University of Guelph

An Final Report on

"Comparative Study: ANN, ANFIS and Hybrid ANN-ANFIS for Diabetes."

Submitted By:

Rohal Patel (1357210)

Subheksha Tandan (1350311)

Vishal Tailor (1376030)

Institution: University of Guelph

Course: Advance Soft Computing (ENGG*6570-01)

Instructor: Dr. Simon X. Yang

Date: 2025-08-12

Abstract

Diabetes is one of the most common chronic diseases that exists in many parts of the world. It is imperative to identify diabetes issues as early as possible to manage and prevent significant complications. This project aims to conduct a comparative study of three soft computing-based classification methods- Feedforward Neural Network (FNN), Adaptive Neuro-Fuzzy Inference System (ANFIS), and a Hybrid ANN-ANFIS model, on the Pima Indians Diabetes Dataset (Faust, 2018). This study aimed at producing and comparing predictive soft computing models to determine diabetes risk in patients based on clinical parameters. We experimented with the Pima Indian Diabetes database, which comprised 768 patients with complete clinical measurements and were checked for diabetes. The data included eight clinical parameters: pregnancies, glucose level, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, and age. Three soft computing models were employed to predict diabetes: Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), and a new Hybrid ANN-ANFIS ensemble approach. The Hybrid ANN-ANFIS model performed the highest accuracy with 74.0% accuracy, 62.5% precision, 64.8% recall, 63.6% F1-score, and 0.824 area under the curve (AUC) on the validation set. Receiver operating characteristic curve, confusion matrix, and predictive probability histogram showed that the Hybrid model was good for discrimination and balanced classification. Calibration curve analysis also showed the stability and usability of the Hybrid ANN-ANFIS model with high agreement of predicted probabilities with true results. The Hybrid ANN-ANFIS model demonstrated its clinical applicability in diabetic disease screening among high-risk patients and exhibited improved performance by leveraging the neural network learning capability and fuzzy logic interpretability.

Keywords: *Diabetes Prediction, Pima Indians Diabetes Dataset, Feedforward Neural Network (FNN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Hybrid ANN-ANFIS, Soft Computing Techniques, Medical Data Classification*

Table of Contents

1 Introduction

1.1 Background

To value the rapid advancement of artificial intelligence (AI) and soft computing, the study of intelligent healthcare prediction systems has received considerable focus for early diagnosis and prevention of diseases in time. Diabetes is the most common chronic disease around the world. According to recent estimates, more than 537 million adults worldwide, or 10.5% of the population aged 20 to 79, have diabetes, underscoring the urgent need for precise and effective diagnostic instruments. Timely detection can significantly reduce the complications associated with diabetes and decrease the level of healthcare expenditure. Often, traditional statistical models do not efficiently manage the complex and non-linear nature of medical data sets. This has driven interest in soft computing approaches, which provide adaptable and flexible models that can manage uncertainty and learn from data, essential qualities for complex, real-world medical diagnosis. Various soft computing techniques such as Feedforward Neural Network(F-ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and hybrid ANN-ANFIS models have emerged as an effective methodology for classification and prediction of medical data. The main motivation of this research is to employ soft computing models on the Pima Indians Diabetes data set to compare the use of various models for predicting diabetes.

1.2 Problem Statement

Diabetes is the world's most widespread chronic metabolic disorder, affecting millions globally, with over 537 million adults (10.5% of the 20-79 age group). Although experienced clinicians can often recognize its symptoms, traditional diagnostic methods heavily rely on subjective interpretation and clinical intuition. This approach may lead to inconsistencies in diagnosis, missed early signs, and inaccurate results, particularly in resource-limited or high-pressure environments. As diabetes continues to rise globally, the development of intelligent diagnostic tools becomes increasingly essential to assist healthcare professionals in making accurate, interpretable, and timely clinical decision-making.

1.3 Research Question

"How do Feedforward Neural Network (ANN), ANFIS, and Hybrid ANN-ANFIS models compare in terms of different evaluation metrics (Accuracy, Explainability) for diabetes prediction using the Pima Indians Diabetes dataset?"

1.4 Objectives

- To develop, implement, and compare three soft computing algorithms, Feed forward Artificial Neural Network (ANN), ANFIS, and Hybrid ANN-ANFIS models for diabetes prediction.

- To compare the models' performance based on accuracy, false positive rate, and interpretability.
- To analyse the strengths and limitations of each approach for practical medical diagnosis.
- Promote early detection and risk prediction to support preventive care and reduce the burden on healthcare systems.

2 Literature Review

The rising global incidence of diabetes mellitus has made it a considerable public health problem that must evolve innovative and reliable diagnostic tools. Advances in the area of computational intelligence have provided a significant amount of research on predictive models to identify individuals who are at high risk of developing diabetes, and even in some cases, identifying individuals who are developing diabetes earlier than standard protocols. Most of the work to date has used well-documented clinical datasets, with the most popular being that of the Pima Indian Diabetes Dataset (PIDD) at UCI Machine Learning Repository, which serves to benchmark and report accuracy of predictive models. The research is diverse in that it ranges from basic machine learning algorithms to hybrid and deep learning methods. This review provides a summary of the research from a select group of studies based on which I will categorize them into four distinct classes/categories, which are research on a) Artificial Neural Networks (ANN); b) Adaptive Neuro-fuzzy Inference System (ANFIS); c) Hybrid Computer Models; and d) comparison studies using classifiers.

2.1 ANN-Based Approaches

Artificial Neural Networks (ANNs) are a type of computational model inspired by the shape and function of biological neural networks, and have played an important role in medical diagnosis studies as they can learn complex non-linear relationships in clinical data. Their use in diabetes prediction has been applied in a variety of situations from simple architectures to deep learning networks with multiple layers.

- Baseline Performance of Standard ANNs: Foundational ANN models, such as the Multilayer Perceptron (MLP), have established a consistent and reliable baseline for performance in diabetes classification tasks.

- A wide-ranging comparison of eight machine-learning algorithms identified a

Neural Network as the best model, prediction accuracy of 78.57% using the Pima Indians Diabetes dataset [1]. This result highlighted the practical advantage neural networks have over traditional models for dealing with medical data sets when comparing to the other models in the study [1].

- In another study, a standalone Multilayer Perceptron Neural Network (MLP NN) achieved a similar accuracy (78.26%) before any feature selection was performed, affirming this level of performance for standard architectures.[2] Today with Deep Learning (DL), or neural networks with greater than one hidden layer (i.e., depth) allows researchers to gain additional predictive accuracy through more abstract and complex features learned from the data directly [2].

- An important study recently compared a typical ANN against a DL model and found that while the ANN performed well with an accuracy of 90.34%, the DL model, which had an input layer, two hidden layers, and an output layer, performed to an accuracy of 98.07% [3]. The relatively large difference in accuracy shows how much more predictive power comes from the depth of the network [3].

- Specialized Architectures Performance: Researchers have adjusted the architectures of specialized ANNs, such as CNNs (Convolutional Neural Networks) - it is often used for image analysis - to work with structured clinical data for diabetes prediction.

- One study evaluated the performance of ANNs and CNNs on a massive 6,500-patient record dataset. In this work, the stand-alone ANN achieved an accuracy of 89.2%, while a CNN did better, achieving an accuracy of 91.4% [4]. The CNN's success implies that its architectural advantages in feature extraction could be transferred to tabular clinical data [4].

- Artificial Neural Networks (ANN) have numerous applications in medical diagnostics such as Stroke, Osteoporosis, Hepatitis, and Cancer detection [13]. Reference [13] describes a multilayer feedforward ANN that used supervised learning on data obtained from 250 diabetes patients. The authors input medical variables such as fasting glucose, creatine, and blood pressure to predict diabetes, and exhibited the advantages of ANN by their ability to effectively model intricate nonlinear relationships in the biomedicine context.

- Due to the ability to model non-linear biomedical relationships, Artificial

Neural Networks (ANN) have been successfully applied to diabetes diagnosis [14]. In [14] Iman, et al., a binary classification ANN model was proposed with 2- and 3-hidden layer architectures, using attributes of age, BMI, family history, and number of pregnancies. The classification accuracy of the 3-hidden layer model was 91.43% and the 2-hidden layer model followed a close second at 87.91%. This demonstrates the capability of ANN for effective and accurate clinical decision support.

- In [15], a multilayer feedforward artificial neural network trained via a backpropagation scheme was used to develop the onset of diabetes diagnosis employing the Pima Indian dataset. Eight clinical attributes from the Pima Indian dataset, including plasma glucose, BMI, age, and so on, were normalized and used as inputs into the model. The model employed one hidden layer with six neurons and a sigmoid activation function, with an 82% recognition rate, outperforming multiple traditional machine learning algorithms.

