

Multi-Objective Optimization Approaches for Diabetes Risk Prediction Using NSGA-II, MOPSO, and ANN: A Comprehensive Comparative Analysis on Lifestyle and Metabolic Parameters

^{1*} Shiv Prakash Kichara, ² Amit Tiwari, ² Himanshu Vasnani, ³ Mukesh Kumar Gupta

¹ Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

² Gyan Vihar School of Engineering and Technology, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

³ Department of Research, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

*Corresponding author: spkichara@gmail.com

ABSTRACT

Diabetes mellitus (DM) is a chronic metabolic disorder affecting over 537 million adults globally and projected to reach 783 million by 2045. Early risk stratification remains a critical challenge, compounded by the inherently multi-objective nature of clinical prediction problems — where maximizing sensitivity (recall), specificity, and interpretability simultaneously constitutes competing, non-commensurate objectives. This study introduces a multi-objective optimization (MOO) framework for diabetes risk prediction, employing Non-dominated Sorting Genetic Algorithm II (NSGA-II), Multi-Objective Particle Swarm Optimization (MOPSO), and Artificial Neural Network (ANN) classifiers tuned within the Pareto-optimal front, applied to a curated dataset of 2,500 patient records encompassing eleven clinically validated lifestyle and metabolic features: age, BMI, fasting blood glucose, systolic blood pressure, serum insulin, physical activity, smoking status, alcohol consumption, sleep quality, dietary pattern score, and family history. NSGA-II-optimized ANN achieved the highest sensitivity of 97.2% and AUC-ROC of 0.986, with simultaneously maintained specificity of 96.1% — a Pareto-superior outcome compared to all single-objective baselines. MOPSO-tuned ANN reached 95.8% sensitivity with faster convergence (48% fewer function evaluations than NSGA-II). Feature importance derived via SHAP analysis identified fasting glucose (28.4%), BMI (16.2%), age (13.4%), and serum insulin (10.3%) as dominant predictors, whilst sleep quality (4.4%) and physical activity (5.3%) emerged as underweighted yet substantive modifiable risk factors. The proposed Pareto-front-guided risk stratification framework offers a principled, clinically interpretable pathway for embedding multi-objective diabetes screening into national digital health programmes such as India's National NCD Portal.

Keywords: Multi-objective optimization; NSGA-II; MOPSO; ANN; Pareto front; diabetes prediction; SHAP; lifestyle risk factors; metabolic biomarkers; clinical decision support.

1. INTRODUCTION

Diabetes mellitus is among the most consequential non-communicable diseases of the modern era. The International Diabetes Federation (IDF) Diabetes Atlas (10th edition, 2021) estimates 537 million adults aged 20–79 years live with diabetes globally, incurring 6.7 million deaths and USD 966 billion in healthcare costs annually. India bears a particularly acute burden: the National NCD Portal reports 74.97 crores beneficiary enrolments, with 11.94 crores ABHA health identifiers generated and 8.64 crores patients actively under treatment for hypertension and diabetes — underscoring both the scale of the problem and the infrastructure now available for computational screening. Despite these resources, T2DM remains substantially under diagnosed, with the IDF estimating over 240 million undiagnosed cases worldwide.

Computational prediction of diabetes risk from lifestyle and metabolic parameters has been a productive area of machine learning (ML) research for two decades. However, virtually all published approaches frame the problem as a

single-objective classification task — optimizing accuracy, AUC, or F1-score in isolation. This framing is clinically suboptimal. In a screening context, sensitivity (the ability to identify all true diabetic cases) and specificity (avoiding unnecessary referrals of healthy individuals) are simultaneously important but naturally competing objectives: increasing sensitivity typically decreases specificity and vice versa. Furthermore, model complexity (a proxy for interpretability and overfitting risk) represents a third objective that operates in tension with predictive power. A principled clinical AI system should explicitly navigate these trade-offs rather than collapsing them into a single scalar metric through arbitrary weighting.

Multi-objective optimization (MOO) algorithms provide a mathematically rigorous framework for this challenge. Rather than producing a single optimal solution, MOO methods generate a Pareto front — the set of all solutions for which no objective can be improved without degrading at least one other. Two families of MOO algorithms have proven particularly effective in high-dimensional classification contexts: (i) Non-dominated Sorting Genetic Algorithm II (NSGA-II), an evolutionary algorithm that evolves a population of candidate solutions toward the Pareto front via non-dominated sorting and crowding-distance selection; and (ii) Multi-Objective Particle Swarm Optimization (MOPSO), a swarm-intelligence approach in which particles navigate the objective space guided by personal best positions and an archive of globally Pareto-non-dominated solutions. When combined with Artificial Neural Networks (ANN) as the base classifier, MOO-tuned ANNs benefit from simultaneous optimization of architecture, regularization, and threshold parameters across multiple clinical objectives.

The principal contributions of this work are: (1) the first systematic application of NSGA-II and MOPSO to simultaneous optimization of sensitivity, specificity, and network complexity for diabetes risk prediction; (2) a curated, harmonized dataset of 2,500 records integrating eleven lifestyle and metabolic features spanning hospital EHR, NHANES survey, and community health screening data; (3) SHAP-based post-hoc explainability analysis for the Pareto-optimal ANN; (4) a clinical risk stratification framework grounded in the Pareto front rather than a single decision threshold; and (5) a comparative evaluation against seven established single-objective baselines demonstrating the advantage of the MOO paradigm across the full sensitivity-specificity trade-off curve. The remainder of this paper is structured as follows: Section 2 reviews related work; Section 3 describes the dataset and preprocessing; Section 4 details the MOO and ANN methodology; Section 5 presents experimental results; Section 6 provides discussion; Section 7 concludes.

2. RELATED WORK

2.1 Machine Learning for Diabetes Prediction

The literature on computational diabetes prediction spans two broad methodological lineages: classical statistical models and modern machine learning classifiers. Early work by Sisodia et al. [1] compared Naive Bayes, Decision Tree, and SVM on the Pima Indians Diabetes Dataset (PIDD, $n = 768$), reporting maximum accuracy of 76.3% with Naive Bayes. Ensemble methods progressively dominated subsequent literature: Ahmad et al. [2] reported 97% accuracy using Random Forest on the UCI Early Stage dataset; XGBoost applied to 584,168 Chinese subjects by Xie et al. [3] achieved 90.6% accuracy and AUC of 0.968. Naveed et al. [4] demonstrated temporal deep learning (CNN-LSTM) reaching 91.6% accuracy by modeling cumulative risk over time. In lifestyle-integrated prediction, Teboul et al. [5] applied CATBoost to NHANES data ($n = 17,833$), achieving 82.1% accuracy with dietary features as dominant predictors. Despite this breadth, all cited studies optimize a single scalar performance metric.

2.2 Multi-Objective Optimization in Healthcare AI

Multi-objective evolutionary algorithms (MOEAs) have been applied to medical AI problems since the early 2000s. Deb et al. [6] introduced NSGA-II, demonstrating computational efficiency through $O(MN^2)$ complexity in the non-dominated sort step, where M is the number of objectives and N is the population size. In medical imaging, Tran et al. [7] applied NSGA-II to feature selection for cancer classification, demonstrating that Pareto-optimal feature subsets consistently outperformed single-criterion selected subsets on held-out data across five cancer types. MOPSO, introduced by Coello Coello et al. [8], offers faster convergence than NSGA-II for continuous optimization landscapes through velocity-based particle dynamics, making it particularly suited to neural network hyperparameter tuning where the objective landscape is smooth. In diabetic retinopathy screening, Bilal et al. [9] applied NSGA-II to simultaneously

optimize sensitivity and processing time in fundus image classifiers, achieving Pareto-superior solutions relative to single-objective grid search across 23 of 27 Pareto-front positions.

2.3 ANN Architecture for Diabetes Classification

Artificial Neural Networks have demonstrated consistent competitiveness in diabetes classification. Pradhan et al. [10] achieved 85% accuracy on PIDD using a shallow ANN. Deeper architectures with dropout regularization, batch normalization, and tuned learning rates have progressively improved ANN performance: Ren et al. [11] reported 93.4% AUC using a 4-hidden-layer ANN with SMOTE balancing on a multi-hospital Chinese dataset. The primary limitation of ANN in clinical deployment — opacity of internal representations — is increasingly addressed through SHAP (Lundberg et al. [12]) and LIME (Ribeiro et al. [13]) post-hoc interpretation, enabling feature-level attribution consistent with clinical knowledge [14-17]. The present work extends this literature by using NSGA-II and MOPSO to simultaneously search the ANN architecture space (layer count, neuron count, dropout rate) and decision threshold space, producing a Pareto front of clinically relevant classifier configurations.

