Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

Optimized Explainable Predictive Models for Risk-Based Prioritization in Type 2 Diabetes Prevention among Women with Prior Gestational Diabetes

Amirthanathan Prashanthan [1,*] and Jenifar Prashanthan [2,*]

[1] The Business Researcher, London, TW7 6RG, United Kingdom; e-mail : prash@thebusinessresearcher.co.uk; prashanthana@hotmail.com
[2] Department of Endocrinology and Diabetes, West Middlesex University Hospital, Middlesex, TW7 6AF, United Kingdom; e-mail : p.jenifar@outlook.com; jenifar.prashanthan@nhs.net
* Corresponding Author : Amirthanathan Prashanthan

Abstract: Women with a history of gestational diabetes mellitus (GDM) face a substantially elevated risk of developing type 2 diabetes mellitus (T2DM), yet healthcare systems lack systematic approaches to prioritize these women for preventive interventions under resource constraints. This proof-of-concept study develops and demonstrates an integrated framework that combines machine-learning–based risk prediction with multi-algorithm optimization to enable evidence-based patient prioritization using synthetic data. A two-phase methodology was implemented using synthetic data from 6,000 women with prior GDM. Phase 1 deployed five classification algorithms (Logistic Regression, Random Forest, XGBoost, LSTM, and CNN-1D) with 10-fold stratified cross-validation and SMOTE-ENN resampling for T2DM risk prediction. Phase 2 implemented 9 optimization algorithms across 10 budget scenarios and 5 priority thresholds, yielding 450 optimization runs. The prioritization framework targets postpartum and interpregnancy follow-up care for women with prior GDM. Logistic Regression achieved the highest predictive performance with an AUC-ROC of 0.9454 (accuracy: 0.8875, recall: 0.8533, F1-score: 0.7913). SHAP analysis identified insulin treatment during pregnancy (mean |SHAP| = 0.099), GDM recurrence history (0.073), and postpartum weight gain (0.063) as the most influential predictors. Linear Programming consistently produced optimal solutions with a mean total priority of 901.47 and 65.26% high-risk coverage. At a 25% budget allocation ($1.95 million), 60.5% of very high-risk women could be prioritized, whereas full high-risk coverage would require a 50% budget allocation. Overall, the proposed framework demonstrates the feasibility of integrating risk prediction with constrained optimization to support resource allocation for T2DM prevention. Clinical validation using real-world prospective data is required prior to practical implementation.

Keywords: Clinical decision support; Explainable artificial intelligence; Gestational diabetes mellitus; Good Health and Well-being; Healthcare resource allocation; Machine learning; Risk prediction; Type 2 diabetes mellitus.

Received: January, 26th 2026
Revised: February, 21st 2026
Accepted: February, 23rd 2026
Published: February, 25th 2026
Curr. Ver.: February, 25th 2026

1. Introduction

Type 2 Diabetes Mellitus (T2DM) affects an estimated 537 million individuals globally and is projected to reach 783 million by 2045 [1]. Beyond glycemic dysregulation, its cascade of micro- and macrovascular complications imposes substantial morbidity, mortality, and economic burden on healthcare systems worldwide [2]. Epidemiological evidence suggests that T2DM is largely preventable through the timely identification of high-risk individuals and the implementation of tailored lifestyle interventions [3], underscoring the need for comprehensive risk stratification.

Gestational Diabetes Mellitus (GDM), defined as glucose intolerance first diagnosed during pregnancy, has emerged as one of the most significant predictors of future T2DM

development [4]. Women with a history of GDM experience a seven-fold greater risk of developing T2DM compared to those with normoglycemic pregnancies, with cumulative incidence rates reaching 50–70% within 10–20 years postpartum [5], [6]. The underlying pathophysiological mechanisms—including reduced pancreatic beta-cell function, insulin resistance, and persistent metabolic dysregulation—extend well beyond the gestational period [7]. The rising prevalence of GDM, paralleling increases in maternal obesity and advanced maternal age, has further expanded the population at heightened risk of T2DM [8].

Despite strong evidence supporting postpartum prevention interventions—including the Diabetes Prevention Program [9], metformin prophylaxis [10], and structured follow-up protocols—clinical translation remains suboptimal. Fewer than 50% of women with prior GDM receive the recommended postpartum glucose testing [11], [12], a gap exacerbated by resource constraints and the absence of systematic methodologies for identifying those who would most benefit from intensive intervention [13].

Machine learning (ML) has rapidly advanced healthcare risk prediction by leveraging the growing availability of electronic health data [14], [15]. SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee [16], provides a game-theoretic framework for transparent feature attribution that is widely adopted in clinical prediction research [17]. To address class imbalance—a pervasive challenge in healthcare prediction—the Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors (SMOTE-ENN) combines synthetic minority oversampling with Edited Nearest Neighbors under-sampling [18], [19]. In parallel, operations research (OR) methods have been applied to healthcare resource allocation [20], [21], with linear programming optimizing human immunodeficiency virus (HIV) prevention resource allocation [22] and mixed-integer programming enabling individualized diabetes management [23]. Metaheuristic and multi-objective optimization approaches have further expanded the methodological landscape for complex healthcare decision-making [24]–[26].

This study aims to develop a comprehensive framework for T2DM prevention in women with prior GDM, with three interconnected objectives. First, we develop and rigorously compare five ML architectures—Logistic Regression, Random Forest, Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM), and one-dimensional Convolutional Neural Network (1D-CNN)—to predict T2DM risk using GDM-relevant clinical and lifestyle features [27]–[30]. Building on these predictions, the second objective systematically evaluates nine optimization algorithms to prioritize women for intervention under realistic budget constraints ranging from 5% to 75% of total intervention costs, while conducting a multi-objective Pareto analysis to balance total priority scores with high-risk coverage. Finally, the study establishes an empirical basis for resource allocation by quantifying budget–coverage relationships and providing model interpretability through SHAP analysis that distinguishes actionable modifiable risk factors from fixed clinical predictors, thereby delivering evidence-informed guidance for healthcare decision-makers [31], [32].

Unlike existing research, which treats prediction and allocation as separate problems, the proposed approach integrates predictive modelling with formal optimization to enable healthcare systems to translate risk scores into actionable intervention priorities. By systematically comparing five ML architectures and nine optimization algorithms across 450 resource scenarios, this study provides empirical evidence for budget-constrained T2DM prevention strategies in this high-risk population. The proof-of-concept framework achieves strong predictive performance (AUC-ROC: 0.9454) while providing transparent and interpretable support for clinical decision-making through SHAP-based explanations that distinguish modifiable from fixed risk factors.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the proposed methodology. Section 4 presents the experimental results. Section 5 discusses the findings, limitations, and future research directions. Finally, Section 6 concludes the paper.

2. Literature Review

2.1 Risk Prediction Models for T2DM

Traditional risk scores incorporating clinical and demographic factors have established baseline performance benchmarks. The Finnish Diabetes Risk Score (FINDRISC) [33]

achieves an AUC of 0.72–0.87, while the Framingham model [34] obtains an AUC of 0.85 for 7-year prediction, and QDScore [35] achieves an AUC of 0.83–0.85. However, Noble et al. [36] noted that most of the 94 reviewed models lacked validation in high-risk subgroups, including women with prior GDM.

GDM-specific prediction models have attempted to address this gap with varying levels of success. Kwak et al. [37] achieved an AUC of 0.73 in Korean women, and Köhler et al. [38] reported an AUC of 0.76 using German registry data. Statistical approaches include Li et al. [39], who reported a Cox model AUROC of 82.8%, and Man et al. [40], who achieved a C-index of 0.68, while Chen et al. [41] applied Poisson regression for risk estimation. Belsti et al. [42] developed antenatal (AUC: 0.76) and postnatal (AUC: 0.85) clinical models.

Omics-integrated approaches have further enhanced predictive accuracy. Allalou et al. [43] identified anthranilic acid and glutamate as predictive metabolites. Khan et al. [44] achieved an AUC of 0.92 by combining real-world and omics data. Lai et al. [45] reported a median AUC of 0.883 using metabolic signatures, and Joglekar et al. [46] demonstrated that microRNAs (miRNAs) improved the AUC from 0.83 to 0.92.

Machine learning methods have demonstrated strong performance in predicting T2DM risk. Early work by Razavian et al. [47] achieved an AUC of 0.80 using convolutional neural networks (CNNs), while Zou et al. [48] reported an AUC of 0.82 with random forests. Lai et al. [45] achieved an AUC of 0.83 using deep learning on electronic health record (EHR) data, and Kopitar et al. [49] showed gradient boosting approaching an AUC of 0.85. Recent GDM-focused ML studies include Houri et al. [50], who reported an XGBoost AUC of 0.85; Kumar et al. [51], who achieved a CatBoost AUC of 0.86; Ilari et al. [52], who reported penalized logistic regression with an AUC of 0.884; and Prashanthan and Prashanthan [53], who achieved an AdaBoost F1-score of 80.4%.

2.2. Optimization Methods in Healthcare Resource Allocation

Operations research (OR) has extensive applications in healthcare. Rais and Viana [20] reviewed optimization models for facility location, staff scheduling, and resource allocation, while Hulshof et al. [21] noted that disease prevention remains relatively underexplored. Linear programming has demonstrated effectiveness across diverse allocation problems. Earnshaw et al. [22] reported a 15–20% improvement in human immunodeficiency virus (HIV) prevention resource allocation, D'Aeth et al. [54] reported savings of 50,750–5,891,608 life-years through optimized COVID-19 hospital prioritization, and Bertsimas et al. [23] demonstrated approximately 20% improvement in personalized diabetes management. Stuart et al. [55] further advanced constrained optimization for evidence-based prioritization of health spending.

Metaheuristic algorithms provide approximate solutions to complex healthcare problems. Genetic algorithms (GAs) have been applied to nurse scheduling [24] and resource allocation [56]. Particle swarm optimization (PSO) has also shown promise; for example, Wang et al. [57] integrated PSO with Markov decision processes for emergency department resource allocation. Ala et al. [58] combined whale optimization with NSGA-II for multi-objective appointment scheduling, producing Pareto-optimal solutions that balance fairness and efficiency.

Hybrid ML–optimization approaches have recently emerged as powerful tools. Mizan and Taghipour [59] integrated predictive models with resource planning to reduce waiting times. Zuo et al. [60] developed a graph-based deep reinforcement learning model (GTARS-DRL) for outpatient scheduling, achieving high computational efficiency (ACPU = 0.762 s) while matching GA solution quality. Akbari-Moghaddam et al. [56] employed ensemble clustering with priority score functions for patient segmentation, while Piyush Ingole [61] applied artificial intelligence–powered constraint-satisfaction algorithms for hospital resource optimization.

Healthcare scheduling optimization has also reduced waiting times through various approaches. Moura and Pinho [62] developed mathematical model–heuristic hybrids for outpatient prioritization across geographic areas. Multi-objective optimization—distinguishing between weighted aggregation and Pareto-optimal approaches [26]—has been applied to chemotherapy scheduling [63], facility siting [64], and operating room scheduling [65].

GDM-to-T2DM prevention optimization remains relatively underexplored. Neuwahl et al. [66] used decision tree modelling to evaluate the cost-effectiveness of screening criteria, finding the International Association of the Diabetes and Pregnancy Study Groups

(IADPSG) criteria cost-effective at ≥23% Diabetes Prevention Program (DPP) participation ($48,588 per quality-adjusted life year [QALY]). Lloyd et al. [67] demonstrated that pregnancy lifestyle intervention is a dominant cost-saving strategy, with a return on investment (ROI) of AU$1.22 per dollar invested.

2.3. Integration of Prediction and Optimization

Despite concurrent developments, the integration of machine learning (ML) for prediction and optimization in healthcare resource allocation remains limited. Bertsimas and Kallus [68] proposed prescriptive analytics advocating decision-focused learning, while Elmachtoub and Grigas [69] developed smart predict-then-optimize frameworks. Recent advances have demonstrated notable progress. Mizan and Taghipour [59] integrated multi-target ML with optimization models for medical resource allocation, achieving a 10.81% improvement in prediction accuracy. Feuerriegel et al. [70] developed a data-driven decision model combining counterfactual inference, ML, and optimization for diabetes prevention, demonstrating potential annual savings of $1.1 billion when applied to U.S. prediabetic populations.

In chronic disease management, early work by Deo [71] combined prediction with simulation, while Liu et al. [72] applied reinforcement learning to diabetes treatment regimens. More recent clinical validation studies have accelerated progress. Wang et al. [73] demonstrated reinforcement learning–based insulin titration optimization (RL-DITR), achieving improved glycemic control with a mean absolute error (MAE) of 1.10 ± 0.03 U compared with standard clinical methods in a proof-of-concept trial published in Nature Medicine. Dénes-Fazakas et al. [74] achieved 73% Time in Range using policy optimization for personalized blood glucose management, while Zhou et al. [75] introduced Duramax, a reinforcement learning framework for cardiovascular disease prevention that achieved a policy value of 93 versus clinicians' 68.