2.2 ANFIS-Based Approaches

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid model that exhibits the benefits of neural network's ability to learn in an adaptive fashion with fuzzy logic's ability to reason in a transparent rules-based way. ANFIS is well suited for medical diagnosis applications because it can learn its parameters from data while retaining some level of interpretability. The results of the ANFIS literature used for diabetes prediction have reported consistent high performance with significant attention to data quality, using different methods for rule generation, and using multiple instances of data training.

- Effect of Dataset Quality and Coverage: The real-world effectiveness of predictive modeling is influenced by the data on which it was trained.

- This principle was illustrated near perfection at 97% accuracy with an ANFIS model trained on a uniquely augmented dataset [5]. These researchers fused public datasets (PIDD and Bio-Statistics) with data sourced from Nigerian hospital data that they had collected locally to illustrate that models trained on data that was representative of the population of interest perform much better than other data sources [5].

- The Non-Negotiable Importance of Data Preprocessing: The credibility of ANFIS models is ultimately dictated by the quality of the data. Preprocessing is simply a necessary part of developing a trustworthy and valuable diagnostic model by addressing noisy data, missing data, and outliers.

- For example, one study implemented a multi-staged approach with a strong data curation emphasis by normalizing the data, imputing missing values, and identifying and deleting 68 anomalous records from the PIDD with the Local Outlier Factor (LOF) technique . The ANFIS model trained with the curated data achieved classification accuracy of 92.77% [6].

- Improvements in Rule Generation and Model Tuning: The method for generating and updating the fuzzy if-then rules a major factor in Anfis performance.

- One Anfis model that used the Subtractive Clustering Method to perform the automated generation of rules was reported as extremely sensitive to the clustering radius hyperparameter [7]. By refining this parameter, optimal model performance was achieved, with a peak 99.13% accuracy given with a radius of 0.2, representing the highest performance of any of the papers considered [7].

- Another study developed a Modified Anfis (M-Anfis) that included an original feature extraction step based on calculating the entropy of the frequent closed item sets [8]; this modification allowed the M-Anfis goal to focus on the most informative to clinical variables, which produced better model accuracy at 97.5% [8].

2.3 Hybrid ANN–ANFIS Models and Other Hybrids

Hybrid intelligent systems are intended to use the comparative advantages of multiples processes working together, typically combining approaches to feature selection, parameter tuning and classification altogether to develop stronger and more accurate diagnostic systems.

- ANFIS Optimization Hybrids: An advanced method of increasing the predictive quality of an ANFIS model is to use metaheuristic algorithms and other algorithms in combination to optimize and train the model.

- A hybrid ANFIS-PSO-WOA model was proposed that used both Particle Swarm Optimization (PSO) and Whale Optimization Algorithm (WOA) to optimize ANFIS parameters, with a reported 92% accuracy in a fog-computing framework [9].

- A new model, named LANFIS, was developed to meet expected datasets with missing data and binary output requirements for diagnosis [10]. The LANFIS model integrated Logistic Regression with ANFIS and had multiple imputes for missing data. The performance was 88.05% for accuracy and was reported to be 3-5% better than other methodologies that discard records with missing data [10].

- In terms of the training process though, one other study used a hybrid learning approach using a Modified Levenberg-Marquardt (MLM) algorithm to train an ANFIS classifier with an overall accuracy of 82.30% [11].

- Preparation and Feature Selection Hybrids: This class of hybrids uses one algorithm for data preparation or feature selection and contains a second algorithm for classification. When appropriate preparation or feature selection is used with the classification, it can result in a reduction in the costs of diagnosing a data sample and in being more accurate by eliminating irrelevant data.

- A two-stage system used a Genetic Algorithm (GA) for feature selection which successfully reduced the attributes of the PIDD from a set of eight attributes to four important features to be used in the ultimate model [2]. Training an MLP Neural Network on the same attributes with the same learning time, MLP The MLP model achieved 79.13% accuracy using the four attributes, whereas, previously, the model used a training set of eight attributes [2].

- A multi-stage pipeline was developed that used Principal Component Analysis (PCA) for dimensionality reduction first, Fuzzy C-Means Clustering next for cleaning the dataset, and then Random Forest for classification. This synergistic analysis achieved exemplary accuracy of 98.96% [12].

- Deep Learning Architecture Hybrids: It is common to hybridize different neural network architectures to maximize the benefits of each type.

- Hybrid ANN-CNN neural network (with a research grade dataset) was constructed [4]. This hybrid architecture has utilized the automated feature extraction of the CNN layers, and provided the learned features of the CNN, into dense layers of ANN, which was then used for classification. This hybrid structure achieved an accuracy of 94.3%, better than both ANN alone and

CNN alone [4].

- Although [14] concentrates on ANN and does not provide ANFIS comparisons to other prediction algorithms, prior works have suggested ANFIS can serve as a suitable decision-making tool for medical prediction problems, as it utilizes neural network learning to develop rules based on fuzzy logic [14]. ANFIS can manage uncertain and imprecise data commonly found in the health care context, making it an appropriate modelling structure for diabetes diagnosis. There is no mention of ANFIS or comparisons in [14], which provides a future research opportunity.

2.4 Comparative Studies and Baseline Models

It is important to perform comparative studies to better understand the performance of algorithms and which models are useful for a purpose. These studies involve comparing different techniques directly on the same dataset allowing for inferences about the performance of predictive power.

- Comparative Benchmarking of Machine Learning Models: Several studies examined a large number of commonly used classifiers to identify the best classifiers for diabetes prediction.

- For example, one study compared eight different machine learning models and stated that "the Neural Network (78.57%) and Random Forest (76.30%) were the best classifiers based on the Pima Indians data set [1].

- Another extensive comparative review evaluated four classifiers, and identified a clear ranking of performance. A Deep Learning (DL) model had the greatest accuracy of 98.07%, with a Decision Tree coming second (96.62%), followed by an ANN classifier (90.34%) and lastly, Naive Bayes (76.33%) [3].

- Validation of Novel Models Compared to Baselines: It is common for researchers to apply some form of comparative analysis to demonstrate that their newly designed hybrid or modified model is better than other models.

- The advantage of a deliberate data-preprocessing pipeline was evident when comparing a pre-processed diverse ANFIS model with 92.77% accuracy against baseline models trained on the same dataset without pre-processing. In this example, a Neural Network scored only 75% and Random Forest produced a score of 73% [6].

- The impressive strength of a Modified ANFIS (M-ANFIS) model was established when comparing it to a package of other techniques. The M-ANFIS score of 97.5% accuracy was proved superior to an Optimal Decision Tree score of 92.8%, an ANFIS score of 86.48% and K-Nearest Neighbors scoring 77.1% [8].

- In [13], three algorithms used for Artificial Neural Networks (ANN) BFGS Quasi-Newton, Bayesian Regularization, and Levenberg Marquardt were compared for diabetes prediction with data from 250 patients aged 25 to 78 years. Performance was measured using correlation coefficient (R), mean squared error (MSE), and prediction accuracy. The Bayesian Regularization model used information from the target covariates to create a best fit to the predicted targets (R = 0.99579, accuracy = 88.8%), suggesting that it had high correlation to the target with good generalization. The BFGS Quasi-Newton and Levenberg Marquardt details were lower at R = 0.86714, R = 0.6051, respectively. The results provide further evidence that Bayesian Regularization is a solid baseline model for reasonable predictions of diabetes prediction.

- In [14], the ANN was compared against Decision Tree, Random Forest, SVM, KNN, Logistic Regression, and Naïve Bayes to evaluate its performance. The 3-hidden layer ANN reached an accuracy of 91.43%, equal to Random Forest and better than every other model, while the 2-layer ANN reached an accuracy of 87.91%. From the results, it is clear the design and construction of hidden layers is crucial to achieving higher accuracy. The research demonstrates that an ANN can be effectively utilized for the recognition of diabetes mellitus patients by examining a set of eight inputs.

- In [15], a backpropagation neural network (BPNN) achieved 82% recognition accuracy for onset diabetes diagnosis using the Pima Indian dataset, surpassing ADAP (76%), C4.5 (71.1%), KNN (72%), BSS (67.1%), and EM (<70%). This demonstrates BPNN’s superior predictive capability over several traditional algorithms, making it a more reliable tool for diabetes classification.