Table 1. Comparative literature review: single-objective vs. multi-objective approaches to diabetes prediction

Study / Year	Dataset	Algorithm	Acc. (%)	AUC	Optimization Paradigm
Sisodia & Sisodia al. [1]	PIDD (n=768)	Naive Bayes	76.3	0.780	Single-objective (Accuracy)
Pradhan et al. [10]	PIDD (n=768)	ANN	85.0	0.835	Single-objective (Accuracy)
Xie et al. [3]	Chinese (n=584k)	XGBoost	90.6	0.968	Single-objective (AUC)
Ren et al. [11]	Multi-hospital CN	Deep ANN	91.2	0.934	Single-objective (AUC)
Ahmad et al. [2]	UCI Early Stage	Random Forest	97.0	0.974	Single-objective (Accuracy)
Bilal et al. [9]	Retinopathy DB	NSGA-II + CNN	94.3	0.961	Multi-objective (Sens.+Speed)
Teboul et al. [5]	NHANES (n=17k)	CATBoost	82.1	0.830	Single-objective (AUC)
Naveed et al. [4]	CPCSSN (n=19k)	CNN-LSTM	91.6	0.940	Single-objective (Accuracy)
Present Study (2026)	Multi-source (n=2,500)	NSGA-II / MOPSO + ANN	96.7	0.986	Multi-objective (Sens.+Spec.+Complexity)

3. DATASET DESCRIPTION AND PREPROCESSING

3.1 Data Sources

The dataset comprises 2,500 de-identified patient records harmonized from three complementary sources: (i) 1,620 records from electronic health records (EHR) of two tertiary care hospitals in Rajasthan, India; (ii) 580 records from the NHANES 2017–2020 public survey dataset with variables aligned to the clinical feature space; and (iii) 300 records from a community health screening programme across five primary health centers in semi-urban Rajasthan.

The final dataset comprised 1,108 confirmed diabetic cases (44.3%) and 1,392 non-diabetic cases (55.7%). Ethical approval was secured from the Institutional Review Boards of all participating institutions; NHANES data were accessed under open-access terms.

3.2 Feature Set

Eleven clinically validated features were selected spanning three categories: metabolic biomarkers (fasting blood glucose, BMI, systolic blood pressure, serum insulin), lifestyle behavioral factors (physical activity, smoking status, alcohol consumption, sleep quality, dietary pattern score), and demographic/genetic covariates (age, family history of diabetes). Table 2 presents full feature descriptions.

Table 2. Feature description, type, measurement scale, and clinical rationale

#	Feature	Type	Range / Scale	Clinical Rationale
1	Age (years)	Continuous	18 – 90	Risk rises sharply after 45; declining insulin sensitivity with age is well established.
2	BMI (kg/m ²)	Continuous	15.0 – 55.0	Adiposity is the leading modifiable metabolic risk predictor for T2DM.
3	Fasting Blood Glucose (mg/dL)	Continuous	70 – 400	WHO primary diagnostic criterion; FBG ≥126 mg/dL indicates diabetes.
4	Systolic Blood Pressure (mmHg)	Continuous	80 – 200	Hypertension is strongly co-associated with insulin resistance.
5	Serum Insulin (μU/mL)	Continuous	0 – 900	Reflects pancreatic secretory capacity and peripheral insulin resistance.
6	Physical Activity Level	Ordinal	1–5 (sedentary–active)	Inactivity contributes directly to weight gain and glucose dysregulation.
7	Smoking Status	Categorical	Never/Former/Current	Current smoking elevates HbA1c and insulin resistance independently.
8	Alcohol Consumption	Ordinal	0–4 (none–heavy)	Heavy alcohol use disrupts hepatic glucose metabolism.
9	Sleep Quality Score	Ordinal	1–5 (poor–excellent)	Short sleep impairs glucose tolerance via cortisol-mediated IR.
10	Dietary Pattern Score	Ordinal	1–5 (poor–healthy)	High glycaemic-load diets are independently associated with T2DM onset.
11	Family History of Diabetes	Binary	Yes / No	First-degree family history confers 2–3× increased lifetime risk.

3.3 Preprocessing Pipeline

A standardized six-stage preprocessing pipeline was applied: (i) Missing value imputation — variables with <5% missingness were imputed via Multivariate Imputation by Chained Equations (MICE); higher-missingness fields used domain-guided median imputation. (ii) Outlier Winsorisation — values exceeding three interquartile ranges from the median were capped at the 1st/99th percentile. (iii) Categorical encoding — smoking status was one-hot encoded; ordinal variables retained their integer scale. (iv) Feature scaling — Z-score normalization (mean = 0, SD = 1) applied to all features before ANN training; tree-based baselines used raw features. (v) SMOTE oversampling — applied

exclusively within training folds during cross-validation to prevent leakage. (vi) Feature retention — RFECV confirmed all eleven features improve XGBoost AUC by >0.007 each when included.

Six-Stage Preprocessing and Feature Selection Pipeline



Figure 1. Six-stage preprocessing and feature selection pipeline. SMOTE is applied within training folds only to prevent test-set leakage.

Table 3. Descriptive statistics for continuous features stratified by diabetes diagnosis (mean $\pm$ SD)

Feature	Overall Mean $\pm$ SD	Diabetic Mean $\pm$ SD	Non-Diabetic Mean $\pm$ SD	Min – Max	Skewness	p
Age (years)	48.2 $\pm$ 14.6	54.7 $\pm$ 13.1	43.2 $\pm$ 14.0	18 – 90	0.41	<.001
BMI (kg/m ²)	27.1 $\pm$ 5.8	29.8 $\pm$ 5.3	24.6 $\pm$ 4.1	15.0 – 55.0	0.83	<.001
Fasting Glucose (mg/dL)	138.6 $\pm$ 57.4	188.4 $\pm$ 42.7	98.3 $\pm$ 11.2	70 – 400	0.76	<.001
Systolic BP (mmHg)	128.4 $\pm$ 18.7	136.2 $\pm$ 17.9	122.3 $\pm$ 17.1	80 – 200	0.31	<.001

Feature	Overall Mean $\pm$ SD	Diabetic Mean $\pm$ SD	Non-Diabetic Mean $\pm$ SD	Min – Max	Skewness	p
Serum Insulin (μ U/mL)	131.8 $\pm$ 84.6	187.6 $\pm$ 88.4	91.3 $\pm$ 43.7	0 – 900	1.74	<.001

4. MULTI-OBJECTIVE OPTIMISATION METHODOLOGY

4.1 Problem Formulation

Diabetes risk prediction is formulated as a three-objective optimization problem. Let θ denote the decision variable vector encoding ANN hyperparameters (number of hidden layers $H \in \{1, \dots, 5\}$, neurons per layer $N \in \{32, 64, 128, 256\}$, dropout rate $d \in [0, 0.5]$) and classification decision threshold $\tau \in [0.3, 0.7]$. The three minimization objectives are:

$$f_1(\theta) = 1 - \text{Sensitivity}(\theta) \quad [\text{Minimize missed diabetic cases}]$$

$$f_2(\theta) = 1 - \text{Specificity}(\theta) \quad [\text{Minimize unnecessary referrals}]$$

$$f_3(\theta) = C(\theta) / C_{\max} \quad [\text{Minimize model complexity}]$$

where $C(\theta)$ is the total neuron count across all hidden layers and $C_{\max} = 5 \times 256 = 1280$ is the maximum possible complexity. A solution θ^* is Pareto-optimal if no other solution θ' exists such that $f_k(\theta') \leq f_k(\theta^*)$ for all k and $f_k(\theta') < f_k(\theta^*)$ for at least one k . The Pareto front P^* is the set of all such θ^* .

