



An enhanced adaptive differential evolution for early diabetes prediction



Lin Chen^a, Jinzhou Cao^{b,*}, Guoqiang Wu^a, Yuanqi Li^a

^a College of Management, Shenzhen University, Shenzhen, China

^b School of Artificial Intelligence, Shenzhen Technology University, Shenzhen, China

ARTICLE INFO

Keywords:

Diabetes prediction
Machine learning
Differential evolution
Feature selection
Hyperparameter optimization

ABSTRACT

Diabetes requires timely detection to mitigate risks and manage disease progression. While predictive modeling has advanced, comprehensive frameworks integrating feature selection and hyperparameter optimization are still needed. This study proposes an enhanced hybrid optimization framework that combines machine learning with an AHDE algorithm to enhance diabetes prediction performance. Experiments on three public diabetes datasets demonstrate that our AHDE-optimized algorithm outperforms the comparative models, achieving recall rates of 0.9226, 0.9894, and 0.9989, with ROC AUC values of 0.9276, 0.9972, and 0.9945, respectively. These findings offer insights for improving early diabetes screening, highlighting the potential applications of the proposed framework in clinical decision support systems.

1. Introduction

Diabetes is a metabolic disorder, characterized by elevated blood sugar levels, which result from insufficient insulin production or ineffective insulin utilization in the body [1]. According to the International Diabetes Federation (IDF) Diabetes Atlas 11th Edition, approximately 589 million adults aged 20–79 years were living with diabetes globally in 2024, representing 11.1% of the world's adult population, with projections indicating an increase to 853 million cases by 2050. Global healthcare expenditure attributable to diabetes reached at least 1 trillion USD in 2024, marking a 338% increase compared to 17 years prior. Most concerning is that an estimated 252 million individuals, nearly 43% of all cases, remain undiagnosed, highlighting a critical need for improved early detection. This study specifically focuses on early prediction of Type 2 diabetes, which accounts for 90–95% of all cases and is strongly associated with modifiable risk factors including obesity, physical inactivity, and dietary patterns. Here, early means detecting undiagnosed cases before the onset of irreversible complications. Unlike Type 1 diabetes with its primarily autoimmune origin, Type 2 diabetes is particularly amenable to prevention through early intervention. Early identification enables timely lifestyle modifications and clinical management to prevent progression to severe complications including renal failure and diabetic retinopathy.

Machine learning methods have been extensively applied to disease diagnosis, encompassing traditional models such as support vector machines (SVMs), logistic regression and K-nearest neighbors [2–5],

ensemble learning approaches including random forests and gradient boosting [6–13], and more recently to neural network architectures including multilayer perceptrons (MLPs), convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformers [14–20]. While deep learning has demonstrated impressive capabilities in representation learning, its demands for large-scale datasets and limited interpretability make it less suitable for clinical settings, where optimized traditional ML models have shown comparable performance on small-to-moderate tabular data [21].

However, the effectiveness of these models critically depends on two optimization tasks: feature selection and hyperparameter optimization [22]. These two tasks present a fundamental challenge that has been largely overlooked in existing research: they are inherently interdependent rather than independent. The optimal feature subset depends on the model's hyperparameter configuration, while the optimal hyperparameters are determined by which features are selected [23]. For instance, in support vector machines, the RBF kernel parameter γ is intrinsically related to feature scales and distributions—different feature subsets may require entirely different γ values to achieve optimal decision boundaries [22]. Similarly, in random forests, the optimal values of hyperparameters such as tree depth and splitting criteria strongly depend on the number and nature of available features [23]. This bidirectional dependency implies that optimizing either component in isolation or in a fixed sequential order inevitably leads to suboptimal solutions.

Early studies commonly treated feature selection and

* Corresponding author.

