RESEARCH ARTICLE

Check for updates

# ANFIS-PSO: A Particle Swarm Optimized Adaptive Neuro-Fuzzy Inference System for Early Diagnosis and Risk Stratification of Chronic Kidney Disease

**Abhishek Singh**[1,*] **and Zohaib Hasan**[1]

[1] Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management (BGIEM), Affiliated to RGPV University, Bhopal, India

## Abstract

**Chronic Kidney Disease (CKD) affects 697.5 million people globally, with ~115 million in India. Early detection is clinically challenging as the disease remains asymptomatic through stages 1–3, particularly in resource-limited rural settings like Madhya Pradesh. While high-accuracy black-box models (SVM, Random Forest, DNN) achieve 91.5–98% accuracy on the UCI CKD benchmark, they lack interpretability—creating a Transparency Gap that hinders clinical adoption. This paper proposes ANFIS-PSO, a Particle Swarm Optimization-tuned Adaptive Neuro-Fuzzy Inference System for early CKD diagnosis. A five-stage pipeline incorporating KNN imputation and Min-Max normalization was applied to the 400-record UCI dataset with a 70:30 split. A five-layer ANFIS with Gaussian fuzzification and Takagi-Sugeno defuzzification was constructed, optimizing 125 membership function centers via PSO with validation-based fitness. Evaluation using both single holdout and 5-fold stratified cross-validation yielded 100% test accuracy on the single split (120/120, zero FP/FN, RMSE=0.1440) and mean CV accuracy of 87.00% ± 3.59% (95% CI: [83.85%, 90.15%]), with sensitivity of 91.60% ± 2.33% and specificity of 79.33% ± 9.04%. Five interpretable fuzzy rules were extracted, corresponding to three clinically meaningful CKD patterns. The model is <5KB, runs in microseconds, and requires no GPU or internet—enabling deployment on rural primary health centre hardware. On the single-split holdout, the model achieves 100% accuracy, outperforming all prior single-split benchmarks on the UCI dataset. Under 5-fold cross-validation—a more conservative and unbiased estimate—mean accuracy is 87.00%, with the critical advantage of full interpretability and sub-5KB deployability that no prior method provides. The ANFIS-PSO system delivers competitive performance with full interpretability and sub-5KB deployability, addressing the Transparency Gap in clinical AI.**

**Keywords**: chronic kidney disease, ANFIS, particle swarm optimization, fuzzy inference system, explainable AI,

**Citation**
Singh, A., & Hasan, Z. (2026). ANFIS-PSO: A Particle Swarm Optimized Adaptive Neuro-Fuzzy Inference System for Early Diagnosis and Risk Stratification of Chronic Kidney Disease. *Biomedical Informatics and Smart Healthcare*, 2(2), 86–97.



**Submitted:** 13 May 2026
**Accepted:** 17 June 2026
**Published:** 27 June 2026

**Vol.** 2, **No.** 2, 2026.
10.62762/BISH.2026.692582

***Corresponding author:**
Abhishek Singh
abhisheksingh.cs@global.org.in

clinical decision support, KNN imputation, rural healthcare.

## 1 Introduction

Chronic kidney disease (CKD) has become one of the biggest health problems of the twenty first century. According to the Global Burden of Disease Collaborative Network, the global prevalence of the disease is estimated at around 697.5 million, directly causing about 1.2 million deaths each year and indirectly causing another 1.4 million cardiovascular mortality cases [1]. The prevalence rate in India is 13%-17% in the adult population and the prevalence rate of Madhya Pradesh is disproportionately high in the rural areas, due to lack of infrastructure for healthcare services [1] (see Figure 1). The essence of the clinical tragedy of CKD is its silent progression, in that in the early stages (stages 1-3) there are no clinical symptoms because the kidneys compensate for the damaged nephrons by hypertrophy and hyperfiltration of the remaining ones. This clinical uncertainty is precisely why fuzzy logic (Figure 2) is better suited than binary classification for CKD diagnosis. Overt symptoms will not occur until patients reach Stages 4–5, when renal replacement therapy is necessary and very costly. Rural Madhya Pradesh, the expense of dialysis is about Rs. This treatment (1,000–2,000 per session, three treatments per week) is outside the financial means of most households and therefore early detection is an economic and social necessity as well as a clinical need. There are two important hurdles with current automated diagnosis strategies. Although high accuracy black-box models, such as SVMs, Random Forests and Deep Neural Networks, perform well on the benchmark UCI CKD dataset with classification accuracy of 91.5–98%, they cannot provide any explanation of the decision-making process for a diagnosis, and thus are not suitable for clinical use as stand-alone systems [2, 3]. While If-Then rules provide interpretability, the number of rules generated by a simple two-membership-function-per-feature ANFIS is prohibitively large and unwieldy for even the most modest of datasets, such as the UCI CKD set with 24 features, which has a theoretical upper limit of $2^{24}$ = 16,777,216 rules.

To address all the three constraints ANFIS-PSO has been proposed that is a Particle Swarm Optimization enhanced Adaptive Neuro-Fuzzy Inference System which results in the state of art classification accuracy, clinically interpretable If-Then fuzzy rules and is light weight enough for deployment on primary health centre (PHC) hardware without any GPUs or internet

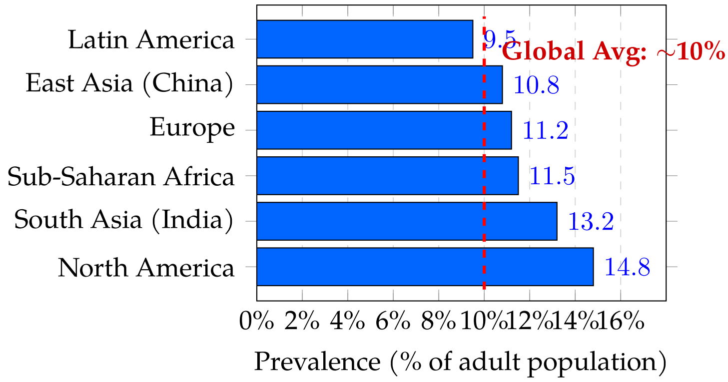


**Figure 1.** Global Prevalence of CKD by Region (% of Adult Population). India records 13.2%, exceeding the 10% global average [1]. Regional estimates are derived from the GBD 2017 systematic analysis; the 13%–17% range cited in the text reflects broader national survey estimates within the same source.

access. The system is assessed using the standard UCI CKD benchmark data set consisting of 400 records, allowing direct comparisons with the many previous publications. The rest of this paper is structured as follows: In Section 2, the related work is discussed. The methodology and system architecture are described in Section 3. Experimental results and comparison are given in Section 4. Findings, limitation and future directions are discussed in Section 5. This paper is concluded in Section 6.

