IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

AN EFFICIENT DEEP NEURAL NETWORK APPROACH FOR DIABETES PREDICTION

Dr. Rajendra Prasad Banavathu[1], Dr.S. Jayaprada[1], Dr. Kalpana Devi Bai Mudavathu[2]

Dept. of CSE(AI&ML), Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India[1]

Dept. of CSE, PSCMR College of Engineering and Technology, Vijayawada, Andhra Pradesh, India[2]

Abstract- Millions of people all over the world endure from the incessant condition of Diabetes. Early detection and action can lower the likelihood of problems and assist avoid or delay its development. Diabetes has been predicted using machine learning algorithms using a variety of characteristics, including demographics, clinical data, and lifestyle factors. Using a mix of patient data, including age, body mass index and more we present an approach based on deep learning to predict the chance of acquiring diabetes. K Nearest Neighbor(KNN), Logistic Regression(LR), Support Vector Machine(SVM), Decision Tree(DT) and Random Forest(RF), Deep Neural Networks (DNN) are some of the algorithms used. Each algorithm's accuracy is calculated along with the model's accuracy. The approach with a high accuracy level is used as the model to predict diabetes. This strategy may help medical professionals make knowledgeable judgements and give patients personalized care. A number of metrics, such as accuracy and F1 score, are used to assess the effectiveness of the suggested model. Using deep learning concepts by training the properties of a deep neural network(DNN), we suggest a method for diagnosing diabetes. with 98.49% prediction accuracy, and 93% F1 Score. The experimental findings show that when using Deep learning approach, the suggested system offers good outcomes. This strategy may help medical professionals make knowledgeable judgements and give patients personalized care.

Keywords: Diabetes, Supervised Learning, DL, Data mining, KNN, SVM, Light GBM, DT, RF, DNN

1. INTRODUCTION

A continual metabolic illness called diabetes impairs the body's capacity to create or use insulin, which causes blood sugar levels to rise. With an estimated 463 million adults having the condition in 2019, it has elevated to a major public health concern. To effectively manage diabetes and lower the risk of consequences like heart disease, blindness, and kidney failure, early detection and intervention are essential. A range of data mining techniques, ML algorithms, DL algorithms and statistical methods are used to locate knowledge and predict future events using both current and historical data. We employed deep neural networks (DNN), a method that has lately become quite popular in deep learning, to predict diabetes mellitus. Diabetes has been correctly diagnosed using Deep Neural Network technology. Using Deep Learning approach i.e., using Deep Neural Networks (DNN) the accuracy was further improved. The goal of this project is to investigate how deep learning algorithms can be used instead of machine learning algorithms model to prognosticate diabetes based on patient complication information from the demographics, some other different source and lifestyle factors. In order to provide preventive and individualized healthcare interventions, deep learning algorithms have emerged as a viable technique for predicting the possibility of getting diabetes.

2. LITERATURE REVIEW

Results from related research that analyzed various healthcare datasets and made predictions using a variety of methods and strategies are presented. Researchers have created and used a variety of prediction models utilizing different data mining techniques, machine learning algorithms, deep learning techniques or even a mix of these techniques. For the study of diabetic data, [1] M. M. Islam, R. Ferdousi, S. Rahman, and Y. B. Humayra, (2020) designed a method employing the likelihood prediction of diabetes at early stage using data mining techniques,. This approach forecasts diabetes as well as the risks that come with it. They draw information from the dataset and describe patterns in simple language. Using a Bayesian network. [3] M.Panwar, A. Acharyya, R A. Shafik, D. Biswas, (2016) used K-Nearest Neighbor (KNN) Based Methodology to predict diabetes.Any healthcare institution can use the system because it is classification-based and cost-effective.[5] D. A. Otchere, T. O. Arbi Ganat, R. Gholāmi, and S. Ridha, used comparative analysis of ANN and SVM models, helps in predicting diabetes. It helps in mitigating overfitting.[7] DD Rufo, TG Debelee, A Ibenthal, WG Negera (2021) used Light GBM technique for this system. They applied light GBM to extract hidden patterns from the dataset for effective classification. This algorithm improved efficiency and have faster training speed. [11] Han, J., Rodriguez, J.C., Beheshti, M, Discovering decision tree based diabetes prediction model. The training of dataset using

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

decision tree is very effective and easier to implement and analyze.[17] Due to the fact that Random Forest creates a random subset of features, it ensures low correlation between decision trees. M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan for Performance Assessment of Random Forest. [17] Deep neural networks with gradual input were utilised by H. M. D. Kabir, M. Abdar, A. Khosravi, et al. to predict diabetes. DNN is capable of learning and modelling complicated, nonlinear relationships between inputs and outputs, as well as generalisations and inferences, as well as the modelling of extremely volatile data.