E-mail address: caojinzhou@sztu.edu.cn (J. Cao).

hyperparameter optimization as a sequential two-stage process: first selecting a feature subset, then optimizing hyperparameters on the fixed feature set. Feature selection methods can be categorized into traditional approaches (filter-based [24,25], wrapper-based [26–28], and embedded-based [29,30] and emerging deep learning-based approaches [31,32]. Following feature subset determination, hyperparameter optimization employs various techniques including grid search, random search, Bayesian optimization, evolutionary algorithms, and other advanced methods to configure parameters that critically influence model behavior [33–37]. Despite differences in feature evaluation mechanisms, these methods share a common assumption: an optimal feature subset can be determined first and then used as fixed input for hyperparameter optimization. However, for medical diagnosis applications, this assumption is fundamentally flawed. The diagnostic value of clinical indicators is highly dependent on model configuration—biomarkers that appear irrelevant under one hyperparameter configuration may become critical under another [38]. More critically, the feature subset selected in the first stage is determined without knowledge of the optimal hyperparameter configuration, while the second stage cannot retrospectively correct the first-stage decisions even when certain hyperparameter configurations require different feature combinations.

Recent research has begun exploring collaborative optimization methods that jointly optimize feature selection and hyperparameter optimization within a unified framework. Khalil et al. employed NSGA-II to simultaneously optimize feature selection and SVM [39], achieving 12–15% accuracy improvement over sequential approaches, though with 20–40% increased computational cost [22]. However, when extending to models with multiple hyperparameters, the search space grows exponentially, rendering population-based genetic algorithms computationally intractable [40,41]. Martinez-de-Pison et al. developed a hybrid genetic algorithm for simultaneous optimization but adopted a “continuation” strategy treating all hyperparameters as continuous variables with rounding for discrete parameters [42]. This approach violates the inherent structure of discrete variables, as standard genetic operators cause offspring to collapse to identical values, substantially reducing diversity and leading to premature convergence [43–45]. Konak et al. observed asymmetric convergence behavior: binary feature spaces converged rapidly due to their finite solution space, while continuous hyperparameter spaces required more extensive exploration [46]. Standard diversity maintenance mechanisms designed for homogeneous continuous spaces cannot independently regulate exploration intensity in heterogeneous subspaces, causing algorithms to remain trapped in local optima once the discrete subspace prematurely converges.

In summary, simultaneous feature-hyperparameter optimization currently faces three interrelated technical challenges: scalability in mixed-variable spaces, effective representation of heterogeneous variables, and diversity preservation under asymmetric convergence. To address these challenges, DE demonstrates distinctive technical advantages among metaheuristic algorithms. DE’s mutation and crossover operators are inherently efficient for continuous variables, while its population-based search mechanism provides intrinsic parallelism for exploring complex optimization landscapes [47]. More importantly, DE possesses a simple algorithmic structure with minimal control parameters, making it more readily extensible than genetic algorithms [48] or particle swarm optimization [49] for handling mixed-variable problems [47]. Despite these advantages, standard DE and its existing variants, when applied to the specific scenario of joint feature-hyperparameter optimization in medical diagnosis, have not systematically resolved the aforementioned three core technical challenges.

Accordingly, this paper proposes an Adaptive Heterogeneous Differential Evolution (AHDE) algorithm designed for simultaneous feature selection and hyperparameter optimization. Unlike existing methods that apply uniform strategies across variable types, AHDE deliberately integrates three synergistic innovations: (1) Heterogeneous encoding

with type-specific operations: employing binary tournament selection for discrete feature decisions and adaptive differential mutation for continuous/discrete hyperparameters, thereby eliminating the need for artificial continuation while preserving the intrinsic structure of each variable type; (2) Entropy-based adaptive diversity control: utilizing Shannon entropy to separately quantify population diversity in discrete and continuous dimensions, dynamically adjusting mutation intensity to balance exploration and exploitation across heterogeneous variable types; (3) Convergence-aware staged strategy switching: monitoring fitness improvement rates and entropy-based diversity to adaptively transition between exploration-oriented and exploitation-oriented strategies, ensuring efficient optimization progress in high-dimensional mixed spaces. The proposed AHDE is evaluated on multiple diabetes datasets and comprehensively compared with various existing methods. Experimental results demonstrate that our method achieves considerable improvements in classification performance, feature selection, and convergence efficiency.

