to achieve high sensitivities^[18]. The limitations of applying classification algorithms to imbalanced datasets are well described^[30–32].

However, many studies on the prediction of readmission using machine learning algorithms do not report the use of any procedure regarding class imbalance. They continue to use imbalanced data to conduct statistical modeling^[33–35]. The resulting model is biased towards the majority class (*i.e.*, non-readmission), achieving high accuracy in the majority class and low accuracy for the minority class^[14].

In the current study, we used 3 different machine learning models (LR, ANN, and EE). ANN^[36] is a deep learning model; LR is a traditional statistical model; and EE combines different ensemble strategies to achieve stronger generalization.

We tried various evaluation metrics in our models (data not fully shown) and finally selected 3 metrics (F1 score, sensitivity, and PPV). The F1 score maintains a balance between the precision (PPV) and recall (sensitivity) for our classifier. The AUC ROC or c-statistic is the standard de facto metric for measuring the discrimination ability of risk prediction models^[14]. The AUC ROC indicates how well the probabilities from the positive classes are separated from the negative classes. The F1 score is the harmonic mean of precision and recall applicable for any point on the ROC curve. The AUC is used for many different levels of thresholds, whereas the F1 score is an optimal fixed pair of precision and recall. The AUC may not be the best measure for predicting imbalanced minority class because it is used to describe how well the model ranks cases over non-cases^[6]. Atrial fibrillation prediction by ANN had an AUC ROC of 0.800 but an F1 score of 0.110^[37].

Since this is an imbalanced dataset with a small positive class (*i.e.*, readmission class), ROC plots are visually useful and provide an overview of a classifier's performance across a wide range of specificities. However, ROC plots in the context of imbalanced datasets can be deceptive owing to an intuitive but wrong interpretation of specificity^[38]. Therefore, F1 score is considered a better evaluation metric over ROC AUC.

Our results suggest that the deep learning model did not

outperform traditional statistical models (LR and EE) because ANN was more prone to over-fitting. The LR and EE models offered relatively consistent prediction performance between the training set and testing set. The results of our study showed that the LR model had a higher F1 than ANN (0.213 vs. 0.183).

The results from other studies^[6,37,39] also indicate that ANN is not necessarily superior to LR. Frizzell *et al.*^[6] found that machine learning approaches did not improve 30-day readmission predictions compared to logistic regression. A simple logistic regression model based on known and clinical risk factors performed nearly as well as the more sophisticated machine learning model^[37].

Based on the review of 10 unique studies that compared ANN and LR models in predicting health-related outcomes in trauma patients, the authors concluded that ANN had better performance than LR for predicting the terminal outcomes of trauma patients with respect to both the AUC and accuracy rate^[40]. However, ANN's coefficients/weights do not have simple interpretations, and there is a higher chance of over-fitting^[41], as we observed. Other studies suggested that both LR and shallow neural networks performed at roughly the same level^[42]. The superior popularity of LR is attributed to the interpretability of model parameters and the ease of use. An obstacle for neural networks is their black-box nature that hinders the model's interpretability.

In general, the interaction between variables does not affect the ANN, but if the goal is to study the causal relationship between variables, LR would be a more suitable choice. ANN is likely to have over-fitting^[38] and not necessarily be more predictive, as we found. However, ANN and LR models can complement each other. In situations where the ANN cannot report individual factors, LR can still provide this information. The LR model facilitates risk stratification for key patient-level outcomes for routinely collected health care data and helps identify a high-risk target population for effectively deploying population-based interventions to reduce readmissions.

In conclusion, we implemented and compared 3 machine learning models for the identification of readmission-related data features and predicting readmission of patients with diabetes. Using the LR model, we identified 14 risk factors for hospital readmission in patients with diabetes and further understood their influences at factor levels.

We found that either an optimal threshold value or oversampling algorithm could effectively correct the imbalance diabetes dataset. The EE and LR models were more suitable for the readmission classification problem in patients with diabetes. The deep machine learning ANN model did not outperform traditional LR and EE models in this study.

Our findings suggest that machine learning models hold promise for integration into clinical workflow to predict readmission. One of the most promising aspects of a future ideal prediction model would be that it allows the identification of patients at high risk for

unplanned readmission. This can then guide intensive or targeted treatment strategies during admission to optimize the discharge plan and follow-up. Ultimately, we hope to significantly reduce readmission, improve patient health, and save medical costs.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Funding

This work was supported in part by the Key Research and Development Program for Guangdong Province (No. 2019B010136001), in part by Hainan Major Science and Technology Projects (No. ZDKJ2019010), in part by the National Key Research and Development Program of China (No. 2016YFB0800803 and No. 2018YFB1004005), in part by National Natural Science Foundation of China (No. 81960565, No. 81260139, No. 81060073, No. 81560275, No. 61562021, No. 30560161 and No. 61872110), in part by Hainan Special Projects of Social Development (No. ZDYF2018103 and No. 2015SF 39), and in part by Hainan Association for Academic Excellence Youth Science and Technology Innovation Program (No. 201515).