Hybrid ML–optimization approaches have gained increasing attention. Sarode et al. [76] reviewed artificial intelligence (AI) integration for dynamic resource allocation in healthcare management systems. Ponsiglione et al. [77] proposed combining simulation models with ML for healthcare management, while a comprehensive narrative review synthesizing 170 studies [78] confirmed ML's growing role in resource optimization, with reported prediction accuracies ranging from 88% to 95%. The prescriptive analytics market in healthcare, valued at $9.53 billion in 2023, is projected to reach $61.92 billion by 2030, reflecting the broader shift toward decision-support systems that recommend optimal actions rather than merely forecasting outcomes.

Despite these advances, GDM-to-T2DM prevention remains relatively underexplored in the context of integrated prediction–optimization approaches. Lee et al. [79] proposed integrating social determinants of health with ML-driven decision support for diabetes case management resource allocation. Tan et al. [80] developed ML models for predicting high healthcare utilizers in diabetes populations for population health initiatives. However, the systematic integration of risk prediction with budget-constrained prioritization of women with prior GDM for post-GDM-to-T2DM prevention has not been adequately addressed, representing a key gap that this study aims to address. Based on the literature review, several critical gaps motivate the present study:

- Absence of prediction–optimization integration: No existing framework seamlessly integrates risk prediction with optimization methods specifically for T2DM prevention resource allocation in women with prior GDM.
- Limited algorithm comparison: The comparative performance of exact methods, metaheuristics, and greedy heuristics for prioritizing women with prior GDM has not been systematically evaluated across diverse resource scenarios.
- Insufficient multi-objective analysis: Prior work has not adequately addressed the multidimensional trade-offs (e.g., health impact, high-risk coverage, efficiency, and equity) inherent in prevention resource allocation.
- Lack of budget–coverage quantification: Healthcare administrators lack empirical evidence quantifying the relationship between budget allocation and achievable coverage among women with prior GDM.
- Limited consideration of intervention amenability: Existing prioritization approaches primarily focus on disease probability without explicitly accounting for modifiable risk factor profiles and the likelihood of response to preventive interventions.

3. Methodology

3.1. Study Design

This study employed a two-phase integrated framework that combines ML-based predictive modeling with multi-algorithm optimization for T2DM risk prediction and prioritization among women with prior GDM. The predictive phase develops and validates risk prediction models, whereas the optimization phase addresses resource allocation under budget constraints. The overall framework is illustrated in Figure 1.

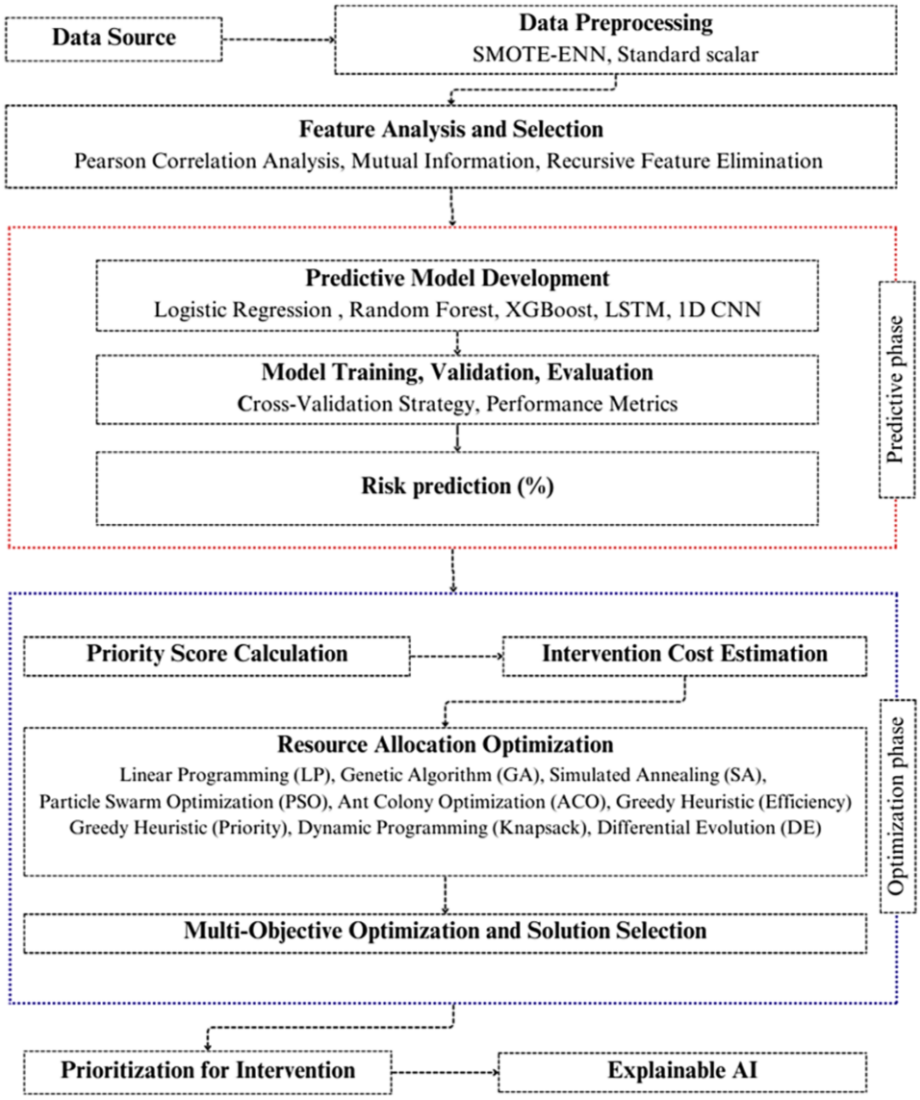


Figure 1. Two-phase integrated framework

The critical innovation lies in the vertical integration between prediction and allocation. Unlike existing studies that treat these components as disconnected activities, the proposed framework systematically links validated risk prediction with formal optimization via nine algorithms—from Linear Programming (LP) to metaheuristics and dynamic programming—while explicitly balancing total priority scores with high-risk coverage to promote both efficiency and equity. This enables healthcare systems to address not only "who is at risk?" but also "who should receive limited resources first?" The framework's modular design supports local adaptation by combining predicted risk with intervention amenability within the composite priority score. In addition, SHAP analysis and Pareto frontier visualization facilitate clinical communication and trade-off assessment. This study is intended as a methodological proof-of-concept using synthetic data; therefore, clinical validation with real-world data is required prior to practical implementation.

3.2. Data Source and Study Population

The study used a publicly available synthetic dataset [81] comprising 6,000 women with prior GDM, including 28 predictor variables and one binary outcome (T2DM_risk), with an observed class imbalance of approximately 3:1 (4,500 non-cases; 1,500 positive cases). The study population represents women 6 months to 5 years postpartum following a GDM pregnancy. The prioritization framework targets postpartum and interpregnancy periods, and interventions during pregnancy were explicitly excluded.

Synthetic data were selected to address real-world constraints, including the scarcity of large longitudinal cohorts, privacy barriers that limit reproducibility, and the difficulty of systematically controlling class distributions for methodological validation. Statistical distributions were derived from multiple epidemiological studies to ensure clinically plausible feature relationships. The dataset satisfies four key methodological requirements. First, the sample size of 6,000 observations (approximately a 214:1 observation-to-feature ratio) substantially exceeds AMIA guidelines and provides 53.6 events per predictor. Second, the 3:1 imbalance reflects a realistic prevalence suitable for SMOTE-ENN validation. Third, the comprehensive feature set spans both modifiable and immutable predictors, directly supporting composite priority scoring. Finally, this feature mix facilitates SHAP interpretability by distinguishing actionable targets from fixed prognostic indicators.

3.3. Data Preprocessing

The dataset was partitioned using stratified sampling into 80% training (n = 4,800) and 20% hold-out testing (n = 1,200). SMOTE-ENN was applied exclusively within the training folds to prevent data leakage, and z-score normalization parameters were derived solely from the training data and subsequently applied to the test set.

SMOTE-ENN was selected for its hybrid design, which combines synthetic minority oversampling with Edited Nearest Neighbors (ENN) noise removal, thereby producing more reliable decision boundaries than standalone SMOTE. This characteristic is particularly important in medical datasets that often exhibit biologically diffuse class boundaries. Compared with ADASYN's more aggressive sampling behavior, the conservative nature of SMOTE-ENN is better aligned with healthcare applications, where excessive false positives may impose unnecessary resource and patient burden costs. The preprocessing pipeline successfully transformed the 3:1 training imbalance into an approximately balanced 50:50 distribution while preserving the original prevalence in the hold-out test set.

3.4. Feature Analysis and Selection

A multi-method feature selection strategy was employed using three complementary techniques: Pearson correlation for linear target associations, mutual information for nonlinear dependencies, and recursive feature elimination (RFE) with a Random Forest base estimator (n_estimators = 100, max_depth = 5) to identify the 15 most informative features. A rank aggregation procedure was then applied by averaging feature ranks across the three methods.

Convergence across these techniques provides robust feature identification that is less biased by any single statistical assumption. The selected 15 features—spanning modifiable factors (e.g., BMI, physical activity, dietary quality) and non-modifiable predictors (e.g., age, ethnicity, family history)—directly support the priority scoring methodology while reducing overfitting risk and improving interpretability. Importantly, the ranked feature importance feeds into the optimization phase through the composite priority score, where highly predictive and modifiable features receive greater intervention amenability weights. This design represents a structured alternative to ad hoc priority rules commonly used in healthcare optimization studies.

3.5. Predictive Model Development

Five ML algorithms were selected (Table 1) to span a diverse modeling spectrum: Logistic Regression (interpretable baseline), Random Forest and Extreme Gradient Boosting (XGBoost) (state-of-the-art ensemble methods), Long Short-Term Memory (LSTM) (sequential dependency evaluation), and one-dimensional Convolutional Neural Network (1D-CNN) (convolutional pattern detection). Although LSTM and 1D-CNN are primarily designed for sequential or temporal data, they were intentionally included to evaluate whether their capacity

for automated feature interaction and nonlinear pattern detection provides any measurable advantage over classical methods when applied to structured cross-sectional feature vectors. This benchmarking question is methodologically relevant for future longitudinal extensions of the proposed framework.

The results indicate that simpler interpretable models (particularly Logistic Regression) were not outperformed, reinforcing their suitability for the present cross-sectional setting. This model diversity enables identification of the appropriate complexity level for GDM-to-T2DM prediction while balancing underfitting risk against overfitting and interpretability constraints. Each model was trained independently to enable rigorous comparative evaluation and to identify the best-performing architecture for downstream optimization integration.

Table 1. Machine learning model configurations and hyperparameters

Algorithm	Key Hyperparameters	Rationale for Selection
Logistic Regression	L2 regularization (C = 0.1); Max iterations: 1,000; Solver: lbfgs	Establishes an interpretable baseline; enables direct coefficient interpretation for clinical stakeholders; computationally efficient; well established in healthcare risk prediction
Random Forest	n_estimators: 100; max_depth: 6; min_samples_split: 20; max_features: sqrt	Captures nonlinear relationships and feature interactions; robust to outliers; provides feature importance rankings; strong performance in imbalanced medical datasets; reduces overfitting through ensemble averaging
XGBoost	n_estimators: 100; max_depth: 4; learning_rate: 0.05; eval_metric: logloss; subsample: 0.8	State-of-the-art gradient boosting; handles class imbalance effectively; sequential error correction improves weak learners; regularization helps prevent overfitting; widely successful in healthcare ML applications
LSTM	LSTM layer: 32 units; Dropout: 0.5; Recurrent dropout: 0.3; Dense layers: 16, 8 units; L2 regularization: 0.01; Batch normalization; Activation: sigmoid	Evaluates potential sequential dependencies in feature representations; dropout and regularization mitigate overfitting in deep architectures; represents an advanced neural network baseline for comparison with simpler models
1D-CNN	Convolutional layers: 32, 64 filters (kernel = 3); Max pooling: pool_size = 2; Global max pooling; Dense layers with dropout (0.5); Optimizer: Adam (lr = 0.001)	Evaluates convolutional pattern detection in feature space; parameter sharing reduces model complexity relative to fully connected networks; pooling extracts salient features; provides an alternative deep learning architecture to LSTM

All models were implemented using scikit-learn 1.2.0, XGBoost 1.7.0, and TensorFlow 2.12.0 with the Keras API. Models were trained using identical 80/20 data splits with 10-fold cross-validation. Deep learning models were trained for up to 50 epochs with early stopping (patience = 10, monitored on the validation loss) and a batch size of 32. Hyperparameters were selected through grid search or targeted manual tuning, with emphasis on predictive performance, computational feasibility, and generalizability.

3.6. Model Evaluation and Risk Prediction

Model performance was evaluated using multiple complementary metrics, including Accuracy, Precision, Recall (Sensitivity), F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Training performance was additionally compared with validation results to detect potential overfitting.

The best-performing model was subsequently used to generate individual T2DM risk probabilities (expressed as percentages) for all women with prior GDM.

3.7. Women with Prior GDM Priority Score Calculation

A composite priority score was constructed by integrating clinical risk with intervention amenability. The modifiable risk factors considered included obesity or excessive postpartum

weight gain, physical inactivity, unhealthy diet, smoking, alcohol intake, absence of postpartum glucose testing, and breastfeeding status. The priority score was computed as:

$$\text{Priority Score} = P_{\text{risk}} \times \left(0.7 + 0.3 \times \frac{N_{\text{modifiable}}}{N_{\text{total}}}\right) \quad (1)$$

where P_{risk} denotes the predicted T2DM risk probability, $N_{\text{modifiable}}$ represents the number of modifiable risk factors present, and N_{total} is the total number of modifiable factors considered.

