ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

Machine Learning based Readmission Prediction for Diabetes Patients using Hospital Readmission Reduction Technique (HRRT)

Received: 14 October 2022, **Revised**: 17 November 2022, **Accepted**: 16 December 2022

G.G.Rajput,
Department of computer science,
Akkamahadevi women's University, Vijayapura,
Karnataka,India 586108
ggrajput@yahoo.co.in

Ashvini Alashetty,
Department of computer science,
Akkamahadevi women's University, Vijayapura,
Karnataka,India 586108
ashwinialashetty@gmail.com

Keywords
Support Vector Machine, Principal Component Analysis, Machine Learning, Diabetes Mellitus, Logistic Regression.

Abstract
Nowadays, more and more patients suffer from still incurable diabetes disease. Every wrong chosen treatment for patients can harm their health and lead to early readmission that costs more money. Therefore, there is a demand for predicting the readmission of patients to increase quality of health care and also to reduce costs. We compared machine learning (ML)-based readmission prediction techniques such as ANN, K-Nearest Neighbor (KNN), support vector machines (SVM), decision trees, Random Forest (RF), logistic regression. In addition to this technique, we developed Hospital Readmission Reduction technique using principal component analysis (PCA) method to predict the exact diabetes patient's readmission. The accuracy of the ANN model was 92.2 percent with PCA, and in comparison, to the other methods, it had a bigger area under the receiver operating characteristic curve (ROC), which may indicate that its applicability is more suitable for predicting readmission. When compared to other machine learning techniques, the ANN model has obtained higher consistency.

1. Introduction

A hospital readmission is when a patient who is discharged from the hospital, gets re-admitted again within certain period of time. By predicting readmission, more attention may be given to treatment of patients with high probability of readmission and so increase the quality of care during hospitalization. Because there is no cure for the diabetes yet1 and diabetic patients can be readmitted in the future, an early readmission prediction can help mostly when it comes to selection of best treatment for the patient.

Data of patient's clinical encounters are being collected naturally with healthcare systems, thus data-driven approach seem to be appropriate for this problem. Machine learning algorithms used for early readmission prediction provide ability to process the data of a lot of patients and may help to find hidden dependencies in the data to outperform basic methods. Various previous studies involving the prediction of hospital readmissions have yielded an accuracy of only about 60-64 percent due to factors such as data imbalance, an overabundance of features, and the sheer unpredictability of medical scenarios. Diabetes mellitus (DM), a major chronic non-communicable disease, affects a large number of people and frequently requires hospitalization due to poor disease management [1-2]. A "readmission" occurs when a patient is readmitted to the same department after being discharged for the same condition within a specified time frame. Inadvertent readmission can be caused by a variety of factors, including incorrect initial diagnosis, relapse, early discharge, and so on [3]. The Centers for Medicare and Medicaid Services now uses the 30-day readmission rate following an index hospitalization as a key hospital performance indicator, and it is being scrutinized as a sign of poor patient care [4-5]. These readmission prediction strategies were discovered to be marginally more accurate than arbitrary guesswork. Many prediction activities, on the other hand, rely heavily on machine learning [6-9]. The proposed work shows how diabetes patients' readmissions may be predicted using machine learning algorithms such as ANN, SVM, Logistic Regression, Decision Tree, KNN and Random Forest.

2. Literature Review

Numerous researchers have discussed numerous important methods for treating diabetes mellitus problems. According to Veena Vijayan et al. [10], an increase in blood sugar levels causes diabetes. A number of computerized information systems using decision trees, SVM, Naive Bayes, and ANN algorithms were chosen for preventing and detecting diabetes. P. Suresh Kumar et al.[11] investigated many data mining techniques for diagnosing diabetes, including Decision Tree, SVM, and Naive Bayes. The main cause of blindness in diabetic people is diabetic retinopathy (DR). A group of machine learning algorithms were also tested by Ridam Pal et al. [12] for their effectiveness. M. Renuka Devi et al, [13] reported the analysis of several mining techniques for diabetes prediction utilizing Naive Bayes, Random Forest, Decision Tree, and J48 algorithms. Rahul Joshi et al's[14] proposal for ML methods that make use of the KNN and Naive Bayes algorithms to predict datasets at an early stage in order to save lives. The results of several methods that have been adopted to increase diagnosis consistency were forecasted by Zhilbert Tafa et al. in their study [15]. In order to predict diseases using PCA, Dhomse Kanchan et al.[16] looked into a variety of machine learning algorithms, including SVM, Nave Bayes, Decision Trees, and PCA.

