Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 9, Issue 2, 2022
DOI Journal: 10.37591/JoAIRA

STM JOURNALS

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Research JoAIRA

Timely Diabetes Possibility Prediction Using AI Techniques

Partha Ghosh*

Abstract

The fastest chronic life-threatening disease affecting more than 422 million people globally is diabetes. The primary causes of type 2 diabetes are lifestyle choices and environmental factors. It is a slow-growing disease which starts to develop metabolic factors long before they evolve into a disease and is formally diagnosed by a fasting sugar test. There are records of indications related to diabetes from 520 patients in the dataset that was used. It contains data on people, such as age, sex and symptoms that may lead to diabetes. I analysed the dataset using Naive Bayes Classifier (NB), Logistic Regression Classifier (LR), J48 Algorithm, Random Forest (RF) and Multi-Layer Perceptron (MLP) Algorithm. After applying ten-fold Cross-Validation and Percentage Split evaluation techniques to this dataset, MLP was found to have the best accuracy. MLP achieved 98% accuracy with this dataset and very few numbers of misclassification. In the medical field, dealing with ambiguous and uncertain data is also a significant concern. In recent years, managing ambiguity in medical data has received more attention. The adaptive neuro-fuzzy inference system (ANFIS) was used in this work to identify diabetes. It combined fuzzy logic's learning capabilities with neural networks to describe uncertainty in expressiveness. To represent uncertain circumstances, fuzzy logic is used, and the model is trained by a neural network. The neural network of ANFIS is based on mathematical computations and is linked with the Takagi–Sugeno fuzzy inference system to tackle complicated problems. The Pima Indian Diabetes Dataset (PIDD) was trained and tested for classification using MATLAB. I have utilised this method to diagnose diabetes by leveraging its great uncertainty-handling capabilities and interpretability to produce good classification results.

Keywords: Diabetes risk, Naive Bayes Classifier, Multi-Layer Perceptron (MLP), random forest, logistic regression classifier, symptom, uncertainty

INTRODUCTION

Diabetes diagnosis is considered a difficult issue for quantitative research. Due to some limitations, some parameters such as A1c [1], hectically, white blood cell count, fatty acid oxidation and hematological index [2] have been shown to be insufficient. Vitamin C intake may increase A1c when predicted by electrophoresis but may decrease values when predicted by chromatography [3]. Most studies have indicated that the inflammatory response during hypertension causes a higher number of white blood cells [4]. To accurately predict diabetes, a single parameter just is not very effective and can be confusing in decision-making practice. In order to efficiently predict diabetes at an early point, various parameters need to be consolidated. In our analysis, important characteristics and associations of different characteristics predict diabetes.

Algorithms for machine learning that can simulate uncertainty and produce useful data for

*Author for Correspondence
Partha Ghosh
E-mail: parth_ghos@rediffmail.com

Assistant Professor, Department of Computer Science and Engineering, Govt. College of Engineering and Ceramic Technology, Kolkata, West Bengal, India

Received Date: August 20, 2022
Accepted Date: September 02, 2022
Published Date: September 09, 2022

Citation: Partha Ghosh. Timely Diabetes Possibility Prediction using AI Techniques. Journal of Artificial Intelligence Research & Advances. 2022; 9(2): 37–47p.

better decision-making would be very helpful. Generally, there are two types of uncertainty: model uncertainty and data (noise) uncertainty (also known as epidemiological uncertainty). The noise between the labels is likely due to measurement inaccuracy, which can lead to alleortic uncertainty. Structure uncertainty and uncertainty in model parameters are the two basic categories of model uncertainty. We determine the kind of model structure to be used in structural uncertainty and then specify our suggested model for extrapolating and/or interpolating. In the second category, which is uncertainty in the model parameters, the best model parameters are chosen to produce predictions that are more precise. In medical science, Bayesian inference, fuzzy systems, Monte Carlo simulations, rough categorization, Dempster-Shaffer theory, and imprecise probability are the most often used strategies for handling uncertainty. Therefore, researchers have proposed a number of methods for diagnosis, one of which is a fuzzy based expert system that is suitable for diagnosis in the medical field considering fuzzy clinical data [5, 6]. The Adaptive Neuro-Fuzzy Inference System (ANFIS) [7], a more potent extension, integrates fuzzy logic's capacity for learning with neural networks to explain uncertainty in ambiguity. Uncertain scenarios are modelled using fuzzy logic, and the neural network learns this model over time.