The remainder of this paper is structured as follows: Section 2 describes the datasets, data preprocessing methods, and the proposed AHDE methodology. Section 3 presents reports the experimental results and performance comparisons. Section 4 discusses the findings, clinical implications, limitations, and future research directions. Section 5 concludes the paper.

2. Materials and methods

Our proposed machine learning optimization framework for early diabetes prediction consists of several main parts: data preparation, machine learning model optimization with AHDE algorithm, and model testing and evaluation. Initially, we carried out data preparation to enhance data quality. This process involved exploratory data analysis, removing outliers, imputing missing values, and standardizing the features. Next, we utilized proposed AHDE algorithm to optimize the hyperparameter configuration of the machine learning model while simultaneously selecting the feature subset. Finally, we trained the machine learning model using the optimized hyperparameters and feature subsets. The model’s performance was evaluated using metrics such as accuracy, precision, recall, F1 score and ROC AUC. The flowchart of the proposed framework is illustrated in Fig. 1.

2.1. Data description

To evaluate early diabetes prediction, we selected three public datasets with distinct characteristics. The Pima Indians Diabetes Dataset (PIDD) [50] serves as a well-established benchmark dataset, focusing on a specific population with basic metabolic indicators. The Laboratory of Medical City Hospital Dataset (LMCH) [51] offers extensive clinical measurements including detailed blood tests and lipid profiles. The Early Stage Diabetes Risk Prediction Dataset (ESDRPD) [52] provides questionnaire-based symptoms and behavioral features. These datasets collectively enable evaluation across different types of medical data (clinical measurements, laboratory tests, and symptom records) and population groups. The detailed descriptions of these datasets are as follows:

- (i) Dataset 1: The PIDD includes medical records from 768 Pima Indian women near Phoenix, Arizona. It contains 268 diabetic and 500 non-diabetic cases. Each patient is described by attributes like pregnancies (Preg), glucose, blood pressure (BP), skin thickness (ST), insulin, body mass index (BMI), diabetes pedigree function (DPF), and age. Further elaboration on the dataset attributes can be found in Table 1
- (ii) Dataset 2: The LMCH consists of 1000 patients, including 103 normal, 53 prediabetic, and 844 diabetic cases. Prediabetic cases were classified as positive samples since prediabetes is defined by fasting glucose of 5.6–6.9 mmol/L or HbA1c of 5.7–6.4%,