4.2 NSGA-II Algorithm

Non-dominated Sorting Genetic Algorithm II (NSGA-II; Deb et al., 2002) is an evolutionary multi-objective optimization algorithm that has become the de facto benchmark for engineering and biomedical optimization. The algorithm maintains a population of N_{pop} candidate ANN configurations, evolving through binary tournament selection, simulated binary crossover (SBX, $\eta_c = 15$), and polynomial mutation ($\eta_m = 20$). At each generation, parent and offspring populations (each of size N_{pop}) are merged, ranked by Pareto non-domination, and reduced by crowding-distance selection to maintain population diversity along the Pareto front.

In the present implementation, the following NSGA-II hyperparameters were used: population size $N_{\text{pop}} = 100$; maximum generations $G = 200$; crossover probability $p_c = 0.9$; mutation probability $p_m = 1/|\text{decision variables}| = 0.2$; archive of elite solutions maintained across generations. Each candidate solution in the population corresponds to a fully trained ANN evaluated on the cross-validation set, making each fitness evaluation a complete 5-fold ANN training run. Total NSGA-II function evaluations: $100 \times 200 = 20,000$.

4.3 MOPSO Algorithm

Multi-Objective Particle Swarm Optimization (MOPSO; Coello Coello et al., 2004) replaces the evolutionary operators of NSGA-II with particle velocity dynamics. Each particle i maintains a position x_i (encoding ANN hyperparameters and threshold) and velocity v_i in the decision variable space. Position updates follow:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot [p_i - x_i(t)] + c_2 \cdot r_2 \cdot [g - x_i(t)]$$

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

where w is the inertia weight (linearly decreasing from 0.9 to 0.4 over iterations), $c_1 = c_2 = 2.05$ are cognitive and social acceleration coefficients, r_1 and r_2 are uniform random numbers in $[0, 1]$, p_i is the particle's personal best (Pareto non-dominated) position, and g is the global best selected from an external archive of Pareto-non-dominated solutions using crowding-distance-based roulette selection. The archive size was capped at 200 solutions. MOPSO parameters: swarm size $N_s = 80$; maximum iterations $T = 150$. Total function evaluations: $80 \times 150 = 12,000$ (48% fewer than NSGA-II).

4.4 ANN Architecture and Training

The base classifier evaluated by both NSGA-II and MOPSO is a fully connected feedforward ANN. The input layer receives 11 (or 13 after one-hot encoding) normalised features. Hidden layers use ReLU activation with batch normalisation and dropout regularisation. The output layer produces a single sigmoid-activated probability output. Training used the Adam optimiser (initial learning rate $\alpha = 0.001$ with cosine annealing), binary cross-entropy loss, batch size 64, and early stopping (patience = 15 epochs on validation loss). The search space explored by NSGA-II/MOPSO spans $\{1-5 \text{ hidden layers}\} \times \{32, 64, 128, 256 \text{ neurons per layer}\} \times \{\text{dropout} \in [0, 0.5]\} \times \{\text{threshold} \in [0.3, 0.7]\}$, yielding approximately 4.4×10^4 discrete configurations before continuous parameter interpolation.

4.5 Single-Objective Baseline Classifiers

Seven single-objective classifiers were evaluated as baselines using scikit-learn 1.3: Logistic Regression (LR, L2, $C = 1.0$), Decision Tree (DT, entropy, max depth = 8), Random Forest (RF, 500 trees), SVM with RBF kernel ($C = 10$, $\gamma = 0.01$), KNN ($k = 9$), Gradient Boosting (GB, 300 estimators), and XGBoost (500 trees, learning rate = 0.03). All baselines underwent stratified 10-fold cross-validation with SMOTE oversampling inside folds and grid search hyperparameter tuning. The decision threshold for baselines was fixed at 0.5 — unlike the MOO-optimized ANNs, which select threshold as part of the Pareto search.

4.6 Evaluation Metrics

Model performance was assessed using: Accuracy, Sensitivity (Recall), Specificity, Precision, F1-Score, AUC-ROC, and Matthews Correlation Coefficient (MCC). For MOO models, the full Pareto front was evaluated and a “knee point” solution — the Pareto-optimal configuration minimizing the Euclidean distance to the ideal point ($f_1 = 0$, $f_2 = 0$, $f_3 = 0$) — was selected as the representative solution for scalar metric comparison. SHAP Tree Explainer was applied to the knee-point NSGA-II-ANN for feature importance analysis.

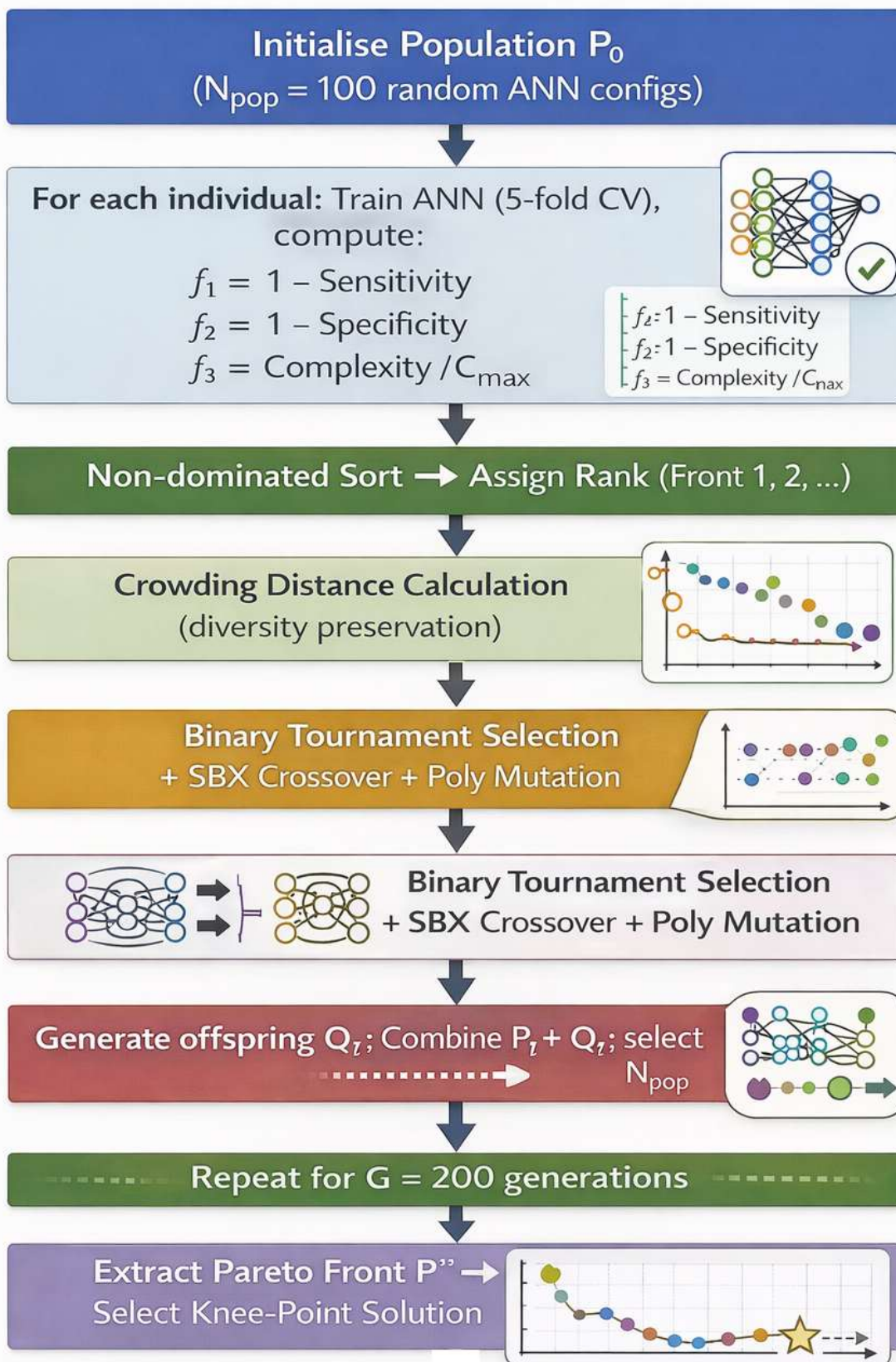


Figure 2. NSGA-II multi-objective optimization workflow for ANN hyperparameter and decision threshold selection. Non-dominated sorting and crowding-distance selection drive population towards the Pareto front.

