A Systematic Literature Review of Adaptive Neuro Fuzzy Models for Early Heart Attack Detection

Siyani Ezra Dogo, Yusuf Musa Malgwi, Etemi Joshua Garba, Caleb Markus*

Department of Computer Science, Modibbo Adama University, Yola

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101000272

Received: 05 October 2026; Accepted: 10 October 2026; Published: 01 November 2026

ABSTRACT

Cardiovascular prediction increasingly combines data-driven learning with interpretable reasoning, yet reported performance is difficult to compare because studies use different datasets, validation schemes, optimization targets and clinical endpoints. This systematic literature review examines adaptive neuro-fuzzy models for early heart attack and cardiovascular risk detection, with particular attention to model architecture, metaheuristic optimization, dataset design, evaluation practice and translation readiness. The review followed the PRISMA 2020 statement and searched Scopus, Web of Science, IEEE Xplore, PubMed, ScienceDirect, SpringerLink and ACM Digital Library for studies published between January 2000 and September 2026. Of 11,321 records identified, 23 studies were retained after deduplication, screening, eligibility assessment and quality appraisal, and their risk of bias was judged with a checklist mapped to the PROBAST domains. Evidence indicates a progression from conventional ANFIS models toward feature-engineered, swarm-optimized, IoMT-enabled and explainability-aware variants. Reported accuracy is often high, but direct ranking is unsafe because Cleveland-derived datasets, which label angiographic coronary artery disease rather than acute myocardial infarction, dominate, external validation remains uncommon, and calibration, subgroup fairness and prospective clinical testing are rarely reported. Optimization methods including particle swarm optimization, genetic algorithms, salp swarm optimization, beetle swarm optimization and hybrid metaheuristics improve search over membership functions, rule parameters and feature subsets, but they also increase computational and reproducibility demands. Rule bases grow exponentially with the number of inputs, and few studies compare ANFIS with tuned ensemble or deep learning baselines under the same validation protocol. The review concludes that future progress should prioritize multi-centre validation, uncertainty quantification, transparent rule extraction, clinically meaningful calibration and reproducible benchmark protocols rather than accuracy alone.

Keywords: ANFIS, adaptive neuro-fuzzy inference system, heart attack detection, cardiovascular disease, metaheuristic optimization, systematic literature review

INTRODUCTION

Cardiovascular disease remains a major global health burden, and early risk recognition is important because myocardial infarction can emerge from interacting demographic, metabolic, behavioural and physiological factors that are not always well represented by simple thresholds. Artificial intelligence has therefore been used to complement conventional risk assessment by modelling nonlinear relations across age, blood pressure, cholesterol, glucose, electrocardiographic variables and other clinical indicators. Adaptive Neuro-Fuzzy Inference Systems are especially relevant because they combine the learning capability of neural networks with fuzzy rules that represent gradual clinical states such as low, moderate and high risk. The architecture introduced by Jang provides a trainable Sugeno-type fuzzy system in which premise and consequent parameters can be learned from data, while fuzzy set theory provides a formal basis for reasoning with partial membership rather than rigid binary boundaries. In cardiovascular applications, this combination is attractive because clinical measurements are noisy, correlated and context dependent, while practitioners still need an understandable account of how risk is inferred. Recent cardiovascular ANFIS studies consequently move

beyond conventional training toward feature selection, swarm optimization, cloud-connected monitoring and interpretable hybrid models (Jang, 1993; Khan & Algarni, 2020; Siontis et al., 2021; Taylan et al., 2023; World Health Organization, 2021; Zadeh, 1965).

The literature nevertheless presents a difficult evidence landscape. High predictive scores are frequently reported, but studies differ substantially in endpoint definition, sample size, feature set, preprocessing, class balance, cross-validation design and whether optimization is applied to inputs, membership functions, rules or classifier parameters. A model evaluated on the small Cleveland dataset cannot be compared directly with one trained on a hospital cohort or long-term post-PCI outcomes, even when both report accuracy. Similarly, optimization can improve internal test performance while increasing the risk of overfitting when feature selection and parameter tuning are performed outside a properly nested validation loop. The strongest recent work therefore points toward a broader concept of model quality that includes discrimination, calibration, sensitivity, specificity, robustness, interpretability and external validity. This review treats performance as an evidence package rather than a single number and distinguishes algorithmic improvement from clinical readiness. Figure 1 summarizes this evidence architecture. That distinction is essential because the intended use is early detection, where false negatives, subgroup performance and transportability across patient populations can matter more than small gains in headline accuracy (Collins et al., 2015; Dehdar Karsidani et al., 2022; Paul & Karn, 2026; Steyerberg et al., 2010; Suri et al., 2022).

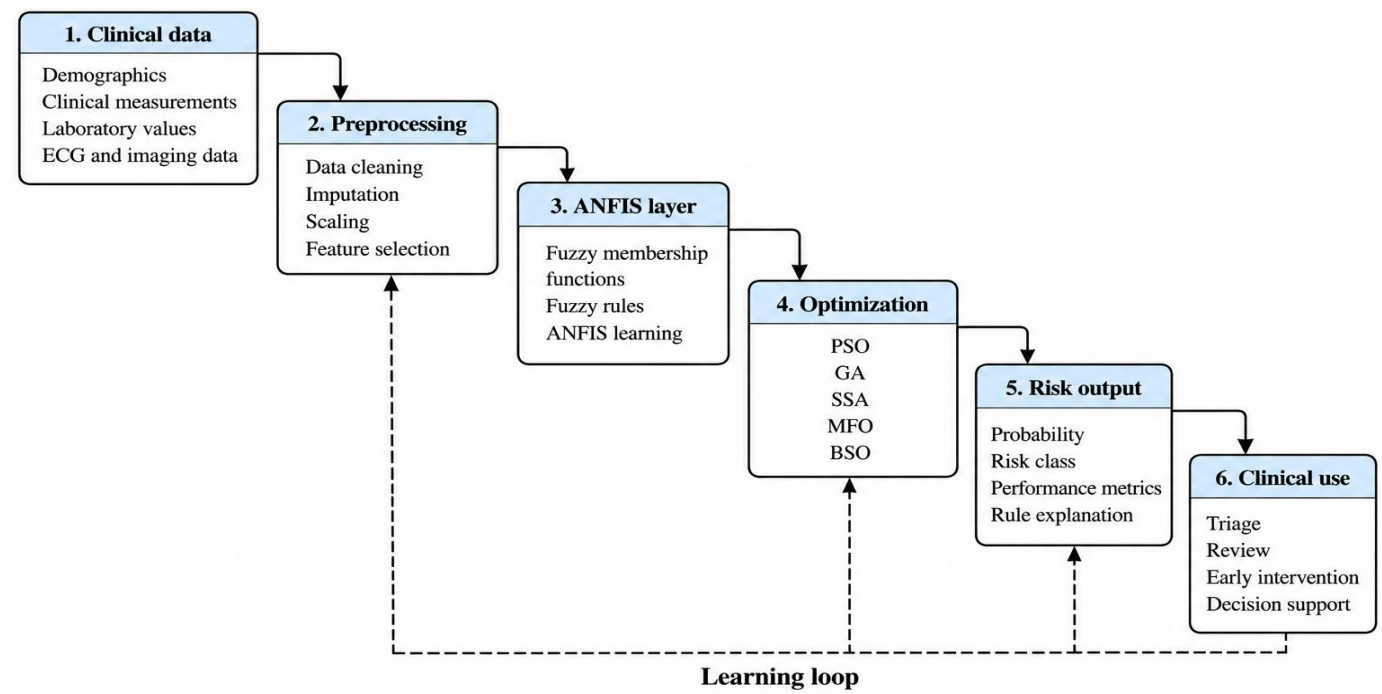

Figure 1: Evidence architecture for adaptive neuro-fuzzy cardiovascular prediction

The central problem is not whether ANFIS can classify cardiovascular data, but whether the current evidence supports reliable, generalizable and clinically interpretable early detection. Conventional ANFIS can become trapped in local optima because premise parameters are commonly adjusted through gradient-based learning, and the number and shape of membership functions can cause rule-base growth as input dimensionality increases. Metaheuristic optimization addresses part of this problem by searching wider parameter spaces, but it introduces new choices involving population size, stopping criteria, random seeds and computational cost. Dataset limitations create a second problem because repeated use of the same UCI-derived records can produce apparent methodological progress without demonstrating transportability to different hospitals, ethnic groups or acquisition settings. A third problem concerns reporting: many studies emphasize accuracy while omitting calibration, confidence intervals, external validation, decision-curve utility and reproducible code. The review therefore aims to examine the application and performance of adaptive neuro-fuzzy models for early heart attack detection and asks four questions: which ANFIS variants are used, which optimization strategies are applied, how models are evaluated, and which gaps limit clinical translation, as set out in Table 1 (Bhuvaneswari Amma, 2013; Feng et al., 2021; Jang, 1993; Jayachitra et al., 2022; Suri et al., 2022).

Table 1: Review objectives and research questions

Objective	Research question
Identify neuro-fuzzy model families	RQ1: What adaptive neuro-fuzzy models have been applied to early heart attack or cardiovascular detection?
Examine optimization strategies	RQ2: What optimization techniques are used to improve neuro-fuzzy models?
Compare datasets and evaluation	RQ3: How are existing neuro-fuzzy models evaluated?
Identify limitations and future needs	RQ4: What research gaps limit current adaptive neuro-fuzzy approaches?