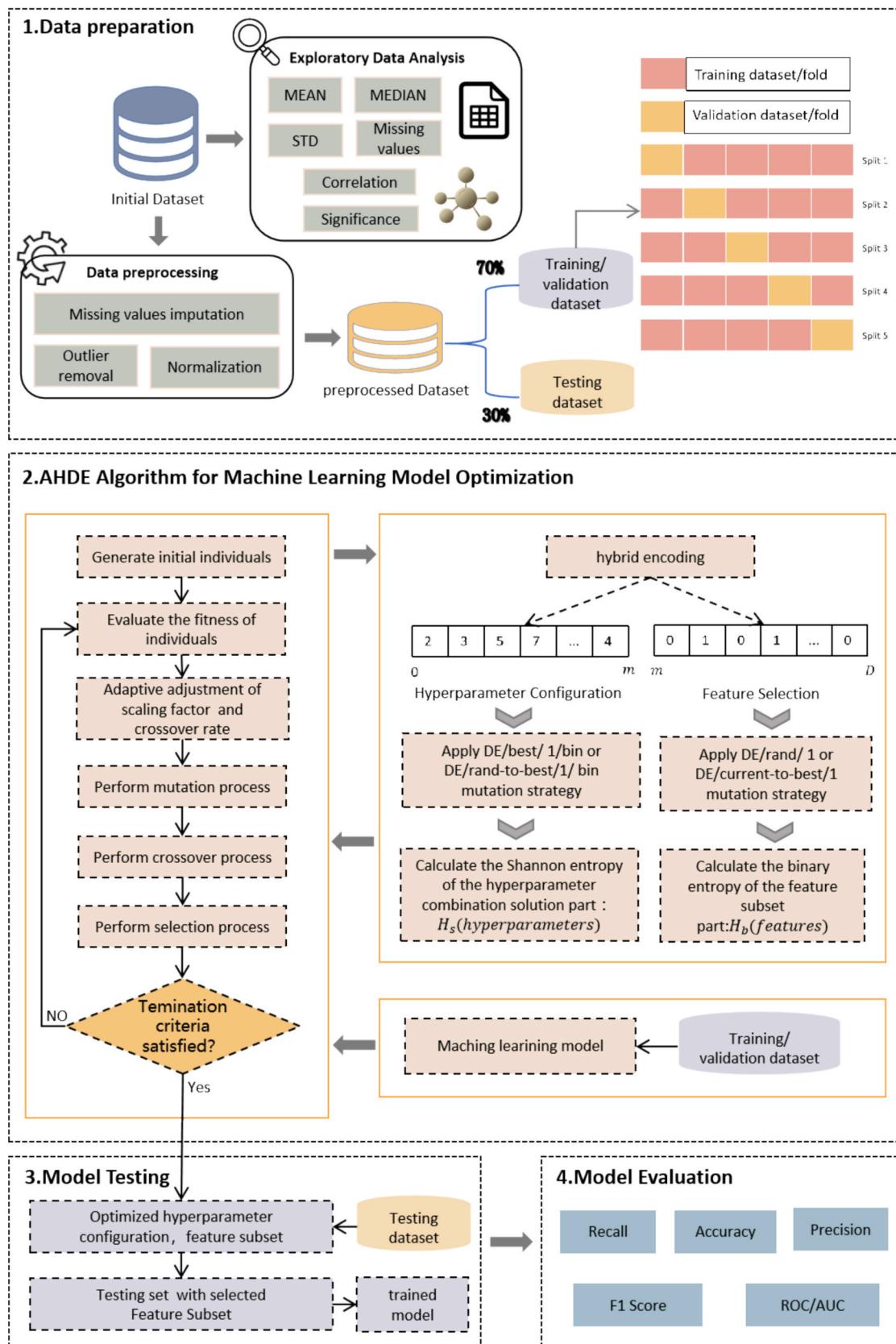


Fig. 1. The proposed machine learning optimization framework for diabetes prediction.

Table 1

Description and raw range of attributes used for Dataset 1, Dataset 2, and Dataset 3.