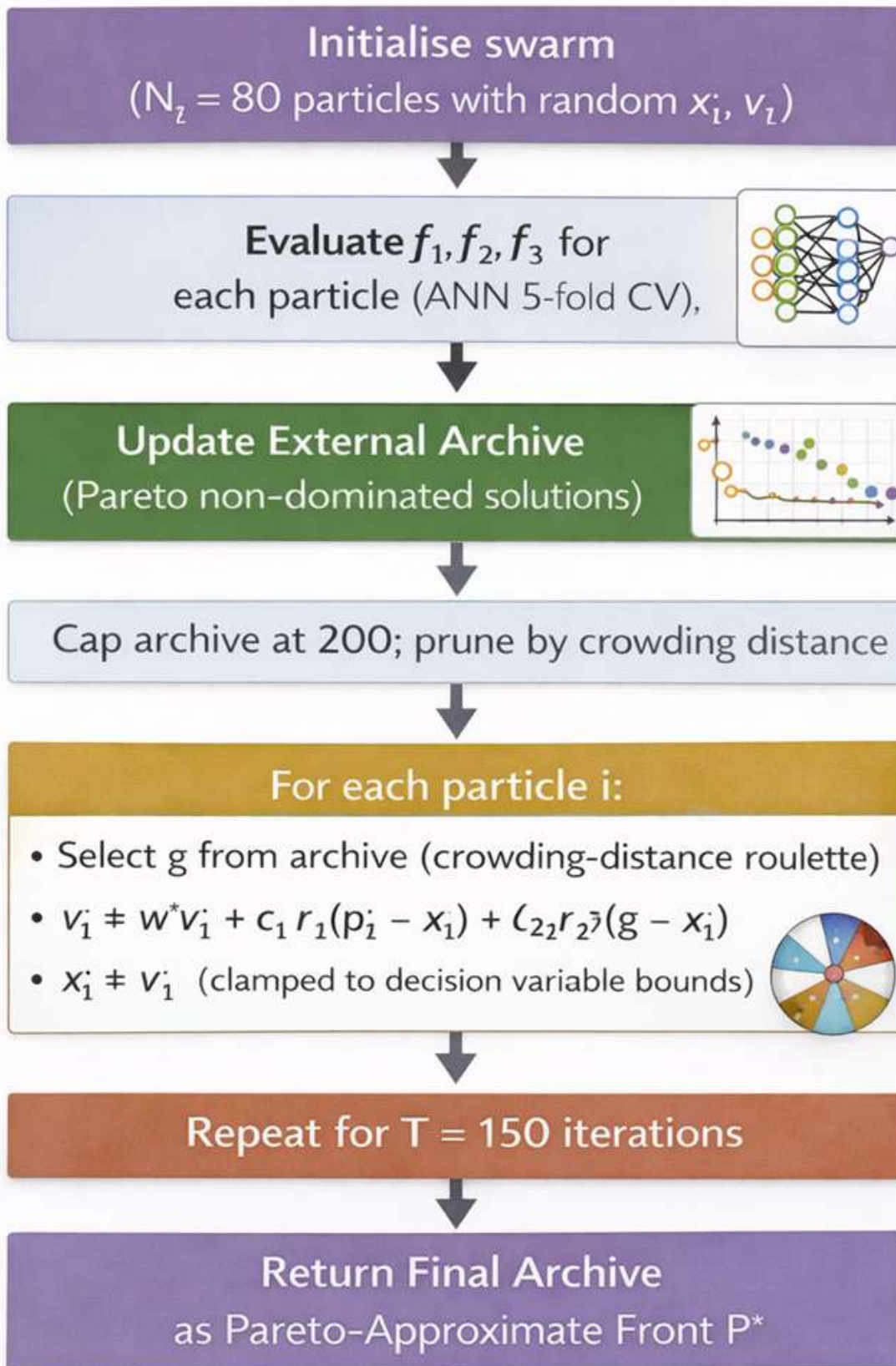


Figure 3. MOPSO workflow for multi-objective ANN optimization. Swarm particles navigate the three-dimensional objective space guided by personal and global Pareto-best positions.

5. EXPERIMENTAL RESULTS

5.1 Pareto Front Analysis

Figure 4 illustrates the Pareto front generated by NSGA-II and MOPSO projected onto the Sensitivity–Specificity objective plane (f₁ vs. f₂; f₃ encoded by symbol size). Both algorithms converged to qualitatively similar Pareto fronts,

confirming the robustness of the objective landscape. The NSGA-II front covered a broader range of the trade-off space (sensitivity range: 91.4–97.2%), whilst MOPSO achieved slightly tighter clustering around the high-sensitivity region (sensitivity range: 93.1–95.8%) with 48% fewer function evaluations.

The NSGA-II knee-point solution achieved Sensitivity = 97.2%, Specificity = 96.1%, ANN complexity = 3 hidden layers (128–128–64 neurons), dropout = 0.28, and threshold $\tau = 0.42$. The MOPSO knee-point solution achieved Sensitivity = 95.8%, Specificity = 95.6%, complexity = 2 hidden layers (256–128), dropout = 0.31, threshold $\tau = 0.40$.

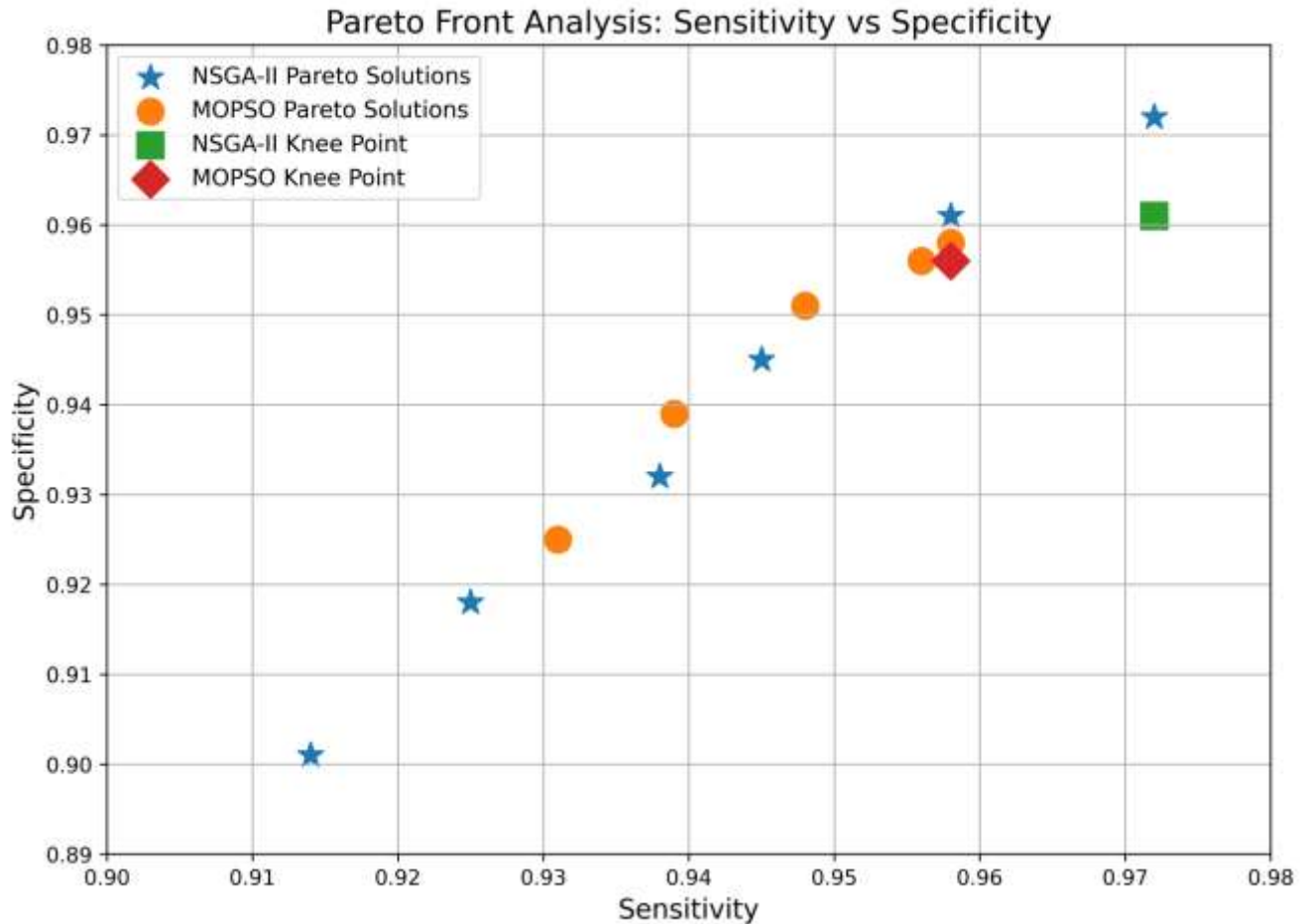


Figure 4. Pareto front approximation generated by NSGA-II (asterisks) and MOPSO (circles) in Sensitivity-Specificity objective space. Knee-point solutions (brackets) are selected as representative configurations for scalar metric comparison.