The US Health Facts Medical Database is taken into consideration in this proposed work. Prior to training the prediction models, the dimensionality of the data was reduced using principal component analysis. Following the reduction, six machine learning algorithms—SVM, Decision Tree, Logistic Regression, K-NN, Ann, and Random Forest—are employed to create prediction models. In order to compare the outcomes, prediction accuracy is taken into account [17–21].

3. Methodology

The methodology, which describes the approach taken to carry out the inquiry and compare the findings, is contained in this part. Because it makes it easier to compare and assess several machine learning algorithms on real-time data, the Jupyter notebook application was employed in the study.

4. Dataset Description

The dataset for this study contained 100,000 medical histories for 70,000 diabetic patients from 130 hospitals in the United States between 1999 and 2008. A label identifying a patient's readmission status, which indicates whether or not a patient was readmitted to the hospital within 30 days, is included in the medical data collection along with 50 risk indicators. The dataset includes clinical records of in-patients with diabetes whose hospital stays ranged from 1 to 14 days, as well as information on the lab tests and medications they received while they were in the hospital. As shown on Fig. 1, 3 classes describing whether and when was patient readmitted are provided for prediction. The model was developed using a total of 59706 records after data cleaning and standardization. The last 23 parameters, including race, sex, age, admission type, admission location, length of stay, and drug usage, were identified as modeling risk factors.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

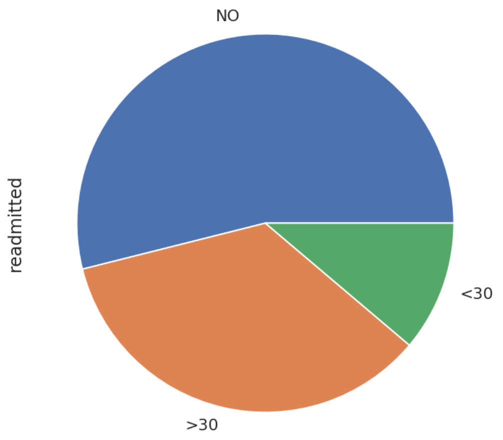


Fig. 1. General demographic distribution of US diabetic data set

Table 1. List Of Attributes In The Initial Dataset (The Dataset Is Also Available At The Website Of Datamining And Biomedical Informatics Lab At Vcu (http://www.cioslab.vcu.edu/)).

Attribute	Type	Description	%missing
Encounter ID	Numeric	Unique identifier of an encounter	0
Patient number	Numeric	Unique identifier of a patient	0
Race	Nominal	Values: Caucasian, Asian, African American, Hispanic, and other	2
Gender	Nominal	Values: male, female, and unknown/invalid	0
Age	Nominal	Grouped in 10-year intervals: 0, 10), 10, 20), …, 90, 100)	0
Weight	Numeric	Weight in pounds	97
Admission type	Nominal	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available	0
Discharge disposition	Nominal	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available	0
Admission source	Nominal	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital	0
Time in hospital	Numeric	Integer number of days between admission and discharge	0
Payer code	Nominal	Integer identifier corresponding to 23 distinct values, for example, Blue Cross/Blue Shield, Medicare, and selfpay Medical	52
Medical specialty	Nominal	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family/general practice, and surgeon	53
Number of lab procedures	Numeric	Number of lab tests performed during the encounter	0

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Number of procedures	Numeric	Numeric Number of procedures (other than lab tests) performed during the encounter	0
Number of medications	Numeric	Number of distinct generic names administered during the encounter	0
Number of outpatient visits	Numeric	Number of outpatient visits of the patient in the year preceding the encounter	0
Number of emergency visits	Numeric	Number of emergency visits of the patient in the year preceding the encounter	0
Number of inpatient visits	Numeric	Number of inpatient visits of the patient in the year preceding the encounter	0
Diagnosis 1	Nominal	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values	0
Diagnosis	Nominal	2 Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values	0
Diagnosis 3	Nominal	Additionalsecondary diagnosis (coded as first three digits of ICD9); 954 distinct values	1
Number of diagnoses	Numeric	Number of diagnoses entered to the system 0%	0
Glucose serum test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured	0
A1c test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured.	0
Change of medications	Nominal	Indicates if there was a change in diabetic medications (either dosage or generic	0

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

		name). Values: “change” and “no change”	
Diabetes medications	Nominal	Indicates if there was any diabetic medication prescribed. Values: “yes” and “no”	0
24 features for medications For the generic names: metformin,repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone,	Nominal	the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter, “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed	0
Readmitted	Nominal	Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission	0

