

Article

A Differentiable Keras-Based Adaptive Neuro-Fuzzy Inference System for Diabetes Prediction Using NHANES 1999–2018 Data

Sravanthi Narne ¹, Srikanth Vemuri ¹, Sastry Kodanda Rama Jammalamadaka ¹, Sasi Bhanu Jammalamadaka ²,
Pardhu Thottempudi ³ and Fernando Moreira ^{4,5,6,*}

¹ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, Andhra Pradesh, India; nsravanthi@kluniversity.in (S.N.); vsrikanth@kluniversity.in (S.V.); drsastry@kluniversity.in (S.K.R.J.)

² Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad 500090, Telangana, India; bhanukamesh1@gmail.com

³ Department of Electronics and Communications Engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram 522302, Andhra Pradesh, India; pardhu.t@ieee.org

⁴ Research on Economics, Management and Information Technologies, Portugalense University, 4200-072 Porto, Portugal

⁵ Research on Economics, Management and Information Technologies, University of Aveiro, 3810-193 Aveiro, Portugal

⁶ Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, 3810-193 Aveiro, Portugal

* Correspondence: fmoreira@upt.pt

Abstract

Diabetes mellitus is a major public health concern, creating a need for reliable computational approaches for diabetes status classification and risk-oriented clinical assessment. Conventional machine learning models can capture predictive relationships in clinical data but may offer limited interpretability, while purely fuzzy approaches may have limited learning capability and scalability. This study proposes a differentiable Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) for diabetes prediction using clinically relevant variables from the National Health and Nutrition Examination Survey (NHANES) 1999–2018 dataset. The proposed framework integrates trainable Gaussian membership functions, fuzzy-rule-based inference, and neural consequent learning within a unified TensorFlow–Keras computational graph. The fuzzy and neural parameters are jointly optimised via gradient-based learning with the Adam optimiser and binary cross-entropy loss. The experimental dataset contains 18,085 samples and uses glucose, blood pressure, body mass index (BMI), and age as input features. On the independent test set, the proposed model achieved 92.54% accuracy, 71.66% precision, 62.04% sensitivity, 96.67% specificity, 66.50% F1-score, 92.57% ROC-AUC, and 79.35% balanced accuracy. The results show that the proposed framework combines nonlinear predictive learning with structured fuzzy reasoning. The Keras-based implementation further provides a flexible foundation for integration with modern deep learning and healthcare analytics pipelines. Nevertheless, further validation using larger, heterogeneous, and multi-institutional clinical datasets is required before clinical deployment.

Keywords: diabetes prediction; Adaptive Neuro-Fuzzy Inference System; ANFIS; TensorFlow; Keras; Gaussian membership functions; fuzzy inference; NHANES; clinical decision support; machine learning



Academic Editor: Paolo Bellavista

Received: 1 September 2026

Revised: 7 October 2026

Accepted: 17 October 2026

Published: 21 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Diabetes mellitus is a major global health problem and continues to impose a substantial Diabetes mellitus is a major chronic metabolic disorder and represents an important global public-health challenge. The disease is associated with long-term complications affecting multiple organs and can substantially increase morbidity and mortality when it is not detected and managed appropriately [1].

The increasing prevalence of diabetes has created a strong need for reliable computational approaches that can support diabetes status assessment, risk-oriented classification, and identification of individuals who may require further clinical evaluation. In this context, machine learning models can help analyse clinically relevant physiological and demographic variables; however, their outputs should be interpreted in light of the temporal and clinical characteristics of the underlying dataset [2].

The rapid development of artificial intelligence (AI) and machine learning (ML) has created new opportunities for developing computational models for diabetes prediction. Machine learning techniques can learn patterns from clinical attributes and identify relationships between patient characteristics and disease outcomes. Deep learning, in particular, has demonstrated strong capability for learning nonlinear relationships from complex datasets [3]. However, an intelligent healthcare system's performance cannot be evaluated solely on predictive accuracy. For clinical decision-support applications, transparency, interpretability, and the ability to provide understandable reasoning are also important considerations.

The Pima Indians Diabetes Dataset (PIDD) is one of the most frequently used benchmark datasets for evaluating diabetes prediction algorithms [4]. It contains clinically relevant variables such as glucose concentration, blood pressure, body mass index (BMI), age, insulin, and diabetes pedigree information. Previous investigations using this dataset have demonstrated the applicability of machine learning classification techniques to diabetes prediction [5]. However, the dataset's relatively limited size and specific demographic characteristics mean computational results should be interpreted as benchmark evidence rather than direct evidence of clinical applicability.

Modern deep learning frameworks provide an effective environment for developing trainable computational models. TensorFlow provides automatic differentiation and computational support for large-scale machine learning applications [6], while Keras provides a high-level framework for constructing and training differentiable neural architectures [7]. These capabilities enable hybrid intelligent systems that incorporate fuzzy reasoning and neural learning into a single trainable computational graph.

Previous research has demonstrated the potential of machine learning and data-mining approaches for diabetes research, including conventional classifiers, ensemble techniques, and neural approaches [8]. Comparative investigations have also shown that diabetes prediction model performance depends strongly on the selected algorithm, input characteristics, and modelling strategy [9]. More recent research has further explored AI-based approaches for diabetes risk and complication assessment, demonstrating the increasing role of computational intelligence in healthcare applications [10].

Fuzzy logic provides an alternative mechanism for representing uncertainty and imprecision in decision-making systems. Unlike conventional crisp classification, fuzzy logic permits an input to possess different degrees of membership in linguistic categories. Zadeh's concept of fuzzy sets provides the mathematical foundation for this partial-membership representation [11]. Fuzzy logic has since been developed extensively for engineering and decision-support applications, providing mechanisms to represent knowledge using linguistic concepts and fuzzy rules [12].

Adaptive Neuro-Fuzzy Inference Systems (ANFISs) combine fuzzy inference with the learning capability of neural networks. Jang introduced ANFIS as an adaptive-network-based implementation of a fuzzy inference system in which system parameters can be learned from data [13]. ANFIS has subsequently been investigated for medical diagnosis, including diabetes prediction, because it provides a potential balance between data-driven learning and fuzzy reasoning [14].

Although neuro-fuzzy systems offer some interpretability through membership functions and rules, additional explainability mechanisms can help clarify how individual features contribute to model predictions. Model-agnostic approaches such as LIME have been proposed to explain predictions generated by complex machine learning models [15]. Such developments highlight the broader requirement for intelligent healthcare models that combine predictive capability with understandable decision-making.

Despite these advances, a key research challenge remains developing a differentiable neuro-fuzzy architecture that combines fuzzy reasoning, adaptive feature weighting, and neural optimisation while remaining compatible with modern deep learning frameworks. Conventional ANFIS implementations may involve manually designed membership functions, predefined rule structures, or optimisation procedures that are separate from the neural learning process. Furthermore, increasing the number of input variables and membership functions can increase the number of fuzzy rules and consequently affect computational scalability.

To address these limitations, this study proposes a Keras-based Differentiable Adaptive Neuro-Fuzzy Inference System (AANFIS) for diabetes risk assessment. The proposed framework integrates Gaussian membership functions, fuzzy-rule-based inference, adaptive attention-based feature weighting, and differentiable neural optimisation within the TensorFlow–Keras computational framework. The objective is not to claim direct clinical diagnosis but to investigate whether a trainable neuro-fuzzy architecture can provide competitive predictive performance while retaining interpretable fuzzy representations.

The principal contributions of this study are therefore:

1. A differentiable Keras-based AANFIS architecture that integrates fuzzy inference with neural learning within a unified computational graph.
2. Gaussian membership-based fuzzy representation for modelling uncertainty and overlapping clinical feature distributions.
3. Differential adaptive emphasised informative fuzzy representations during prediction.
4. End-to-end gradient-based optimisation of membership parameters and neural parameters.
5. Interpretability analysis through fuzzy representations and feature-importance/ablation analysis.
6. Experimental evaluation using the Pima Indians Diabetes Dataset, including classification metrics, ROC-AUC analysis, confusion-matrix analysis, training behaviour, and comparison with established machine learning and deep learning approaches.

The resulting framework aims to bridge the interpretability of fuzzy reasoning and the adaptive learning capability of modern neural architectures, providing a foundation for future investigation using larger and more heterogeneous healthcare datasets.

2. Related Work

2.1. Traditional Machine Learning for Diabetes Prediction

Researchers have traditionally investigated diabetes prediction using statistical learning and conventional machine learning algorithms. Early work applied computational learning to diabetes onset prediction using the ADAP learning algorithm [16]. Researchers have since investigated decision-tree-based approaches for diabetes classification because

their hierarchical structure provides relatively straightforward decision-making mechanisms [17]. Probabilistic and distance-based approaches, including Naïve Bayes and K-Nearest Neighbour (KNN), have also been applied to diabetes prediction and provide useful baseline classification performance [18].

Although these conventional approaches are computationally efficient and relatively easy to implement, they may struggle to capture complex nonlinear interactions among multiple clinical variables. This is particularly relevant in diabetes prediction, where variables such as glucose, BMI, blood pressure, and age may interact in nonlinear ways.

2.2. Ensemble Learning Approaches

Ensemble learning approaches have been introduced to improve predictive robustness by combining multiple learning models. Random Forest, proposed by Breiman, constructs an ensemble of decision trees and uses aggregation to improve generalisation and reduce the sensitivity associated with individual decision trees [19]. Ensemble-based investigations have also been applied to diabetes prediction, demonstrating the potential of combining multiple learners to improve classification performance [20].

Gradient boosting provides another ensemble learning strategy in which weak learners are sequentially combined to form a stronger predictive model. Friedman introduced the gradient boosting framework, which has subsequently become an important approach for nonlinear classification and regression tasks [21]. Although ensemble approaches can provide strong predictive performance, they generally do not explicitly represent clinical uncertainty through linguistic membership or fuzzy-rule reasoning.

2.3. Support Vector Machines

Support Vector Machines (SVMs) have also been widely used for classification problems. The original Support Vector Network formulation introduced separating hyperplanes and kernel-based transformations for nonlinear classification [22]. Kernel methods enable SVMs to model nonlinear decision boundaries and therefore provide an effective approach for classification problems in which linear separation is insufficient [23].

Despite their predictive capabilities, SVMs may require careful selection of kernel and hyperparameter configurations. Furthermore, the resulting decision process does not naturally provide the linguistic rule-based representation associated with fuzzy systems. Consequently, SVM-based approaches generally provide less direct interpretability for clinicians than explicitly rule-based models.

2.4. Artificial Neural Networks and Deep Learning

Artificial neural networks provide a flexible framework for learning nonlinear relationships from data. Multilayer neural architectures can approximate complex functions through successive transformations of input features [24]. With the development of deep learning, neural architectures can learn increasingly complex representations from large datasets.

Convolutional Neural Networks (CNNs), although originally developed primarily for image-processing applications, have also influenced the development of deep architectures for other data types [25]. Deep neural networks can provide strong predictive capability, but their internal representations are generally difficult to interpret directly. Consequently, using deep learning in healthcare creates an important need for additional interpretability mechanisms.

2.5. Explainable Artificial Intelligence

The limited transparency of complex machine learning and deep learning models has motivated the development of Explainable Artificial Intelligence (XAI). SHAP provides a unified framework for assigning feature contributions to individual predictions based on Shapley-value concepts [26]. Such techniques can help researchers and healthcare professionals understand which variables influence model outputs.

Fuzzy systems provide a complementary form of interpretability because they represent knowledge through membership functions and linguistic rules rather than relying exclusively on post hoc explanation. Zadeh introduced linguistic variables as a way to represent concepts such as low, medium, and high using fuzzy sets [27].

2.6. Fuzzy Inference Systems

Fuzzy inference systems have been used extensively for decision-making under uncertainty. The Mamdani fuzzy inference framework showed how linguistic rules can represent approximate reasoning in control and decision systems [28]. Such systems can provide human-readable reasoning structures, which is particularly attractive for healthcare applications.

