Readmission Prediction among Diabetic Patients Based on Expert System

Sushma Jaiswal[1], Priyanka Gupta[1], Avinash Kumar[2], Sushil Dohare[3], Bhgah Y. Adam[4], Hamza Abdullah M. Adam[4], Priyanshu Agarwal[5], Saptadeepa Kalita[5*]

[1] Guru Ghasidas Vishwavidyalaya Bilaspur (C.G.), India
[2]SITAICS, Rastriya Raksha University, Gujarat ,India
[3]Jazan University, College of Public Health & Tropical Medicine, Epidemiology Departmental, Jazan, Saudi Arabia
[4]Health Education and Promotion Department College of Public Health and Tropical Medicine, Jazan University , Saudi Arabia
[5]Department of Computer Science and Engineering , Sharda University,Uttar Pradesh, India
[*]Corresponding author: saptadeepakalita@gmail.com

doi: 10.48047/ecb/2023.12.si4.1011

Abstract

Diabetes Mellitus depicts a group of chronic metabolic disorders affecting more than 450 million people worldwide. Diabetes is a hazardous and incurable disease. The root cause of this health complication is yet to be ascertained. Many researchers have worked on this area to make it easier for medical professionals to examine it. In this study, recent literature was analyzed. It is to be accounted that different Machine Learning models have obtained better results and performs well in many diabetes-associated tasks. This research investigates the causes of readmission and hospital readmission within 30 days of deliverance amongst diabetic patients. The proposed work is implemented on the Diabetes Readmission dataset acquired from the UCI Repository in the area of medical science contains 55 attribute and 100000 records of the patients. The model obtains benchmarked score which is 98% for the stacking model and Long Short-Term Memory (LSTM) gives 98.5% precision. The framework assists in classifying whether the diabetic patient will be readmitted to the hospital within 30 days or not.

Keywords: Machine Learning, Classification, stacking, LSTM, Diabetes Mellitus.

1. Introduction

Diabetes is a chronic, incurable disease. It seems to be very difficult to eradicate this disease, so it is preferable that these diseases do not occur. It is critical to maintain a healthy lifestyle in order to avoid disease. Many people are affected when they are young, and many more when they are old. However, it is dangerous in both cases. Diabetes comes under the category of metabolic diseases termed by hyperglycemia introduced by deficiencies in insulin secretion, action, or both. Diabetes diseases become even more dangerous with increasing age because in what quantity, how much sugar is to be eaten, it’s hard to decide. To regulate the amount of glucose insulin has an important role in the body which is released by the pancreas. As stated by etiopathology of diabetes, there are three different forms of the disease based on the clinical institutions. They are classified as Type 1 diabetes, Type 2 diabetes, and gestational diabetes mellitus. People spend huge amount of money in Diabetes prevention. The Government launches many schemes for diabetes prevention to help the people. Based on the International Diabetes Federation (IDF) report, it is observed that approximately 450 million people suffer from this severe disease, which will increase to 493 million by 2045. This research work developed one of the robust and accurate sets of techniques for predicting diabetes readmission within 30 days [27][28]. Diabetes is commonly occurring in people due to augmentation of busy and bad lifestyles; the early recognizance of diabetes is the sole way to stay away from danger. The models have been developed with different classifiers, and after comparing them, the best model will be considered. Data preprocessing takes a very important place for developing the model [29][30]. Diabetes Prevention Program is initiated by India and other countries so that the people with diabetes and in the group of diabetic individuals in India and different evaluative countries are getting holds the advantages of the schemes. It is vital for a person suffering from diabetes requires maintaining of their blood glucose levels in a normal range. Moreover, hyperglycemia or hypoglycemia leads to short and long-term difficulty in micro vascular and macro vascular, incorporating neuropathy, nephropathy, retinopathy, heart stroke, cardiovascular disease, and peripheral vascular diseases [31][32].

This work's the major contributions of this paper are as follows.

- This work tries to produce an efficient method for distinguishing diabetes readmission using readmission dataset [33].
- This work also tries to create predictive classifiers and models to investigate real diabetic patients' readmission rates [34].
- This work vitally put various parameters that could be used to identify if there is need of readmission or not [35].

The purpose of this study is to investigate and classify whether a patient is admitted to the hospital or not, as well as the efficacy of standard care for outdoor diabetes patients on 30-day hospital readmission rates [36][37].