Dataset 1(PIDD)		
Feature	Description	Range
Preg	Number of times pregnant	(0–17)
Glucose	Plasma glucose concentration(2 h oral test) (mg/dL)	(44.0–199.0)
BP	Diastolic blood pressure(mm Hg)	(24.0–122.0)
ST	Triceps skin fold thickness(mm)	(7.0–63.0)
Insulin	2-Hour serum insulin(μU/ml)	(14.0–846.0)
BMI	Body mass index(weight in kg)/(height in m) ²	(18.2–67.1)
DPF	a function which scores likelihood of diabetes based on family history	(0.08–2.42)
Age	Age (years)	(21–81)
Class	Class variable(0 if non-diabetic,1 if diabetic)	(0–1)
Dataset 2 (LMCH)		
ID	Patient identification number	–
AGE	Age of the patient (years)	–
Urea	Blood urea level (mg/dL) – measures kidney function	–
Cr	Creatinine ratio – indicator of kidney function (mg/dL)	–
HbA1c	Glycated hemoglobin – indicates average blood sugar level over past 2–3 months (%)	–
Chol	Total cholesterol level (mg/dL)	–
TG	Triglycerides level (mg/dL) – type of blood fat	–
HDL	High-density lipoprotein cholesterol (mg/dL) – “good” cholesterol	–
LDL	Low-density lipoprotein cholesterol (mg/dL) – “bad” cholesterol	–
VLDL	Very low-density lipoprotein cholesterol (mg/dL)	–
BMI	Body Mass Index – weight(kg)/(height(m)) ²	–
Dataset 3(ESDRPD)		
Age	Age (years)	(16–90)
Gender	0 if Female,1 if Male	(0–1)
Polyuria	0 if no polyuria,1 if yes	(0–1)
Polydipsia	0 if no polydipsia,1 if yes	(0–1)
Sudden weight loss	0 if no sudden weight loss,1 if yes	(0–1)
Weakness	0 if no weakness,1 if yes	(0–1)
Polyphagia	0 if no polyphagia,1 if yes	(0–1)
Genital thrush	0 if no genital thrush,1 if yes	(0–1)
Visual blurring	0 if no visual blurring,1 if yes	(0–1)
Itching	0 if no itchiness,1 if yes	(0–1)
Irritability	0 if no irritability,1 if yes	(0–1)
Delayed healing	0 if no delayed healing,1 if yes	(0–1)
Partial paresis	0 if no partial paresis,1 if yes	(0–1)
Muscle stiffness	0 if no muscle stiffness,1 if yes	(0–1)
Alopecia	0 if no alopecia,1 if yes	(0–1)
Obesity	0 if not obese,1 if yes	(0–1)
Class	Class variable (0 if non-diabetic,1 if diabetic)	(0–1)

representing impaired glucose metabolism requiring clinical intervention. The dataset records demographic information, anthropometric measurements, and blood test results for each patient. Blood tests include glucose, creatinine ratio (Cr), urea, total cholesterol, Low-Density lipoprotein (LDL), Very Low-Density lipoprotein (VLDL), triglycerides, High-Density lipoprotein (HDL), and Glycated hemoglobin (HbA1c). Detailed attribute definitions are provided in Table 1

(iii) Dataset 3: The ESDRPD is from Sylhet diabetes Hospital, collected via a questionnaire for early diabetes risk prediction. It includes 520 cases (320 diabetic) with 16 predictive attributes and 1 class attribute, as detailed in Table 1

2.2. Data preprocessing

2.2.1. Exploratory data analysis and outlier detection

In this study, we conducted exploratory data analysis (EDA) on Dataset 1 (PIDD) and Dataset 2 (LMCH), with the results summarized in Table 2. We employed the interquartile range (IQR) method to identify and remove outliers. Since all features in Dataset 3 are binary (yes/no) except for AGE, no exploratory analysis results in a tabular format were produced.

2.2.2. Missing value analysis and imputation

During the Exploratory Data Analysis (EDA), we identified a significant number of missing values: 111 for pregnancies, 374 for insulin, and 227 for skin thickness. Notably, only “number of pregnancies” can reasonably be zero, while “skin thickness” and “insulin” should typically be non-zero under normal physiological conditions. Although insulin ($r = 0.13$) and skin thickness ($r = 0.075$) showed relatively lower correlation coefficients compared to glucose ($r = 0.47$) and BMI ($r = 0.29$) in the correlation heatmap (Fig. 2), both features demonstrated statistically significant associations with diabetes outcome ($p < 0.0001$, Table 3). Moreover, insulin plays a crucial pathophysiological role in diabetes development, and skin thickness reflects adiposity-related metabolic changes. Therefore, we retained these features despite their missing values.

To impute missing values, we categorized data into three classes based on clinical diabetes criteria: glucose levels below 5.6 mmol/L (healthy), 5.6–6.9 mmol/L (prediabetes), and above 6.9 mmol/L (diabetes). We used class mean values to fill missing skin thickness and insulin data. This approach considers physiological variations, providing a personalized and realistic treatment of missing data, reducing imputation bias.

