RESEARCH ARTICLE

SAMRIDDHI Volume 12, Issue 2, 2020 Print ISSN: 2229-7111 Online ISSN: 2454–5767

An Ensemble Approach for the Prediction of Diabetes

Sunit Kumar Mishra*, Arvind Kumar Tiwari

Kamla Nehru Institute of Technology, Sultanpur, Sultanpur-228118, Uttar Pradesh, India

Abstract

Diabetes is a very common disease in the world. If diabetes is detected in the early stage, it can be cured easily. Several machine learning techniques are available to predict diabetes in an earlier stage using a data set. This paper presents a review of several machine learning-based methods to predict diabetes. This paper provides the comparative analysis of Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, an ensemble of CNN and LSTM, and an ensemble of CNN and BLSTM to predict diabetes. This paper proposed an ensemble approach of CNN and LSTM to predict diabetes and provides an accuracy of 97.14%, precision of 97.30%, recall of 96.30%, F1-score of 96.79%, and AUC value of 0.97. The comparative analysis shows that the performance of the proposed approach is better in comparison to Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, ensemble of CNN and LSTM, and ensemble of CNN and BLSTM for the prediction of diabetes.

Keywords: ANN, CNN, LSTM, BLSTM, SVM, AUC.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2020); DOI: 10.18090/samriddhi.v12i02.11

Corresponding Author: Sunit Kumar Mishra, Kamla Nehru Institute of Technology, Sultanpur, Sultanpur-228118, Uttar Pradesh, India, e-mail: suniteng@gmail.com

How to cite his article: Mishra, S.K., & Tiwari, A.K. (2020). An Ensemble Approach for the Prediction of Diabetes. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 12(2), 122-129.

Source of support: Nil

Conflict of interest: None

Introduction

Diabetes is a common disease in our society. Every third person is affected from this serious disease. This is caused by irregular lifestyle, bad eating habits, lack of exercise, and pregnancy. In the human body blood, sugar level is controlled by the insulin hormone released by the pancreas. When due to any reason, secretion of insulin hormone becomes irregular, blood sugar level also affected. In this way a person may be affected by diabetes. The patients affected by diabetes can be cured by regular exercise and by adopting a healthy lifestyle. To control blood sugar levels some medicine may be given or insulin may be given explicitly. To know whether a person is affected by diabetes, some diagnosis is required. If we came to know about the disease in early stage, we may prevent this harmful disease. For early stage, prediction machine learning techniques have been used.[1] Machine learning techniques learn from data set to predict outcomes. Some data is used as training data which is used to train and then we can perform prediction using test data.[2] For early-stage diabetes prediction, various researchers have been used Support Vector Machine,[3] Naive Bayes,[4] Artificial Neural Network,[5] Decision tree,[6,7] K nearest Neighbour,[8] Long Short Term Memory (LSTM).[9]