3.8. Intervention Cost Estimation

Individualized intervention costs were estimated using a tiered cost model:

$$C_{\text{intervention}} = \left(C_{\text{base}} + (N_{\text{modifiable}} \times C_{\text{factor}})\right) \times R_{\text{multiplier}} \quad (2)$$

where the base cost (C_{base}) was set to \$500, the cost per factor (C_{factor}) to \$200, and the risk multiplier ($R_{\text{multiplier}}$) to 1.5 for women with prior GDM whose predicted risk probability exceeded 70%. Budget scenarios ranging from 5% to 75% of the total intervention cost were evaluated.

Importantly, the probability thresholds used in this study (e.g., >70% for the risk multiplier) are intended for methodological prioritization purposes to support resource allocation analysis and should not be interpreted as strict clinical decision cut-offs. Calibration against prospective clinical data will be required before operational deployment.

3.9. Resource Allocation Optimization

Nine optimization algorithms were implemented to solve the selection problem for women with prior GDM under budget constraints. Table 2 summarizes the algorithms, key parameters, and implementation details.

Table 2. Optimization algorithms and key parameters.

Algorithm	Key Parameters and Implementation Details
Linear Programming (LP)	Continuous relaxation of binary selection with HiGHS solver; thresholding at 0.5 for binary decisions
Genetic Algorithm (GA)	Population size = 100; generations = 50; two-point crossover (probability = 0.7); mutation (probability = 0.2; individual bit flip = 0.05); tournament selection (size = 3)
Simulated Annealing (SA)	Initial temperature = 1000; cooling rate = 0.995; 5000 iterations; greedy initialization based on efficiency ratio
Particle Swarm Optimization (PSO)	50 particles; 100 iterations; inertia weight = 0.7; cognitive coefficient = 1.5; social coefficient = 1.5
Ant Colony Optimization (ACO)	30 ants; 50 iterations; pheromone evaporation rate = 0.5; heuristic based on priority-to-cost ratio
Greedy Heuristic (Efficiency)	Sequential selection by descending priority-to-cost ratio until budget exhaustion
Greedy Heuristic (Priority)	Sequential selection by descending priority score until budget exhaustion
Dynamic Programming (Knapsack)	0–1 knapsack formulation; cost scaling factor = 100 for computational feasibility
Differential Evolution (DE)	Continuous optimization with penalty for constraint violation; 50 iterations; parallel evaluation

Each algorithm was evaluated across 10 budget scenarios (5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, and 75% of total cost) and five priority thresholds (0.0, 0.2, 0.3, 0.4, and 0.5), resulting in 450 optimization runs. The total intervention cost was \$7,806,250.

3.10. Multi-Objective Optimization and Solution Selection

Pareto frontier analysis was conducted using two objectives: Total Priority Score and High-Risk Coverage (percentage of women with prior GDM with predicted risk >70%

selected). Non-dominated solutions were identified, and a composite scoring function was used for final solution selection:

$$CS = (0.4 \times NP) + (0.3 \times NC) + (0.2 \times NE) + (0.1 \times BU) \quad (3)$$

where CS represents the Composite Score, NP denotes Normalized Priority, NC refers to Normalized Coverage, NE indicates Normalized Efficiency, and BU stands for Budget Utilization.

3.11. Model Interpretability

SHapley Additive exPlanations (SHAP) analysis was performed to enhance model interpretability. A Kernel SHAP explainer was constructed using k-means clustering (k = 50) on a background sample of 200 women with prior GDM. SHAP values were then generated for 50 randomly selected women with prior GDM to quantify feature-level contributions to individual risk predictions.

4. Implementation and Results

4.1. Feature Selection Results

The multi-method feature selection approach identified 15 predictors with the highest discriminative power for T2DM risk. The combined feature ranking is presented in Figure 2. The selected top 15 features include insulin treatment during pregnancy, elevated HbA1c during pregnancy, macrosomia (birth weight > 3.5 kg), large for gestational age, pregnancy complications (hypertensive disorders), history of recurrence of GDM, abnormal oral glucose tolerance test (OGTT) results, pregnancy complications (preterm delivery), obesity or unhealthy postpartum weight gain, high pre-pregnancy BMI, physical inactivity, instrumental delivery, gestational weight gain, unhealthy diet, and older maternal age.

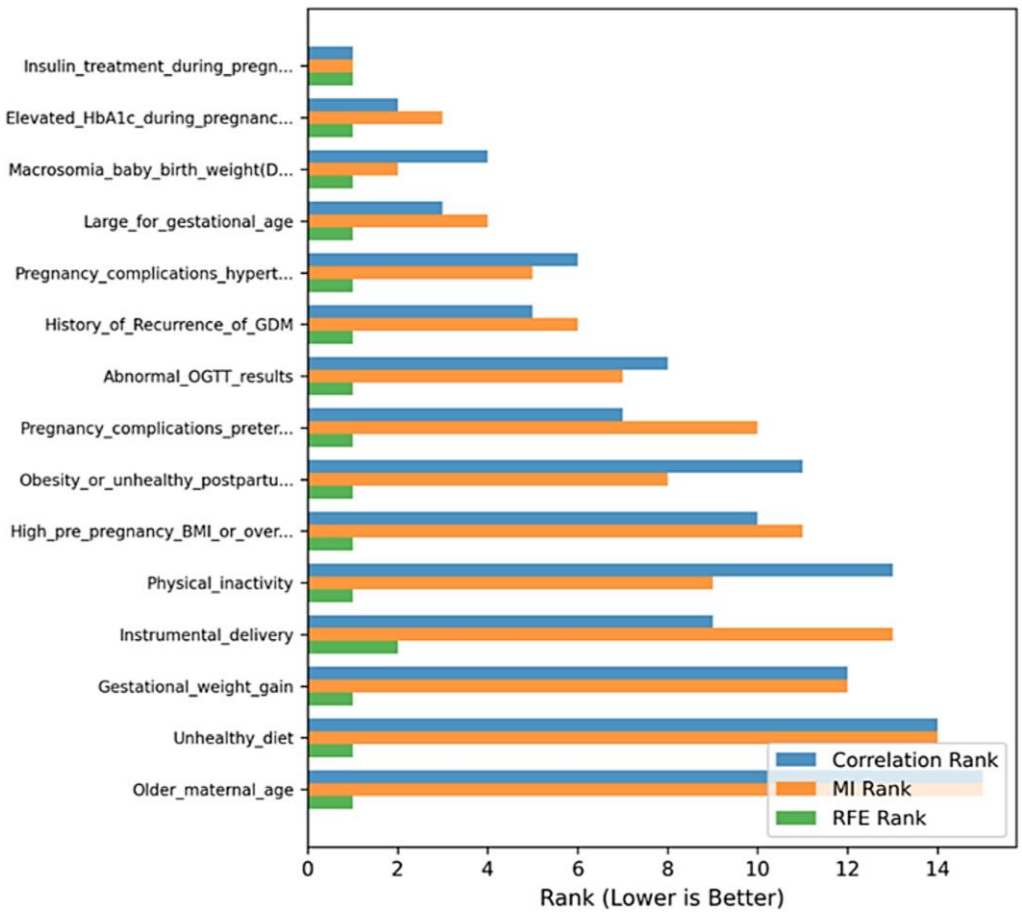


Figure 2. Combined feature ranking.

4.2. Predictive Model Performance

Five classification algorithms were evaluated using 10-fold stratified cross-validation. The performance metrics for all models are summarized in Table 3.

Table 3. Cross-validation performance of predictive models

Model	AUC-ROC (Mean ± SD)	Accuracy	Precision	Recall	F1-Score	Selected
Logistic Regression	0.9431 ± 0.0081	0.88	0.73	0.85	0.79	Best
Random Forest	0.9374 ± 0.0102	0.87	0.72	0.83	0.78	—
XGBoost	0.9294 ± 0.0169	0.86	0.74	0.82	0.77	—
LSTM	0.9335 ± 0.0089	0.87	0.73	0.84	0.78	—
CNN-1D	0.9246 ± 0.0167	0.85	0.74	0.81	0.76	—

Logistic Regression achieved the highest mean AUC-ROC of 0.9431 ± 0.0081, indicating strong discriminative performance and stability across folds. Random Forest attained a mean AUC-ROC of 0.9374 ± 0.0102, while XGBoost achieved 0.9294 ± 0.0169. Among the deep learning models, LSTM achieved an AUC-ROC of 0.9335 ± 0.0089 and CNN-1D achieved 0.9246 ± 0.0167. Based on these results, Logistic Regression was selected as the final prediction model.

The final Logistic Regression model, trained on the fully resampled training set, was evaluated on the independent hold-out test set (n = 1,200). The model demonstrated strong performance, achieving an accuracy of 0.8875, precision of 0.7378, recall of 0.8533, F1-score of 0.7913, and AUC-ROC of 0.9454. The confusion matrix is shown in Figure 3.

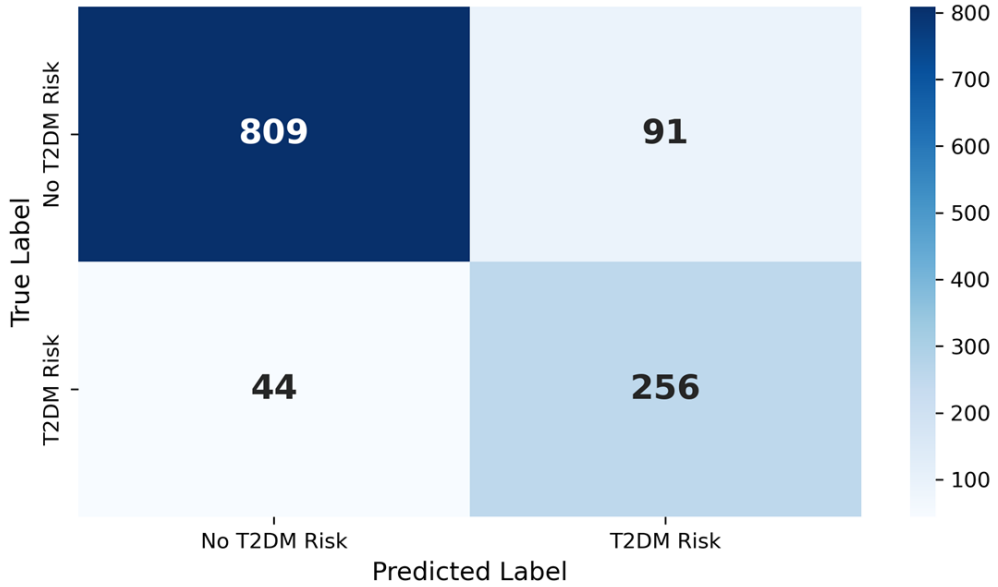


Figure 3. Confusion matrix of the logistic regression model

4.2.1. Ablation Study: Impact of SMOTE-ENN and Hyperparameter Optimization

To isolate the individual and combined contributions of SMOTE-ENN and hyperparameter optimization, an ablation study was conducted using Logistic Regression under four conditions:

- Baseline (default hyperparameters, no resampling)
- SMOTE-ENN only
- Hyperparameter optimization only
- Combined SMOTE-ENN and optimization

All configurations were evaluated using identical train–test splits with 10-fold cross-validation to ensure a fair comparison. The results are summarized in Table 4.

Table 4. Performance comparison across SMOTE-ENN and hyperparameter optimization conditions

Configuration	AUC-ROC	Accuracy	Precision	Recall	F1-Score	Δ AUC (%)
Condition 1: Baseline (No resampling; Default params)	0.8745	0.8367	0.7892	0.6733	0.7270	Baseline
Condition 2: SMOTE-ENN only (Default params)	0.8923	0.8450	0.7645	0.8133	0.7879	+2.0%
Condition 3: Optimization only (No resampling; Tuned params)	0.9212	0.8623	0.7245	0.7267	0.7724	+5.3%
Condition 4: Final model (SMOTE-ENN; Optimization)	0.9454	0.8817	0.7342	0.8533	0.7901	+8.1%

Note: All configurations were evaluated on the identical test set (n = 1,200) using Logistic Regression. Δ AUC (%) represents relative improvement over the baseline.

The ablation analysis provides several important insights. The baseline model (Condition 1) achieved moderate performance (AUC-ROC = 0.8745; recall = 0.6733), reflecting the difficulty of identifying positive cases under the 3:1 class imbalance. Applying SMOTE-ENN alone (Condition 2) substantially improved recall to 0.8133 (+20.9%), although with reduced precision (0.7645) and only modest AUC-ROC improvement (0.8923, +2.0%). Hyperparameter optimization alone (Condition 3) yielded a stronger AUC-ROC improvement (0.9212, +5.3%), but recall remained comparatively lower (0.7267), indicating that parameter tuning alone cannot fully address class imbalance.

