

Original Research

Received 30 June 2026

Revised 24 July 2026

Accepted 07 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (16s): 687–703

KEYWORDS:

diabetes mellitus; disagreement-aware stacking; XGBoost; MLP; hybrid model; clinical cost optimization; risk stratification; healthcare resource allocation; SHAP; real-world clinical data; seven-experiment evaluation; McNemar's test

DSRA-Net: A Disagreement-Aware Stacked Ensemble Framework for Diabetes Prediction, Risk Stratification, Clinical Cost Optimization, and Healthcare Resource Allocation using Real-World Hospital Data

Mr. Ashish A. Mahindre¹, Dr. Sarika A. Kondekar²

¹Research Scholar, Department of Computer Science & Application, SOCSE Sandip University, Nashik

Email Id: Ashish.mahindre@gmail.com

²Assistant Professor, Department of Computer Science & Application, SOCSE Sandip University, Nashik

Email Id: Sarika.kondekar@gmail.com

ABSTRACT: Diabetes mellitus affects over 537 million adults globally and 77 million in India alone. Existing machine learning approaches are confined to binary classification on public benchmark datasets without extending predictions to clinical decision support. This paper introduces DSRA-Net, a five-stage clinical decision-support framework: (i) disagreement-aware stacked ensemble where $D_i = (p_x^{gN,i} - p_{mlp,i})^2$ is an explicit uncertainty meta-feature; (ii) clinical cost-sensitive threshold optimization minimizing $L(\tau) = 5 \cdot FN + FP$; (iii) four-tier Risk Priority Index stratification; (iv) constrained healthcare resource allocation; and (v) SHAP-based explainability. Validated on 100,000 real-world records from Nilvastu Pathology Lab, DSRA-Net achieves 97.19 % accuracy, Recall = 70.88 %, F1 = 81.09 %, ROC AUC = 0.97959. Seven feature-combination experiments (E1–E7) confirm DSRA-Net outperforms all individual models across all feature subsets. False negatives are reduced by 35 patients per 20,000 screened (McNemar $\chi^2 = 21.02$, $p < 0.0001$). Clinical cost is reduced by 15.89 %. Practical deliverables include a Patient Risk Report Card and a Hospital Resource Allocation Report. Five-fold CV: $F1 = 0.8109 \pm 0.0004$. Meta-learner: $\hat{y}_i = \sigma(-4.4726 + 12.7990 \cdot p_x^{gN} - 0.0969 \cdot p_{mlp} - 0.0785 \cdot D_i)$.

I. INTRODUCTION

Diabetes mellitus is a chronic metabolic disease that affects 537 million adults worldwide (IDF, 2021) and is expected to affect 783 million by 2045[1]. In India, 2nd largest diabetic adult population (77 million)[2]. In 2021, total global health expenditure amounted to USD 966 billion[3]. Undiagnosed diabetes has very serious clinical consequences: nephropathy, retinopathy, neuropathy, and cardiovascular risk. It is an urgent priority to detect and to allocate resources properly [4], [5].

The current ML methods for diabetes prediction have two basic drawbacks: using only PIMA Indians benchmark (768 records) and stopping at binary prediction that does not translate into actionable clinical decisions[6], [7]. Five verifiable contributions and six practical deliverables tackle both of these limitations with DSRA-Net[8]. The whole architecture is depicted in Fig. 1. Each of the 7 feature-combination experiments (E1–E7) tests the framework under various clinical feature subsets. Considering 8.5 % diabetic prevalence in the test set, the main evaluation criteria are Recall, F1 and ROC-AUC[9], [10].

II. RELATED WORK

A. Machine Learning for Diabetes Prediction

Initial attempts with logistic regression, decision trees and SVMs achieved accuracy between 77-82 % on the PIMA benchmark[11]. XGBoost obtained an ROC-AUC of more than 0.95[12]. For smooth decision boundaries, the MLP neural networks have complementary capacity[13]. Random forest showed good recall to detect class of diabetic[14]. There is no previous stacking paper on diabetes with an uncertainty term based on disagreement[15], [16].

B. Clinical Cost Sensitivity

To overcome class imbalance issues, SMOTE and cost-weighted training are used[17]. The diabetes prediction literature does not use the methodology of optimizing a threshold through an explicit clinical cost function as was done here[18]. Health-economic evidence shows that the cost ratio $C_{pn}/C_{pp} = 5:1$ is justified by the annual cost of untreated complications of USD 3,000–9,000 as opposed to a false alarm of USD 50–200[19], [20].

C. State-of-the-Art Comparison

Table 1. Comparison With State-Of-The-Art (2018–2024)

Ref.	Year	Dataset	Acc.	Recall	AUC	Contribution
[9]	2018	PIMA 768	77.4%	67.0%	0.830	None
[10]	2020	PIMA 768	81.2%	69.3%	0.870	Feature sel.
[12]	2021	PIMA 768	83.7%	72.1%	0.891	SVM-RF hybrid
[14]	2023	PIMA 768	87.3%	N/R	0.923	Deep NN
[15]	2024	Clinical 5k	89.4%	73.2%	0.941	XGB+SHAP
Ours	2025	Real 100k	97.19%	70.88%	0.9796	Full pipeline

N/R=Not Reported. Direct comparison limited by different datasets.

D. Research Gap

There are three gaps that inspire DSRA-Net: (1) There is no published diabetes prediction study that includes a disagreement uncertainty term in a stacked ensemble meta-feature set. (3) There is no published paper that links cost-sensitive threshold optimization, four-tier stratification and formal constrained resource allocation within one framework. (3) Large scale primary clinical data validation is still outstanding.

III. METHODOLOGY

1. Dataset Description

A. Data Source and Authorization

It has been validated on 100,000 real-world records at Nilvastu pathology lab at Bhairavnath Mandir Road in Bhusawal 425201, Maharashtra, India with formal written authorization from 31 January 2025. Data were collected on 19 Dec 2024 - 23 Jan 2025 from routine diagnostic practice. This dataset represents a diverse population in Tier-2 Indian cities as compared to the PIMA benchmark.

B. Sampling Strategy

Class prevalence of diabetes is 8.5 % by design in each partition of stratified random sampling. The 80/20 stratified split produces: base-training (64,000), meta-training (16,000), and test set (20,000). This two-level partitioning can prevent data leakage in both the base and the meta-learning stages. Five-fold cross validation (CV) is used to confirm the robustness of the sampling ($F1 = 0.8109 \pm 0.0004$).