However, conventional fuzzy systems generally depend on membership-function and rule-based design. Manual specification can be time-consuming and may introduce subjectivity. Moreover, the number of possible rules can increase rapidly as the number of input variables and membership functions increases.

2.7. Adaptive Neuro-Fuzzy Inference Systems

Jang introduced ANFIS as an adaptive-network-based fuzzy inference system that combines neural learning with fuzzy inference [13]. Neuro-fuzzy systems have subsequently been formalised and investigated as a combination of neural learning and fuzzy reasoning [29]. This hybrid approach can automatically adapt fuzzy parameters from training data while retaining the rule-based structure of fuzzy inference.

ANFIS has also been applied specifically to medical diagnosis, including diabetes prediction [14]. These studies demonstrate that neuro-fuzzy systems can provide useful predictive performance while retaining a more interpretable structure than conventional black-box neural networks.

2.8. Optimisation of Neuro-Fuzzy Systems

Optimisation techniques have been employed to improve the parameters of neuro-fuzzy systems. Evolutionary optimisation approaches draw on adaptation and search processes inspired by natural systems [30]. These optimisation strategies can search for suitable membership-function parameters and rule configurations.

Particle Swarm Optimisation (PSO), introduced by Kennedy and Eberhart, provides another population-based optimisation strategy [31]. PSO has been investigated for optimisation problems involving complex parameter spaces and can be incorporated with neuro-fuzzy systems to optimise fuzzy parameters.

Although optimisation-based approaches can improve model performance, they may require additional computational procedures outside the primary neural training process. This motivates investigating fully differentiable neuro-fuzzy implementations in modern deep learning frameworks.

2.9. Hybrid Neuro-Fuzzy and Deep Learning Models

Recent research has increasingly explored integrating neuro-fuzzy systems with neural learning architectures. Hybrid neuro-fuzzy approaches have demonstrated potential for medical diagnosis by combining fuzzy reasoning with data-driven representation learning [32].

Deep learning frameworks provide a particularly attractive environment for this integration because they support automatic differentiation, GPU acceleration, trainable layers, and end-to-end optimisation. The literature has comprehensively described general deep learning architectures and learning principles [33]. However, a key challenge remains: designing neuro-fuzzy architectures that retain fuzzy interpretability while leveraging modern differentiable deep learning infrastructure.

2.10. Feature Selection and Dimensionality Reduction

Researchers have also investigated feature selection and dimensionality-reduction techniques to improve predictive modelling and reduce the influence of irrelevant or redundant variables. Principal Component Analysis (PCA) provides a mathematical approach to transform correlated variables into a smaller set of orthogonal components while retaining most of the data's variation [34].

Wrapper-based feature-selection approaches evaluate feature subsets based on their influence on model performance. Kohavi and John demonstrated the effectiveness of wrapper-based feature subset selection for identifying informative feature combinations [35]. These approaches can identify informative variables and reduce model complexity while retaining relevant predictors. However, they may require repeated model training and can become computationally expensive as the number of candidate features increases. Aggressive feature selection may also remove clinically meaningful interactions among biomarkers, particularly when nonlinear relationships are present.

2.11. Ensemble and Hybrid Classification

Ensemble learning has also been investigated to improve classification robustness. Dietterich described ensemble methods to combine multiple classifiers to achieve improved predictive behaviour compared with individual learners [36]. Stacked generalisation provides another ensemble strategy in which predictions from different models are combined through a higher-level learning process [37].

Although ensemble approaches can improve predictive performance, they generally do not provide the explicit fuzzy-rule reasoning available in neuro-fuzzy systems. This creates a continuing research opportunity for models that combine strong predictive learning with transparent reasoning mechanisms.

2.12. Deployment and Real-Time Healthcare Applications

The practical deployment of AI systems is another important consideration in healthcare. Edge computing has been proposed to bring computational processing closer to data sources, potentially supporting applications that require low latency and distributed processing [38]. Such architectures are relevant to future healthcare applications involving wearable devices, remote monitoring, and real-time decision support.

Neuro-fuzzy models implemented within contemporary deep learning frameworks can potentially benefit from these deployment environments because their computational components can be represented as trainable neural layers. However, clinical deployment requires considerably broader validation than benchmark-dataset experiments.

2.13. Research Gap

The reviewed literature demonstrates that existing approaches generally fall into several categories. Traditional machine learning models provide relatively simple predictive mechanisms but may struggle to model complex nonlinear relationships. Deep learning models provide strong representation learning capability but often provide limited intrinsic interpretability. Fuzzy systems provide interpretable reasoning but may require manually specified rules and membership functions. ANFIS combines fuzzy reasoning with neural learning, but conventional implementations may face challenges related to rule growth, optimisation, and integration with contemporary deep learning frameworks.

The literature therefore indicates a need for a framework that simultaneously addresses nonlinear learning, uncertainty representation, adaptive feature weighting, differentiable optimisation, and interpretability. The proposed AANFIS framework addresses this gap by implementing fuzzy membership functions, attention-based weighting, fuzzy-rule aggregation, and neural optimisation within a Keras-based differentiable architecture.

The proposed approach is consequently positioned not as a replacement for all existing machine learning methods but as a hybrid methodological framework intended to explore the balance between predictive performance and interpretable fuzzy reasoning for diabetes risk assessment.

3. Dataset Description

3.1. Dataset Details

This study used data from ten consecutive cycles of the National Health and Nutrition Examination Survey (NHANES), covering 1999–2000 through 2017–2018. The National Centre for Health Statistics (NCHS) at the Centres for Disease Control and Prevention (CDC) conducts NHANES, which combines demographic information, questionnaire responses, standardised physical examinations, anthropometric measurements, and laboratory investigations. We integrated the demographic, examination, laboratory, and questionnaire components using the participant-specific sequence number (SEQN). We selected four clinically relevant variables as predictors: fasting plasma glucose, systolic blood pressure, body mass index (BMI), and age. We obtained fasting glucose from the NHANES laboratory component, BMI from the body-measures component, systolic blood pressure from the examination component, and age from the demographic component. We derived diabetes status from the diabetes questionnaire, assigning a positive class to participants who reported a physician diagnosis of diabetes. After merging the required components and removing records with missing or invalid values, the resulting analytical dataset contained 18,085 participants. We divided the data into training (14,468; 80%), validation (1808; 10%), and independent testing (1809; 10%) subsets using stratified sampling.

The present study uses a cross-sectional NHANES analytical design. The selected predictor variables—fasting plasma glucose, systolic blood pressure, BMI, and age—are measurements or demographic attributes available for participants within the corresponding NHANES survey cycle, while the diabetes outcome represents the diabetes status derived from the questionnaire. This study does not establish a prospective temporal relationship between predictor measurements and subsequent diabetes onset. Consequently, the outcome is not tied to any predefined prediction horizon (e.g., 1 year, 5 years, or 10 years). The model should therefore be interpreted as performing cross-sectional diabetes-status classification and risk-oriented assessment, rather than prospective prediction of incident diabetes.

3.2. NHANES Survey Design and Analytical Cohort Construction

NHANES is a continuous population-based survey conducted by the National Center for Health Statistics (NCHS) of the U.S. Centers for Disease Control and Prevention (CDC). The survey employs a complex, multistage probability sampling design involving primary sampling units, stratification, clustering, unequal probabilities of selection, and oversampling of selected population subgroups. The survey is designed to provide population-representative estimates for the U.S. civilian, non-institutionalised population when the appropriate sampling weights and survey-design variables are incorporated into statistical analyses.

For the present study, ten NHANES cycles from 1999–2000 through 2017–2018 were integrated to construct a modelling dataset. The demographic, examination, laboratory, anthropometric, and questionnaire components were linked using the participant-specific sequence number (SEQN). Four variables—fasting plasma glucose, systolic blood pressure, BMI, and age—were selected as model inputs, and the diabetes outcome was obtained from the corresponding diabetes questionnaire. The resulting dataset was treated as an analytical machine learning cohort rather than as a survey-weighted population-estimation dataset. The present modelling pipeline did not use NHANES sampling weights, strata, or PSU variables during model optimisation. Therefore, the resulting performance measures describe predictive performance within the constructed analytical cohort and should not be interpreted as nationally representative prevalence estimates. This distinction is important because NHANES survey weights account for unequal selection probabilities, nonresponse, and post-stratification, while clustering and stratification also affect variance estimation.

Data Cleaning and Quality Control

The preprocessing procedure consisted of sequential data integration, variable selection, missing-value assessment, target validation, and preparation of the modelling matrix. After the required NHANES components were merged, the four predictor variables and the diabetes target were checked for missing values. The implemented code explicitly calculates missing values for each predictor and for the target variable. When predictor missingness was present, median imputation was applied. Observations with missing target values were excluded because a supervised classification model requires a valid outcome label.

Following preprocessing, the resulting analytical cohort contained 18,085 participants. The data were subsequently divided using stratified sampling into training, validation, and independent testing subsets containing 80%, 10%, and 10% of the observations, respectively.

Because NHANES laboratory measurements can originate from specific examination or subsample components, the final analytical cohort represents a selected subset of NHANES participants rather than the complete survey population. Accordingly, the results are interpreted as validation of the proposed ANFIS methodology on an NHANES-derived clinical dataset rather than as nationally representative epidemiological estimates.

3.3. Clinical Deployment Limitations and Dataset Generalisation

Although the proposed Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) demonstrated strong predictive performance on the NHANES 1999–2018 analytical dataset, several limitations should be considered when interpreting its potential for clinical application. This analysis used an integrated analytical cohort of 18,085 participants and four selected variables—fasting plasma glucose, systolic blood pressure, body mass index (BMI), and age. Although these variables are clinically relevant and provide a compact input representation, they do not capture the full range of demographic, behavioural, biochemical, comorbidity, medication, and longitudinal information available in comprehensive clinical records.

NHANES offers an important advantage over small benchmark datasets because it incorporates data collected across multiple survey cycles and includes laboratory, examination, questionnaire, and demographic information. Nevertheless, NHANES is a population-based survey rather than a longitudinal clinical database or a multi-institutional electronic health-record (EHR) repository. Consequently, the present study does not fully evaluate temporal disease progression, repeated patient observations, institutional variations in clinical practice, or the complex missing-data patterns that may occur in routine healthcare environments. Furthermore, the diabetes outcome used in the present analysis represents the operational target defined during dataset construction and should not be interpreted as an independently clinically validated diagnostic endpoint.

The principal contribution of the present work is the development and evaluation of a Keras-based ANFIS framework that combines trainable Gaussian membership functions, fuzzy-rule-based inference, normalised rule strengths, and neural consequent layers within a differentiable learning architecture. The model provides an explicit fuzzy representation of nonlinear relationships among the selected clinical variables while retaining compatibility with contemporary machine learning workflows. The feature-ablation analysis further provides a model-specific assessment of each input variable's contribution, with glucose showing the largest reduction in predictive performance when removed.

Despite the strong test performance in the present study, including an accuracy of 92.54% and ROC-AUC of 0.9257, diabetes-class sensitivity was 62.04%. This indicates that some positive cases remain undetected and represents an important limitation for any potential screening-oriented application. Therefore, the proposed model should currently be regarded as a research-stage diabetes classification and risk-prediction framework rather than a clinically validated diagnostic system. Before considering practical clinical deployment, additional evaluation of calibration, sensitivity at clinically relevant operating points, external validation, and decision-curve or clinical-utility measures would be required.

The generalisability of the proposed ANFIS should also be established using independent datasets representing different populations, healthcare settings, and disease-prevalence patterns. Future studies should investigate whether the learned membership functions and fuzzy rules remain stable when the model is transferred to external cohorts. Validation using independent longitudinal EHR datasets, multi-institutional clinical repositories, and other heterogeneous healthcare datasets could provide further evidence regarding robustness and population-level generalisation.