SYSTEMATIC REVIEW METHODOLOGY

2.1 Review Protocol and Information Sources

A structured review protocol was used to reduce selection drift and make the synthesis reproducible. Figure 2 shows the sequence of the protocol, and Table 2 lists the seven information sources and their roles in the review. The protocol was informed by PRISMA 2020 and software-engineering review guidance, but was adapted to the mixed clinical and computational character of neuro-fuzzy prediction research. Searches were organized around three concept blocks: adaptive neuro-fuzzy methods, cardiovascular or heart-attack outcomes, and prediction or diagnosis. Seven information sources were targeted because no single database provides complete coverage of biomedical engineering, computer science and clinical prediction research. Searches covered publications from January 2000 to September 2026, a window that begins well before the earliest cardiovascular ANFIS applications in the retained set and captures the recent shift toward optimized and hybrid variants, and the final search was run on 28 September 2026. The searches returned 11,321 records, of which 2,541 came from Scopus, 2,103 from Web of Science, 1,892 from IEEE Xplore, 1,654 from PubMed, 1,402 from ScienceDirect, 1,021 from SpringerLink and 708 from the ACM Digital Library (Table 2 and Figure 3). PubMed captures medical and cardiovascular studies, IEEE Xplore and ACM Digital Library cover computational systems, while Scopus, Web of Science, ScienceDirect and SpringerLink broaden multidisciplinary retrieval. Search results were deduplicated before title and abstract screening. Full texts were then assessed against predefined eligibility criteria, followed by quality appraisal and structured extraction. The protocol deliberately separated study inclusion from performance interpretation so that a high reported accuracy could not compensate for weak validation or unclear methodology. Search strings, exclusion reasons, quality scores and extraction fields were retained as an audit trail, consistent with reproducibility principles for systematic reviews and prediction-model research (Collins et al., 2015; Kitchenham & Charters, 2007; Moher et al., 2009; Page et al., 2021; Wolff et al., 2019).

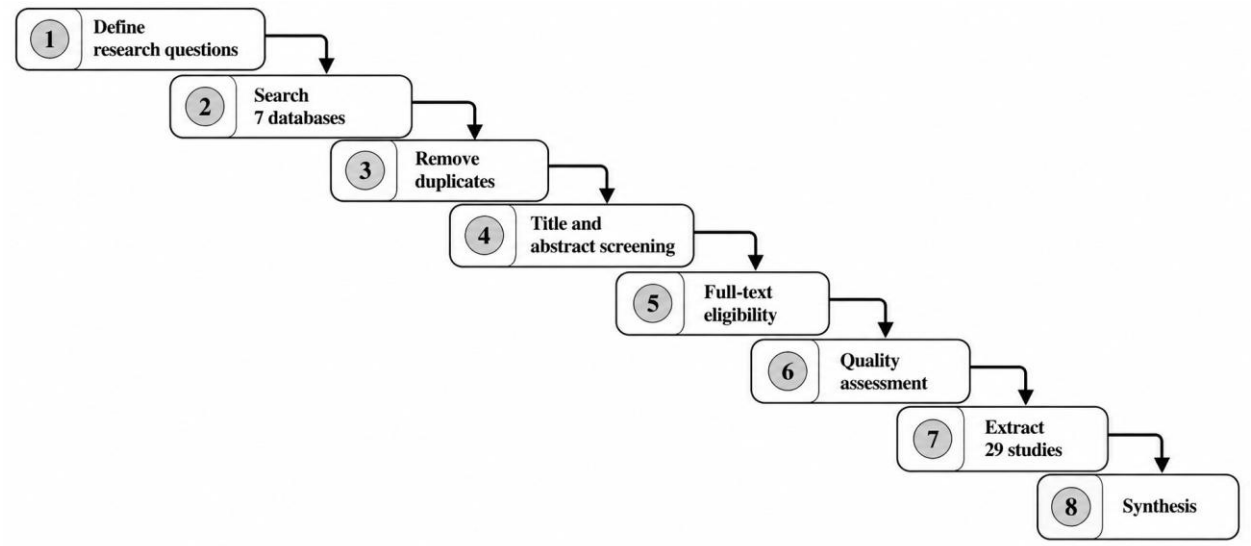

Figure 2: Systematic review protocol showing the sequence from research questions to synthesis

Table 2: Electronic information sources

Source	Primary coverage	Review role	Records retrieved
Scopus	Multidisciplinary	Broad citation-indexed retrieval	2,541
Web of Science	Multidisciplinary	Cross-disciplinary quality-controlled retrieval	2,103
IEEE Xplore	Engineering and computing	ANFIS, optimization and IoMT studies	1,892
PubMed	Biomedical research	Clinical cardiovascular evidence	1,654
ScienceDirect	Engineering and health	Applied AI and signal-processing studies	1,402
SpringerLink	Computing and medicine	Neuro-fuzzy and optimization literature	1,021
ACM Digital Library	Computer science	AI and data-driven system research	708
Total			11,321

2.2 Search Strategy and Eligibility

The core search expression combined synonyms rather than relying on a single model label: ("adaptive neuro fuzzy" OR "ANFIS" OR "neuro fuzzy") AND ("heart attack" OR "heart disease" OR "cardiovascular disease" OR "coronary artery disease" OR "myocardial infarction") AND ("prediction" OR "detection" OR "diagnosis" OR "risk"). Database-specific syntax was adjusted for title, abstract and keyword fields while preserving the same conceptual structure. Table 3 summarizes the inclusion and exclusion criteria. Studies were eligible when they presented an empirical neuro-fuzzy or ANFIS-based cardiovascular prediction model, reported enough methodological detail to understand the model and evaluation, and provided quantitative outcomes. Closely related cardiovascular event prediction studies were retained when ANFIS was a central predictive component, including post-PCI event prediction and ECG-based coronary risk detection. Purely theoretical fuzzy papers, non-cardiovascular applications, papers without empirical evaluation, non-scholarly web content and inaccessible reports were excluded. The scope was intentionally broad enough to capture the evolution from conventional ANFIS to optimized and hybrid variants, while remaining narrow enough for meaningful synthesis. Foundational sources were used to interpret model design but were not counted among the 23 retained empirical studies (Bhuvaneswari Amma, 2013; Jang, 1993; Khan & Algarni, 2020; Taylan et al., 2023; Zadeh, 1965).

Each criterion had a specific purpose. ANFIS had to be a central predictive component rather than a peripheral comparator, because the review question concerns neuro-fuzzy design choices. Heart attack was operationalized broadly to include coronary artery disease, myocardial infarction and closely related cardiac events, because few ANFIS studies target acute myocardial infarction alone and most rely on heart disease labels from benchmark datasets; the consequences of this choice are examined in Section 6. Quantitative evaluation was required so that model behaviour could be compared, and peer-reviewed publication was required to limit the inclusion of unvetted performance claims. Studies needed at least an English title and abstract and a full text from which methods and results could be extracted reliably, because the reviewers could not appraise methodological detail otherwise. This restriction, together with the decision not to search grey literature, may have excluded relevant regional work, a limitation acknowledged in Section 11.

Table 3: Eligibility criteria

Inclusion criteria	Exclusion criteria
Empirical ANFIS or neuro-fuzzy cardiovascular prediction	No neuro-fuzzy component
Heart disease, CAD, myocardial infarction or closely related cardiac event	Non-cardiovascular target
Quantitative evaluation reported	No empirical performance evaluation
Peer-reviewed article or conference paper	Blog, commentary or non-scholarly source

Sufficient methodological detail	Model or dataset insufficiently described
English title and abstract, with a full text from which methods and results could be extracted	Full text unavailable or not extractable

2.3 Study Selection and Quality Assessment

The selection process identified 11,321 records across the seven sources. After removal of 2,017 duplicates, 9,304 unique records entered title and abstract screening. At this stage, 8,673 records were excluded because they were not related to cardiovascular disease or ANFIS, were not empirical studies, addressed the wrong population or outcome, or were reviews, editorials or non-scholarly sources. Six hundred and thirty-one full texts were assessed, and 608 were excluded because they had no ANFIS component, lacked sufficient methodological detail, reported no quantitative evaluation, addressed a non-cardiovascular application or could not be retrieved in full. Twenty-three studies remained for quality appraisal and synthesis, as shown in Figure 3. Screening, full-text assessment and quality appraisal were carried out independently by two reviewers, with disagreements resolved through discussion with a third reviewer. Quality assessment used seven questions, listed in Table 4, covering clarity of objectives, dataset description, model specification, optimization transparency, evaluation metrics, comparative analysis and limitations. Each item was scored 1 for yes, 0.5 for partial and 0 for no. Studies were not ranked by accuracy during selection because this would favour small or highly curated datasets. Instead, methodological transparency and relevance determined inclusion, while performance was analysed later in relation to dataset and validation context (Collins et al., 2015; Kitchenham & Charters, 2007; Page et al., 2021; Whiting et al., 2011; Wolff et al., 2019).

Because reporting quality and risk of bias are different properties, each checklist item was also mapped to the corresponding domain of the PROBAST tool for prediction model studies, which judges risk of bias in participants, predictors, outcome and analysis (Wolff et al., 2019), as shown in Table 4. A study was judged at high risk of bias in a domain when any item mapped to that domain scored 0. The analysis domain is where the weaknesses discussed in Sections 7 and 8 are captured, including small samples relative to the number of model parameters, single holdout splits, unhandled class imbalance and optimization performed outside the resampling loop (Riley et al., 2019). Quality and risk of bias did not determine inclusion, but they shaped interpretation, so performance figures from studies at high risk of bias in the analysis domain are treated in this review as proof of concept rather than as evidence of generalizable accuracy.