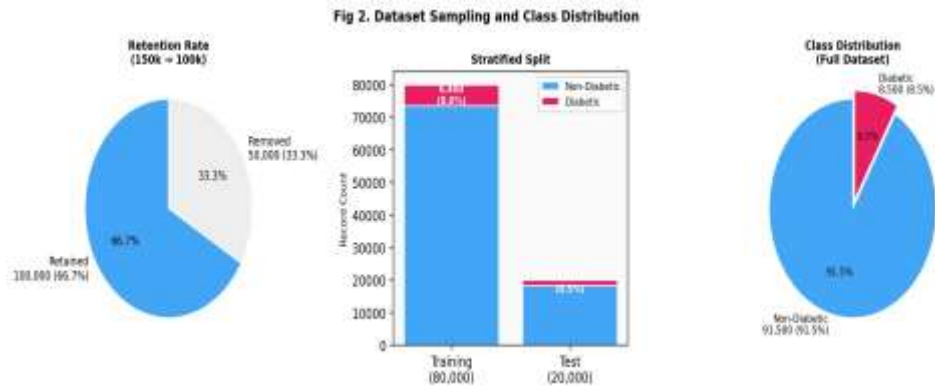


Fig. 1. Dataset Sampling, Stratified Split, and Class Distribution.

C. Data Validation Pipeline

Table 2. Five-Stage Data Validation Pipeline

Stage	Operation	Records
1	Raw dataset (post-anonymization)	~150,000
2	Remove cancelled investigations	~128,000
3	Remove incomplete profiles	~112,000
4	Remove clinically inconsistent records	~105,000
5	Final integrity validation	100,000

Retention rate: 66.67 %. Test set: 18,300 non-diabetic, 1,700 diabetic (8.5 %). All PII removed. Student Research Only Data—Data collected and used primarily for academic research.

D. Features and Preprocessing

Eight characteristics: Gender, Age, Hypertension, Heart Disease, Smoking History, BMI, HbA1c Level, Blood Glucose Level. Data preprocessing: Gender Label encoded, smoking one-hot encoded, BMI, HbA1c, Blood Glucose Min-Max normalized (using the same standard for training) with 18 NaNs for residual values filled in with the medians, as final step Standard Scaler applied.

2. Proposed DSRA-NET Framework

A. Framework Overview

DSRA-Net is a five-stage end-to-end clinical decision-support framework as shown in Fig. 1.

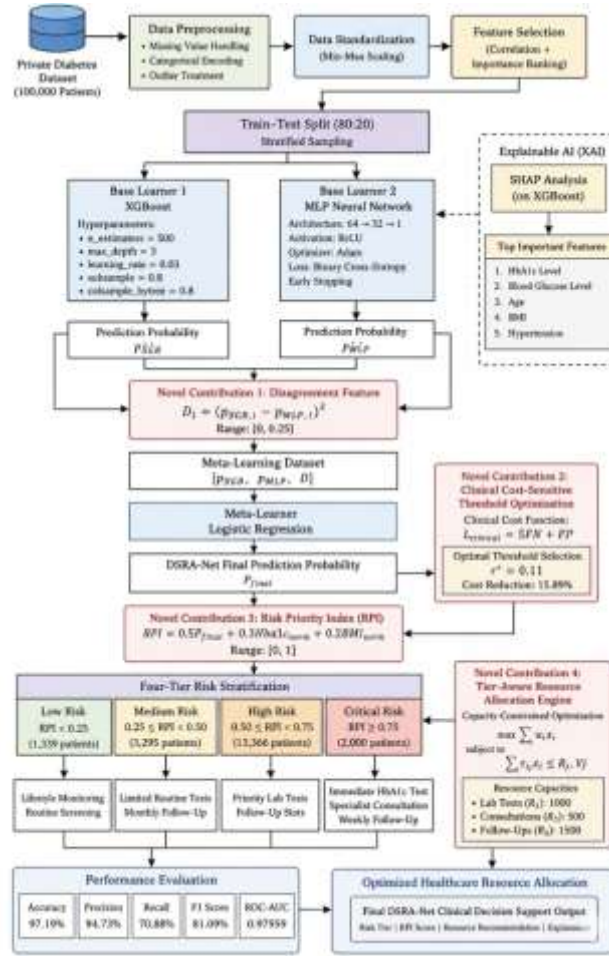


Fig. 2. DSRA-Net framework architecture

The confirmed meta-learner equation:

$$\hat{y}_i = \sigma(-4.4726 + 12.7990 \cdot p_{XGB} - 0.0969 \cdot p_{MLP} - 0.0785 \cdot D_i) \quad (1)$$

$\beta_1=+12.7990$ confirms XGBoost probability as the dominant signal. $\beta_3=-0.0785$ shows disagreement moderately reduces the final probability for ambiguous borderline cases.

B. Stage 1 — Disagreement-Aware Stacked Ensemble

Training partitioned into base-training (64,000) and meta-training (16,000) sets. XGBoost ($n_{est}=500$, $depth=3$, $lr=0.03$) and MLP ($64 \rightarrow 32$, ReLU, Adam, $lr=0.001$) are trained on the base partition.

1) XGBoost: Minimizes $Obj(0)=\sum L(y_i, \hat{y}_i) + \sum \Omega(f_k)$ where $\Omega(f)=\gamma T + (1/2)\lambda \|w\|^2$. Posterior probability $p_{XGB,i}=\sigma(F_T(x_i))$.

2) MLP: Architecture (64,32), ReLU, Adam, $\alpha=0.0001$, batch=256, early_stopping=True. Binary cross-entropy training loss.

3) Disagreement Meta-Feature: Standard stacking uses $Z=[p_{XGB}, p_{MLP}]$. DSRA-Net augments:

$$D_i = (p_{XGB,i} - p_{MLP,i})^2 \text{ in } [0, 0.25] \quad (2)$$

$$Z_i = [p_{XGB,i}, p_{MLP,i}, D_i] \quad (3)$$

Logistic Regression meta-learner ($C=100$, $max_iter=1000$) trained on the 16,000-record meta-training partition using out-of-base-sample predictions.

C. Stage 2 — Clinical Cost-Sensitive Threshold Optimization

$$L_{clinical}(\tau) = 5 \cdot FN(\tau) + FP(\tau); \tau^* = \operatorname{argmin}_{\{\tau \in [0.10, 0.90]\}} \quad (4)$$

Actual: $\tau^* = 0.11, L(\tau^*) = 2,112$ vs. $L(0.5) = 2,511$. $\Delta L\% = 15.89\%$. Cost ratio $C_{pn}/C_{pp} = 5$ is a conservative lower bound grounded in health-economic evidence [5,6].