5. Feature Selection

We decide not to fill in the missing values because the variable weight covers nearly 97 percent of them. Approximately 40% of the missing values in the medical specialty and variable Payer code have been removed. There are very few missing values for the variables relating to race, diagnosis 1, 2, and 3, and gender. In comparison to other properties that we dropped, we chose to drop those of these attributes that had missing values. Every entry has the same value and refers to variables (drugs with names like examide and citoglipton). As a result, these cannot provide any discriminating or interpretative information for predicting readmission. After removing redundant attributes, finally we took 23 attributes for consideration. Each of the 23 attributes for 23 drugs (or drug combinations) in the dataset indicates if the drug has changed throughout the patient's current hospital stay. Changing a diabetic patient's medication after admission has been linked to a lower readmission rate, according to prior study. We obviously count all the adjustments performed for each patient and classify this as a new feature. The objective was to simplify the model regardless of whatever drug was altered and potentially discover a correlation with several alterations. String values were utilized for a number of items in the original dataset, including gender, race, medication changes, and each of the 23 drugs used. We interpret the variables as numeric variables to better reflect their nature and fit them into our model. For a sample, we altered "No" (no change) and "Ch" (changed) in the "medication change" feature to 0 and 1, respectively. The patient is regarded as seeking if they are readmitted to the clinic within 30 days. Actually, there are three possibilities for the variable: more than 30, 30 and No Readmission (refer fig 1). Only two categories (Fig 2)—readmission after 30 days and no readmission—are considered in our system's binary classification in order to make it more straightforward.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

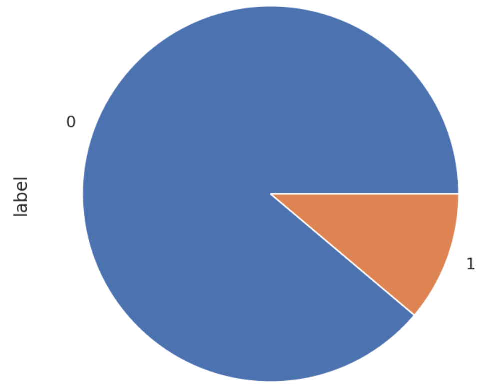


Fig 2 : Binary-classification of proposed system

In the dataset, the subject in issue had up to three diagnoses (primary, secondary, and additional). However, due to the fact that each of these contains 700–900 unique ICD codes, it is really challenging to include and comprehend them in the model. Therefore, we used this information to classify these diagnosis codes into 9 different types of illnesses. Circulatory, Respiratory, Digestive, Diabetes, Injury, Musculoskeletal, Genitourinary, Neoplasms, and Others are among these nine categories. All diagnoses, including primary, secondary, and additional illnesses, are categorized; however, only the primary diagnosis was employed in our model. In this paper, six machine learning algorithms are used to predict the readmission rate of diabetic patients. These six algorithms are Logistic regression, decision tree, random forest, K-NN, SVM, and ANN.

6. Proposed Hospital Readmission Reduction Technique

We have two cases in this technique:

- Dataset reduction using PCA with six ML algorithms
- ii) Dataset reduction without PCA with six ML algorithms

Fig.3 provides an illustration of Hospital Readmission Reduction technique.