The Adaptive Neuro Fuzzy Inference System (ANFIS) is used in this study to improve classification accuracy by taking adaptability into account via neural networks integrated with fuzzy systems. This system offers various benefits. Without requiring human experience, it may portray the behaviour of a complex system using fuzzy if-then logic, fast convergence time, it uses membership functions and has Ability to learn rapidly and capture the non-linear structure of a process. One of the health care systems with the highest cost-effectiveness is fuzzy-based modelling. It provides precise solutions by removing uncertainties through powerful reasoning capabilities. The Adaptive Neuro-Fuzzy Inference System (ANFIS) combines fuzzy logic and neural network principles in a single frame, allowing for both interpolation and learning. These combined properties enable effective estimation of nonlinear functions. The Takagi-Suzeno Fuzzy Inference System is combined with the neural network in ANFIS, which is based on computations that may address challenging issues.

Complementing these expert systems with knowledge and diagnostic data can provide efficient diagnostic results with minimal error.

RELATED WORK

Shetty *et al*. planned KNN and the Naïve Bayes strategy which have been used for diabetes prediction [8]. As a specialist software application, their methodology was implemented, where users provide feedback in the context of medical records and conclude whether the patient is diabetic or not. Various algorithms were applied on different types of datasets by Singh and Lakshmiganthan [9]. They used algorithms from KNN, Random Forest and Naïve Bayesian. For evaluation, the k-fold cross-validation method was used. For the diagnosis of diabetes, Ahmed used patient data and treatment planning dimensions; Naïve Bayes, Logistic and J48, these three algorithms were combined in the methods [10]. Antony *et al.* used medical evidence to predict diabetes [11]. After pre-processing the data, Naïve Bayes, function-based multilayer perceptual (MLP), and decision-tree-based random forest (RF) procedures were applied. To exclude additional features, a correlation-based feature selection approach was used. A model of learning was then used to predict whether the patient was diabetic. Compared to other machine learning algorithms, the results were enhanced when the Naïve Bayes algorithm was used through a pre-processing system. By expanding the Pima Indian Diabetes Dataset for early diabetes prediction, Amina *et al*. compared different data mining algorithms [12]. Talha *et al.* suggest a clear relationship between diabetes and body mass index (BMI) and the amount of glucose extracted through the A-priority procedure [13]. For diabetes prediction, artificial neural networks (ANN), random forests (RF) and k-means clustering techniques were introduced. The ANN method gave a maximum accuracy of 75.7%. Islam *et al.* had shown that the random forest register is accurate, and that 97.4% accuracy can be achieved [14]. They compared it with decision trees and other learners. Both 10-fold cross-validation and 80/20 split were used.

In addition, some researchers attempted to measure uncertainty and ambiguity in medical diagnostic data sets by using properties of neutrosophic logic [15], neutrosophic graphs [16], hypergraphs [17], and other soft computing techniques. Adequate analysis of medical diagnosis data sets using three-way fuzzy concept lattice calculus was detailed by Singh [18].

EXPERIMENTAL RESULTS AND DISCUSSION

Dataset Used

The dataset was created using a thorough questionnaire that was filled out under a doctor's supervision. Other diabetic factors were considered for this study to determine the diagnosis of pre-diabetic individuals. From the machine learning repository at the University of California, Irvine (UCI), a dataset pertaining to diabetic medicine was gathered [19]. In addition, the Pima Indian Diabetes Dataset (PIDD) was also used [7, 20].