Authors' contributions

Conception and design by XXY, LCY, and ZYC; Administrative support by WX and BF; Collection and assembly of data: XXY; Data analysis and interpretation: XXY; Manuscript writing: XXY and ZYC; Final approval of manuscript by all the authors.

References

- [1] Prevention for Disease Control. *National diabetes statistics report*. [Online]. Available from: <https://www.cdc.gov/diabetes/library/features/diabetes-stat-report.html>. [Accessed on 1 May 2020].
- [2] Lage MJ, Boye KS. The relationship between HbA1c reduction and healthcare costs among patients with type 2 diabetes: Evidence from a U.S. claims database. *Curr Med Res Opin* 2020; **36**(9): 1441-1447.
- [3] American Diabetes Association. Glycemic targets: Standards of medical care in diabetes-2019. *Diabetes Care* 2019; **42**(Suppl 1): S61-S70.
- [4] Unnikrishnan R, Mohan V. Diabetes in the tropics: Prevalent, increasing and a major public health problem. *Trans R Soc Trop Med Hyg* 2016; **110**(5): 263-264.
- [5] Gan D. *International diabetes federation diabetes atlas*. [Online]. Available from: <https://idf.org/e-library/epidemiology-research/diabetes-atlas.html>. [Accessed on 1 May 2020].
- [6] Ghasemi-Dehnoo M, Amini-Khoei H, Lorigooini Z, Rafieian-Kopaei M. Oxidative stress and antioxidants in diabetes mellitus. *Asian Pac J Trop Med* 2020; **13**(10): 431-438.
- [7] Aghdam AR, Watson J, Miah SJ, Cliff C. A novel information sharing framework for people living with type-2 diabetes in the context of a group education program. *Health Inf Sci Syst* 2021; **9**(1): 1-16.
- [8] Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM. Classification and prediction of diabetes disease using machine learning paradigm. *Health Inf Sci Syst* 2020; **8**(1): 1-14.
- [9] Rodriguez-Gutierrez R, Herrin J, Lipska KJ, Montori VM, Shah ND, McCoy RG. Racial and ethnic differences in 30-day hospital readmissions among US adults with diabetes. *JAMA Netw Open* 2019; **2**(10): e1913249.
- [10] Auerbach AD, Kripalani S, Vasilevskis EE, Sehgal N, Lindenauer PK, Metlay JP, et al. Preventability and causes of readmissions in a national cohort of general medicine patients. *JAMA Intern Med* 2016; **176**(4): 484-493.
- [11] Strack B, DeShazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ, et al. Impact of HbA1c measurement on hospital readmission rates: Analysis of 70 000 clinical database patient records. *Biomed Res Int* 2014; **2014**: 781670. doi: 10.1155/2014/781670.
- [12] Parikh RB, Manz C, Chivers C, Regli SH, Braun J, Draugelis ME, et al. Machine learning approaches to predict 6-month mortality among patients with cancer. *JAMA Netw Open* 2019; **2**(10): e1915997.
- [13] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI (US)* 1995; **2**: 1137-1143.
- [14] Artetxe A, Beristain A, Grana M. Predictive models for hospital readmission risk: A systematic review of methods. *Comput Methods Programs Biomed* 2018; **164**: 49-64.
- [15] Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. 3rd ed. New York: John Wiley & Sons; 2013.
- [16] Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus logistic regression for 30 days all-cause readmission prediction. *Sci Rep* 2019; **9**(1): 1-11.
- [17] Esteva A, Kuprel B, Novoa RA, Ko J, Thrun S. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **546**(7660): 686-686.
- [18] Turgeman L, May JH. A mixed-ensemble model for hospital readmission. *Artif Intell Med* 2016; **72**: 72-82.
- [19] Liu TY. Easyensemble and feature selection for imbalance data sets. *Proc Int Joint Conf Bioinform Syst Biol Intell Comput* 2009; **2009**: 517-520.
- [20] Pineda-Cortel MRB, Baybayan K, Bello PL, Camenforte LL, Ching SJ, Conti K, et al. Genetic association analysis of rs7754840 and rs7756992 polymorphisms in the *CDKALI* gene and gestational diabetes mellitus in selected Filipino pregnant women. *Asian Pac J Reprod* 2021; **10**(4): 155-161.
- [21] Robbins TD, Keung SLC, Sankar S, Randeva H, Arvanitis T. Risk factors for readmission of inpatients with diabetes: A systematic review. *J Diabets Complications* 2019; **33**(5): 398-405.