The combined approach (Condition 4) produced the best overall performance, achieving an AUC-ROC of 0.9454 (+8.1%), high recall (0.8533), recovered precision (0.7342), and the highest accuracy (0.8817). These findings suggest that SMOTE-ENN and hyperparameter optimization address complementary aspects of model performance—namely class imbalance mitigation and decision boundary refinement—and that their joint application is important for achieving robust performance in imbalanced healthcare prediction tasks.

4.3. Risk Prediction

The optimal model was used to generate T2DM risk probabilities for all women with prior GDM. The resulting risk distribution is illustrated in Figure. 4.

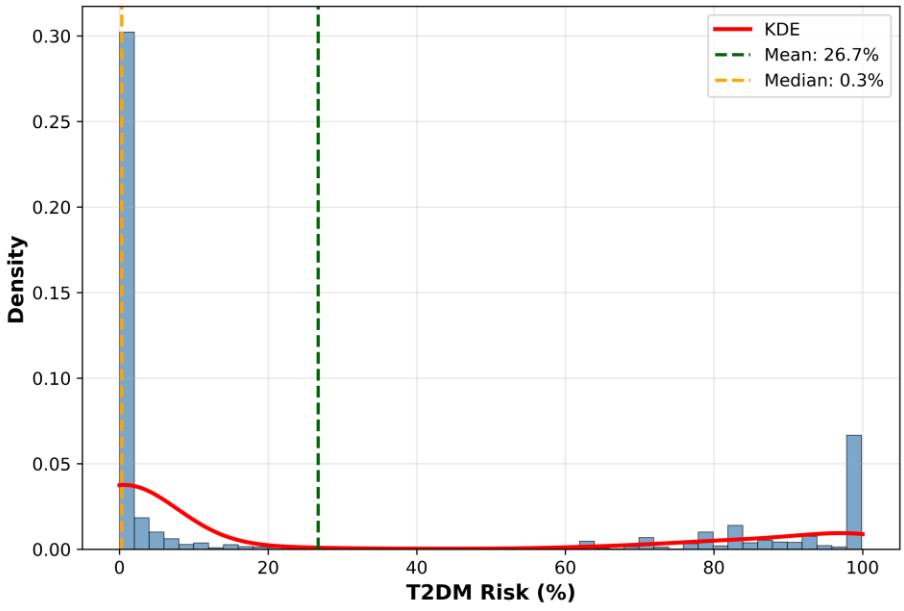


Figure 4. Risk distribution

The predicted T2DM risk distribution among women with prior GDM is distinctly bimodal and strongly right-skewed. Most individuals cluster near the low-risk region (median = 0.3%), while a smaller subgroup clusters near the upper-risk boundary. The substantial divergence between the median and the mean (26.7%) highlights the disproportionate influence of a relatively small high-risk subgroup on the overall distribution.

This pattern indicates that the model effectively stratifies the population into a large, low-risk majority and a smaller, yet clinically important, high-risk group. Such separation supports a risk-guided rather than uniform approach to post-GDM surveillance. Based on the predicted distribution, approximately one quarter of the cohort may warrant more intensive follow-up for T2DM prevention.

4.4. Feature Importance Analysis

SHapley Additive exPlanations (SHAP) analysis was performed to quantify feature-level contributions to model predictions. The mean absolute SHAP values for the selected predictors are summarized in Table 5.

Table 5. Feature importance based on mean absolute SHAP values.

Rank	Feature	Mean |SHAP|
1	Insulin treatment during pregnancy	0.099
2	History of recurrence of GDM	0.073
3	Obesity or unhealthy postpartum weight gain	0.063
4	Elevated HbA1c during pregnancy	0.055
5	Large for gestational age	0.031
6	Macrosomia (birth weight > 3.5 kg)	0.031
7	Physical inactivity	0.029
8	Pregnancy complications (hypertensive)	0.028
9	Abnormal OGTT results	0.016
10	Unhealthy diet	0.014

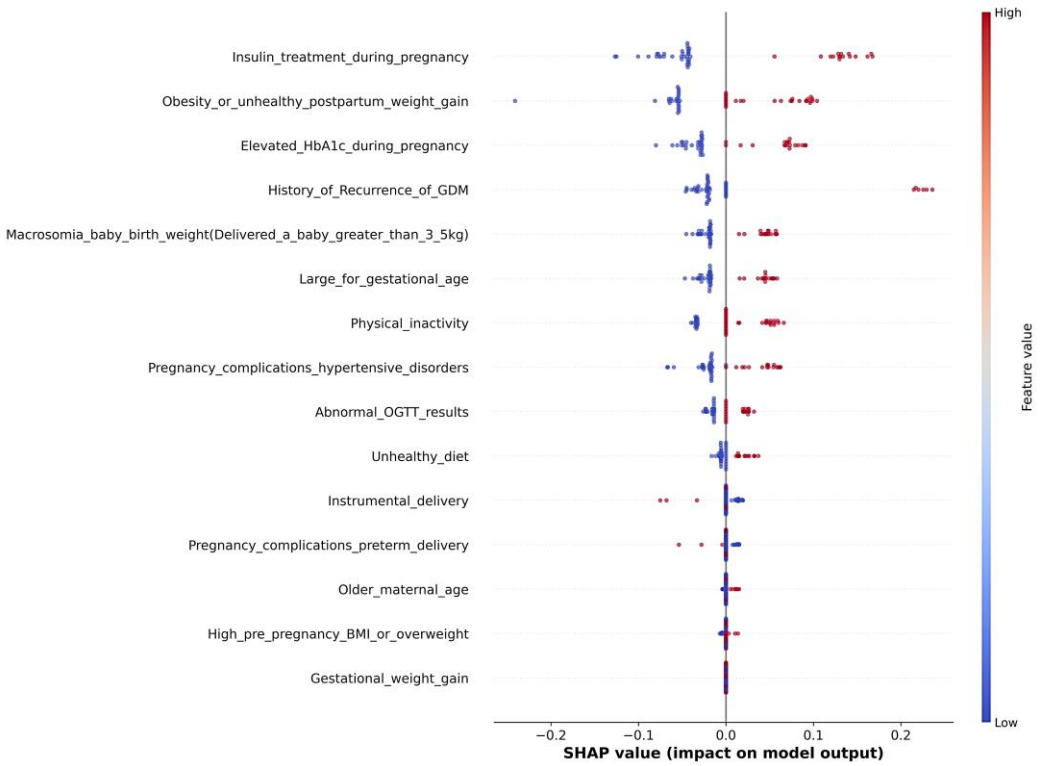


Figure 5. SHAP summary plot

Insulin treatment during pregnancy emerged as the most influential predictor (mean |SHAP| = 0.099), followed by history of GDM recurrence (0.073), obesity or unhealthy postpartum weight gain (0.063), and elevated HbA1c during pregnancy (0.055). Overall, the importance profile reflects a combination of clinical severity markers and modifiable lifestyle factors. Figure 5. SHAP summary plot illustrating feature contributions to T2DM risk predictions. Features are ordered by mean absolute SHAP value. Red points indicate higher feature values, whereas blue points indicate lower values. Positive SHAP values correspond to increased predicted risk.

4.5. Optimization Algorithm Performance

Nine optimization techniques were evaluated across 10 budget scenarios and five priority thresholds, resulting in a total of 450 optimization runs. Table 6 summarizes the aggregated performance metrics for each algorithm.

Table 6. Optimization algorithm performance summary (mean across all scenarios)

Algorithm	Mean Priority	Mean Women with Prior GDM	High-Risk Coverage	Efficiency
Linear Programming	901.47	1,199	65.26%	0.00044
Simulated Annealing	901.38	1,200	65.24%	0.00044
Dynamic Programming	901.37	1,200	65.24%	0.00044
Greedy (Efficiency)	901.37	1,200	65.24%	0.00044
Ant Colony	890.05	1,188	64.63%	0.00043
Greedy (Priority)	869.63	1,117	63.37%	0.00042
Genetic Algorithm	774.53	1,397	56.81%	0.00039
Particle Swarm	704.25	1,159	51.75%	0.00036

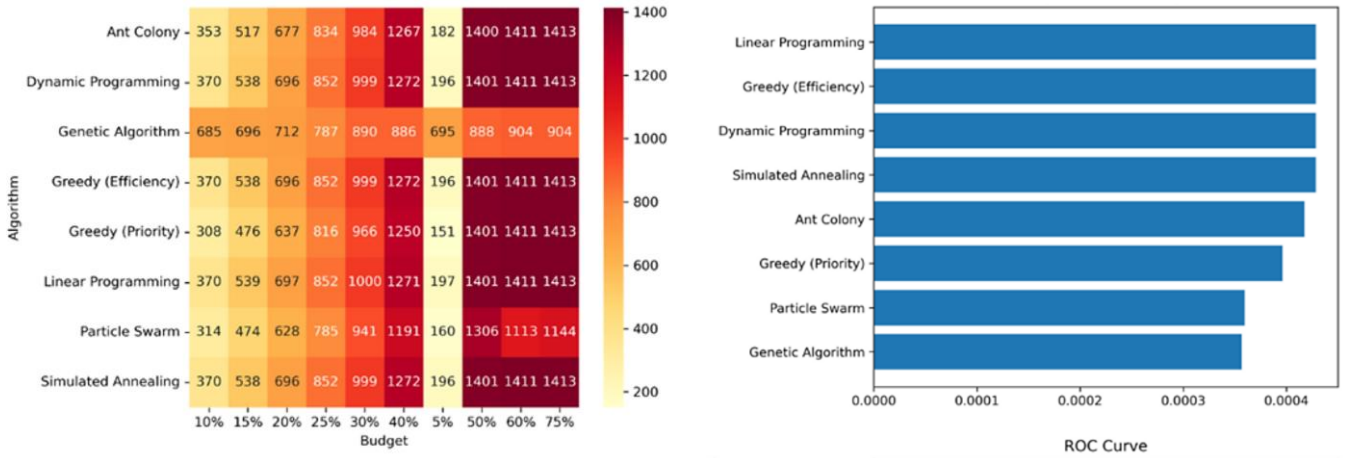
Linear Programming (LP) achieved the highest mean overall priority score (901.47) while maintaining consistent performance across budget levels. Simulated Annealing (901.38) and Dynamic Programming (901.37) demonstrated near-equivalent performance, with Greedy (Efficiency) matching Dynamic Programming. The Genetic Algorithm (GA), despite a lower mean priority score (774.53), selected, on average, a larger number of women (1,397), suggesting broader solution-space exploration. Particle Swarm Optimization (PSO) exhibited the lowest mean performance (704.25). The Differential Evolution (DE) algorithm was excluded from further analysis due to technical constraints encountered during parallel computation. Table 7 presents the optimal solutions identified for each budget scenario.

Table 7. Optimal solutions by budget scenario

Budget	Amount	Best Algorithm	Women with Prior GDM	Priority	High-Risk Coverage
5%	$390,313	Genetic Algorithm	908	695.17	50.5%
10%	$780,625	Genetic Algorithm	2,958	685.33	48.5%
15%	$1,170,938	Genetic Algorithm	2,970	695.72	49.3%
20%	$1,561,250	Genetic Algorithm	2,997	712.17	50.7%
25%	$1,951,563	Linear Programming	1,051	852.32	60.5%
30%	$2,341,875	Linear Programming	1,238	999.75	71.3%
40%	$3,122,500	Simulated Annealing	1,603	1,271.57	93.2%
50%	$3,903,125	Linear Programming	2,123	1,400.71	100.0%
60%	$4,683,750	Simulated Annealing	2,749	1,410.86	100.0%
75%	$5,854,688	Linear Programming	3,633	1,413.17	100.0%

At lower budget levels (5–20%), the GA tended to identify solutions that selected more women with prior GDM but yielded lower total priority scores. For moderate to high budget levels (25–75%), LP and Simulated Annealing consistently identified the highest-quality solutions. At the 25% budget level ($1,951,563), LP selected 1,051 women with prior GDM,

achieving a total priority score of 852.32 and 60.5% high-risk coverage. At the highest budget level (75%), full coverage of high-risk women (100%) was achievable, with 3,633 women prioritized. The heatmap in Fig. 6(a) visualizes total priority scores achieved by each algorithm across budget levels, while Fig. 6(b) illustrates the mean algorithmic efficiency (priority per dollar spent).



(a) (b)

Figure 6. Optimization results across budget scenarios. (**a**) Heatmap of total priority scores achieved by each algorithm at different budget levels; (**b**) Mean algorithmic efficiency measured as priority per dollar spent.

4.6. Multi-Objective Optimization and Pareto Analysis

Pareto frontier analysis identified four non-dominated solutions that jointly optimize total priority and high-risk coverage. These solutions were observed at the 75% budget level using Linear Programming (LP), Simulated Annealing (SA), Greedy (Efficiency), and Dynamic Programming (DP), each achieving a total priority of 1,413.17 with 100% high-risk coverage.

The Pareto frontier, shown in Figure 7, highlights the trade-off between maximizing total priority and achieving high-risk coverage. At lower budget levels, clear trade-offs emerge between maximizing aggregate priority and ensuring comprehensive coverage of the highest-risk women with prior GDM. As the budget increases, the solution space converges toward full high-risk coverage.