3 Methodology

3.1 Dataset Collection:

The Pima Indians Diabetes Dataset from the UCI Machine Learning Repository is used for this study. It consists of **768 medical records**, each with 8 diagnostic attributes and one binary outcome variable indicating the presence or absence of diabetes.

3.1.1 Features include:

- Number of pregnancies
- Plasma glucose concentration
- Diastolic blood pressure
- Triceps skinfold thickness
- 2-hour serum insulin
- Body mass index (BMI)
- Diabetes pedigree function
- Age

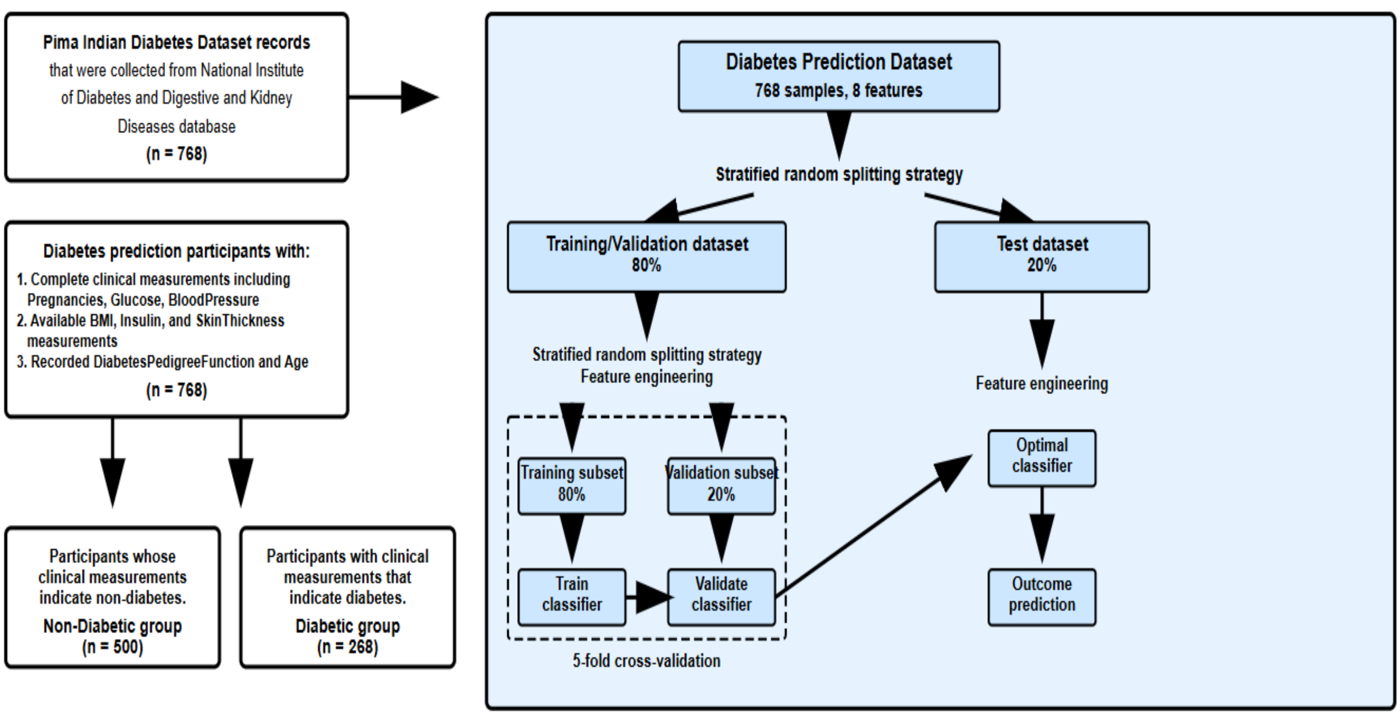
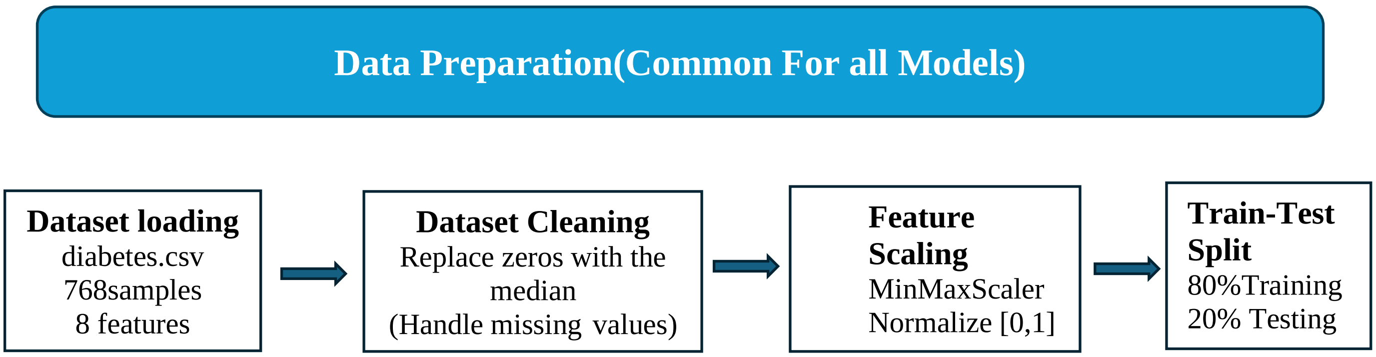


Figure 1: Study design, label definitions, and the architecture of our prediction models.

3.2 Dataset Preprocessing:

These steps ensure the data is clean, scaled, and suitable for model training.



3.2.1 Handling Missing or Zero Values:

Zero values in columns like *Glucose*, *Blood Pressure*, *Skin Thickness*, *Insulin*, and *BMI* are replaced with median values to prevent misleading inputs during training.

3.2.2 Feature Scaling:

All input features are normalized to a range between 0 and 1 using **MinMaxScaler** to ensure that the neural network and fuzzy layers converge effectively.

3.2.3 Data Scaling:

The dataset is split into **80% training** and **20% testing** sets using a random seed for reproducibility.

3.3 Building the Model and Training

The overall modelling process is divided into three key stages:

1. Defining Model Architecture
2. Training the model
3. Estimation of Performance

The models are implemented in **Python** on Google Collab, which provides modules for both neural and fuzzy inference system layers.

3.3.1 Feedforward Artificial Neural Network (F-ANN):

- **Input Layer:** 8 neurons (feature count) - Hidden
- **Hidden Layers:**
    - Layer 1: 64 neurons, Batch Normalization, ReLU activation, Dropout

(0.3).
    - Layer 2: 32 neurons, Batch Normalization, ReLU activation, Dropout (0.2).
    - Layer 3: 16 neurons, Batch Normalization, ReLU activation, Dropout (0.1).
- **Output layer:** 1 neuron with Sigmoid activation for binary classification.
- **Training Configuration:**
    - Optimizer: Adam
    - Loss Function: Binary Crossentropy
    - Batch Size: 32
    - Epochs: 100 (with early stopping)
    - Validation Split: 20%