Related Work

In the literature, various researchers have been proposed machine learning approaches for the prediction of diabetes. In paper,[10] the authors have been proposed SVM-based approach by using the data set of disease in Saudi Arabia to observe obesity and predict diabetes chances in a person. The authors of paper[11] have been proposed a diabetes prediction model based on boosting algorithms. They performed non-parametric testing using two algorithms, Adaboost and Logitboost on test data of 35669 individuals and got area under characteristics curve 0.99. The authors of paper[12] worked on the concept of classification. They used Pima Indians data set and obtained a precision value of 0.770 and recall value 0.775.The authors of paper[13] proposed a prediction model that used a simple K-means algorithm and C4.5 algorithm using Pima Indians diabetes data and achieved an accuracy of 93.5%.The authors of paper[14] have been proposed SVM based model and got an accuracy of 85%. The authors of paper[15] have been proposed logistic regression on separate heart, breast cancer, and diabetes data sets and got an accuracy of 80.77%. Authors of paper [16] performed classification technique, decision tree J48, and achieved Area under ROC is 0.98.The authors of paper[17]used DT, SVM and NB classification methods on Pima Indians Diabetes data sets. NB gave an accuracy of 76.30%. Authors of paper[18] proposed that deep ANN gave 95.14% accuracy using a data set of 9483 diabetes patients. Authors of paper[19] used Pima Indian data set (PIDD) and ANN and achieved an accuracy of 92%. A paper[20] proposed SVM on PIDD and achieved an accuracy of 89%. The authors of paper[21] proposed a combination of Logistic Regression and Random Forest classifier, which gave an accuracy of 94.25% and AUC 0.95 on data set taken from NHE Survey of 6561 respondents. A study[22] used OCTA image database with a Logistic regression-based model and found 95.01% sensitivity. In a sydy authors[23] used a multilayer feed-forward network to predict diabetes on PIDD. The authors of paper[24] proposed the Gradient Boosting-based method for predicting and diagnosing future diabetes risk and got the accuracy 86%. Researchers[25] used the backpropagation algorithm with Pima Indian diabetes data set and achieved an accuracy of 91%.The authors of paper[26] used random forest method on ICSM data set and got accuracy of 77.7%. The authors of paper[27] proposed ANN model with large clinical data set and achieved accuracy of 94%. The authors of paper[28] proposed logistic regression technique on the data set maintained by NIDDK Diseases and achieved 78% accuracy.The authors of paper[29] proposed an ensemble learning approach with four different data sets and achieved accuracy of 75.78%. The authors of paper[30] considered two predictive models support vector regression and Gaussian process, one for short term glucose control and second for long term glucose control using the data set consists of 15 diabetic patients. The authors of paper[31] have been performed extraction of rules from SVM using ensemble learning approach on CHNS data and got precision of 94.2% and recall 93.9%. The authors of paper[32] proposed deep learning methods on the concentration of glucose, on clinical data. Two layer networks can be used. First layer performs prediction whereas second layer is used for correctness. The authors of paper[33] proposed rule based classification system to predict diabetes using PIDD for diabetes and got accuracy of 81.97%. The authors of paper[35] used a deep learning network for Hypoglycemia: Euglycemia: Hyperglycemia patients and got highest accuracy of 79.97%, 81.89% and 62.72% respectively. The authors of paper[36] proposed CART(classification and regression trees) algorithm on PIDD and got accuracy of 93.6%. The authors of paper[37] used multi layered feed forward neural network on clinical data set of 10 patients, for real time predictions to predict the rise or fall in glucose level in every 90 minutes. The authors of paper[38] proposed neural network on PIDD data set and achieved an accuracy of 83.3%. The authors of paper[39] proposed two algorithms: Artificial neural network, to predict the rate of fasting blood sugar, and the second is decision tree take decision based on symptoms. The algorithms were applied to the clinical data set of 500 patients and improved 84.8% with feature extraction. The authors of paper[40] proposed LMT (logistic model tree) based classification techniques for the prediction of diabetes in a patient in early-stage, on clinical data and got the accuracy of 96.38%. The authors of paper[41] proposed a semi-supervised machine learning algorithm Laplacian support vector machine on Pima Indians data set and achieved accuracy of 82.29%. Authors of paper[42] have been compared machine learning techniques like KNN and naïve Bayes. Authors of paper[43] have been developed two models to predict prediabetes one is Artificial Neural Network and the other is SVM. Data is taken from KNHANES. They got area under the curve using SVM is 0.731 and using ANN is 0.729. The authors of paper[44] proposed multilayer neural network with a backpropagation model for the prediction of diabetes on 6142 patients and get the sensitivity 86.04%. The authors of paper[45] used deep learning techniques to diagnose diabetes and got accuracy 84.95%, specificity 83.45%, sensitivity 86.44% and AUC of 0.8540[46] Used Rapid-I's to analyze Pima Indians Diabetes data set. They used an ID3 decision tree to predict diabetes with 80% accuracy[47] used an extreme learning machine algorithm that uses single-layer feed-forward neural network and points out future perspectives of ELM, and gave an accuracy of 77.63% on UCI data set. A study[48] perform the classification of HRV and diabetic signals by using long short-term memory and convolutional neural network or a combination of both to extract features of input HRV data which was treated as input to SVM and got the accuracy of 95.7%.The authors of paper[49] proposed a convolutional neural network to predict diabetes on the Brigham and Women's Hospital data set set and got AUC of 0.97. The paper[50] proposed a recurrent deep neural network (RNN) model on PIMA Indian diabetes data set and got an accuracy of 81%.

Computational Intelligence Techniques

Various methods are used for the prediction of diabetes. This paper proposed an ensemble approach of LSTM and CNN to predict the early stage of diabetes.

Long Short Term Memory (LSTM)

LSTM is another deep learning model used for sequential information proposed by the SeppHochreiter and JuergenSchmidhuber. LSTM is a RNN architecture that remembers values over arbitrary intervals. It is used to solve the problem of vanishing gradient problem. LSTM cell can be seen in Figure 1.