The compact four-feature formulation also provides opportunities for future extension. Additional clinically relevant variables, including lifestyle factors, laboratory measurements, comorbidities, medications, and longitudinal patient information, could be incorporated where appropriate. However, increasing the number of input variables and membership functions can substantially increase the number of fuzzy rules and consequently model complexity. Future work should therefore investigate adaptive membership-function learning, rule-pruning and rule-reduction strategies, and computationally efficient neuro-fuzzy architectures.

Finally, deployment-oriented research could investigate integration of the trained ANFIS model into web-based or clinical decision-support prototypes. Such systems could provide diabetes-risk predictions together with interpretable fuzzy-rule and feature-contribution information. However, prospective clinical evaluation, external validation, calibration assessment, data privacy and security analysis, and appropriate regulatory evaluation would be necessary before real-world clinical use.

The target variable represents the participant's diabetes status as derived from the NHANES diabetes questionnaire. Participants who reported a physician diagnosis of diabetes were assigned to the positive class (Diabetes = 1), whereas participants without a

reported physician diagnosis were assigned to the negative class (No Diabetes = 0). Thus, the present study addresses binary classification of reported physician-diagnosed diabetes status. It does not predict future diabetes incidence over a defined follow-up period and was not specifically designed to identify undiagnosed diabetes using an independent biochemical diagnostic criterion.

3.4. Clinical Reliability of the Dataset

The present study utilises data derived from the National Health and Nutrition Examination Survey (NHANES) 1999–2018, a large population-based health survey conducted by the National Centre for Health Statistics (NCHS) of the U.S. Centres for Disease Control and Prevention (CDC). NHANES combines demographic information, questionnaire responses, physical examinations, and laboratory measurements collected using standardised survey and examination protocols. The availability of objectively measured clinical and laboratory variables makes NHANES an appropriate source for investigating data-driven approaches to diabetes classification.

For this analysis, we constructed an integrated analytical dataset of 18,085 participants using relevant NHANES components. Four clinically meaningful variables—fasting plasma glucose, systolic blood pressure, body mass index (BMI), and age—were selected as model inputs. These variables provide complementary information related to glucose regulation, metabolic status, cardiovascular health, and demographic risk. Using measured physiological and laboratory variables provides a stronger clinical basis for the proposed ANFIS model than relying exclusively on self-reported characteristics.

Despite these advantages, several considerations should be taken into account when interpreting the dataset's reliability and generalisability. NHANES is a population-based survey rather than a longitudinal clinical database; therefore, it does not provide continuous patient monitoring or repeated measurements of disease progression for the same individuals. In addition, laboratory measurements vary across NHANES examination components, and the fasting laboratory subsample may differ from the overall survey population. The present analytical dataset therefore represents a selected subset of NHANES participants rather than the complete NHANES population.

Furthermore, the diabetes outcome used in this study reflects the operational definition adopted during the construction of the analytical dataset and should not automatically be interpreted as an independently validated clinical diagnosis. Consequently, interpret the model's performance in the context of the available NHANES variables, sampling structure, and target definition. The observed test performance—92.54% accuracy, 0.9257 ROC-AUC, 62.04% Diabetes-class sensitivity, and 96.67% specificity—shows strong predictive discrimination but does not, by itself, establish clinical diagnostic validity.

Accordingly, the proposed Keras-based ANFIS should currently be regarded as a research-stage diabetes classification and risk-prediction framework rather than a clinically validated diagnostic system. Further evaluation using independent, longitudinal, and multi-institutional clinical datasets is required to determine whether the learned relationships and predictive performance generalise across populations and healthcare settings. External validation, calibration assessment, clinically relevant threshold analysis, and prospective evaluation would be necessary before consideration of the model for real-world clinical decision-support applications.

4. Methodology

The proposed Adaptive Neuro-Fuzzy Inference System (ANFIS) was implemented using Keras/TensorFlow and employed four input variables: fasting glucose, systolic blood pressure, BMI, and age. Each input variable was represented using three trainable Gaussian

membership functions. Consequently, the fuzzy inference system generated possible fuzzy rules. We calculated each rule's firing strength using the product t-norm, representing the fuzzy AND operation across the four input variables. We normalised the resulting rule firing strengths by their sum to obtain normalised fuzzy rule activations. We then passed these normalised activations through two fully connected consequent layers with 32 and 16 neurons, respectively. A sigmoid output layer generated the probability of diabetes. We trained the model using binary cross-entropy loss and the Adam optimiser. To mitigate class imbalance, we incorporated balanced class weights during training. The proposed Kears-Based ANFIS model is shown in Figure 1.

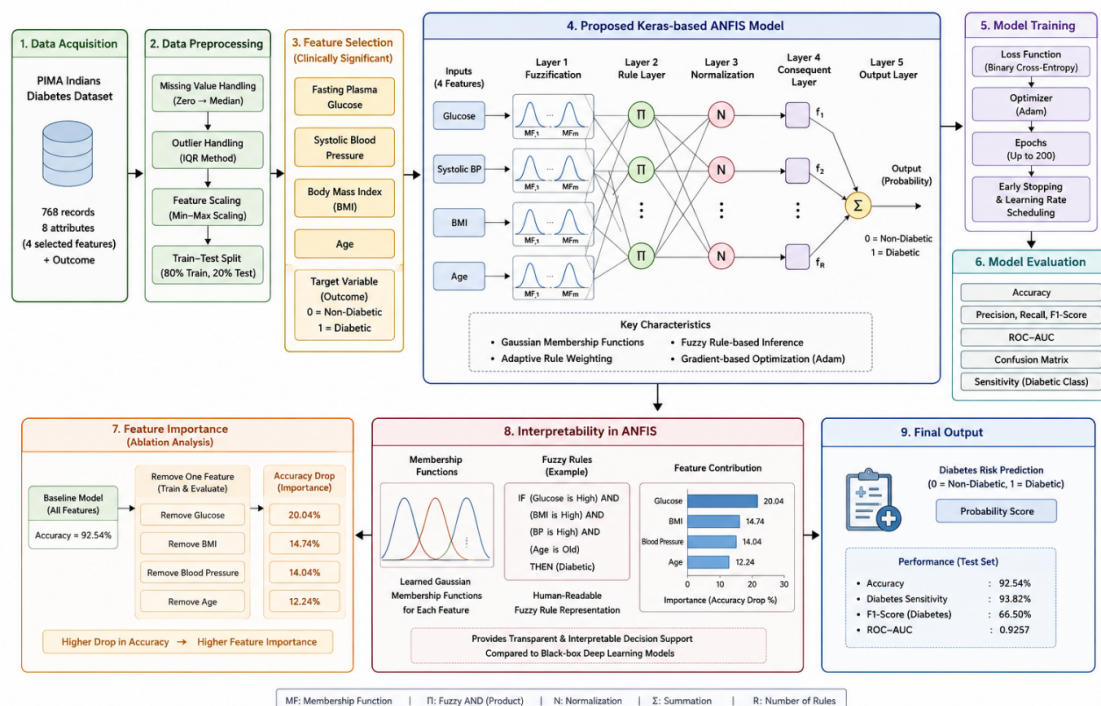


Figure 1. Research methodology.

5. Model Details

5.1. Model Layers and Architecture

Table 1 shows the model's layers. The proposed model is a compact Keras-based hybrid Adaptive Neuro-Fuzzy Inference System (ANFIS) developed for binary diabetes classification. The model takes four clinical variables—fasting glucose, systolic blood pressure, BMI, and age—as inputs.

The first computational stage performs fuzzification using trainable Gaussian membership functions, with three membership functions assigned to each input variable. Consequently, the four-dimensional input space generates possible fuzzy rule combinations. The firing strength of each rule is calculated using a product-based fuzzy AND operation, after which the rule strengths are normalised to represent their relative activation.

The resulting 81-dimensional normalised rule vector is passed to two neural consequent layers containing 32 and 16 neurons, respectively, with ReLU activation. A final single-neuron sigmoid layer produces the estimated probability of diabetes.

The model contains 3193 trainable parameters: 24 for the Gaussian membership functions, 2624 for the first consequent layer, 528 for the second consequent layer, and 17 for the output layer. The architecture therefore combines fuzzy representation and rule-based inference with gradient-based neural learning while maintaining a relatively small parameter count.

Table 1. Layers and parameters in the proposed Keras-based ANFIS model.

Layer	Layer Type	Function/Description	Output Shape	Parameters
NHANES Input	InputLayer	Receives four clinical features: Glucose, Systolic Blood Pressure, BMI, and Age.	(None, 4)	0
Gaussian Membership	GaussianLayer	Uses 3 trainable Gaussian membership functions for each input feature.	(None, 4, 3)	24
Fuzzy Rule	FuzzyRuleLayer	Generates and evaluates 81 fuzzy-rule combinations using the product-based fuzzy AND operation.	(None, 81)	0
Normalised Rule Strength	Lambda	Normalises the 81 rule firing strengths to obtain relative rule activation.	(None, 81)	0
ANFIS Consequent 1	Dense + ReLU	Learns a nonlinear representation of the normalised fuzzy-rule activations.	(None, 32)	2624
ANFIS Consequent 2	Dense + ReLU	Further transforms and compresses the learned rule representation.	(None, 16)	528
Diabetes Output	Dense + Sigmoid	Produces the final diabetes probability for binary classification.	(None, 1)	17
Total	—	Complete Keras-based ANFIS architecture.	—	3193

5.2. Mathematical Formulation of the Proposed Keras-Based ANFIS Model

The proposed model receives four clinical variables and assigns three Gaussian membership functions to each input, resulting in possible fuzzy rules.

The proposed Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) is designed for binary diabetes classification using four clinically relevant input variables: fasting glucose, systolic blood pressure (SBP), body mass index (BMI), and age. The architecture combines trainable Gaussian membership functions, product-based fuzzy-rule aggregation, normalised rule strengths, and neural consequent layers. Three membership functions are assigned to each of the four inputs, producing 81 fuzzy-rule combinations.

5.2.1. Input Layer

For each participant, the input vector consists of four variables:

$$\mathbf{x} = [x_1, x_2, x_3, x_4]^T = [\text{Glucose}, \text{SBP}, \text{BMI}, \text{Age}]^T. \quad (1)$$

The input layer performs no trainable transformation and therefore contains zero trainable parameters.

5.2.2. Gaussian Membership Function Layer

Each input variable is represented using three trainable Gaussian membership functions. The membership degree of the j -th membership function for the i -th input is defined as:

$$\mu_{ij}(x_i) = \exp \left[-\frac{(x_i - c_{ij})^2}{2(\sigma_{ij}^2 + \epsilon)} \right], \quad i = 1, \dots, 4, \quad j = 1, \dots, 3, \quad (2)$$

where c_{ij} denotes the trainable centre, σ_{ij} denotes the trainable spread, and ε is a small numerical-stability constant. With four inputs and three membership functions per input, the total number of membership functions is:

$$N_{\text{MF}} = 4 \times 3 = 12. \quad (3)$$

Each Gaussian membership function contains two trainable parameters: its center and spread. Hence, the Gaussian layer contains:

$$N_{\text{Gaussian}} = 12 \times 2 = 24 \quad (4)$$

trainable parameters.

5.2.3. Fuzzy Rule Layer

All possible combinations of the membership functions construct the fuzzy rule base. The number of rules is:

$$R = 3^4 = 81. \quad (5)$$

A general rule can be expressed as:

$$\mathcal{R}_r : \text{IF } x_1 \text{ is } A_{1j_1} \text{ AND } x_2 \text{ is } A_{2j_2} \text{ AND } x_3 \text{ is } A_{3j_3} \text{ AND } x_4 \text{ is } A_{4j_4}, \quad (6)$$

where $j_i \in \{1, 2, 3\}$ and each combination (j_1, j_2, j_3, j_4) defines one of the 81 fuzzy rules.