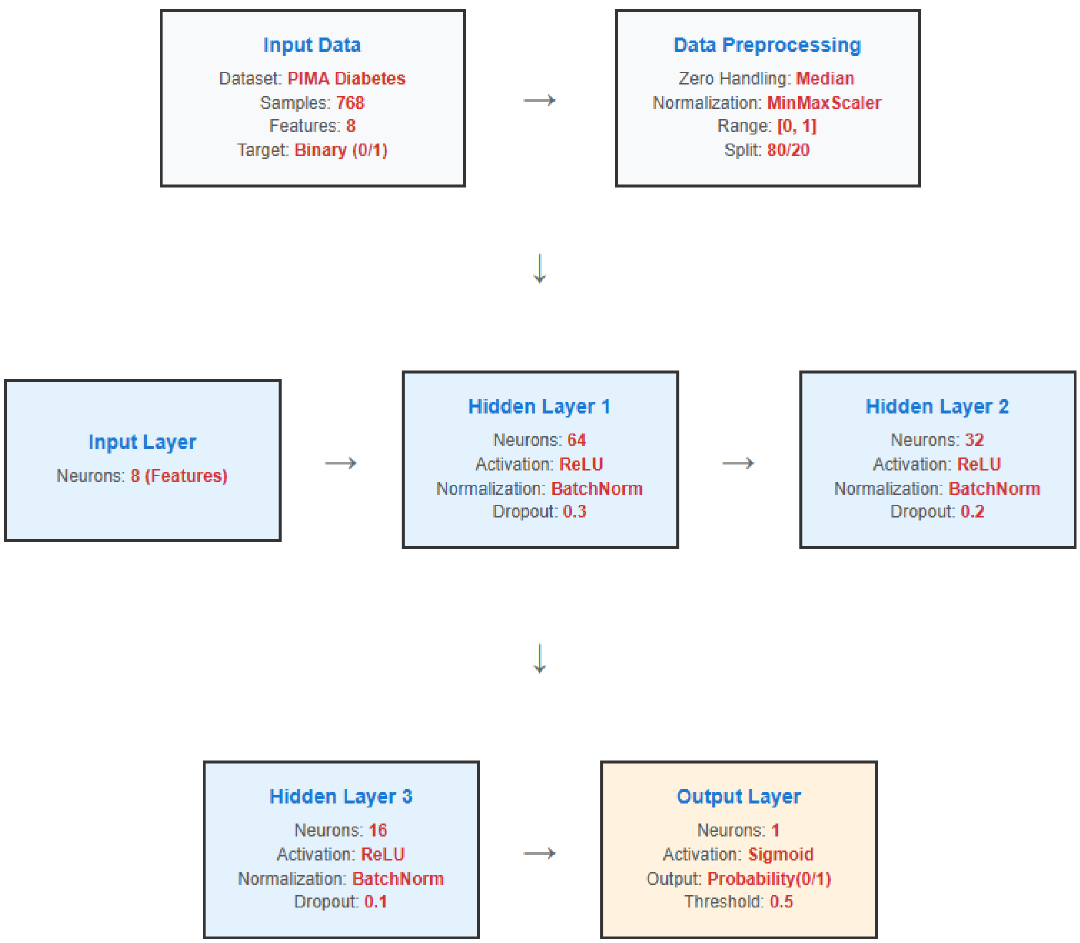


Figure 2: Feedforward Artificial Neural Network (F-ANN) Block Diagram

3.3.2 Adaptive network-based fuzzy inference system (ANFIS):

Architecture:

- Input Variables: 8 (corresponding to dataset features)
- Membership Functions: 5 triangular functions per input
- Rules: 5 fuzzy rules - Learning: Hybrid
- learning algorithm (gradient descent + least squares)
- Defuzzification: Weighted average method

Implementation Details:

- Membership function initialization: Data-driven
- Rule consequence: Linear functions
- Training: Least squares optimization

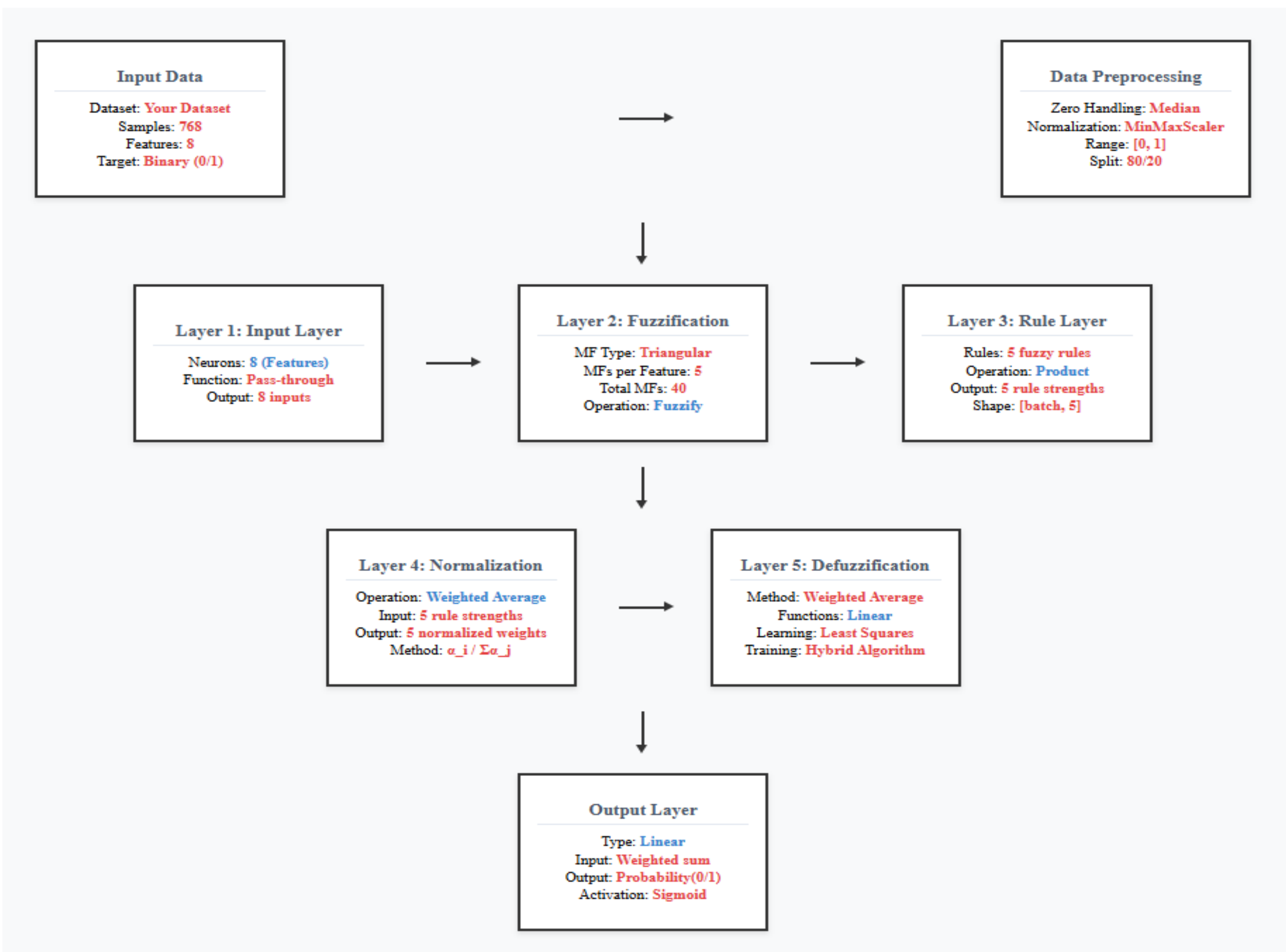


Figure 3: Adaptive Neuro-Fuzzy Inference System (ANFIS) Block Diagram

3.3.3 Hybrid ANN-ANFIS:

Ensemble Strategy:

- Component 1: ANN model (as described above)
- Component 2: ANFIS model (as described above)
- Combination: Weighted average (equal weights: 0.5 each)
- Final Output: Ensemble probability prediction