In the LSTM Model shown in Figure 1, it represents input gate, ft represents forget gate, ot represents output gate, σ represents sigmoid function, Wx represents weight

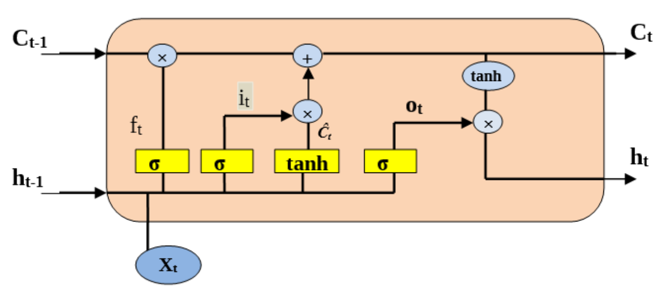

Figure 1: LSTM Model

for respective gate(x) neurons, ht-1 represents the output of previous LSTM unit, xt represents output at the current timestamp, bx represents biases for respective gates(x), Ct = Cell state at timestamp t and Ĉt = Candidate for cell state at timestamp t. In the first step of LSTM model, it is decided which information is to throw from cell state. This is decided by the sigmoid layer, which is also called "forget gate" layer. It takes h_{t-1} , x_t as input and generates an output between 0 to 1 for each cell state (C_{t-1}) numbers. A' 1' denote accept this and '0' denotes reject this.

$$f_t = \sigma (W_{fh} * h_{t-1} + W_{fx} * x_t] + b_f) \quad ...(3.1.1)$$

In the next step it is decided which new information is to be stored in the cell state. It consists of two layers, sigmoid layer, which is also called a input gate and the second is tanh layer. Input gate decides values to be updated. Tanh layer generates a candidate values vector $\hat{C}_t$. In the next step these two are combined to generate update for the state.

$$i_t = \sigma (W_{ih} * h_{t-1} + W_{ix} * x_t] + b_i) \quad ...(3.1.2)$$

$$\hat{C}_t = \tanh(W_{ch} * h_{t-1} + W_{cx} * x_t] + b_C) \quad ...(3.1.3)$$

Now the old value of cell state C_{t-1}, is updated to new value of cell state C_t. Multiply C_{t-1} by f_t the add $i_t * \hat{C}_t$ to it. This is new updated candidate value, scaled according to decided valued to update each state.

$$C_t = \sigma (f_t * C_{t-1} + i_t * \hat{C}_t] \quad ...(3.1.4)$$

In the last step it is decided that what will be the output. Final output depends upon cell state. Sigmoid layer decided the part of cell state, which will be the output. Then apply tanh and take the product of it by the sigmoid gate output to get the final output.

$$o_t = \sigma (W_{oh} * h_{t-1} + W_{ox} * x_t] + b_o) \quad ...(3.1.5)$$

$$h_t = o_t * \tanh(C_t) \quad ...(3.1.6)$$

Convolutional Neural Network (CNN)

The Convolutional Neural Network involves two operations named as convolution and pooling as feature extractors. The output of this sequence of operations is connected to a fully connected layer, same as a multi-layer perceptron. There are basically two kinds of pooling used, such as max-pooling and average-pooling. Max-pooling selects the maximum number of values in input feature map region of each step, and average-pooling selects the average number of values in the region. In literature various researchers have been used Convolutional neural network for the prediction of diabetes. When we use CNN for numbers instead of images, then we use 1-Dimensional array to represent the values. In order to perform classification, the data set is considered as NumPy arrays. Each row of the matrix shows the one set of features. We can say that each row is vector and represents a feature set. We use stride =1 so that each time window shifts by 1. The max-pooling selects maximum values of pooling window. Flatten is used to reduce the dimension of 3–D data in CNN layers so that this data is given as input to the dense layers. Dense layers works similar to simple neural network and a single output is obtained. In dense layers, 'Relu' is used as an activation function and in output layer, it uses sigmoid function (Figure 2).

Materials and Methods

Here, in this paper, the description of the diabetes data set, methodology, and the proposed approach to predict diabetes and the performance evaluation of the proposed method have been provided.

Data Description

In this paper, the PIMA Indian diabetes data set[51] was used which was taken from Kaggle(https://www.kaggle.com/uciml/pima-indians-diabetes-database).[52] It is made to predict diabetes in women more than 21 years of age. It contains eight attributes or input variables and one output variable. The attributes are as follows:

Pregnancies: It represents number of pregnancies of a woman. During pregnancy, the glucose level of women may increase which is called gestational diabetes. If women got a pregnant number of times, the gestational diabetes may leads to diabetes mellitus.