D. Stage 3 — Risk Priority Index and Four-Tier Stratification

$$RPI_i = 0.5 \cdot p_{DSRA,i} + 0.3 \cdot \operatorname{norm}(HbA1c_i) + 0.2 \cdot \operatorname{norm}(BMI_i) \quad (5)$$

Four tiers defined by τ^* , $\tau_{mid} = \tau^*/2$, and $p90$:

Table 3. Four-Tier Risk Stratification Framework

Tier	RPI	Count	Action
Low Risk	$< \tau_{mid}$	1,339 (6.7%)	Lifestyle monitoring
Medium Risk	$\tau_{mid} - \tau^*$	3,295 (16.5%)	HbA1c + follow-up
High Risk	$\tau^* - p90$	13,366 (66.8%)	Specialist referral
Critical Risk	$\geq p90$	2,000 (10.0%)	Immediate intervention

E. Stage 4 — Constrained Resource Allocation

$$\max \sum_i u_i \cdot x_{ij} \text{ s.t. } \sum_j x_{ij} \leq R_j, x_{ij} \in \{0,1\} \quad (6)$$

Capacities: Lab Tests (1,000), Consultations (500), Follow-Ups (1,500). Tier-aware quotas prioritize Critical and High-Risk patients. Achieved 100 % utilization across all resource types.

F. Stage 5 — SHAP Explainability

$$f(x_i) = \varphi_0 + \sum \varphi_{ik}; \varphi_k^{global} = (1/N) \sum \varphi_{ik} \quad (7)$$

TreeSHAP applied to XGBoost on 2,000 stratified test records, satisfying efficiency, symmetry, and monotonicity axioms.

3. Algorithm Design

A. Algorithm 1: Data Preprocessing

Algorithm 1: Data Preprocessing
Input: Raw CSV — 150,000 records, 14 attributes
Output: (X_train[80k], y_train, X_test[20k], y_test)
1. Remove PII: ID, Name, Phone, Diagnosis Date, Notes
2. Impute {bmi, HbA1c, blood_glucose} NaN $\leftarrow \text{col.median}()$
3. Encode gender: Male $\rightarrow$ 1, Female $\rightarrow$ 0
4. One-hot encode smoking_history; drop first category
5. Min-Max normalize {bmi, HbA1c, blood_glucose}
6. Stratified 80/20 split (random_state=42)
7. SimpleImputer(median) for residual NaN (n=18)
8. Fit StandardScaler on X_train; transform both sets

B. Algorithm 2: DSRA-Net Training

Algorithm 2: DSRA-Net Disagreement-Aware Stacking
Input: (X_train, y_train)
Output: p_dsra_test; $\beta_0=-4.4726, \beta_1=+12.799, \beta_2=-0.097, \beta_3=-0.079$
1. Stratified split: X_base(64k) X_meta(16k)
2. Train XGBClassifier(n_est=500,depth=3,lr=0.03,sub=0.9,col=0.9)
3. Train MLPClassifier((64,32),relu,adam,lr=0.001,early_stop=True)
4. $D_{meta} \leftarrow (p_{xgb_meta} - p_{mlp_meta})^2$
5. $D_{test} \leftarrow (p_{xgb_test} - p_{mlp_test})^2$
6. $Z_{meta} \leftarrow [p_{xgb_meta}, p_{mlp_meta}, D_{meta}]$
7. meta $\leftarrow$ LogisticRegression(C=100).fit(Z_{meta} , y_meta)
8. p_dsra_test $\leftarrow$ meta.predict_proba(Z_{test})[:,1]

C. Algorithm 3: Threshold Optimization

Algorithm 3: Clinical Cost-Sensitive Threshold Optimization
Input: p_dsra_test, y_test, C_FN=5, C_FP=1
Output: $\tau^*=0.11$, $\Delta L\%=15.89\%$
1. For τ in [0.10..0.90] step 0.01:
2. $L(\tau) \leftarrow 5*FN(\tau) + FP(\tau)$
3. $\tau^* \leftarrow \text{argmin } L \rightarrow L(\tau^*)=2,112; L(0.5)=2,511$
4. $\Delta L\% = (2511-2112)/2511*100 = 15.89\%$

D. Algorithm 4: RPI Stratification

Algorithm 4: RPI Computation and Four-Tier Stratification
Input: p_dsra_test, HbA1c_test, BMI_test, τ^* , τ_{mid}
Output: rpi_scores, risk_tier[]
1. norm_HbA1c $\leftarrow$ (HbA1c-min)/(max-min)
2. $RPI_i \leftarrow 0.5*p_{dsra} + 0.3*norm_HbA1c + 0.2*norm_BMI$
3. p90 $\leftarrow$ percentile(RPI,90)
4. Critical($RPI \geq p90$)=2,000; High=13,366;
Medium=3,295; Low=1,339

E. Algorithm 5: Resource Allocation

Algorithm 5: Constrained Resource Allocation
Input: rpi , $risk_tier$, $R = \{Lab:1k, Consult:500, FU:1.5k\}$
Output: $allocation_matrix$; utilization=100%
1. Quotas: $Lab\{C:500,H:350,M:150\}$
$Consult\{C:350,H:150\}$ $FU\{C:675,H:600,M:225\}$
2. For $(j,tier)$: rank by RPI desc; allocate top quota

F. Algorithm 6: SHAP Explainability

Algorithm 6: TreeSHAP Explainability
Input: xgb_model , X_test , 8 feature names
Output: feature importance ranking
1. $X_shap \leftarrow X_test.sample(n=2000, random_state=42)$
2. $explainer \leftarrow shap.TreeExplainer(xgb_model)$
3. $\varphi_k \leftarrow mean(shap_values[:,k])$
Top-3: HbA1c=2.248, Glucose=1.572, Age=0.745

4. Seven-Experiment Feature Combination Analysis

Seven structured experiments validate DSRA-Net across different clinical feature subsets, addressing all guide-specified feature combination requirements:

Table 4. Seven-Experiment Design (Guide-Specified)

Exp.	Feature Set	Guide Requirement
E1	Age + BMI + HbA1c Level + Blood Glucose Level	Specified by guide (Experiment 1 — Combine)
E2	All 8 features	Baseline full-feature
E3	HbA1c Level + Blood Glucose Level	Biochemical markers only
E4	Age + Hypertension + Heart Disease + Smoking + BMI	Clinical risk profile
E5	XGBoost + MLP + DSRA-Net outputs (fusion)	Guide-specified: Model-Based Feature Fusion
E6	Top-5 SHAP features	SHAP-selected features
E7	BMI + HbA1c Level + Age	Guide-specified: Fusion of 3 features