2. Literature Review

Machine Learning (ML), Artificial Intelligence (AI) and Deep Learning (DL) have proven to be some favorable approaches with optimistic results. According to the World Health Organization (WHO) in 1980 the number of diabetic people recorded is 108 million and in 2014 it enhanced with 422 million. As the medical sector produces a vast amount of useful data, including patient records, electronic medical records of diabetic patients, information on diagnoses and treatments, etc. This can be used as a crucial resource for knowledge extraction that will facilitate systems for cost reduction and decision-making. Several authors contribute their knowledge in healthcare area such as diabetes prediction, classification of COVID 19 patients, breast cancer prediction, and heart disease classification[1]. The author suggested a new LASSOQ model, which ensures better glycemic control than standard formula. It gives an expert system for diabetes prediction. [2] The author focused on deep convolutional neural network (DCNN) and achieved 86 % accuracy and 0.91 areas under curve result. [3] In this section the author introduced convolutional recurrent neural network (CRNN) model for glucose level forecasting using Silico dataset and a clinical dataset. Specially used for capture patterns of features of multi-dimensional time series data. The moderate RNN have the ability to analyzing the prevenient sequential data and producing the predictive BG. The proposed CRNN method represents surpassing result in forecasting BG levels (RMSE and MARD) in the silico and clinical observation [4]. This paper implemented two novel feature extraction techniques in order to find the best risk-factors and fed the features into a machine learning model for prediction of type 2 diabetes. The recommended methods have been defined and gathered the clinical data over a long period of time, known as the San Antonio Heart Study. The framework results in achieving 95.94% accuracy in predicting whether the person will get affected by type 2 diabetes in upcoming 7–8 years or not [5]. This research work used Pima Indian Diabetes dataset as well as a regional dataset from Bombay. The classification approaches followed for prediction are logistic regression, naïve bayes, KNN, ID3 DT and C4.5, DT. Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO) technique for feature reduction and found that PCA outperform better [6]. This work developed Support Vector Machine (SVM) framework for predicting diabetes. Techniques like chi-squared test, extra trees along with feature selection techniques are used that gained 83.20% accuracy, 87.20% sensitivity and 79% specificity [7]. It focused on DL technique convolutional neural network Long Short-Term Memory (CNN-LSTM) and PID dataset is utilized for prognostication of diabetes mellitus. The overall accuracy achieved by this model was 68-74%. Fathi et al. [8] developed a decision support system for Type 1 diabetic person. The framework equipped properly medical records from 15 contributors with physiologically workable licensed model parameters [9]. Multiple K-nearest neighbor algorithms was proposed to predict postprandial glucose profile that occurs because the nominal therapy and suggested a correction to proper timing and/or quantity of the meal bolus. This model has been validated based on the adult in silico population of the UVA/PADOVA Type I diabetes simulator [10]. A Chinese dataset from 2009 to 2015 was used in this study and achieved 58.24% F1 score, sensitivity 91%, and G-mean 86.69%. It shows the importance of pre-processing for the training a machine learning based model [11]. In this framework the author focuses on developing a statistical model for diabetes prediction in order obtain better classification accuracy score compared to others. The various features are extracted by PCA and then a linear regression method is implemented on these newly formed features. This classifier succeeded to achieve an accuracy score of 82.1% for prediction of diabetes mellitus [12]. The author uses two datasets: first is 130-us and the second is Pima Indian Diabetes (PID) dataset. These two datasets are retrieved from UCI machine learning repository. In case of PID dataset the model gives better accuracy of 93.62% whereas for 130-US it gives 88.56% accuracy [13]. A stacked auto encoder, Deep Learning framework was developed in this paper for classification of Type 2 Diabetes. The framework is implemented on PID Dataset contains 768 records and 8 features. The author achieved accuracy of 86.26% for prediction of diabetes mellitus [14-15]. This study proves stacking meta classifier can lead to achieve better accuracy of 93.62% in case of PID dataset. In case of large dataset such as 130-us hospital the model based on ensemble method yields 88.56% accuracy. Thus, from the comparison it is seen that the model achieves higher accuracy in ensemble method in comparison to single prediction algorithm [16]. Various ML models are studied and proposed DT-CART-Boosting yields better accuracy of 62.9% in comparison to DT-Chaid, NB-TAN, NN Bagging, DT-CART-Boosting, Ensemble. considered the diabetes Readmission dataset from UCI machine learning warehouse. When formulating a strategy to relatively low readmissions, the measurement of HbA1C in diabetic patients can be a useful indicator of readmission rates [17]. Implemented LR, MLP, NB classifier, decision trees, and SVMS; SVM gives 95% accuracy [18]. To rearrange the data, variables were converted from categorical to binary (0/1). This work concluded that patients suffering from diabetes may be at higher risk for early readmission due to poor health literacy, health system failure, as well as loss of control over illness [19].

3. Methodology

Diabetes mellitus is a widely known disease; classification of diabetes mellitus readmission has been effectuated using a variety of methods. After familiarization, the data collection and setting the methodology is explained explicitly in this section. This work has been carried out using several components that includes preprocessing, feature selection and using different types of machine learning models. Moreover, an ensemble model has been developed to enhance the outcomes achieved by the classifiers. The framework of the proposed model is shown in Fig. 1. The diabetes patients Readmission prediction has been carried out in various stages which has been covered in below subsections [38][39].

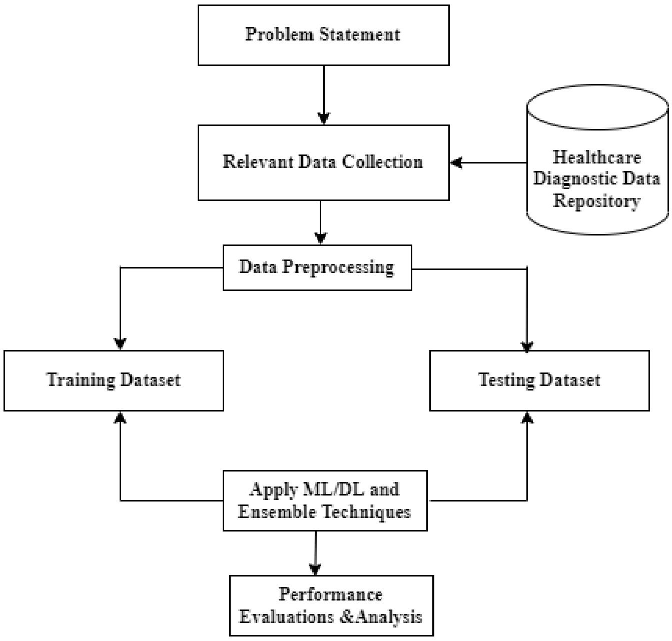


Figure 1 The framework of the proposed model

3.1 Dataset Acquisition

The Readmission Diabetes Dataset has been considered for this framework. The diabetes mellitus dataset is used as input for further processing. The Readmission Diabetes Mellitus (RDM) dataset is acquired from the UCI Repository in the area of medical science contains 55 attribute and 100000 records of the patients [20-22]. The dataset contains some missing values and some features of the dataset are not important so it has been discarded because of irrelevancy. The dataset contains categorical, numerical and free-text columns. The records of the columns will be discussed in more detail in below sections. The most useful columns here are readmitted also called target column, it keeps the details of the patient that the patient stays in the hospital within 30 days, greater than 30 days or not readmitted.