Data Interpretation

The dataset did not contain any missing values. This dataset's data form contains 16 features that were utilised to forecast results, with class variables +ve denoting diabetes in a patient and class variables -ve denoting absence of diabetes.

We have 16 features used to predict the class of diabetes as shown in Table 1. There is clear data with two different outcomes in all characteristics with the exception of age. The age of the patients is between 16 and 90 years.

Table 2 displays the first five rows. The data is in a streamlined format, with variable values in columns, using an observation in each row. The feature size is (520, 17). 38.5% of the patients did not have diabetes and 61.5% had diabetes, as shown in Figure 1.

Patient gender split was 37% for women and 63% for men. 90% of women had diabetes while 45% of men had diabetes.

Table 1. 16-features or attributes.

Attributes	**Description**
Age (years)	16–90
Sex	1. Male, 2. Female
Polyuria	1. Yes, 2. No.
Polydipsia	1. Yes, 2. No.
Sudden weight loss	1. Yes, 2. No.
weakness	1. Yes 2. No.
Polyphagia	1. Yes 2. No.
Genital thrush	1. Yes 2. No.
Visual blurring	1. Yes 2. No.
Itching	1. Yes 2. No.
Irritability	1. Yes 2. No.
Delayed Healing	1. Yes 2. No.

Partial paresis	1. Yes 2. No.
Muscle stiffness	1. Yes, 2. No.
Alopecia	1. Yes 2. No.
Obesity	1. Yes 2. No.
Class	1. Positive 2. Negative

Table 2. Display first five rows.

	0	1	2	3	4
Age	40	58	41	45	60
Gender	Male	Male	Male	Male	Male
Polyuria	No	No	Yes	No	Yes
Polydipsia	Yes	No	No	No	Yes
Sudden weight loss	No	No	No	Yes	Yes
weakness	Yes	Yes	Yes	Yes	Yes
Polyphagia	No	No	Yes	Yes	Yes
Genital thrush	No	No	No	Yes	No
Visual blurring	No	Yes	No	No	Yes
Itching	Yes	No	Yes	Yes	Yes
Irritability	No	No	No	No	Yes
Delayed Healing	Yes	No	Yes	Yes	Yes
Partial paresis	No	Yes	No	No	Yes
Muscle stiffness	Yes	No	Yes	No	Yes
Alopecia	Yes	Yes	Yes	No	Yes
Obesity	Yes	No	No	No	Yes
Class	Positive	Positive	Positive	Positive	Positive

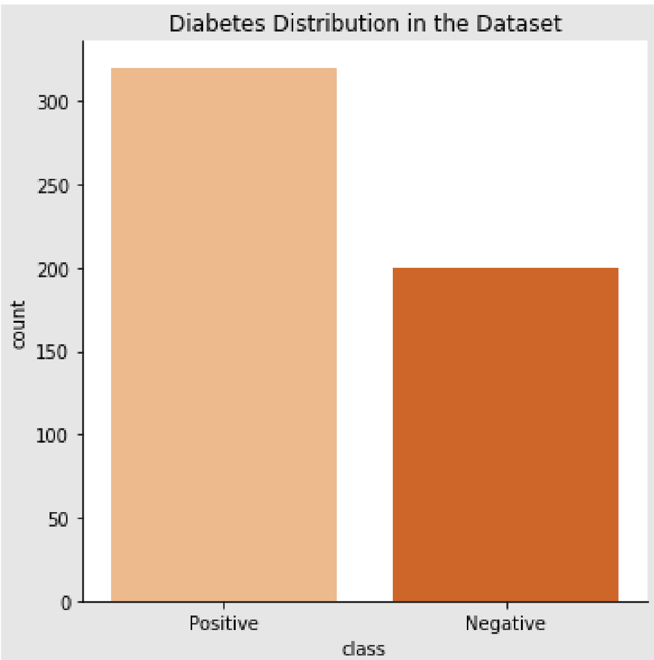

Figure 1. Visualization for distribution of diabetes in the sample.

