Deep Learning Techniques for the Prediction of Diabetes: A Review

Sunit Kumar Mishra, Arvind Kumar Tiwari
Kamla Nehru Institute of Technology, Sultanpur, India

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

Abstract Diabetes is a verycommon disease in the world. If diabetes is detected in early stage, it can be cured easily. Several machine learning techniques are available to predict diabetes in earlier stage using data set. This paper presents 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, ensemble of CNN and LSTM and ensemble of CNN and BLSTM to predict diabetes by taking a dataset.

1 INTRODUCTION

Diabetes is a common disease in our society. Every third person is affected from this serious disease. This is caused by irregular life style, bad eating habits, and lack of exercise and also during pregnancy. In human body blood sugar level is controlled by insulin hormone released by 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 from diabetes. The patients affected from diabetes can be cured by regular exercise, and by adopting healthy lifestyle. To control blood sugar level some medicine may be given or insulin may be given explicitly. To know whether a person is affected from 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 (Kerner & Bruckel,2014). Machine learning techniques learn from dataset to predict outcomes. Some data is used as a training data which is used to train and then we can perform prediction using test data (Bottou,2014).For early stage diabetes prediction the various researchers have been used Support Vector Machine(Vishwanathan et al.,2002),Naive Bayes (Rish,2001), Artificial Neural Network (Wang,2003), Decision tree (Safavian et al.,1991)(Pal,2005),K nearest Neighbour (Liao & Vemuri ,2002),LSTM(Long Short Term Memory)(Sherstinsky ,2020).

2 RELATED WORK

In the literaturevarious researchers have been proposed machine learning approaches for prediction of diabetes. In paper authors (Aljumah et al.,2013) authors has been proposed SVM based approach by using the Dataset of disease in Saudi Arabia to observe obesityand predicts chances of diabetes in a person. The authors (Chen & Pan ,2018) have been proposed 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 (Mercaido et al.,2017) worked on the concept of classification. They used Pima Indians dataset and obtained precision value 0.770 and recall value 0.775.The authors (Patil et al.,2010) proposed a prediction Model which used simple K-means algorithm and C4.5 algorithmusingPima Indians diabetes data and achieved an accuracy of 93.5%.The authors (Kavakiotis et al.,2017) have been proposed SVM based model and got the accuracy of 85%. The authors (Kohli & Arora,2018) have been proposed logistic regression on separate datasets of heart, breast cancer and diabetesand got accuracy of 80.77%.The authors (Perveen et al.,2016) performed classification technique, decision tree J48 and achieved Area under ROC is 0.98.The authors (Sisodia, 2018)used DT, SVM and NB classification methods on Pima Indians Diabetes datasets. NB gave accuracy of 76.30%. The authors (Kowsher et al., 2019) proposed that deep ANN gave 95.14%

Mishra, S. and Tiwari, A.
Deep Learning Techniques for the Prediction of Diabetes: A Review.
DOI: 10.5220/0010567400003161
In *Proceedings of the 3rd International Conference on Advanced Computing and Software Engineering (ICACSE 2021)*, pages 232-237
ISBN: 978-989-758-544-9