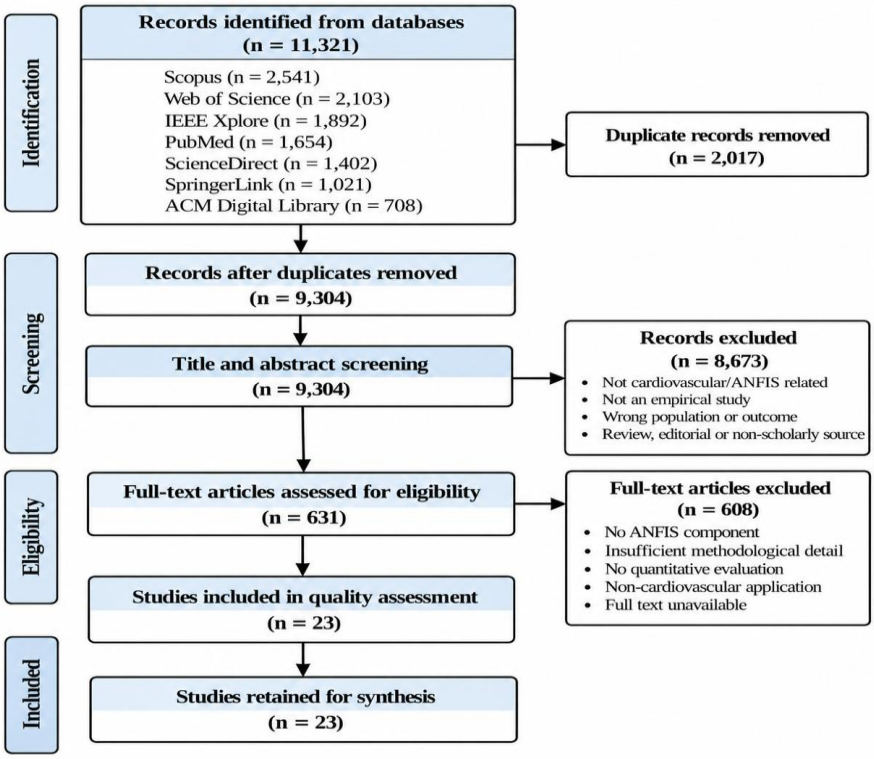

Figure 3: PRISMA-style flow of study identification, screening, eligibility assessment and final retention of 23 studies

Table 4: Quality assessment checklist

Code	Assessment question	PROBAST domain	Score
QA1	Is the research objective clearly stated?	Applicability	0, 0.5, 1
QA2	Is the dataset adequately described?	Participants and outcome	0, 0.5, 1
QA3	Is the neuro-fuzzy architecture reproducible?	Predictors	0, 0.5, 1
QA4	Is optimization adequately specified?	Analysis	0, 0.5, 1
QA5	Are appropriate evaluation metrics reported?	Analysis	0, 0.5, 1
QA6	Are meaningful comparator models used?	Analysis	0, 0.5, 1
QA7	Are limitations or threats discussed?	Overall judgement	0, 0.5, 1

Note. Items scored 1 (yes), 0.5 (partial) or 0 (no). A score of 0 on any item mapped to a PROBAST domain was taken to indicate high risk of bias in that domain.

2.4 Data Extraction and Synthesis

A standardized extraction matrix, summarized in Table 5, captured bibliographic information, clinical endpoint, dataset source, sample size, input features, preprocessing, ANFIS architecture, optimization target, comparator models, validation design, performance metrics, interpretability mechanism, reported limitations and reproducibility information. This structure was necessary because the same algorithm label can conceal materially different pipelines. For example, PSO may select features in one study, tune membership functions in another, or optimize a meta-classifier in a hybrid ensemble. Similarly, a study may report 99 percent accuracy using a single holdout split while another reports lower accuracy under repeated cross-validation or an independent cohort. The synthesis therefore combined descriptive mapping with thematic analysis rather than meta-analysis. Quantitative pooling was not attempted because outcome definitions, datasets and validation protocols were too heterogeneous for a defensible common effect estimate. Instead, the review compared model families, optimization strategies, dataset patterns and evaluation maturity, and then mapped recurring weaknesses to future research priorities. This approach follows the principle that heterogeneity should be explained rather than hidden by inappropriate aggregation (Higgins et al., 2022; Page et al., 2021; Riley et al., 2019; Steyerberg et al., 2010; Suri et al., 2022).

Table 5: Data extraction framework

Field	Information extracted
Study	Author, year, venue
Clinical target	Heart disease, CAD, MACCE, ECG risk
Data	Dataset, sample size, feature type
Model	ANFIS variant and architecture
ANFIS configuration	Membership function type and number, rule generation, training algorithm
Optimization	Algorithm and optimization target
Validation	Holdout, k-fold, external
Metrics	Accuracy, sensitivity, specificity, F1, AUC
Interpretability	Rules, SHAP, feature importance
Limitations	Bias, generalization, computational issues

RESULTS

The 23 retained studies show a clear shift from conventional ANFIS classification toward optimization-intensive and hybrid architectures. Early studies concentrated on proving that fuzzy-neural learning could classify Cleveland-style cardiovascular records, often using a reduced subset of clinical features. Later work introduced preprocessing and dimensionality reduction, including principal component analysis, fuzzy c-means clustering and wavelet-based ECG feature extraction. From 2020 onward, the dominant direction became

metaheuristic tuning, IoMT integration and multi-stage pipelines in which feature selection and ANFIS parameter optimization are jointly addressed. Particle swarm optimization, genetic algorithms, salp swarm optimization, beetle swarm optimization, moth-flame variants and hybrid search strategies appear repeatedly because they can explore non-convex parameter spaces without requiring differentiability. More recent studies also begin to treat interpretability as an explicit evaluation dimension rather than assuming that fuzzy rules are automatically clinically meaningful. Figure 4 shows the temporal concentration of retained studies, Figure 5 shows their geographical distribution, while Table 6 summarizes representative evidence. The distribution should not be interpreted as a publication-rate census because it reflects the review eligibility criteria rather than all cardiovascular AI research (Bhuvaneswari Amma, 2013; Holle, 2016; Khan & Algarni, 2020; Paul & Karn, 2026; Singh et al., 2021).

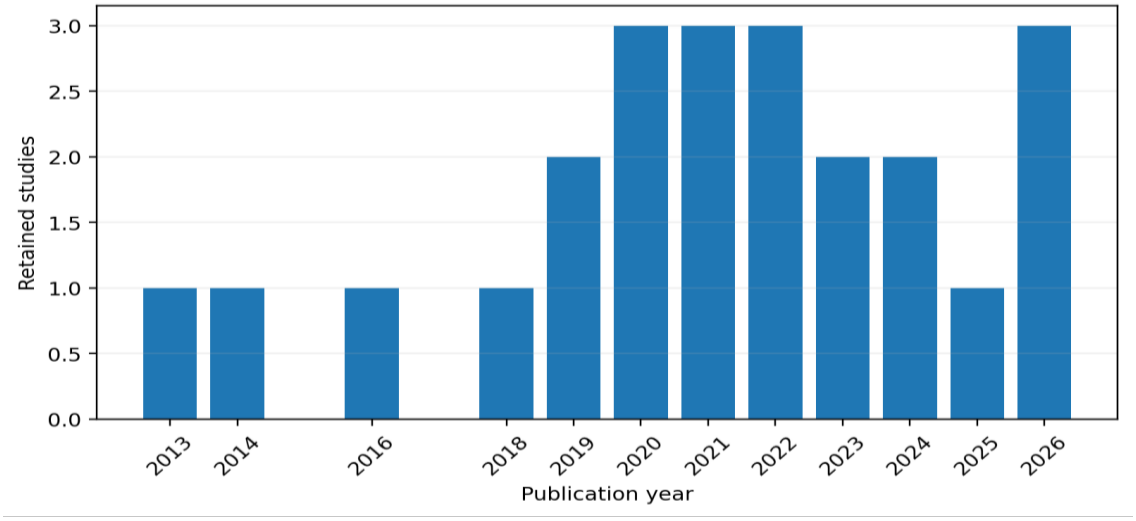

Figure 4: Publication-year distribution of the 23 retained studies

As shown in Figure 5, the 23 retained studies come from twelve countries, and India is the largest single contributor with eight studies (34.8%). China, Iran, Saudi Arabia and the United States contribute two studies each (8.7% each), while Brazil, Egypt, Indonesia, Iraq, Latvia, South Korea and Turkey contribute one study each (4.3% each). By region, Asia contributes 12 studies (52.2%) and the Middle East seven (30.4%), which together account for 82.6% of the evidence, while North America contributes two (8.7%) and Europe and South America one each (4.3% each); Turkey, Iran, Iraq, Saudi Arabia and Egypt are grouped here as the Middle East. Each study was assigned to the country of its first author affiliation, so the figure describes where the research was produced and not where the patient data were collected. This distinction matters because cardiovascular risk profiles, measurement practice and disease prevalence differ between populations, and a model trained on a public benchmark or a single hospital cohort cannot be assumed to transport to other settings (Collins et al., 2015; Steyerberg et al., 2010). The geographical spread of the research therefore does not remove the need for external and multi-centre validation emphasized in later sections, because the evidence base remains dominated by a small number of benchmark datasets.