3.2 Feature Engineering

It is a process to formulate the appropriate input dataset, because all machine learning model use these input data to create output. In this segment, the fabrication of feature for predictive system has been carried out. In each fragment, novel variable has been added to the data frame and then sustaining of track id one for which columns is the desired columns of the data frame that is utilized as part of the predictive model features. After that division of this segment has been done into numerical features, categorical features and extra features. The numerical features in the readmission dataset and the missing numbers were accomplished with a question mark. It can be replaced with a nan representation. To handle this value, the work using Multivariate Imputation by Chained Equations *(*MICE) imputing technique. The most convenient type of features to take is numerical features for building the model because it is easy to use and produce better result, also these features do not need any transformation. The numerical columns utilized are time_in_hospital, num_lab_proocedures, num_procedures, num_medications, number_outpatient, number_emergency, number_inpatient, number_diagnoses.

It is observed that some records of the columns such as race, payer_code, and medical_speciality is missing. The mentioned columns are categorical then the best solution is to add another categorical type for unknown or missing data. The one-hot encoding technique is used to convert the categorical features in the dataset into numerical features; here a new column has to be created for each unique value. The value of the column value will be 1 if the unique value is present in the sample, if not, the value of the column will be 0. There are two extra features age and weight of the person. In this dataset these are categorical in the nature normally it is numerical only. One choice could be to generate categorical data, since there is a natural order to this value, it might make more sense to transform these to numerical data. Creating a variable is better than creating an ordinal feature. So that it can be known whether the value of weight is filled or not. The appearance of a variable might be predictive regardless of the null value.

3.3 Training / Validation/ Test samples

In this segment, the data has been analyzed and the features are generated from the categorical data. After performing these methods, the data is now splinted. Splitting the data means: how does the model behave for unseen data. Normally data is split into three parts that are Training Sample, Validating Sample and Test Sample.

Training samples: In this phase some parts of the dataset are taken for training, so that we can train the model.

Validation samples: In this phase some samples of data are taken from training data so that the model can be improved.

Test samples: In this phase some parts of the dataset are used to observe the performance of the classifiers. For designing the model, the splitting of the data is done as 70% train, 15% validation, 15% test. Now, here the aim is to investigate what percent of individuals who are hospitalized within 30 days. And it is termed as prevalence; all three groups have same level of prevalence. Here it could be see that with the help of data visualization techniques, the dataset has an imbalanced data where negative cases are more than positive cases and due to which the model is not able to produce the correct result. For this, the work has given more weight to positive cases. For this there are various approaches which are generally used – applying appropriate evaluation metrics for models generated using imbalanced data can be useful. For samples that are drawn from larger sample which is more dominant class: resampling can also be a good option, here two techniques are used oversampling the minority class or under sampling the majority class. Since there exist a lot of class imbalanced so this work chose subsample data. After sub-sample the data have been balanced and which the model will produce better result. Many machine learning classifiers do not give good results if the size of the dataset is (0-100, vs 1-100000), for dealing with that here the propose work could scale the data using scikit learn's standard scaler which removes the mean and scales to unit variance. Now build a scaler using all the training data. Original dataset shape is (0-54635, 1-5071) while new dataset shape becomes (0-54635, 1-54635) after sub-sample the data.

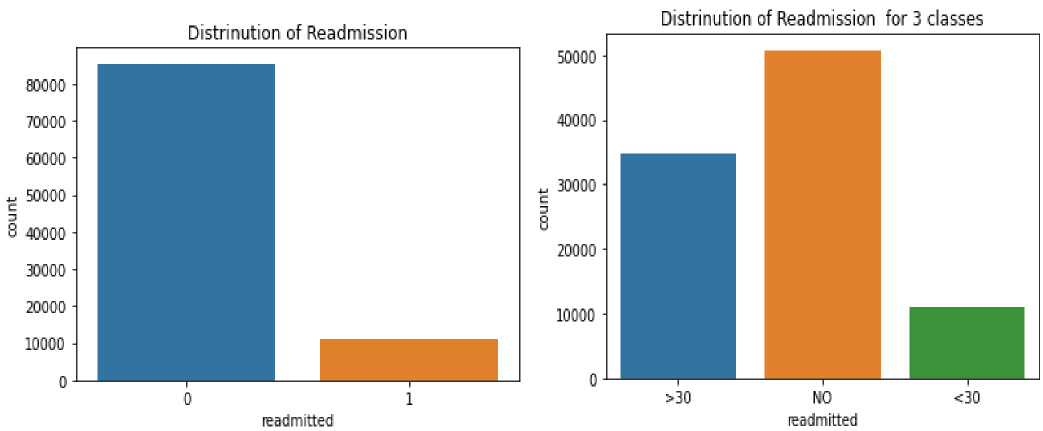


Figure 2 Distribution of the dataset

3.4 Model Consideration

In this context, this paper trains the ML models and use hyperparameter tuning techniques for optimizing them. Based on literature, four most known machine learning prediction techniques were deliberated. ML enables the artificial intelligence to learn and extract patterns from data to create well-organized and appropriate models.

Best model selected on the basis of performance and validation set. In this section, the paper would compare different machine learning models using hyperparameter tuning.