3. PROPOSED SYSTEM

All of the methods like Naïve bayes, K-Nearest Neighbors, SVM, Decision trees, Random Forest adopted here are overfitting-sensitive and even cannot work well for large amount of data. The algorithms we used here are also giving an accurate output with an efficient accuracy but the accuracy is less than 90%. To solve that issue, we employed the deep learning techniques, which is less vulnerable to overfitting, can work well with large datasets and even capable of producing greater accuracy results for small datasets. As DNN here gives an average accuracy of 98.49%, we choose DNN is best among the above-mentioned algorithms. Hence, here we proposed deep neural networksto train our model. In between the input, output layersof a deep neural network (DNN), it is a complex neural network, with many hidden layers for proposed model deep neural networks are built for anticipating results and finding connections and patterns in the data set. different learning techniques are used to decide the conclusion in this case. The elements, or nodes, model deep neural network are interconnected, from input to output layer. The accuracy of the output generation depends on the hidden interconnected layer strength. There are many hidden layers and numerous neurons in each hidden layer of a proposed model deep neural network. Fig.1 displays a straightforward deep neural network architecture. The activation function plays a crucial function in deep neural networks. Furthermore, essential for eliminating the infinitely large linearly weighted sum from neurons are activation functions.

We employ deep learning to predict diabetes mellitus in the manner described below.

a. Collection of Data
b. Data Preparation
c. Deep Neural Network implementation
d. Evaluation Criteria

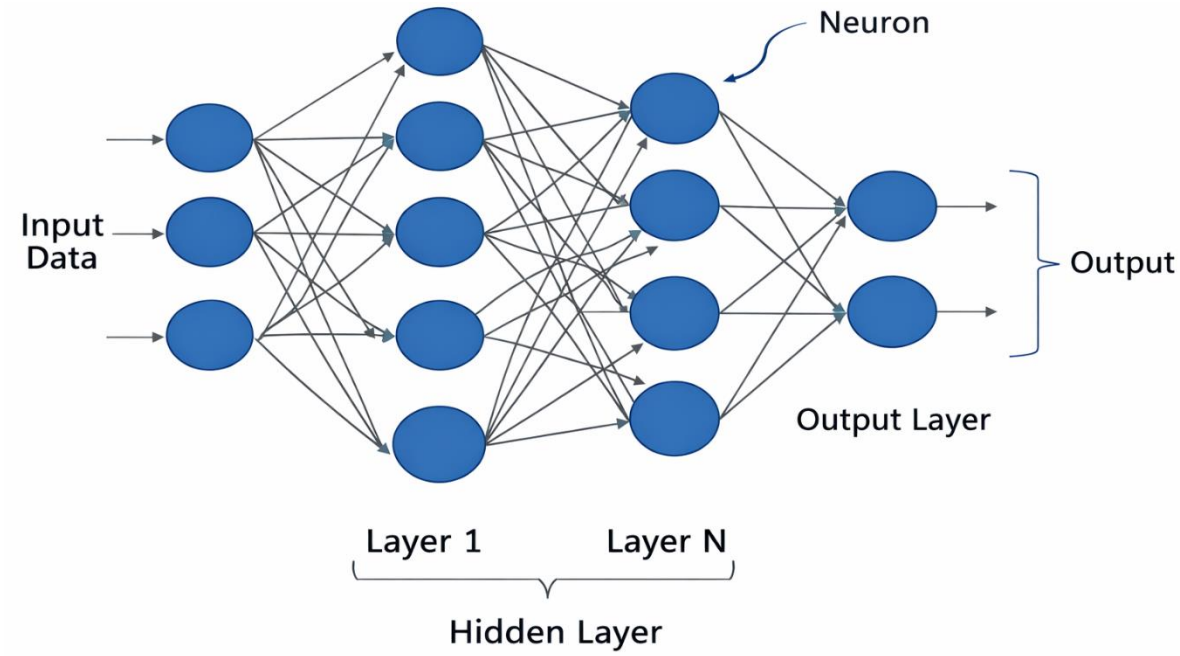


Fig 1. Deep Neural Network

a) **Collection of Data:** The dataset is collected from all female patients with ages ranging from 20 to 85, and its features are examined. The dataset has 9 attributes and 768 records. The output label is represented by the last attribute. The dataset do not have any missing values. The distribution of outcome is, it has the data about 500 patients having diabetes and 268 patients not having diabetes. The following Fig 2 represents the attributes and top 5 records present in the dataset.