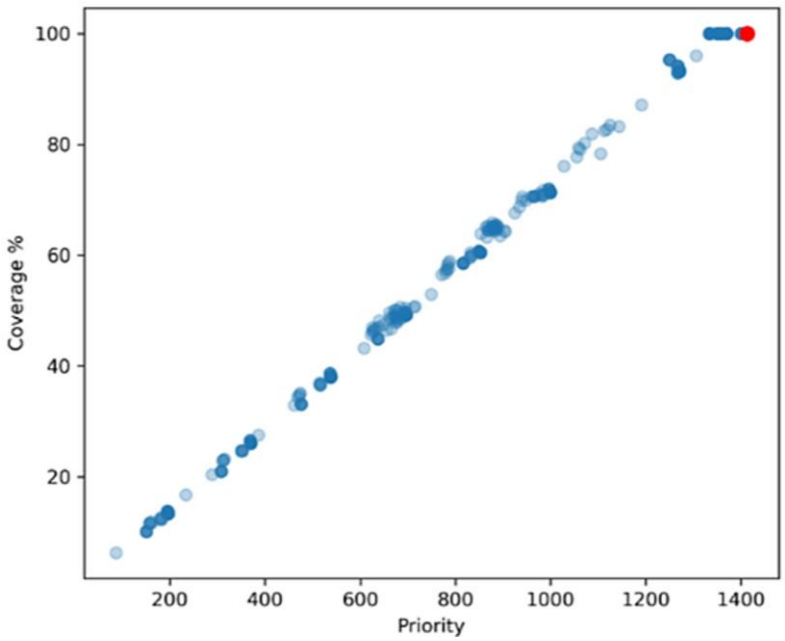


Figure 7. Pareto frontier analysis.

4.7. Optimal Resource Allocation Analysis

4.7.1. Selection of the Recommended Solution

The optimization framework evaluated multiple algorithms across budget scenarios ranging from 5% to 75% of total intervention costs. Although the composite score provides a useful aggregate indicator that combines priority, coverage, efficiency, and budget utilization, selecting a practically deployable solution requires consideration of feasibility constraints and public health objectives. Table 8 summarizes the highest composite score solution for each budget category.

Table 8. Best composite score by budget category.

Budget	Budget Amount ($)	Algorithm	Women with Prior GDM Selected	Total Priority	Cost Used ($)	Budget Utilization (%)	High-Risk Coverage (%)	Efficiency	Composite Score
5%	390,312.50	GA	2,881	655.67	3,735,150.00	956.96	46.46	0.000176	1.3517
10%	780,625.00	GA	2,958	685.33	3,814,900.00	488.70	48.53	0.000180	0.8997
15%	1,170,937.50	GA	2,970	695.72	3,867,450.00	330.29	49.34	0.000180	0.7468
20%	1,561,250.00	GA	2,997	712.17	3,921,200.00	251.16	50.66	0.000182	0.6769
25%	1,951,562.50	LP	1,051	852.32	1,951,450.00	99.99	60.50	0.000437	0.6964
30%	2,341,875.00	LP	1,222	996.63	2,341,950.00	100.00	71.91	0.000426	0.7670
40%	3,122,500.00	LP	1,582	1,267.09	3,121,800.00	99.98	94.17	0.000406	0.9025
50%	3,903,125.00	LP	2,123	1,400.71	3,903,550.00	100.01	100.00	0.000359	0.9391
60%	4,683,750.00	LP	2,748	1,410.86	4,683,250.00	99.99	100.00	0.000301	0.9191
75%	5,854,687.50	LP	1,782	1,370.21	3,469,650.00	59.26	100.00	0.000395	0.9041

Inspection of Table 8 shows that GA solutions at lower budget levels (5–20%) achieve higher composite scores (0.6769–1.3517). However, these solutions substantially exceed their allocated budget limits, with utilization ranging from 251% to 957% of the designated budget. Consequently, despite favorable composite scores, these configurations are not feasible for real-world deployment under fixed public health budget constraints.

The 25% budget scenario using LP emerges as a practically balanced solution when considering both optimization performance and implementation feasibility. This configuration achieves 99.99% budget utilization, ensuring near-complete deployment of allocated resources without exceeding fiscal constraints. Table 9 summarizes the key characteristics of this recommended configuration.

Table 9. Recommended solution characteristics (25% budget, LP).

Parameter	Value
Budget Allocated	$1,951,562.50
Budget Utilized	$1,951,450.00
Budget Utilization Rate	99.99%
Women with Prior GDM Selected	1,051
Total Priority Score	852.32
High-Risk Coverage	60.50%
Efficiency (Priority per Dollar)	0.000437
Composite Score	0.6964
Execution Time	0.20 seconds

Table 10 provides a comparative analysis of algorithm performance at the 25% budget level.

Table 10. Optimal solutions by budget scenario

Algorithm	Women with Prior GDM Selected	Total Priority	Budget Utilization (%)	High-Risk Coverage (%)	Efficiency	Composite Score
Linear Programming	1,051	852.32	99.99	60.50	0.000437	0.6964
Simulated Annealing	1,051	852.32	99.99	60.50	0.000437	0.6964
Greedy (Efficiency)	1,051	852.32	99.99	60.50	0.000437	0.6964
Ant Colony	1,045	831.49	99.99	59.75	0.000426	0.6840
Greedy (Priority)	936	816.41	100.00	58.62	0.000418	0.6733
Genetic Algorithm	2,931	691.37	195.34	48.90	0.000181	0.6098
Particle Swarm	1,459	386.24	99.97	27.59	0.000198	0.3708

4.7.2 Characteristics of Prioritized Women

Analysis of the 1,051 selected women revealed consistent clinical patterns among the highest-priority cases. Individuals with priority scores exceeding 0.91 generally exhibited T2DM risk probabilities above 99.7% (very high risk) and presented multiple modifiable risk factors, including obesity, physical inactivity, and unhealthy dietary patterns. The estimated intervention cost for these high-priority individuals was approximately $2,250 per patient.

The mean predicted risk probability among the selected cohort was substantially higher than the overall population average, indicating that the prioritization framework effectively concentrates resources on individuals with elevated projected risk. Table 11 lists the top 10 women with prior GDM who are prioritized for intervention.

Table 11. Top 10 prioritized women with prior GDM for intervention

Patient ID	T2DM Risk Probability	Priority Score	Intervention Cost ($)
1886	0.998465	0.912883	2250.0
1405	0.998465	0.912883	2250.0
2601	0.998465	0.912883	2250.0
3546	0.998465	0.912883	2250.0
548	0.998465	0.912883	2250.0
4733	0.998465	0.912883	2250.0
627	0.997881	0.912348	2250.0
5477	0.997804	0.912278	2250.0
929	0.997804	0.912278	2250.0
3443	0.997804	0.912278	2250.0

Figure 8 presents a SHAP waterfall plot illustrating the individual feature contributions to the model prediction for a representative high-risk case. This visualization enhances interpretability by enabling clinicians to identify the primary drivers of predicted risk at the individual level, thereby supporting more informed clinical assessment and targeted preventive planning.

5. Discussion

5.1. Primary Findings

This study developed and validated an integrated framework combining machine-learning–based risk prediction with multi-algorithm optimization to prevent T2DM among women with prior GDM. The findings demonstrate the feasibility and practical relevance of this two-phase approach for evidence-informed prioritization under resource constraints. Importantly, the primary contribution of this work lies not in achieving the highest AUC (although the predictive performance is competitive), but in systematically integrating risk prediction with optimization-based prioritization within realistic budget constraints.

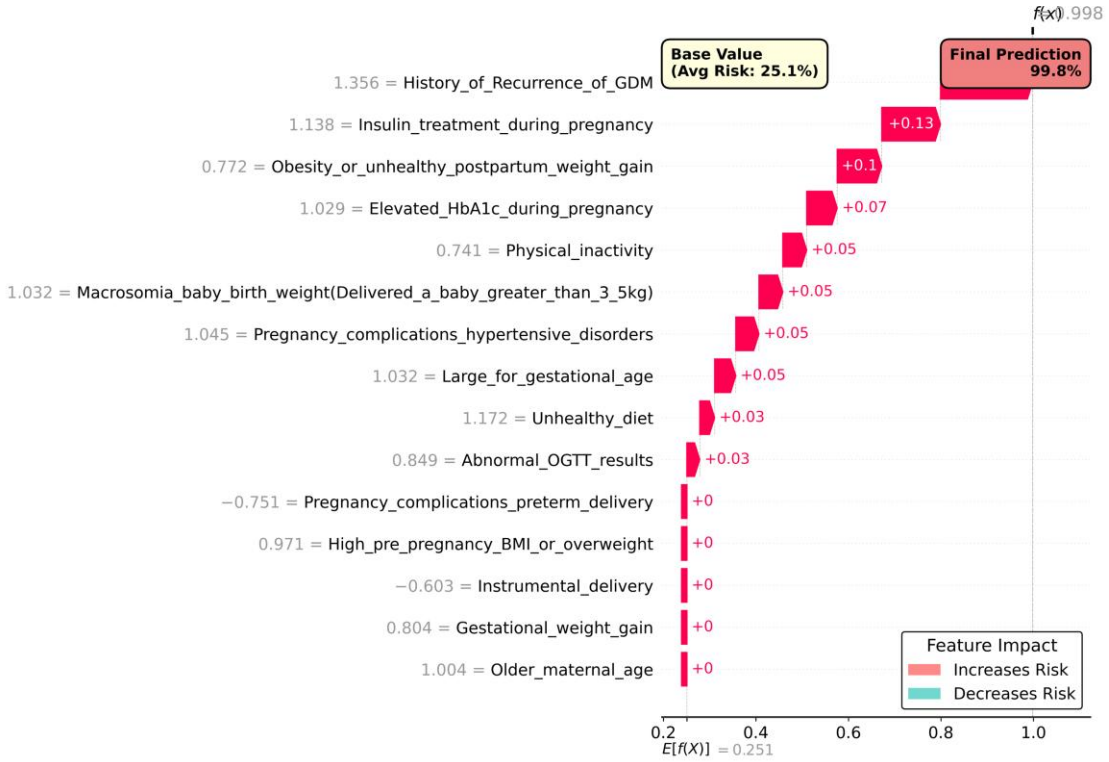


Figure 8. SHAP waterfall plot for a representative high-risk woman with prior GDM.

5.1.1. Predictive model performance.

The Logistic Regression model demonstrated strong discriminative performance, achieving an AUC-ROC of 0.9454 on the independent hold-out test set and outperforming more complex approaches, including deep learning architectures. This observation aligns with prior evidence indicating that, for structured tabular clinical data with well-defined predictors, classical machine learning models often perform comparably to—or in some cases better than—deep learning models while maintaining superior interpretability and computational efficiency. The model also achieved high recall (0.8533), which is particularly valuable in screening contexts where identifying true positive cases is prioritized over minimizing false positives.

5.1.2. Key risk factors.

SHAP analysis identified insulin treatment during pregnancy as the most influential predictor of subsequent T2DM risk (mean |SHAP| = 0.099). This result is clinically plausible, as insulin requirement during pregnancy typically reflects more severe glucose intolerance and underlying beta-cell dysfunction. History of recurrent GDM (|SHAP| = 0.073) emerged as the second most important predictor, highlighting the cumulative metabolic burden associated with repeated gestational hyperglycemia. Notably, several modifiable lifestyle factors—including obesity or unhealthy postpartum weight gain (|SHAP| = 0.063), physical inactivity (|SHAP| = 0.029), and unhealthy diet (|SHAP| = 0.014)—also contributed meaningfully to risk estimates, underscoring potential targets for preventive intervention.

5.1.3. Optimization performance.

Among the nine optimization techniques evaluated, Linear Programming (LP) consistently achieved the highest total priority scores across budget scenarios, with a mean priority of 901.47 and 65.26% high-risk coverage. The deterministic optimality properties of LP appear well suited to this resource allocation structure. Simulated Annealing and Dynamic Programming produced near-equivalent results, indicating robustness across established optimization paradigms. Interestingly, the computationally simpler Greedy (Efficiency) heuristic achieved performance comparable to that of Dynamic Programming, suggesting that, for this specific problem structure, highly complex metaheuristics may not always yield additional practical benefit.

5.1.4. Budget-dependent resource allocation.

The relationship between budget allocation and high-risk coverage reveals important operational trade-offs for healthcare planning. At a 25% budget allocation ($1.95 million), approximately 60.5% of very high-risk women with prior GDM could be prioritized for intervention. Full coverage (100%) of high-risk women was achieved at the 50% budget level ($3.90 million), after which marginal gains diminished. These quantified budget–coverage relationships provide actionable planning insight for healthcare administrators when setting funding levels and coverage targets.

5.1.5. Pareto-optimal solutions.

Multi-objective optimization identified four Pareto-optimal solutions that jointly maximized total priority and high-risk coverage. These solutions emerged at the 75% budget level, each achieving comparable performance (priority = 1,413.17; coverage = 100%). This convergence suggests that at higher budget levels multiple algorithms can reach similar optimal frontiers, whereas under tighter budget constraints algorithm selection becomes more consequential.