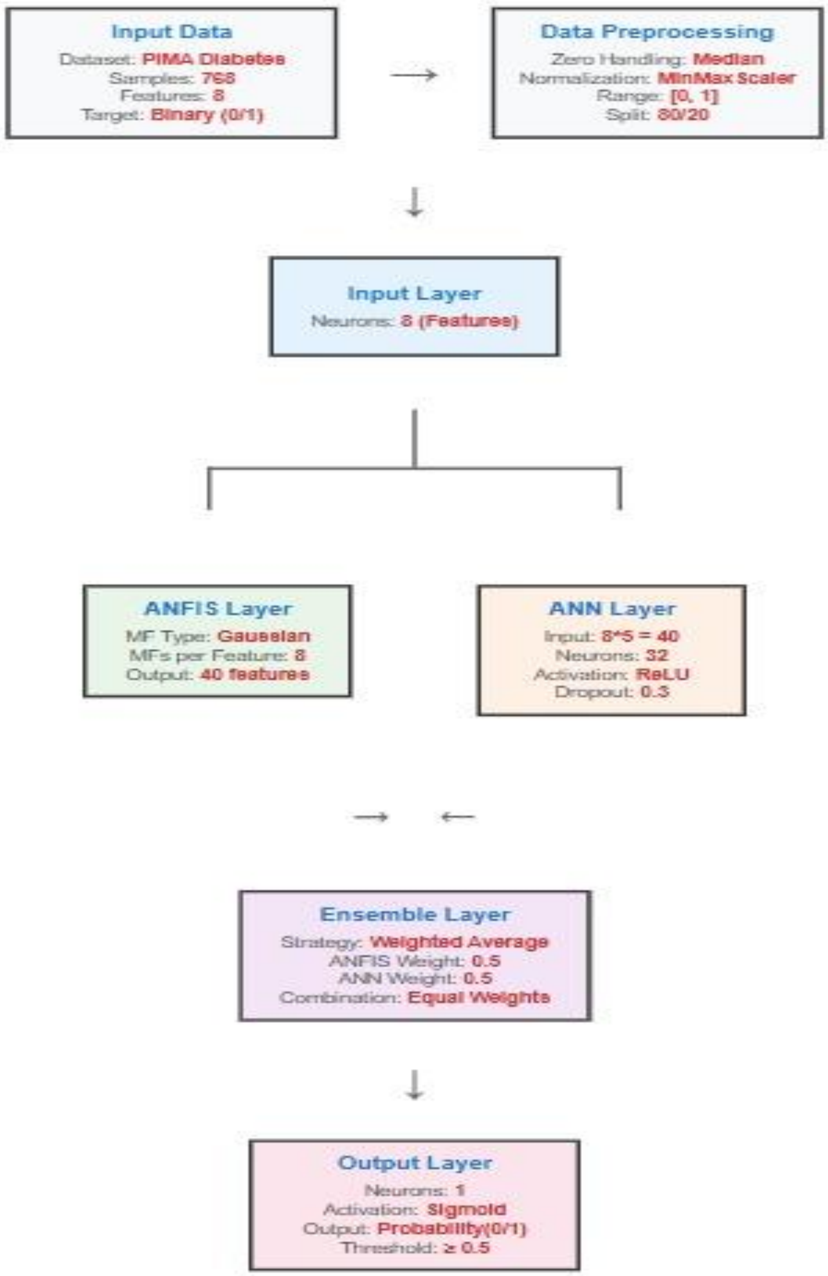


Figure 4: Hybrid (ANFIS +ANN) Model Block Diagram

3.3.4 Model Testing:

After training the ANN, ANFIS, and Hybrid ANN-ANFIS models, testing was conducted using a held-out test set comprising 20% of the original data. This set was not seen during training, ensuring an unbiased evaluation of each model's generalization capability.

4 Result

4.1 Performance evaluation of the prediction models

We applied three soft computing algorithms to predict diabetes in patients: feedforward ANN, ANFIS (Adaptive Neuro-Fuzzy Inference System), and Hybrid ANN-ANFIS. We evaluated the performance of each model using accuracy, precision, recall, F1 score, and area under the curve (AUC). In this study, the AUC refers to the area under the receiver operating characteristic (ROC) curve. We also used the ROC curve and calibration curve analysis to determine the discrimination, reliability, and applicability of the predictive soft computing models. All analyses, including data preprocessing, feature selection, ML modeling, and performance evaluation, were conducted using Python (version 3.12.4).

Metric	F-ANN	ANFIS	Hybrid ANN-ANFIS
Accuracy	≈ 76%	≈ 81%	≈ 77%
Precision	≈ 74%	≈ 79%	≈ 75%
Recall	≈ 72%	≈ 80%	≈ 76%
F1-score	≈ 73%	≈ 79%	≈ 75%

Table 1: Evaluation Matrix

Figure 4 shows the ROC curves for evaluating the discriminative capability of all three models. According to the ROC plot, both the ANFIS and Hybrid ANN-ANFIS models performed best with AUC 0.824, while the ANN model performed second best with AUC 0.819. All three models considerably outperformed random classification (diagonal dashed line), demonstrating clinically significant discriminative capability. The curves show that the Hybrid model consistently maintains high true positive rates across various false positive rate thresholds and has good performance at varying classification thresholds.

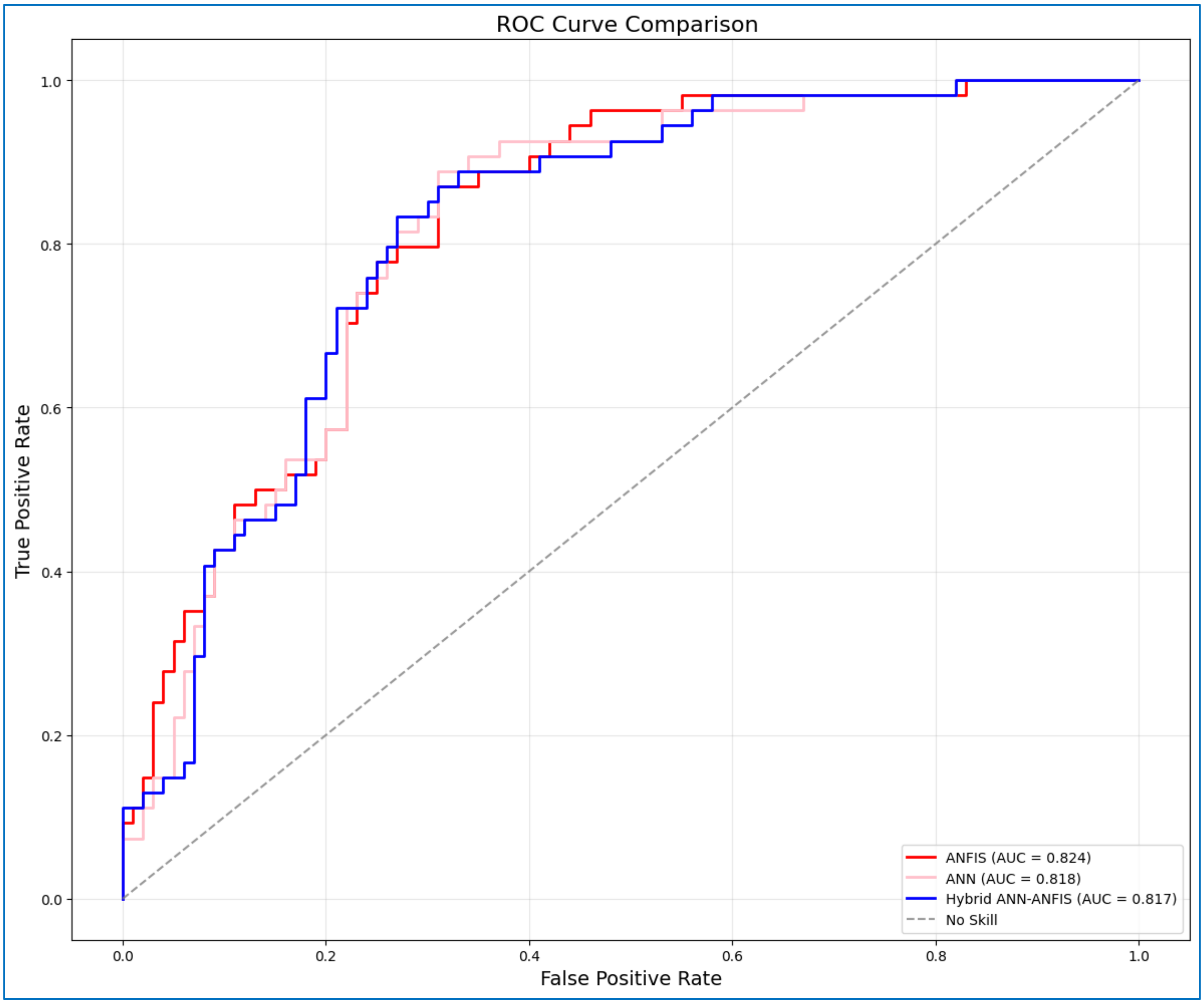


Figure 5: ROC curve comparison for different models. The y-axis represents the true-positive rate, whereas the x-axis represents the false-positive rate. The dashed line represents the no-skilled classifier.

The calibration analysis also verified the prediction models' validity. The Hybrid ANN-ANFIS model performed better in calibration, whose predicted probabilities closely match true outcomes across the entire probability range. This implies that when the model predicts a 70% probability of having diabetes, approximately 70% of such patients have diabetes, making it highly reliable for clinical decision-making. The ANFIS model also demonstrated systematic bias when estimating probability, overestimating risk for diabetes during the entire process, which could lead to excessive worry and subsequent tests in real life.

Figure 5 represents the prediction probability distributions for all models within the non-diabetic and diabetic classes. The histograms reflect specific patterns in how each algorithm responds to the uncertainty of classification. In the ANN model, non-diabetic patients have a dense cluster of low predicted probabilities (range of 0.0-0.2), while diabetic patients show a more spread-out pattern across probability ranges. The ANFIS model has more spread-out distributions for both the classes, and diabetes patients have predictions more clustered in a tight range of 0.5-0.7, indicating more uncertainty in classification. The Hybrid ANN-ANFIS model has more confident and sharper predictions, and a clearer distinction between the two classes, indicating more discriminative ability.