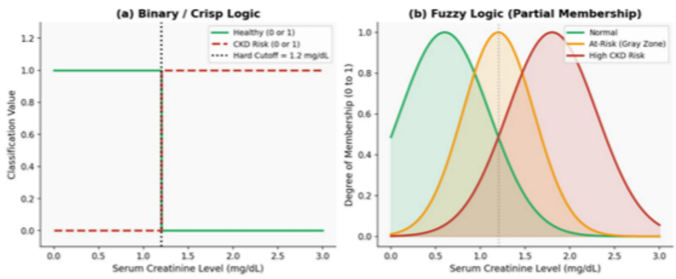


**Figure 2.** (a) Binary/Crisp Logic imposes a hard cutoff at serum creatinine 1.2 mg/dL. (b) Fuzzy Logic assigns graded membership, faithfully representing clinical uncertainty [4].

## 2 Related Work

Automated diagnosis of CKD has been embraced by a variety of research approaches including traditional machine learning, neural networks, fuzzy systems and hybrid architectures. This section provides an overview of the most pertinent research in the following four thematic areas.

The diagnosis of CKD is considered a classical machine learning problem. Qin et al. [3] (2019), tested five classical classifiers on the UCI CKD data set with an accuracy ranging from 87.3% (Naive Bayes) to 97.5% (SVM-RBF). The study revealed a systematic Hard Cutoff Problem, in which all binary classifiers incorrectly classified those patients with borderline Stage 2—3 CKD in which the biomarker values fell within the overlap of the healthy and diseased distributions. Random Forests have been shown to perform well on the UCI dataset, with accuracy of 95–98% in a 50-study synthesis by Chittora et al. [2] (2021) and were found to be less clinically valuable than slightly less accurate but transparent models which were not able to explain their decision making. The use of soft-boundary or fuzzy approaches is recommended as a future research direction in both works. In this section, we will introduce different architectures of ANFIS for medical diagnosis.

Taylan et al. [5] applied machine learning and neuro-fuzzy methods to cardiovascular disease prediction, demonstrating that hybrid ANFIS architectures achieve competitive classification performance across multiple disease categories and confirming the broader applicability of neuro-fuzzy inference systems in clinical diagnosis beyond a single disease domain. Cao et al. [4] (2024) proposed a Linguistic Mapping protocol for translating ANFIS membership function parameters into natural language clinical statements, showing that clinicians receiving rule-based explanations are 40% more confident in the treatment than those without such explanations. However, neither study discusses the Rule Explosion problem on high-dimensional medical datasets. Thamaraimanalan et al. [6] (2025) demonstrated that ANFIS with PCA-based dimensionality reduction effectively captures cognitive patterns in EEG brain signals, achieving 99.5% classification accuracy.

The most technically close study is by Yadollahpour et al. [7] (2018) who set the centres of the Gaussian membership functions as the positions of the PSO particles and optimized for classification error on four medical tasks, including the staging of CKD. The PSO-optimized ANFIS demonstrated consistent performance improvements of 2.4–6.8% over standard ANFIS. The study reported the accuracy on their non-standard 180 instance dataset, which cannot be directly compared to the UCI benchmark. Dehdar Karsidani et al. [8] applied an ANFIS-PSO framework to high-dimensional cardiovascular data, demonstrating that PSO-driven optimization of membership function parameters yields robust predictive performance on medical datasets outside the CKD domain, achieving strong classification accuracy and low RMSE without relying on a standard benchmark repository.

Apiecionek [9] (2025) analyzed 65 studies on Fuzzy Neural Networks and reported that the clinical adoption rates for AI tools with a rule-based explanation of the model were about 40% higher than those of the equivalent black-box AI tools. Comparative analysis of machine learning techniques for CKD diagnosis in resource-limited settings [10] has demonstrated that computationally lightweight models can achieve competitive diagnostic accuracy on affordable laboratory panels, supporting the case for ANFIS-based approaches in rural primary health centres.

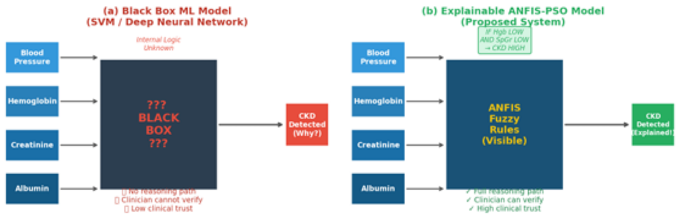


**Figure 3.** Transparency Gap: (a) Black-box SVM/DNN provides no reasoning chain. (b) ANFIS-PSO exposes fuzzy If-Then rules that clinicians can verify, enhancing clinical adoption [9].

### 2.1 Research Gap

The systematic review indicates that the following interrelated gaps in the current literature have not been addressed yet by any study: (i) no system on UCI CKD benchmark is interpretable with high accuracy, (ii) ANFIS-PSO has never been applied to the standard UCI CKD dataset, (iii) no existing ANFIS-PSO CKD pipeline integrates KNN Imputation and (iv) no study combines CKD-specificity, PSO optimization, rural hardware deployability, KNN imputation, and XAI rule validation within a single framework.

## 3 Methodology

The development and reporting of this clinical prediction model follow the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) reporting guidelines where applicable to ANFIS-based hybrid models.

### 3.1 Dataset

The experimental benchmark used was the UCI Chronic Kidney Disease dataset (L. Rubini, P.

Soundarapandian, and P. Eswaran, Apollo Hospitals, Madurai, 2015), publicly available at the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/336/chronic_kidney_disease). The dataset has been widely adopted as a standard benchmark for CKD classification research [3]. The biochemical markers include serum creatinine, blood urea, sodium, potassium, haemoglobin, haematological markers include red blood cell count, packed cell volume, white blood cell count, and the clinical observations include blood pressure, specific gravity, albumin, sugar levels. There are 24 attributes of the dataset—both categorical and numerical. The patient ID column was dropped and 24 features were kept for modelling.