Glucose: It represents glucose concentration in blood. If glucose concentration in blood increases than a certain value then it may cause diabetes.

Blood Pressure: It represents BP(diastolic in mm Hg). Higher diastolic blood pressure increases the risk of diabetes.

BMI: It represents body mass index (weight (kg)/height (m)^2). It determines the obesity of the patient. Hence it is an important metric to predict diabetes.

Skin Thickness: It represents skin thickness (mm). In case of varying ratio of muscle mass and fat mass BMI is not adequate parameter to assess obecity which may lead to diabetes. Hence triceps skinfold thickness plays an impotant role to predict the patient may be diabetic or not.

Insulin: It represents serum insulin(mu U/mL). It is 2 hour serum insulin which indicates that how the body of a person respond on taking food.

Diabetes Pedigree Function: It is a function which determines the probability of diabetes on the basis of diabetic family history of a person.

Age: It is generally observed that person having age greater than 60 years are more prone to diabetes.

Performance Evaluation

In this paper the Accuracy, Precision, Recall, F1 Score and AUC are used to measure the proposed approach's performance. Accuracy of the models is determined by confusion metrics through *K*-Fold cross-validation. The Confusion matrix

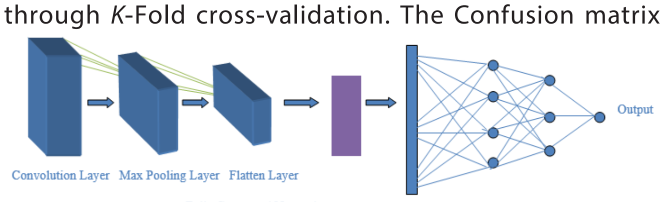

Figure 2: CNN Model

consists of True Positive (TP), True Negative (TN), and False-positive (FP) and False Negative (FP) where:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad \text{... (4.2.1)}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{... (4.2.2)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{... (4.2.3)}$$

$$\text{F1 score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{precision}} \quad \text{... (4.2.4)}$$

AUC-ROC Curve

AUC-ROC curve is used for measuring the performance of classification problems. ROC represents a curve of probability, and AUC defines the degree of separability. It defines the capability of a model to distinguish between classes. Higher the AUC value better will be the prediction.

ROC curve can be drawn by using TPR(true positive rate, also known as recall) and FPR(False positive rate). See Figure 3.

$$\text{TPR/Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{... (4.2.5)}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad \text{... (4.2.6)}$$

$$\text{FPR} = 1 - \text{Specificity} \quad \text{... (4.2.7)}$$

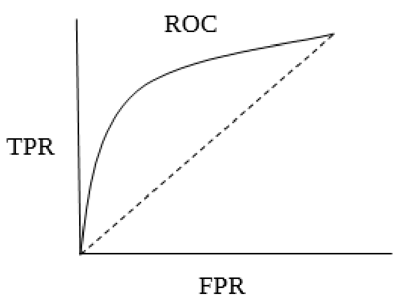

Figure 3: ROC-AUC Curve

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad \text{... (4.2.8)}$$

Methodology

In this paper, the data set is loaded on Jupiter Notebook. Then data cleaning is performed using Python queries. All the analysis is carried out on Jupiter Notebook on Anaconda. For the comparative analysis this paper uses the Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, ensemble of CNN and LSTM andensemble of CNN and BLSTM.

An Ensemble Approach of CNN with LSTM for the Prediction of Diabetes

This paper proposed an ensemble approach of CNN with LSTM for the prediction of diabetes.In the proposed model input is given in the form of 3-dimensional numpy array to the convolutional layer having kernel size one and stride is also one. After the convolutional layer Max pooling Layer is used to extract features by taking a maximum of pooling window. After the max-pooling layer LSTM layer is used, flatten layer is used to convert data into 1-dimensional data then after dense layers are used, which end on the output layer. In the hidden layers, activation function 'Relu' is used whereas in the output layer Sigmoid is used. The loss function used is 'binary_crossentropy,' and optimization 'Adam' is used See (Figure-4) of the proposed approach for the prediction of diabetes.