Data Manipulation

To perform tasks for machine learning and correlation tasks, datasets were converted from non-numeric labels to numerical items. So, first we have removed the numerical data (Age) and converted the remaining categorical data into numerical data as shown in Table 3.

Table 3. Deleted numerical data ("Age") and dataset was converted into numeric items.

	0	1	2	3	4
Gender	1	1	1	1	1
Polyuria	0	0	1	0	1
Polydipsia	1	0	0	0	1
Sudden weight loss	0	0	0	1	1
weakness	1	1	1	1	1
Polyphagia	0	0	1	1	1
Genital thrush	0	0	0	1	0
Visual blurring	0	1	0	0	1
Itching	1	0	1	1	1
Irritability	0	0	0	0	1
Delayed Healing	1	0	1	1	1
Partial paresis	0	1	0	0	1
Muscle stiffness	1	0	1	0	1
Alopecia	1	1	1	0	1
Obesity	1	0	0	0	1
Class	1	1	1	1	1

Again, we added "Age" back to the dataset as shown in Table 4.

Table 4. Transformed numerical data set with "Age" column.

	0	1	2	3	4
Gender	1	1	1	1	1
Polyuria	0	0	1	0	1
Polydipsia	1	0	0	0	1
Sudden weight loss	0	0	0	1	1
weakness	1	1	1	1	1
Polyphagia	0	0	1	1	1
Genital thrush	0	0	0	1	0
Visual blurring	0	1	0	0	1
Itching	1	0	1	1	1
Irritability	0	0	0	0	1
Delayed Healing	1	0	1	1	1
Partial paresis	0	1	0	0	1
Muscle stiffness	1	0	1	0	1
Alopecia	1	1	1	0	1
Obesity	1	0	0	0	1
Class	1	1	1	1	1
Age	40	58	41	45	60

It is observable from Figure 2 that there was a low correlation between age (0.11 correlation coefficient) and diabetes. Polyuria and polydipsia, with correlations of 0.67 and 0.65, respectively, had the highest association with diabetes. Delayed healing and itching had the lowest association with diabetes, with a correlation of –0.01 and 0.05, respectively. Most of the characteristics had low diabetes association. In addition, we transformed the properties of age into categorical variables (1.15–25, 2.43, 3.36–45, 4.46–55, 5.56–65, and 6.00 above the age of 65) and examined whether diabetes was related to age groups as that is shown in Table 5.

First, we plot a bar graph to look at the distribution of age groups to test whether there is an association between age group and diabetes. Within the age group attribute, we then plot the distribution of diabetes as shown in Figure 3.