To validate our methodological choice, we compared six imputation strategies across four classifiers: SVM, KNN, XGB, and RF. The strategies evaluated were: (1) Origin (zero-imputation for missing values), (2) Mean imputation, (3) Median imputation, (4) KNN imputation, (5) Logistic Regression (LR) imputation, and (6) Class mean imputation (our proposed glucose-stratified approach).

Our analysis across four classifiers and different imputation methods reveals several key findings (Table 4). Performance improvements were most evident in SVM and KNN classifiers, where F1-scores increased from 0.6491 and 0.6271 (Origin method) to 0.7616 and 0.7843 (Class mean method), representing improvements of 17.3% and 25.1%, respectively. The Class mean method demonstrated consistent superior performance across all metrics, achieving the highest accuracy of 0.8220 with KNN while maintaining accuracy above 0.79 across other classifiers. These results validate the effectiveness of our glucose-stratified class-mean imputation approach.

2.2.3. Feature standardization

Finally, we applied z-score standardization to all features to prepare for modeling. Each feature was transformed to zero mean and unit variance using the formula: $z = (x - \mu)/\sigma$ where x is the original value, μ is the mean, and σ is the standard deviation. This standardization eliminates scale differences across features, ensuring that features with different units and ranges (e.g., glucose in mmol/L vs. age in years) contribute equitably to model learning, which is essential for distance-based algorithms and gradient descent optimization.

2.3. AHDE algorithm for machine learning model optimization

In this section, we introduce an innovative hybrid model, AHDE-ML, for early diabetes prediction using machine learning and the AHDE algorithm. This approach simultaneously optimizes feature selection and hyperparameters, incorporating strategies to enhance the Differential Evolution (DE) algorithm's performance.

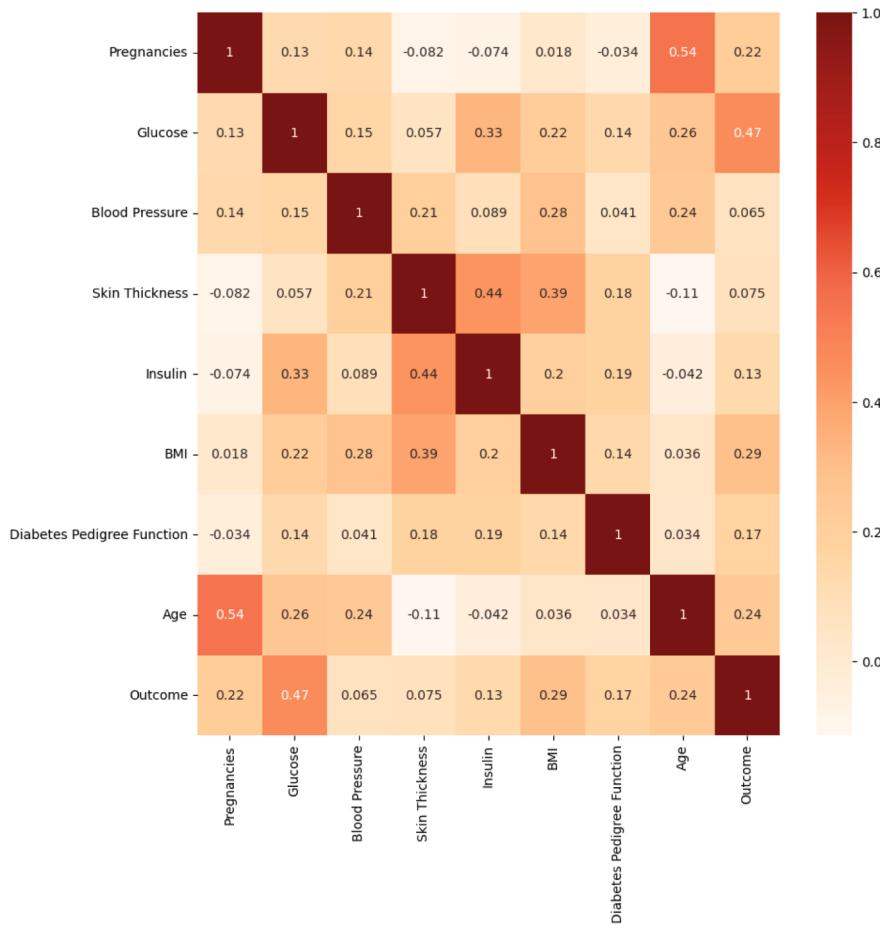
A critical design decision underlying AHDE is the choice of