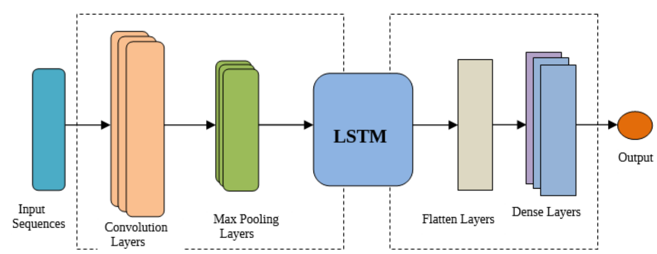

Figure 4: Proposed approach for prediction of diabetes

Table 5.1: Comparative analysis of Machine Learning-Based Approaches

ML Model	*Accuracy*	*Precision*	*Recall*	*F1 Score*	*AUC*
NAIVE BAYES	0.9035	0.8896	0.8991	0.8931	0.9608
RANDOM FOREST	0.9661	0.9540	0.9706	0.9617	0.9863
SVM	0.9283	0.9265	0.9134	0.9193	0.9741
KNN	0.9061	0.9201	0.8671	0.8915	0.9595
ANN	0.9675	0.9691	0.9569	0.9626	0.9670
LSTM	0.9531	0.9446	0.9512	0.9474	0.9534
BLSTM	0.9610	0.9600	0.9515	0.9552	0.9609
CNN	0.9661	0.9672	0.9578	0.9619	0.9658
CNN+LSTM (Proposed Approach)	0.9714	0.9730	0.9635	0.9679	0.9711
CNN+BLSTM	0.9674	0.9646	0.9630	0.9635	0.9673

Results and Comparative Analysis

This paper used the PIMA Indian diabetes data set taken from Kaggle and uploaded it onto the Jupitor Notebook. After data cleaning and pre-processing, the data is converted into a numpy array required for machine learning models. Here 10 fold cross-validation is used for the performance evaluation. For the comparative analysis, this paper uses the Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, the ensemble of CNN and LSTM, and ensemble of CNN and BLSTM for the prediction of diabetes (See Table-1).

From the comparative analysis, it is observed that the proposed approach and ensemble of CNN and LSTM performed better in comparison to Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, and ensemble of