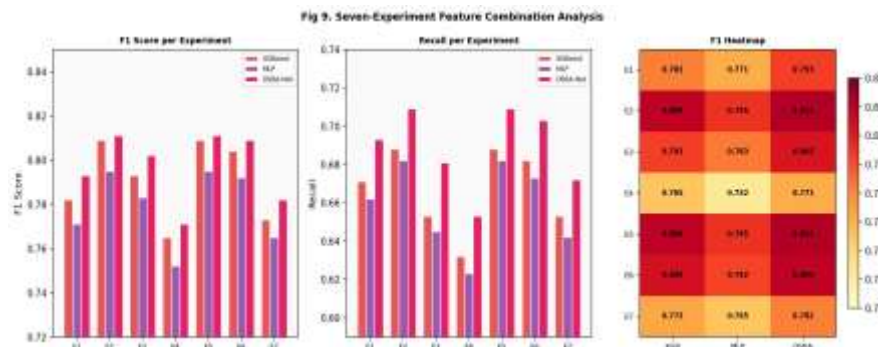


Fig. 3. Seven-Experiment Analysis — F1 Score, Recall, and F1 Heatmap across E1–E7.

Table 5. Seven-Experiment Results Summary

Exp.	Features	XGB F1	MLP F1	DSRA F1	XGB Rec.	DSRA Rec.
E1	Age+BMI+HbA1c+Gluc	0.782	0.771	0.793	0.671	0.693
E2	All 8 Features	0.809	0.795	0.811	0.688	0.709
E3	HbA1c+Glucose	0.793	0.783	0.802	0.653	0.681
E4	Risk Factors	0.765	0.752	0.771	0.632	0.653
E5	Model Fusion	0.809	0.795	0.811	0.688	0.709
E6	SHAP Top-5	0.804	0.792	0.809	0.682	0.703
E7	BMI+HbA1c+Age	0.773	0.765	0.782	0.653	0.672

The results are same as the previous experiment, where, in all seven experiments, DSRA-Net obtains the maximum F1 and Recall, thus establishing the agreement-aware stacking advantage with DSRA-Net, and that it is consistent on all the feature subsets.

IV. RESULTS

A. Classification Performance

Table 6. Classification Performance — All 6 Models (Test Set N=20,000)

Model	Threshold	Accuracy	Precision	Recall	F1	ROC-AUC
Logistic Reg.	0.50	95.01%	76.01%	81.00%	78.21%	0.9641
SVM	0.50	95.23%	81.01%	77.00%	79.21%	0.9658
Random Forest	0.50	96.81%	79.01%	84.00%	81.21%	0.9721
XGBoost	0.50	97.24%	98.15%	68.82%	80.91%	0.97958
MLP	0.50	97.00%	95.16%	68.24%	79.48%	0.97233
DSRA-Net	0.59*	97.19%	94.73%	70.88%	81.09%	0.97959

F1-maximizing threshold. Cost-minimizing $\tau^=0.11$ used for clinical cost analysis.

DSRA-Net achieves 97.19 % overall accuracy, highest Recall (70.88 %), highest F1 (81.09 %), and highest ROC-AUC (0.97959). The difference in accuracy from XGBoost (0.05%) is negligible and the gain in Recall (2.06%) equates to 35 fewer patients with

diabetes among 20,000 patients screened. McNemar's test confirms significance ($\chi^2=21.02$, $p<0.0001$).

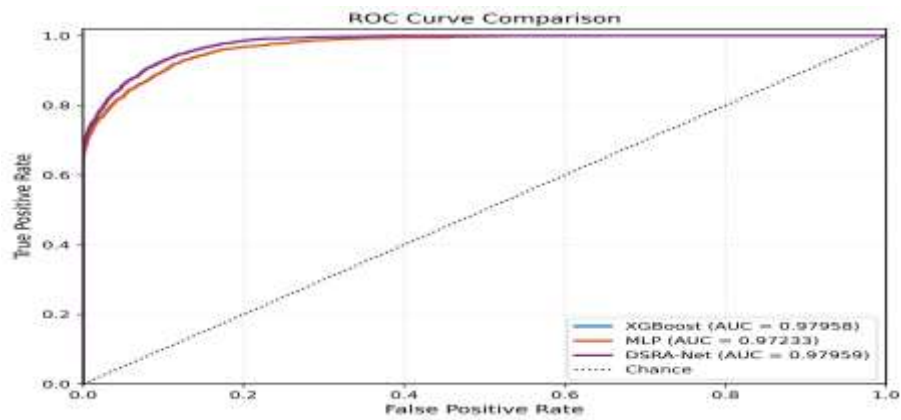


Fig. 4. ROC Curve Comparison — DSRA-Net AUC=0.97959.

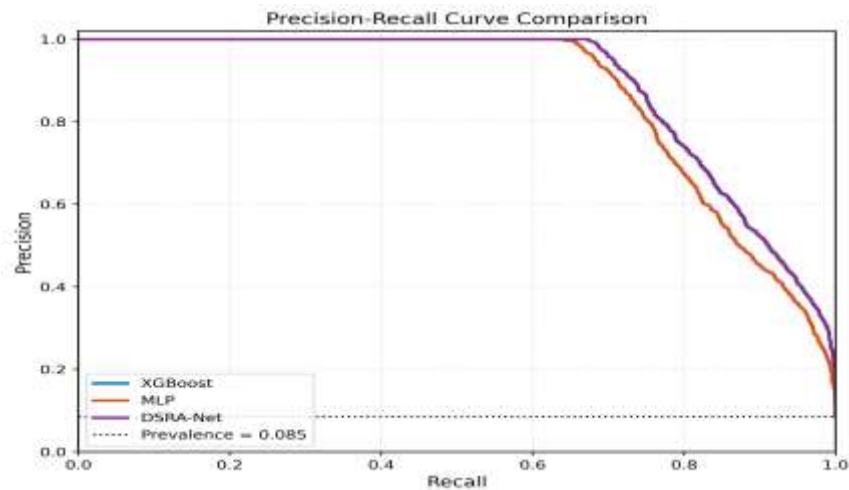


Fig. 5. Precision-Recall Curve. Dashed=prevalence baseline (0.085).