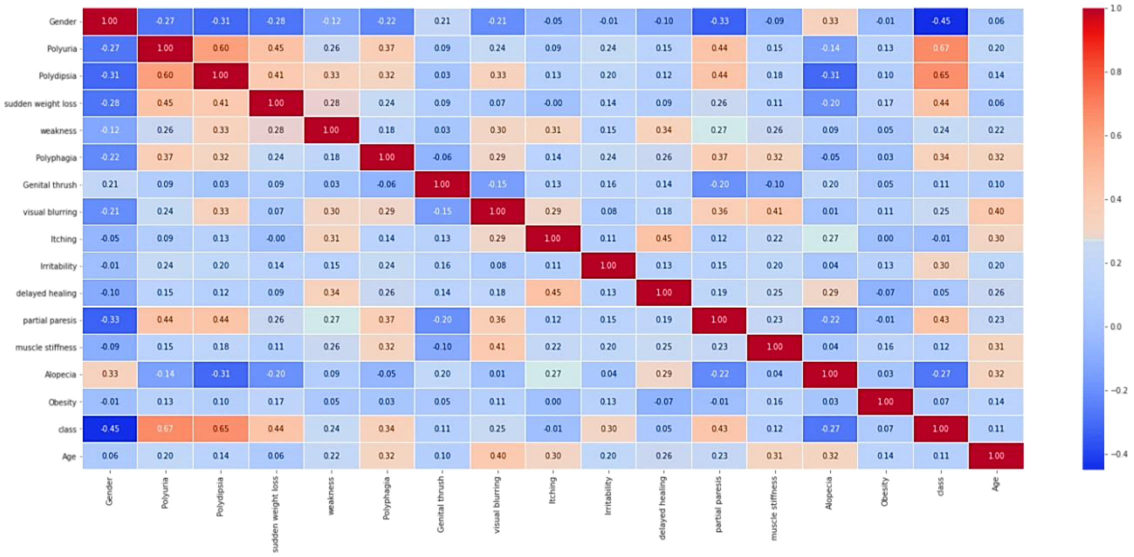

Figure 2. Correlation matrix heat-map for the early diabetes prediction dataset.

Table 5. The age attribute changed to categorical variables.

	0	1	2	3	4
Gender	1	1	1	1	1
Polyuria	0	0	1	0	1
Polydipsia	1	0	0	0	1
Sudden weight loss	0	0	0	1	1
Weakness	1	1	1	1	1
Polyphagia	0	0	1	1	1
Genital thrush	0	0	0	1	0
Age Group	3	6	3	3	5
Visual blurring	0	1	0	0	1
Itching	1	0	1	1	1
Irritability	0	0	0	0	1
Delayed Healing	1	0	1	1	1
Partial paresis	0	1	0	0	1
Muscle stiffness	1	0	1	0	1
Alopecia	1	1	1	0	1
Obesity	1	0	0	0	1
Class	1	1	1	1	1
Age	40	58	41	45	60

From the percentage of 'age group' and 'class distribution', we can conclude that in age group 4 (i.e., 46–55 years), the patients had more symptoms of diabetes. As such, the association between age group and diabetes is not statistically significant because diabetes is distributed equally across all age groups.

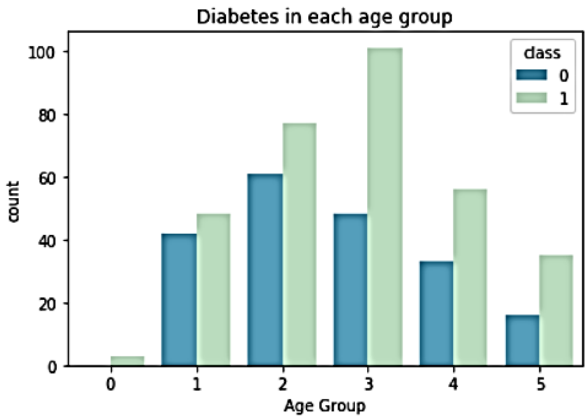

Figure 3. Age group and diabetic class distribution.

Subsequently, a chi-square test for age and diabetes was performed, resulting in a p-value of 0.076. P-values >0.05, thus we do not disprove the null hypothesis that there is not a relationship between age group and diabetes.

The obtained p-value was less than 0.05 when the chi-square test was performed to examine the association between polyuria and diabetes. As a result, we declare that, as illustrated in Figure 4, there is a significant association between polyuria and diabetes and reject the null hypothesis.

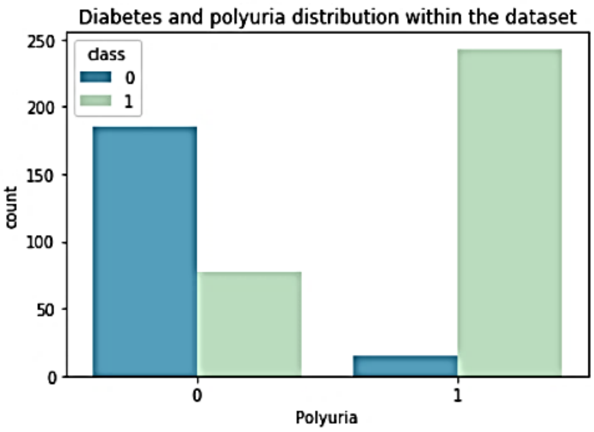

Figure 4. Distribution of diabetes in patients who had polyuria.