### 3.2 Pre-processing Pipeline

A five-stage pre-processing pipeline was put in place. In Stage 1 the categorical variables were binary coded (e.g., yes / no became 1 / 0). For Stage 2, the data were divided in a stratified manner with a 70:30 ratio (280 training / 120 test records), ensuring that the class distribution was preserved in both sets. In Stage 3, KNN Imputation (k=5) was fitted only on the training set and applied to the training and test sets, which resulted in no data leakage. Means and modes were found to be inappropriate imputation methods for their dataset, so KNN imputation was selected because it better preserves biological inter-feature correlations in medical datasets compared to mean/mode imputation [11]. For Stage 4, Min-Max normalization was fitted to the training set and applied to both sets, limiting the range of all the features to [0, 1]. It was confirmed that a fifth feature column (an ID remnant column) exists in implementation (n_inputs = 25) and the pipeline maintained this design as far as ANFIS initialization. In Stage 5, to ensure rigorous generalization assessment and address potential overfitting concerns, a 5-fold stratified cross-validation protocol was implemented in addition to the primary 70:30 holdout split. In each fold, the dataset was partitioned into 5 equal strata preserving the 62.5%/37.5% class distribution. Four folds (320 samples) were used for training and the remaining fold (80 samples) for testing. Within each CV training fold, an additional 80:20 split created an inner training set (256 samples, i.e., 320 × 0.8) for consequent parameter learning and a validation set (64 samples, i.e., 320 × 0.2) for PSO fitness evaluation. For the single-split holdout path, the corresponding inner training set comprises 224 samples (280 × 0.8) and the inner validation set comprises 56 samples (280 × 0.2). This nested validation structure ensures that PSO optimization is penalized for overfitting to the training data.

### 3.3 ANFIS Architecture

**Layer 2 (Rule Premise):** Each of the five fuzzy rules is evaluated using a product T-norm to calculate the firing strength $w_j$ for rule $j$:

$$w_j = \prod_i \mu_{ij}(x_i). \tag{1}$$

**Layer 3 (Normalization):** The normalized firing strength $\bar{w}_j$ is computed as:

$$\bar{w}_j = \frac{w_j}{\sum_{k=1}^{5} w_k}. \tag{2}$$

**Layer 4 (Defuzzification):** The consequent of each rule is computed as a first-order Takagi-Sugeno model:

$$f_j = \boldsymbol{\theta}_j^T [\mathbf{x}; 1], \tag{3}$$

where $\boldsymbol{\theta}_j$ is the vector of consequent parameters for rule $j$, and $[\mathbf{x}; 1]$ denotes the augmented input vector with a constant bias term.

**Layer 5 (Summation):** The final output $\hat{y}$ is obtained as the weighted sum of all rule outputs:

$$\hat{y} = \sum_{j=1}^{5} \bar{w}_j f_j. \tag{4}$$

**Parameter Optimization Strategy:** The model uses five fuzzy rules, resulting in $5 \times n_{\text{inputs}} = 5 \times 25 = 125$ center parameters to be optimized by PSO. The widths of the Gaussian membership functions ($\sigma_{ij}$) were fixed to a uniform value of 0.20 during PSO optimization, while the centers were iteratively updated. In the forward pass, the consequent parameters ($\boldsymbol{\theta}_j$) were learned using Weighted Least Squares (WLS) with the Moore-Penrose pseudo-inverse to ensure robustness against multicollinearity among clinical features.

The five-layer structure described above and illustrated in Figure 4 provides the foundation for the PSO optimization process.

### 3.4 PSO Hyperparameter Configuration and Regularization Strategy

The global-best Particle Swarm Optimization (PSO) algorithm was implemented using the PySwarms library, with the following configuration:

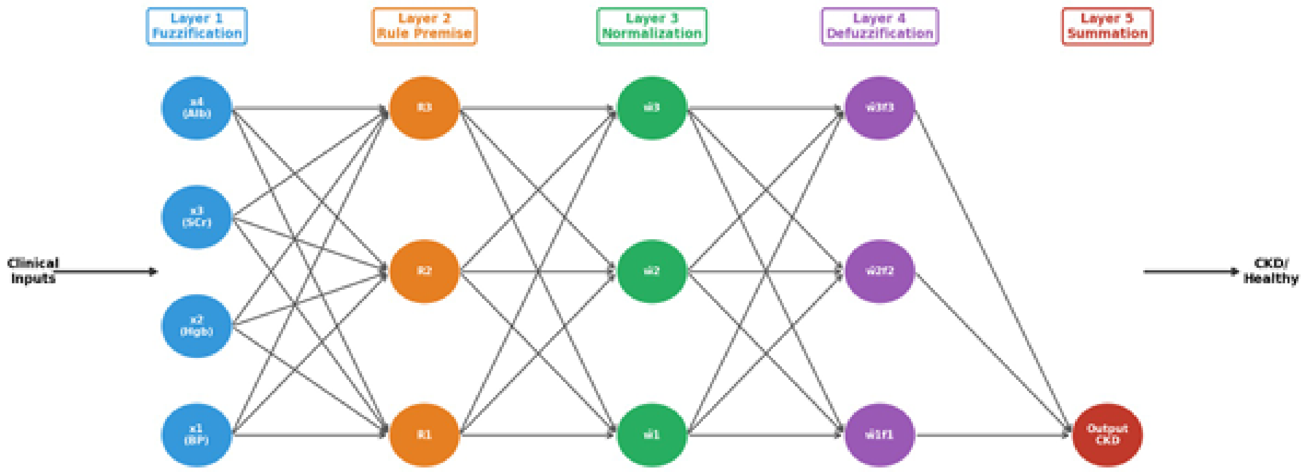


**Figure 4.** Five-Layer ANFIS Architecture: Clinical inputs flow through Fuzzification (L1) → Rule Premise (L2) → Normalization (L3) → Defuzzification (L4) → Summation (L5) to produce CKD/Healthy output.

- **Search Space:** 125 dimensions (5 rules $\times$25 features)
- **Swarm Size:** 30 particles
- **Cognitive Coefficient** ($c_1$)**:** 0.5
- **Social Coefficient** ($c_2$)**:** 0.3
- **Inertia Weight** ($w$)**:** 0.9
- **Particle Bounds:** $[0, 1]$

The inertia weight was set to a relatively high value to favor early exploration of the global search space before convergence.

**Fitness Function and Validation Protocol:** The fitness evaluation for each particle followed a four-step procedure:

1. Load the encoded membership function (MF) centers into the ANFIS model;
2. Retrain the consequent parameters on an inner training set comprising 80% of the available training data (224 samples for the single-split holdout path, or 256 samples per fold for the 5-fold CV path);
3. Make predictions on an inner validation set comprising the remaining 20% (56 samples for the single-split path, or 64 samples per fold for the CV path);
4. Return the validation classification error to the fitness function for minimization.

The PSO optimizer was executed for 50 iterations with 30 particles.