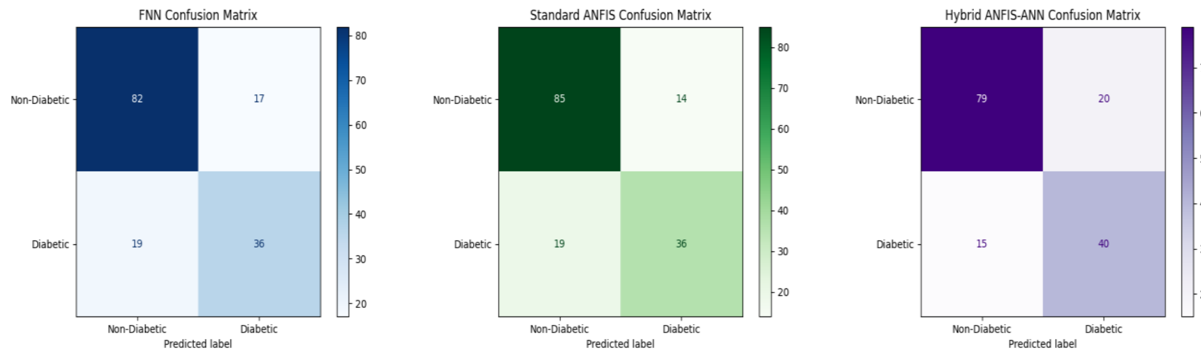


Figure 6: Confusion matrices for different models

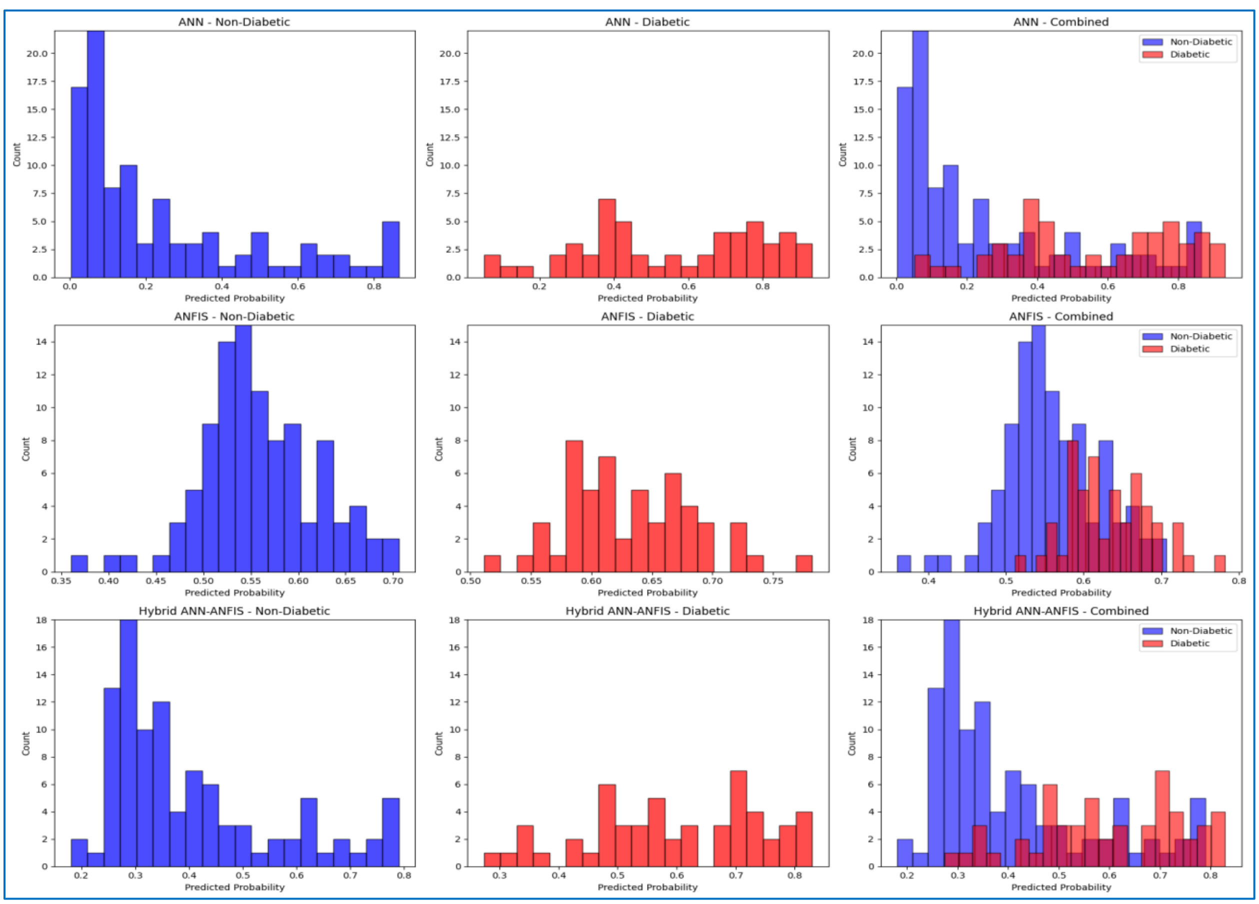


Figure 7: Probability histograms for the different models. The left panels display histograms of the predicted probabilities for the "not diabetes" and the Middle one with "diabetics" classes. The rightmost panels combine the predicted probabilities of both

classes and show the overall distribution. The x-axis represents the predicted probability, and the y-axis represents the sample count, reflecting the confidence of the model in its predictions.

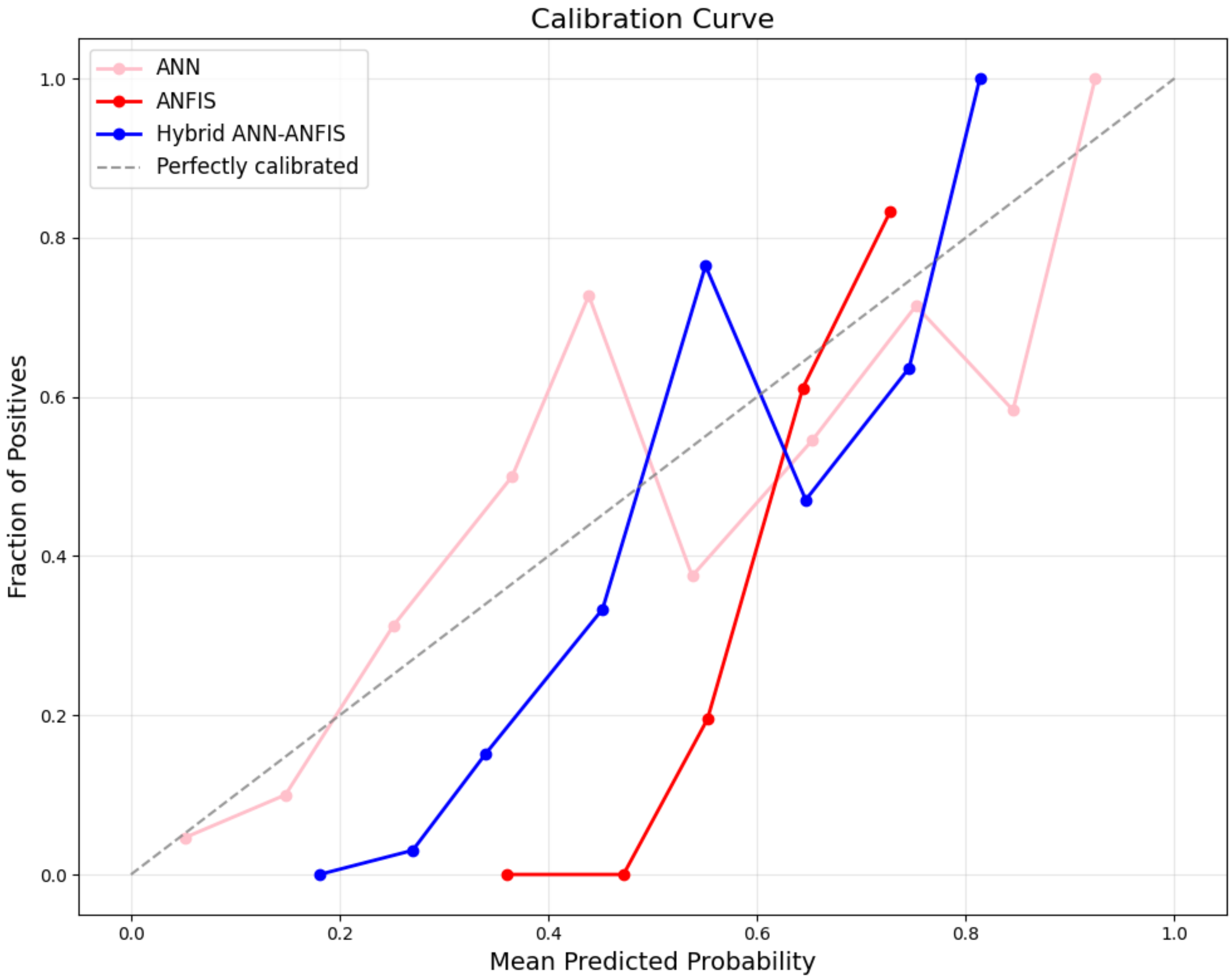


Figure 8: OC curve comparison for different models. The y-axis represents the true-positive rate, whereas the x-axis represents the false-positive rate. The dashed line represents the no-skilled classifier.