Diabetes Prediction Using Various Classifiers

I have analysed using Naive Bayes Classifier (NB), Logistic Regression Classifier (LR), J48 Algorithm [10], Random Forest (RF) and Multi-layer Perceptron (MLP) algorithms, as shown in the Table 6. After using the ten-fold cross-validation and percentage split evaluation methodologies, as shown in Table 6, it was discovered that MLP had the best accuracy on this dataset. For widespread use within the medical field for the prediction of complex symptoms, the multi-layer perceptron (MLP) was de facto the neural network of preference. In order to train the model and evaluate how well it worked, data was divided into training and test sets. 20% of the dataset was used for test data, while the remaining 80% was divided into training sets.

The Random Forest Regressor can achieve 97.4% accuracy but our MLP cannot. With using UCI Machine Learning dataset [19] we achieve 98% of accuracy for MLP.

The size of the hidden layer was set to three layers of 12 nodes and a cumulative iteration of 1000 was used by our design. The test data is predicted after the training of the data and the success of the technique is finally analysed.

Table 6. Comparison of performance using 10-fold cross-validation.

Performance Parameters	Class	Algorithms				
		NB	***LR***	***J48***	***RF***	***MLP***
Precision	Weighted Average	0.879	0.924	0.957	0.974	0.985
Recall	Weighted Average	0.874	0.924	0.956	0.974	0.980
F-measure	Weighted Average	0.875	0.924	0.956	0.974	0.980

Confusion Matrix:

[[44 2]
[0 58]]

Classification Report:

	precision	**recall**	**f1-score**	**support**
0	1.00	0.96	0.98	46
1	0.97	1.00	0.98	58
Accuracy			0.98	104
macro avg	0.98	0.98	0.98	104
weighted avg	0.98	0.98	0.98	104

We observed that two (2) out of 104 patients were incorrectly labelled with a confusion matrix, resulting in 98% accuracy and 98% F-score, which are strong predictive measures.

Proposed Methodology to Deal with Uncertainty

To access the Pima Indian Diabetes Dataset intrinsic degree of uncertainty and estimate the likelihood of diabetes, an expert system based on fuzzy logic was created in the current study, after the design of a classification model based on ANFIS.

The obtained results were then compared to the multi-layer perceptron-based categorization approach (MLP). The findings demonstrated that the ANFIS model offers greater classification accuracy.

Pima Indian Diabetes Dataset was used to create the expert system (PIDD). Men and women of various age groups are susceptible to diabetes mellitus. Each patient's input was recorded in order to calculate the likelihood that they would develop diabetes, including their age, body mass index (BMI), diabetic pedigree function (DPF), glucose level, and insulin level. These test results were transformed into fuzzy data with output values between 0 and 1, taking linguistic medical assumptions into account. Thereafter, a knowledge-based decision was taken to derive the result from the fuzzy values. With diffusivity, final values were obtained as crisp values on an arbitrary scale of the fuzzy output variable that describes the probability of a diabetes diagnosis.

Expert System with Fuzzy Logic

Identifying the input and output characteristics was the initial stage in the suggested system. In accordance with clinical likelihood, outputs were classified in descending order of severity: very low, low, medium, high, and very high. Membership values were defined to obscure the input attributes. And the trapezoidal membership function was used to perform the fuzzification.

The fuzzy operator 'AND' or 'OR' was applied to the precedent and then inferred in the next step, i.e., from the precedent to the resultant and then the resultant was added to the entire defined rule. The opinions of professionals in the medical area served as the foundation for these regulations. The Fuzzy Logic MATLAB Toolbox was used for this purpose.

The following stage was diffusion to provide a single output value that was sharp using the centroid approach. Thus, the result gave us the possibility of diagnosis.