accuracy using dataset of 9483 diabetes patients. The authors (Srivastava et al.,2019) used Pima Indian dataset(PIDD) and ANN and achieved the accuracy of 92%. The authors (Kaur & Kumari,2020) proposed SVM on PIDD and achieved accuracy of 89%.The authors (Maniruzzaman et al.,2020) proposed combination of Logistic Regression and Random Forest classifier which gave accuracy of 94.25% and AUC 0.95 on dataset taken from NHE Survey of 6561 respondents. The authors (Alam et al.,2019) used OCTA image database with Logistic regression-based model and found 95.01% sensitivity. The authors (Zhang et al.,2018) used Multilayer feed forward network to predict diabetes on PIDD. The authors (Birjais et al.,2019) proposed the Gradient Boosting based method for the prediction and diagnosis of future diabetes risk and got the accuracy 86%.The authors (Durairaj et al.,2015) used Back propagation algorithm with Pima Indian diabetes dataset and achieved an accuracy of 91%.The authors (Dagliati et al.,2018) used random forest method on ICSM dataset and got accuracy of 77.7%.The authors (Donsa et al.,2015) proposed ANN model with large clinical dataset and achieved accuracy of 94%. The author (Dwivedi,2018) proposed logistic regression technique on the dataset maintained by NIDDK Diseases and achieved 78% accuracy. The authors (Fitriyani et al.,2019) proposed an ensemble learning approach with four different datasets and achieved accuracy of 75.78%.The authors (Georga et al., 2013) considered two predictive models support vector regression and Gaussian process, one for short term glucose control and second for long term glucose control using thedatasetconsists of 15 diabetic patients. The authors (Han et al,2014) 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 (Jankovic et al.,2016) 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 (Karthikeyan et al., 2019) proposed rule based classification system to predict diabetes using PIDD for diabetes and got accuracy of 81.97%. The authors (Mhaskar et al.,2017) 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 (Nilashi et al., 2017) proposed CART(classification and regression trees) algorithm on PIDD and got accuracy of 93.6%. The authors (Pappada et al., 2011) used multi layered feed forward neural network on clinical dataset of 10 patients, for real time predictions to predict the rise or fall in glucose level in every 90 minutes.The authors (Rakshit et al.,2017) proposed neural network on PIDD dataset and achieved accuracy of 83.3%. The authors (Rashid et al.,2016) proposed two algorithms one is Artificial neural network, to predict rate of fasting blood sugar and the second is decision tree take decision on the basis of symptoms. The algorithms was applied on the clinical data set of 500 patients and got accuracy of 84.8% with feature extraction.The authors (Tama & Rhee, 2019) proposed LMT (logistic model tree) based classification techniques for prediction of diabetes in a patient in early stage, on clinical data and got the accuracy 96.38%.The authors (Wu J et al., 2009) proposed a semi supervised machine learning algorithm Laplacian support vector machine on Pima indians dataset and achieved accuracy of 82.29%.The authors (Zheng et al., 2017) have been compared machine learning techniques like KNN and naïve Bayes.The authors (Choi et al., 2014) have been developed two models to predict pre-diabetes one is Artificial Neural Network and the other is SVM. Data is taken from KNHANES. They got area under curve using SVM is 0.731 and using ANN is 0.729.The authors (Park & Edington, 2001) proposed multilayer neural network with back propagation model for the prediction of diabetes on 6142 patients and get the sensitivity 86.04%.The authors (Wu M et al.,2019) used deep learning techniques to diagnose diabetes and got accuracy 84.95% ,specificity 83.45%, sensitivity 86.44% and AUC of 0.8540. Authors (Han et al.,2008) used Rapid-I's to analyze Pima Indians Diabetes Dataset. Theyused ID3 decision tree to predict diabetes with 80% of accuracy. Authors (Ding et al.,2015) used extreme learning machine algorithm which uses single layer feed forward neuralnetwork and also points out future perspectives of ELM and gave accuracy of 77.63% on UCI dataset.The authors (Swapna et al., 2018) performs 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 (Yang & Wright , 2018) proposed convolutional neural network to predict diabetes on the Brigham and Women's Hospital dataset set and got AUC of 0.97.The authors (Ramesh et al.,2017)proposed Recurrent Deep Neural Network (RNN) model on PIMA Indian diabetes datasetand got the accuracy of

81%.The authors (Hasan et al.,2020) proposed multilayer perceptron model to predict diabetes using AUC values as the performance parameter and got AUC value 0.95 on PIDD. The authors (Naz & Ahuja, 2020)proposed deep learning model to predict diabetes on PIDD and got accuracy of 98.07 percent. The authors (Rehman et al., 2020) proposed DELM based deep learning technique to predict diabetes got accuracy of 92.8 percent on sample of 4500. The authors (Srivastava et al.,2021) proposed ABC-DNN model to predict diabetes using PIDD and got accuracy of 94.74 percent. The authors (Bora et al., 2021) proposed a deep learning model for predicting the risk of development of diabetic retinopathy on a set of 575431 eyes and got ROC value 0.79.