Table1. Dataset Attributes and Values

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

Index	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

b) DATA PREPARATION: Any system's effectiveness is based on the data's standard. So, we examine the dataset to see whether any missing values are there. Using bar graphs, shown in Fig 3. we looked at the range of each attribute.

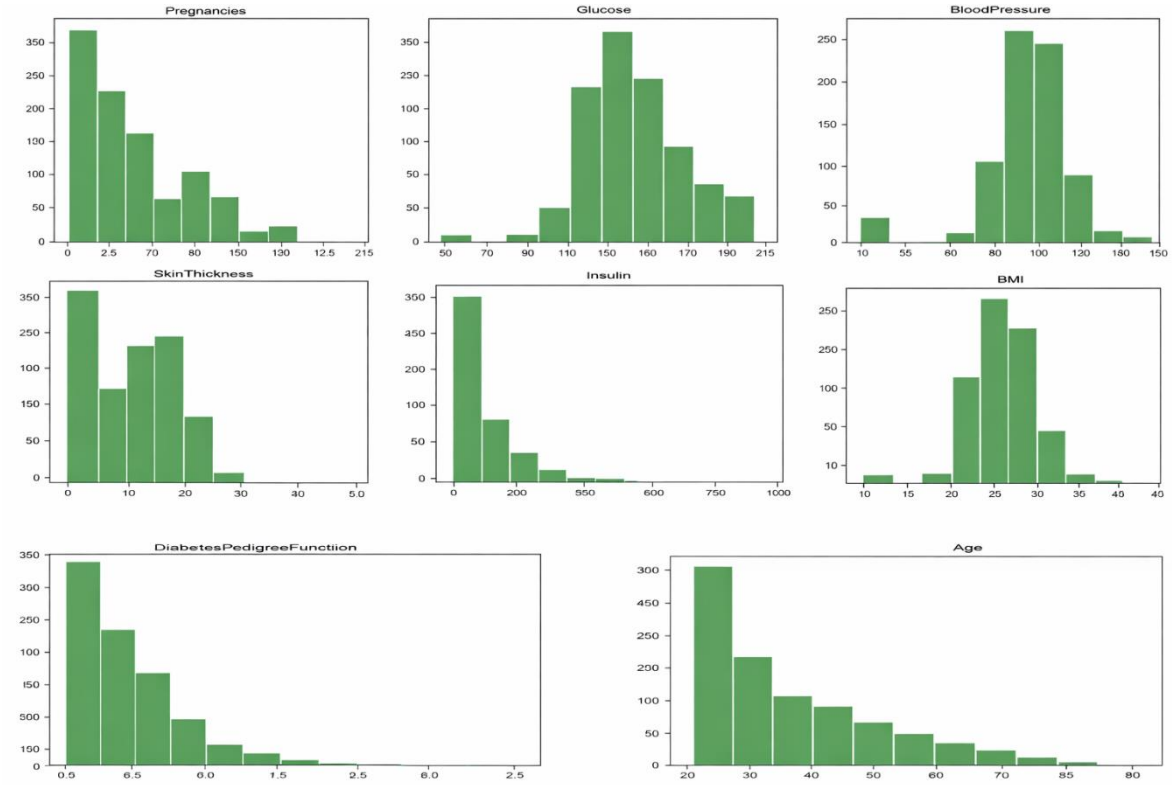


Figure 2. Bar Graphs of every Attribute

Heat maps shown in Fig 5. were used to determine the relationships between the attributes, and scatterplots shown in Fig 4. were used to show how each attribute was distributed. To calculate the system's performance, we employed the validation approaches.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