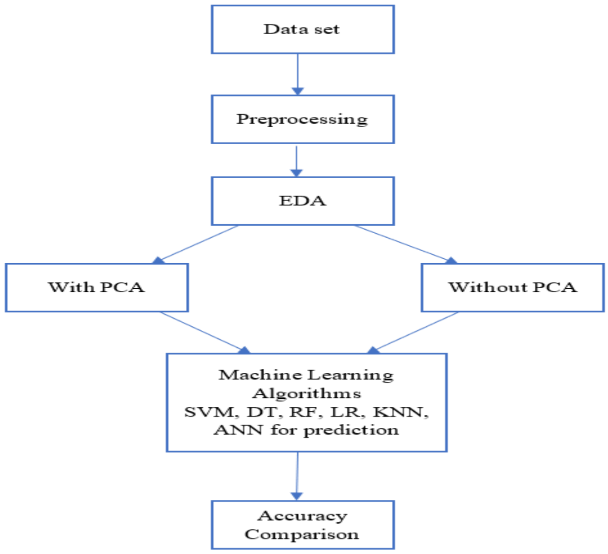


Fig. 3 Flow chart of the projected work

Apply one-shot encoding to the diabetes transformations dataset from the United States. Reduce the dimensionality of the modified data using PCA. Machine learning algorithms and neural networks are used to train the dimensionally reduced dataset. Evaluate the proposed model's effectiveness and performance in comparison to other methods. This work proposed six machine learning algorithms such as Artificial neural networks, K-NN, Decision Tree, Random Forest, Logistic Regression and SVM. On the US dataset, all of these algorithms were used. The data was separated into two categories: training data and testing data, which accounted for 70% and 30% of the total data, respectively. The main evaluation parameter that we used in this work was prediction accuracy. Equation 1 can be used to calculate accuracy. In this case, accuracy refers to the algorithm's overall success rate.

$$\text{Accuracy rate} = \frac{(TP+TN)}{P+N} \qquad (1)$$

Figure 4 shows confusion matrix interpreting true positives (TP), false positives(FP), true negatives(TN) and false negatives(FN)

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

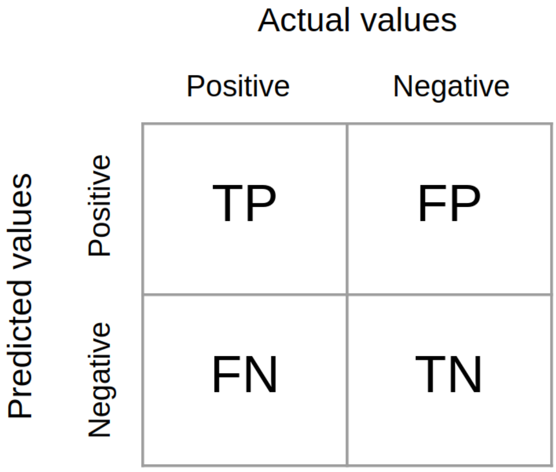


Fig. 4: Confusion Matrix

Then, TPR and FPR are defined as follows:

$$TPR = TP / TP + FN \quad (2)$$

$$FPR = FP / FP + TN$$

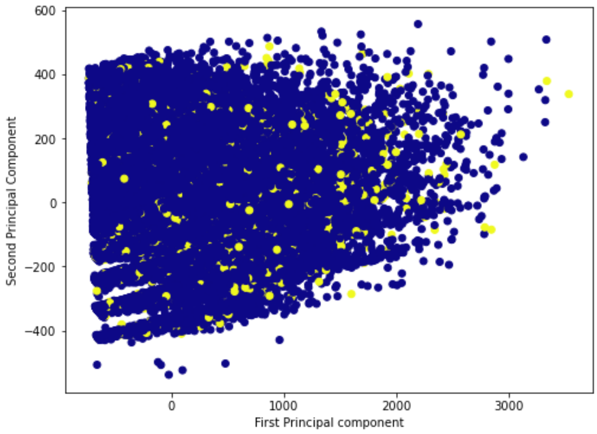


Fig. 5. Scattering plot for first principal component and second principal component

To forecast diabetes, various machine learning algorithms have been used, and the best results have been obtained. They are detailed below.

i) Logistic regression

Logistic regression is one of the simplest and extensively cast-off in machine learning techniques for two-class classification. Before loading the dataset, divide the available columns into dependent and independent variables. Using the function train test split (), divide the dataset. The three requirements are features, target, and test set size. The dataset is divided in half, 70:30, which means that 30% of the data will be used to test the models and 70% will be used to train the models. The confusion matrix is constructed by adding the total number of correct and incorrect guesses for each class. The confusion matrix is shown as an array object in Table 2. This matrix has a dimension of 2 by 2 due to the binary classification used in the model. The results show that 155 and 1482 are incorrect predictions, while 1674 and 142 are correct.

Table 2. The Confusion Matrix Of Logistic Regression

	Positive	Negative
Positive	1674	1482
Negative	155	142

The trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) for various parameters is provided by the Receiver Operating Characteristic curve (ROC). The percentage of observations for which a positive outcome was accurately anticipated is known as the true positive rate. The percentage of observations that are falsely expected to be positive is known as the false positive rate. The ROC curve for Logistic regression is exemplified in Fig. 5