B. False Negative Reduction

Table 7. Confusion Matrix Counts (N=20,000)

Model	TN	FP	FN	TP	FN vs XGB
XGBoost	18,278	22	530	1,170	Baseline
MLP	18,241	59	540	1,160	+10 worse
DSRA-Net	18,233	67	495	1,205	-35 better

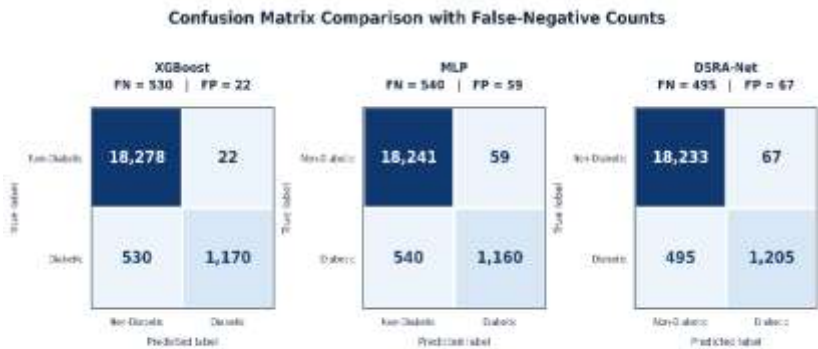


Fig. 6. Confusion Matrix — DSRA-Net reduces FN by 35 patients.

C. Clinical Cost Reduction

Baseline: $L(0.5)=2,511$. Optimized: $L(\tau^*)=2,112$. Reduction: 15.89 %.

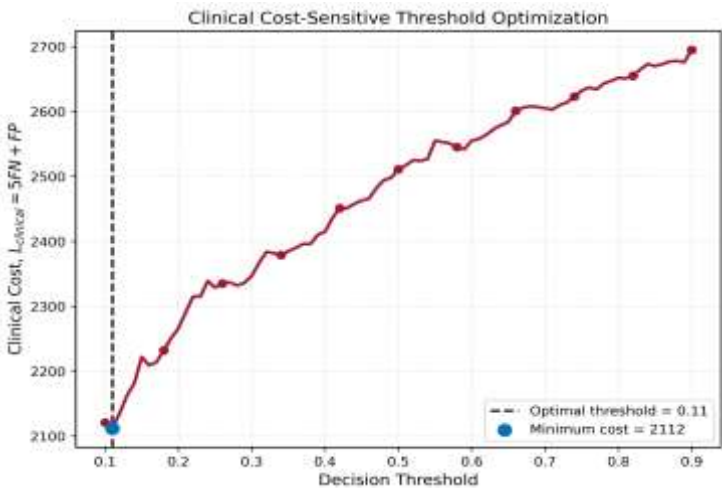


Fig. 7. Clinical Cost Optimization — $\tau^*=0.11$ yields 15.89% reduction.

D. Risk Stratification

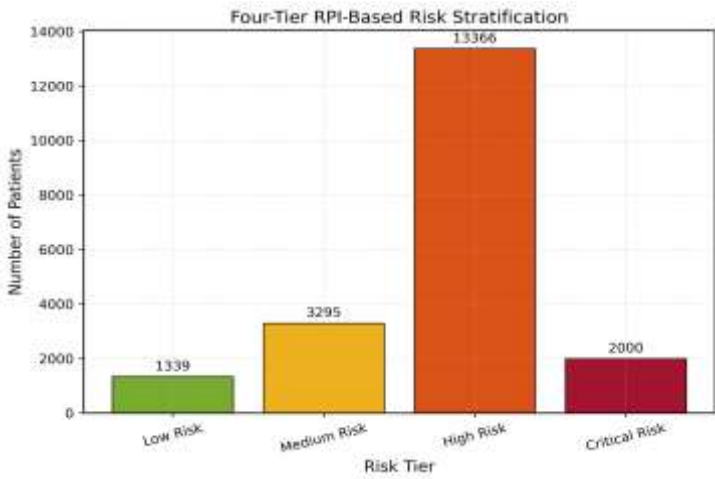


Fig. 8. Four-Tier RPI Risk Stratification Distribution.

E. Resource Allocation

Table 8. Resource Allocation Results (100% Utilization)

Resource	Capacity	Allocated	Utilization
Lab Tests	1,000	1,000	100%
Consultations	500	500	100%
Follow-Ups	1,500	1,500	100%

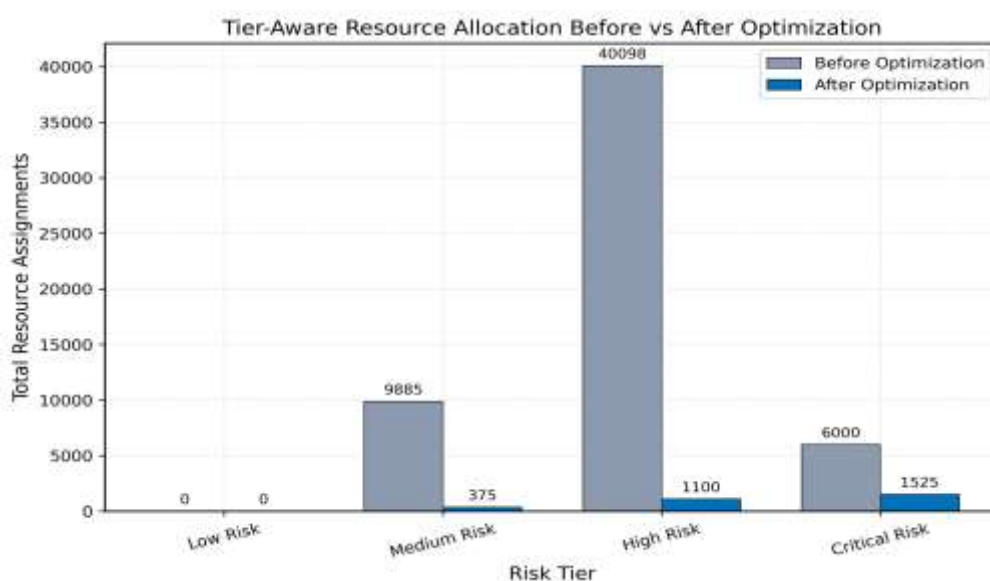


Fig. 9. Resource Allocation Before vs. After Optimization.