Adaptive Neuro-fuzzy Inference System

The Takagi-Suzeno Fuzzy Inference System is the foundation of ANFIS [7, 21]. This architecture consists of five layers of input features and one output. The first and second layers, respectively, are the Fuzzification and Rule layers. The third layer, which normalised the values, provides input to the fourth layer. The final layer receives diffuse-fixed data and produces the output.

The learner element was represented by Layers I and IV. The first layer factor determines the membership functions which in this case was trapezoidal. A hybrid algorithm was implemented for training and optimization that uses the least squares estimation method to update the coefficients of the output, thus optimizing the base and resulting parameters and faster than the back-propagation algorithm, which only updates the fundamental parameters.

Training and Testing

For the initial training of the network, input and output vectors were inserted into 80% of PIDD values. The fuzzy inference system was generated by choosing the hybrid optimization algorithm and the trapezoidal membership function 3-3-3-3-3.

The remaining 20% of the PIDD was used to test the ANFIS model upon completion of the training phase. Root Mean Square Error (RMSE) values computed during the training and testing processes were used to confirm the model's performance. Both the training and testing procedures were repeated for different epoch values.

RESULT

Chances of Diabetes

Based on the probability or severity of diabetes, the fuzzy model's outputs were divided into five groups: very low, low, moderate, high, and very high.

Hybrid Algorithm-based ANFIS Model Demonstration

In addition, the generated ANFIS model was trained and tested in the MATLAB Fuzzy Logic Toolbox. The performance efficiency of the model was validated through the RMSE values determined for the different EPOCH. In different EPOCHs (e.g., 10, 20, 50, 100, 150), we obtained the same RMSE (0.2276) for both the training and testing phases.

In order to categorize related the Pima Indian Diabetes Dataset, a multi-layer perceptron (MLP) neural network was created. In order to compare the confusion matrix's results to those produced by the ANFIS model, statistical measures like accuracy, sensitivity, and specificity were assessed. The results are shown in Table 7.

Table 7. Performance validation of different classification models.

Classification model →	**MLP**	**ANFIS [7, 21]**
Performance Metrics ↓		
Sensitivity (%)	98.7	93
Specificity (%)	72.5	74.17
Accuracy (%)	85.3	*87.24*

The obtained results showed that the ANFIS classification model had better accuracy than the MLP classification model.

CONCLUSION

From the confusion matrix of Diabetes prediction using various classifiers, out of 104 samples, only 2 patients were classified incorrectly, resulting in 98% accuracy and 98% percentile F-score. These are markers of successful prediction. Using machine learning techniques, both specific and less specific diabetes symptoms can be used for early detection of diabetes. The accuracy of the MLP machine learning model makes it a superior suite. Therefore, diabetes can be predicted using the MLP machine learning model at an early stage.

Uncertainty in the data and in the model are both factors to be considered while studying uncertainty. Data uncertainty can be caused by measurement noise, transmission noise, and missing values, among other things. Uncertainty about a model's architecture and parameters makes it difficult to anticipate future data. Quantification of uncertainty helps in increasing confidence in the results obtained by various methods. To make appropriate conclusions in the medical field, researchers must understand how to cope with ambiguity in both data and models. The most popular method is the fuzzy system. These techniques are straightforward and effective at getting rid of different kinds of uncertainties.

A fuzzy based expert method has been used to determine the likelihood of diabetes mellitus in the early stages as discussed in the proposed methodology section. This method has the ability to make decisions by accepting accurate and unambiguous medical data simulating human control logic and thus ensures high accuracy, unlike traditional methods. ANFIS was used in conjunction with a hybrid algorithm to further enhance the fuzzy system. The results discussed, proved to classify the Pima Indian Diabetes Dataset at better accuracy than other classification algorithms such as MLP. The suggested strategy reduces complexity while also being cost-effective. The models are to be compared with different machine learning techniques to provide efficient diagnostics in the future.