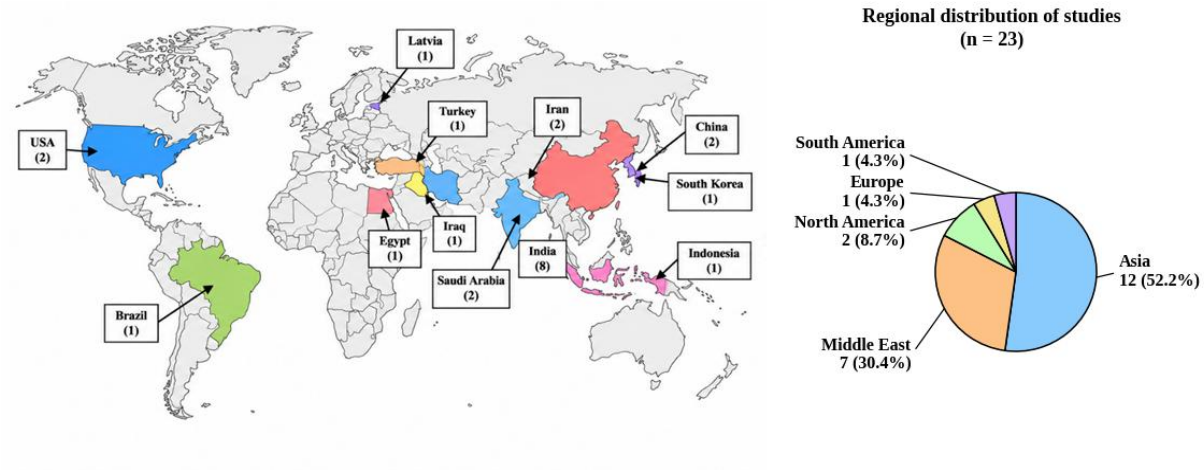

Figure 5: Geographical distribution of the 23 retained studies

Table 6: Characteristics of the 23 retained studies

Study	Year	Model	Dataset/context	Optimization or enhancement	Reported result
Bhuvaneswari Amma	2013	PCA-ANFIS	Cleveland UCI	PCA	Accuracy 93.2%
Ziasabounchi & Askerzade	2014	ANFIS	UCI heart disease	Hybrid learning	Accuracy 92.3%
Holle	2016	ANFIS	Cleveland UCI	Learning-rate tuning	Accuracy 65.7%
Manogaran et al.	2018	MKL-ANFIS	Metabolic/heart data	Multiple-kernel learning	High sensitivity/specificity*
Puspita & Abadi	2019	Wavelet-ANFIS	ECG	Wavelet extraction	ECG diagnosis
Rajinikanth & Pavithra	2019	Bat-ANFIS	Heart disease data	Modified bat optimization	Optimized classification
Khan & Algarni	2020	MSSO-ANFIS	IoMT clinical data	Modified SSA	Accuracy 99.45%
Polatgil	2020	FCM-ANFIS	Heart disease data	Fuzzy c-means	Improved classification
Sood et al.	2020	LDA-ANFIS	Dew-cloud data, dengue patients	LDA	CHD risk monitoring in dengue
Feng et al.	2021	GA-ANFIS	UCI heart data	Genetic algorithm	Accuracy 91.25%
Singh et al.	2021	BSO-ANFIS	Heart/multi-disease	BSO + crow search	Accuracy 99.1%
Jayachitra et al.	2022	TANFIS	11 UCI datasets	MFO + GOA	Accuracy 99.76%
Dehdar Karsidani et al.	2022	ANFIS-PSO	PCI cohort	PSO	Accuracy 89%
Taylan et al.	2023	ANFIS + ML	Hospital CVD data	Hybrid learning	ANFIS training 96.56%
Rajabzadeh et al.	2023	ANFIS	MASHAD cohort (n = 7,385)	Compared with LR and FDA	Accuracy 83.4%
Sumarlinda et al.	2024	ANFIS comparison	500 records	Comparative ML	Early CVD prediction
Bhagat et al.	2024	ANFIS/TANFIS	Cleveland	Tuning comparison	Comparative accuracy
“A Novel Approach”	2024	SANFIS	Cleveland	Hybrid AITH2O	Accuracy 99.28%
Gopi & Arun	2026	ANFIS + PCA/MI	Heart disease data	Feature extraction	Accuracy 95.5%
Paul & Karn	2026	SC-ANFIS + SHAP	Cleveland/Statlog	Subtractive clustering	Accuracy 99.87/98.80%
Shadhan et al.	2026	Breakpoint ANFIS	200 patients	Segmented modelling	ANFIS superior to OLS
Jeya et al.	2026	ANFIS-PSO	Clinical data	PSO	Optimized prediction
Narigina et al.	2026	Interval type-2 ANFIS layer	PTB-XL ECG	Deep backbone with fuzzy experts	Interpretable cardiac recognition

Note. Results are as reported by the original authors and are not comparable across rows, because datasets, endpoints and validation designs differ. Studies using the Cleveland or Statlog data predict angiographic coronary artery disease rather than acute myocardial infarction (Section 6). LR = logistic regression; FDA = flexible discriminant analysis.

ADAPTIVE NEURO-FUZZY MODEL FAMILIES

The retained evidence can be organized into four model families, as shown in Figure 6, and Table 7 compares their main advantages, risks and evidence needs. Conventional ANFIS studies use the standard five-layer architecture with data-driven adjustment of membership and consequent parameters. Their main value is conceptual clarity, but they can struggle with local minima and rule growth. Feature-engineered ANFIS pipelines address dimensionality before inference, for example through PCA, clustering, wavelets or discriminant analysis. This can reduce computational burden and improve signal-to-noise ratio, although feature extraction must be confined to the training process to avoid leakage. Metaheuristic-optimized ANFIS models explicitly search membership functions, rule parameters or feature subsets using population-based algorithms. These studies often report the largest gains, but they also create more hyperparameters and increase the need for repeated runs and nested validation. Hybrid and ensemble ANFIS models integrate cloud monitoring, IoMT, multiple learners or explanation layers, moving the field toward systems rather than isolated classifiers. The progression is technically important because it shows that current research is no longer asking whether neuro-fuzzy learning works, but how it should be optimized, validated, explained and embedded in clinical workflows (Bhuvaneswari Amma, 2013; Jang, 1993; Khan & Algarni, 2020; Paul & Karn, 2026; Singh et al., 2021).

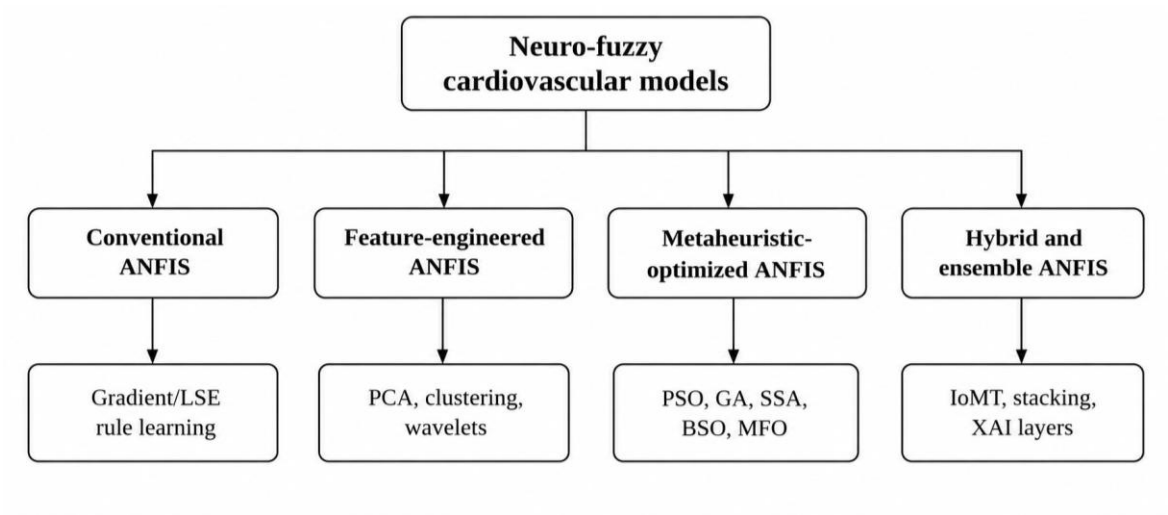

Figure 6: Taxonomy of adaptive neuro-fuzzy cardiovascular models identified in the retained literature

Table 7: Comparison of neuro-fuzzy model families

Model family	Primary advantage	Main risk	Best evidence need
Conventional ANFIS	Transparent baseline architecture	Local optima and rule growth	Strong validation baseline
Feature-engineered ANFIS	Reduced dimensionality and noise	Feature-selection leakage	Nested preprocessing
Metaheuristic ANFIS	Global parameter search	Overfitting and compute cost	Repeated nested validation
Hybrid/ensemble ANFIS	Combines complementary learners	Complexity and reduced traceability	Ablation and external validation
Explainability-aware ANFIS	Rules plus post-hoc explanation	Explanation may not equal clinical validity	Clinician-centred validation

Beyond the family a model belongs to, its internal configuration determines size, interpretability and overfitting risk, and Table 8 summarizes the main choices. Membership functions are usually Gaussian, generalized bell, triangular or trapezoidal. Smooth Gaussian and bell functions suit gradient-based tuning, whereas triangular and trapezoidal functions are easier for clinicians to read as value ranges. Rules are generated by grid partitioning, subtractive clustering or fuzzy c-means clustering. Grid partitioning creates one

rule for every combination of input membership functions, so with *n* inputs and *m* functions per input the rule base contains m^n rules. For the 13 Cleveland attributes, two functions per input already yield 8,192 rules and three yield 1,594,323, far more than the 303 records available to estimate them (Detrano et al., 1989; Jang, 1993). This curse of dimensionality explains why most retained studies either reduce the inputs first, through PCA, mutual information or discriminant analysis, or generate rules from the data through subtractive clustering or fuzzy c-means, which ties the number of rules to the number of clusters rather than to the number of inputs (Bhuvaneswari Amma, 2013; Gopi & Arun, 2026; Paul & Karn, 2026; Polatgil, 2020). Training most often uses Jang's hybrid rule, combining least-squares estimation for consequent parameters with gradient descent for premise parameters, which metaheuristic search then replaces or supplements (Section 5). Because these choices shape both performance and the readability of the final rule base, studies should report them explicitly; in the retained set they were often described only in part, which limited how far configurations could be compared.