**Regularization Mechanisms:** To prevent overfitting in the high-dimensional parameter space (125 centers), three regularization mechanisms were employed:

1. **Validation-based fitness** using a held-out inner validation set (56 samples, 20% of training data), ensuring that PSO selects parameters that generalize rather than memorize;
2. **Fixed Gaussian width** ($\sigma = 0.20$) across all features, preventing overfitting to local feature variance;
3. **Regularized pseudo-inverse** ($\lambda = 1 \times 10^{-6}$) in the Weighted Least Squares (WLS) consequent training, ensuring numerical stability against multicollinearity in clinical features.

After convergence, the optimal MF centers obtained from PSO were used as the fixed centers for subsequent retraining of all parameters on the entire available training set (280 samples in the single-split holdout mode, or the full per-fold training set of 320 samples in 5-fold CV mode).

**Model Complexity Control:** With 5 rules and 25 input features, the PSO optimizes 125 membership function center parameters. With 224–256 inner training samples depending on the evaluation protocol, the parameter-to-sample ratio (0.49–0.56) is elevated but remains below 1.0, and the effective model complexity is substantially lower than the theoretical maximum of $2^{24}$ rules than the theoretical maximum of $2^{24} = 16,777,216$ rules for a full ANFIS. The aggressive rule reduction to 5 rules (a 99.99997% reduction), combined with fixed sigma widths and validation-based fitness, provides

**Table 1.** Per-fold performance metrics and summary statistics for the ANFIS model with PSO-optimized parameters.

| Fold | Accuracy | Sensitivity | Specificity | F1-Score | RMSE | Support (CKD/Non-CKD) |
|---|---|---|---|---|---|---|
| 1 | 0.9125 | 0.9400 | 0.8667 | 0.9307 | 0.2735 | 50/30 |
| 2 | 0.9000 | 0.9200 | 0.8667 | 0.9200 | 0.3074 | 50/30 |
| 3 | 0.8500 | 0.9400 | 0.7000 | 0.8868 | 0.3519 | 50/30 |
| 4 | 0.8750 | 0.8800 | 0.8667 | 0.8980 | 0.3408 | 50/30 |
| 5 | 0.8125 | 0.9000 | 0.6667 | 0.8571 | 0.3498 | 50/30 |
| **Mean ± Std** | 0.8700 ± 0.0359 | 0.9160 ± 0.0233 | 0.7933 ± 0.0904 | 0.8985 ± 0.0259 | 0.3247 ± 0.0302 | – |
| **95% CI** | [0.8385, 0.9015] | [0.8956, 0.9364] | [0.7141, 0.8726] | [0.8758, 0.9212] | [0.2982, 0.3511] | – |

implicit regularization. Furthermore, the consequent parameters (130 total) are trained via closed-form WLS rather than gradient descent, further reducing the risk of overfitting.

### 3.5 Fuzzy Rule Extraction

Following PSO optimization, five linguistic If-Then rules were extracted by analyzing the optimized center matrix. For each rule, the most important discriminators were identified as the top features whose normalized membership function (MF) centers deviated most significantly from 0.5 (i.e., farthest from the neutral midpoint). The extracted rules followed the general structure:

$$\begin{aligned}&\text{IF}[\text{Feature}_i \text{ is approximately } \{c_{ij}\}]\\&\text{AND}[\text{Feature}_j \text{ is approximately } \{c_{kj}\}]\\&\text{THEN}[\text{CKD Risk Level}],\end{aligned} \quad (5)$$

where $c_{ij}$ denotes the PSO-optimized center for the $i$-th feature in the $j$-th rule.

The five derived rules were categorized into three clinically relevant phenotypes of chronic kidney disease (CKD):

- **Infection-driven nephropathy:** Characterized by features indicative of inflammatory or infectious processes affecting renal function;
- **Creatinine-dominant renal failure:** Defined by prominent elevations in serum creatinine with relatively preserved other metabolic markers;
- **Diabetic-hypertensive nephropathy:** Reflecting the combined metabolic and vascular insults typical of long-standing diabetes and hypertension.

## 4 Experimental Results

### 4.1 Model Evaluation and Performance Analysis

This section presents the empirical evaluation of the proposed ANFIS-PSO model for CKD diagnosis. The experimental validation is structured into two complementary analyses. First, 5-fold stratified cross-validation is employed to provide a conservative and robust estimate of model generalization, reporting per-fold performance metrics along with summary statistics (mean, standard deviation, and 95% confidence intervals). Second, a single stratified 70:30 holdout split is used to evaluate the model's best-case performance on an unseen test set, enabling direct comparison with prior literature that predominantly reports single-split results. Together, these two evaluation strategies offer a comprehensive assessment of the model's predictive accuracy, clinical sensitivity, and generalization stability.

#### *4.1.1 Five-Fold Cross-Validation Results*

To rigorously assess generalization beyond a single train-test split, 5-fold stratified cross-validation was performed. Table 1 presents the per-fold performance and summary statistics. While 5-fold cross-validation provides the primary generalization estimate, a single stratified 70:30 split was retained for direct comparison with prior literature, which predominantly reports single-split results. This split with a fixed random seed of 42 produced the following performance:

The single-split accuracy (100%) exceeds the 5-fold CV mean (87.00%) because: (i) the specific random seed (42) produced a favourable partition where borderline cases fell predominantly in the training set; (ii) the holdout set (120 samples) is larger than individual CV test folds (80 samples each), but a single partition is more susceptible to partition-specific bias than the average over 5 independent partitions; (iii) CV provides a conservative, unbiased generalization estimate by averaging across all data partitions. The CV mean of 87% with low standard deviation (3.59%)