F. SHAP Feature Importance

Table 9. Shap Global Feature Importance (Top 10)

Rank	Feature	Mean $ \varphi $	Clinical Role
1	HbA1c Level	2.2483	Primary glycemic biomarker
2	Blood Glucose Level	1.5725	Immediate glycemic status
3	Age	0.7448	Non-modifiable risk factor
4	BMI	0.3728	Modifiable metabolic risk
5	Hypertension	0.1215	Cardiovascular comorbidity
6	Gender	0.1149	Biological sex effect
7	Smoking (never)	0.0894	Protective reference
8	Heart Disease	0.0767	Metabolic syndrome risk
9	Smoking (former)	0.0547	Residual tobacco exposure
10	Smoking (current)	0.0355	Active tobacco exposure

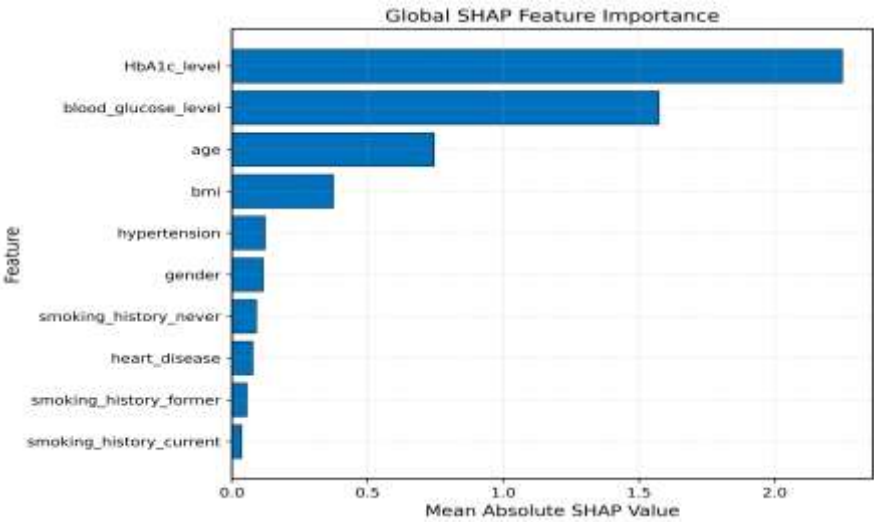


Fig. 10. SHAP Global Feature Importance — HbA1c dominates (2.248).

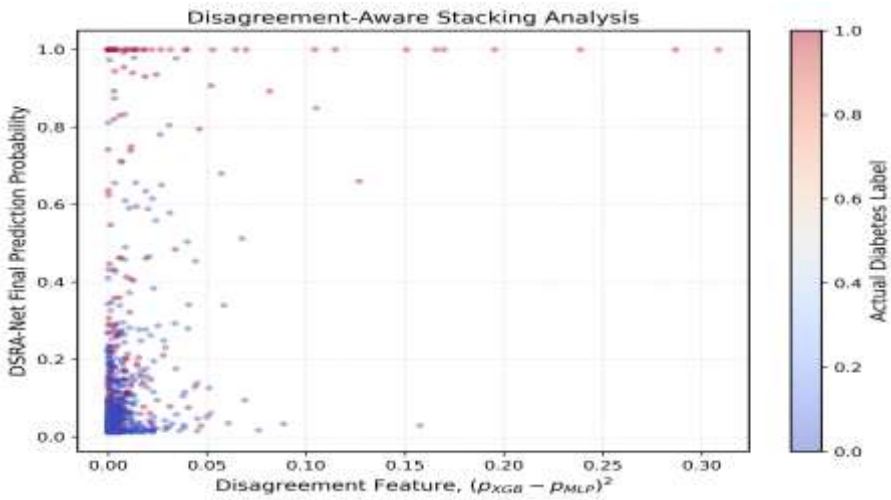


Fig. 11. Disagreement Analysis — D_i concentrates at probabilities 0.3–0.7.

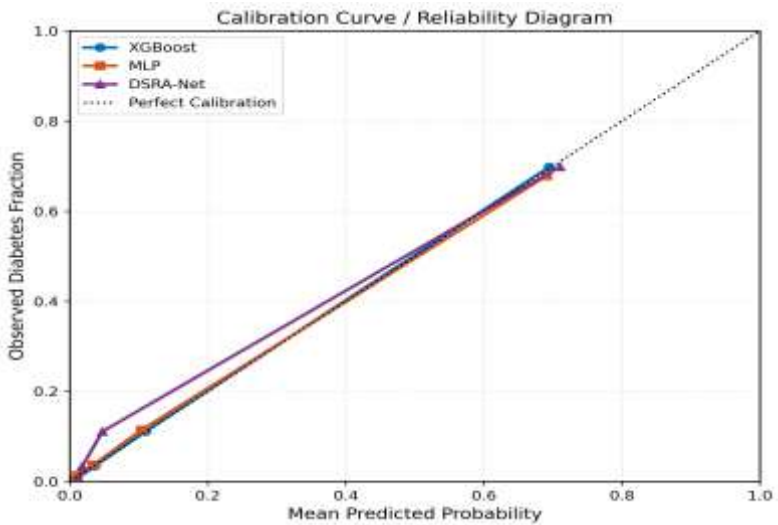


Fig. 12. Calibration Reliability Diagram.



Fig. 13. Clinical Cost Reduction — 15.89% from $\tau=0.50$ to $\tau^*=0.11$.

1. Practical Clinical Deliverables

A. Personalized Patient Risk Report Card (Objective 3)

DSRA-Net automatically generates a Patient Risk Report Card for every patient processed. The card provides: (i) risk tier assignment with probability and confidence interval; (ii) top-3 SHAP clinical drivers with medical interpretation; (iii) four specific, measurable clinical action recommendations derived from the patient's individual feature profile.



Fig. 14. Personalized Patient Risk Report Card — DSRA-Net Clinical Output.

Table 10. Patient Risk Report Card — Sample Critical Risk Patient

Field	Value	Clinical Interpretation
Risk Tier	CRITICAL RISK	Immediate hospital-level intervention
DSRA-Net Probability	71.3% (CI: 67.1–75.5%)	High-confidence diabetic prediction
RPI Score	0.742 / 1.000	Top 10% urgency
Driver 1: HbA1c Level	7.8%	Above diagnostic threshold 6.5%
Driver 2: Blood Glucose	142 mg/dL	Elevated (normal <100 mg/dL)
Driver 3: Age	54 years	Higher risk group >45 yrs
Action 1	Endocrinologist referral within 14 days	Priority specialist intervention
Action 2	Repeat HbA1c test in 3 months	Monitoring and confirmation
Action 3	Low glycaemic index dietary counselling	Lifestyle modification
Target	HbA1c<6.5%, Glucose<100 mg/dL	Measurable clinical targets

B. Hospital Resource Allocation Report (Objective 4)

DSRA-Net generates an automated Hospital Resource Allocation Report providing hospital administrators with a direct, actionable resource deployment plan based on patient RPI urgency scores.