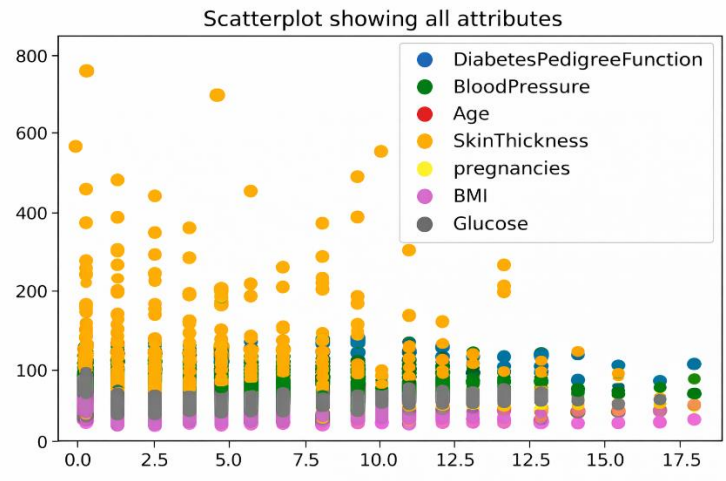


Figure 3. Scatterplot representing Attributes

Table 2 Correlation Matrix of Diabetes Dataset Features

Feature	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
Pregnancies	1.000	0.125	0.141	-0.082	-0.074	0.018	-0.034	0.544	0.222
Glucose	0.125	1.000	0.153	0.057	0.331	0.221	0.137	0.263	0.467
Blood Pressure	0.141	0.153	1.000	0.207	0.089	0.281	0.041	0.239	0.066
Skin Thickness	-0.082	0.057	0.207	1.000	0.437	0.392	0.184	-0.113	0.075
Insulin	-0.074	0.331	0.089	0.437	1.000	0.198	0.181	-0.042	0.130
BMI	0.018	0.221	0.281	0.392	0.198	1.000	0.141	0.036	0.292
Diabetes Pedigree Function	-0.034	0.137	0.041	0.184	0.181	0.141	1.000	0.034	0.174
Age	0.544	0.263	0.239	-0.113	-0.042	0.036	0.034	1.000	0.238
Outcome	0.222	0.467	0.066	0.075	0.130	0.292	0.174	0.238	1.000

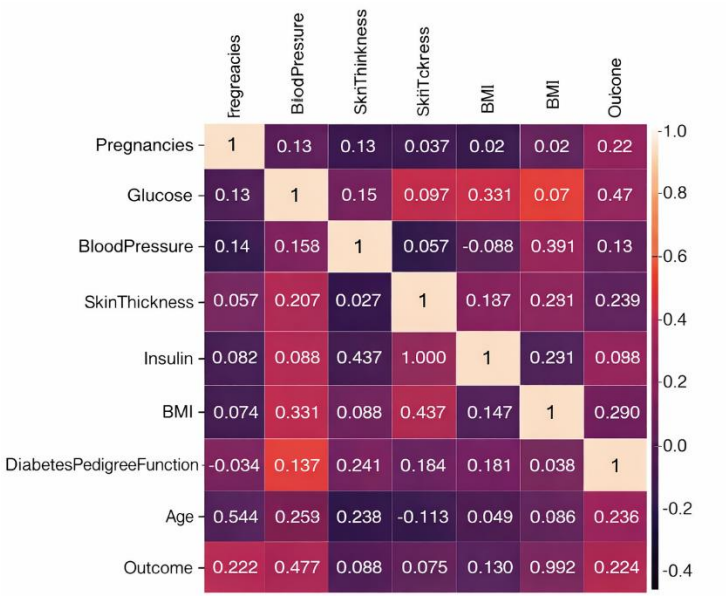


Figure4. Heatmap showing correlation values of Attributes

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

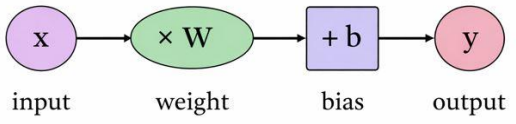
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

c) DEEP NEURAL NETWORK IMPLEMENTATION

The neural network's hidden layer was selected for this system, and it has four hidden layers with a total of 16, 14, 12, and 16 neurons, respectively. For the diabetes prediction, we test a variety of hidden layers and various neurons in a variety of layers. The number of neuron nodes in each hidden layer should be 16, 14, 12, and 16 for the best results. The hidden layer should be 4. Fig. 6 depictsthe design of the deep neural network that we created to forecast diabetes. In a proposed deep neural network, neuron nodes calculate sum of the weight of the inputs based on given formula below, added bias,and then choose whether or not to "fire" them. So, a particular neuron will look like this.