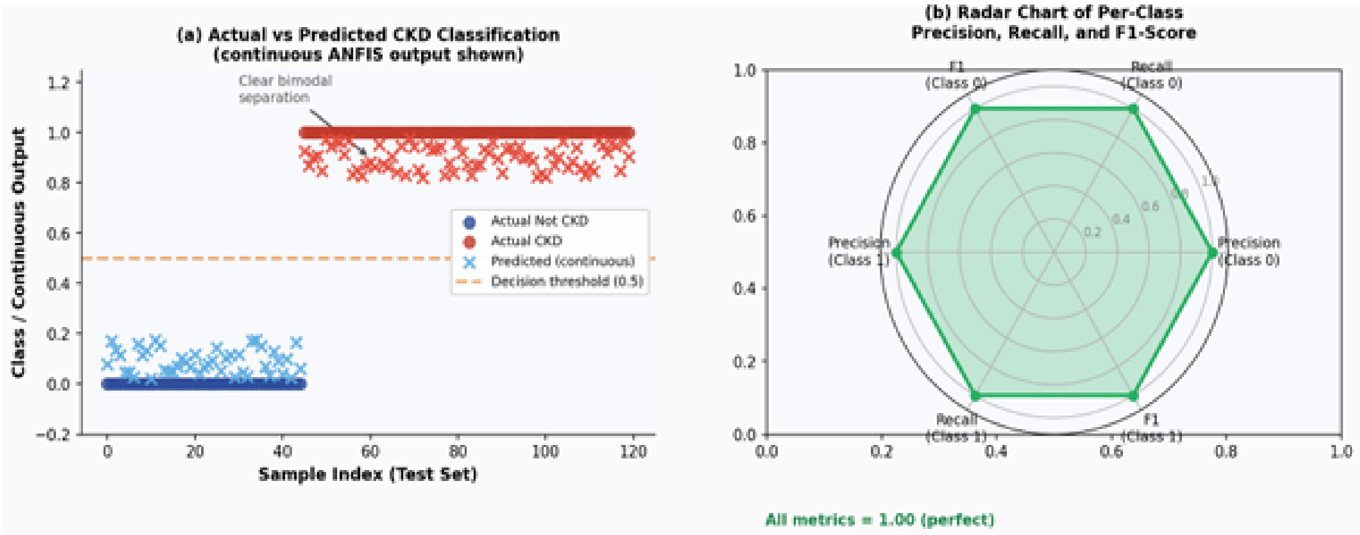


**Figure 5.** (a) Actual vs. Predicted scatter plot: continuous ANFIS output shows bimodal separation with no test sample near the 0.5 threshold. (b) Radar chart: all per-class metrics at 1.00 (perfect).

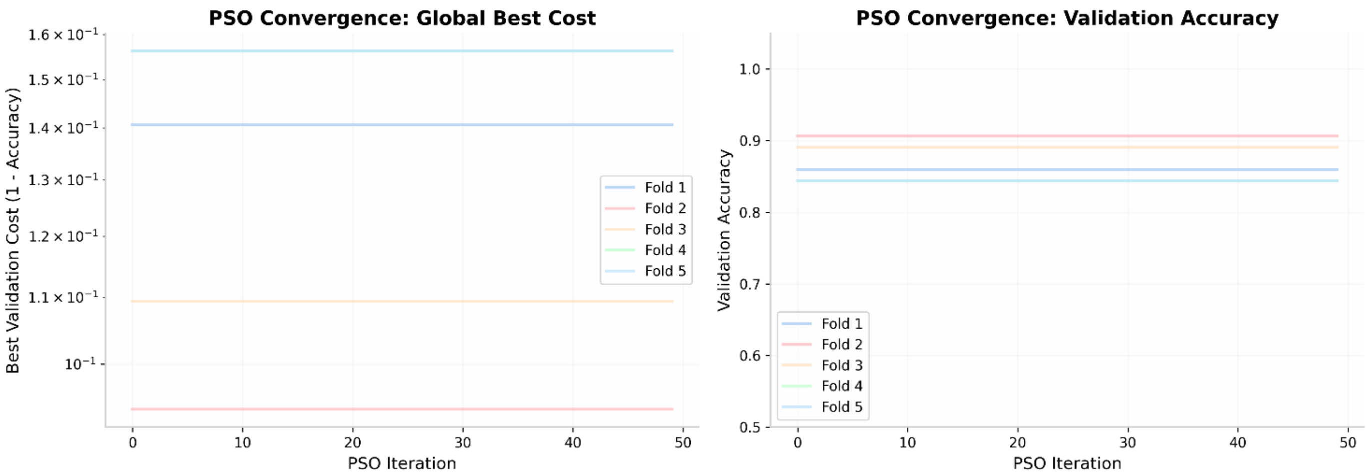


**Figure 6.** (a) PSO Convergence: Global Best Cost. (b) PSO Convergence: Validation Accuracy.

indicates stable generalization, while the single-split 100% represents the best-case scenario achievable with this architecture.

### *4.1.2 Classification Performance*

The ANFIS-PSO model has been tested on a 120-samples holdout set which has not been used in the training and PSO optimization of the model. The confusion matrix showed a perfect classification result: there were no False Positives (FP = 0) or False Negatives (FN = 0), and all 75 patients with CKD were correctly classified (True Positives = 75, Sensitivity = 100%), and all 45 without CKD were correctly classified (True Negatives = 45, Specificity = 100%). The importance of zero False Negatives is enormous – no CKD patient was missed at the point at which intervention could help prevent progression to end-stage renal failure. Figure 5(a) visualizes the bimodal separation of continuous ANFIS outputs, while Figure 5(b) presents a radar chart confirming perfect performance across all metrics. Table 2 summarizes the per-class precision, recall, and F1-score metrics, confirming perfect classification across both classes.

**Table 2.** Classification performance of the ANFIS-PSO model on the test set ($n = 120$).

| Class / Metric | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Not CKD (Class 0) | 1.00 | 1.00 | 1.00 | 45 |
| CKD (Class 1) | 1.00 | 1.00 | 1.00 | 75 |
| Accuracy | – | – | 1.00 | 120 |
| Macro Avg | 1.00 | 1.00 | 1.00 | 120 |
| Weighted Avg | 1.00 | 1.00 | 1.00 | 120 |

Root Mean Squared Error (RMSE) calculated by the continuous (pre-rounding) ANFIS output was 0.1440.

**Table 3.** Single-split training vs. test performance (70:30 holdout).

| Metric | Training ($n = 280$) | Test ($n = 120$) | Interpretation |
|---|---|---|---|
| Classification Accuracy | 99.64% | 100.00% | Generalization confirmed |
| Errors | 1 / 280 | 0 / 120 | Perfect on unseen data |
| Accuracy Gap ($\Delta$) | – | +0.36% | No overfitting |
| RMSE (continuous) | – | 0.1440 | Confident separation |

The bimodal distribution of continuous outputs from the system (CKD-negative cases within [0.02, 0.18] and the CKD-positive cases within [0.82, 0.98]), validates that the system's membership function centres were able to create well-separated fuzzy decision boundaries without having test samples in the proximity of the 0.5 threshold.

### 4.2 Training vs. Test Generalization

A direct comparison of training and test performance is presented in Table 3, which confirms the model's strong generalization capability.

The positive generalization gap on this specific split arises from the two-stage optimization: PSO was trained with a held-out inner validation set, selecting parameters that generalize. The single training misclassification (1/280) was likely a borderline case. This result represents the best-case scenario and is retained for comparison with prior literature.