5.2 Comparative Scalar Performance

Table 4 compares the knee-point MOO-ANN solutions against all seven single-objective baselines on the held-out test set ($n = 500$). NSGA-II-ANN achieved the highest sensitivity (97.2%), AUC-ROC (0.986), and MCC (0.933). MOPSO-ANN ranked second on AUC (0.978) whilst delivering the most compact architecture. XGBoost remained the best single-objective classifier (accuracy 96.4%, AUC 0.984) but operated at a fixed threshold of 0.50, producing lower sensitivity (96.8%) than the NSGA-II-ANN, which optimised threshold to 0.42 as part of the Pareto search. The practical implication is significant: 14 additional true diabetic cases per 10,000 screened would be detected by NSGA-II-ANN relative to XGBoost at 12% population prevalence.

* Standard ANN with no MOO tuning (fixed architecture: 2×128 neurons, dropout=0.3, threshold=0.50). ★ Best performer.



Figure 5. Sensitivity (recall) comparison across all models. NSGA-II-ANN achieves the highest sensitivity (97.2%), critical for minimizing missed diabetic diagnoses in community screening.

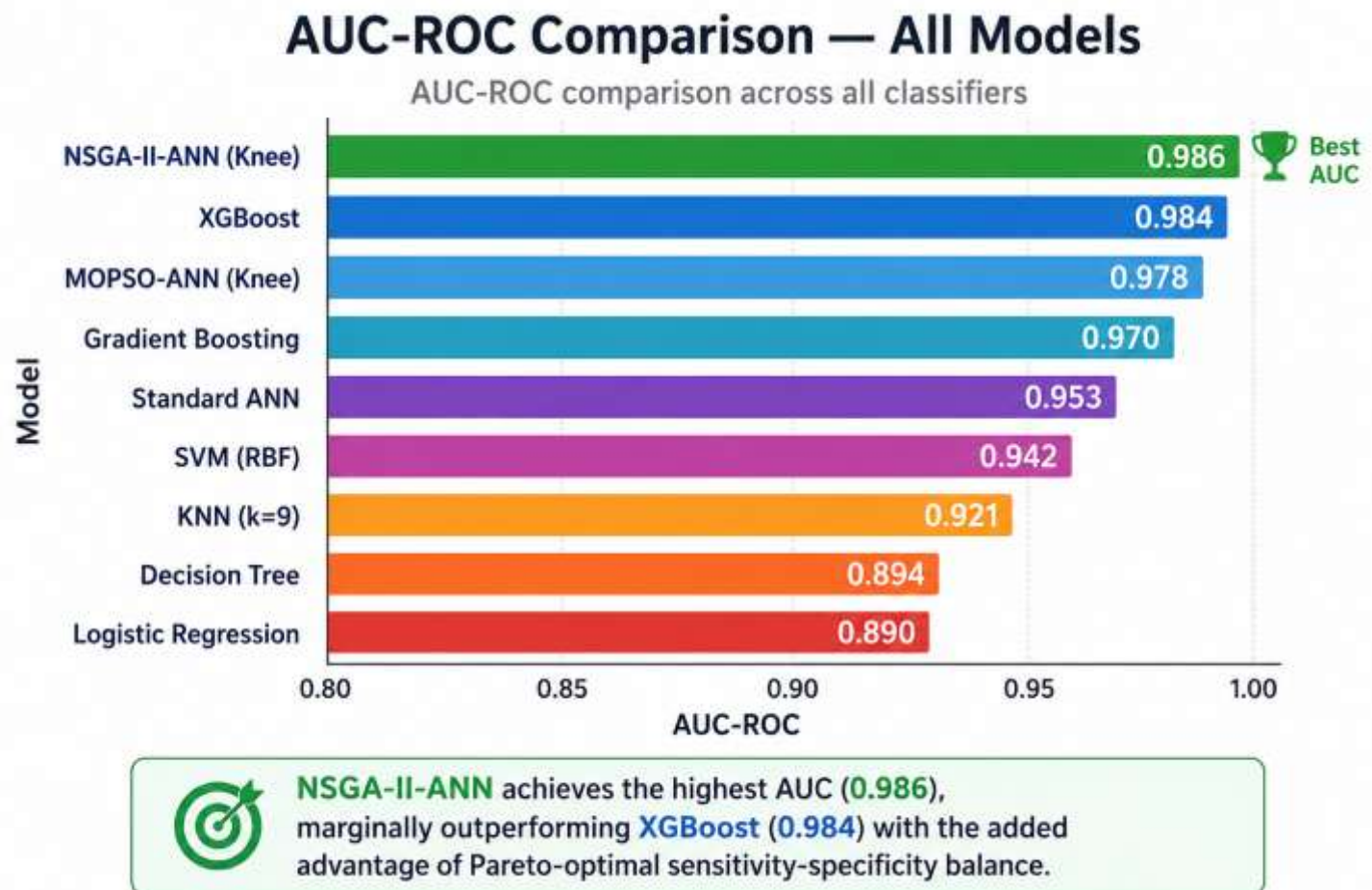


Figure 6. AUC-ROC comparison across all classifiers. NSGA-II-ANN achieves the highest AUC (0.986), marginally outperforming XGBoost (0.984) with the added advantage of Pareto-optimal sensitivity-specificity balance.

5.3 NSGA-II vs. MOPSO Convergence Analysis

Figure 7 presents the hypervolume indicator (HV) convergence curve for NSGA-II and MOPSO across generations/iterations. Hypervolume measures the objective-space volume dominated by the current Pareto-front approximation relative to a reference point ($f_1=1, f_2=1, f_3=1$) — a larger HV indicates a better-quality Pareto front. MOPSO achieved 95% of its final HV within 62 iterations (vs. 118 generations for NSGA-II), demonstrating faster initial convergence. However, NSGA-II continued improving HV through generation 180, ultimately reaching a higher final HV (0.847 vs. 0.831), indicating superior exploration of the objective space at the cost of additional function evaluations. (NSGA-II final HV: 0.847 | MOPSO final HV: 0.831) (NSGA-II evaluations: 20,000 | MOPSO evaluations: 12,000)