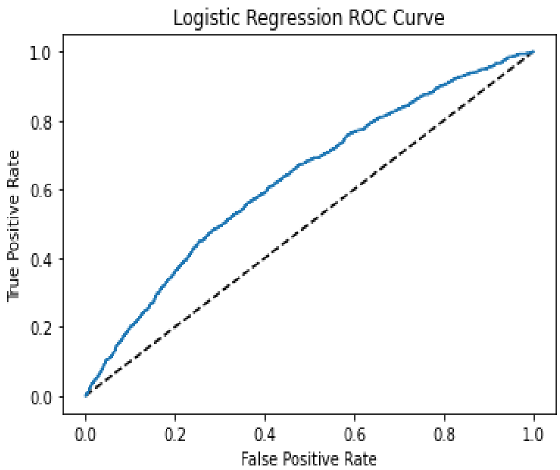


Fig. 6 ROC curve for Logistic regression

ii) Support Vector Machine (SVM)

When compared to other classifiers such as decision trees and logistic regression, SVM has extremely high accuracy. It employs a well-known kernel technique to manage nonlinear input spaces. The dataset contains 23 features as well as target data. Diabetes is classified in this data as either Positive or Negative. To better understand model performance, divide the dataset into a training set and a test set after loading it. To divide the dataset, use the train test split function (). The three requirements are features, target, and test set size. The random state can also be used to select records at random. The linear kernel of the SVC () function

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

is bypassed. Fit is then applied (). Table 3 depicts the confusion matrix of SVM. The scattered plot for component1 vs component 2 for readmission of patient status using the SVM algorithm is given in Fig.6.

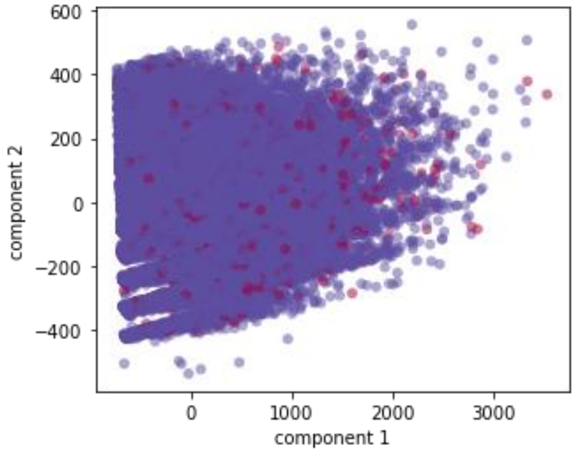


Fig. 7 Scattering plot for readmission status using SVM algorithm

Table 3. The confusion matrix of SVM

	Positive	Negative
Positive	10901	1041
Negative	238	176

iii). Decision tree algorithm

The libraries required to create a decision tree must be imported prior to loading the data set. To load the data, use pandas' read CSV () function. The various values of the confusion matrix will be correctly identified 2065 positive class data items. The model correctly identified 196 data elements from the negative class. The model correctly identified negative class data items from 1908. The model misclassified 224 positive class data points as negative class data points which are represented in Table 4. The ROC curve for the decision tree algorithm is illustrated in Fig. 7

Table 4. Confusion Matrix of Decision Tree

	Positive	Negative
Positive	2065	1908
Negative	224	196

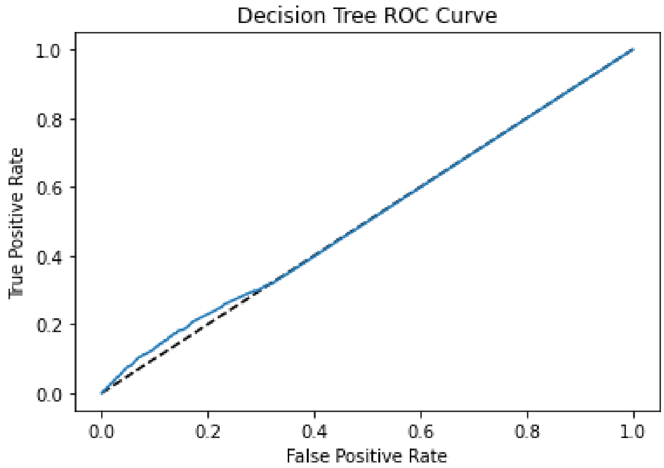


Fig. 8 ROC curve for decision tree algorithm

iv) Random forest

A random forest is a meta estimator that averages data as it fits various decision tree classifiers to various subsamples of a dataset in order to improve predicted accuracy and decrease over fitting. The amount of the sub-samples is controlled by the max samples option if bootstrap=True (the default) is specified; otherwise, each tree is constructed using the entire dataset. With the practice set (x,y), I constructed a forest of trees to help the estimator obtain parameters. The confusion matrix from Random Forest is presented in Table.5. Fig. 8 shows the ROC curve for the random forest algorithm.