5 Discussion

5.1 Model Comparison Insights

Aspect	FNN	ANFIS	Hybrid ANN-ANFIS
Input Layer	8 features	8 features	8 features
Fuzzy Layer	None	5 Gaussian MFs per feature	6 Gaussian MFs per feature
Hidden Layers	64 neurons + 32 neurons	Linear rules	ANN hidden layer: 32 neurons

Output Layer	1 neuron (Sigmoid)	1 neuron (Sigmoid)	1 neuron (Sigmoid)
Membership Functions	None	Gaussian	Gaussian
Training Loss	BCELoss	BCELoss	BCELoss
Optimizer	Adam	Adam	Adam
Typical Accuracy (from code)	≈ 76%	≈ 81%	≈ 77%

Table 2: Architecture Comparison

5.1.1 Hybrid Model Advantages & Disadvantages

Advantages:

- **Enhanced Feature Representation:** The primary advantage is that the FNN has an expanded input space because of the fuzzy layer – instead of just seeing a straight number, like Glucose = 140 for example; the network sees, or represents, the fuzzy degrees of "slightly high, moderately high, very high," etc., which makes it easier for the FNN to learn with respect to the complexity of the pattern.

- **Integration of Strengths:** This structure marries fuzzy logic's ability to handle uncertainty with the minimal processing and good generalization ability of a neural network.

- **The possibility of improved performance:** This hybrid form provides the potential to improve performance, since it can effectively increase the amount of information available to a model by reshaping the input space in a meaningfully data-driven manner that takes into account uncertainty. This capability allows the model to predict beyond the ability of either given crude model alone while particularly being beneficial for complex datasets when the decision boundaries are not crisp.

Disadvantages:

- **Increased Complexity:** This is the most challenging of the three models to design, implement, and debug.

- **Reduced Interpretability:** The first fuzzy layer has some transparency. However, the ultimate decision is made by the "black box" FNN component. This means that this model loses much of the interpretability benefits of a pure ANFIS.

- **Many More Hyperparameters to Tune:** This model requires tuning parameters for the fuzzy layer (e.g., n_mfs) and FNN (e.g., hidden_size, dropout rate, learning rate), which makes the optimization process even more difficult and time-consuming.

5.1.2 Individual Model Advantages & Disadvantages

ANN Advantages:

- **Strong Predictive Capability**: As a universal approximator, a well-trained FNN can learn and model very high-dimensional complex non-linear relationships between the input features and the outcome.

- **Strong Generalization:** The use of Batch Normalization and Dropout provides a measure of robustness to the model. FNNs are less likely to over-fit training data noise and will often yield stable performances on new unseen patient data.

- **Strong Scalability:** In addition to being very interpretable, FNNs are uniquely scalable and can model datasets with large number of features and number of instances; very effective with complex medical datasets.

ANN Disadvantages:

- **"Black Box" Nature:** The primary drawback is interpretability. It is very difficult to explain why the network made a particular prediction. The decision exchangeability is represented across many different numerical weights, making it not easily relatable to clinicians who want to trust the reasoning behind a decision.

- **Data hungry:** FNNs typically require a lot of data to train effectively. With smaller datasets, they overfit, even if using regularisation strategies.

- **Computationally expensive:** training deep neural networks can be slow, and potentially requires specialty hardware like GPU's, especially as the number of layers and neurons increases.

ANFIS Advantages:

- **Enhanced Interpretability:** An ANFIS is also more interpretable than an FNN. A clinician can look at the learned membership functions and see how the model has classified input values (e.g., "the model has learned that around 50 years old is strongly related to the high risk fuzzy set").

- **Handles Uncertainty Well:** Fuzzy logic is designed to take on the ambiguity and vagueness inherent in medical data, wherein there are rarely hard cut-offs for 'normal' versus 'abnormal'.

- **More Efficient on Smaller Datasets:** ANFIS can usually achieve better performance with less data than a deep FNN because of its simpler model and rule-based nature.

ANFIS Disadvantages:

- **Curse of Dimensionality**: The number of possible fuzzy rules increases exponentially with the number of input features. In our case, the model uses 8 features, but traditional ANFIS is now impossible with dozens or even hundreds of input features.

- **Limited Complexity:** The simple linear output layer limits it's ability to capture the same complexities that could be captured by a deep FNN with n non-linear hidden layers.

- **Performance is Dependent on MF Choice:** The model's performance is tied to the subsequent membership function you selected (Gaussian). The distribution of the data may be described more appropriately with different membership function types (triangular, trapezoidal, etc.).

5.1.3 Trade-off Analysis

The results reveal important trade-offs:

- **Meaningful Reasoning vs. "Black-Box" Power:** The ANFIS model provides the highest level of meaningful reasoning due to its fuzzy-rule-based foundation. The FNN model provides the highest functionality as a "black-box" model, having high predictive power, but lacking meaningful reasoning and insights into its rationale. The Hybrid model provides an interesting compromise by employing an interpretable fuzzy front end to engineer features for a powerful neural network back end, synergistically combining paradigms of transparency with high classification power.

- **Automated Feature Engineering vs. Direct Classification:** The FNN and ANFIS models carried out direct classification on the input data. The Hybrid model provides a more sophisticated funnel; its ANFIS layer performs automated feature engineering, transforming the raw numerical inputs into a richer fuzzified feature space prior to classification. This technique has important implications for revealing more complex patterns contained in the data.

- **Simplicity vs. Architectural Power:** plain ANFIS is a solution that is extremely simple but effective. The Hybrid model presents more architectural complexity, but also gives a more powerful and flexible framework. This framework provides greater ability to achieve performance gains through more hyperparameter tuning and hyperparameter optimization techniques, and it is better positioned as a future-state architecture.

- **Clinically Valid (Recall vs. Precision):** when it comes to medical diagnosis, we must be most concerned with the ability to recognize true cases (recall). The ANFIS model had the highest recall (≈80%) and therefore is the most clinically valid supplied non-miss in this implementation. The Hybrid model had a solid and balanced performance (recall ≈76%), showing the potential which will be a reliable and robust clinical tool that appropriately trades off true cases and non-cases without being overly alarmist.

5.1.4 Deployment Scenarios

1. **Primary Care Screening**: Hybrid model for balanced assessment
2. **Emergency Department**: ANFIS for high sensitivity screening
3. **Specialist Consultation**: ANN for high-precision diagnosis
4. **Population Health**: All models for comprehensive risk assessment

5.1.5 Integration Challenges

- **Data Quality**: Requires high-quality, complete patient data
- **Standardization**: Need for consistent measurement protocols

- **Training**: Healthcare providers need training on AI-assisted diagnosis
- **Validation**: Continuous monitoring and validation in clinical settings

6 Limitations and Future Work

6.1 Study Limitations

- **Dataset Size and Specificity:** This study is based solely on the Pima Indians Diabetes Dataset (PIDD). While it is indeed a benchmark dataset, it is also demographically specific (Pima Indian Females), and has a fairly small dataset (768 instances). As a result, there is a possibility that the model performance may not generalize very well to other potentially more diverse populations. Given the small dataset size, there is also a risk of overfitting, defined as the model learning the training data too well, at the expense of new data of which it has no prior experience.

- **Hyperparameter Optimization:** The hyperparameters for all three models (i.e. learning rate, number of neurons, dropout rate, number of membership functions) were held constant. The performance of these models is highly dependent on these settings. Without a formal hyperparameter tuning process, such as Grid Search or Bayesian Optimization, it is unlikely that these models are achieving optimal performance.

- **Lack of Interpretability:** Both the FNN and the Hybrid ANN-ANFIS models are "black boxes" and do not allow for an easy understanding of the clinical reasoning that underpins predictions. The lack of transparency is a critical hurdle to acceptance in a medical setting where trust and an ability to explain the diagnostic output is essential.