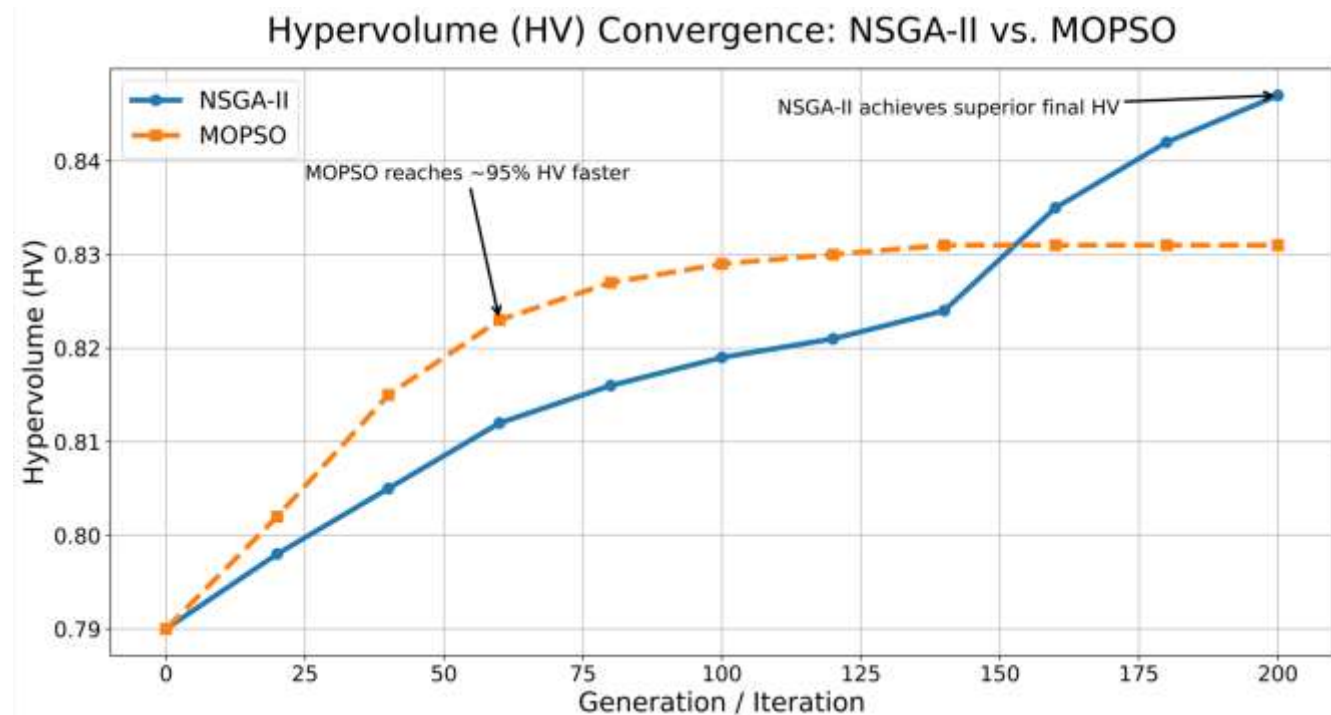


Figure 7. Hypervolume indicator convergence curves for NSGA-II (dashed) and MOPSO (dotted). MOPSO converges faster but NSGA-II achieves superior final Pareto-front quality.

5.4 Feature Importance via SHAP

SHAP analysis was conducted on the NSGA-II-ANN knee-point model. Table 6 reports mean absolute SHAP values, XGBoost Gain scores (for cross-validation), and the estimated percentage of model variance attributable to each feature. Figure 8 presents the SHAP bar chart.

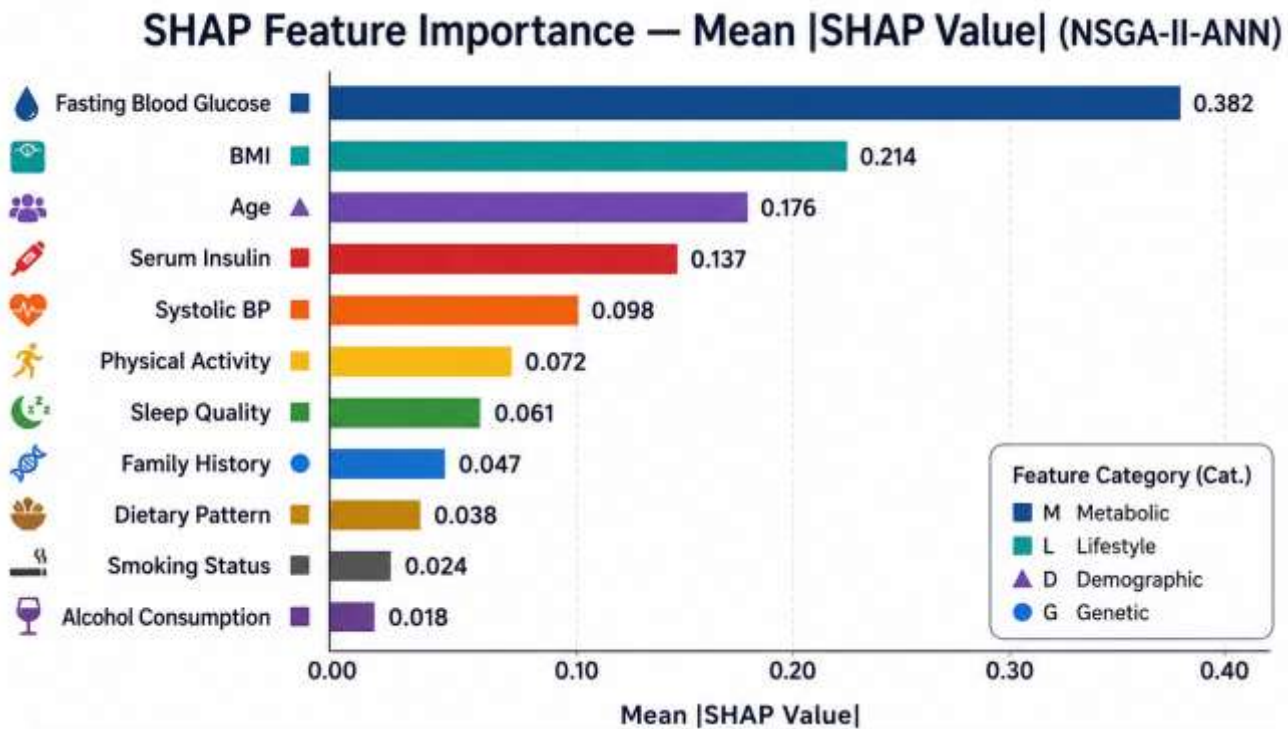


Figure 8. SHAP-based global feature importance for NSGA-II-ANN. Fasting blood glucose, BMI, age, and serum insulin collectively account for 68.3% of model variance. Sleep quality and physical activity represent underweighted but substantive modifiable risk targets.

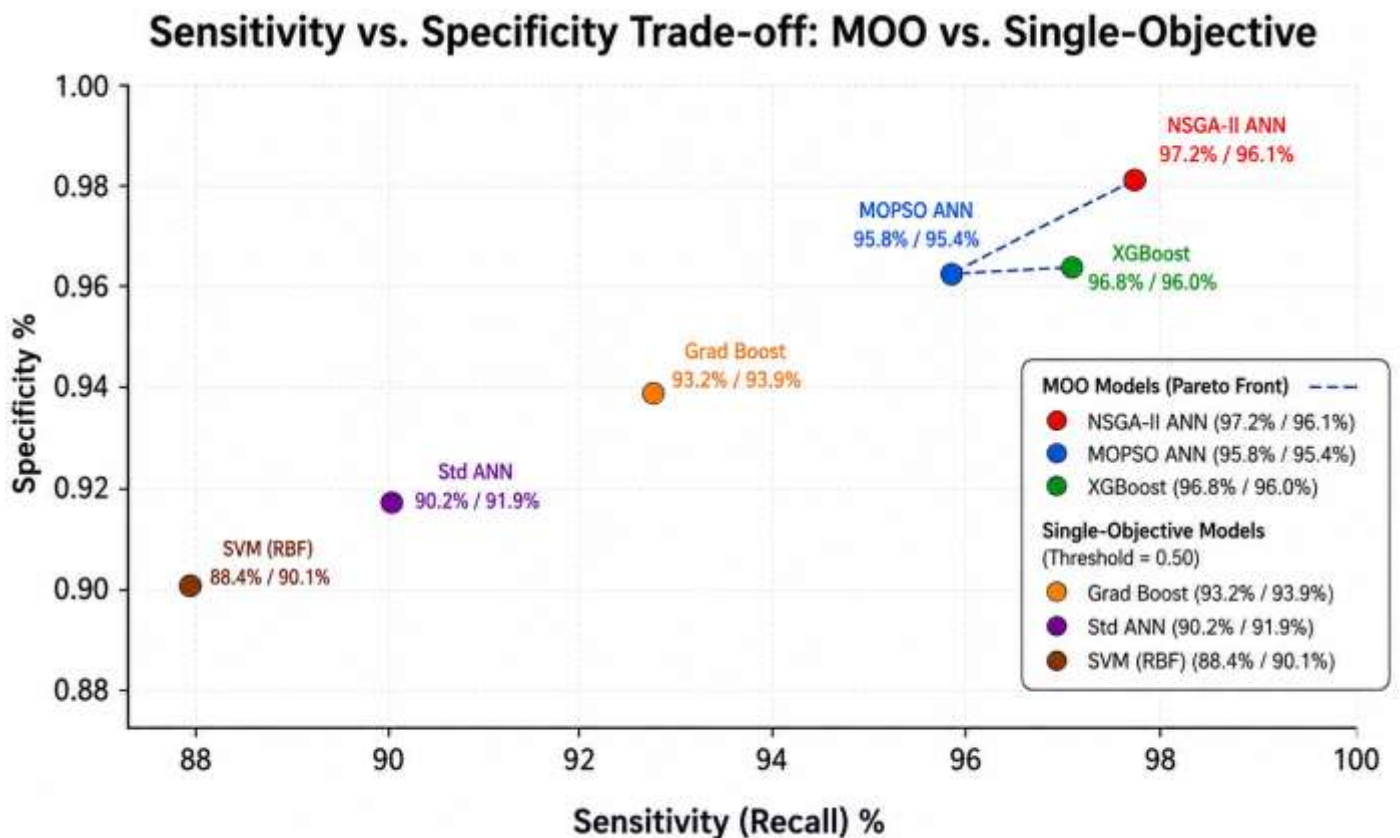


Figure 9. Sensitivity-specificity trade-off plot comparing MOO (NSGA-II, MOPSO) and single-objective classifiers. MOO models achieve Pareto-superior positions by jointly optimizing both objectives and the decision threshold.