3.5 Model Evaluation

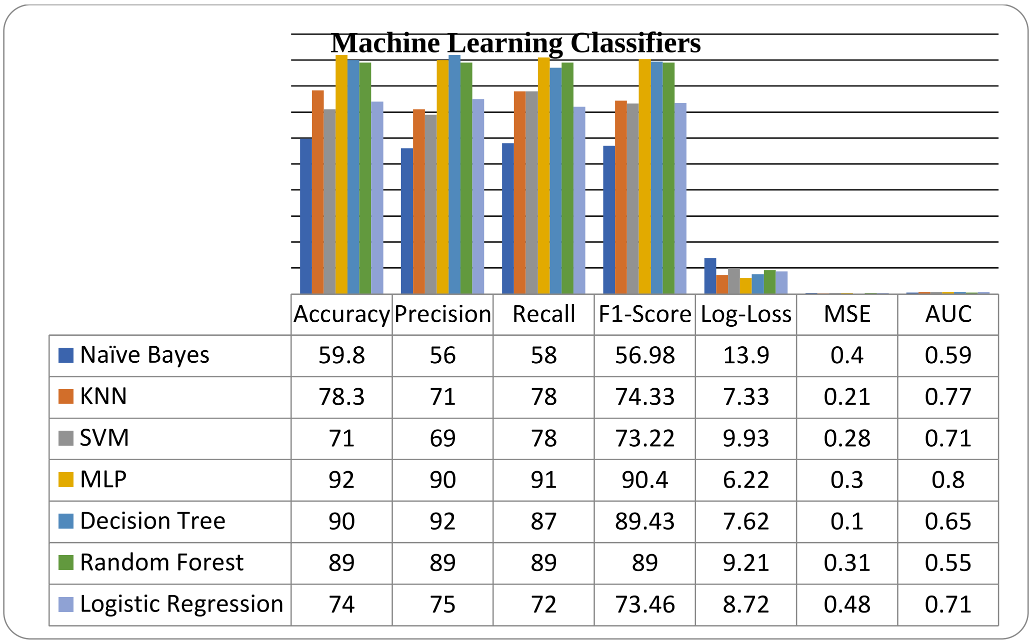
In this article various machine learning, ensemble techniques and deep learning models are utilized. ML models such as Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, Multilayer Perceptron, Decision Tree, Random Forest, Logistic Regression, is used for developed system that can predict whether they will be readmitted in 30 days or not. The combination of ML models such as, Boosting: Gradient Boosting, Adaboost and XGBoost model are consider and averaging techniques, and voting, stacking methods are used. In Deep Learning model LSTM and BLSTM models are evaluated.

Table 1 Sample performance analysis of the algorithms

Algorithm	**Result (Mean ± SD)**						
	Accuracy	Precision	Recall	F1-Score	Log loss	MSE	AUC
Machine Learning Classifiers							
Naïve Bayes	59.8±0.3	56±0.5	58±0.4	56.98	13.90±0.01	0.40±0.13	0.59±0.30
KNN	78.3±0.5	71±0.7	78±0.4	74.33	7.33±0.12	0.21±0.52	0.77±0.22
SVM	71±0.3	69±0.8	78±0.3	73.22	9.93±0.01	0.28±0.23	0.71±0.23
MLP	92±0.7	90±0.5	91±0.7	90.4	6.22±0.5	0.30±0.56	0.80±00.04
DT	90±0.3	92±0.5	87±0.3	89.43	7.62±0.3	0.10±0.34	0.65±0.32
RF	89±0.6	89±0.6	89±0.2	89	9.21±0.7	0.31±0.21	0.55±0.23
LR	74±0.5	75±0.8	72±0.5	73.46	8.72±0.2	0.48±0.41	0.71±0.27
Ensemble Learning Models (Average Method)							
LR+RF	94±0.3	97±0.5	91±0.6	93.90	2.80±0.04	0.06±0.02	0.81±0.11
RF+KNN+SVM	95±0.4	96±0.3	90±0.7	92.90	1.92±0.04	0.08±0.11±	0.91±0.03
Ensemble Learning Models (Max Voting Method)							
LR+RF	92±0.4	91±0.6	90±0.8	90.49	2.93±0.3	0.09±0.01	0.91±0.05
RF+KNN+SVM	96±0.5	97±0.4	95±0.3	95.98	2.50±0.6	0.07±0.02	0.90±0.05
Ensemble Learning Models (Stacking Method)							
LR+RF	84±0.8	96±0.5	72±0.5	82.28	5.40±0.31	0.15±0.4	0.79±0.20
RF+KNN+SVM	**98±0.5**	96±0.3	97±0.4	96.49	1.90±0.04	0.13±0.7	0.90±0.01
Boosting Techniques							
Gradient Boosting	91±0.7	90±0.3	93±0.6	91.47	3.5±0.4	0.08±0.5	0.8±0.05
Adaboost	87±0.5	84±0.5	85±0.5	84.49	5.33±0.64	0.8±0.3	0.7±0.04
XGBoost	94±0.4	94±0.2	94±0.6	94	2.2±0.05	0.07±0.21	0.9±0.08
Deep Learning Models							
LSTM	95±0.5	**98.5±0.3**	97±0.5	97.74	2.5±0.3	0.06±0.21	0.96±0.02
BLSTM	95±0.2	97±0.6	95±0.7	95.98	1.95±0.03	0.05±0.32	0.98±0.01