5.2. Contextual Comparison: Model Performance in Literature and Methodological Positioning

The present study contributes to the T2DM prevention literature in women with prior GDM by proposing an integrated framework that combines machine-learning–based risk prediction with multi-algorithm optimization for budget-constrained prioritization. Rather than focusing solely on predictive accuracy, the framework emphasizes the operational translation of risk estimates into actionable resource allocation strategies.

Table 12 provides a contextual comparison with representative studies across both prediction and optimization domains. The comparison highlights that most prior work has addressed either risk prediction or resource optimization in isolation, whereas relatively few studies have attempted to bridge these components within a unified decision-support pipeline.

Table 12. Contextual comparison: model performance in literature and methodological positioning

Study	Domain/Scope	Best Algorithm	Key Metrics	Key Findings	Integration with Resource Allocation
Khan et al. [44]	T2DM prediction (GDM→T2DM)	Decision Tree	AUC: 0.92; Acc: 91%	Real-world + omics	Prediction only
Li et al. [39]	T2DM prediction (postpartum)	Cox model	AUROC: 82.8%	Non-invasive clinical features	Prediction only
Lai et al. [45]	T2DM prediction (metabolic)	Not specified	AUC: 0.883	Metabolic signatures at 6–9 weeks postpartum	Prediction only
Joglekar et al. [46]	T2DM prediction (miRNA)	Penalized logistic regression	AUC: 0.92 (miRNA); 0.83 (clinical)	miRNAs enhance prediction	Prediction only
Kumar et al. [51]	T2DM prediction (pregnancy)	CatBoost	AUC: 0.86	Mid-pregnancy clinical features	Prediction only
Houri et al. [50]	T2DM prediction (glucose)	XGBoost	AUC: 0.85; Acc: 91%	Glucose metabolism patterns	Prediction only
Ilari et al. (2022) [46]	T2DM prediction (GDM history)	L2-LogReg	AUC: 0.884; F1: 0.83	ML comparison in GDM women	Prediction only
Belsti et al. [52]	T2DM prediction (antenatal/postnatal)	Not stated	AUC: 0.76–0.85	Antenatal and postnatal models	Prediction only
D'Aeth et al. [54]	Hospital resource optimization	Linear Programming	50,750–5.89M life-years saved	LP improves allocation efficiency	Optimization only (LP)
Mizan & Taghipour [59]	Medical resource allocation	ML + optimization	Reduced waiting times	Hybrid approach	Optimization only (multi-objective)
Stuart et al. [55]	Health resource policy	Constrained optimization	Evidence-based allocation	Budget-constrained prioritization	Optimization only (constrained)
Zuo et al. [60]	Outpatient scheduling	GTARS-DRL vs GA	ACPU: 0.762 s	Deep RL matched GA quality	Optimization only (scheduling RL)

Study	Domain/Scope	Best Algorithm	Key Metrics	Key Findings	Integration with Resource Allocation
Zhou et al. [75]	CVD prevention	Duramax RL	Policy value: 93 vs 68	Long-term preventive optimization	Optimization only (treatment RL)
Wang et al. [73]	T2DM insulin titration	RL-DITR	MAE: 1.10±0.03 U	Improved insulin dosing	Optimization only (dosage RL)
Lee et al. [82]	Acute dyspnea treatment	PROP-RL	Mortality: 3.8% → 2.2%	Robust offline RL policy	Optimization only (treatment RL)
Arora et al. [83]	System-wide healthcare allocation	ML ensemble	Framework study	Integrated resource framework	Both
Frommeyer et al. [84]	Precision medicine DTR	Various RL	46 studies reviewed	Multi-specialty RL applications	Optimization-focused review
Wu et al. [85]	Healthcare operations	RL methods	321 documents	Multi-level optimization	Optimization-focused review
Teo et al. [86]	Generative AI in medicine	Foundation models	Review	End-to-end AI pipelines	Both
Li et al. [87]	Medical AI scaling	Context-switching paradigm	Adaptive inference	Cross-context robustness	Both
Neuwahl et al. [66]	GDM screening & prevention	Decision Tree	ICER: $48,588–$122,279/QALY	Cost-effectiveness	Both (economic focus)
Lloyd et al. [67]	GDM prevention	Decision modeling	ROI: AU$1.22 per $1	Cost-saving strategy	Both (economic focus)
This study	Integrated T2DM prediction + prioritization	Logistic Regression + LP	AUC: 0.9454; Acc: 88.75%; F1: 0.79; Priority: 901.47	Integrated predictive–prescriptive framework with SHAP interpretability	Both (ML prediction + budget-constrained optimization)

Overall, the comparison indicates that prior studies have predominantly emphasized either predictive modeling accuracy or standalone optimization. In contrast, the present work focuses on the operational linkage between risk estimation and budget-constrained prioritization. The results suggest that such integration can provide complementary decision support for population-level T2DM prevention planning among women with prior GDM. Importantly, the contribution of this study should be interpreted as methodological integration rather than solely predictive superiority. While the achieved AUC-ROC is competitive relative to prior work, the primary added value lies in demonstrating how risk predictions can be systematically translated into resource allocation decisions under explicit budget constraints.

5.2.1. Competitive Performance in Risk Prediction with Enhanced Applicability

With respect to predictive performance, the Logistic Regression model achieved an AUC-ROC of 0.9454, indicating competitive discrimination relative to leading studies in the literature. Importantly, the principal contribution of this work lies not in pursuing the highest possible AUC, but in integrating risk prediction with optimization-based prioritization under explicit resource constraints—a gap that remains insufficiently addressed in translational research.

Khan et al. [44] reported an AUC of 0.92 using Decision Trees combined with omics data, while Joglekar et al. [46] achieved comparable performance (AUC: 0.92) through circulating miRNA integration. However, these approaches rely on specialized biomarker assays. In contrast, the present model utilizes routinely collected clinical variables, which may enhance practical deployability, particularly in resource-constrained settings. Similarly, Lai et al. [45] achieved an AUC of 0.883 using metabolic signatures measured at 6–9 weeks postpartum, which also require additional laboratory assessments.

When compared with machine learning models based on standard clinical features—such as Kumar et al. [51] using CatBoost (AUC: 0.86), Houri et al. [44] using XGBoost (AUC: 0.85), Ilari et al. [52] using L2-penalized Logistic Regression (AUC: 0.884), and Li et al. [39] using Cox models (AUROC: 82.8%)—the proposed model demonstrates strong and competitive discrimination while maintaining interpretability through SHAP analysis. Belsti et al. [42] reported AUC values ranging from 0.76 to 0.85 for antenatal and postnatal models, further situating the present results within the upper range of reported performance for T2DM prediction among women with prior GDM.

5.2.2. Bridging the Optimization Gap in Preventive Care

Recent healthcare resource allocation studies have largely emphasized operational efficiency rather than preventive prioritization. D'Aeth et al. [54] demonstrated the effectiveness of Linear Programming (LP) for hospital bed allocation, while Mizan and Taghipour [59] explored hybrid ML–optimization approaches to reduce patient waiting times, and Stuart et al. [55] applied constrained optimization for evidence-based health resource allocation. Zuo et al. [60] further investigated deep reinforcement learning for outpatient scheduling. However, these studies did not explicitly address budget-constrained patient prioritization for disease prevention.

More recent reinforcement learning approaches have shown promise in treatment optimization. Zhou et al. [75] reported a 6% reduction in cardiovascular disease risk through long-term lipid control (policy value: 93 vs. clinicians' 68), Wang et al. [73] achieved improved insulin titration with 90.2% physician adoption, and Lee et al. [82] reported mortality reduction from 3.8% to 2.2% in acute dyspnea management. While these studies optimize treatment policies, they do not directly address preventive prioritization under explicit budget constraints.

Comprehensive frameworks (Arora et al. [83]), systematic reviews (Frommeyer et al. [84]; Wu et al. [85]), and recent advances in generative AI [86] and context-switching paradigms [87] have broadened the methodological landscape. Nevertheless, the integration of prediction with budget-constrained preventive prioritization for women with prior GDM remains limited.

The few GDM-focused economic optimization studies—Neuwahl et al. [66] and Lloyd et al. [67]—primarily evaluated cost-effectiveness (ICER: $48,588–$122,279/QALY; ROI: AU$1.22 per $1 invested) rather than operational patient selection under resource constraints.

In this context, the present study systematically evaluates nine optimization algorithms across 10 budget scenarios. LP consistently achieved the highest mean total priority (901.47) with 65.26% high-risk coverage. The analysis further indicates that full coverage of high-risk women with prior GDM becomes achievable at the 50% budget level ($3.90 million), providing quantitative planning insight for healthcare administrators. Additionally, the identification of Pareto-optimal solutions at the 75% budget level supports flexible multi-objective decision-making.

5.2.3. Novel Integrated Framework for Translational Impact

Unlike prior studies that typically examine prediction and optimization separately (see Table 12), the proposed framework establishes a continuous translational pathway from risk stratification to resource allocation. While Arora et al. [83] outlined a system-level resource framework and recent advances in generative AI [86] and medical AI scaling [87] have improved prediction and deployment capabilities, few studies have specifically integrated ML-based T2DM risk prediction with budget-constrained prioritization for women with prior GDM.

The proposed framework combines interpretable risk estimation (via SHAP) with systematic optimization benchmarking to address key operational questions: (i) which women should be prioritized for intervention; (ii) how many high-risk individuals can be covered under varying budget levels; and (iii) how trade-offs among coverage, efficiency, and resource utilization can be balanced. By empirically evaluating these dimensions through structured algorithm comparison and budget sensitivity analysis, the study provides a practical blueprint for evidence-informed prioritization in preventive care settings.

5.3. Strengths

Several methodological strengths support the robustness of the present study. The validation strategy—combining 10-fold stratified cross-validation with an independent hold-out test set—provides reliable generalization estimates. Importantly, SMOTE-ENN resampling was applied exclusively within training folds, minimizing the risk of data leakage. The comprehensive benchmarking of nine optimization algorithms across 450 scenarios offers detailed comparative insight for healthcare decision contexts. Pareto frontier analysis explicitly acknowledges competing objectives in resource allocation, enabling flexible decision-making rather than enforcing a single prescriptive solution.

Model interpretability is enhanced through SHAP analysis, which clearly identifies modifiable risk factors that may support targeted preventive strategies. The composite priority

score—integrating both clinical risk and intervention amenability—helps ensure that selected individuals are not only high risk but also potentially responsive to lifestyle interventions. Finally, the sub-second execution time of LP (<0.5 seconds) indicates practical feasibility for near–real-time clinical decision support environments.

5.4. Limitations

The primary limitation of this study is the reliance on synthetic rather than real-world clinical data. Although the dataset was carefully constructed to reflect plausible epidemiological patterns, synthetic data cannot fully capture real-world complexity, measurement error, and unobserved confounding. Consequently, the observed AUC-ROC of 0.9454 should be interpreted cautiously and may represent an upper-bound estimate; real-world performance would likely decline in the presence of missing data, incomplete documentation, and loss to follow-up. In addition, single-source development limits generalizability across populations with differing demographics, healthcare access, and risk factor distributions, underscoring the need for external validation.

A further limitation is the absence of a comprehensive calibration assessment. While discrimination metrics were reported, the study did not assess agreement between predicted probabilities and observed event rates—an important consideration for clinical decision-making, where absolute risk estimates inform resource allocation. Calibration plots, calibration slope, calibration-in-the-large, and decision curve analysis were not performed; therefore, the potential for systematic risk misestimation cannot be excluded.

The optimization framework also has technical limitations that currently preclude direct clinical deployment. Most notably, the Genetic Algorithm did not strictly enforce budget constraints, resulting in infeasible solutions that exceeded the budget limits (e.g., up to 957% utilization under a 5% constraint). Formal Integer Linear Programming (ILP) or Mixed-Integer Linear Programming (MILP) formulations were not implemented, and feasibility repair mechanisms or penalty-based constraint handling were not incorporated. These aspects require further methodological refinement before operational use.

Moreover, the cross-sectional dataset captures predictors at a single time point without longitudinal follow-up linked to observed T2DM incidence, which limits confidence in long-term prognostic validity. Future work incorporating longitudinal trajectories and time-varying covariates could improve risk estimation and enable adaptive prevention strategies. The simplified tiered intervention cost model—based primarily on the count of modifiable risk factors—does not fully reflect real-world cost heterogeneity across healthcare systems, delivery modalities, or patient adherence patterns.

Finally, modeling T2DM risk as a binary outcome omits clinically relevant intermediate glycemic states. Multi-class classification or time-to-event survival modeling could provide more nuanced risk stratification across the continuum of glycemic deterioration. Collectively, these limitations indicate that the present work should be interpreted as a methodological proof-of-concept requiring further real-world validation prior to clinical deployment.

5.5. Clinical Implications

Despite the above limitations, the proposed framework illustrates the potential utility of risk-stratified postpartum surveillance by directing preventive resources toward women at elevated T2DM risk. The finding that approximately 60.5% high-risk coverage may be achieved at 25% of total potential spending provides a quantitative reference point for budget planning and payer discussions.

SHAP-based interpretability further supports individualized intervention planning. For example, women identified with obesity-related risk profiles may be prioritized for dietary and weight-management programs, whereas those with prominent physical inactivity signals may benefit from structured exercise interventions. Such differentiation may improve alignment between risk drivers and preventive strategies.