Table 8: ANFIS configuration choices and their implications

Component	Common options	Effect on the model	What studies should report
Membership function type	Gaussian, generalized bell, triangular, trapezoidal	Smoothness of tuning and readability as clinical ranges	Type and fitted parameters for each input
Number of membership functions	Typically two to five per input	Rule count grows as m^n under grid partitioning	Number per input and how it was chosen
Rule generation	Grid partitioning, subtractive clustering, fuzzy c-means	Controls rule-base size and interpretability	Method, cluster radius or number of clusters, final rule count
Training algorithm	Hybrid least squares and gradient descent, Levenberg-Marquardt, metaheuristic search	Convergence, local optima and computational cost	Algorithm, epochs, stopping rule and search settings
Input reduction	PCA, mutual information, LDA, wrapper selection	Fewer rules but risk of selection leakage	Method and whether it was nested within training folds
Inference type	Type-1 Sugeno, interval type-2	Treatment of uncertainty in membership grades	Fuzzy system type and output computation

OPTIMIZATION TECHNIQUES

Optimization is the most distinctive methodological theme in the recent literature. Table 9 compares the main optimization approaches by typical target, strength and limitation. Standard ANFIS training commonly combines least-squares estimation for consequent parameters with gradient descent for premise parameters, which is efficient but sensitive to initialization and local minima. Genetic algorithms encode candidate parameter sets and use selection, crossover and mutation to search globally, while particle swarm optimization updates candidate solutions according to individual and collective best positions. Salp swarm optimization models leader-follower movement and has been adapted for heart-disease feature selection and ANFIS parameter tuning. Beetle swarm optimization, moth-flame optimization, grasshopper optimization and hybrid arithmetic or Harris-hawk variants extend the same principle by balancing exploration and exploitation differently. The critical issue is not the biological metaphor but the optimization target and validation discipline. Figure 7 separates four targets: input features, premise parameters, rule structure and consequent parameters. A method that tunes all four on the same validation data can overfit even when test accuracy rises. Stronger studies should therefore report objective functions, search bounds, population size, iterations, random

seeds, computational budget and whether optimization occurred entirely inside training folds (Dehdar Karsidani et al., 2022; Feng et al., 2021; Kennedy & Eberhart, 1995; Mirjalili et al., 2017; Singh et al., 2021).

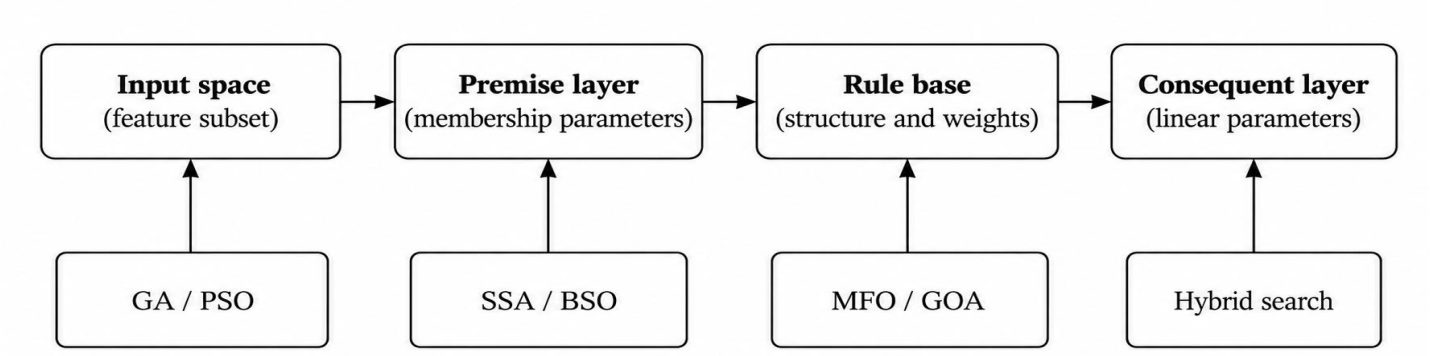

Figure 7: Optimization targets within an ANFIS cardiovascular prediction pipeline and representative metaheuristic families

Table 9: Comparison of optimization approaches

Approach	Typical target	Strength	Limitation
PSO	Membership parameters, features	Simple continuous search and rapid convergence	Premature convergence possible
GA	Rules, membership functions, features	Flexible representation and global search	Computationally intensive
SSA/MSSO	Features and ANFIS parameters	Exploration-exploitation balance	Sensitive to implementation choices
BSO	ANFIS parameters	Strong nonlinear search	Limited independent replication
MFO/GOA hybrids	TANFIS tuning	Broad search with hybrid operators	Many hyperparameters
Subtractive clustering	Rule generation	Controls rule-base construction	Radius selection affects complexity

DATASETS, CLINICAL FEATURES AND PREPROCESSING

Dataset choice is the largest source of comparability risk. Table 10 summarizes the main data contexts, their typical inputs, common preprocessing and key concerns. The Cleveland heart disease dataset is repeatedly used because it is public, compact and historically important, but repeated optimization on the same benchmark can encourage implicit overfitting at the research-community level. Some retained studies broaden the evidence base using the MASHAD cohort, hospital cardiovascular records, ECG signals, IoMT sensor data, post-PCI follow-up or multiple UCI datasets. These sources differ in clinical endpoint, prevalence, measurement quality and feature semantics. Common structured predictors include age, sex, chest-pain type, resting blood pressure, cholesterol, fasting glucose, ECG findings, maximum heart rate and exercise-related ST changes, while ECG-based studies extract temporal or frequency-domain features. Preprocessing practices include missing-value handling, normalization, PCA, mutual information, clustering, wavelet transforms and SMOTE. These steps can materially change performance, so they should be treated as part of the model rather than neutral preparation. External validation is especially important because a classifier trained on curated benchmark data may not preserve calibration or sensitivity when applied to a hospital population with different disease prevalence or measurement practices (Detrano et al., 1989; Dua & Graff, 2019; Gopi & Arun, 2026; Puspita & Abadi, 2019; Rajabzadeh et al., 2023).

The dominance of the Cleveland data also creates an endpoint problem that bears directly on the aim of this review. The Cleveland records were collected in the 1980s from patients referred for coronary angiography, and their label indicates angiographic coronary artery disease, defined as more than 50 percent diameter narrowing in a major vessel, rather than acute myocardial infarction or a future cardiac event (Detrano et al., 1989). The widely used Statlog version contains 270 of these records. Models trained on these data therefore learn to recognize established coronary disease in a small, referred, symptomatic and largely male population from one era and setting, which is a different task from the early detection of heart attack in the general population. Only a minority of retained studies used larger or more recent data, such as the MASHAD cohort of 7,385 participants (Rajabzadeh et al., 2023), hospital records (Taylan et al., 2023), post-PCI outcomes (Dehdar Karsidani et al., 2022) or the PTB-XL electrocardiogram corpus (Narigina et al., 2026). Results from benchmark studies should therefore be read as evidence about classifying coronary disease labels, and claims about heart attack detection require validation on real clinical cohorts with myocardial infarction outcomes. Resampling methods such as SMOTE (Chawla et al., 2002) and feature selection (Guyon & Elisseeff, 2003) must also be confined to training folds, since applying them before splitting lets information from the test data leak into model development (Kohavi, 1995).

Table 10: Dataset and preprocessing patterns

Data context	Typical inputs	Common preprocessing	Key concern
Cleveland/UCI	13 structured clinical variables	Scaling, PCA, feature selection	Small benchmark and repeated reuse
Hospital cohort	Demographic, biochemical, anthropometric	Cleaning, balancing, holdout	Transportability and prevalence
ECG/HRV	Waveforms and derived intervals	Wavelets, segmentation, feature extraction	Signal quality and patient-level splitting
IoMT/wearables	Continuous sensor and clinical data	Streaming cleaning, feature selection	Latency, missingness and privacy
Post-PCI cohort	Clinical and procedural variables	Normalization and longitudinal outcome coding	Small event count and temporal validity

PERFORMANCE EVALUATION

Accuracy remains the most frequently reported metric, but it is insufficient for early heart-attack detection because class imbalance and asymmetric clinical consequences can make a high accuracy misleading. Table 11 sets out the main metrics, what each measures, its clinical interpretation and the cautions that apply. Sensitivity is crucial when the priority is to avoid missed high-risk patients, while specificity measures the ability to limit false alarms and unnecessary follow-up. Precision and F1 score become useful when positive cases are less common, and AUC summarizes discrimination across thresholds (Fawcett, 2006; Hanley & McNeil, 1982). Yet none of these measures indicates whether predicted probabilities are calibrated. A clinically useful risk model should also be assessed with calibration plots, calibration slope or intercept, Brier score and, where a decision threshold is meaningful, decision-curve analysis. Figure 8 illustrates the uneven reporting of metrics across dataset contexts in the retained evidence, and Table 6 shows the same pattern, since accuracy is the only metric reported consistently across the retained studies while sensitivity, specificity, F1 score and AUC appear in only some of them. A like-for-like comparison of these clinically important measures is therefore not possible from published results. Benchmark studies frequently emphasize accuracy, whereas clinical cohort studies are more likely to include sensitivity, specificity and AUC. Direct performance ranking is therefore inappropriate unless the same data, endpoint and validation protocol are used. Confidence intervals and repeated resampling are also important because single-split estimates can vary substantially in small datasets

(Dehdar Karsidani et al., 2022; Paul & Karn, 2026; Steyerberg et al., 2010; Taylan et al., 2023; Van Calster et al., 2019).