4. Result and Discussion

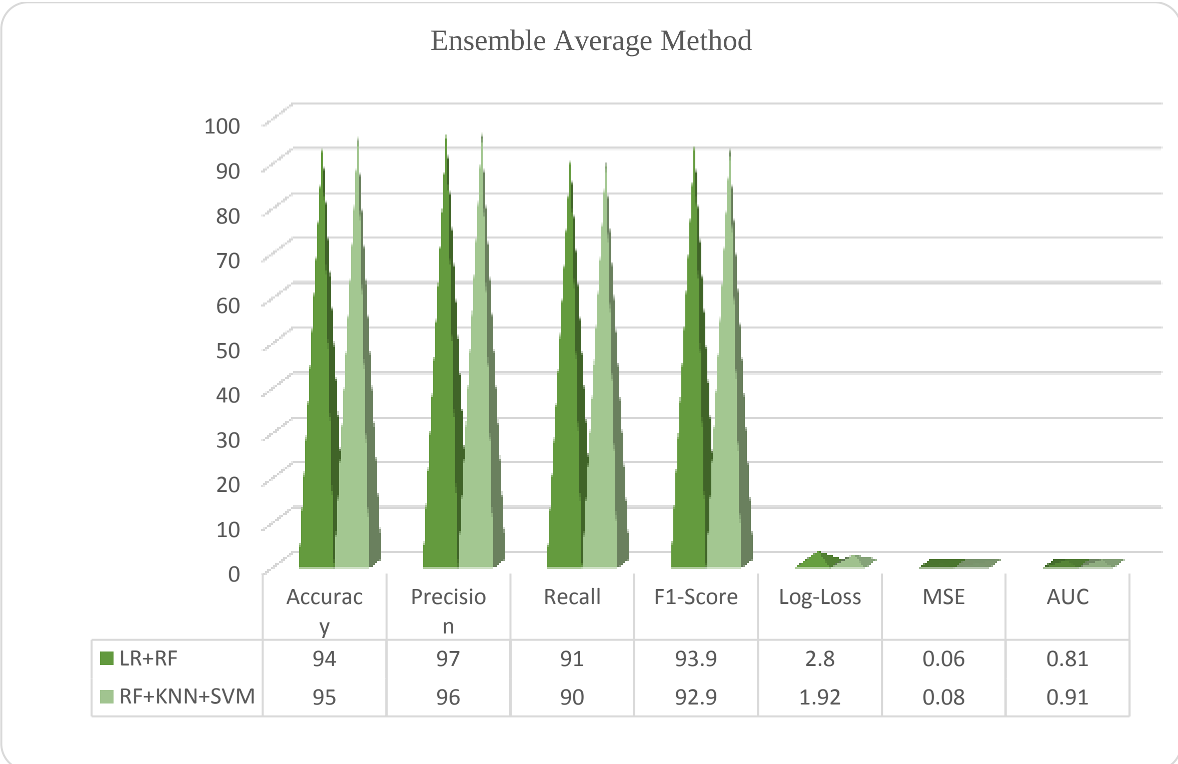
Based on the practical implementation it is observed that the presented ensemble techniques for diabetes prediction whether the person have diabetes or not. It can be used substantial tool for the disease diagnosis steps. It is clearly shown in the Table 1 that ensemble stacking method gives 98% accuracy while in deep learning model LSTM provide 98.5% precision. Hyperparameter tuning effect the results of the models because different model has their own statistics and they perform well when they utilize their resources. This work has also shown the mean and standard deviation of the results so that the output could be analyzed in a better way. Different ML and DL techniques are considered for the experiment of this research. Deep learning model gives satisfactory results with hyperparameter tuning.



	Accuracy	Precision	Recall	F1-Score	Log-Loss	MSE	AUC
Naïve Bayes	59.8	56	58	56.98	13.9	0.4	0.59
KNN	78.3	71	78	74.33	7.33	0.21	0.77
SVM	71	69	78	73.22	9.93	0.28	0.71
MLP	92	90	91	90.4	6.22	0.3	0.8
Decision Tree	90	92	87	89.43	7.62	0.1	0.65
Random Forest	89	89	89	89	9.21	0.31	0.55
Logistic Regression	74	75	72	73.46	8.72	0.48	0.71

Figure 3 Machine learning classifiers results comparison using different parameters

In Fig. 3 Multilayer Perceptron algorithms gives satisfactory result while Naïve Bayes gives less accuracy as compare to others. The hyperparameter are tuned according to the classifiers so the model gives better prediction rate.



	Accuracy	Precision	Recall	F1-Score	Log-Loss	MSE	AUC
LR+RF	94	97	91	93.9	2.8	0.06	0.81
RF+KNN+SVM	95	96	90	92.9	1.92	0.08	0.91

Figure 4 Result comparison of ensemble average method using different parameters

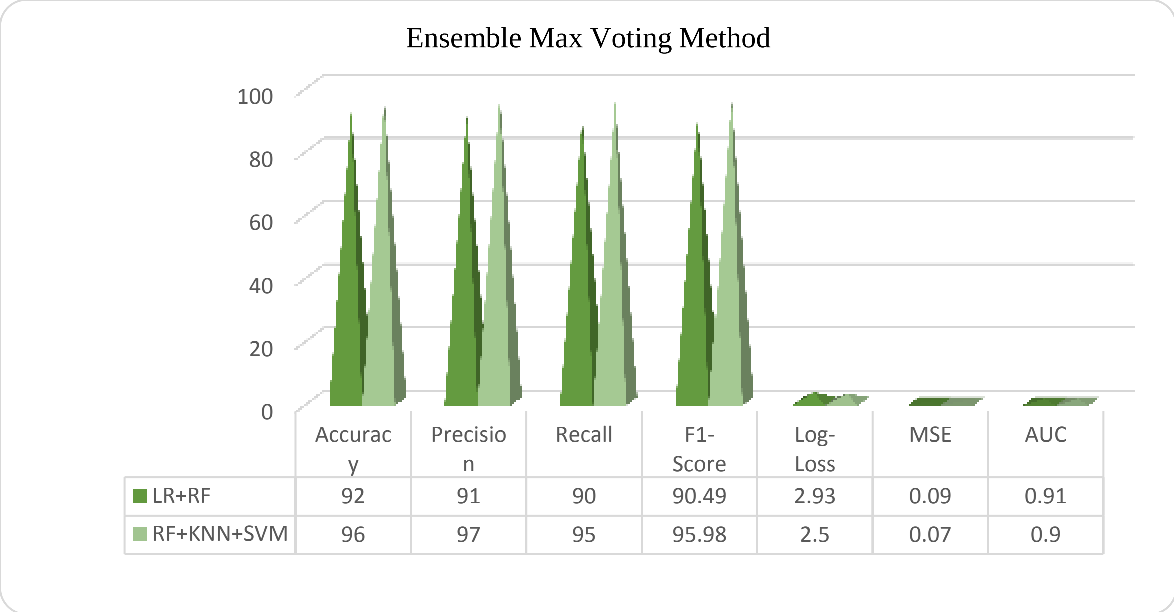


Figure 5 Result comparison of ensemble max voting method using different parameters