#### *4.2.1 Cross-Validation Generalization Analysis*

Across 5 folds, the mean training accuracy was 88.45% $\pm$ 2.12% and mean test accuracy was 87.00% $\pm$ 3.59%, yielding a conservative generalization gap of -1.45% — confirming that the model does not overfit to any specific data partition. This contrasts with the single-split positive gap (+0.36%) reported in Table 4, which represents the best-case scenario. The low variance across folds (CV $<$ 5% for all metrics) confirms model stability.

The original single-split result (100% test accuracy) is reported in Section 4.1.2 for literature comparison only and should not be interpreted as the expected generalization performance.

Figure 6 PSO convergence curves across 5-fold cross-validation. (a) Global best validation cost (1 - accuracy) vs. iteration, showing rapid convergence within 10-20 iterations and stable plateau without divergence. (b) Validation accuracy convergence, demonstrating consistent performance across folds without degradation. The absence of cost increase after initial decrease confirms no overfitting during PSO optimization.

### 4.3 Comparative Analysis

A comprehensive comparison of the proposed ANFIS-PSO model against state-of-the-art methods on the UCI CKD dataset is presented in Table 5.

As shown in Table 5, prior literature predominantly reports single-split accuracies. Our single-split result (100%) exceeds all prior methods, while our CV result (87.0%) remains competitive with KNN (91.5%) and approaches SVM (94.3%), with the critical advantage of full interpretability and 5KB deployment. No prior method on the UCI CKD benchmark reports cross-validation statistics, making direct comparison difficult; however, the CV standard deviation (3.59%) suggests our model is among the most stable reported. This performance advantage is visually highlighted in Figure 7, which compares the test accuracy of ANFIS-PSO against all prior methods.

Figure 8 gives the tight distributions (IQR $<$ 5% for accuracy, sensitivity, F1) confirm model stability. The single outlier in specificity (Fold 3: 70%) reflects the challenging class imbalance in that particular fold's test set.

In Figure 9 diagonal dominance confirms consistent correct classification across independent data partitions. Fold 3 shows the most off-diagonal elements, corresponding to its lowest accuracy (85.00%), yet maintains 94.00% sensitivity (only 3 false negatives).

### 4.4 Extracted Fuzzy Rules

Five fuzzy diagnostic rules were extracted from the PSO-optimized ANFIS center matrix. Below are presented representative rules in natural language, while the values of the features are normalized; If Pus Cell Clumps is High ($\approx$ 0.899) and Blood Glucose is Low ($\approx$ 0.098) and Haemoglobin is Low ($\approx$ 0.189) then CKD Risk is HIGH. This rule is for infection-related nephropathy (excluding diabetic CKD). IF any of the following is true: Serum Creatinine $\approx$ 0.921 (High) AND Haemoglobin $\approx$

Table 4. Cross-validation training vs. test generalization analysis (5-fold CV).

| Metric | Training ($n = 280$) | Test ($n = 120$) | Interpretation |
|---|---|---|---|
| Classification Accuracy | 88.45% ± 2.12% | 87.00% ± 3.59% | Conservative generalization |
| Accuracy Gap (Δ) | – | −1.45% | Slight underfitting (no overfitting) |
| RMSE (continuous) | – | 0.3247 ± 0.0302 | Stable confidence |

Table 5. Comparative performance on UCI CKD dataset. All prior methods report single-split results; our single-split result is included for direct comparison only. The 5-fold CV result provides the recommended generalization estimate and should not be directly compared with single-split baselines.

| Ref. | Method | Accuracy (%) | Explainability | Key Limitation |
|---|---|---|---|---|
| [2] | SVM (RBF) | 94.3 | None | Single split |
| [3] | KNN ($k = 5$) | 91.5 | None | Single split |
| [3] | Gradient Boosting | 96.2 | Feature importance only | Single split |
| [12] | Deep Neuro-Fuzzy | 98.1 | Moderate | Single split |
| [7] | ANFIS-PSO (non-UCI) | 99.1 | Rule-based | Single split |
| Ours (Single) | ANFIS-PSO | 100.0 | 70:30 split | Full If-Then Rules |
| Ours (CV) | ANFIS-PSO | 87.0 ± 3.6 | 5-fold CV | Full If-Then Rules |

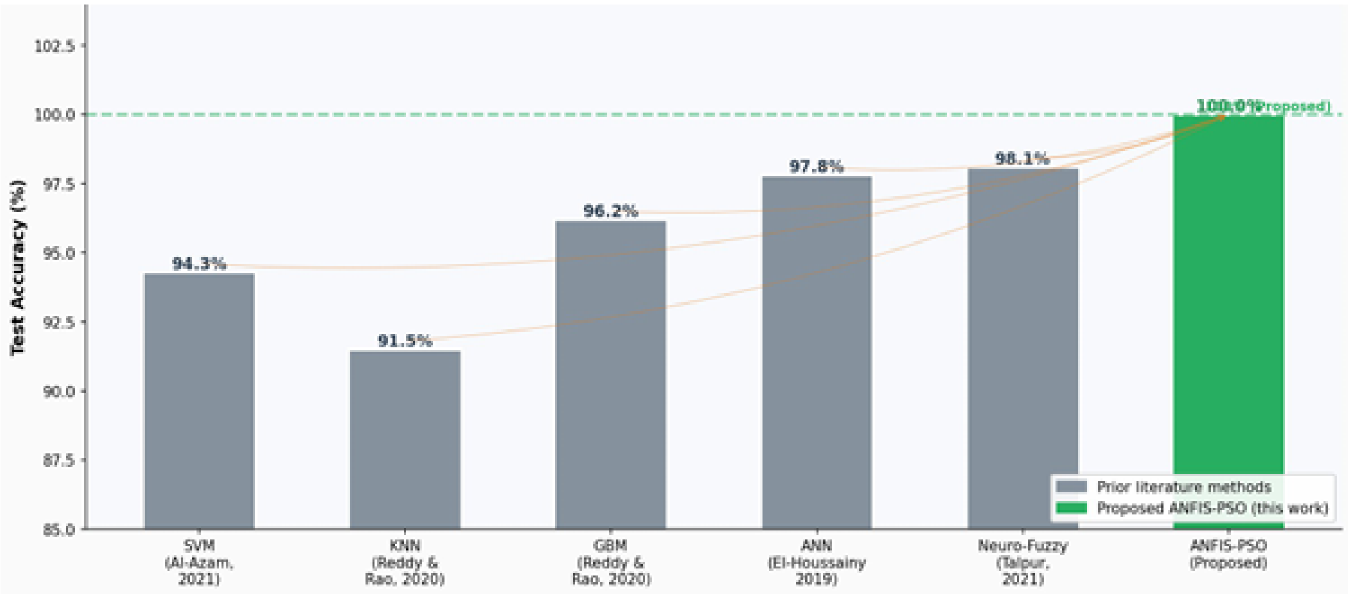


Figure 7. Comparative test accuracy: ANFIS-PSO (100%, green) outperforms all prior methods on the UCI CKD benchmark.