- [22] Nguyen TH, Nguyen TH, Nguyen VM, Nguyen TLP, Tran TVA, Do AD, et al. Antidiabetic and antioxidant activities of red seaweed *Laurencia dendroidea*. *Asian Pac J Trop Biomed* 2019; **9**(12): 501-509.
- [23] Shi BY. The importance and strategy of diabetes prevention. *Chronic Dis Transl Med* 2016; **2**(4): 204-207.
- [24] Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, Hadaegh F, Steyerberg EW, Khalili D. A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *J Clin Epidemiol* 2016; **71**: 76-85.
- [25] Shameer K, Johnson KW, Yahi A, Miotto R, Li L, Ricks D, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: A case-study using Mount Sinai Heart Failure Cohort. *Pac Symp Biocomput* 2017; **2017**: 276-287.
- [26] Engoren M, Schwann TA, Habib RH. Elevated hemoglobin A1c is associated with readmission but not complications. *Asian Cardiovasc Thorac Ann* 2014; **22**(7): 800-806.
- [27] Bakeri H, Wakefield D, Dulipsingh L. Is there a role of hemoglobin A1C in predicting hospital readmission rates for patients with diabetes. *J Endocrinol Diabetes Obes* 2018; **1**(2): 3.
- [28] Jones CE, Graham LA, Morris MS, Richman JS, Hollis RH, Wahl TS, et al. Association between preoperative hemoglobin A1c levels, post-operative hyperglycemia, and readmissions following gastrointestinal surgery. *JAMA Surg* 2017; **152**(11): 1031-1038.
- [29] Magny-Normilus C, Nolido NV, Borges JC, Brady M, Labonville S, Williams D, et al. Effects of an intensive discharge intervention on medication adherence, glycemic control, and readmission rates in patients with type 2 diabetes. *J Patient Saf* 2021; **17**(2): 73-80.
- [30] Li DC, Hu SC, Lin LS, Yeh CW. Detecting representative data and generating synthetic samples to improve learning accuracy with imbalanced data sets. *PloS One* 2017; **12**(8): e0181853.
- [31] Colborn KL, Bronsert M, Amioka E, Hammermeister K, Henderson WG, Meguid R. Identification of surgical site infections using electronic health record data. *Am J Infect Control* 2018; **46**(11): 1230-1235.
- [32] Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics* 2013; **14**(1): 106.
- [33] Agrawal D, Chen CB, Dravenstott RW, Stromblad CT, Schmid JA, Darer JD, et al. Predicting patients at risk for 3-day postdischarge readmissions, ED visits, and deaths. *Med Care* 2016; **54**(11): 1017-1023.
- [34] Jamei M, Nisnevich A, Wetchler E, Sudat S, Liu E. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PloS One* 2017; **12**(7): e0181173.
- [35] Mesgarpour M, Chausalet T, Chahed S. Ensemble risk model of emergency readmissions (ER-MER). *Int J Med Inform* 2017; **103**: 65-77.
- [36] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for Electronic Health Record (EHR) analysis. *IEEE J Biomed Health Inform* 2017; **22**(5): 1589-1604.
- [37] Premanand T, Colborn KL, Smith DE, Xing F, Debashis G, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of Incident Atrial Fibrillation. *JAMA Netw Open* 2020; **3**(1): 284-288.
- [38] Takaya S, Marc R. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One* 2015; **10**(3): e0118432.
- [39] Averitt AJ, Natarajan K. Going deep: The role of neural networks for renal survival and beyond. *Kidney Int Rep* 2018; **3**(2): 242-243.
- [40] Hassanipour S, Ghaem H, Arab-Zozani M, Seif M, Fararouei M, Abdzadeh E, et al. Comparison of artificial neural network and logistic regression models for prediction of outcomes in trauma patients: A systematic review and meta-analysis. *Injury* 2019; **50**(2): 244-250.
- [41] Renganathan V. Overview of artificial neural network models in the biomedical domain. *Bratisl Lek Listy* 2019; **120**(7): 536-540.
- [42] Desai RJ, Wang SV, Muthiah V, Thomas E, Sebastian S. Comparison of machine learning methods with traditional models for use of administrative claims with Electronic Medical Records to predict heart failure outcomes. *JAMA New Open* 2020; **3**(1): 447-454.