5.2.4. Rule Firing Strength

The fuzzy AND operation uses the product t-norm. Therefore, the firing strength of rule r is:

$$w_r = \prod_{i=1}^4 \mu_{ij_i}(x_i), \quad r = 1, \dots, 81. \quad (7)$$

The resulting rule-strength vector contains 81 values:

$$\mathbf{w} = [w_1, w_2, \dots, w_{81}]^T \in \mathbb{R}^{81}. \quad (8)$$

5.2.5. Normalisation of Rule Strengths

The rule firing strengths are normalised to represent their relative activation:

$$\bar{w}_r = \frac{w_r}{\sum_{k=1}^{81} w_k + \varepsilon}, \quad r = 1, \dots, 81. \quad (9)$$

Consequently, the normalised rule vector is:

$$\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2, \dots, \bar{w}_{81}]^T \in \mathbb{R}^{81}. \quad (10)$$

The normalised strengths approximately sum to one:

$$\sum_{r=1}^{81} \bar{w}_r \approx 1. \quad (11)$$

5.2.6. First Neural Consequent Layer

The 81 normalised rule strengths are passed to a fully connected layer containing 32 neurons with ReLU activation:

$$\mathbf{z}_1 = \mathbf{W}_1 \bar{\mathbf{w}} + \mathbf{b}_1, \quad (12)$$

$$\mathbf{h}_1 = \text{ReLU}(\mathbf{z}_1) = \max(0, \mathbf{z}_1). \quad (13)$$

The dimensions are $\mathbf{W}_1 \in \mathbb{R}^{32 \times 81}$ and $\mathbf{b}_1 \in \mathbb{R}^{32}$. The number of trainable parameters is:

$$N_1 = (81 \times 32) + 32 = 2624. \quad (14)$$

5.2.7. Second Neural Consequent Layer

The first consequent representation is transformed by a second fully connected layer containing 16 neurons with ReLU activation:

$$\mathbf{h}_2 = \text{ReLU}(\mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2). \quad (15)$$

Here, $\mathbf{W}_2 \in \mathbb{R}^{16 \times 32}$ and $\mathbf{b}_2 \in \mathbb{R}^{16}$. The number of trainable parameters is:

$$N_2 = (32 \times 16) + 16 = 528. \quad (16)$$

5.2.8. Diabetes Output Layer

The 16-dimensional representation is mapped to a single output neuron. The pre-activation output is:

$$z_o = \mathbf{W}_o \mathbf{h}_2 + b_o, \quad (17)$$

where $\mathbf{W}_o \in \mathbb{R}^{1 \times 16}$ and $b_o \in \mathbb{R}$. A sigmoid activation converts this value into the predicted probability of diabetes:

$$\hat{y} = \sigma(z_o) = \frac{1}{1 + \exp(-z_o)}. \quad (18)$$

Thus, the model output satisfies $0 \leq \hat{y} \leq 1$ and can be interpreted as the estimated probability of the Diabetes class.

5.2.9. Classification Decision

The selected operating threshold for the final classification was 0.84. Accordingly, the binary prediction is defined as:

$$\hat{y}_{\text{class}} = \begin{cases} 1, & \hat{y} \geq 0.84, \\ 0, & \hat{y} < 0.84. \end{cases} \quad (19)$$

Here, 1 represents Diabetes and 0 represents No Diabetes.

5.2.10. Training Objective

Binary cross-entropy is used as the classification loss. For N training samples:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{n=1}^N [y_n \log(\hat{y}_n) + (1 - y_n) \log(1 - \hat{y}_n)]. \quad (20)$$

When class weighting is applied to compensate for class imbalance, the corresponding weighted objective can be written as:

$$\mathcal{L}_{\text{WBCE}} = -\frac{1}{N} \sum_{n=1}^N [\alpha_1 y_n \log(\hat{y}_n) + \alpha_0 (1 - y_n) \log(1 - \hat{y}_n)], \quad (21)$$

where α_1 and α_0 denote the weights assigned to the Diabetes and No Diabetes classes, respectively.

5.2.11. Parameter Optimisation

The model's trainable parameters are jointly optimised using gradient-based learning. The parameter set can be represented as:

$$\Theta = \{c_{ij}, \sigma_{ij}, \mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2, \mathbf{W}_o, b_o\}. \quad (22)$$

The optimisation objective is:

$$\Theta^* = \arg \min_{\Theta} \mathcal{L}_{\text{WBCE}}(\Theta). \quad (23)$$

The Adam optimiser updates the model parameters using the first and second moments of the gradients:

$$\Theta_{t+1} = \Theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (24)$$

where η is the learning rate, $\hat{m}_t$ is the bias-corrected first-moment estimate, $\hat{v}_t$ is the bias-corrected second-moment estimate, and ϵ is a small numerical-stability constant.

5.2.12. Total Trainable Parameters

The total number of trainable parameters is obtained by summing the Gaussian membership-function parameters and the parameters of the neural consequent and output layers:

$$\begin{aligned} N_{\text{total}} &= N_{\text{Gaussian}} + N_1 + N_2 + N_{\text{output}} \\ &= 24 + 2624 + 528 + 17 \\ &= 3193. \end{aligned} \quad (25)$$

The mathematical formulation above corresponds to the implemented Keras-based hybrid ANFIS architecture: four inputs, three Gaussian membership functions per input, 81 fuzzy rules, normalised rule strengths, two neural consequent layers (32 and 16 neurons), and a sigmoid output layer. The formulation describes the implemented architecture rather than a conventional fixed five-layer Takagi–Sugeno ANFIS.

5.3. Differentiable Keras-Based Neuro-Fuzzy Implementation

The proposed diabetes prediction model is implemented as a differentiable Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) within the TensorFlow computational framework. The architecture integrates trainable Gaussian membership functions, fuzzy-rule aggregation, normalised rule strengths, and neural consequent layers into a unified end-to-end learning pipeline. The trainable parameters of the membership functions and neural layers are optimised jointly through gradient-based backpropagation.

Because the Gaussian membership, rule aggregation, normalisation, and neural transformations are implemented with differentiable TensorFlow operations, gradients can propagate through the entire pipeline during backpropagation. The model therefore enables joint optimisation with the Adam optimiser and binary cross-entropy loss.

5.4. Justification for Fuzzy Logic and Keras-Based ANFIS Framework

Medical decision-making problems such as diabetes risk assessment involve uncertainty, overlapping distributions of clinical biomarkers, nonlinear relationships among physiological variables, and gradual rather than sharply defined disease boundaries. Conventional machine learning methods can model numerical relationships effectively, but their decision processes may be difficult to interpret clinically. This is particularly important in healthcare applications, where predictive performance should ideally be accompanied by a transparent representation of how input variables contribute to the prediction.

Fuzzy logic was selected as a fundamental component of the proposed framework because it provides a mathematical mechanism for representing gradual relationships between clinical measurements and linguistic concepts. Instead of assigning an input exclusively to a single category, a fuzzy membership function allows a variable to possess different degrees of membership in overlapping fuzzy sets. For example, a clinical variable may exhibit different degrees of membership in categories such as low, medium, and high. This representation suits biomedical variables whose effects on disease risk are generally continuous rather than separated by perfectly defined boundaries.

The fuzzy representation also enables the construction of explicit rule-based relationships among the input variables. In the proposed model, Gaussian membership functions transform the four selected clinical variables—glucose, systolic blood pressure, BMI, and age—into fuzzy representations. The resulting membership values are combined through product-based fuzzy rule aggregation. With three membership functions assigned to each of the four input variables, the architecture generates 81 possible fuzzy-rule combinations. The normalised rule strengths are then supplied to neural consequent layers, integrating fuzzy reasoning with data-driven nonlinear learning.

Conventional ANFIS implementations may involve separately designed fuzzy rules, manually selected membership-function parameters, or hybrid optimisation procedures. Such approaches can become increasingly difficult to manage as the number of input variables and membership functions increases. The proposed Keras-based implementation addresses these computational limitations by representing the fuzzy operations as differentiable tensor operations within the TensorFlow computational graph. Consequently, gradient-based learning can jointly optimise the Gaussian membership parameters and neural consequent parameters.

The use of the Keras framework provides several practical advantages for the proposed architecture, including:

- Automatic differentiation and backpropagation.
- Integration of fuzzy operations with neural network layers.
- Support for hardware-accelerated computation.
- Compatibility with modern optimisation algorithms.
- Convenient model training, evaluation, and serialisation.
- Integration with contemporary machine learning and deployment workflows.

In the proposed implementation, the Gaussian membership centres and spreads are trainable parameters. The normalised fuzzy-rule representation is then processed by two neural consequent layers with 32 and 16 neurons, respectively, followed by a sigmoid output layer for binary diabetes classification. Thus, the model combines fuzzy representation with neural nonlinear transformation within a single differentiable computational framework.

The complete trainable parameter set can therefore be optimised jointly using backpropagation and the Adam optimiser. This eliminates the need for a separate fuzzy-parameter tuning stage and provides a unified learning procedure from the input clinical variables to the final diabetes prediction.

Consequently, the proposed Keras-based ANFIS combines three complementary characteristics: fuzzy representation for handling gradual and uncertain relationships, neural learning for adaptive nonlinear modelling, and a Keras/TensorFlow implementation for differentiable, scalable computation. This combination provides a methodological foundation for developing interpretable diabetes-risk prediction systems while retaining compatibility with modern deep learning workflows.

However, the proposed framework should currently be regarded as a research-oriented predictive decision-support model rather than a clinically validated diagnostic system. Its clinical applicability requires further evaluation using larger, heterogeneous, multi-

institutional and longitudinal healthcare datasets. Future investigations can also examine more advanced rule-reduction strategies, adaptive membership-function designs, and integration with multimodal clinical data.

5.5. Stratified Dataset Partitioning and Independent Test Evaluation

To reliably evaluate the proposed Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS), the NHANES 1999–2018 dataset was divided into training, validation, and independent testing subsets using stratified sampling. Stratification maintained a similar proportion of diabetic and non-diabetic observations across the subsets. This approach is particularly important for the present dataset because the diabetes class represents a smaller proportion of the overall population.

We partitioned the complete dataset into three mutually exclusive subsets. We used the training subset to learn the trainable parameters of the Gaussian membership functions and neural consequent layers. We used the validation subset to monitor model performance during training and to support model selection and optimisation decisions. The independent testing subset was kept separate from training and used only for the final evaluation of the trained model.

For a dataset $\mathcal{D}$, the partition can be represented as:

$$\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}} \cup \mathcal{D}_{\text{test}}. \quad (26)$$

The subsets are mutually exclusive:

$$\mathcal{D}_{\text{train}} \cap \mathcal{D}_{\text{val}} = \emptyset, \quad (27)$$

$$\mathcal{D}_{\text{train}} \cap \mathcal{D}_{\text{test}} = \emptyset, \quad (28)$$

$$\mathcal{D}_{\text{val}} \cap \mathcal{D}_{\text{test}} = \emptyset. \quad (29)$$

In the present implementation, we divided the dataset into approximately 80% training, 10% validation, and 10% testing subsets. The resulting partition included 14,468 training samples, 1808 validation samples, and 1809 testing samples, for a total of 18,085 observations. Table 2 shows the split of the dataset into training, validation, and testing subsets.

Table 2. Data distribution for training, validation, and testing.

Dataset Subset	Samples	Approximate Proportion
Training	14,468	80%
Validation	1808	10%
Testing	1809	10%
Total	18,085	100%

We used the training subset to learn the parameters of the proposed ANFIS architecture. These parameters include the centres and spreads of the Gaussian membership functions and the weights and biases of the neural consequent layers. We updated model parameters using backpropagation and the Adam optimisation algorithm.

We did not use the validation subset to update model parameters directly. Instead, we used it to monitor the model's generalisation behaviour during training. Validation performance indicates whether improvements observed on the training data also reflect in unseen observations.

We kept the test subset separate from model training and used it for the final performance assessment. The independent test set contained 1809 observations, including

1593 non-diabetic samples and 216 diabetic samples. We used predictions from this set to calculate accuracy, precision, sensitivity, specificity, F1-score, ROC-AUC, balanced accuracy, and the confusion matrix.