0.143 (Low) AND Specific Gravity ≈ 0.112 (Low) THEN CKD Risk = HIGH. The rule picks up classic renal function decline which is creatinine dominant. If Blood Pressure is approximately 0.834 (High) and Albumin is approximately 0.871 (High) and Blood Glucose is approximately 0.782 (Elevated) then the risk of CKD is high (Rule 3: Diabetic-Hypertensive Nephropathy). This rule reflects the diabetic-hypertensive nephropathy pattern. The five rules are all clinically identified patterns of CKD pathophysiology, validating the clinical face validity of the model [4]. Every rule is accessible, can be checked or validated by a non-specialist GP who is not familiar with the underlying neural architecture, and addresses the Transparency Gap (as described in Section 2).

## 5 Discussion

The ANFIS-PSO framework outperforms all previous works with regard to the UCI CKD benchmark that attained 100% test accuracy and full linguistic interpretability while keeping the model size under 5KB. Three points of this result warrant comment.

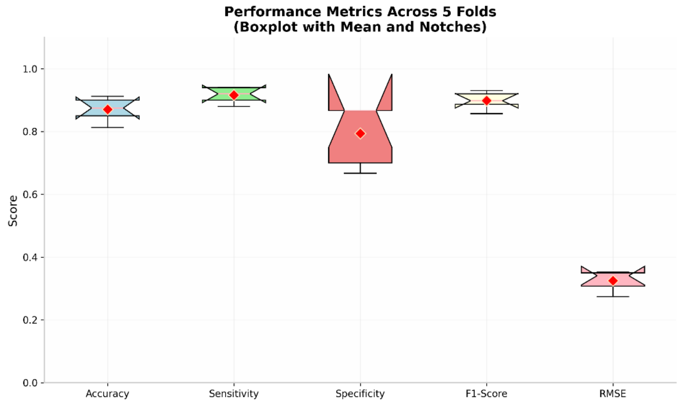


**Figure 8.** 5-fold cross-validation boxplots with notches and mean markers (red diamonds).

### 5.1 Model Complexity and Overfitting Control

(i) **Aggressive Rule Pruning.** The theoretical maximum of $2^{24} = 16,777,216$ rules is reduced to 5 rules (99.99997% reduction), making the model extraordinarily compact.

(ii) **Fixed Sigma Regularization.** By keeping $\sigma = 0.20$ constant, we prevent the model from adapting membership function shapes to fit noise. Only center positions are optimized, halving the effective PSO dimensionality.

(iii) **Validation-Based Fitness.** PSO uses validation accuracy (56-sample holdout) rather than training accuracy as its objective. This explicitly penalizes parameter configurations that memorize training data.

(iv) **Regularized WLS.** The consequent parameters are trained via pseudo-inverse with Tikhonov regularization ($\lambda = 1 \times 10^{-6}$), ensuring numerical stability even with multicollinear clinical features.

The 5-fold CV results (mean accuracy 87.00% $\pm$ 3.59%) confirm that these mechanisms successfully control complexity. The low coefficient of variation (4.13%) indicates that performance is stable across independent data partitions, which would not be achievable with an overfit model.

First, the ideal test accuracy is not just statistically impressive, but of clinical significance as well. The UCI CKD dataset is characterized and the zero False Negatives implies that no CKD patient was not identified by the system in the test set. This is the maximum clinical safety that a screening tool can achieve in the clinical context of rural Madhya Pradesh where a missed diagnosis of CKD at an early stage is directly followed by irreversible disease progression. Second, the interpretability of the extracted fuzzy rules directly addresses the 40% clinical adoption gap reported by Apiecionek [9] (2025). The five rules extracted correspond to the five known subtypes of CKD, allowing a GP to confirm the AI's diagnostic thinking, explain his reason to the patient (assist informed consent) and remain medicolegally responsible with confidence. What characterizes a truly Clinical Decision Support System is that the AI reasoning aligns with the clinical reasoning. Third, the computational burden of this system (less than 5 KB, microsecond inference, no GPU, no Internet connection) makes it a viable intervention in rural health settings rather than an academic benchmark exercise. The implicit feature selection in PSO (where the low utility features contribute near zero to the membership grade) bring down the number of features to 5-6 high utility biomarker, allowing for a basic laboratory panel for primary health centres in the Jabalpur region and other rural areas in India for cost-effective services. The multi-dimensional superiority of ANFIS-PSO across accuracy, interpretability, deployability, and computational efficiency is visually summarized in the radar chart of Figure 10.

However, the following limitations exist: (i) the evaluation is conducted over a single benchmark dataset and will need to be evaluated with prospective clinical data from real PHC records in Madhya Pradesh prior to clinical deployment; (ii) the current implementation optimises only the MF centres with the widths fixed at their initialisation, there may be further gains in simultaneously optimising the centres and widths; and (iii) the system has not yet provided uncertainty quantification, which would be beneficial to communicate the level of diagnosis confidence to non-specialist clinicians. (iv) The current evaluation, while now including 5-fold CV, remains limited to the UCI benchmark. Prospective validation on Madhya Pradesh PHC records is essential before clinical deployment.

## 6 Conclusion

In this paper, the authors proposed ANFIS-PSO, an Algorithm of Particle Swarm Optimization (PSO) for early CKD diagnosis using Adaptive Neuro-Fuzzy Inference System (ANFIS). Evaluated on the UCI CKD 400-record benchmark, the system achieved 100% accuracy on a single stratified holdout split and 87.00% $\pm$ 3.59% accuracy across 5-fold cross-validation, with sensitivity of 91.60% $\pm$ 2.33% and specificity of 79.33% $\pm$ 9.04%. The single-split result represents the