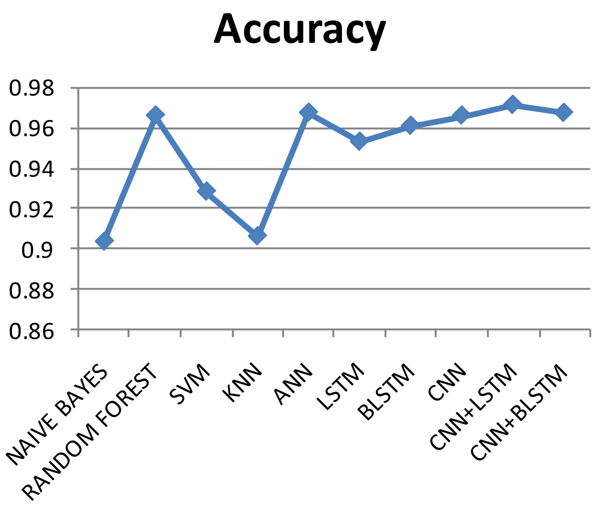

Figure 5: Comparison of accuracy of machine learning models

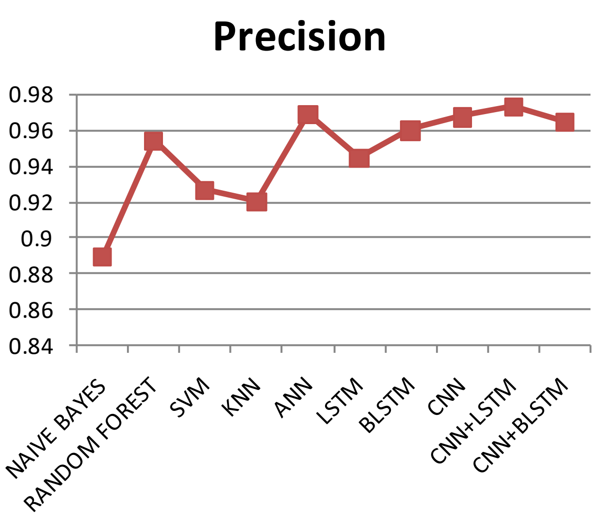

Figure 6: Comparison of precision of machine learning models

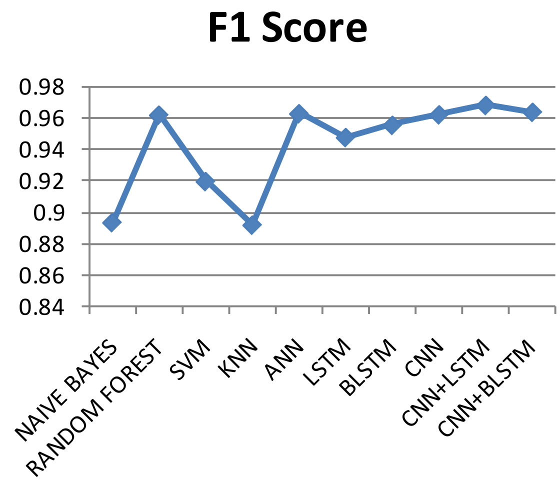

Figure 7: Comparison of F1-score of machine learning models

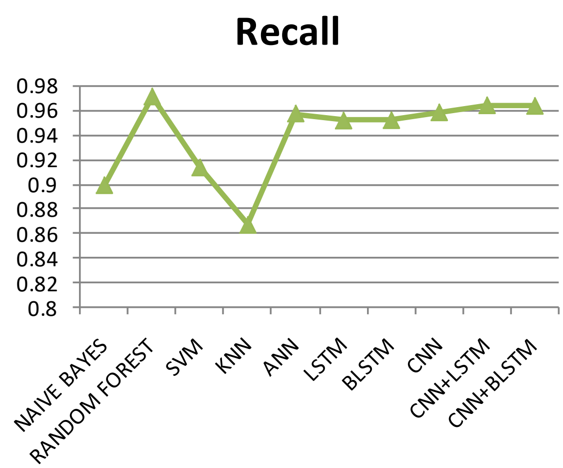

Figure 8: Comparison of recall of machine learning models

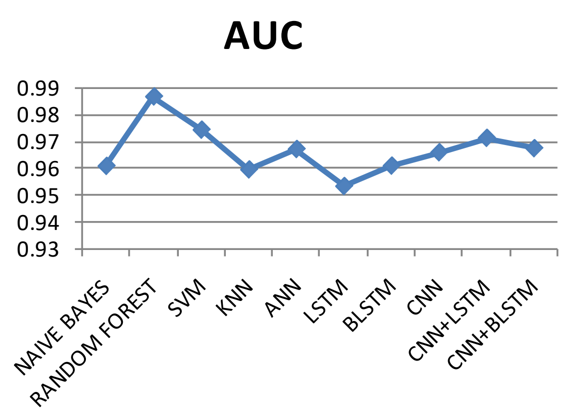

Figure 9: Comparison of AUC value of machine learning models

CNN and BLSTM. The proposed approach, an ensemble of CNN and LSTM provides an accuracy of 97.14%, precision of 97.30%, recall of 96.30%, F1 score of 96.79% and AUC of 0.97 for the prediction of diabetes (See Table 1 and Figures 5 to 9).

Conclusions

To predict the early stage of diabetes is one of the most challenging and important task. If diabetes is detected in an early stage, it can be cured easily. Several machine learning techniques are available to predict diabetes in an earlier stage using data set. This paper has been presented a review of several machine learning-based for the prediction of diabetes. This paper also provided the comparative analysis of Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, ensemble of CNN and LSTM and ensemble of CNN and BLSTM for the prediction of diabetes. This paper proposed an ensemble approach of CNN and LSTM to predict diabetes and provided an Accuracy of 97.14%, Precision of 97.30%, recall of 96.30%, F1-Score of 96.79%, and AUC of 0.97. The comparative analysis showed that the proposed approach performs better than Naive Bayes, ANN, SVM, KNN, Random Forest, LSTM, CNN, BLSTM, ensemble of CNN and LSTM and ensemble of CNN and BLSTM for the prediction of diabetes.

References

[1] Kerner, W., and J. Brückel. “Definition, classification and diagnosis of diabetes mellitus.” Experimental and Clinical Endocrinology & Diabetes 122.07 (2014): 384-386.

[2] Bottou, Léon. “From machine learning to machine reasoning.” Machine learning 94.2 (2014): 133-149.

[3] Vishwanathan, S. V. M., and M. NarasimhaMurty. “SSVM: a simple SVM algorithm.” Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN’02 (Cat. No. 02CH37290). Vol. 3.IEEE, 2002.

