ISSN 2085-4552

Application of the ANFIS Model in Predicting Diabetes Mellitus Disease

Aprilia Nurfazila[1], Hetty Rohayani[2]
[1,2]Informatics Study Program,Muhammadiyah University Jambi, Jambi, Indonesia
[1] aprilianurfazila5@gmail.com, [2] hettyrohayani@gmail.com

Accepted 19 December 2025
Approved 07 January 2026

***Abstract*— This study presents the application of the Adaptive Neuro-Fuzzy Inference System (ANFIS) model for predicting Diabetes Mellitus using two primary input features, namely glucose level and body mass index (BMI). The research employs a quantitative experimental approach using the public diabetes dataset obtained from Kaggle. The data underwent preprocessing steps, including cleaning, normalization, and splitting into training and testing subsets. The ANFIS model was designed with fuzzification, rule-based inference, and a hybrid learning algorithm to optimize membership function parameters. Model evaluation was conducted using accuracy, precision, recall, and F1-score. The results show that the ANFIS model achieved an accuracy of 69.70% on the test dataset, demonstrating strong sensitivity in detecting diabetic cases but generating a notable number of false positives. These findings indicate that ANFIS has potential as an early-screening decision support tool, although further optimization and additional features are required to enhance predictive performance.**

Index Terms*— ANFIS; diabetes prediction; fuzzy logic; machine learning; medical diagnosis.

I. Introduction

Diabetes Mellitus (DM), particularly Type 2 Diabetes, remains one of the major health challenges in Indonesia and globally. According to the International Diabetes Federation (IDF), approximately 19.47 million adults in Indonesia were living with diabetes in 2021 [1]. The dataset employed in this study, the Pima Indians Diabetes Dataset, specifically represents a population with a high prevalence of Type 2 Diabetes, which is generally associated with insulin resistance and lifestyle factors.

One of the main problems in managing DM is the diagnostic process, which is often delayed and complex, as it involves various risk factors and medical parameters such as blood glucose levels, blood pressure, body mass index (BMI), family history, and patient age [2]. Furthermore, patient data characteristics are often non-linear and uncertain (fuzzy), making them difficult to model conventionally using simple mathematical approaches.

In this context, artificial intelligence-based systems are increasingly being used to assist the diagnostic process. An expert system is one branch of artificial intelligence that makes extensive use of specialized knowledge to solve problems [3]. An expert system consists of two main parts: the Development Environment and the Consultant Environment [4]. However, traditional expert systems, such as the Certainty Factor method or Rule-Based Systems, still have limitations as they are unable to adapt to new data and cannot handle complex relationships between variables [5].

The Adaptive Neuro-Fuzzy Inference System (ANFIS) model is a method developed or implemented from a Fuzzy Inference System and NNW (Neural Network) [6]. The ANFIS method can overcome the difficulties faced by ANN (Artificial Neural Network) methods, namely determining the number of layers, and fuzzy methods, specifically determining the rules to be used [7]. In general, Fuzzy Logic has the ability to process vague data with if-then rules, while ANNs are capable of learning from data and adjusting parameters to improve system accuracy [8]. By combining these two methods, ANFIS is able to form an adaptive, accurate, and efficient predictive model for analyzing complex patterns in medical data.

Recent research in Indonesia indicates that the application of ANFIS can yield more accurate results in predicting Diabetes Mellitus compared to other methods such as Naïve Bayes or Decision Tree. For example, research by D. Kurniawan et al. (2024) showed that the ANFIS model was able to increase diagnostic accuracy to 95% in classifying diabetes risk based on clinical patient data [9]. Furthermore, other studies have also proven the effectiveness of ANFIS in other medical domains, such as predicting stunting and heart disease, demonstrating its ability to map non-linear relationships between symptoms and diagnostic outcomes [10].

The implementation of the ANFIS model in a web-based platform is also considered to have great potential in supporting digital health services in Indonesia. With a web-based system, the public can perform early screening independently, anytime and anywhere, thus supporting government efforts in the digital transformation of the healthcare sector [11].

Therefore, this research is titled "Application of the Adaptive Neuro-Fuzzy Inference System (ANFIS) Model in the Prediction of Diabetes Mellitus," which aims to develop an intelligent web-based system capable of accurately and efficiently predicting diabetes risk, as well as serving as a decision support solution for medical professionals and the public.