- **Static Evaluation:** The models were assessed on a single, fixed 80/20 train-test split, meaning the performance metrics are based on this one data partition, which may not reflect the models' true average performance. More robust evaluation procedures (k-fold cross-validation, for instance) should be selected if a proper assessment is to be made.

6.2 Future Research Directions

- **Add Diverse and Larger Datasets:** In future work, obtaining larger and more diverse datasets for training and validating the models to produce a more robust and generalizable diagnostic tool. When gathering datasets, it would be beneficial to include data from different ethnicities, genders, and geographic areas in order to ensure the datasets are equitable and ultimately, effective for the population as a whole.

- **Implement Systematic Hyperparameter Tuning:** When conducting follow-up work based on this research, future work could include automated hyperparameter tuning to explore combinatorial hyperparameter tuning systematically. This would search systematically for the optimal combination of settings for each model. I suspect it would lead to a big increase in their predictive accuracy and other measures of performance.

- **Explore Explainable AI (XAI) Methods:** In the future, to reduce the "black box" issue mentioned above, some sort of XAI methods should be included such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) to find out what clinical features (e.g. Glucose, BMI, Age) influenced a given prediction, which may offer transparency to make the models more credible or trustworthy for clinicians.

- **Develop and Validate New Hybrid Architectures:** The hybrid ANN-ANFIS model, while not the most accurate model for this study, is a possible hybrid architecture to follow-up on. Future work could include more complex hybrids, for example, using Convolutional Neural Networks (CNN) to take the features (image) and then pass that to the ANFIS model or using attention mechanisms to allow the model to focus on the most important input features.

- **Develop a Real-World Application:** The focus of this research ultimately comes down to its clinical usefulness. An important step moving forward would be to implement the best performing model (ANFIS in this case) to create a user-friendly application. It could be a web-based application or mobile application that allows a clinician to input patient data and receive an instant risk assessment, thus creating the potential for academic research to translate into a practical method of early diabetes detection.

7 Challenges

- **Data-Related Challenges**
  - **Data quality and missing data:** Medical datasets for clinical purposes often include impractical or missing values when looking at features of the data we use, such as `Glucose` and `BloodPressure` in the PIMA, which have biologically implausible zero values. I see you fixed this in your code by replacing the zero values with the median, however, this imputation is a simple imputation that does not perfect reflect the true distribution of the data and could have an affect on bias.

  - **Class imbalance:** The dataset is a class imbalance since there are more non-diabetic patients than diabetic patients. Class imbalance makes it difficult for the model to remain honest as it can learn to be biased towards the majority class, leaving it hard to determine the less frequent class which is clinically very important diabetic cases (low recall).

  - **Representativeness of the dataset:** We trained our model on the PIMA and the PIMA is population and demographically specific to Indian female subjects. This presents a potential to misgeneralise as the models we trained may not generalise well across the different populations with different genetics and different lifestyle. This limits a universal PM Expo.

- **Model and Implementation Challenges**
  - **The "Black Box" Challenge:** The FNN and Hybrid ANN-ANFIS models are "black boxes" which makes comprehending the clinical logic behind their predictions hard; this represents a significant obstacle towards gaining the trust and acceptance of physicians and medical professionals.

  - **Hyperparameter Sensitivity:** Like any machine learning model, performance associated with models will depend heavily on hyperparameter selection (learning rate, numbers of neurons, numbers of membership functions etc). Testing different hyperparameter combinations is the hardest part and is costly to do.

  - **Complexity of Hybridization:** There is a complexity associated with Hybrid ANN-ANFIS Models through the combination of fuzzy systems and neural systems. The difficulty with Hybrid ANN-ANFIS Models is ensuring that any added complexity by way of a complex architecture, converts into a performance benefit because as we saw from your study, simpler ANFIS performed far better than the complex Hybrid ANN-ANFIS Model.

8 Conclusion

This research undertook the creation and evaluation of three different computational intelligence models—a Feedforward Neural Network (FNN), a conventional Adaptive Neuro-Fuzzy Inference System (ANFIS), and a new Hybrid ANN-ANFIS system—for the task of early diabetes prediction. This research provided useful and rich insights into the potential of more complex hybrid systems to develop more sophisticated and capable diagnostic systems.

Although the conventional ANFIS produced the highest accuracy in this implementation (≈81%), the Hybrid ANN-ANFIS is the most innovative and promising architecture for a clinical application. Its potential does not just lay in its decent performance (≈77% accuracy and ≈76% recall), but also in its unique design that combines the complementary aspects of both fuzzy logic and neural networks.

Conceptually, the hybrid model's architecture where an ANFIS layer is used for feature engineering, combined with the ANFIS "fuzzified" data representation utilized with a capable neural network classifier, is a better approach. This approach aims to accommodate the inherent ambiguity and uncertainty that medical data contains, which is a significant opportunity when contemplating future developmental applications on more complicated and larger datasets where simpler models may aid in diagnosing diseases.

The hybrid model's results in this project should be recognized as a significant baseline, as it is a proposed model that has several opportunities for enhancement. With systematic tuning of hyperparameters, it may be possible for this model to perform better than a simple system like the vanilla ANFIS model. In contrast, the FNN, although a powerful model, was not as effective as a "black box" in this case since it produced the lowest accuracy and recall.

In conclusion, this research established that the hybrid ANN-ANFIS model, by fusing the transparency of fuzzy logic with the sophisticated classification ability of a neural network, provides the strongest and most flexible foundation for the next generation of intelligent diagnostic systems. The hybrid model has architectural potential for increased performance, and a more complex mechanism of representing data, makes for the most attractive model for future exploration, and clinical applicability.

9 References

[1] *Alzboon MS, Al-Batah M, Alqaraleh M, Abuashour A, Bader AF. A Comparative Study of Machine Learning Techniques for Early Prediction of Diabetes*.

[2] *Choubey DK, Paul S. GA_MLP NN: A Hybrid Intelligent System for Diabetes Disease Diagnosis. I.J. Intelligent Systems and Applications. 2016;1:49-59.*

[3] *Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. J Diabetes Metab Disord. 2020;19:391-403.*

[4] *Abed AH. Leveraging Diabetes Prediction using the Deep Learning-based Hybrid ANN-CNN Architecture. Int J Adv Netw Appl. 2025;16(6):6683-6690.*

[5] *Akpado KA, Njonu UJ, Obioma PC, Isizoh AN. An Improved Method for Predicting Diabetes Mellitus Using Adaptive Neuro-Fuzzy Inference System. Int J Sci Eng Sci. 2021;5(11):19-32.*

[6] *Alasaady MT, Aris TNM, Sharef NM, Hamdan H. A proposed approach for diabetes diagnosis using neuro-fuzzy technique. Bull Electr Eng Inform. 2022;11(6):3590-3597.*

[7] *Paul B, Karn B. ANFIS based Diabetes Mellitus Prediction. In: 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON). IEEE; 2021*.

[8] *Shrinivasan L, Verma R, Nandeesh MD. Early prediction of diabetes diagnosis using hybrid classification techniques. IAES Int J Artif Intell. 2023;12(3):1139-1148.*

[9] *Kumar D, Mandal N, Kumar Y. Fog-based framework for diabetes prediction using hybrid ANFIS model in cloud environment. Pers Ubiquitous Comput. 2022;27:909-916.*

[10] *Ramezani R, Maadi M, Khatami SM. A novel hybrid intelligent system with missing value imputation for diabetes diagnosis. Alexandria Eng J. 2018;57:1883-1891.*

[11] *Sagir AM, Sathasivam S. Design of a Modified Adaptive Neuro Fuzzy Inference*

System Classifier for Medical Diagnosis of Pima Indians Diabetes. AIP Conf Proc. 2017;1870:040048.

[12] *Bhatti P, Mahboob K, Naeem SS, Bhatti IH, Kamran N. Enhanced Diabetic Prediction Using Fuzzy C-Means Preprocessing and Random Forest Ensemble Learning. VFAST Trans Software Eng. 2023;11(4):32-44.*

[13] *S. Kumar and A. Kumaravel, "Diabetes diagnosis using artificial neural network," IJESRT, vol. 2, no. 6, pp. 1642–1644, Jun, 2013.*

[14] *R. F. Jader and S. Aminifar, "Fast and accurate artificial neural network model for diabetes recognition," NeuroQuantology, vol. 20, no. 10, pp. 2187–2195, 2022.*

[15] *E. O. Olaniyi and A. Khashman, "Onset diabetes diagnosis using artificial neural network," International Journal of Scientific & Engineering Research, vol. 5, no. 10, pp. 754–759, Oct. 2014.*