[4] Rish, Irina. “An empirical study of the naive Bayes classifier.”IJCAI 2001 workshop on empirical methods in artificial intelligence. Vol. 3.No. 22. 2001.

[5] Wang, Sun-Chong. “Artificial neural network.”Interdisciplinary computing in java programming. Springer, Boston, MA, 2003. 81-100.

[6] Safavian, S. Rasoul, and David Landgrebe. “A survey of decision tree classifier methodology.” IEEE transactions on systems, man, and cybernetics 21.3 (1991): 660-674.

[7] Pal, Mahesh. “Random forest classifier for remote sensing classification.” International Journal of Remote Sensing 26.1 (2005): 217-222.

[8] Liao, Yihua, and V. RaoVemuri. “Use of k-nearest neighbor classifier for intrusion detection.” Computers & security 21.5 (2002): 439-448.

[9] Sherstinsky, Alex. “Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network.” arXiv preprint arXiv:1808.03314 (2018).

[10] Aljumah, Abdullah A., Mohammed Gulam Ahamad, and Mohammad KhubebSiddiqui. “Application of data mining: Diabetes health care in young and old patients.” Journal of King Saud University-Computer and Information Sciences 25.2 (2013): 127-136.

[11] Chen, Peihua, and Chuandi Pan. “Diabetes classification model based on boosting algorithms.”BMC bioinformatics 19.1 (2018): 109.

[12] Mercaldo, Francesco, Vittoria Nardone, and Antonella Santone. “Diabetes mellitus affected patients classification and diagnosis through machine learning techniques.” Procedia computer science 112 (2017): 2519-2528.

[13] Patil, Bankat M., Ramesh Chandra Joshi, and DurgaToshniwal. “Hybrid prediction model for type-2 diabetic patients.” Expert systems with applications 37.12 (2010): 8102-8108.

[14] Kavakiotis, Ioannis, et al. “Machine learning and data mining methods in diabetes research.” Computational and structural biotechnology journal 15 (2017): 104-116.

[15] Kohli, Pahulpreet Singh, and ShriyaArora. “Application of Machine Learning in Disease Prediction.”2018 4th International Conference on Computing Communication and Automation (ICCCA).IEEE, 2018.

[16] Perveen, Sajida, et al. “Performance analysis of data mining classification techniques to predict diabetes.” Procedia Computer Science 82 (2016): 115-121.

[17] Sisodia, Deepti, and Dilip Singh Sisodia. “Prediction of diabetes using classification algorithms.”Procedia computer science 132 (2018): 1578-1585.

[18] Kowsher, Md, et al. “Prognosis and Treatment Prediction of Type-2 Diabetes Using Deep Neural Network and Machine Learning Classifiers.” 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019.

[19] Srivastava, Suyash, et al. “Prediction of Diabetes Using Artificial Neural Network Approach.” Engineering Vibration, Communication and Information Processing.Springer, Singapore, 2019.679-687.

[20] Kaur, Harleen, and Vinita Kumari. “Predictive modelling and analytics for diabetes using a machine learning approach.” Applied Computing and Informatics (2018).

[21] Maniruzzaman, Md, et al. “Classification and prediction of diabetes disease using machine learning paradigm.” Health Information Science and Systems 8.1 (2020): 7.

[22] Alam, Minhaj, et al. “Supervised machine learning based multi-task artificial intelligence classification of retinopathies.” Journal of clinical medicine 8.6 (2019): 872.

[23] Zhang, Yinghui, et al. “A feed-forward neural network model for the accurate prediction of diabetes mellitus.” International Journal of Scientific and Technology Research 7.8 (2018): 151-155.

[24] Birjais, Roshan, et al. “Prediction and diagnosis of future diabetes risk: a machine learning approach.” SN Applied Sciences 1.9 (2019): 1112.

[25] Durairaj, M., and G. Kalaiselvi. “Prediction of diabetes using back propagation algorithm.”International Journal of Emerging Technology and Innovative Engineering 1.8 (2015).

[26] Dagliati, Arianna, et al. “Machine learning methods to predict diabetes complications.” Journal of diabetes science and technology 12.2 (2018): 295-302.

[27] Donsa, Klaus, et al. “Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges.” Smart Health. Springer, Cham, 2015.237-260.

[28] Dwivedi, Ashok Kumar. “Analysis of computational intelligence techniques for diabetes mellitus prediction.” Neural Computing and Applications 30.12 (2018): 3837-3845.

[29] Fitriyani, Norma Latif, et al. “Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension.” IEEE Access 7 (2019): 144777-144789.