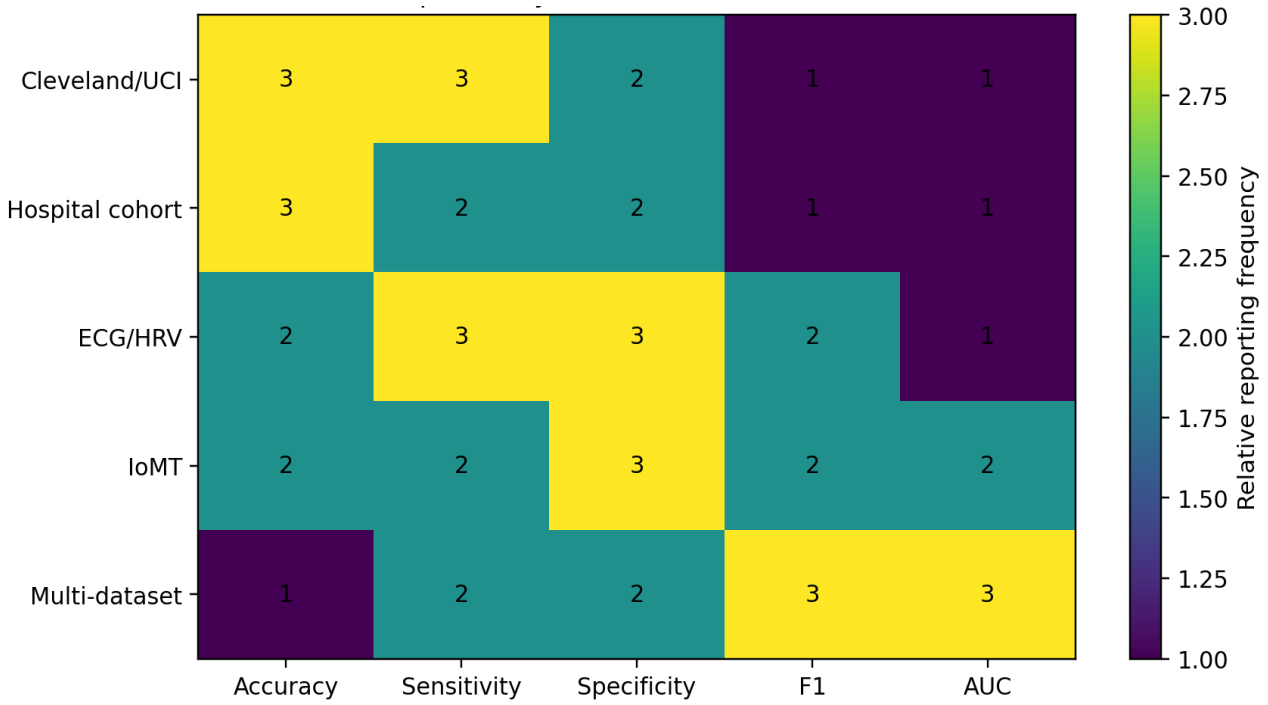

Figure 8: Evidence comparability matrix

Table 11: Performance evaluation framework

Metric	What it measures	Clinical interpretation	Review caution
Accuracy	Overall correct classifications	General classification success	Can hide minority-class errors
Sensitivity	True-positive rate	Ability to detect high-risk cases	Threshold dependent
Specificity	True-negative rate	Ability to avoid false alarms	Threshold dependent
Precision	Positive predictive value	Reliability of positive alerts	Prevalence dependent
F1 score	Balance of precision and recall	Useful under imbalance	Ignores true negatives
AUC	Rank discrimination	Threshold-independent ranking	Does not assess calibration
Calibration/Brier	Probability reliability	Supports risk communication	Rarely reported in retained studies

CHALLENGES AND RESEARCH GAPS

The first major gap is generalization. Many high-performing models are evaluated on small public datasets, often after intensive feature selection and optimization, but are not tested on independent populations. This creates uncertainty about whether gains reflect a robust cardiovascular signal or adaptation to a familiar benchmark. The second gap is validation design. Metaheuristic feature selection and hyperparameter search can leak information when performed before data splitting or outside cross-validation folds, and few studies explicitly describe nested validation. The third gap is interpretability. ANFIS is often described as inherently interpretable because it uses fuzzy rules, yet a large rule base with many membership functions can be difficult for clinicians to inspect. Recent work that combines subtractive clustering with SHAP (Lundberg & Lee, 2017) is a useful step because it links rule construction to feature-level explanations, but explanation quality still

requires clinical validation. The fourth gap is incomplete reporting of calibration, uncertainty, fairness and subgroup performance. These omissions matter because early detection systems may behave differently across age, sex, comorbidity and acquisition settings even when aggregate accuracy is high (Collins et al., 2015; Paul & Karn, 2026; Riley et al., 2019; Suri et al., 2022; Van Calster et al., 2019).

A fifth gap concerns reproducibility and computational transparency. Metaheuristic studies frequently report the algorithm name and final score but provide limited information about initialization, random seeds, search bounds, stopping rules, repeated runs or hardware. Without these details, it is difficult to determine whether performance is stable or dependent on a favourable run. A sixth gap is clinical deployment evidence. IoMT and cloud-based ANFIS frameworks demonstrate technical feasibility for remote monitoring, but deployment requires attention to missing sensor data, latency, privacy, cybersecurity, model updating and human oversight. A seventh gap is outcome definition. Heart disease, coronary artery disease, myocardial infarction risk and post-PCI MACCE are clinically related but not interchangeable targets. Combining them without distinction can exaggerate the apparent size and consistency of the evidence base. Finally, few studies evaluate net clinical benefit or prospective workflow impact. The field therefore needs to shift from algorithm-centred demonstrations toward prediction-model studies that clearly define intended use, target population, decision threshold and consequences of false-positive and false-negative outputs (Dehdar Karsidani et al., 2022; Khan & Algarni, 2020; Page et al., 2021; Siontis et al., 2021; Wolff et al., 2019). Table 12 summarizes each research gap, its current limitation and the research required to address it.

Clinical usability and real-time performance also receive less attention than they deserve. Once trained, an ANFIS model is computationally light, since inference requires only the evaluation of membership functions and a weighted combination of rule outputs, which makes it suitable for bedside, mobile or wearable use. The cost lies in training, where metaheuristic search may require thousands of model evaluations and must be repeated whenever the model is updated. IoMT and dew-cloud frameworks show that streaming deployment is technically feasible (Khan & Algarni, 2020; Sood et al., 2020), but the retained studies rarely report latency under realistic load, usability testing with clinicians or integration with electronic health records. A rule base that is readable in principle is only useful at the bedside if the number of rules is small, the membership functions correspond to recognized clinical ranges, and the output is presented in a way that supports rather than replaces clinical judgement (Topol, 2019).

Table 12: Research gaps and required evidence

Research gap	Current limitation	Required research
External validity	Benchmark-heavy evidence	Multi-centre external validation
Optimization leakage	Search often poorly nested	Nested cross-validation and locked test sets
Calibration	Discrimination dominates reporting	Calibration curves, Brier score and recalibration
Interpretability	Rule bases may be large	Rule pruning plus clinician-tested explanations
Fairness	Aggregate metrics dominate	Subgroup and intersectional evaluation
Reproducibility	Search settings often incomplete	Seeds, code, bounds and repeated-run reporting
Clinical utility	Few prospective evaluations	Decision-curve and workflow studies
Outcome heterogeneity	Different cardiac endpoints combined	Endpoint-specific evidence synthesis

FUTURE RESEARCH DIRECTIONS

Future research should treat neuro-fuzzy cardiovascular prediction as a staged translation problem rather than a competition for the highest accuracy. The first stage is methodological strengthening: preprocessing, feature selection and optimization should be nested within resampling, and performance should be reported with uncertainty intervals (Harrell, 2015). The second stage is probabilistic reliability: ANFIS outputs should be

calibrated so that a predicted risk has a meaningful relationship to observed event frequency. The third stage is interpretability: rule extraction, membership functions and post-hoc explanations such as SHAP or LIME (Lundberg & Lee, 2017; Ribeiro et al., 2016) should be simplified and tested with clinicians for consistency, usefulness and potential automation bias. The fourth stage is external validation on real clinical data with myocardial infarction outcomes, across hospitals, demographic groups and acquisition platforms. The fifth stage is prospective deployment, where the model is evaluated as part of a clinical workflow rather than as an isolated classifier. Figure 9 summarizes this evidence-to-translation pathway. Metaheuristic innovation remains useful, but future contributions should justify why a new optimizer changes clinically relevant performance, computational efficiency or interpretability instead of only adding another search metaphor. Benchmark leadership should therefore be replaced by evidence maturity as the main criterion for progress (Collins et al., 2015; Paul & Karn, 2026; Siontis et al., 2021; Van Calster et al., 2019; Wolff et al., 2019). Table 13 translates these priorities into recommended directions and expected contributions.