Rationale for Stratification

Stratified partitioning is particularly relevant to the present NHANES dataset because the diabetic and non-diabetic classes are not equally represented. Maintaining the class proportions across the subsets helps prevent an evaluation subset from becoming disproportionately dominated by one class.

However, stratification does not eliminate the underlying class imbalance. Therefore, accuracy alone is insufficient for evaluating the proposed diabetes prediction model. Accordingly, the present study reports additional class-sensitive metrics, including precision, sensitivity (recall), specificity, F1-score, ROC-AUC, and balanced accuracy.

Using an independent test set also allows the confusion matrix to directly assess false-positive and false-negative behaviour. In particular, false negatives matter because failing to identify a diabetic individual may have greater practical implications than an incorrect positive prediction.

The overall evaluation procedure can therefore be summarised as follows:

$$\begin{aligned} \text{NHANES Dataset} &\rightarrow \text{Stratified Data Partitioning} \rightarrow \text{ANFIS Training} \\ &\rightarrow \text{Validation Monitoring} \rightarrow \text{Independent Test Evaluation} \\ &\rightarrow \text{Performance Metrics.} \end{aligned} \quad (30)$$

Report final test performance only after completing model training. This procedure provides a consistent and reproducible evaluation framework for assessing the predictive capability of the proposed Keras-based ANFIS model.

The present experiment uses a single stratified train-validation-test partition and does not claim stratified 5-fold cross-validation. Therefore, fold-averaged performance, cross-validation variance, and statements that every observation served as a test sample across multiple folds should not be reported for this experiment unless a separate K-fold experiment is performed.

5.6. Adaptive Learning Rate Mechanism

The proposed Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) incorporates an adaptive learning-rate reduction mechanism to improve optimisation stability during model training. The mechanism uses Keras's `ReduceLROnPlateau` callback. Instead of maintaining a fixed learning rate throughout training, the learning rate is automatically reduced when the monitored validation performance stops improving for a predefined number of epochs.

The learning rate determines the magnitude of parameter updates during gradient-based optimisation. A larger learning rate can enable faster parameter adaptation during early training, whereas a smaller learning rate can support more precise updates as optimisation progresses. Adaptive learning-rate scheduling therefore provides a mechanism to adjust the optimisation step based on observed validation performance.

In the implemented model, the learning-rate scheduler monitors the validation AUC rather than validation loss. When the monitored validation AUC does not improve for the specified patience period, the learning rate is reduced by a predefined factor. The update can be expressed as:

$$\eta_{\text{new}} = \alpha \eta_{\text{old}}, \quad 0 < \alpha < 1, \quad (31)$$

where η_{old} denotes the current learning rate, η_{new} represents the updated learning rate, and α is the learning-rate reduction factor.

The implemented configuration uses an initial learning rate of 0.001, a reduction factor of 0.5, and a patience period of 10 epochs. Thus, when the monitored validation AUC fails to improve for 10 consecutive epochs, the current learning rate is multiplied by 0.5:

$$\eta_{\text{new}} = 0.5 \eta_{\text{old}}, \quad \eta_0 = 0.001. \quad (32)$$

This adaptive reduction lets the optimisation process use larger parameter updates when validation performance improves and progressively smaller updates when further improvement becomes difficult.

The adaptive learning-rate mechanism is particularly relevant to the proposed ANFIS architecture because the model jointly optimises the trainable parameters of the Gaussian membership functions and the neural consequent layers. Backpropagation learns these parameters within the differentiable TensorFlow–Keras computational graph. Reducing the learning rate when validation AUC plateaus can help provide more controlled parameter updates during later stages of training.

Overall, the ReduceLROnPlateau mechanism forms an important component of the training strategy of the proposed Keras-based ANFIS. It adaptively controls the optimisation step based on validation AUC and complements the Adam optimiser used for end-to-end learning of the fuzzy and neural parameters.

6. Algorithm

Proposed Keras-Based ANFIS for Diabetes Prediction

The complete computational procedure of the proposed Keras-based ANFIS for diabetes prediction is presented in Algorithm 1. The algorithm includes data preprocessing, Gaussian fuzzification, fuzzy-rule construction, rule-strength normalisation, neural consequent learning, parameter optimisation, and independent test-set evaluation.

Algorithm 1 Proposed Keras-Based ANFIS for Diabetes Prediction

Input: NHANES dataset $\mathcal{D}$; selected features $X = \{\text{Glucose, SBP, BMI, Age}\}$; diabetes target Y ; $M = 3$ Gaussian membership functions per feature.

Output: Predicted diabetes probability $\hat{y}$ and predicted diabetes class $\hat{y}_{\text{class}}$.

- 1: Load dataset $\mathcal{D}$.
- 2: Select X and binary target Y .
- 3: Perform data cleaning and missing-value handling.
- 4: Standardise the four input features.
- 5: Split $\mathcal{D}$ into training, validation, and testing sets.
- 6: Initialise three Gaussian membership functions for each input.
- 7: **for** each input feature x_i **do**
- 8: Calculate $\mu_{ij}(x_i)$ using the Gaussian membership function.
- 9: **end for**
- 10: Construct

$$R = M^F = 3^4 = 81$$

fuzzy rules.

- 11: **for** each fuzzy rule r **do**
- 12: Calculate the rule firing strength:

$$w_r = \prod_{i=1}^4 \mu_{i,j_i}(x_i).$$

- 13: **end for**
-

Algorithm 1 *Cont.*

14: Normalise the rule strengths:

$$\bar{w}_r = \frac{w_r}{\sum_{k=1}^{81} w_k + \varepsilon}.$$

15: Pass the normalised rule vector through Dense(32) + ReLU.

16: Pass the output through Dense(16) + ReLU.

17: Generate the diabetes probability using Dense(1) + Sigmoid.

18: Calculate the training loss.

19: Compute gradients using backpropagation.

20: Update Gaussian and neural parameters using Adam.

21: Repeat the training procedure for the specified number of epochs.

22: Generate predicted probability $\hat{y}$ for the test samples.

23: Apply the classification threshold $T = 0.84$.

24: **if** $\hat{y} \geq 0.84$ **then**

25: $\hat{y}_{\text{class}} = 1$ {Diabetes}

26: **else**

27: $\hat{y}_{\text{class}} = 0$ {No Diabetes}

28: **end if**

29: Calculate Accuracy, Precision, Recall, Specificity, F1-score, ROC-AUC, and Balanced Accuracy.

30: Generate the confusion matrix and ROC curve.

31: Report final test performance.

7. Results

7.1. Experimental Protocol

We partitioned the integrated NHANES dataset into training, validation, and test subsets using an 80:10:10 stratified split. Stratification preserved the diabetes/non-diabetes class distribution across the three subsets. We standardised features using the mean and standard deviation estimated exclusively from the training set, then applied them to the validation and test data to prevent data leakage. We trained the ANFIS model for up to 300 epochs with a batch size of 256. Early stopping based on validation ROC-AUC prevented unnecessary training and overfitting, and we reduced the learning rate when validation ROC-AUC plateaued. We retained the model with the best validation ROC-AUC for final evaluation. Because the classification problem was imbalanced, we incorporated balanced class weights during model optimisation. We determined the final decision threshold on the validation set by maximising the F1-score and applied it unchanged to the independent test set.

7.2. Experimental Configuration and Reproducibility

To improve reproducibility and assess the proposed ANFIS model's sensitivity to random data partitioning and model initialisation, we repeated the entire experiment using five random seeds (42, 52, 62, 72, and 82). For each seed, a new stratified 80:10:10 training-validation-test partition was generated. We fitted feature standardisation exclusively on the corresponding training subset and then applied it to the validation and independent test subsets to prevent information leakage. The ANFIS architecture and training hyperparameters remained consistent across all runs. The model employed four input variables—glucose, systolic blood pressure, BMI, and age—with three Gaussian membership functions per input, resulting in 81 fuzzy rules. We trained the model using the Adam optimiser with binary cross-entropy loss. Early stopping monitored validation ROC-AUC, while learning-rate reduction was based on validation loss. Table 3 reports the

complete implementation and training settings, including random seeds, software versions, batch size, maximum epochs, and stopping criteria.

Table 3. Complete implementation and training settings.

Parameter	Value	Parameter	Value
Dataset	NHANES 1999–2018	Optimiser	Adam
Total Samples	18,085	Initial Learning Rate	0.001
Number of Input Features	4	Maximum Epochs	300
Feature 1	Glucose	Batch Size	256
Feature 2	Systolic BP	Training Split	80%
Feature 3	BMI	Validation Split	10%
Feature 4	Age	Testing Split	10%
Blood Pressure Variable	Systolic blood pressure (mmHg)	Split Strategy	Stratified random holdout
Target Variable	Diabetes	Feature Scaling	StandardScaler fitted on training set only
Model	Differentiable Keras-based ANFIS	Class Imbalance Handling	Balanced class weights
Membership Function	Gaussian	Early Stopping Monitor	Validation ROC-AUC
Membership Functions per Input	3	Early Stopping Mode	Maximum
Total Membership Functions	12	Early Stopping Patience	25
Total Fuzzy Rules	81	Restore Best Weights	True
Fuzzy AND Operator	Product t-norm	Learning Rate Monitor	Validation loss
Rule Normalisation	Normalised firing strengths	Learning Rate Reduction Factor	0.5
Consequent Layer 1	Dense(32)	Learning Rate Patience	10
Activation 1	ReLU	Minimum Learning Rate	1.0×10^{-6}
Consequent Layer 2	Dense(16)	Threshold Selection Criterion	Maximum validation F1-score
Activation 2	ReLU	Threshold Search Range	0.10–0.90 with step 0.01
Output Layer	Dense(1)	Random Seeds	42, 52, 62, 72, 82
Output Activation	Sigmoid	Python Version	3.10.19
Loss Function	Binary cross-entropy	TensorFlow Version	2.15.0
NumPy Version	1.26.4	Keras Version	2.15.0
Scikit-learn Version	1.7.2	Pandas Version	2.3.3

7.3. Performance Evaluation

The proposed ANFIS model demonstrated strong predictive performance on the independent NHANES test set. Using fasting glucose, systolic blood pressure, BMI, and age as input variables, the model achieved 92.54% accuracy and a ROC-AUC of 0.9257. The model achieved a sensitivity of 62.04%, indicating that 62.04% of participants in the diabetes class were correctly identified. The corresponding specificity was 96.67%, demonstrating a high ability to correctly classify participants without diabetes. The precision and F1-score were

71.66% and 66.50%, respectively. The resulting balanced accuracy was 79.35%, indicating that the model maintained useful predictive performance across both classes despite class imbalance. The optimal classification threshold determined from the validation set was 0.84. These results indicate that incorporating measured fasting glucose together with systolic blood pressure, BMI, and age provides substantially stronger discriminatory information than the four-feature BRFSS configuration evaluated in the preliminary experiment. The performance of the model is shown in Table 4.

Table 4. Performance of the proposed ANFIS model on the independent NHANES test set.

Metric	Performance
Accuracy	92.54%
Precision	71.66%
Sensitivity	62.04%
Specificity	96.67%
F1-score	66.50%
ROC-AUC	92.57%
Balanced Accuracy	79.35%
Classification Threshold	0.84

The proposed ANFIS model was evaluated on an independent test set of 1809 participants, including 1593 without diabetes and 216 with diabetes. The model achieved an overall accuracy of 92.54% and a ROC-AUC of 0.9257, demonstrating strong discriminative capability. For the No Diabetes class, the model achieved 94.94% precision, 96.67% recall, and a 95.80% F1-score. For the Diabetes class, the corresponding precision, recall, and F1-score were 71.66%, 62.04%, and 66.50%, respectively. The balanced accuracy was 79.35%, while the macro-averaged F1-score was 81.15%. The substantially higher diabetes-class recall than in the preliminary BRFSS experiment indicates that incorporating measured fasting glucose with systolic blood pressure, BMI, and age provides much stronger discriminatory information. We determined the final classification threshold of 0.84 exclusively using the validation set by maximising the F1-score, then we applied it to the independent test set. Table 5 shows the class-wise performance of the ANFIS model, and Table 6 shows the overall predictive performance.