REFERENCES

1. Cobos Leopoldo. Unreliable hemoglobin A1C (HBA1C) in a patient with new onset diabetes after transplant (nodat). Endocr Pract. 2018; 24: 43–44.
2. Dorcely Brenda, Karin Katz, Ram Jagannathan, Chiang Stephanie S, Babajide Oluwadare, Goldberg Ira J, Michael Bergman. Novel biomarkers for prediabetes, diabetes, and associated complications. Diabetes, Metab Syndr Obes: Targets Ther. 2017; 10: 345–361.
3. Radin Michael S. Pitfalls in hemoglobin A1c measurement: when results may be misleading. J Gen Intern Med. 2014; 29(2): 388–394.
4. Merad-Boudia Hamza Nadjib, Majda Dali-Sahi, Youcef Kachekouche, Nouria Dennouni-Medjati. Hematologic disorders during essential hypertension. Diabetes Metab Syndr: Clin Res Rev. 2019; 13(2): 1575–1579.
5. Toğaçar Mesut, Burhan Ergen, Zafer Cömert. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. Comput Biol Med. 2020; 121: 103805.
6. Rundo Leonardo, Lucian Beer, Stephan Ursprung, Paula Martin-Gonzalez, Florian Markowetz, Brenton James D, Mireia Crispin-Ortuzar, Evis Sala, Ramona Woitek. Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering. Comput Biol Med. 2020; 120: 103751.
7. Karaboga Dervis, Ebubekir Kaya. Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif Intell Rev. 2019; 52(4): 2263–2293.
8. Shetty Deeraj, Kishor Rit, Sohail Shaikh, Nikita Patil. Diabetes disease prediction using data mining. In 2017 international conference on innovations in information, embedded and communication systems (ICIIECS), IEEE. 2017; 1–5.
9. Singh Asmita, Lakshmiganthan R. Impact of different data types on classifier performance of random forest, Naive Bayes, and k-nearest neighbors algorithms. Int J Adv Comput Sci Appl. 2018; 8(12): 1–10.
10. Ahmed Tarig Mohamed. Using data mining to develop model for classifying diabetic patient control level based on historical medical records. J Theor Appl Inf Technol. 2016; 87(2): 316–323.
11. Asir Antony Gnana Singh Danasingh. Diabetes prediction using medical data. Journal of Computational Intelligence in Bioinformatics (JCIB). 2017; 10(1): 1–8.

12. Azrar Amina, Ali Y, Awais M, Khurram Zaheer. Data mining models comparison for diabetes prediction. Int J Adv Comput Sci Appl. 2018; 9(8): 320–323.
13. Alam Talha Mahboob, Muhammad Atif Iqbal, *et al.* A model for early prediction of diabetes. Inform Med Unlocked. 2019; 16: 100204.
14. Faniqul Islam M, Rahatara Ferdousi, Sadikur Rahman, Humayra Yasmin Bushra. Likelihood prediction of diabetes at early stage using data mining techniques. In Computer Vision and Machine Intelligence in Medical Image Analysis. Singapore: Springer; 2020; 113–125.
15. Singh Prem Kumar. Three-way fuzzy concept lattice representation using neutrosophic set. Int J Mach Learn Cybern. 2017; 8(1): 69–79.
16. Kandasamy Vasantha, Ilanthenral K, Florentin Smarandache. Neutrosophic graphs: a new dimension to graph theory. Infinite Study; 2015.
17. Somu Nivethitha, Gauthama Raman MR, Kannan Kirthivasan, Shankar Sriram VS. Hypergraph based feature selection technique for medical diagnosis. J Med Syst. 2016; 40(11): 1–16.
18. Singh Prem Kumar. Medical diagnoses using three-way fuzzy concept lattice and their Euclidean distance. Comput Appl Math. 2018; 37(3): 3283–3306.
19. UCI Machine Learning Repository. Early stage diabetes risk prediction dataset. [Online]. Data Set. Available from https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.
20. Kaggle. Pima Indians Diabetes Database: Predict the onset of diabetes based on diagnostic measures [Online]. Available from https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
21. J-SR. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern. 1993; 23(3): 665–685.