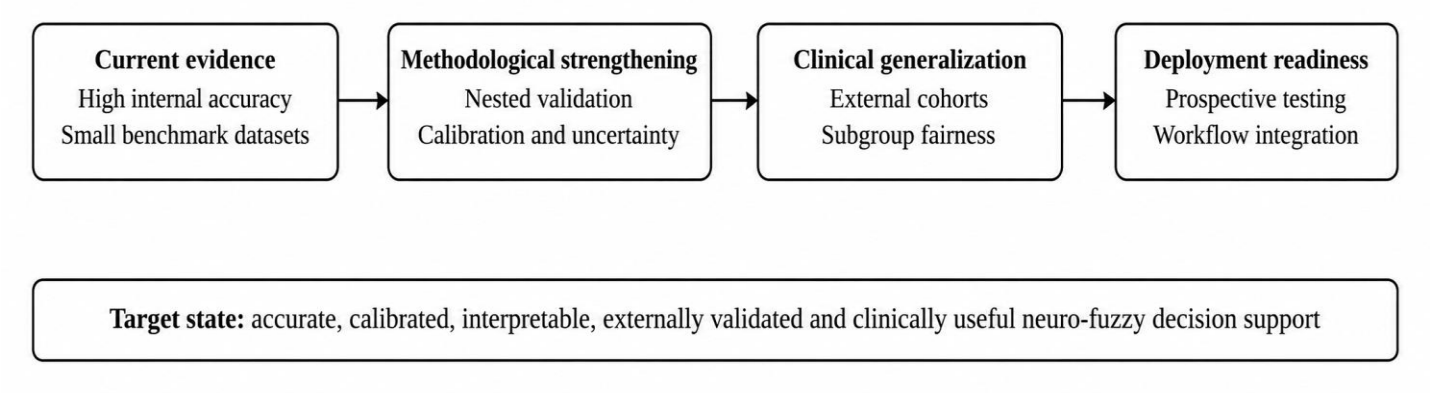

Figure 9: Evidence-to-clinical-translation roadmap for adaptive neuro-fuzzy heart attack detection

Table 13: Future research agenda

Priority	Recommended direction	Expected contribution
Validation	Nested and external validation	More credible generalization estimates
Calibration	Probability calibration and uncertainty	Safer risk communication
Interpretability	Rule pruning, SHAP and clinician evaluation	Actionable transparency
Data diversity	Multi-centre and longitudinal cohorts	Improved transportability
Optimization	Ablation and compute-aware search	Evidence that complexity adds value
Fairness	Subgroup performance assessment	Reduced hidden performance disparities
Deployment	Prospective decision-support trials	Evidence of real clinical utility

DISCUSSION

The synthesis answers RQ1 by showing that adaptive neuro-fuzzy cardiovascular prediction has evolved through conventional, feature-engineered, metaheuristic-optimized and hybrid system families. RQ2 is answered by the broad use of PSO, GA, SSA, BSO, MFO, GOA and hybrid search, with optimization applied to features, membership functions, rule parameters and classifier settings. RQ3 shows that evaluation is still dominated by discrimination metrics, particularly accuracy, while calibration, uncertainty and external validation remain comparatively weak. RQ4 identifies a consistent gap between algorithmic performance and clinical evidence. The strongest contribution of ANFIS is not simply that it can achieve high accuracy, but that it provides a structured interface between nonlinear learning and fuzzy reasoning. This makes it particularly suitable for cardiovascular risk factors that vary continuously and interact in complex ways. However, interpretability is conditional rather than automatic. A model with dozens of inputs and hundreds of fuzzy rules can become as difficult to inspect as other complex models. Optimization should therefore be paired with rule-

base control, feature parsimony and transparent validation rather than treated as an end in itself (Dehdar Karsidani et al., 2022; Jang, 1993; Paul & Karn, 2026; Pekaslan et al., 2020; Taylan et al., 2023).

The evidence also clarifies why reported scores above 99 percent should be interpreted cautiously. Several highly optimized studies use the Cleveland dataset or related public datasets that are small by modern prediction-model standards. Repeated methodological experimentation on the same benchmark can indirectly tune research choices to dataset-specific characteristics even when each individual study uses a nominal test split. In contrast, cohort studies with thousands of participants or long-term clinical outcomes may report lower performance but provide stronger evidence about real-world variation. This does not invalidate benchmark research, which remains valuable for controlled comparison, but it changes the meaning of the result. The appropriate question is whether an optimization strategy improves performance under rigorous resampling and then retains that benefit on independent data. Future reviews should therefore distinguish proof-of-concept performance, internal validation, temporal validation, geographic validation and prospective evaluation. Such an evidence ladder would make it easier to compare methodological novelty without confusing technical optimization with clinical effectiveness (Detrano et al., 1989; Khan & Algarni, 2020; Riley et al., 2019; Steyerberg et al., 2010; Suri et al., 2022).

How ANFIS compares with modern machine learning is a question the retained evidence answers only in part. Several studies compare ANFIS with classical learners or statistical models such as logistic regression and discriminant analysis (Rajabzadeh et al., 2023; Sumarlinda et al., 2024; Taylan et al., 2023), but few compare it with strong ensemble methods such as random forests (Breiman, 2001) or boosting (Freund & Schapire, 1997), or with support vector machines (Cortes & Vapnik, 1995) and decision trees (Quinlan, 1986), under the same resampling protocol, and fewer still with deep learning. On small tabular datasets such as Cleveland, well-tuned ensembles are demanding baselines, and an advantage reported for an optimized ANFIS over an untuned comparator says little about relative merit (Kuhn & Johnson, 2013). Deep learning dominates where the input is a raw signal such as the electrocardiogram (Rajkomar et al., 2019; Siontis et al., 2021), and the most recent neuro-fuzzy work responds by placing an interpretable fuzzy layer on top of a deep backbone rather than competing with it directly (Narigina et al., 2026). Hybrid metaheuristic designs, including GA-ANFIS, PSO-ANFIS and swarm-tuned variants, sit between these positions, trading higher training cost for wider parameter search while keeping a rule-based output (Dehdar Karsidani et al., 2022; Feng et al., 2021; Singh et al., 2021), and related hybrid fuzzy designs have combined neural networks with fuzzy multi-criteria weighting for heart failure risk (Samuel et al., 2017). Table 14 summarizes these trade-offs. The case for ANFIS therefore rests less on accuracy than on the combination of competitive performance on small structured data with an inspectable rule base, and future studies should demonstrate this with matched comparisons against tuned ensembles and, where signals are used, deep networks.

Table 14: Comparison of ANFIS with alternative approaches for cardiovascular prediction

Approach	Main strength	Main limitation	Typical role
Conventional ANFIS	Rule-based, inspectable output that works on small tabular data	Rule growth with inputs and local optima	Interpretable baseline
Metaheuristic hybrid ANFIS (GA, PSO, swarm)	Wider search over parameters, rules and features	Training cost, overfitting risk and reproducibility	Optimized interpretable models
Ensemble learners (random forest, boosting)	Strong accuracy on tabular data with robust defaults	Less transparent, needs post-hoc explanation	Demanding comparator baseline
Kernel and tree learners (SVM, decision tree)	Well understood and efficient	SVM opaque; single trees unstable	Classical comparators
Deep learning	Learns directly from ECG and other signals	Needs large data and is opaque	Signal-based detection and backbone for neuro-fuzzy layers

THREATS TO VALIDITY AND REVIEW LIMITATIONS

This review has limitations. Search sensitivity depends on terminology, and some relevant papers may describe neuro-fuzzy systems without using ANFIS in searchable fields. Database coverage also differs, particularly for regional journals and conference proceedings, and the requirement for an English title and abstract, together with the exclusion of grey literature, may have omitted relevant regional work. The evidence base itself is a further constraint, since most retained studies relied on small, public or single-centre datasets whose labels do not correspond to acute heart attack, which limits what any review of this literature can conclude about real clinical populations. The retained studies are heterogeneous in target outcome, dataset, validation and reporting, which prevents statistically defensible meta-analysis and means the synthesis is primarily comparative and thematic. Publication bias may favour studies reporting strong improvements, especially when new metaheuristic algorithms are proposed. Quality scoring reduces but does not eliminate reviewer judgement. The PRISMA counts provide a transparent selection narrative, yet they should be understood as the review dataset rather than a universal census of all ANFIS cardiovascular publications. One historically influential hybrid article was marked as retracted and was therefore treated only as contextual evidence, not as support for effectiveness. Finally, several recent 2026 studies have limited citation history and require independent replication. These limitations were mitigated through predefined eligibility criteria, explicit quality questions, separation of foundational references from retained empirical studies, and cautious interpretation of cross-study performance differences (Collins et al., 2015; Higgins et al., 2022; Kitchenham & Charters, 2007; Page et al., 2021; Whiting et al., 2011).

REPRODUCIBILITY AND OPEN RESEARCH

Reproducibility should become a core evaluation criterion for optimized ANFIS studies. At minimum, authors should publish the exact dataset version, inclusion rules, preprocessing code, train-test partitions or cross-validation seeds, feature-selection procedure, membership-function definitions, optimization bounds, population size, iteration count, objective function and software environment. When patient-level data cannot be shared, synthetic examples, data dictionaries and executable pipelines can still expose the analytical workflow, and open tooling that binds preprocessing to model fitting, such as scikit-learn pipelines (Pedregosa et al., 2011), makes leakage easier to avoid and to audit. Repeated metaheuristic runs should be summarized with mean, standard deviation and confidence intervals rather than only the best run. Ablation studies should separate gains from preprocessing, feature selection, optimizer choice and ensemble combination. Model cards or structured reporting templates can document intended use, contraindications, subgroup performance and update procedures. These practices are especially important for clinical prediction because reproducibility is linked to safety: an unstable model can appear strong in a single experiment but fail when retrained or moved to a different population. Open, versioned workflows would allow future systematic reviews to compare evidence at the pipeline level rather than reconstructing incomplete methods from prose (Collins et al., 2015; Page et al., 2021; Siontis et al., 2021; Van Calster et al., 2019; Wolff et al., 2019).