II. METHOD

A. Type of Research

This study is classified as quantitative experimental research, which aims to develop and test a diabetes prediction model using the Adaptive Neuro-Fuzzy Inference System (ANFIS). The quantitative approach was chosen because this research focuses on the analysis of numerical data derived from the dataset and measures model performance using statistical metrics such as accuracy, precision, recall, and F1-score. This research is conducted experimentally because the researcher builds, trains, and tests the model to obtain empirical results.

B. Source and Type of Data

The data utilized in this study is secondary data from the Pima Indians Diabetes Dataset, consisting of 768 samples. Based on preliminary analysis, non-physiological zero values (indicating missing values) were identified in the **Glucose** (5 instances) and **BMI** (11 instances) attributes, which were subsequently addressed during the pre-processing stage. The attributes used in this research are as follows:

- **Pregnancies**: number of pregnancies,
- **Glucose**: Plasma glucose concentration at 2 hours in an oral glucose tolerance test (OGTT). This indicates that the data reflects the body's response to a glucose load, rather than merely fasting blood sugar [12].
- **Blood Pressure**: Diastolic blood pressure (mm Hg).
- **BMI**: Body Mass Index (weight in kg / (height in m)²).
- **Outcome**: Target variable (0 = Negative, 1 = Positive for Diabetes).

This dataset is public in nature, allowing it to be used for academic research without licensing restrictions. The type of data employed is secondary data, as it was previously collected by a third party (Kaggle) and is being reused for further analysis.

C. Proposed System Workflow

The research workflow is systematically designed to process raw data into diagnostic decisions. The process initiates with data pre-processing to address missing values in the Glucose and BMI attributes, as well as data normalization. Subsequently, the data is split into training data (70%) and testing data (30%) [13]. The core stage involves constructing the ANFIS structure, where premise and consequent parameters are trained until the model achieves convergence. Finally, the model is evaluated using the Confusion Matrix metric to measure detection sensitivity and precision.

D. ANFIS Architecture

This study implements a first-order Sugeno-type Adaptive Neuro-Fuzzy Inference System (ANFIS) model. The network structure consists of five layers with the following mathematical formulations:

Layer 1 (Fuzzification): Each node in this layer is adaptive and is associated with a membership function:

$$O1,i = \mu Ai(x)$$

Where X is the input (BMI or Glucose) and A_I is the linguistic label (Low, Medium, High) [14].

Layer 2 (Rules): Each node computes the *firing strength* of the rule using the multiplication (AND) operation:

$$O_i^2 = wi = \mu_{A_i}(x)\mu_{B_i}(y)\ i = 1,2$$

Layer 3 (Normalization): Calculates the ratio of the i-th rule's firing strength to the total firing strength:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}\ i = 1,2$$

Layer 4 (Defuzzification): Each node computes the rule's contribution to the crisp output using a linear function:

$$O_{4,i} = \overline{w_l} f_i = \overline{w_l}(k_{i1}x_1 + k_{i2}x_2 + k_{i0})$$

Where $\{ki1,\ ki2,\ ki0\}$ are the optimized consequent parameters [15].

Layer 5 (Output): The total output is computed as the summation of all incoming signals:

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{w_1 + w_2}$$

III. RESULT AND DISCUSSIONS

A. General Dataset Description

The dataset used is the Diabetes Mellitus dataset (often known as the Pima Indians Diabetes Dataset), which contains data from health examinations of several female patients of Pima Indian descent. This dataset has eight input variables and one output variable (Outcome) that indicates whether the patient was diagnosed with diabetes (1) or not (0).

Table 3.1 Dataset

No	Attribute Name	Description
1	Pregnancies	Number of pregnancies
2	Glucose	Plasma glucose concentration
3	Blood Pressure	Diastolic blood pressure (mm Hg)
4	Skin Thickness	Triceps skinfold thickness (mm)
5	Insulin	Serum insulin level (mu U/ml)
6	BMI	Body Mass Index
7	Diabetes Pedigree Function	Diabetes pedigree function
8	Age	Patient age
9	Outcome	Target class: 1 = Diabetes, 0 = Not Diabetes

B. *Membership Function Characteristics*

The ANFIS model maps numerical input variables into fuzzy sets to address the uncertainty inherent in medical data. Figure 3.1 and Figure 3.2 illustrates the membership functions for the Glucose and BMI variables.