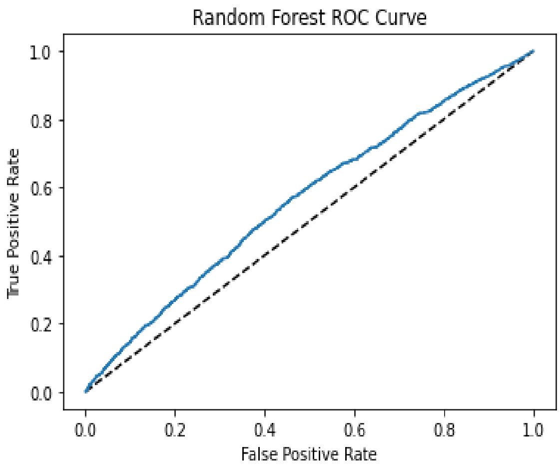


Fig. 9 ROC curve for the random forest algorithm

Table 5. Confusion Matrix of Random Forest

	Positive	Negative
Positive	1937	2036
Negative	208	212

v). KNN Algorithm

It is a classification technique that uses a similarity or distance measure to classify a new sample. The measure includes three distance measures: Euclidean distance, Manhattan distance, and

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Minkowski distance. KNN analysis starts with importing pandas, NumPy, and matplotlib.pyplot. Then the train and test data sets are imported. Finally, a confusion matrix is generated. The error values of trained and test data sets comparison using the KNN algorithm is illustrated in Fig. 9. The testing and training accuracy for the number of neighbors using the KNN algorithm is exemplified in

Fig. 10 and Fig.11 represents the ROC curve using the KNN algorithm.

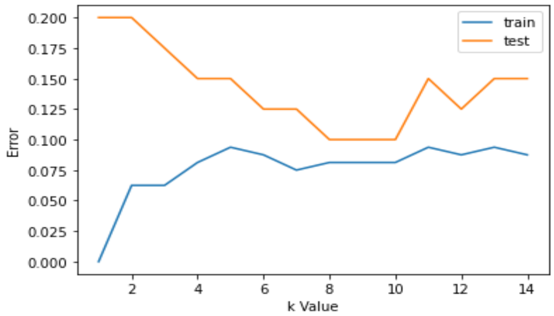
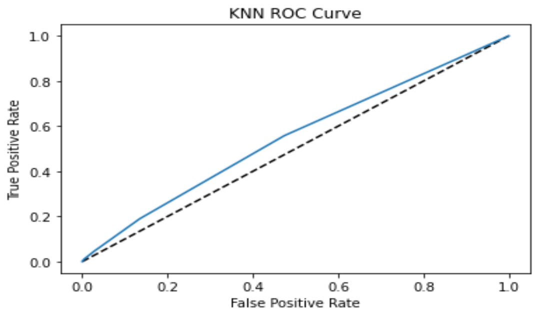


Fig. 10 Training and test data sets comparison using KNN algorithm



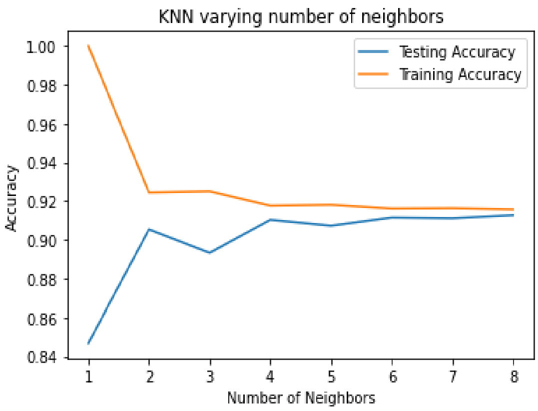


Fig. 11 Testing and training accuracy for a number of
neighbors using KNN

vi). ANN Algorithm

Supervised learning is used by artificial neural networks to categorise input data into the desired outcome. To start, a neural network folds and recognises data as an input to the system in diabetes categorization. The training algorithm is selected, and the system is taught using a specified training dataset. After training, ANN is tested to gauge the network's response, which updates whether the disease is correctly diagnosed or not. The ANN algorithm is used to create the graphic in Fig. 12 that compares the anticipated readmission data to the actual data.