Table 2

Statistical summary of dataset 1 (PIDD) and dataset 2 (LMCH).

Dataset 1(PIDD)									
Feature	Preg	Glucose	BP	ST	Insulin	BMI	DPM	Age	
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.08	21.00	
1st Qu	1.00	99.00	62.00	0.00	0.00	7.30	0.24	24.00	
Median	3.00	117.00	72.00	23.00	30.50	32.00	0.37	29.00	
Mean	3.85	120.89	69.11	20.54	79.80	31.99	0.47	33.24	
3rd Qu	6.00	140.25	80.00	32.00	127.25	36.60	0.63	41.00	
Max	17.00	199.00	122.00	99.00	846.00	67.10	2.42	81.00	
Zeros	111.00	5.00	35.00	227.00	374.00	11.00	0.00	0.00	

Dataset 2(LMCH)											
Feature	ID	AGE	Urea	Cr	HbA1c	Chol	TG	HDL	LDL	VLDL	BMI
Min	1.00	20.00	0.50	6.00	0.90	0.00	0.30	0.20	0.30	0.10	19.00
1st Qu	125.75	51.00	3.70	48.00	6.50	4.00	1.50	0.90	1.80	0.70	26.00
Median	300.50	55.00	4.60	60.00	8.00	4.80	2.00	1.10	2.50	0.90	30.00
Mean	340.50	53.53	5.12	68.94	8.28	4.86	2.35	1.20	2.61	1.85	29.58
3rd Qu	550.25	59.00	5.70	73.00	10.20	5.60	2.90	1.30	3.30	1.50	33.00
Max	800.00	79.00	38.90	800.00	16.00	10.30	13.80	9.90	9.90	35.00	47.75

**Fig. 2.** Correlation Matrix Heatmap of Features in Dataset 1 (PIDD).

optimization objective, which must reflect the clinical realities of medical diagnosis. In medical diagnostic applications, classification errors exhibit asymmetric clinical costs. Missed diagnoses in diabetes prediction can lead to delayed intervention and disease progression to irreversible complications, significantly increasing both mortality and healthcare costs [53]. Conversely, false positives primarily result in additional confirmatory tests with minimal clinical harm [54]. This cost asymmetry is well-established in medical decision theory [55], with

misclassification cost ratios estimated at 5:1 to 10:1 in diabetes screening [56]. Following clinical decision support guidelines [57], we prioritize recall in our fitness function to minimize life-threatening missed diagnoses while maintaining acceptable precision. To optimize this objective across the heterogeneous search space of discrete feature subsets and continuous hyperparameters, AHDE integrates three core mechanisms: a heterogeneous encoding for mixed-type variables, an entropy-based adaptive strategy for dynamic diversity control, and a

Table 3

Statistical Significance of Scaled P-Values for Predictor and Outcome Variables in Dataset 1 (PIDD).