Y= □ (xi * wi) + bias



d) **Evaluation Criteria**

The confusion matrix is used to visualize how well supervised learning systems perform. It is an overview of the classification problem's predicted result. This is a list of terms and their definitions as they relate to the confusion matrix:

Table 3 Confusion Matrix Representation

Prediction / Actual	Actual: No	Actual: Yes
Predicted: No	True Negative (TN) – Correctly predicted negative cases	False Positive (FP) – Incorrectly predicted positive cases
Predicted: Yes	False Negative (FN) – Incorrectly predicted negative cases	True Positive (TP) – Correctly predicted positive cases

Using the confusion matrix, the performance of the system can easily be calculated. The accuracy, F1 score are calculated as shown in below Table

Table 4 Confusion Matrix and Performance Metrics

Actual / Predicted	Positive (+)	Negative (−)
Actual Positive (+)	TP – True Positive	FN – False Negative (Type II Error)
Actual Negative (−)	FP – False Positive (Type I Error)	TN – True Negative

Table 5 Classification Evaluation Metrics

Metric	Formula	Meaning
Sensitivity (Recall)	TP / (TP + FN)	Ability of the model to correctly identify positive cases
False Negative Rate (FNR)	FN / (TP + FN)	Proportion of positives incorrectly predicted as negative
False Positive Rate (FPR)	FP / (FP + TN)	Proportion of negatives incorrectly predicted as positive
Specificity	TN / (FP + TN)	Ability of the model to correctly identify negative cases
Precision	TP / (TP + FP)	Accuracy of positive predictions
False Discovery Rate (FDR)	FP / (TP + FP)	Probability that a positive prediction is incorrect
False Omission Rate (FOR)	FN / (FN + TN)	Probability that a negative prediction is incorrect
Negative Predictive Value (NPV)	TN / (FN + TN)	Accuracy of negative predictions
Accuracy	(TP + TN) / (TP + TN + FP + FN)	Overall correctness of the model
F1 Score	2TP / (2TP + FP + FN)	Harmonic mean of Precision and Recall