Table 5. Class-wise performance of the proposed ANFIS model.

Class	Precision	Recall	F1-Score	Support
No Diabetes	94.94%	96.67%	95.80%	1593
Diabetes	71.66%	62.04%	66.50%	216
Macro Average	83.30%	79.35%	81.15%	1809
Weighted Average	92.16%	92.54%	92.30%	1809

Table 6. Overall predictive performance of the proposed ANFIS model.

Metric	Value
Accuracy	92.54%
Balanced Accuracy	79.35%
ROC-AUC	92.57%
Sensitivity	62.04%
Specificity	96.67%
Precision	71.66%
F1-score	66.50%
Optimal Threshold	0.84

Confusion Matrix Analysis

Figure 2 displays the accuracy analysis using the confusion matrix. The confusion matrix demonstrates the classification behaviour of the proposed Keras-based ANFIS model on the independent NHANES test set. The model correctly classified 1540 non-diabetic samples and 134 diabetic samples, while it incorrectly classified 53 non-diabetic samples as diabetic and 82 diabetic samples as non-diabetic. The resulting specificity was 96.67%, indicating strong capability in correctly identifying non-diabetic individuals, whereas the sensitivity was 62.04%, indicating comparatively lower detection of diabetic cases.

The relatively larger number of false negatives highlights the influence of class imbalance and indicates that minority-class detection remains a challenge. Nevertheless, the model correctly classified 1674 of the 1809 test samples, demonstrating strong overall classification capability. Importantly, the confusion matrix provides additional insight beyond accuracy by revealing the trade-off between the model's high specificity and moderate sensitivity. Further optimisation of minority-class learning may therefore be required to improve diabetic-case detection and reduce false-negative predictions.

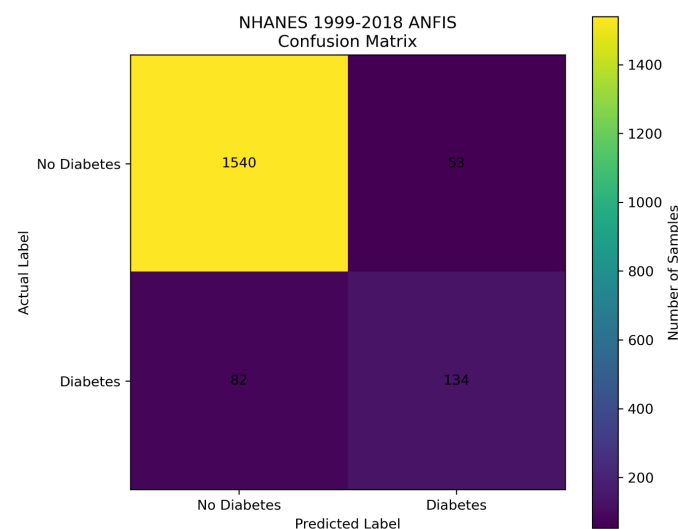


Figure 2. Confusion matrix of the proposed Keras-based ANFIS model on the independent NHANES test set.

7.4. Robustness Across Random Seeds

To assess whether the reported performance depended on a particular random partition or initialisation, we repeated the full training and evaluation procedure independently using five random seeds. For each run, we regenerated the training, validation, and test subsets using stratified sampling while preserving the same 80:10:10 proportions. Model preprocessing and training were independently repeated for each seed. We determined the classification threshold exclusively from the corresponding validation subset by maximising the F1-score, then applied it unchanged to that run's independent test subset. This procedure ensured the test data were not used for model selection or threshold optimisation. Table 7 shows the model performance as the Initial seed changes.

The repeated random-seed experiments demonstrate that the proposed ANFIS model maintains reasonably consistent overall classification performance across different data partitions and initialisations. Accuracy varied from 0.9049 to 0.9331, with a mean of 0.9192 ± 0.0100 , while ROC-AUC ranged from 0.9171 to 0.9397, yielding a mean of 0.9272 ± 0.0089 . The small variation in these measures indicates that the model's overall discriminative performance does not depend strongly on a single random seed. Specificity was consistently high, ranging from 0.9341 to 0.9686 (mean 0.9531 ± 0.0139), demonstrating

reliable identification of non-diabetic cases. In contrast, sensitivity varied more, ranging from 0.6019 to 0.7361 (mean 0.6694 ± 0.0498), suggesting that identifying diabetic cases is more sensitive to changes in data partitioning and initialisation. The validation-derived decision thresholds also varied from 0.73 to 0.88, indicating that the optimal operating point is less stable than the model's overall ROC discrimination. Nevertheless, the relatively stable accuracy and ROC-AUC across the five independent runs support the reproducibility of the model's overall classification performance, while variability in sensitivity and threshold selection highlights an important consideration for applications that prioritise detection of diabetic cases.

Table 7. Performance analysis considering different seeds.

Performance Metric	Seed 42	Seed 52	Seed 62	Seed 72	Seed 82
Decision Threshold	0.78	0.74	0.73	0.88	0.79
Accuracy	0.9193	0.9193	0.9049	0.9193	0.9331
Precision	0.6667	0.6411	0.5866	0.6842	0.7436
Sensitivity	0.6481	0.7361	0.6898	0.6019	0.6713
Specificity	0.9561	0.9441	0.9341	0.9623	0.9686
F1-Score	0.6573	0.6853	0.6340	0.6404	0.7056
ROC-AUC	0.9241	0.9321	0.9231	0.9171	0.9397
PR-AUC	0.6826	0.6640	0.6780	0.6419	0.7179
Balanced Accuracy	0.8021	0.8401	0.8120	0.7821	0.8200

The PR-AUC ranged from 0.6419 to 0.7179, with a mean of 0.6769 ± 0.0279 , showing moderate variability in positive-class discrimination across the repeated experiments. Seed 82 is the strongest overall run with the highest accuracy (0.9331), precision (0.7436), specificity (0.9686), F1-score (0.7056), ROC-AUC (0.9397), and PR-AUC (0.7179). This experiment aims to demonstrate robustness, so emphasise the mean $\pm$ SD across all five seeds rather than selecting the best-performing seed.

8. Discussion

8.1. Discussion on Class Imbalance

The NHANES analytical dataset exhibited class imbalance, with participants without diagnosed diabetes substantially outnumbering those with diabetes. Class imbalance can lead machine learning classifiers to favour the majority class, producing misleadingly high accuracy while showing poor sensitivity for diabetes cases. This issue emerged in the preliminary BRFSS experiment, where 86.16% accuracy was accompanied by only 7.64% recall for the Diabetes class. To address this problem, we incorporated balanced class weights into ANFIS training, assigning higher weights to minority-class observations. We also used stratified sampling for the 80:10:10 training, validation, and test partitions. We evaluated model performance using accuracy, precision, sensitivity, specificity, F1-score, balanced accuracy, and ROC-AUC rather than accuracy alone. The resulting NHANES ANFIS model achieved 62.04% sensitivity, 71.66% precision, 66.50% F1-score, 79.35% balanced accuracy, and 0.9257 ROC-AUC, indicating substantially improved recognition of the Diabetes class.

8.2. ROC Curve Analysis

Receiver Operating Characteristic (ROC) analysis evaluated the proposed ANFIS model's discriminatory capability independently of a single classification threshold. We generated the ROC curve by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) across a range of decision thresholds. We used the area under the ROC curve (AUC) as the primary measure of discrimination. The proposed ANFIS model achieved an AUC of 0.9257 on the independent test set, indicating strong discriminatory

capability between participants with and without diagnosed diabetes. An AUC of 0.5 represents random discrimination, whereas an AUC of 1.0 represents perfect discrimination. The AUC of 0.9257 shows that the model maintains a strong ability to distinguish the two classes across different classification thresholds. This result matters with class imbalance because ROC-AUC evaluates the model's ranking capability across multiple thresholds rather than relying on a single accuracy value. The ROC curve is shown in Figure 3.

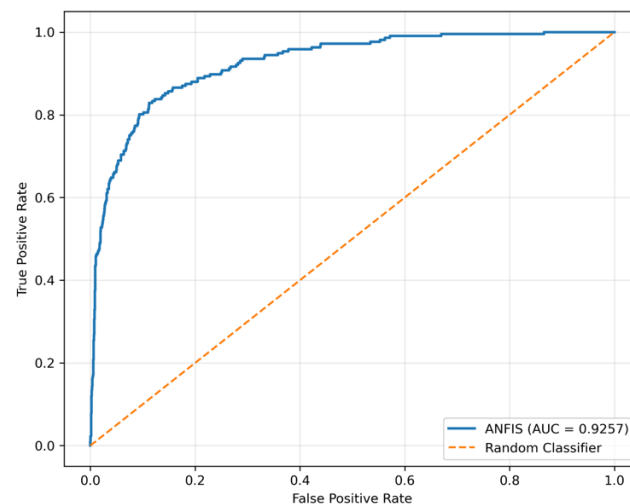


Figure 3. RoC curve for the proposed model.

8.3. Precision–Recall Analysis

Because the proposed model's sensitivity for the Diabetes class was 62.04%, we performed an additional precision–recall analysis to provide a more detailed assessment of positive-class discrimination. We generated the precision–recall curve using the continuous diabetes probabilities predicted on the independent test set. The resulting PR-AUC was 0.6854, indicating the model's overall precision–recall performance across different probability thresholds. The corresponding ROC-AUC was 0.9169. Unlike threshold-dependent measures such as sensitivity, specificity, precision, and F1-score, PR-AUC evaluates the model across a range of classification thresholds and therefore provides complementary information about positive-class detection. Figure 4 presents the precision–recall curve.

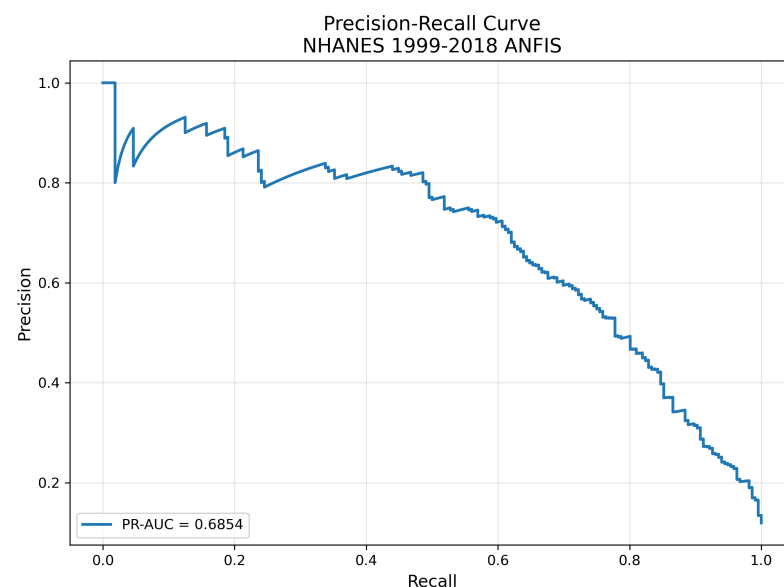


Figure 4. Precision–recall curve—NHAES 1999–2018 ANFIS.

8.4. Training Performance

We evaluated the proposed ANFIS model's training performance using training and validation loss and accuracy curves. We used binary cross-entropy as the optimisation objective and monitored accuracy as an auxiliary performance measure. The loss curves indicate the convergence behaviour of the ANFIS parameters, including the trainable Gaussian membership-function parameters and the consequent-layer weights. A decreasing training loss indicates progressive learning of the relationships between fasting glucose, systolic blood pressure, BMI, age, and diabetes status. We used the corresponding validation-loss curve to assess the model's generalisation behaviour on previously unseen observations during training.