[30] Georga, Eleni I., et al. “Short-term vs. long-term analysis of diabetes data: Application of machine learning and data

mining techniques." 13th IEEE International Conference on BioInformatics and BioEngineering.IEEE, 2013.

[31] Han, Longfei, et al. "Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes." IEEE journal of biomedical and health informatics 19.2 (2014): 728-734.

[32] Jankovic, Marko V., et al. "Deep prediction model: The case of online adaptive prediction of subcutaneous glucose." 2016 13th Symposium on Neural Networks and Applications (NEUREL). IEEE, 2016.

[33] Karthikeyan, R., P. Geetha, and E. Ramaraj. "Rule Based System for Better Prediction of Diabetes." 2019 3rd International Conference on Computing and Communications Technologies (ICCCT).IEEE, 2019.

[34] Maniruzzaman, Md, et al. "Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm." Computer methods and programs in biomedicine 152 (2017): 23-34.

[35] Mhaskar, Hrushikesh N., Sergei V. Pereverzyev, and Maria D. van der Walt. "A deep learning approach to diabetic blood glucose prediction." Frontiers in Applied Mathematics and Statistics 3 (2017): 14.

[36] Nilashi, Mehrbakhsh, et al. "An analytical method for diseases prediction using machine learning techniques." Computers & Chemical Engineering 106 (2017): 212-223.

[37] Pappada, Scott M., et al. "Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes." Diabetes technology & therapeutics 13.2 (2011): 135-141.

[38] Rakshit, Somnath, et al. "Prediction of Diabetes Type-II Using a Two-Class Neural Network." International Conference on Computational Intelligence, Communications, and Business Analytics.Springer, Singapore, 2017.

[39] Rashid, Tarik A., Saman M. Abdullah, and RezhnaMirza Abdullah. "An intelligent approach for diabetes classification, prediction and description."Innovations in Bio-Inspired Computing and Applications.Springer, Cham, 2016.323-335.

[40] Tama, BayuAdhi, and Kyung-Hyune Rhee. "Tree-based classifier ensembles for early detection method of diabetes: an exploratory study." Artificial Intelligence Review 51.3 (2019): 355-370.

[41] Wu, Jiang, et al. "A semi-supervised learning based method: Laplacian support vector machine used in diabetes disease diagnosis." Interdisciplinary Sciences: Computational Life Sciences 1.2 (2009): 151-155.

[42] Zheng, Tao, et al. "A machine learning-based framework to identify type 2 diabetes through electronic health records." International journal of medical informatics 97 (2017): 120-127.

[43] Choi, SooBeom, et al. "Screening for prediabetes using machine learning models." Computational and mathematical methods in medicine 2014 (2014).

[44] Park, Jin, and Dee W. Edington. "A sequential neural network model for diabetes prediction." Artificial intelligence in medicine 23.3 (2001): 277-293.

[45] Wu, Meiqi, et al. "A deep learning method to more accurately recall known lysine acetylation sites." BMC bioinformatics 20.1 (2019): 49.

[46] Han, Jianchao, Juan C. Rodriguez, and Mohsen Beheshti. "Diabetes data analysis and prediction model discovery using rapidminer." 2008 Second international conference on future generation communication and networking. Vol. 3.IEEE, 2008.

[47] Ding, Shifei, et al. "Extreme learning machine: algorithm, theory and applications." Artificial Intelligence Review 44.1 (2015): 103-115.

[48] Swapna, G., R. Vinayakumar, and K. P. Soman. "Diabetes detection using deep learning algorithms." ICT Express 4.4 (2018): 243-246.

[49] Yang, Boyi, and Adam Wright. "Development of deep learning algorithms to categorize free-text notes pertaining to diabetes: convolution neural networks achieve higher accuracy than support vector machines." arXiv preprint arXiv:1809.05814 (2018).

[50] Ramesh, Sushant, Ronnie D. Caytiles, and N. C. S. Iyengar. "A Deep Learning Approach to Identify Diabetes." Advanced Science and Technology Letters 145 (2017): 44-49.

[51] Thomas, Jensia, et al. "Machine Learning Approach For Diabetes Prediction." *International Journal of Information* 8.2 (2019).

[52] https://www.kaggle.com/uciml/pima-indians-diabetes-database

[53] Moolayil, Jojo, Jojo Moolayil, and Suresh John. *Learn Keras for Deep Neural Networks*. Apress, 2019.