From a health system perspective, the observation that full high-risk coverage becomes achievable at approximately the 50% budget level offers an evidence-informed planning benchmark. However, these estimates should be validated in real-world cohorts before being used for policy decisions.

5.6. Future Work

Prospective cohort validation following women with prior GDM over 5–10 years is a critical next step to establish real-world performance. Such studies would enable comprehensive calibration assessment, including calibration plots, calibration slope, calibration-in-the-large, and decision curve analysis, thereby complementing the discrimination metrics reported here. From an optimization standpoint, the framework would benefit from reformulation using ILP or MILP with strict constraint enforcement, explicit feasibility filtering, and transparent reporting of feasible solution rates to support fair algorithm comparison and guaranteed budget adherence.

Multi-site external validation across diverse healthcare systems is also necessary to evaluate model transportability and identify potential recalibration needs, particularly given that calibration often degrades across populations even when discrimination remains acceptable. Integration with formal health economic modeling could further enable cost-effectiveness evaluation incorporating QALYs and long-term cost savings.

Future methodological extensions may include longitudinal modeling with time-varying covariates, heterogeneous treatment effect modeling, and algorithmic fairness auditing across demographic subgroups to support equitable precision prevention. Embedding the framework within electronic health record–based clinical decision support systems would facilitate real-world deployment and enable continuous learning from observed outcomes.

5.7. Ethical Considerations

AI-driven prioritization systems raise important fairness and governance considerations. Prior to deployment, rigorous subgroup performance evaluation across demographic dimensions—including race/ethnicity, socioeconomic status, and geography—is necessary to mitigate the risk of amplifying existing healthcare disparities.

Transparent and explainable predictions, combined with appropriate human oversight, appeal mechanisms, and procedural safeguards, are essential for responsible clinical use. Any real-world implementation would require institutional review board approval and alignment with established healthcare AI governance frameworks (e.g., WHO, NIH, FDA), alongside ongoing monitoring for algorithmic bias and unintended consequences.

6. Conclusions

This study presents an integrated framework combining machine learning–based risk prediction with multi-algorithm optimization for T2DM prevention among women with prior GDM. The Logistic Regression model achieved strong discriminative performance (AUC = 0.9454), while SHAP analysis highlighted insulin treatment during pregnancy, GDM recurrence history, and postpartum weight status as influential predictors. Across optimization experiments, Linear Programming consistently produced high-quality prioritization solutions, with approximately 60.5% high-risk coverage achievable at a 25% budget level and full coverage observed at 50% allocation within the simulated setting. By linking predictive modeling with formal resource allocation, the framework illustrates a potential pathway from risk estimation to operational decision support.

Importantly, the present study should be interpreted as a methodological proof-of-concept based on synthetic data. Substantial real-world validation—including prospective evaluation, external generalizability testing, calibration assessment, and workflow integration—remains necessary before clinical adoption. Nevertheless, the publicly available framework provides a structured foundation for future validation studies aimed at determining whether integrated predictive–prescriptive approaches can translate into measurable improvements in population-level diabetes prevention and patient outcomes..

Author Contributions: Conceptualization: A.P. and J.P.; Methodology: A.P.; Validation: J.P.; Formal analysis: A.P.; Investigation: J.P.; Data curation: A.P.; Writing—original draft preparation: J.P.; Writing—review and editing: A.P.; Visualization: A.P.; Project administration: A.P. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Data Availability Statement: Code available**:** https://github.com/DataInsighty/Machine-Learning-and-Multi-Algorithm-Optimization-Framework-for-Type-2-Diabetes-Mellitus-.git, Data source available: https://doi.org/10.34740/kaggle/dsv/13183093.

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

References

[1] D. Magliano and E. J. Boyko, *IDF Diabetes Atlas*, 10th ed. Brussels: International Diabetes Federation, 2021.
[2] Y. Zheng, S. H. Ley, and F. B. Hu, "Global aetiology and epidemiology of type 2 diabetes mellitus and its complications," *Nat. Rev. Endocrinol.*, vol. 14, no. 2, pp. 88–98, Feb. 2018, doi: 10.1038/nrendo.2017.151.
[3] Diabetes Prevention Program Research Group, "Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin," *N. Engl. J. Med.*, vol. 346, no. 6, pp. 393–403, Feb. 2002, doi: 10.1056/NEJMoa012512.
[4] L. Bellamy, J.-P. Casas, A. D. Hingorani, and D. Williams, "Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis," *Lancet*, vol. 373, no. 9677, pp. 1773–1779, May 2009, doi: 10.1016/S0140-6736(09)60731-5.
[5] C. Kim, K. M. Newton, and R. H. Knopp, "Gestational Diabetes and the Incidence of Type 2 Diabetes," *Diabetes Care*, vol. 25, no. 10, pp. 1862–1868, Oct. 2002, doi: 10.2337/diacare.25.10.1862.
[6] E. Vounzoulaki, K. Khunti, S. C. Abner, B. K. Tan, M. J. Davies, and C. L. Gillies, "Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis," *BMJ*, p. m1361, May 2020, doi: 10.1136/bmj.m1361.
[7] T. A. Buchanan and A. H. Xiang, "Gestational diabetes mellitus," *J. Clin. Invest.*, vol. 115, no. 3, pp. 485–491, Mar. 2005, doi: 10.1172/JCI200524531.
[8] Y. Zhu and C. Zhang, "Prevalence of Gestational Diabetes and Risk of Progression to Type 2 Diabetes: a Global Perspective," *Curr. Diab. Rep.*, vol. 16, no. 1, p. 7, Jan. 2016, doi: 10.1007/s11892-015-0699-x.
[9] R. E. Ratner *et al.*, "Prevention of Diabetes in Women with a History of Gestational Diabetes: Effects of Metformin and Lifestyle Interventions," *J. Clin. Endocrinol. Metab.*, vol. 93, no. 12, pp. 4774–4779, Dec. 2008, doi: 10.1210/jc.2008-0772.
[10] V. R. Aroda *et al.*, "The Effect of Lifestyle Intervention and Metformin on Preventing or Delaying Diabetes Among Women With and Without Gestational Diabetes: The Diabetes Prevention Program Outcomes Study 10-Year Follow-Up," *J. Clin. Endocrinol. Metab.*, vol. 100, no. 4, pp. 1646–1653, Apr. 2015, doi: 10.1210/jc.2014-3761.
[11] A. McGovern *et al.*, "Diabetes screening after gestational diabetes in England: a quantitative retrospective cohort study," *Br. J. Gen. Pract.*, vol. 64, no. 618, pp. e17–e23, Jan. 2014, doi: 10.3399/bjgp14X676410.
[12] A. Tovar, L. Chasan-Taber, E. Eggleston, and E. Oken, "Postpartum screening for diabetes among women with a history of gestational diabetes mellitus.," *Prev. Chronic Dis.*, vol. 8, no. 6, p. A124, Nov. 2011, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22005617
[13] K. K. Nielsen, A. Kapur, P. Damm, M. de Courten, and I. C. Bygbjerg, "From screening to postpartum follow-up – the determinants and barriers for gestational diabetes mellitus (GDM) services, a systematic review," *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 41, Dec. 2014, doi: 10.1186/1471-2393-14-41.
[14] Z. Obermeyer and E. J. Emanuel, "Predicting the Future — Big Data, Machine Learning, and Clinical Medicine," *N. Engl. J. Med.*, vol. 375, no. 13, pp. 1216–1219, Sep. 2016, doi: 10.1056/NEJMp1606181.
[15] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digit. Med.*, vol. 1, no. 1, p. 18, May 2018, doi: 10.1038/s41746-018-0029-1.
[16] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *arXiv*, Nov. 2017. [Online]. Available: http://arxiv.org/abs/1705.07874
[17] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.
[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, no. February 2017, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
[19] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.
[20] A. Rais and A. Viana, "Operations Research in Healthcare: a survey," *Int. Trans. Oper. Res.*, vol. 18, no. 1, pp. 1–31, Jan. 2011, doi: 10.1111/j.1475-3995.2010.00767.x.
[21] P. J. H. Hulshof, N. Kortbeek, R. J. Boucherie, E. W. Hans, and P. J. M. Bakker, "Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS," *Heal. Syst.*, vol. 1, no. 2, pp. 129–175, Dec. 2012, doi: 10.1057/hs.2012.18.
[22] S. R. Earnshaw, K. Hicks, A. Richter, and A. Honeycutt, "A linear programming model for allocating HIV prevention funds with state agencies: a pilot study," *Health Care Manag. Sci.*, vol. 10, no. 3, pp. 239–252, Jun. 2007, doi: 10.1007/s10729-007-9017-8.
[23] D. Bertsimas, N. Kallus, A. M. Weinstein, and Y. D. Zhuo, "Personalized Diabetes Management Using Electronic Medical Records," *Diabetes Care*, vol. 40, no. 2, pp. 210–217, Feb. 2017, doi: 10.2337/dc16-0826.
[24] U. Aickelin and K. A. Dowsland, "An indirect Genetic Algorithm for a nurse-scheduling problem," *Comput. Oper. Res.*, vol. 31, no. 5, pp. 761–778, Apr. 2004, doi: 10.1016/S0305-0548(03)00034-0.
[25] A. Lameesa, M. Hoque, M. S. Bin Alam, S. F. Ahmed, and A. H. Gandomi, "Role of metaheuristic algorithms in healthcare: a comprehensive investigation across clinical diagnosis, medical imaging, operations management, and public health," *J. Comput. Des. Eng.*, vol. 11, no. 3, pp. 223–247, May 2024, doi: 10.1093/jcde/qwae046.
[26] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscip. Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004, doi: 10.1007/s00158-003-0368-6.

[27] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. W. Irianandа, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 23–38, May 2024, doi: 10.62411/faith.2024-11.

[28] A. O. Eboka *et al.*, "Resolving Data Imbalance Using a Bi-Directional Long-Short Term Memory for Enhanced Diabetes Mellitus Detection," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 95–109, May 2025, doi: 10.62411/faith.3048-3719-73.

[29] M. Behera, D. Mishra, A. Sarangi, and S. Sarangi, "Diabetes Prediction Using Hybrid 1D CNN and PeeHole Long Short-Term Memory Algorithm," in *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Feb. 2025, pp. 232–237. doi: 10.1109/ESIC64052.2025.10962760.

[30] F. Mustofa, A. N. Safriandono, A. R. Muslikh, and D. R. I. M. Setiadi, "Dataset and Feature Analysis for Diabetes Mellitus Classification using Random Forest," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 41–48, Jan. 2023, doi: 10.33633/jcta.v1i1.9190.

[31] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using shapley additive explanation and application for real data in hospital," *Comput. Methods Programs Biomed.*, vol. 214, p. 106584, Feb. 2022, doi: 10.1016/j.cmpb.2021.106584.

[32] O. A. Abioye and M. E. Irhebhude, "Big Data-Driven Health Risk Stratification: A Health Index-Based Approach Using Feature Importance and PySpark," *J. Comput. Theor. Appl.*, vol. 2, no. 4, pp. 456–469, Mar. 2025, doi: 10.62411/jcta.12327.

[33] J. Lindström and J. Tuomilehto, "The Diabetes Risk Score," *Diabetes Care*, vol. 26, no. 3, pp. 725–731, Mar. 2003, doi: 10.2337/diacare.26.3.725.

[34] P. W. F. Wilson, "Prediction of Incident Diabetes Mellitus in Middle-aged Adults," *Arch. Intern. Med.*, vol. 167, no. 10, p. 1068, May 2007, doi: 10.1001/archinte.167.10.1068.

[35] J. Hippisley-Cox, C. Coupland, J. Robson, A. Sheikh, and P. Brindle, "Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore," *BMJ*, vol. 338, no. mar17 2, pp. b880–b880, Mar. 2009, doi: 10.1136/bmj.b880.

[36] D. Noble, R. Mathur, T. Dent, C. Meads, and T. Greenhalgh, "Risk models and scores for type 2 diabetes: systematic review," *BMJ*, vol. 343, no. nov28 1, pp. d7163–d7163, Nov. 2011, doi: 10.1136/bmj.d7163.

[37] S. H. Kwak *et al.*, "Prediction of type 2 diabetes in women with a history of gestational diabetes using a genetic risk score," *Diabetologia*, vol. 56, no. 12, pp. 2556–2563, Dec. 2013, doi: 10.1007/s00125-013-3059-x.

[38] M. Köhler, A. G. Ziegler, and A. Beyerlein, "Development of a simple tool to predict the risk of postpartum diabetes in women with gestational diabetes mellitus," *Acta Diabetol.*, vol. 53, no. 3, pp. 433–437, Jun. 2016, doi: 10.1007/s00592-015-0814-0.

[39] W. Li *et al.*, "Nomograms for incident risk of post-partum type 2 diabetes in Chinese women with prior gestational diabetes mellitus," *Clin. Endocrinol. (Oxf).*, vol. 90, no. 3, pp. 417–424, Mar. 2019, doi: 10.1111/cen.13863.

[40] B. Man, A. Schwartz, O. Pugach, Y. Xia, and B. Gerber, "A clinical diabetes risk prediction model for prediabetic women with prior gestational diabetes," *PLoS One*, vol. 16, no. 6, p. e0252501, Jun. 2021, doi: 10.1371/journal.pone.0252501.