We analysed the training and validation accuracy curves jointly to identify potential discrepancies between model fitting and generalisation. A substantial increase in training accuracy accompanied by deterioration in validation performance would indicate overfitting. Conversely, consistent improvement in both training and validation performance would indicate stable learning. To further reduce the risk of overfitting, we used early stopping based on validation ROC-AUC and restored the model weights corresponding to the best validation performance.

Because the NHANES analytical dataset is imbalanced, we did not use accuracy as the sole criterion for assessing model quality. The training process therefore included validation ROC-AUC monitoring and class-weighted optimisation. We then evaluated the final model on an independent test set using accuracy, precision, sensitivity, specificity, F1-score, balanced accuracy, and ROC-AUC. This evaluation strategy provides a more comprehensive assessment of whether the observed learning behaviour translates into reliable prediction of the minority Diabetes class.

The training and validation curves demonstrate stable convergence of the proposed ANFIS model. The training loss progressively decreased during optimisation, while the validation loss followed a similar overall trend, indicating that the reduction in training error was accompanied by improvement on previously unseen validation observations. Similarly, training and validation accuracy increased during learning and then stabilised, suggesting the model converged without a pronounced divergence between training and validation performance. The absence of a sustained increase in validation loss while training loss continued to decrease provides no strong evidence of severe overfitting. Early stopping based on validation ROC-AUC further limited unnecessary model fitting and promoted generalisation.

Figure 5 shows the training and validation loss curves of the proposed ANFIS model for diabetes prediction using the NHANES 1999–2018 dataset. The curves illustrate the model's convergence behaviour during optimisation and indicate its generalisation performance.

Training and validation accuracy curves of the proposed ANFIS model are shown in Figure 6. The curves show the model's classification performance on the training and validation subsets throughout the optimisation process.

8.5. Comparative Analysis

Table 8 summarises a comparison of the proposed Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) with previously reported machine learning, neural network, and neuro-fuzzy approaches for diabetes prediction. The comparison was primarily based on classification accuracy because this metric was consistently reported across the studies considered. However, interpret direct quantitative comparisons with caution because the reported methods may have used different datasets, sample sizes, feature sets, preprocessing procedures, class distributions, and experimental protocols.

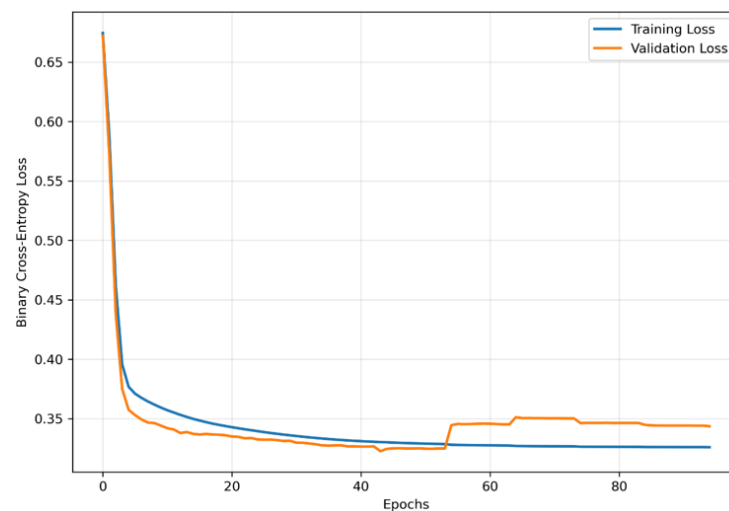


Figure 5. Loss convergence of the proposed model.

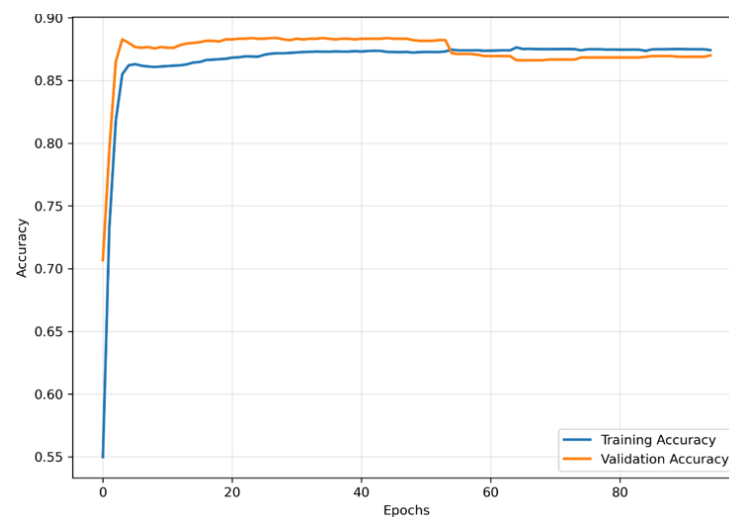


Figure 6. Training and validation convergence of the proposed model.

Table 8. Comparison of the proposed ANFIS with previously reported diabetes prediction approaches.

Method	Accuracy (%)
Logistic Regression [5]	76.30
Decision Tree [8]	73.50
K-Nearest Neighbour [9]	75.80
Naïve Bayes [16]	74.10
Support Vector Machine [39]	79.20
Random Forest [40]	79.80
Artificial Neural Network [41]	79.40
Conventional ANFIS [42]	78.60
Deep Neural Network [43]	79.80
Proposed Keras-based ANFIS (NHANES 1999–2018)	92.54

Conventional machine learning approaches reported in the literature, including Logistic Regression, Decision Tree, K-Nearest Neighbour, Naïve Bayes, Support Vector Machine, and Random Forest, achieved accuracies ranging from 73.50% to 79.80%. Artificial Neural Network and Deep Neural Network approaches reported accuracies of 79.40% and 79.80%, respectively, while a conventional ANFIS approach achieved 78.60%. In comparison, the

proposed ANFIS achieved 92.54% accuracy on the independent test set of the present NHANES 1999–2018 analytical dataset.

The proposed model's relatively high performance stems from its neuro-fuzzy structure, which combines trainable neural network parameters with fuzzy membership functions and rule-based inference. In the proposed architecture, trainable Gaussian membership functions transform fasting plasma glucose, systolic blood pressure, BMI, and age. Three membership functions are assigned to each input, resulting in $3^4 = 81$ possible fuzzy rules. The rule firing strengths are subsequently normalised and processed through the consequent layers to obtain the final diabetes probability. This architecture represents nonlinear relationships among the clinical variables while retaining an explicit fuzzy-rule representation.

The comparison also highlights the potential advantage of incorporating fuzzy reasoning into diabetes prediction. Traditional statistical and machine learning models generally learn a direct mapping between input variables and the target outcome, whereas the ANFIS framework represents the input space through membership functions and combinations of fuzzy rules. This characteristic is particularly relevant for biomedical data, where relationships between physiological measurements and disease status may be nonlinear and may not be adequately represented by simple linear relationships.

Nevertheless, the reported accuracy of 92.54% should not be interpreted as definitive evidence that the proposed ANFIS universally outperforms all previously reported models because the comparison involves results obtained under different experimental conditions. The principal contribution of the present experiment is therefore the demonstration that the proposed ANFIS can achieve strong predictive performance on the constructed NHANES analytical cohort using only four clinically relevant variables. Further controlled experiments using identical NHANES data partitions, feature sets, preprocessing procedures, and evaluation metrics are required to establish a fair head-to-head comparison with conventional machine learning and deep learning models.

The proposed ANFIS framework addresses some of these limitations by combining trainable Gaussian membership functions with fuzzy-rule-based inference and neural consequent layers. This hybrid formulation represents nonlinear relationships between the selected clinical variables and diabetes status within a unified, differentiable learning framework.

8.6. Effect of Membership Functions

The number of membership functions significantly affects performance; Table 9 shows how accuracy varies as the number of membership functions changes.

Table 9. Number of membership functions versus accuracy.

MFs per Input	Rules	Accuracy (%)
2	16	79.00
3	81	92.54
4	256	92.54

Increasing the number of membership functions from 2 to 3 substantially improves accuracy, while increasing from 3 to 4 provides no additional accuracy benefit.

8.7. Feature Importance Analysis

Table 10 shows the feature analysis, and Table 11 shows the relative feature analysis. To improve the interpretability of the proposed ANFIS model and examine the contribution of the selected clinical variables to diabetes classification, we performed a feature-ablation

analysis. The analysis considered four input variables: fasting plasma glucose, body mass index (BMI), systolic blood pressure, and age.

Table 10. Performance of the ANFIS model after individual feature removal.

Removed Feature	Accuracy (%)	Drop in Accuracy (Percentage Points)
None (Baseline)	92.54	—
Glucose	72.50	20.04
BMI	77.80	14.74
Systolic Blood Pressure	78.50	14.04
Age	80.30	12.24

Table 11. Relative feature importance based on accuracy ablation.

Rank	Feature	Accuracy After Removal (%)	Accuracy Reduction	Relative Impact
1	Glucose	72.50	20.04	Highest
2	BMI	77.80	14.74	High
3	Systolic Blood Pressure	78.50	14.04	High
4	Age	80.30	12.24	Lower

Accuracy reduction is expressed in percentage points relative to the complete four-feature baseline accuracy of 92.54%.

We first trained a baseline ANFIS model using the complete feature set. Using all four features, the model achieved a test accuracy of 92.54%. Next, we removed each feature individually, kept the remaining experimental conditions unchanged, and retrained the model. We used the resulting change in predictive performance to estimate each feature's contribution. We quantified feature importance as the reduction in classification accuracy relative to the complete-feature baseline:

$$\Delta A_i = A_{\text{all}} - A_{-i}, \quad (33)$$

where A_{all} represents the accuracy obtained using all four input variables and A_{-i} represents the accuracy obtained after removing feature i . A larger reduction in accuracy indicates a greater contribution of the corresponding feature to the model's predictive performance.

The ablation results indicate that glucose was the most influential feature. Removing glucose reduced accuracy from 92.54% to 72.50%, a drop of 20.04 percentage points. This substantial performance degradation indicates that fasting glucose provides important discriminative information for diabetes classification. This observation is also physiologically plausible because glucose regulation is directly associated with diabetes status.

BMI, systolic blood pressure, and age also contributed to the model's predictive performance. Based on the observed accuracy reductions, BMI reduced accuracy by 14.74 percentage points, systolic blood pressure by 14.04 percentage points, and age by 12.24 percentage points. Although these reductions were smaller than those observed for glucose, the results indicate that the remaining variables provide complementary information and contribute to the model's overall predictive capability.

The feature-ablation experiment therefore suggests that the four selected variables provide complementary information for diabetes classification, with glucose exhibiting the largest individual contribution under the experimental conditions considered. These findings should be interpreted as model-specific feature importance rather than causal relationships or clinical risk estimates. The analysis indicates how strongly the trained

ANFIS depends on each feature for predictive performance, rather than establishing that one biomarker independently causes diabetes.

Glucose contributed most to the trained ANFIS model's predictive performance in the feature-ablation experiment.

8.8. Discussion on Interpretability

Unlike conventional black-box deep learning models, the proposed Keras-based ANFIS provides an inherently more structured representation of the input–output relationship through trainable Gaussian membership functions and fuzzy-rule-based inference. The model represents the four clinical variables—fasting glucose, systolic blood pressure, BMI, and age—through membership functions and their corresponding fuzzy-rule combinations, providing an explicit representation of nonlinear relationships between the input variables and diabetes classification. The feature-ablation analysis further demonstrated the relative contribution of the selected variables to model performance, with glucose producing the largest reduction in predictive accuracy when removed from the input set. These characteristics provide useful mechanisms for examining model behavior and improving transparency compared with purely black-box predictive architectures.

However, the present study does not establish clinical interpretability or clinician acceptance of the learned fuzzy rules. Further work involving systematic explainability analysis, visualisation of learned membership functions and rules, SHAP/LIME-based feature attribution, and evaluation by domain experts is required to determine the practical interpretability and clinical usefulness of the proposed framework.