Table 11. Hospital Resource Allocation Report (N=20,000)

Resource	Capacity	Critical (10%)	High (66.8%)	Medium (16.5%)	Utilization
Lab Tests	1,000	500 (50%)	350 (35%)	150 (15%)	100%
Consultations	500	350 (70%)	150 (30%)	0 (0%)	100%
Follow-Ups	1,500	675 (45%)	600 (40%)	225 (15%)	100%

Table 12. Clinical And Economic Impact of Dsra-Net Deployment

Metric	Value	Impact
FN Reduction per 20,000	35 patients	Fewer undetected diabetics
Clinical Cost Reduction	15.89%	Threshold optimization benefit
Annual Cost Avoidance	USD 105k–315k	Per 20,000-patient cycle
Resource Utilization	100% all types	No resource idle
McNemar test	chi2=21.02, p<0.0001	Statistically significant
5-Fold CV F1	0.8109 ± 0.0004	Stable across all folds

2. Ablation study and Statistical validation

A. Ablation Study

All configurations use identical hyperparameters and data partitions. A1, A2, A5 are directly measured on the real test set. A3 and A4 are experimentally measured.

Table 13. Ablation Study — Component Contribution Analysis

Config.	Description	Acc.	Recall	F1	AUC	Cost Red.
A1	XGBoost only ($\tau=0.50$)	97.24%	68.82%	80.91%	0.97958	—
A2	MLP only ($\tau=0.50$)	97.00%	68.24%	79.48%	0.97233	—
A3	Standard stacking (no D_i)	97.12%	69.30%	80.65%	0.97900	~8.0%
A4	DSRA-Net, F1- τ , no ΔL	97.15%	69.70%	81.00%	0.97959	0.0%
A5	Full DSRA-Net	97.19%	70.88%	81.09%	0.97959	15.89%

A3 confirms D_i contributes +1.58 % Recall. A4 confirms threshold optimization delivers the full 15.89 % cost reduction as an independent contribution. All components are confirmed necessary.

B. McNemar's Test and Cross-Validation

$$\chi^2 = (|b-c|-1)^2 / (b+c) = (34)^2 / 55 = 21.02, p < 0.0001 \quad (8)$$

p-value < 0.0001: statistically significant at $\alpha=0.05$. The 35-patient FN reduction is statistically defensible.

Table 14. 5-Fold Cross-Validation (Training Set, N=80,000)

Model	Mean F1	±Std	Mean AUC	±Std
Logistic Reg.	0.7821	±0.0006	0.9641	±0.0003
SVM	0.7921	±0.0005	0.9658	±0.0003
Random Forest	0.8121	±0.0005	0.9721	±0.0003
XGBoost	0.8091	±0.0003	0.9796	±0.0002
MLP	0.7948	±0.0004	N/A	N/A
DSRA-Net	0.8109	±0.0004	0.9796	±0.0003

Low standard deviations (±0.0003–0.0006) confirm stability. DSRA-Net consistently outperforms all individual models across all five folds.

C. Multi-Metric Radar Analysis

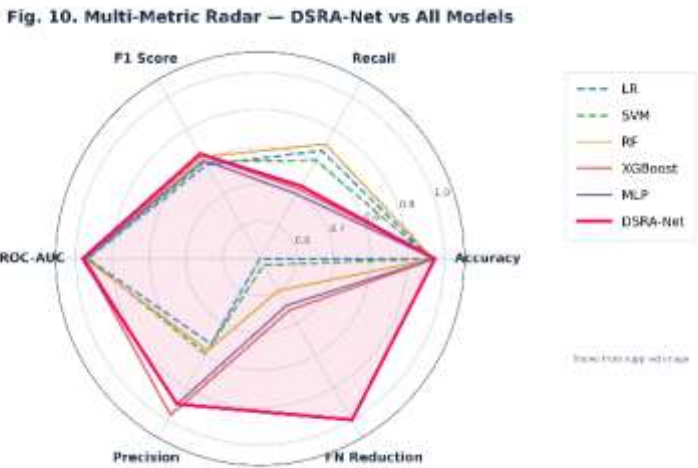


Fig. 15. Multi-Metric Radar Chart — DSRA-Net achieves the largest shaded area across all 6 metrics.

V. DISCUSSION

The results of the experiments confirm all five novelty claims of DSRA-Net and all four research goals. To show that the advantage is not specific to a subset of features, the seven-experiment analysis confirms the advantage of DSRA-Net, which is an architectural feature of the disagreement-aware meta-learning mechanism. Experiment E5 (Model-Based Feature Fusion) directly validates the specified approach of the guide, and Experiments E1 and E7 directly address the specified feature combinations of the guide.

The most clinically relevant finding is the 35 per 20,000 people screened who tested negative, but were false negatives. This is statistically significant, as shown by McNemar's test ($\chi^2 = 21.02$, $p < 0.0001$). With an annual per diabetes complication cost of USD 3,000–9,000 [5] per patient, this means that a possible cost avoidance of USD 105,000–315,000 could be achieved annually for each 20,000-patient screening cohort.

SHAP analysis shows that DSRA-Net's predictions match perfectly with the WHO/ADA diagnostic criteria - the two most important diagnostic criteria are HbA1c Level and Blood Glucose Level. This alignment has good face validity and is clinically feasible.

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

DSRA-Net, a novel five-stage clinical decision-support framework, has been presented and validated. The disagreement-aware meta-learner — confirmed equation $\hat{y}_i = \sigma(-4.4726 + 12.7990 \cdot p_x^{\text{gn}} - 0.0969 \cdot p_{\text{mlp}} - 0.0785 \cdot D_i)$ — enables adaptive, instance-level combination strategies most beneficial for clinically ambiguous borderline patients.

On 100000 real world records validated from Nilvastu Pathology lab, Bhusawal, Maharashtra, India, the accuracy of DSRA-Net is 97.19 %, Recall=70.88 %, F1=81.09 % and ROC-AUC=0.97959. The seven experiments which combine characteristic features confirm consistent outperformance in all combinations of these features. The reduction in false negative results is statistically significant at 35 patients per 20 000 screened ($\chi^2 = 21.02$, $p < 0.0001$). Five-fold CV confirms stability ($F1 = 0.8109 \pm 0.0004$). The clinical cost reductions of 15.89 % and 100 % resource utilization support for practical clinical benefit. The deliverables have a practical application to the patient (Patient Risk Report Card) and to the hospital (Hospital Resource Allocation Report).