[41] L.-W. Chen *et al.*, "Combined analysis of gestational diabetes and maternal weight status from pre-pregnancy through post-delivery in future development of type 2 diabetes," *Sci. Rep.*, vol. 11, no. 1, p. 5021, Mar. 2021, doi: 10.1038/s41598-021-82789-x.

[42] Y. Belsti *et al.*, "Development of a risk prediction model for postpartum onset of type 2 diabetes mellitus, following gestational diabetes; the lifestyle InterVention in gestational diabetes (LIVING) study," *Clin. Nutr.*, vol. 43, no. 8, pp. 1728–1735, Aug. 2024, doi: 10.1016/j.clnu.2024.06.006.

[43] A. Allalou *et al.*, "A Predictive Metabolic Signature for the Transition From Gestational Diabetes Mellitus to Type 2 Diabetes," *Diabetes*, vol. 65, no. 9, pp. 2529–2539, Sep. 2016, doi: 10.2337/db15-1720.

[44] S. R. Khan *et al.*, "The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes," *Diabetologia*, vol. 62, no. 4, pp. 687–703, Apr. 2019, doi: 10.1007/s00125-018-4800-2.

[45] H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, "Predictive models for diabetes mellitus using machine learning techniques," *BMC Endocr. Disord.*, vol. 19, no. 1, p. 101, Dec. 2019, doi: 10.1186/s12902-019-0436-6.

[46] M. V. Joglekar *et al.*, "Postpartum circulating microRNA enhances prediction of future type 2 diabetes in women with previous gestational diabetes," *Diabetologia*, vol. 64, no. 7, pp. 1516–1526, Jul. 2021, doi: 10.1007/s00125-021-05429-z.

[47] N. Razavian, S. Blecker, A. M. Schmidt, A. Smith-McLallen, S. Nigam, and D. Sontag, "Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors," *Big Data*, vol. 3, no. 4, pp. 277–287, Dec. 2015, doi: 10.1089/big.2015.0020.

[48] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, vol. 9, Nov. 2018, doi: 10.3389/fgene.2018.00515.

[49] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, "Early detection of type 2 diabetes mellitus using machine learning-based prediction models," *Sci. Rep.*, vol. 10, no. 1, p. 11981, Jul. 2020, doi: 10.1038/s41598-020-68771-z.

[50] O. Houri *et al.*, "Prediction of Type 2 Diabetes Mellitus According to Glucose Metabolism Patterns in Pregnancy Using a Novel Machine Learning Algorithm," *J. Med. Biol. Eng.*, vol. 42, no. 1, pp. 138–144, Feb. 2022, doi: 10.1007/s40846-022-00685-9.

[51] M. Kumar *et al.*, "Machine Learning–Derived Prenatal Predictive Risk Model to Guide Intervention and Prevent the Progression of Gestational Diabetes Mellitus to Type 2 Diabetes: Prediction Model Development Study," *JMIR Diabetes*, vol. 7, no. 3, p. e32366, Jul. 2022, doi: 10.2196/32366.

[52] L. Ilari *et al.*, "Unraveling the Factors Determining Development of Type 2 Diabetes in Women With a History of Gestational Diabetes Mellitus Through Machine-Learning Techniques," *Front. Physiol.*, vol. 13, Feb. 2022, doi: 10.3389/fphys.2022.789219.

[53] J. Prashanthan and A. Prashanthan, "Predicting the future risk of developing type 2 diabetes in women with a history of gestational diabetes mellitus using machine learning and explainable artificial intelligence," *Prim. Care Diabetes*, vol. 19, no. 6, pp. 658–666, Dec. 2025, doi: 10.1016/j.pcd.2025.09.006.

[54] J. C. D'Aeth *et al.*, "Optimal national prioritization policies for hospital care during the SARS-CoV-2 pandemic," *Nat. Comput. Sci.*, vol. 1, no. 8, pp. 521–531, Aug. 2021, doi: 10.1038/s43588-021-00111-1.

[55] R. M. Stuart *et al.*, "How to do (or not to do)… health resource allocations using constrained mathematical optimization," *Health Policy Plan.*, vol. 38, no. 1, pp. 122–128, Jan. 2023, doi: 10.1093/heapol/czac096.

[56] P. C. Chu and J. E. Beasley, "A Genetic Algorithm for the Multidimensional Knapsack Problem," *J. Heuristics*, vol. 4, no. 1, pp. 63–86, Jun. 1998, doi: 10.1023/A:1009642405419.

[57] C.-H. Wang, R. Tian, K. Hu, Y.-T. Chen, and T.-H. Ku, "A Markov decision optimization of medical service resources for two-class patient queues in emergency departments via particle swarm optimization algorithm," *Sci. Rep.*, vol. 15, no. 1, p. 2942, Jan. 2025, doi: 10.1038/s41598-025-86158-w.

[58] A. Ala, F. E. Alsaadi, M. Ahmadi, and S. Mirjalili, "Optimization of an appointment scheduling problem for healthcare systems based on the quality of fairness service using whale optimization algorithm and NSGA-II," *Sci. Rep.*, vol. 11, no. 1, p. 19816, Oct. 2021, doi: 10.1038/s41598-021-98851-7.

[59] T. Mizan and S. Taghipour, "Medical resource allocation planning by integrating machine learning and optimization models," *Artif. Intell. Med.*, vol. 134, p. 102430, Dec. 2022, doi: 10.1016/j.artmed.2022.102430.

[60] J. Zuo, Y. Jin, and W. Liu, "Outpatient scheduling problem in smart hospital with two-agent deep reinforcement learning algorithm," *Discov. Comput.*, vol. 27, no. 1, p. 41, Nov. 2024, doi: 10.1007/s10791-024-09474-1.

[61] Piyush Ingole, "Optimizing Resource Allocation in Hospitals Using Predictive Analytics and Information Systems," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 1s, pp. 400–415, Dec. 2024, doi: 10.52783/jisem.v10i1s.224.

[62] A. Moura and M. Pinho, "A Scheduling Optimization Approach to Reduce Outpatient Waiting Times for Specialists," *Healthcare*, vol. 13, no. 7, p. 749, Mar. 2025, doi: 10.3390/healthcare13070749.

[63] F. Ben Abdelaziz, B. Aouni, and R. El Fayedh, "Multi-objective stochastic programming for portfolio selection," *Eur. J. Oper. Res.*, vol. 177, no. 3, pp. 1811–1823, Mar. 2007, doi: 10.1016/j.ejor.2005.10.021.

[64] W. Zhang, K. Cao, S. Liu, and B. Huang, "A multi-objective optimization approach for health-care facility location-allocation problems in highly developed cities such as Hong Kong," *Comput. Environ. Urban Syst.*, vol. 59, pp. 220–230, Sep. 2016, doi: 10.1016/j.compenvurbsys.2016.07.001.

[65] V. Roshanaei, C. Luong, D. M. Aleman, and D. R. Urbach, "Collaborative Operating Room Planning and Scheduling," *INFORMS J. Comput.*, vol. 29, no. 3, pp. 558–580, Aug. 2017, doi: 10.1287/ijoc.2017.0745.

[66] S. J. Neuwahl, A. J. Sharma, P. Zhang, and T. J. Hoerger, "Postdelivery Intervention to Prevent Type 2 Diabetes and the Cost-Effectiveness of Screening Criteria for Gestational Diabetes," *Prev. Chronic Dis.*, vol. 19, p. 220055, Dec. 2022, doi: 10.5888/pcd19.220055.

[67] M. Lloyd *et al.*, "Long-term cost-effectiveness of implementing a lifestyle intervention during pregnancy to reduce the incidence of gestational diabetes and type 2 diabetes," *Diabetologia*, vol. 66, no. 7, pp. 1223–1234, Jul. 2023, doi: 10.1007/s00125-023-05897-5.

[68] D. Bertsimas and N. Kallus, "From Predictive to Prescriptive Analytics," *Manage. Sci.*, vol. 66, no. 3, pp. 1025–1044, Mar. 2020, doi: 10.1287/mnsc.2018.3253.

[69] A. N. Elmachtoub and P. Grigas, "Smart 'Predict, then Optimize,'" *Manage. Sci.*, vol. 68, no. 1, pp. 9–26, Jan. 2022, doi: 10.1287/mnsc.2020.3922.

[70] M. Kraus, S. Feuerriegel, and M. Saar-Tsechansky, "Data-Driven Allocation of Preventive Care with Application to Diabetes Mellitus Type II," *Manuf. Serv. Oper. Manag.*, vol. 26, no. 1, pp. 137–153, Jan. 2024, doi: 10.1287/msom.2021.0251.

[71] R. C. Deo, "Machine Learning in Medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.

[72] Y. Liu, B. Logan, N. Liu, Z. Xu, J. Tang, and Y. Wang, "Deep Reinforcement Learning for Dynamic Treatment Regimes on Medical Registry Data," in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, Aug. 2017, pp. 380–385. doi: 10.1109/ICHI.2017.45.

[73] G. Wang *et al.*, "Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial," *Nat. Med.*, vol. 29, no. 10, pp. 2633–2642, Oct. 2023, doi: 10.1038/s41591-023-02552-9.

[74] L. Dénes-Fazakas, L. Szilágyi, L. Kovács, A. De Gaetano, and G. Eigner, "Reinforcement Learning: A Paradigm Shift in Personalized Blood Glucose Management for Diabetes," *Biomedicines*, vol. 12, no. 9, p. 2143, Sep. 2024, doi: 10.3390/biomedicines12092143.

[75] Y. Zhou *et al.*, "Optimizing long term disease prevention with reinforcement learning: a framework for precision lipid control," *npj Digit. Med.*, vol. 8, no. 1, p. 553, Aug. 2025, doi: 10.1038/s41746-025-01951-1.

[76] H. J. Sarode, M. S. Patil, N. Patil, N. Bhagwat, S. S. Yewale, and P. Balwadkar, "Integrating AI for Dynamic Resource Allocation and Workflow Optimization in Healthcare Management Systems," *Front. Heal. Informatics*, vol. 13, no. 3, pp. 6028–6041, 2024, [Online]. Available: https://healthinformaticsjournal.com/index.php/IJMI/article/view/549

[77] A. M. Ponsiglione *et al.*, "Combining simulation models and machine learning in healthcare management: strategies and applications," *Prog. Biomed. Eng.*, vol. 6, no. 2, p. 022001, Apr. 2024, doi: 10.1088/2516-1091/ad225a.

[78] S. S. Dhanda *et al.*, "Advancement in public health through machine learning: a narrative review of opportunities and ethical considerations," *J. Big Data*, vol. 12, no. 1, p. 154, Jul. 2025, doi: 10.1186/s40537-025-01201-x.

[79] S.-Y. Lee *et al.*, "Integrating Social Determinants of Health in Machine Learning–Driven Decision Support for Diabetes Case Management: Protocol for a Sequential Mixed Methods Study," *JMIR Res. Protoc.*, vol. 13, p. e56049, Sep. 2024, doi: 10.2196/56049.

[80] J. K. Tan *et al.*, "Machine Learning–Based Prediction for High Health Care Utilizers by Using a Multi-Institutional Diabetes Registry: Model Training and Evaluation," *JMIR AI*, vol. 3, p. e58463, Oct. 2024, doi: 10.2196/58463.

[81] A. Prashanthan and J. Prashanthan, "Early Prediction of GDM," *Kaggle*, 2026. https://www.kaggle.com/dsv/14445931

[82] H. Lee, M.-B. Park, and Y.-J. Won, "AI Machine Learning–Based Diabetes Prediction in Older Adults in South Korea: Cross-Sectional Analysis," *JMIR Form. Res.*, vol. 9, pp. e57874–e57874, Jan. 2025, doi: 10.2196/57874.

[83] K. Arora, P. P, H. Yadav, K. Kavitha, J, B. R. Mishra, and K. S. Kumar, "Big Data and Machine Learning for Healthcare Resource Allocation And Optimization," *South East. Eur. J. Public Heal.*, pp. 3998–4005, Jan. 2025, doi: 10.70135/seejph.vi.3821.

[84] T. C. Frommeyer *et al.*, "Reinforcement Learning and Its Clinical Applications Within Healthcare: A Systematic Review of Precision Medicine and Dynamic Treatment Regimes," *Healthcare*, vol. 13, no. 14, p. 1752, Jul. 2025, doi: 10.3390/healthcare13141752.

[85] Q. Wu, J. Han, Y. Yan, Y.-H. Kuo, and Z.-J. M. Shen, "Reinforcement learning for healthcare operations management: methodological framework, recent developments, and future research directions," *Health Care Manag. Sci.*, vol. 28, no. 2, pp. 298–333, Jun. 2025, doi: 10.1007/s10729-025-09699-6.

[86] Z. L. Teo *et al.*, "Generative artificial intelligence in medicine," *Nat. Med.*, vol. 31, no. 10, pp. 3270–3282, Oct. 2025, doi: 10.1038/s41591-025-03983-2.

[87] M. M. Li *et al.*, "Scaling medical AI across clinical contexts," *Nat. Med.*, vol. 32, no. 2, pp. 439–448, Feb. 2026, doi: 10.1038/s41591-025-04184-7.