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

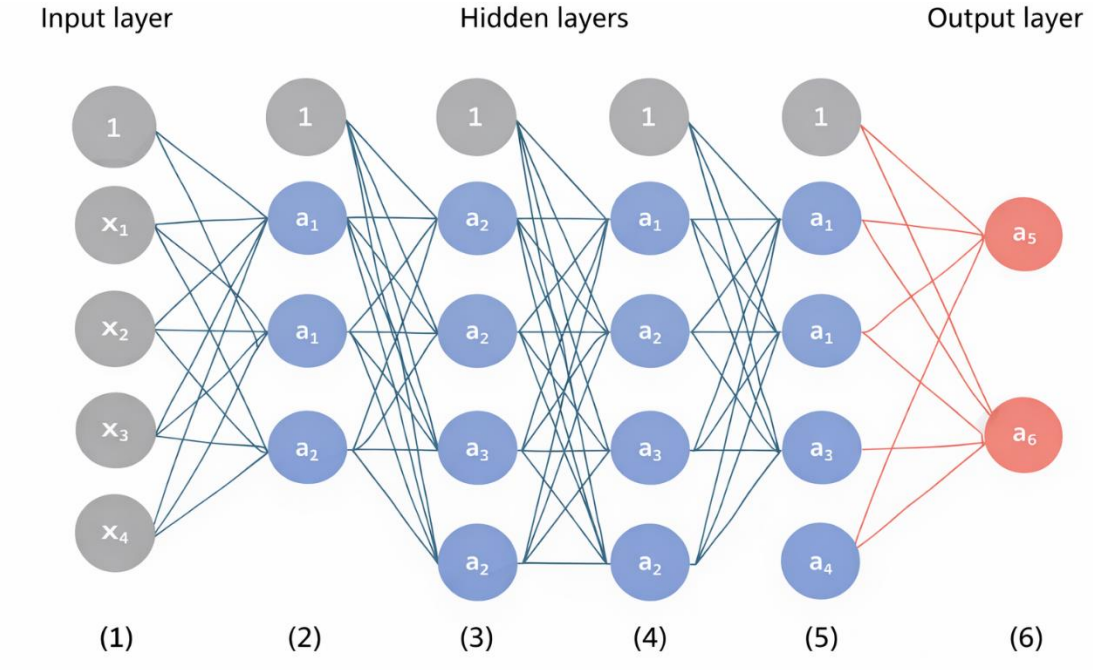


Fig7. Hidden Layers in Deep Learning

4. RESULTS AND DISCUSSION

As we employed a variety of data mining, machine learning and Deep learning techniques, we evaluated the accuracy and F1 Score of each strategy. Performance of these algorithms were compared, and the efficiency of each approach was demonstrated as a consequence. Here a comparison of both ML and DL algorithms are done. Our observations show that DNN offered the highest accuracy of 98.49% and F1 Score of 95 was found out of all of these. The results about the accuracy of different algorithms are shown below in the tabular form and in the form of a bar graph shown in Fig 8. With these results we conclude that DNN provided the best Accurate result among the mentioned algorithms.

Table 1: Accuracy and F1 Score of Algorithms

S.No	Algorithm	Accuracy (%)	F1 Score
0	KNN	76.03	69
1	SVM	78.50	74
2	Logistic Regression	79.32	77
3	Light GBM	82.53	80
4	Classification	85.72	83
5	Decision Tree	90.66	89
6	Random Forest	95.52	92
7	DNN	98.49	95

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13536

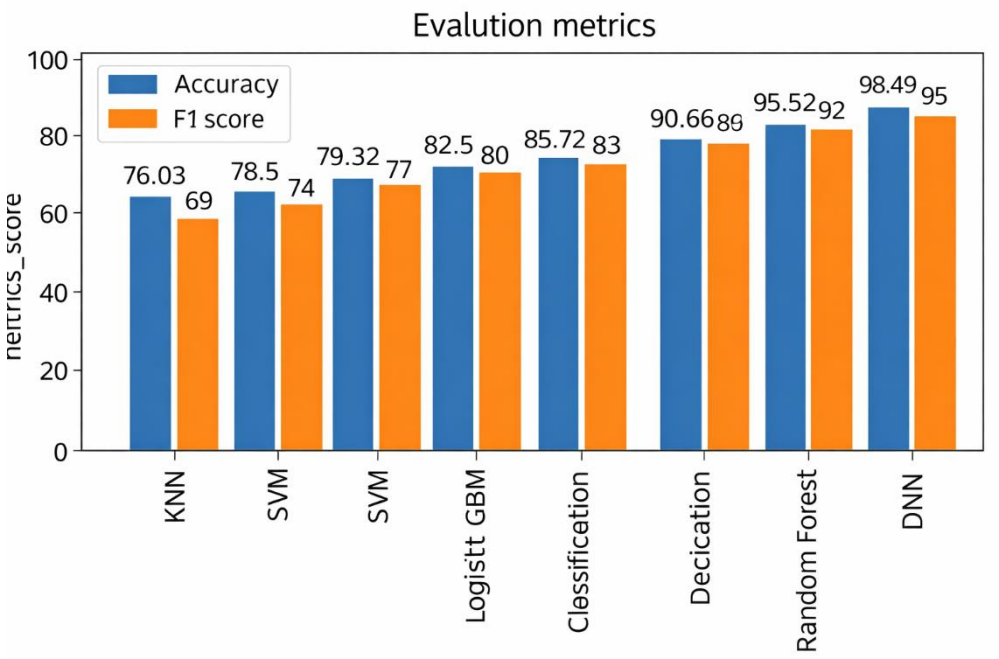


Fig 8. Evaluation Metrics of different Algorithms

5. CONCLUSION

In conclusion, diabetes prediction using deep learning algorithms has yielded promising results. To generate precise forecasts, these algorithms consider a variety of factors like age, body mass index, blood sugar levels. SVM has the F1 Score of 74, with an accuracy rate of 78.50%. The accuracy of KNN is only 76.03%. whereas the accuracy for Logistic Regression is about 79.32% and F1 Score is 77 we further used Mild GBM, with an F1 Score of 80 and an accuracy of 82.53%. Despite being the most widely used classification method, the classification's final accuracy rate on our data set is just 85.72%, and its F1 score is 83. The decision tree method was then used, and it produced results with an accuracy of 90.66% and an F1 Score of 89. Thereafter, we tested Random Forest, which had an F1 Score of 92 and an accuracy of 95.52%. As a result, we explored employing Deep Learning Algorithms next. Deep neural networks (DNN) were deployed, and they provided an accuracy of 98.49% and an F1 Score of 95. This demonstrates that DNN is the best classification method for predicting diabetes. It is crucial to remember that even while deep learning algorithms can make precise predictions, their guidance and diagnosis should always be sought from qualified medical personnel. These algorithms should be used as a supplemental tool to increase the precision of managing and diagnosing diabetes.