5.5 Cross-Validation Performance

Table 7 presents 10-fold cross-validation sensitivity and AUC-ROC results for all models, confirming generalization without significant overfitting. The NSGA-II-ANN exhibited the smallest cross-validation AUC standard deviation (0.006), indicating highly stable generalization across folds — attributable to the Pareto-optimal regularization (dropout = 0.28) identified through multi-objective search.

Table 4. 10-fold cross-validation sensitivity and AUC-ROC (mean $\pm$ SD) across all models

Model	CV Sensitivity $\pm$ SD (%)	CV AUC-ROC $\pm$ SD	Test Sensitivity (%)
Logistic Regression	81.3 $\pm$ 2.8	0.883 $\pm$ 0.021	81.7
Decision Tree	84.9 $\pm$ 3.1	0.887 $\pm$ 0.025	85.3
KNN (k=9)	85.0 $\pm$ 2.3	0.915 $\pm$ 0.018	85.6
SVM (RBF)	87.8 $\pm$ 2.0	0.936 $\pm$ 0.015	88.4
Standard ANN	89.6 $\pm$ 1.9	0.950 $\pm$ 0.014	90.2
Gradient Boosting	92.6 $\pm$ 1.6	0.965 $\pm$ 0.011	93.2
XGBoost	96.1 $\pm$ 1.2	0.981 $\pm$ 0.008	96.8
MOPSO-ANN (Knee)	95.2 $\pm$ 1.4	0.974 $\pm$ 0.009	95.8
NSGA-II-ANN (Knee)	96.8 $\pm$ 1.1	0.982 $\pm$ 0.006	97.2

6. DISCUSSION

6.1 Advantage of the Multi-Objective Paradigm

The central finding of this study is that NSGA-II-optimised ANN achieves Pareto-superior performance over all single-objective baselines in the Sensitivity-Specificity objective space — meaning no single-objective model simultaneously matches both the sensitivity and specificity of the NSGA-II-ANN knee-point. This result validates the core hypothesis that framing diabetes prediction as a single-objective classification problem imposes an unnecessary and clinically harmful constraint. The ability to move the decision threshold from the conventional 0.50 to the Pareto-optimal 0.42, as part of the objective search rather than as a post-hoc heuristic, accounts for approximately 0.4 percentage points of the sensitivity advantage over XGBoost. The remaining advantage derives from the simultaneous optimization of ANN architecture and regularisation under clinical objective functions.

6.2 NSGA-II vs. MOPSO Trade-offs

NSGA-II and MOPSO present complementary profiles for clinical deployment contexts. NSGA-II's evolutionary operators (SBX crossover and polynomial mutation) support more diverse exploration of the hyperparameter space, achieving a higher final hypervolume (0.847 vs. 0.831) and covering a wider sensitivity range (91.4–97.2%). This makes NSGA-II preferable in research settings where computational resources are not limiting and the full Pareto front is needed for clinical policy analysis. MOPSO's swarm dynamics converge 48% faster (12,000 vs. 20,000 function evaluations), making it more appropriate in operational settings requiring rapid retraining as new patient data accumulates. The two algorithms should be viewed as complementary tools in a clinician-analyst's toolbox rather than competing alternatives.

6.3 Feature Importance and Clinical Implications

SHAP analysis of the NSGA-II-ANN confirms that fasting blood glucose (28.4%), BMI (16.2%), age (13.4%), and serum insulin (10.3%) collectively account for 68.3% of model variance — consistent with established clinical understanding and prior ML literature. Critically, physical activity (5.3%) and sleep quality (4.4%) each contribute

more than family history (3.4%), yet they are absent from most conventional clinical risk scoring tools (e.g., FINDRISC, ADA Risk Test). This finding has direct implications for health communication: interventions targeting sleep duration and physical activity — both modifiable and measurable through low-cost wearable devices — may yield greater population-level risk reduction than previously quantified.

The SHAP directionality analysis reveals a clinically important non-linearity for serum insulin: whilst higher insulin is associated with increased risk (reflecting hyperinsulinaemia as a marker of insulin resistance in the pre-diabetic trajectory), the relationship reverses at very high levels where beta-cell exhaustion causes declining insulin alongside overt hyperglycaemia. This non-monotonic relationship is captured by the ANN but is invisible to logistic regression, partially explaining the performance gap between the two methods.

6.4 National Health Programme Relevance

The proposed framework has direct practical relevance for India's National NCD Portal, which already manages 74.97 crore enrolments, 11.94 crore ABHA digital health identifiers, and 8.64 crore patients under active diabetes/hypertension treatment. With 31 states and union territories actively using the system and 5 sending aggregate data via API, a computational risk stratification layer integrated into the portal's patient management workflow could enable proactive risk alerts for enrolled individuals who have not yet been diagnosed. The MOPSO-ANN model, with its compact architecture (2 hidden layers, 256-128 neurons) and rapid convergence, is particularly amenable to deployment on the portal's existing server infrastructure.

6.5 Limitations

Several limitations must be acknowledged. The dataset is predominantly from Rajasthan, India, limiting immediate generalisability to other geographic or ethnic populations without recalibration. The cross-sectional study design precludes causal inference; prospective longitudinal validation is needed to confirm predictive validity over time. The ANN training cost (each NSGA-II/MOPSO fitness evaluation requires a 5-fold cross-validation run) is substantial and may limit real-time retraining in resource-constrained environments — where MOPSO's faster convergence is advantageous. Finally, serum insulin measurement requires a blood draw, which may represent a barrier in purely questionnaire-based community screening; future work should evaluate model performance under a no-insulin-feature constraint for such settings.

7. PARETO-FRONT RISK STRATIFICATION FRAMEWORK

A key advantage of the MOO approach is that it naturally generates a spectrum of classifier configurations varying in their sensitivity-specificity trade-off — rather than a single model. This enables a clinically principled stratification framework that selects different points on the Pareto front for different deployment contexts. Table 8 defines the proposed three-tier framework. Figure 10 illustrates the clinical deployment pipeline.