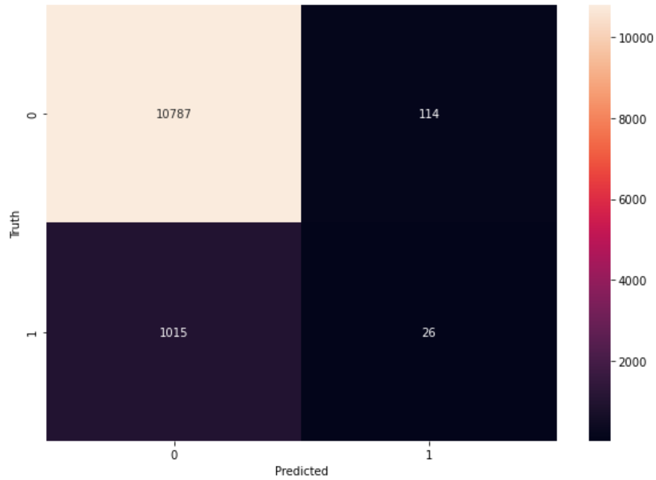


Fig. 12 Readmission Predicted with truth graph using ANN algorithm

7. Results and Discussion

The proposed work compiled in the Jupiter notebook tool to assess the precision of data prediction. Using internal cross-validation 5-folds, experiments are run. In this work, metrics of accuracy, F-measure, recall, precision, and ROC (Receiver Operating Curve) are used. These measurements are listed in Table. 6. Table 7 provides examples of the accuracy of prediction with and without the use of PCA. The suggested work is connected to previous, comparable efforts, and the findings are succinctly presented in Table 8.

Recall: Out of all the positive classes, how many instances were identified correctly

Recall = TP / (TP + FN) (3)

Precision: Out of all the predicted positive instances, how many were predicted correctly.

Precision = TP / (TP + FP) (4)

F-Score: From Precision and Recall, F-Measure is computed and used as metrics sometimes. F – Measure is nothing but the harmonic mean of Precision and Recall.

F-Score = (2 * Recall * Precision) / (Recall + Precision) (5)

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Table 6. Comparative Performance of Classification Algorithms

Algorithms	Precision	Recall	F-Measure	Accuracy
SVM	91.00 %	1.00 %	95.00%	91.49%
Random forest	91.11%	1.00%	95.00%	91.49%
Decision Tree	92.00%	95.00%	93.00%	87.00%
Logistic Regression	91.00%	1.00%	95.00%	91.00%
KNN	91.00%	1.00%	95.00%	91.49%

Table. 7 Prediction Accuracy with and without PCA

Algorithm	Before PCA	After PCA
Logistic Regression	0.9128	0.9128
Decision Tree	0.8686	0.8871
Random Forest	0.9125	0.9124
SVM	0.9141	0.9128
ANN	0.9190	0.9214
KNN	0.875	0.91

Table 8 Comparison of accuracy of various algorithms with the proposed system

Reference	Algorithm used	Dataset Used	Accuracy%
Ayman Mir et al[18]	SVM, KNN, simple CART, Naïve Bays	Pima Indian Diabetes Dataset	SVM - 79.13 %
Muhammad Azeem et al[19]	Six ML algorithms are used to predict disease.	PIDD data set.	SVM and KNN - 77%
Atik Mahabub et al [20]	Eleven different ML algorithms are used for the prediction of diabetes.	PIDD data set.	Precision f mean and Recall. Has to produce 86% accuracy for a voting classifier.
Sneha et al [21]	RF, SVM, NB, DT, KNN	UCI machine repository archive.ics.uci.edu-Diabetes (2500, 15)	Naviey bays algorithm 82.3%
Deepti et al [22]	NB, SVM, and DT	PIDD data set	Naviey Bays algorithm 76.3%
Amani Yahyani et al [23]	Compared conventional ML method with DL approach with CNN, SVM, RF, and CNN	PIDD	RF - 83.67% highest accuracy.
Sivaranjanis et al [24]	SVM and RF PCA - feature selection method.	PIDD	Random Forest - 83%
Proposed work	PCA is applied initially then machine learning algorithms are applied	US Health Facts Medical Database	ANN with PCA provides the highest accuracy of 92.12 %

8. Conclusion

Prediction research in healthcare has the potential to change how researchers and doctors interpret and act upon medical data. Here, we created and analyzed ML-based readmission prediction approaches to anticipate the readmission risks of diabetic patients for the 101766 records of diabetic patients taken into consideration. The model was developed using a total of 59706 records after data cleaning and standardization. This suggested solution uses six machine learning algorithms for predictive analytics. The primary goal of this work is to create diabetes detection classification systems. A variety of machine learning techniques are used to conduct classification on the dataset, with ANN achieving the greatest accuracy of 92.12%. When the five algorithms' performance indices were evaluated, the ANN model performed the best. Its applicability is more suited for predicting readmission since it has a larger area under the receiver operating characteristic curve (ROC) than the other techniques. The ANN model has achieved higher consistency when compared to other machine learning methods. It is possible to extend this research to predict the possibility of non-diabetics developing diabetes in the upcoming years.