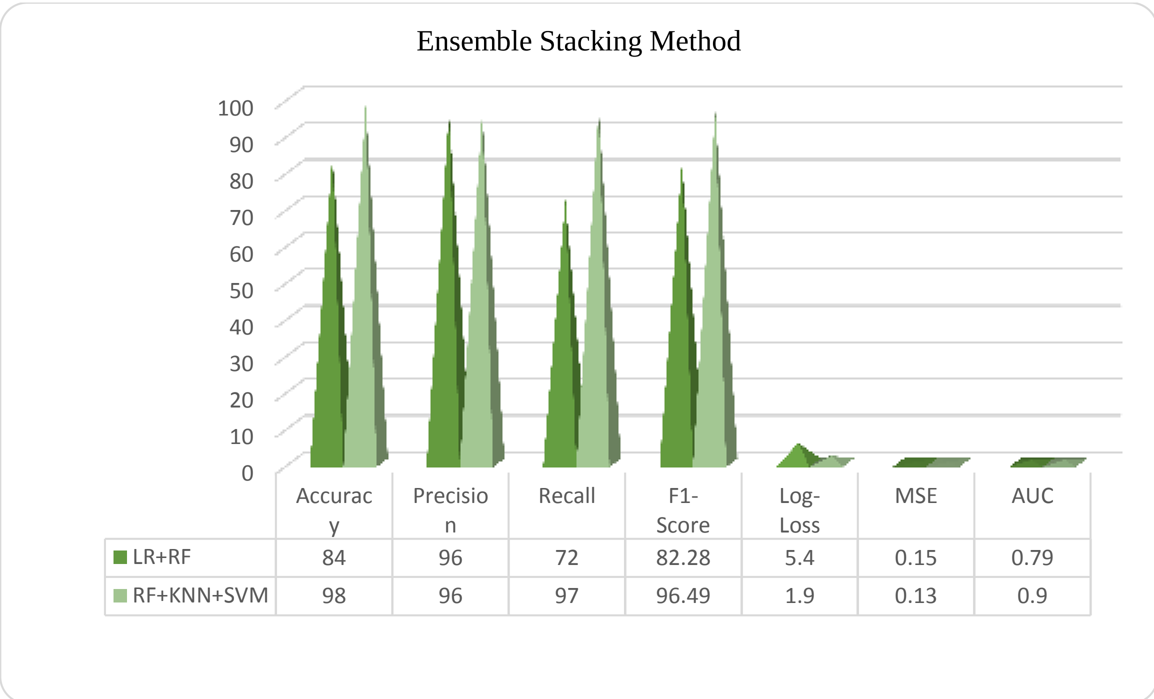


Figure 6 Result comparison of ensemble stacking method using different parameters

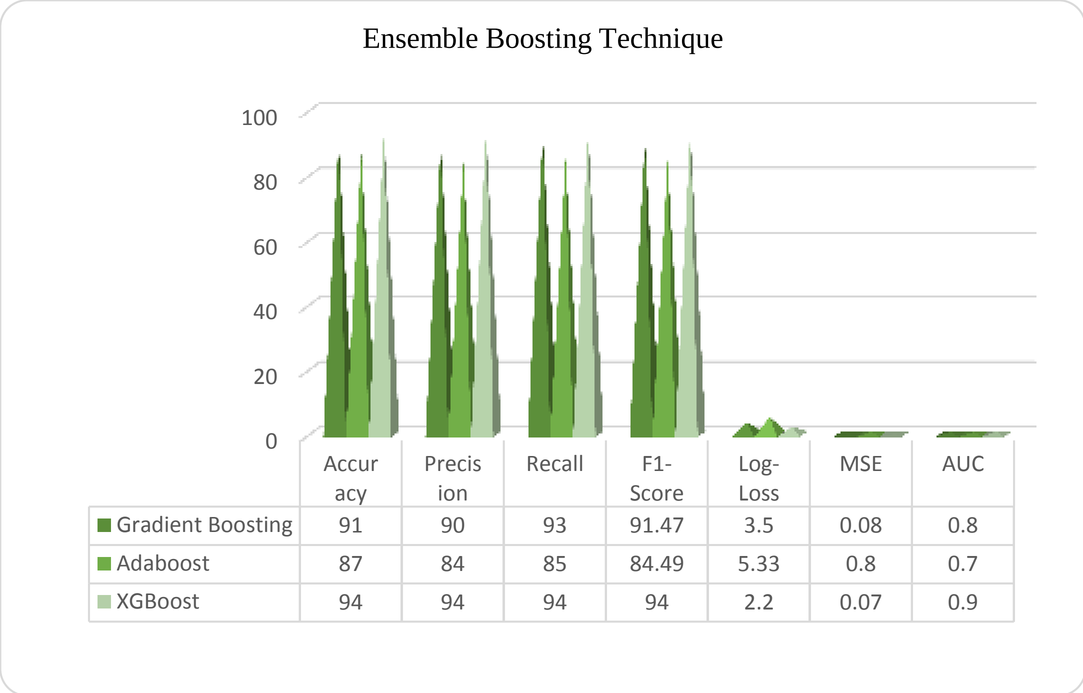


Figure 7 Result comparison of ensemble boosting techniques using different parameters

The Fig. 4 demonstrated the ensemble average method results. Ensemble approach average method gives better precision result 97% and also the mean squared error achieved by the ensemble average method is very less. In Fig. 5 Ensemble max voting method results are represented. The graph had shown that 96% accuracy and 90% area under curve. Fig. 6 represented Ensemble stacking method which gives 98% highest prediction accuracy as compare to other related work. In Fig. 7 the ensemble boosting algorithms are represented XGBoost model gives 94% accuracy and 0.07 mean squared error. Table 2 depict the comparison of different classifiers vs our proposed method.