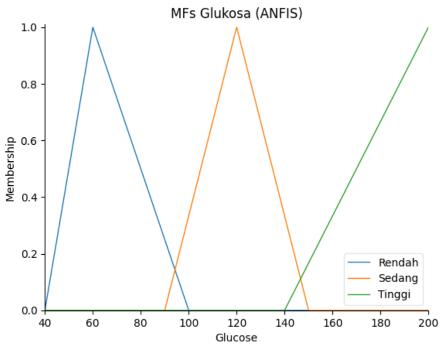

Figure 1 MFs Glucose

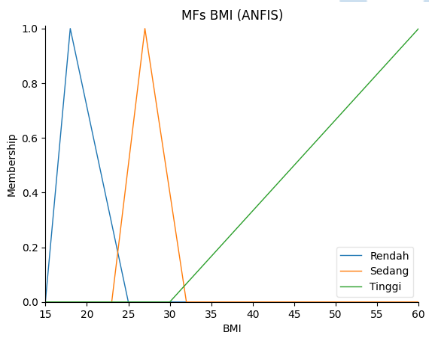

Figure 2 MFs BMI

Triangular membership functions (*trimf*) are utilized to partition the input domain into three linguistic labels: 'Low', 'Medium', and 'High'. The presence of overlap between sets (for instance, at Glucose levels of 140–150 mg/dL) enables the system to perform interpolative reasoning, wherein patients with borderline glucose levels are not categorized rigidly; instead, they possess a partial degree of membership across both categories.

C. *Model Performance Evaluation*

Based on testing conducted on 231 test samples (constituting 30% of the dataset), the model's performance was evaluated using the Confusion Matrix presented in Figure 3.3.

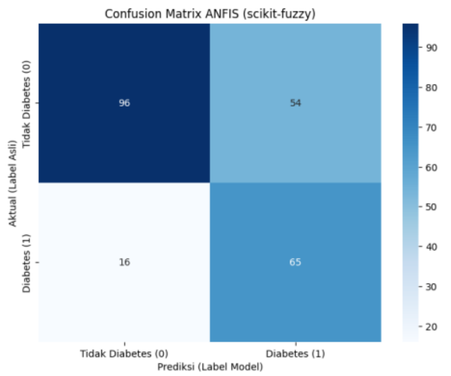

Figure 3 Confusion Matrix

As indicated in the figure above, the model yielded a total accuracy of 69.70%. However, evaluation metrics do not rely solely on accuracy. The model successfully identified 65 out of 81 positive diabetes cases, resulting in a Sensitivity (Recall) value of 80.2%.

D. *Comparative Analysis and Discussion*

The accuracy achieved in this study (69.70%) is notably lower compared to the reference study [9], which reached 95%. This disparity in results highlights two significant findings:

1. **Impact of the Number of Features:** The reference study [9] utilized all 8 clinical attributes of the dataset, whereas this study restricted the input to only two variables (Glucose and BMI). This reduction in accuracy confirms that although Glucose and BMI are primary indicators, a high-precision diabetes diagnosis requires additional variables, particularly Insulin and Family History (Diabetes Pedigree Function).
2. **Clinical Trade-off:** Although the model exhibits low precision (0.55), indicating a high number of False Positives (54 cases), it demonstrates a sufficiently high sensitivity (0.80). In the context of preventive medicine, high sensitivity is prioritized for early screening tools to minimize missed positive patients (False Negatives). False Positive errors can be tolerated as they will be confirmed through subsequent laboratory tests, whereas False Negatives pose a fatal risk as the patient remains undetected.

IV. CONCLUSION

The application of the ANFIS model utilizing two primary input features (Glucose and BMI) on the Pima Indians dataset yielded a moderate accuracy of 69.70%. Although this accuracy is lower than that of the reference model [9], the system successfully achieved a sensitivity of 80%, rendering it a viable potential tool for pre-screening (early detection) to triage high-risk patients prior to further medical examination.

For future development aimed at enhancing clinical validity, the following are recommended:

1. **Integration of Insulin Variable:** Incorporating the 'Insulin' attribute as a third input is highly recommended, given the direct role of this hormone in the pathophysiology of Type 2 Diabetes.
2. **Statistical Threshold Determination:** Substituting manual fuzzy interval determination with statistical methods (such as C-Means Clustering) to define 'Low/Medium/High' boundaries that are more adaptive to the actual distribution of patient data.