CONCLUSION

This systematic literature review examined 23 empirical studies of adaptive neuro-fuzzy approaches for early heart attack and cardiovascular prediction. The evidence shows a clear evolution from conventional ANFIS toward feature-engineered, metaheuristic-optimized, IoMT-enabled and explainability-aware models. Optimization can improve the search for useful features, membership functions and rule parameters, and several studies report strong discrimination. However, the review also shows that the most important unresolved problem is not raw predictive accuracy. It is the gap between internal benchmark performance and trustworthy clinical generalization. Repeated use of small public datasets whose labels describe angiographic coronary disease rather than acute heart attack, few matched comparisons with ensemble and deep learning methods, incomplete reporting of optimization procedures, limited calibration analysis, scarce external validation and weak prospective evidence restrict confident clinical interpretation. Future work should therefore combine neuro-fuzzy interpretability with rigorous nested validation, uncertainty quantification, calibration, multi-centre testing, subgroup analysis and reproducible implementation. ANFIS remains a promising framework because it can connect adaptive learning with human-readable fuzzy reasoning, but its

clinical value will depend on evidence quality and transparent evaluation rather than increasingly complex optimization alone (Dehdar Karsidani et al., 2022; Jang, 1993; Paul & Karn, 2026; Suri et al., 2022; Taylan et al., 2023).

REFERENCES

1. Bhagat, R. K., Boadh, R., & Kumari, M. (2024). A comparative study of ANFIS and TANFIS hybrid techniques for prediction of heart disease. *African Journal of Biomedical Research, 27*(3S). https://doi.org/10.53555/AJBR.v27i3S.3763
2. Bhuvaneswari Amma, N. G. (2013). An intelligent approach based on principal component analysis and adaptive neuro fuzzy inference system for predicting the risk of cardiovascular diseases. In *2013 Fifth International Conference on Advanced Computing (ICoAC)* (pp. 241-245). IEEE. https://doi.org/10.1109/ICoAC.2013.6921957
3. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321-357. https://doi.org/10.1613/jair.953
5. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis: The TRIPOD statement. *Annals of Internal Medicine, 162*(1), 55-63. https://doi.org/10.7326/M14-0697
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273-297. https://doi.org/10.1007/BF00994018
7. Dehdar Karsidani, S., Farhadian, M., Mahjub, H., & Mozayanimonfared, A. (2022). Intelligent prediction of major adverse cardiovascular events following percutaneous coronary intervention using ANFIS-PSO model. *BMC Cardiovascular Disorders, 22*, Article 389. https://doi.org/10.1186/s12872-022-02825-0
8. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J. J., Sandhu, S., Guppy, K. H., Lee, S., & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology, 64*(5), 304-310. https://doi.org/10.1016/0002-9149(89)90524-9
9. Dua, D., & Graff, C. (2019). *UCI machine learning repository*. University of California, Irvine, School of Information and Computer Sciences.
10. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010
11. Feng, J., Wang, Q., & Li, N. (2021). An intelligent system for heart disease prediction using adaptive neuro-fuzzy inference systems and genetic algorithm. *Journal of Physics: Conference Series, 2010*, Article 012172. https://doi.org/10.1088/1742-6596/2010/1/012172
12. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119-139. https://doi.org/10.1006/jcss.1997.1504
13. Gopi, S., & Arun, M. (2026). Heart disease prediction using adaptive neuro-fuzzy inference system model with feature extraction and data pre-processing. *Journal of Multiscale Modelling*, Article 2640017. https://doi.org/10.1142/S1756973726400172
14. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157-1182.
15. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29-36. https://doi.org/10.1148/radiology.143.1.7063747
16. Harrell, F. E. (2015). *Regression modeling strategies* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-19425-7
17. Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2022). *Cochrane handbook for systematic reviews of interventions* (Version 6.3). Cochrane.

18. Holle, K. F. H. (2016). Diagnosis penyakit jantung menggunakan adaptive neuro-fuzzy inference system [Heart disease diagnosis using adaptive neuro-fuzzy inference system]. *MATICS, 8*(2). https://doi.org/10.18860/mat.v8i2.3537
19. Jang, J.-S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665-685. https://doi.org/10.1109/21.256541
20. Jayachitra, S., Prasanth, A., Haleem, S. L. A., Amin, S. M., & Shaik, K. (2022). An efficient clinical support system for heart disease prediction using TANFIS classifier. *Computational Intelligence, 38*. https://doi.org/10.1111/coin.12487
21. Jeya, R., Abdul Karim, H., Muruganantham, B., Mansor, S., & Tan, Y. F. (2026). Soft computing-based heart disease prediction framework (SC-HDP) using ANFIS. *Journal of Internet Services and Information Security*. https://doi.org/10.58346/JISIS.2026.I2.015
22. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95, International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968
23. Khan, M. A., & Algarni, F. (2020). A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. *IEEE Access, 8*, 122259-122269. https://doi.org/10.1109/ACCESS.2020.3006424
24. Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report EBSE-2007-01). Keele University and Durham University.
25. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137-1145). Morgan Kaufmann.
26. Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. https://doi.org/10.1007/978-1-4614-6849-3
27. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates.
28. Manogaran, G., Varatharajan, R., & Priyan, M. K. (2018). Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. *Multimedia Tools and Applications, 77*(4), 4379-4399. https://doi.org/10.1007/s11042-017-5515-y
29. Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., & Mirjalili, S. M. (2017). Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software, 114*, 163-191. https://doi.org/10.1016/j.advengsoft.2017.07.002
30. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097
31. Narigina, M., Romanovs, A., & Merkuryev, Y. (2026). A neuro-fuzzy digital twin for interpretable cardiac disease recognition. *Applied Sciences, 16*(15), Article 7371. https://doi.org/10.3390/app16157371
32. A novel approach for heart disease prediction using hybridized AITH2O algorithm and SANFIS classifier. (2024). *Network: Computation in Neural Systems, 36*(1). https://doi.org/10.1080/0954898X.2024.2404915
33. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71
34. Paul, B., & Karn, B. (2026). An interpretable hybrid subtractive clustering-based adaptive neuro-fuzzy inference system for heart disease prediction. *Computer and Decision Making: An International Journal, 3*(2S), 171-206.
35. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

36. Pekaslan, D., Chen, C., Wagner, C., & Garibaldi, J. M. (2020). Performance and interpretability in fuzzy logic systems: Can we have both? In *Information processing and management of uncertainty in knowledge-based systems* (pp. 571-584). Springer. https://doi.org/10.1007/978-3-030-50146-4_42
37. Polatgil, M. (2020). Detection of heart disease with ANFIS and fuzzy c-means. *Bilişim Teknolojileri Dergisi, 13*(4).
38. Puspita, I., & Abadi, A. M. (2019). Diagnosing heart disease using wavelet transformation and adaptive neuro fuzzy inference system based on electrocardiogram. *EKSAKTA: Journal of Sciences and Data Analysis, 19*(1). https://doi.org/10.20885/eksakta.vol19.iss1.art7
39. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81-106. https://doi.org/10.1007/BF00116251
40. Rajabzadeh, Z., Akbari Sharak, N., Esmaeili, H., Ghayour-Mobarhan, M., Shakeri, M. T., & Pasdar, Z. (2023). Assessment of the performances of adaptive neuro-fuzzy inference system and two statistical methods for diagnosing coronary artery disease. *Medical Journal of the Islamic Republic of Iran, 37*, 392-398.
41. Rajinikanth, N., & Pavithra, L. (2019). Heart disease prediction and classification using optimization based adaptive neuro fuzzy inference system. *Journal of Emerging Technologies and Innovative Research, 6*(3), 677-683.
42. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259
43. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778
44. Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: Part II. Binary and time-to-event outcomes. *Statistics in Medicine, 38*(7), 1276-1296. https://doi.org/10.1002/sim.7992
45. Samuel, O. W., Asogbon, G. M., Sangaiah, A. K., Fang, P., & Li, G. (2017). An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Systems with Applications, 68*, 163-172. https://doi.org/10.1016/j.eswa.2016.10.020
46. Shadhan, N. A.-H. T., Ali, B. K., & Al-Azzawi, S. F. (2026). Breakpoint adaptive neural fuzzy inference regression model for predicting the risk of cardiovascular diseases complications. *Letters in Biomathematics, 13*(2).
47. Singh, P., Kaur, A., Batth, R. S., Kaur, S., & Gianini, G. (2021). Multi-disease big data analysis using beetle swarm optimization and an adaptive neuro-fuzzy inference system. *Neural Computing and Applications, 33*, 12745-12765. https://doi.org/10.1007/s00521-021-05798-x
48. Siontis, K. C., Noseworthy, P. A., Attia, Z. I., & Friedman, P. A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology, 18*(7), 465-478. https://doi.org/10.1038/s41569-020-00503-2
49. Sood, S. K., Kaur, S., & Chahal, K. K. (2020). An intelligent framework for monitoring dengue fever risk using LDA-ANFIS. *Journal of Ambient Intelligence and Smart Environments, 12*(1), 5-20. https://doi.org/10.3233/AIS-200547
50. Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology, 21*(1), 128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2
51. Sumarlinda, S., Rahmat, A. B., Awang Long, Z. B., & Lestari, W. (2024). The comparative early prediction model for cardiovascular disease using machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 10*(1).
52. Suri, J. S., Bhagawati, M., Paul, S., Protogeron, A., Sfikakis, P., Kitas, G., et al. (2022). Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review. *Computers in Biology and Medicine, 142*, Article 105204. https://doi.org/10.1016/j.compbiomed.2021.105204
53. 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), Article 117. https://doi.org/10.3390/biology12010117

54. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44-56. https://doi.org/10.1038/s41591-018-0300-7
55. Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*, Article 230. https://doi.org/10.1186/s12916-019-1466-7
56. Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*(8), 529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
57. Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51-58. https://doi.org/10.7326/M18-1376
58. World Health Organization. (2021). *Cardiovascular diseases (CVDs)* [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
59. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X
60. Ziasabounchi, N., & Askerzade, I. (2014). ANFIS based classification model for heart disease prediction. *International Journal of Electrical & Computer Sciences, 14*(2), 7-12.