Table 2 Comparative analysis of the classifiers state-of-art method

Authors & Year	Model	Result
Hammoudeh A. et al. [23] 2018	CNN	AUC 95% Accuracy 92% F1-score 92%
Cuong, Le et al. [24] 2021	XGBOOST	Accuracy 96%, Recall rate 92%, Precision 98% F1 score 90%
Goudjerkan, T et al. [25] 2019	MLP	Accuracy 95% Precision 93% Recall 99% AUC 95%.

Jouhari S. et al. [26] 2020	DNN	Accuracy 95.2% AUC97.4%
Proposed Work	**Stacking (RF+KNN+SVM)** **LSTM**	**Accuracy 98%** **98.5% Precision**

5. Conclusion

There are many records and the number of features in it is also very high in the readmission dataset. That's why it needs feature engineering process. It is very important to know which feature is important. Considering this as the base, it was found that many such features can also be swallowed. And in many records the number of missing values is very much high which is replaced by the mean method. Due to preprocess of readmission data, the model is producing very good result as compared to others. This implementation is extremely beneficial for healthcare professionals, doctors, clinical practitioners and patients. People don't understand the complications of diabetes diseases due to lack of literacy and awareness about their health. In this study, this paper has developed ML and DL model also put in practice the ensemble techniques. The ensemble techniques achieved 98% accuracy for stacking model and Long Short-Term Memory (LSTM) gives 98.5% precision. Deep Learning models and ensemble learning methods gives satisfactory result because the dataset is adequate in size and required pre-processing and feature engineering methods.

Conflict of interest

There is no conflict of interest as per best of knowledge.

References

[1] G. Noaro, G. Cappon, M. Vettoretti, G. Sparacino, S. D. Favero and A. Facchinetti, "Machine-Learning Based Model to Improve Insulin Bolus Calculation in Type 1 Diabetes Therapy," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 1, pp. 247-255, Jan., doi: 10.1109/TBME.2020.3004031, 2021

[2] Alex, S.A., Nayahi, J.J.V., Shine, H. et al. Deep convolutional neural network for diabetes mellitus prediction. Neural Comput & Applic , https://doi.org/10.1007/s00521-021-06431-7, 2021

[3] K. Li, J. Daniels, C. Liu, P. Herrero and P. Georgiou, "Convolutional Recurrent Neural Networks for Glucose Prediction," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 603-613, Feb., (2020), doi: 10.1109/JBHI.2019.2908488, 2020

[4] M. S. Islam, M. K. Qaraqe, S. B. Belhaouari and M. A. Abdul-Ghani, "Advanced Techniques for Predicting the Future Progression of Type 2 Diabetes," in IEEE Access, vol. 8, pp. 120537-120547, doi: 10.1109/ACCESS.2020.3005540, 2020

[5] Choubey D. K., Kumar, P., Tripathi, S. et al. "Performance evaluation of classification methods with PCA and PSO for diabetes", Netw Model Anal Health Inform Bioinforma 9, Vol. 5, https://doi.org/10.1007/s13721-019-0210-8, 2020

[6] Ramesh J, Aburukba R, Sagahyroon A., "A remote healthcare monitoring framework for diabetes prediction using machine learning", Healthc Technol Lett. 2 Vol. 8(3), pp. 45-57, doi: 10.1049/htl2.12010, 2021

[7] Sainte S. L. M., Aburahmah L, Almohaini R, Saba T. “Current Techniques for Diabetes Prediction: Review and Case Study”, *Applied Sciences*, Vol. 9(21):4604, https://doi.org/10.3390/app9214604, 2019

[8] Fathi AE, Kearney RE, Palisaitis E, Boulet B, Haidar A., “A Model-Based Insulin Dose Optimization Algorithm for People with Type 1 Diabetes on Multiple Daily Injections Therapy”, IEEE Trans Biomed Eng., Apr; Vol. 68(4), pp., 1208-1219, doi: 10.1109/TBME.2020.3023555, 2021

[9] E. M. Aiello, C. Toffanin, M. Messori, C. Cobelli and L. Magni, "Postprandial Glucose Regulation via KNN Meal Classification in Type 1 Diabetes," in IEEE Control Systems Letters, vol. 3, no. 2, pp. 230-235, doi: 10.1109/LCSYS.2018.2844179, 2019

[10] Wei, X., Jiang, F., Wei, F., Zhang J., Liao, W., and Cheng S., “An Ensemble Model for Diabetes Diagnosis in Large-scale and Imbalanced Dataset”, Proceedings of the Computing Frontiers Conference, pp.71-78, https://doi.org/10.1145/3075564.3075576, 2017

[11] H, Roopa & T, Asha. (2019). A Linear Model Based on Principal Component Analysis for Disease Prediction. IEEE Access. PP. 1-1, 10.1109/ACCESS, 2931956, 2019

[12] Alehegn, M., Joshi, R. R., & Mulay, P., “Diabetes Analysis and Prediction Using Random Forest, KNN, Naïve Bayes And J48: An Ensemble Approach”, Int. J. Sci. Technol. Res, Vol. 8(9), pp. 1346-1354, 2019

[13] Hochreiter S, and Schmidhuber J., “Long short-term memory”, Neural Comput 9(8):1735–1780, https://doi.org/10.1162/neco. 1997.9.8.1735, 1997

[14] Kannadasan, K. Damodar R. E., Venkatanareshbabu K., “Type 2 diabetes data classification using stacked autoencoders in deep neural networks”, Clinical Epidemiology and Global Health, Volume 7, **(** Issue 4, 530 – 535, 2019

[15] Alehegn, M., Joshi, R. R., & Mulay, P. Diabetes Analysis and Prediction Using Random Forest, KNN, Naïve Bayes And J48: An Ensemble Approach. Int. J. Sci. Technol. Res, 8(9)**,** 1346-1354, 2019

[16] H. N. Pham et al., "Predicting Hospital Readmission Patterns of Diabetic Patients using Ensemble Model and Cluster Analysis," 2019 International Conference on System Science and Engineering (ICSSE), 2019, pp. 273-278, doi: 10.1109/ICSSE.2019.8823441.