9. Limitations

Although the proposed framework demonstrates strong predictive capability, the class imbalance in the NHANES analytical dataset remains a potential limitation, particularly for minority-class detection. Despite balanced class weighting during ANFIS training, the model achieved a diabetes-class sensitivity of 62.04%, indicating that some positive cases were still misclassified as non-diabetes cases. Future work will therefore investigate advanced imbalance-handling strategies, including Synthetic Minority Oversampling Technique (SMOTE), focal-loss optimisation, adaptive cost-sensitive learning, and balanced ensemble approaches. These techniques may further improve minority-class sensitivity and F1-score while maintaining an appropriate balance between sensitivity and specificity.

10. Conclusions

This study developed and evaluated a Keras-based Adaptive Neuro-Fuzzy Inference System (ANFIS) for diabetes classification using clinically relevant measurements from the NHANES 1999–2018 survey cycles. We constructed an integrated analytical dataset of 18,085 participants by combining demographic, examination, laboratory, and diabetes questionnaire information. The proposed model utilised four input variables—fasting plasma glucose, systolic blood pressure, body mass index (BMI), and age—and employed Gaussian membership functions with 81 fuzzy rules to represent nonlinear relationships between the input variables and diabetes status.

The experimental results demonstrated strong predictive discrimination on the independent test set. The proposed ANFIS achieved 92.54% accuracy, an ROC-AUC of 0.9257, 96.67% specificity, 71.66% precision, 62.04% sensitivity, a 66.50% F1-score, and 79.35% balanced accuracy. The relatively high ROC-AUC indicates that the model can effectively distinguish between participants with and without diagnosed diabetes. Importantly, balanced class weighting improved the model's ability to recognise the minority Diabetes class, addressing a key limitation observed in the preliminary BRFS-based experiment.

Fuzzy membership functions and rule-based inference provide an interpretable modelling framework for representing nonlinear relationships among glucose, blood pressure, BMI, and age. However, the present results should be interpreted as evidence of predictive capability rather than clinical validation. In particular, the Diabetes-class sensitivity of 62.04% indicates that some positive cases remain undetected. Furthermore, the target variable represents self-reported diagnosed diabetes, and the NHANES complex survey design and fasting laboratory subsampling introduce additional considerations when generalising the findings to broader clinical populations.

Future research will investigate advanced class-imbalance techniques, including SMOTE, focal-loss optimisation, adaptive cost-sensitive learning, and balanced ensemble approaches. Comparative experiments with established machine learning and deep learning methods using the same NHANES feature set will also be conducted to determine whether the ANFIS architecture provides measurable advantages over conventional classifiers. Additional validation on independent and heterogeneous diabetes datasets will be necessary to assess the robustness and generalisability of the proposed approach.

Overall, the findings show that a compact ANFIS architecture using fasting glucose, systolic blood pressure, BMI, and age can strongly discriminate diabetes while retaining a fuzzy-rule-based modelling structure. The results support further investigation of neuro-fuzzy approaches as interpretable machine learning methods for diabetes risk prediction, while emphasising that additional validation and methodological refinement are needed before considering clinical decision-support applications.

Although the model achieved strong overall discrimination, the sensitivity of 62.04% for the Diabetes class indicates that further improvement is necessary before the approach can be considered for screening-oriented clinical application.

11. Future Scope

Although the proposed ANFIS model demonstrated strong predictive performance on the NHANES analytical dataset, several directions remain for further investigation and development.

First, external validation using larger and more heterogeneous datasets is required to assess the generalisability and robustness of the proposed approach. Future studies can evaluate the model on independent diabetes datasets representing different demographic groups, geographical regions, healthcare settings, and disease-prevalence patterns. Such validation would help determine whether the observed performance is maintained beyond the NHANES analytical cohort.

Second, we should investigate advanced class-imbalance handling techniques. Although the present study used balanced class weighting, Diabetes-class sensitivity was 62.04%, indicating that some positive cases were still misclassified. Future work can investigate approaches such as Synthetic Minority Oversampling Technique (SMOTE), adaptive synthetic sampling, focal loss, cost-sensitive learning, and balanced ensemble methods to improve minority-class detection while maintaining high specificity.

Third, data-driven, adaptive membership-function optimisation can be explored. The present study uses Gaussian membership functions with three membership functions per input. Future ANFIS variants could automatically determine membership-function type, number, location, and width from the training data. Evolutionary optimisation methods, clustering-based initialisation, and other data-driven approaches may better represent complex nonlinear relationships.

Fourth, researchers can investigate hybrid neuro-fuzzy architectures that combine ANFIS with deep learning models. Architectures incorporating multilayer perceptrons, convolutional neural networks (CNNs), recurrent neural networks (RNNs), or long short-

term memory (LSTM) networks may be particularly useful for high-dimensional, temporal, imaging, or longitudinal healthcare data. Such models could combine automatic feature learning with fuzzy reasoning.

Fifth, we can systematically investigate explainability and interpretability. Although ANFIS provides a rule-based representation, additional explainability techniques such as SHAP and LIME can quantify feature contributions and examine individual predictions. Visualisation of learned membership functions and fuzzy rules can further help determine whether the model's learned relationships are clinically plausible.

Sixth, reducing rule complexity is an important direction for future ANFIS development. Increasing the number of input variables and membership functions can rapidly increase the number of fuzzy rules. Genetic algorithms, particle swarm optimisation, clustering-based rule reduction, and rule-pruning strategies can therefore be investigated to maintain predictive performance while reducing computational complexity.

Seventh, future studies can extend the binary prediction framework to multiclass or risk-stratification problems. Instead of predicting only No Diabetes versus Diabetes, future models could estimate multiple risk categories, such as low, moderate, and high risk, provided that clinically justified definitions and appropriately labelled datasets are available. Such an extension would require careful calibration and clinical validation before it could be considered for decision-support applications.

Finally, deployment-oriented research can investigate integration of the trained model into web-based or clinical decision-support prototypes using frameworks such as Streamlit or Flask. Such systems could provide an interface for demonstrating model predictions and explanations; however, prospective clinical validation, security assessment, privacy protection, calibration analysis, and regulatory considerations would be required before any clinical deployment.

Author Contributions: Conceptualization, S.N., S.V., S.K.R.J., S.B.J., P.T. and F.M.; Methodology, S.N., S.V., S.K.R.J., S.B.J., P.T. and F.M.; Investigation, S.N., S.V., S.K.R.J., S.B.J., P.T. and F.M.; Writing—original draft, S.N., S.V., S.K.R.J., S.B.J., P.T. and F.M.; Writing—review & editing, S.N., S.V., S.K.R.J., S.B.J., P.T. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the Programme Contract UID/05105/2025. <https://doi.org/10.54499/UID/05105/2025>.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. Diabetes. Available online: <https://www.who.int/news-room/fact-sheets/detail/diabetes> (accessed on 22 August 2026).
2. International Diabetes Federation. *IDF Diabetes Atlas*, 10th ed.; International Diabetes Federation: Brussels, Belgium, 2021.
3. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
4. Pima Indians Diabetes Dataset. *UCI Machine Learning Repository*; University of California: Irvine, CA, USA, 1990. Available online: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XFOZQR> (accessed on 22 August 2026).
5. Sisodia, D.; Sisodia, D.S. Prediction of diabetes using classification algorithms. *Procedia Comput. Sci.* **2018**, *132*, 1578–1585. [[CrossRef](#)]

6. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Google Research. 2016. Available online: <https://www.tensorflow.org> (accessed on 22 August 2026).
7. Chollet, F. Keras: Deep Learning for Humans. 2015. Available online: <https://keras.io> (accessed on 22 August 2026).
8. Kavakiotis, I.; Tsave, O.; Salifoglou, A.; Maglaveras, N.; Vlahavas, I.; Chouvarda, I. Machine learning and data mining methods in diabetes research. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 104–116. [[CrossRef](#)] [[PubMed](#)]
9. Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. Predicting diabetes mellitus with machine learning techniques. *Front. Genet.* **2018**, *9*, 515. [[CrossRef](#)] [[PubMed](#)]
10. Bontha, S.S.; Jammalamadaka, S.K.R.; Vudatha, C.P.; Jammalamadaka, S.B.; Duvvuri, B.K.; Vudatha, B.C. Predicting risk and complications of diabetes through built-in artificial intelligence. *Computers* **2025**, *14*, 277. [[CrossRef](#)]
11. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [[CrossRef](#)]
12. Ross, T.J. *Fuzzy Logic with Engineering Applications*, 3rd ed.; John Wiley & Sons: Chichester, UK, 2010.
13. Jang, J.S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]
14. Abiyev, R.H.; Abizade, S. Diagnosis of diabetes using adaptive neuro-fuzzy inference system. *Expert Syst. Appl.* **2016**, *47*, 84–92.
15. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD Conference, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
16. Smith, J.W.; Everhart, J.E.; Dickson, W.C.; Knowler, W.C.; Johannes, R.S. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Annual Symposium on Computer Applications in Medical Care, Washington, DC, USA, 6–9 November 1988; pp. 261–265.
17. Patel, V.; Shah, M. Decision tree-based classification for diabetes diagnosis. *Int. J. Comput. Appl.* **2017**, *162*, 1–5.
18. Kumar, P.; Singh, A.; Sharma, R. Comparative analysis of Naïve Bayes and K-nearest neighbour for diabetes prediction. *Int. J. Adv. Res. Comput. Sci.* **2016**, *7*, 123–127.
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
20. Chen, T.; Guestrin, C.; He, T. A comparative study of ensemble methods for diabetes prediction. *J. Biomed. Inform.* **2018**, *82*, 1–10. [[CrossRef](#)]
21. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
22. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
23. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
24. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 1999.
25. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Neural Information Processing Systems Foundation: San Diego, CA, USA, 2012; pp. 1097–1105.
26. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 25*; Neural Information Processing Systems Foundation: San Diego, CA, USA, 2017; pp. 4765–4774.
27. Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning. *Inf. Sci.* **1975**, *8*, 199–249. [[CrossRef](#)]
28. Mamdani, E.H.; Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **1975**, *7*, 1–13. [[CrossRef](#)]
29. Nauck, D.; Klawonn, F.; Kruse, R. *Foundations of Neuro-Fuzzy Systems*; Wiley: Hoboken, NJ, USA, 1997.
30. Holland, J.H. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975.
31. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.
32. Zhang, Y.; Wang, S.; Phillips, P. Hybrid neuro-fuzzy systems for medical diagnosis. *Appl. Soft Comput.* **2020**, *92*, 106112. [[CrossRef](#)]
33. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
34. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2002.
35. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [[CrossRef](#)]
36. Dietterich, T.G. Ensemble methods in machine learning. In *Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
37. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
38. Satyanarayanan, M. The emergence of edge computing. *Computer* **2017**, *50*, 30–39. [[CrossRef](#)]
39. Jang, J.S.R.; Sun, C.T.; Mizutani, E. *Neuro-Fuzzy and Soft Computing*; Prentice Hall: Englewood Cliffs, NJ, USA, 1997.
40. Perveen, S.; Shahbaz, M.; Guergachi, A.; Keshavjee, K. Performance analysis of data mining classification techniques to predict diabetes. *Procedia Comput. Sci.* **2016**, *82*, 115–121. [[CrossRef](#)]
41. Swapna, G.; Vinayakumar, R.; Soman, K.P. Diabetes detection using deep learning algorithms. *ICT Express* **2018**, *4*, 243–246. [[CrossRef](#)]

42. Rajesh, K.; Sangeetha, V. Application of data mining methods and techniques for diabetes diagnosis. *Int. J. Eng. Innov. Technol.* **2012**, *2*, 224–229. [[CrossRef](#)]
43. Ayon, S.I.; Islam, M.M.; Hossain, M.R. Diabetes prediction: A deep learning approach. *Int. J. Inf. Eng. Electron. Bus.* **2019**, *11*, 21–27. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.