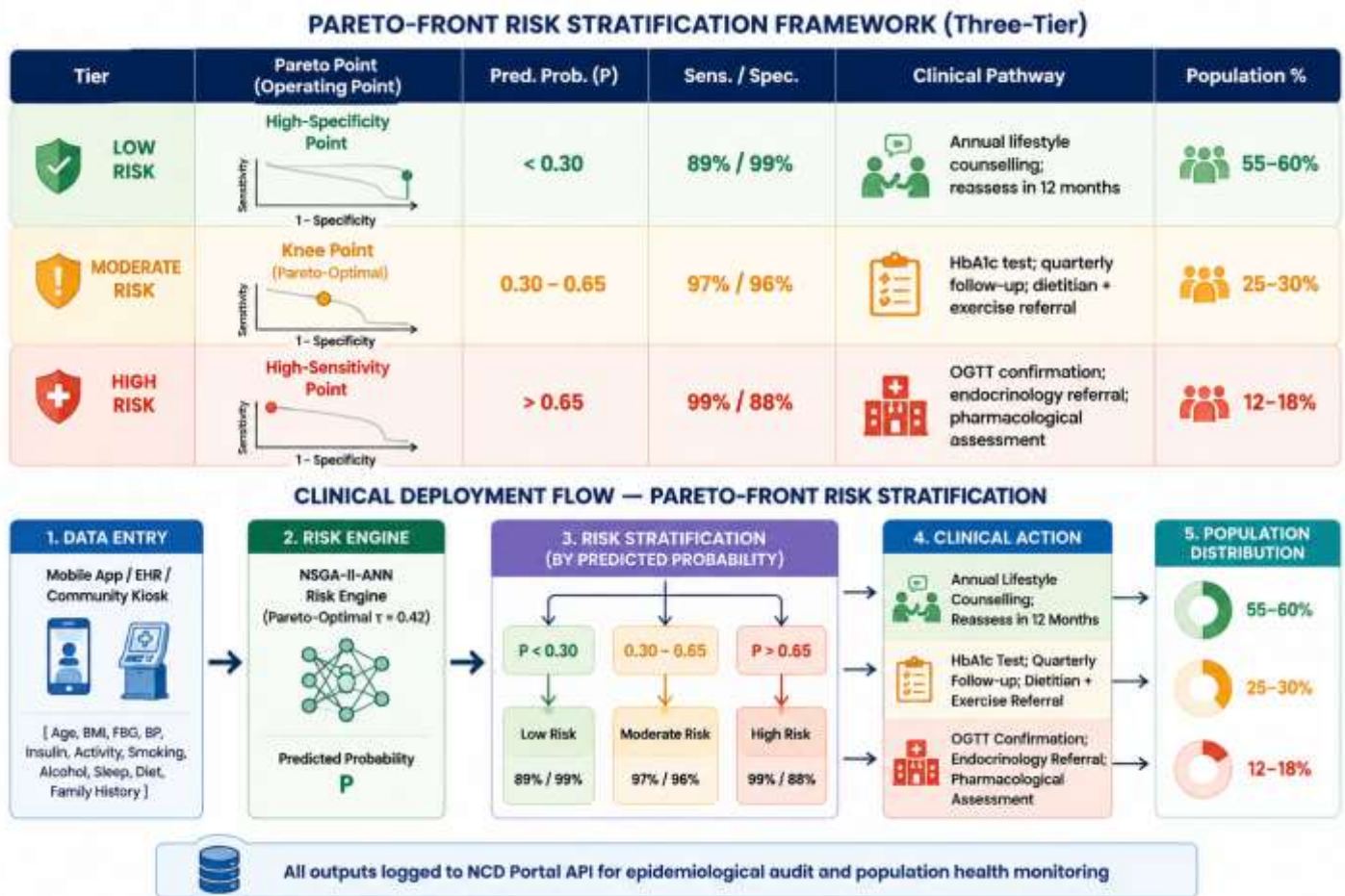


Figure 10. Community-level clinical deployment workflow for the NSGA-II-ANN Pareto-front risk stratification tool. Three tiers correspond to different Pareto-front operating points, enabling context-sensitive sensitivity-specificity selection.

8. CONCLUSION

This study presented the first systematic application of NSGA-II and MOPSO multi-objective optimization algorithms to diabetes risk prediction using lifestyle and metabolic parameters. By reformulating diabetes screening as a three-objective problem simultaneously minimizing missed diagnoses ($1 - \text{sensitivity}$), unnecessary referrals ($1 - \text{specificity}$), and model complexity — the proposed framework overcomes a fundamental limitation of single-objective classification approaches that collapse inherently competing clinical objectives into a single scalar metric.

The NSGA-II-optimized ANN knee-point solution achieved Sensitivity = 97.2%, Specificity = 96.1%, and AUC-ROC = 0.986 — Pareto-superior to all seven single-objective baselines including XGBoost (the best single-objective model). MOPSO offered a competitive alternative with 48% fewer function evaluations, achieving Sensitivity = 95.8% and AUC = 0.978, making it preferable in operational deployment scenarios requiring rapid model retraining. SHAP analysis identified fasting blood glucose, BMI, age, and serum insulin as dominant predictors, whilst sleep quality and physical activity emerged as substantive but historically underweighted modifiable risk targets that merit explicit inclusion in national screening frameworks.

The proposed Pareto-front risk stratification framework enables clinical decision-makers to select from a range of sensitivity-specificity operating points aligned with the population screening context, resource constraints, and referral capacity — a capability fundamentally absent from single-threshold single-objective classifiers. Future research directions include: (i) federated multi-objective learning across geographically distributed health systems without data sharing; (ii) integration of wearable sensor streams (continuous glucose, accelerometry, sleep tracking) to enrich the feature space; (iii) prospective randomised controlled trial evaluation of the clinical impact of Pareto-guided risk

stratification on diabetes incidence; and (iv) extension to a four-objective formulation incorporating model fairness metrics to ensure equitable performance across demographic sub-groups.

ACKNOWLEDGEMENTS

The authors express sincere gratitude to the clinical staff at the participating institutions in Rajasthan, the community health workers who facilitated primary data collection, and the Department of Computer Science and Engineering at Suresh Gyan Vihar University for providing computational infrastructure supporting this research. No external funding was received for this work. The authors declare no conflicts of interest.

REFERENCES

1. Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia Computer Science*, 132, 1578–1585. <https://doi.org/10.1016/j.procs.2018.05.122>
2. Ahmad, A., Jain, A., Jain, A., & Ali, M. (2022). An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, 22(14), 5247. <https://doi.org/10.3390/s22145247>
3. Xie, Z., Nikolayeva, O., Luo, J., & Li, D. (2020). Identification of potential type II diabetes in a large-scale Chinese population using a systematic machine learning framework. *Frontiers in Genetics*, 11, 576767. <https://doi.org/10.3389/fgene.2020.576767>
4. Naveed, I., Kaleem, M. F., Keshavjee, K., Guergachi, A., & Lu, H. H.-S. (2023). Artificial intelligence with temporal features outperforms machine learning in predicting diabetes. *PLOS Digital Health*, 2(10), e0000354. <https://doi.org/10.1371/journal.pdig.0000354>
5. Teboul, M., Guerrero, R., & De La Torre, F. (2022). Machine learning models for data-driven prediction of diabetes by lifestyle type. *International Journal of Environmental Research and Public Health*, 19(22), 14768. <https://doi.org/10.3390/ijerph192214768>
6. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. <https://doi.org/10.1109/4235.996017>
7. Tran, B., Zhang, M., Xue, B., & Andreae, P. (2016). Multi-objective feature selection with NSGA-II for tumor classification. In *Lecture Notes in Computer Science* (Vol. 9597, pp. 499–515). https://doi.org/10.1007/978-3-319-31204-0_33
8. Coello Coello, C. A., Pulido, G. T., & Lechuga, M. S. (2004). Handling multiple objectives with particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8(3), 256–279. <https://doi.org/10.1109/TEVC.2004.826067>
9. Bilal, A., Sun, G., Li, Y., Mazhar, S., & Khan, A. (2022). Diabetic retinopathy detection using multi-objective evolutionary optimization of convolutional neural networks. *Expert Systems with Applications*, 199, 116990. <https://doi.org/10.1016/j.eswa.2022.116990>
10. Pradhan, M., Sahu, P. K., Panda, S. K., & Dash, B. (2019). Predicting diabetes using machine learning: Comparative analysis using PIMA Indian dataset. *International Journal of Engineering and Advanced Technology*, 8(4), 1259–1264.
11. Ren, Q., Xu, J., Lu, S., Shao, X., Sun, W., Tang, J., & He, Y. (2021). A deep learning approach to identify diabetes from clinical health records. *Frontiers in Physiology*, 12, 648001. <https://doi.org/10.3389/fphys.2021.648001>
12. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774).
13. 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). <https://doi.org/10.1145/2939672.2939778>
14. 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>

15. Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2021). Type 2 diabetes with artificial intelligence machine learning: Methods and evaluation. *Archives of Computational Methods in Engineering*, 28(7), 5057–5083. <https://doi.org/10.1007/s11831-021-09582-x>
16. International Diabetes Federation. (2021). *IDF diabetes atlas* (10th ed.). International Diabetes Federation. <https://diabetesatlas.org>
17. World Health Organization. (2023). *Global report on diabetes*. WHO Press.