[17] M. Gowsalya, K. Krushitha and C. Valliyammai, "Predicting the risk of readmission of diabetic patients using MapReduce," 2014 Sixth International Conference on Advanced Computing (ICoAC), 2014, pp. 297-301, doi: 10.1109/ICoAC.2014.7229729.

[18] Samah Alajmani and Hanan Elazhary, “Hospital Readmission Prediction using Machine Learning Techniques” International Journal of Advanced Computer Science and Applications (IJACSA), 10(4), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100425

[19] Daniel J. Rubin, Kelly Donnell-Jackson, Ram Jhingan, Sherita Hill Golden, Anuradha Paranjape, Early readmission among patients with diabetes: A qualitative assessment of contributing factors, Journal of Diabetes and its Complications, Volume 28, Issue 6, 2014, Pages 869-873, ISSN 1056-8727, https://doi.org/10.1016/j.jdiacomp.2014.06.013.

[20] Rohollah Ramezani, Mansoureh Maadi, Seyedeh Malihe Khatami, “A novel hybrid intelligent system with missing value imputation for diabetes diagnosis, Alexandria Engineering Journal, Volume 57, Issue 3, 2018, Pages 1883-1891, ISSN 1110-0168, https://doi.org/10.1016/j.aej.2017.03.043.

[21] Shang, Y., Jiang, K., Wang, L. *et al.* The 30-days hospital readmission risk in diabetic patients: predictive modeling with machine learning classifiers. BMC Med Inform Decis *Mak* 21**,** 57 (2021). https://doi.org/10.1186/s12911-021-01423-y.

[22] Dwivedi, A. K. Analysis of computational intelligence techniques for diabetes mellitus prediction. Neural Comput & Applic 30, 3837-3845 (2018). https://doi.org/10.1007/s00521-017-2969-9

[23] Ahmad Hammoudeh, Ghazi Al-Naymat, Ibrahim Ghannam, Nadim Obied, "Predicting Hospital Readmission among Diabetics using Deep Learning", Procedia Computer Science, Volume 141, Pages 484-489, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.10.138, 2018

[24] Cuong, Le & Wang, Dong. A comparison of Machine Learning Methods to predict Hospital Readmission of Diabetic Patient. Studies of Applied Economics, 39. 10.25115/eea.v39i4.4475, 2017

[25] Goudjerkan, Ti'Jay & Jayabalan, Manoj. Predicting 30-day Hospital Readmission for Diabetes Patients Using Multilayer Perceptron. International Journal of Advanced Computer Science and Applications, 10. 268-275. 10.14569/IJACSA.2019.0100236, 2019

[26] Jauhari, Sarthak & Shukla, Shikhar & Tripathi, Surya. EmbPred30: Assessing 30-days Readmission for Diabetic Patients using Categorical Embeddings, 2020.

[27] Kalita, S., Singh, R. C., Abidi, A. I., Sawhney, H., & Noida, G. (2023). Dental Caries Grade Classification Based on Improved Very Deep Convolutional Neural Network. Journal of Data Acquisition and Processing, 38(2), 2794.

[28] Sawhney, Rahul, et al. "A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease." Decision Analytics Journal 6 (2023): 100169.

[29] Srivastava, Swapnita, et al. "An Ensemble Learning Approach For Chronic Kidney Disease Classification." Journal of Pharmaceutical Negative Results (2022): 2401-2409.

[30] Irfan, Daniyal, et al. "Prediction of Quality Food Sale in Mart Using the AI-Based TOR Method." Journal of Food Quality 2022 (2022

[31] Pramanik, Sabyasachi, et al. "A novel approach using steganography and cryptography in business intelligence." Integration Challenges for Analytics, Business Intelligence, and Data Mining. IGI Global, 2021. 192-217.

[32] Narayan, Vipul, et al. "To Implement a Web Page using Thread in Java." (2017).

[33] Paricherla, Mutyalaiah, et al. "Towards Development of Machine Learning Framework for Enhancing Security in Internet of Things." Security and Communication Networks 2022 (2022).

[34] Tyagi, Lalit Kumar, et al. "Energy Efficient Routing Protocol Using Next Cluster Head Selection Process In Two-Level Hierarchy For Wireless Sensor Network." Journal of Pharmaceutical Negative Results (2023): 665-676.

[35] Faiz, Mohammad, et al. "Improved Homomorphic Encryption for Security in Cloud using Particle Swarm Optimization." Journal of Pharmaceutical Negative Results (2022): 4761-4771.

[36] Narayan, Vipul, A. K. Daniel, and Pooja Chaturvedi. "E-FEERP: Enhanced Fuzzy based Energy Efficient Routing Protocol for Wireless Sensor Network." Wireless Personal Communications (2023): 1-28.

[37] Babu, S. Z., et al. "Abridgement of Business Data Drilling with the Natural Selection and Recasting Breakthrough: Drill Data With GA." Authors Profile Tarun Danti Dey is doing Bachelor in LAW from

Chittagong Independent University, Bangladesh. Her research discipline is business intelligence, LAW, and Computational thinking. She has done 3 (2020).

[38] Narayan, Vipul, et al. "Enhance-Net: An Approach to Boost the Performance of Deep Learning Model Based on Real-Time Medical Images." Journal of Sensors 2023 (2023).

[39] NARAYAN, VIPUL, A. K. Daniel, and Pooja Chaturvedi. "FGWOA: An Efficient Heuristic for Cluster Head Selection in WSN using Fuzzy based Grey Wolf Optimization Algorithm." (2022).