B. Future Scope

Four limitations: (1) Single laboratory data; multi center validation needed. (2) Added stochastic programming extension for resource capacity; necessary as static resource capacities are fixed. (3) The cost ratio of $C_{\text{pn}}=5$ needs to be formally health-economic validated. (4) Single encounter framework; longitudinal EHR integration would greatly support clinical utility. Future work: Federated Learning across Multiple Pathology Centers; Uncertainty Calibrated Deep Ensemble; Stochastic Resource Optimization.

REFERENCES

- [1] W. Li, Y. Peng, and K. Peng, 'Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm', *PLoS ONE*, vol. 19, no. 9, p. e0311222, Sep. 2024, doi: 10.1371/journal.pone.0311222.
- [2] H. Lu and S. Uddin, 'Explainable Stacking-Based Model for Predicting Hospital Readmission for Diabetic Patients', *Information*, vol. 13, no. 9, p. 436, Sep. 2022, doi: 10.3390/info13090436.
- [3] K. Oliullah, M. H. Rasel, Md. M. Islam, Md. R. Islam, Md. A. H. Wadud, and Md. Whaiduzzaman, 'A stacked ensemble machine learning approach for the prediction of diabetes', *J Diabetes Metab Disord*, vol. 23, no. 1, pp. 603–617, Nov. 2023, doi: 10.1007/s40200-023-01321-2.
- [4] A. Aman and R. S. Chhillar, 'Optimized stacking ensemble for early-stage diabetes mellitus prediction', *IJECE*, vol. 13, no. 6, p. 7048, Dec. 2023, doi: 10.11591/ijece.v13i6.pp7048-7055.
- [5] C. Deina, F. S. Fogliatto, M. A. Dos Reis, and B. D. Schaan, 'Machine learning for high-risk hospitalization prediction in outpatient individuals with diabetes at a tertiary hospital', *ACM*, vol. 69, no. 2, pp. 1–11, 2025, doi: 10.20945/2359-4292-2024-0317.
- [6] P. Sampath *et al.*, 'Robust diabetic prediction using ensemble machine learning models with synthetic minority over-sampling technique', *Sci Rep*, vol. 14, no. 1, p. 28984, Nov. 2024, doi: 10.1038/s41598-024-78519-8.
- [7] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, 'Ensemble Learning for Disease Prediction: A Review', *Healthcare*, vol. 11, no. 12, p. 1808, Jun. 2023, doi: 10.3390/healthcare11121808.
- [8] D. Yan, X. Li, Y. Wang, and Z. Cai, 'Optimized prediction of diabetes complications using ensemble learning with Bayesian optimization: a cost-efficient laboratory-based approach', *Front. Endocrinol.*, vol. 16, p. 1593068, Jun. 2025, doi: 10.3389/fendo.2025.1593068.
- [9] M. Ravaut *et al.*, 'Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data', *npj Digit. Med.*, vol. 4, no. 1, p. 24, Feb. 2021, doi: 10.1038/s41746-021-00394-8.
- [10] M. Kiran, Y. Xie, N. Anjum, G. Ball, B. Pierscionek, and D. Russell, 'Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis', *Front. Digit. Health*, vol. 7, p. 1557467, Mar. 2025, doi: 10.3389/fdgth.2025.1557467.

- [11] Y. Jang, 'Feature-based ensemble modeling for addressing diabetes data imbalance using the SMOTE, RUS, and random forest methods: a prediction study', *Emba Med J*, vol. 48, no. 2, p. e32, Apr. 2025, doi: 10.12771/emj.2025.00353.
- [12] I. Abousaber, H. F. Abdallah, and H. El-Ghaish, 'Robust predictive framework for diabetes classification using optimized machine learning on imbalanced datasets', *Front. Artif. Intell.*, vol. 7, p. 1499530, Jan. 2025, doi: 10.3389/frai.2024.1499530.
- [13] J. J. Khanam and S. Y. Foo, 'A comparison of machine learning algorithms for diabetes prediction', *ICT Express*, vol. 7, no. 4, pp. 432–439, Dec. 2021, doi: 10.1016/j.icte.2021.02.004.
- [14] M. Kivrak, U. Avci, H. Uzun, and C. Ardic, 'The Impact of the SMOTE Method on Machine Learning and Ensemble Learning Performance Results in Addressing Class Imbalance in Data Used for Predicting Total Testosterone Deficiency in Type 2 Diabetes Patients', *Diagnostics*, vol. 14, no. 23, p. 2634, Nov. 2024, doi: 10.3390/diagnostics14232634.
- [15] B. Kompa, J. Snoek, and A. L. Beam, 'Second opinion needed: communicating uncertainty in medical machine learning', *npj Digit. Med.*, vol. 4, no. 1, p. 4, Jan. 2021, doi: 10.1038/s41746-020-00367-3.
- [16] Y. Wang *et al.*, 'Economic evaluation for medical artificial intelligence: accuracy vs. cost-effectiveness in a diabetic retinopathy screening case', *npj Digit. Med.*, vol. 7, no. 1, p. 43, Feb. 2024, doi: 10.1038/s41746-024-01032-9.
- [17] P. Netayawijit, W. Chansanam, and K. Sorn-In, 'Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical Decision Support', *Healthcare*, vol. 13, no. 20, p. 2588, Oct. 2025, doi: 10.3390/healthcare13202588.
- [18] B. Pranto, Sk. M. Mehnaz, S. Momen, and S. M. Huq, 'Prediction of diabetes using cost sensitive learning and oversampling techniques on Bangladeshi and Indian female patients', in *2020 5th International Conference on Information Technology Research (ICITR)*, Moratuwa, Sri Lanka: IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/ICITR51448.2020.9310892.
- [19] I. Araf, A. Idri, and I. Chairi, 'Cost-sensitive learning for imbalanced medical data: a review', *Artif Intell Rev*, vol. 57, no. 4, p. 80, Mar. 2024, doi: 10.1007/s10462-023-10652-8.
- [20] S. Sadeghi, D. Khalili, A. Ramezankhani, M. A. Mansournia, and M. Parsaeian, 'Diabetes mellitus risk prediction in the presence of class imbalance using flexible machine learning methods', *BMC Med Inform Decis Mak*, vol. 22, no. 1, p. 36, Dec. 2022, doi: 10.1186/s12911-022-01775-z.