3 MATERIALS AND METHODS

Here, in this paper the description of diabetes dataset, methodology to compare the performance of various machine learning models has been provided.

3.1 Data Description

In this paper the PIMA Indian diabetes dataset (Thomas et al., 2019) was used which was taken from Kaggle(https://www.kaggle.com/uciml/pima-indians-diabetes-database) . 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 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 obesity which may lead to diabetes. Hence triceps skinfold thickness plays an important 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.

3.2 Performance Evaluation

In this paper the Accuracy, Precision, Recall, F1 Score and AUCare used to measure the performance of the proposed approach.Accuracy of the models is determined by confusion metrics through *K*-Fold cross validation. The Confusion matrix consists of True Positive (TP), True Negative (TN), and Falsepositive (FP) and False Negative (FP) where:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 score} = \frac{2*(\text{Recall} * \text{Precision})}{\text{Recall} + \text{precision}} \quad (4)$$

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 1.

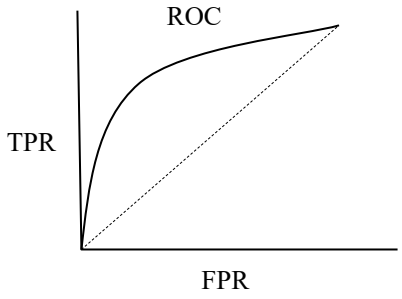

Figure 1.

$$\text{TPR/Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

$$FPR = 1 - \text{Specificity} \quad (7)$$

$$FPR = \frac{FP}{TN + FP} \quad (8)$$

3.3 Methodology

In this paper, the dataset 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 LSTMandensembleof CNN and BLSTM.

4 COMPARATIVE ANALYSIS

This paper used the PIMA Indian diabetes dataset taken from Kaggle and uploaded onto the Jupitor Notebook. After data cleaning and pre-processing the data is converted into numpy array which is 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, ensemble of CNN and LSTM and ensemble of CNN and BLSTM for the prediction of diabetes (See Table - 1).

Table 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	0.9714	0.9730	0.9635	0.9679	0.9711
CNN+BLSTM	0.9674	0.9646	0.9630	0.9635	0.9673

From the comparative analysis it is observed that deep learningapproachesperformed better in comparison tosimple machine learning algorithms like Naive Bayes, ANN, SVM, KNN, Random Forest etc.

5 CONCLUSIONS

To predict the early stage of diabetes is one of the most challenging and important task. If diabetes is detected in early stage, it can be cured easily. Several machine learning techniques are available to predict diabetes in earlier stage using data set. This paper has been presentedareview of several machine learning based for the prediction of diabetes.Thispaper 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 by using PIDD. The comparative analysis shown that deep learning approaches performs better in comparison simple machine learning algorithms like Naive Bayes, ANN, SVM, KNN, Random Forest etc.

REFERENCES

Alam, M., Le, D., Lim, J. I., Chan, R. V., & Yao, X. (2019). Supervised machine learning based multi-task artificial intelligence classification of retinopathies. Journal of clinical medicine, 8(6), 872.

Aljumah, A. A., Ahamad, M. G., &Siddiqui, M. K. (2013). Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University-Computer and Information Sciences, 25(2), 127-136.

Birjais, R., Mourya, A. K., Chauhan, R., &Kaur, H. (2019). Prediction and diagnosis of future diabetes

risk: a machine learning approach. SN Applied Sciences, 1(9), 1112.

Bora, A., Balasubramanian, S., Babenko, B., Virmani, S., Venugopalan, S., Mitani, A., ... &Bavishi, P. (2021). Predicting the risk of developing diabetic retinopathy using deep learning. The Lancet Digital Health, 3(1), e10-e19.

Bottou, L. (2014). From machine learning to machine reasoning. Machine learning, 94(2), 133-149.

Chen, P., & Pan, C. (2018). Diabetes classification model based on boosting algorithms. BMC bioinformatics, 19(1), 109.

Choi, S. B., Kim, W. J., Yoo, T. K., Park, J. S., Chung, J. W., Lee, Y. H., ... & Kim, D. W. (2014). Screening for prediabetes using machine learning models. Computational and mathematical methods in medicine, 2014.

Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., ... &Bellazzi, R. (2018). Machine learning methods to predict diabetes complications. Journal of diabetes science and technology, 12(2), 295-302.

Ding, S., Zhao, H., Zhang, Y., Xu, X., &Nie, R. (2015). Extreme learning machine: algorithm, theory and applications. Artificial Intelligence Review, 44(1), 103-115.

Donsa, K., Spat, S., Beck, P., Pieber, T. R., &Holzinger, A. (2015). Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges. In Smart Health (pp. 237-260). Springer, Cham.

Durairaj, M., &Kalaiselvi, G. (2015). Prediction of diabetes using back propagation algorithm. International Journal of Emerging Technology and Innovative Engineering, 1(8).

Dwivedi, A. K. (2018). Analysis of computational intelligence techniques for diabetes mellitus prediction. Neural Computing and Applications, 30(12), 3837-3845.

Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2019). Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. IEEE Access, 7, 144777-144789.

Georga, E. I., Protopappas, V. C., Mougiakakou, S. G., & Fotiadis, D. I. (2013, November). Short-term vs. long-term analysis of diabetes data: Application of machine learning and data mining techniques. In 13th IEEE International Conference on BioInformatics and BioEngineering (pp. 1-4). IEEE.

Han, J., Rodriguez, J. C., &Beheshti, M. (2008, December). Diabetes data analysis and prediction model discovery using rapidminer. In 2008 Second international conference on future generation communication and networking (Vol. 3, pp. 96-99). IEEE.

Han, L., Luo, S., Yu, J., Pan, L., & Chen, S. (2014). 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), 728-734.

Hasan, M. K., Alam, M. A., Das, D., Hossain, E., &Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, 76516-76531.

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

Jankovic, M. V., Mosimann, S., Bally, L., Stettler, C., &Mougiakakou, S. (2016, November). Deep prediction model: The case of online adaptive prediction of subcutaneous glucose. In 2016 13th Symposium on Neural Networks and Applications (NEUREL) (pp. 1-5). IEEE.

Karthikeyan, R., Geetha, P., &Ramaraj, E. (2019, February). Rule Based System for Better Prediction of Diabetes. In 2019 3rd International Conference on Computing and Communications Technologies (ICCCT) (pp. 195-203). IEEE.

Kaur, H., &Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. Applied computing and informatics.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., &Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal, 15, 104-116.

Kerner, W., &Brückel, J. (2014). Definition, classification and diagnosis of diabetes mellitus. Experimental and clinical endocrinology & diabetes, 122(07), 384-386.

Kohli, P. S., &Arora, S. (2018, December). Application of machine learning in disease prediction. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-4). IEEE.

Kowsher, M., Turaba, M. Y., Sajed, T., &Rahman, M. M. (2019, December). Prognosis and Treatment Prediction of Type-2 Diabetes Using Deep Neural Network and Machine Learning Classifiers. In 2019 22nd International Conference on Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.

Liao, Y., &Vemuri, V. R. (2002). Use of k-nearest neighbor classifier for intrusion detection. Computers & security, 21(5), 439-448.

Maniruzzaman, M., Kumar, N., Abedin, M. M., Islam, M. S., Suri, H. S., El-Baz, A. S., &Suri, J. S. (2017). Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. Computer methods and programs in biomedicine, 152, 23-34.

Maniruzzaman, M., Rahman, M. J., Ahammed, B., &Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. Health Information Science and Systems, 8(1), 7.

Mercaldo, F., Nardone, V., &Santone, A. (2017). Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia computer science, 112, 2519-2528.

Mhaskar, H. N., Pereverzyev, S. V., & van der Walt, M. D. (2017). A deep learning approach to diabetic blood glucose prediction. Frontiers in Applied Mathematics and Statistics, 3, 14.

Naz, H., &Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders, 19(1), 391-403.

Nilashi, M., bin Ibrahim, O., Ahmadi, H., &Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques. Computers & Chemical Engineering, 106, 212-223.

Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

Pappada, S. M., Cameron, B. D., Rosman, P. M., Bourey, R. E., Papadimos, T. J., Olorunto, W., &Borst, M. J. (2011). Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. Diabetes technology & therapeutics, 13(2), 135-141.

Park, J., &Edington, D. W. (2001). A sequential neural network model for diabetes prediction. Artificial intelligence in medicine, 23(3), 277-293.

Patil, B. M., Joshi, R. C., &Toshniwal, D. (2010). Hybrid prediction model for type-2 diabetic patients. Expert systems with applications, 37(12), 8102-8108.

Perveen, S., Shahbaz, M., Guergachi, A., &Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. Procedia Computer Science, 82, 115-121.

Rakshit, S., Manna, S., Biswas, S., Kundu, R., Gupta, P., Maitra, S., & Barman, S. (2017, March). Prediction of Diabetes Type-II Using a Two-Class Neural Network. In International Conference on Computational Intelligence, Communications, and Business Analytics (pp. 65-71). Springer, Singapore.

Ramesh, S., Caytiles, R. D., &Iyengar, N. C. S. (2017). A Deep Learning Approach to Identify Diabetes. Advanced Science and Tech Letter, 145, 44-49.

Rashid, T. A., Abdullah, S. M., & Abdullah, R. M. (2016). An intelligent approach for diabetes classification, prediction and description. In Innovations in Bio-Inspired Computing and Applications (pp. 323-335). Springer, Cham.

Rehman, A., Athar, A., Khan, M. A., Abbas, S., Fatima, A., &Saeed, A. (2020). Modelling, simulation, and optimization of diabetes type II prediction using deep extreme learning machine. Journal of Ambient Intelligence and Smart Environments, (Preprint), 1-14.

Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).

Safavian, S. R., &Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics, 21(3), 660-674.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. Physica D: Nonlinear Phenomena, 404, 132306.

Sisodia, D., &Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. Procedia computer science, 132, 1578-1585.

Srivastava, A. K., Kumar, Y., & Singh, P. K. (2021). Artificial Bee Colony and Deep Neural Network-Based Diagnostic Model for Improving the Prediction Accuracy of Diabetes. International Journal of E-Health and Medical Communications (IJEHMC), 12(2), 32-50.

Srivastava, S., Sharma, L., Sharma, V., Kumar, A., &Darbari, H. (2019). Prediction of Diabetes Using Artificial Neural Network Approach. In Engineering Vibration, Communication and Information Processing (pp. 679-687). Springer, Singapore.

Swapna, G., Vinayakumar, R., &Soman, K. P. (2018). Diabetes detection using deep learning algorithms. ICT Express, 4(4), 243-246.

Tama, B. A., & Rhee, K. H. (2019). Tree-based classifier ensembles for early detection method of diabetes: an exploratory study. Artificial Intelligence Review, 51(3), 355-370.

Thomas, J., Joseph, A., Johnson, I., & Thomas, J. (2019). Machine learning approach for diabetes prediction. International Journal of Information, 8(2).

Vishwanathan, S. V. M., &Murty, M. N. (2002, May). SSVM: a simple SVM algorithm. In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (Vol. 3, pp. 2393-2398). IEEE.

Wang, S. C. (2003). Artificial neural network. In Interdisciplinary computing in java programming (pp. 81-100). Springer, Boston, MA.

Wu, J., Diao, Y. B., Li, M. L., Fang, Y. P., & Ma, D. C. (2009). A semi-supervised learning based method: Laplacian support vector machine used in diabetes disease diagnosis. Interdisciplinary Sciences: Computational Life Sciences, 1(2), 151-155.

Wu, M., Yang, Y., Wang, H., &Xu, Y. (2019). A deep learning method to more accurately recall known lysine acetylation sites. BMC bioinformatics, 20(1), 49.

Yang, B., & Wright, A. (2018). 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.

Zhang, Y., Lin, Z., Kang, Y., Ning, R., &Meng, Y. (2018). A feed-forward neural network model for the accurate prediction of diabetes mellitus. International Journal of Scientific and Technology Research, 7(8), 151-155.

Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics, 97, 120-127.