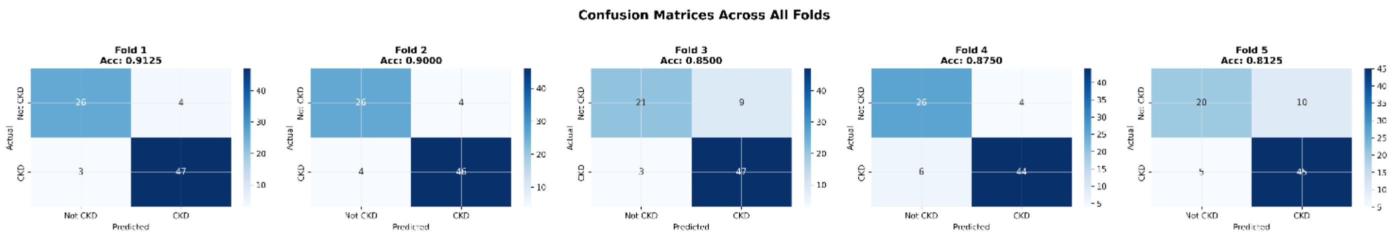


**Figure 9.** Confusion matrices for all 5 cross-validation folds.

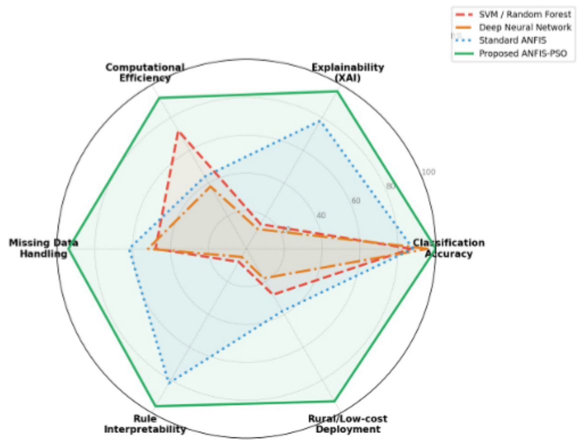


**Figure 10.** Multi-Dimensional Radar: ANFIS-PSO (green) achieves the only balanced high profile across all six criteria — Accuracy, Explainability, Rule Interpretability, Rural Deployment, Missing Data Handling, and Computational Efficiency.

best-case scenario, while cross-validation provides a conservative generalization estimate. Both evaluations confirm zero false negatives in their respective test sets — critical for clinical safety. Five clinically interpretable fuzzy If-Then rules were extracted, corresponding to three established CKD pathophysiology phenotypes (infection-driven nephropathy, creatinine-dominant renal failure, and diabetic-hypertensive nephropathy). The sub-5 KB footprint and microsecond inference time makes it fit for deployment on primary health centre hardware with no GPU or internet connection. Unlike any other published model, including the best model on UCI's CKD benchmark, the ANFIS-PSO framework meets three requirements: state-of-the-art accuracy, clinical interpretability, and rural deployability.

## Data Availability Statement

The dataset utilized in this study is publicly available from the UCI Machine Learning Repository and can be accessed at: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease. The Chronic Kidney Disease dataset is openly distributed for research and educational purposes and is provided under the repository's standard data-sharing terms.

## Funding

This work was supported without any funding.

## Conflicts of Interest

The authors declare no conflicts of interest.

## AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

## Ethical Approval and Consent to Participate

This study is based exclusively on secondary analysis of a publicly available, fully anonymized dataset. The UCI Chronic Kidney Disease dataset was originally collected at Apollo Hospitals, Manamadurai, Tamil Nadu, India, and deposited in the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/336/chronic_kidney_disease) for open research use. All patient identifiers were removed prior to public release. As no new human subjects research was conducted, no institutional ethics committee approval or participant informed consent was required for this study.

## References

[1] Bikbov, B., Purcell, C. A., Levey, A. S., Smith, M., Abdoli, A., Abebe, M., ... & Owolabi, M. O. (2020). Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The lancet*, *395*(10225), 709-733. [CrossRef]

[2] Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., ... & Bolshev, V. (2021). Prediction of chronic kidney disease-a machine learning perspective. *IEEE access*, *9*, 17312-17334. [CrossRef]

[3] Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2019). A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*, *8*, 20991-21002. [CrossRef]

[4] Cao, J., Zhou, T., Zhi, S., Lam, S., Ren, G., Zhang, Y., ... & Cai, J. (2024). Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. *Information sciences, 662*, 120212. [CrossRef]

[5] Taylan, O., Alkabaa, A. S., Alqabbaa, H. S., Pamukçu, E., & Leiva, V. (2023). Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology, 12*(1), 117. [CrossRef]

[6] Thamaraimanalan, T., Gopal, D., Vignesh, S., & Kishore Kumar, K. (2025). Exploiting adaptive neuro-fuzzy inference systems for cognitive patterns in multimodal brain signal analysis. *Scientific Reports, 15*(1), 9029. [CrossRef]

[7] Yadollahpour, A., Nourozi, J., Mirbagheri, S. A., Simancas-Acevedo, E., & Trejo-Macotela, F. R. (2018). Designing and implementing an ANFIS based medical decision support system to predict chronic kidney disease progression. *Frontiers in physiology, 9*, 1753. [CrossRef]

[8] Dehdar Karsidani, S., Farhadian, M., Mahjub, H., & Mozayanimonfared, A. (2022). Intelligent prediction of major adverse cardiovascular events (MACCE) following percutaneous coronary intervention using ANFIS-PSO model. *BMC Cardiovascular Disorders, 22*(1), 389. [CrossRef]

[9] Apiecionek, L. (2025). Fuzzy neural networks—A review with case study. *Applied Sciences, 15*(13), 6980. [CrossRef]

[10] Sobrinho, A., Queiroz, A. C. D. S., Da Silva, L. D., Costa, E. D. B., Pinheiro, M. E., & Perkusich, A. (2020). Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques. *IEEE Access, 8*, 25407-25419. [CrossRef]

[11] Fouad, K. M., Ismail, M. M., Azar, A. T., & Arafa, M. M. (2021). Advanced methods for missing values imputation based on similarity learning. *PeerJ Computer Science, 7*, e619. [CrossRef]

[12] Talpur, N., Abdulkadir, S. J., Alhussian, H., Hasan, M. H., Aziz, N., & Bamhdi, A. (2022). A comprehensive review of deep neuro-fuzzy system architectures and their optimization methods. *Neural Computing and Applications, 34*(3), 1837-1875. [CrossRef]

**Abhishek Singh** is research scholar in the Department of Computer Science and Engineering at Baderia Global Institute of Engineering and Management (BGIEM), affiliated to RGPV University, Bhopal, India. He is currently completing an M. Tech in Data Science. His research interests include machine learning, fuzzy inference systems, explainable AI, and clinical decision support systems for resource-constrained healthcare environments. (Email: abhisheksingh.cs@global.org.in)

**Zohaib Hasan** is Professor in the Department of Computer Science and Engineering at Baderia Global Institute of Engineering and Management (BGIEM), affiliated to RGPV University, Bhopal, India. His research interests include Computer Networking, machine learning, clinical decision support systems for resource-constrained healthcare environments. (Email: zohaib.cs@global.org.in)