REFERENCES

[1] A. Prasetyo, "Analisis Peningkatan Kasus Diabetes di Indonesia Berdasarkan Data IDF 2021," Jurnal Kesehatan Masyarakat Indonesia, vol. 9, no. 2, 2023.

[2] N. Rahmansyah, S. A. Lusinia, dan I. Ilmawati, "Analisa Prediksi Penyakit Diabetes Menggunakan Metode Naïve Bayes dan K-NN," Innovative: Journal of Social Science Research, vol. 5, no. 1, 2024.

[3] S. Agustina, H. Rohayani, N. Marthiawati, dan K. Kurniawansyah, "Rancang Bangun Sistem Pakar untuk Mendiagnosa Kerusakan Handphone Android dengan Metode Forward Chaining," Journal of Informatics Management and Information Technology, vol. 3, no. 3, pp. 103-111, Jul. 2023.

[4] M. Alfareza, H. Rohayani, K. Kurniawansyah, dan N. Marthiawati, "Rancang Bangun Sistem Pakar Untuk Mendiagnosa Kerusakan Hardware Komputer dengan Metode Forward Chaining," Journal of Informatics Management and Information Technology, vol. 3, no. 3, pp. 97-102, Jul. 2023.

[5] S. Sugiono dan A. Junior, "Klasifikasi Sistem Pakar untuk Mendiagnosa Penyakit Diabetes Menggunakan Metode Fuzzy Sugeno," Jurnal Pendidikan dan Konseling (JPDK), vol. 4, no. 5, 2022.

[6] S. Rahmah, W. Witanti, and P. N. Sabrina, "Prediksi Penjualan Obat Menggunakan Metode Adaptive Neuro-Fuzzy Inference System (ANFIS)," J. Inform. Merdeka Pasuruan (JIMP), vol. 7, no. 3, Des. 2022.

[7] D. A. Lusia, K. Semathea, E. Sumarminingsih, and A. Efendi, "Perbandingan Metode Jaringan Saraf Tiruan, Fuzzy, dan ANFIS pada Peramalan Data Inflasi Indonesia," J. Teknol. Inf. Ilmu Komput. (JTIIK), vol. 11, no. 3, pp. 653-660, Jun. 2025.

[8] M. F. Rahman, "Implementasi Metode Adaptive Neuro-Fuzzy Inference System (ANFIS) dalam Sistem Prediksi Penyakit Jantung," Jurnal Ilmiah Teknologi dan Rekayasa, vol. 6, no. 1, 2023.

[9] D. Kurniawan, H. Prabowo, dan E. Widodo, "Prediksi Risiko Diabetes Menggunakan Model Adaptive Neuro-Fuzzy Inference System (ANFIS)," Jurnal Teknologi Informasi dan Komputer (JTIK), vol. 10, no. 2, 2024.

[10] A. Lestari dan R. Putra, "Penerapan ANFIS untuk Prediksi Stunting di Asia Tenggara," Jurnal Data Sains dan Kecerdasan Buatan Indonesia, vol. 3, no. 1, 2024.

[11] F. B. Bahtiarullah, A. Homaidi, dan F. Santoso, "Pengembangan Sistem Prediksi Penyakit Diabetes Berbasis Web," E-Link: Jurnal Teknik Elektro dan Informatika, vol. 19, no. 2, 2024.

[12] D. Nopriyani, H. Rohayani, dan Z. Akbar, "Penerapan Algoritma K-Means untuk Mengidentifikasi Pola Penjualan Frozen Food yang Paling Populer," Bulletin of Computer Science Research, vol. 5, no. 1, pp. 98-104, Des. 2024.

[13] M. F. Aditya, A. Pramuntadi, and D. P. Kusumaningrum, "Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1104-1110, Jul. 2024.

[14] A. Lenteraningati, J. H. Jaman, and C. Rozikin, "Sistem Pakar Diagnosis Penyakit Diabetes Melitus Menggunakan Metode Fuzzy Inference System (Mamdani)," *Jurnal Informatika dan Komputer*, vol. 7, no. 2, pp. 150-159, 2024.

[15] S. M. Putri and E. Santoso, "Optimasi Fungsi Keanggotaan Fuzzy Menggunakan Algoritma Genetika pada Diagnosis Diabetes," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 9, pp. 4455-4463, 2023