REFERENCES

[1]. M. M. Islam, R. Ferdousi, S. Rahman, and Y. B. Humayra, “Likelihood prediction of diabetes at early stage using data mining techniques,” in Computer Vision and Machine Intelligence in Medical Image Analysis, pp. 113–212, Springer, Singapore, 2020

[2]. K. Rajesh and V. Sangeetha, “Application of Data Mining Methods and Techniques for Diabetes Diagnosis”, International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012.

[3]. M. Panwar, A. Acharyya, R A. Shafik, D. Biswas, “K-Nearest Neighbor Based Methodology for Accurate International Symposium on Embedded Computing and System Design (ISED),pp. 132-136,2016.

[4]. R. Hajizadeh, A. Ali, and M. Ezoji, “Mutual neighborhood and modified majority voting based KNN classifier for multi categories classification Pattern Analysis and Applications,” Pattern Analysis and Applications, pp. 1–21, 2022.

[5]. D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, “Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: comparative analysis of ANN and SVM models,” Journal of Petroleum Science and Engineering, vol. 200,Article ID 108182, 2021

[6]. G. A. Pethunachiyar, “Classification of diabetes patients using kernel based support vector machines,” in Proceedings of the 2020 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–4, Coimbatore, India, June 2020.

DOI: 10.17148/IARJSET.2026.13536

[7]. DD Rufo, TG Debelee, A Ibenthal, WG Negera - Diagnostics Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM) 2021

[8]. J. Zhu, Q. Xie, K. Zheng. "An Improved Early Detection Method of Type-2 Diabetes Mellitus Using Multiple Classifier Systems". Information Sciences, volume 292, pages 1-14, 2015.

[9]. Priyam, A., Gupta, R., Rathee, A., Srivastava, S.: Comparative analysis of decision tree classification algorithms. Int. J. Current Eng. Technol. 3, 334–337, 2277–4106 (2013). arXiv:ISSN

[10]. Esposito, F., Malerba, D., Semeraro, G., Kay, J.: A comparative analysis of methods for pruning decision trees. IEEE Trans. Pattern Anal. Mach. Intell. 19, 476–491 (1997).

[11]. Han, J., Rodriguez, J.C., Beheshti, M.: Discovering decision tree based diabetes prediction model. In: International Conference on Advanced Software Engineering and Its Applications, pp. 99–109. Springer (2008)

[12]. Dhomse Kanchan B., M.K.M., 2016. Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis, in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, IEEE. pp. 5– 10.

[13]. M. Heydari, M. Teimouri, Z. Heshmati, and S. M. Alavinia, "Comparison of various classification algorithms in the diagnosis of type diabetes in Iran," International Journal of Diabetes in Developing Countries, pp. 1-7, 2015

[14]. M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan, "Performance Evaluation of Random Forests and Artificial Neural Networks for the Classification of Liver Disorder," in Proc. International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, pp. 1-5, 2018.

[15]. Ho, T.K.: Random decision forests (PDF). In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995, pp. 278–282 (1995).Archived from the original (PDF) on 17 April 2016. Retrieved 5 June 2016

[16]. M. Kumari, Dr. R. Vohra, and A. Arora, "Prediction of Diabetes using Bayesian Network," International Journal of Computer Science and Information Technologies, vol. 5, pp. 5174-5178, 2014.

[17]. H. M. D. Kabir, M. Abdar, A. Khosravi et al., "Spinalnet: deep neural network with gradual input," IEEE Transactions on Artificial Intelligence, pp. 1–13, 2022.

[18]. R. Kumar, M. Saraswat, D. Ather et al., "Deformation adjustment with single real signature image for biometric verification using CNN," Computational Intelligence and Neuroscience, vol. 2022, Article ID 4406101, 12 pages, 2022.

[19]. Ahmad F, Isa NA, Hussain Z, and Osman MK, "Intelligent medical disease diagnosis using improved hybrid genetic algorithm--multilayer perceptron network,"Journal of Medical Systems, vol. 37, Apr. 2013.