References

[1] Y. Shang et al., "The 30-days hospital readmission risk in diabetic patients: predictive modeling with machine learning classifiers," BMC Med. Inform. Decis. Mak., vol. 21, no. 2, pp. 1–11, 2021.

[2] S.J.Kassin MT, Owen RM, Perez SD, Leeds I, Cox JC, Schnier K, Sadiraj V, "No TitleRisk factors for 30-day hospital readmission among general surgery patients," pp. 322–30, 2012.

[3] L. P. Stefan MS, Pekow PS, Nsa W, Priya A, Miller LE, Bratzler DW, Rothberg MB, Goldberg RJ, Baus K, "No TitleHospital performance measures and 30-day readmission rates," J Gen Intern Med, pp. 377–85, 2013.

[4] Report to the Congress promoting greater efficiency in Medicare. Washington, DC," 2007.

[5] V. K. Bhuvan, Malladihalli S., Ankit Kumar, Adil Zafar, "No TitleIdentifying diabetic patients with high risk of readmission," 2016.

[6] H. Lu, S. Uddin, F. Hajati, M. A. Moni, and M. Khushi, "A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus," Appl. Intell., vol. 52, no. 3, pp. 2411–2422, 2022.

[7] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, "Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction," Complex Intell. Syst., no. 0123456789, 2021.

[8] M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm," Heal. Inf. Sci. Syst., vol. 8, no. 1, pp. 1–14, 2020.

[9] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Med. Inform. Decis. Mak., vol. 19, no. 1, pp. 1–16, 2019.

[10] Veena Vijayan V. And Anjali C, "Prediction and Diagnosis of Diabetes Mellitus, A Machine Learning Approach" ,2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) 2015.

[11] P. Suresh Kumar and V. Umatejaswi, "Diagnosing Diabetes using Data Mining Techniques", International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017.

[12] Ridam Pal ,Dr. Jayanta Poray, and Mainak Sen, , "Application of Machine Learning Algorithms on Diabetic Retinopathy", 2017 2nd IEEE International Conference On Recent Trends In Electronics Information & Communication Technology, May 19-20, 2017.

[13] Dr. M. Renuka Devi and J. Maria Shyla, "Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, pp 727-730, 2016.

[14] Rahul Joshi and Minyechil Alehegn, "Analysis and prediction of diabetes diseases using machine learning algorithm": Ensemble approach, International Research Journal of Engineering and Technology Volume: 04 Issue: 10 | Oct -2017.

[15] Zhilbert Tafa and Nerxhivan Pervetica, "An Intelligent System for Diabetes Prediction", 4th Mediterranean Conference on Embedded Computing MECO – 2015 Budva, Montenegro.

[16] Prof. Dhomse Kanchan B. and Mr. Mahale Kishor M. "Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis". International Conference on Global Trends in Signal Processing, Information Computing and Communication 2016

[17] N. V. N. R. P. and K. C. Das M. Gopala Krishnamurthy, D. Dinakar, I. M. Chhabra, P. Kishore"Diabetic prediction analysis using Machine Learning algorithms" Engineering Vibration, Communication and Information Processing, vol. 478. Springer Singapore, 2019.

[18] A. Mir and S. N. Dhage, "Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare," Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018, pp. 1–6, 2018.

[19] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, "Prediction of diabetes using machine learning algorithms in healthcare," ICAC 2018 - 2018 24th IEEE Int. Conf. Autom. Comput. Improv. Product. through Autom. Comput., no. September, pp. 1–6, 2018.

[20] A. Mahabub, "A robust voting approach for diabetes prediction using traditional machine learning techniques," SN Appl. Sci., vol. 1, no. 12, 2019.

[21] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," J. Big Data, vol. 6, no. 1, 2019.

[22] Deepti Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 1578–1585, 2018.

[23] A. Yahyaoui, A. Jamil, J. Rasheed, and M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," 1st Int. Informatics Softw. Eng. Conf. Innov. Technol. Digit. Transform. IISEC 2019 - Proc., no. 2, pp. 1–4, 2019.