Feature	Preg	Glucose	BP	ST	Insulin	BMI	DPM	Age	Outcome
Preg	0	0.0218	>0.1	>0.1	>0.1	>0.1	>0.1	<0.0001	0.0113
Glucose	0.0218	0	0.0004	0.0029	<0.0001	0.012	>0.1	<0.0001	<0.0001
BP	>0.1	0.0004	0	0.0001	>0.1	<0.0001	>0.1	<0.0001	0.012
ST	>0.1	0.0029	0.0001	0	0.0003	<0.0001	>0.1	0.0003	<0.0001
Insulin	>0.1	<0.0001	>0.1	0.0003	0	<0.0001	>0.1	<0.0001	<0.0001
BMI	>0.1	0.012	<0.0001	<0.0001	<0.0001	0	>0.1	>0.1	<0.0001
DPM	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1	0	>0.1	0.012
Age	<0.0001	<0.0001	<0.0001	0.0003	<0.0001	>0.1	>0.1	0	<0.0001
Outcome	0.0113	<0.0001	0.012	<0.0001	<0.0001	<0.0001	0.012	<0.0001	0

Table 4Performance Comparison of Different Classifiers and Imputation Methods [^{*} denotes statistical significance ($p < 0.05$) compared to Class mean; Bold: best performance].

Classifier	Method	Accuracy	Precision	Recall	F1-Score
SVM	Origin	0.7403*	0.6167*	0.6852*	0.6491*
	Mean	0.7488	0.7616	0.7552	0.7518
	Median	0.7465*	0.7392*	0.7483*	0.7437*
	KNN	0.7528	0.7463	0.7551	0.7506
	LR	0.7455*	0.7385*	0.7478*	0.7431*
	Class mean	0.7588	0.7585	0.7652	0.7616
KNN	Origin	0.7143*	0.5781*	0.6852*	0.6271*
	Mean	0.812	0.7602	0.7410	0.7843
	Median	0.8063*	0.7550*	0.7325*	0.7435*
	KNN	0.8045*	0.7534	0.7765	0.7412*
	LR	0.8052*	0.7542	0.7310*	0.7424*
	Class mean	0.8220	0.7702	0.7510	0.7843
XGB	Origin	0.7974**	0.7890*	0.7652*	0.7616*
	Mean	0.8018	0.7890	0.7890	0.7839
	Median	0.8018	0.7825	0.7748	0.7786*
	KNN	0.7965*	0.7892	0.7815	0.7853
	LR	0.7878*	0.7808*	0.7732*	0.7770*
	Class mean	0.8018	0.7990	0.7890	0.7951
RF	Origin	0.7857	0.6909*	0.7037*	0.6972*
	Mean	0.7874	0.7886	0.7958	0.8013
	Median	0.7850*	0.8058	0.7889*	0.7854*
	KNN	0.7935	0.7890	0.7956	0.7923
	LR	0.7842*	0.7785*	0.7872*	0.7818*
	Class mean	0.7974	0.7986	0.8058	0.8013

dual mutation strategy for stage-based exploration-exploitation balancing.

The three core design principles of AHDE are grounded in established theories from evolutionary computation and optimization. The heterogeneous encoding strategy addresses the fundamental challenge that feature selection (discrete binary space) and hyperparameter optimization (continuous real-valued space) exhibit fundamentally different landscape characteristics requiring distinct search operators [58]. Mixed-integer optimization theory demonstrates that unified representations for heterogeneous variables enable more efficient exploration compared to separate sub-problem decomposition [59].

The entropy-based adaptive strategy responds to a critical requirement in evolutionary algorithms: maintaining appropriate population diversity to prevent premature convergence while ensuring efficient convergence to optimal solutions [60]. For problems involving simultaneous optimization of continuous and discrete variables, different variable types require distinct diversity quantification approaches [61]. Shannon entropy provides a principled information-theoretic measure that naturally adapts to the distribution characteristics of different variable types [62], making it particularly suitable for our heterogeneous search space combining feature selection and hyperparameter optimization.

The dual-phase mutation strategy implements a time-varying exploration-exploitation schedule grounded in two theoretical frameworks: the No Free Lunch (NFL) theorem [63] and convergence theory in evolutionary algorithms [64]. The NFL theorem establishes that no single optimization strategy performs optimally across all problem instances, implying that algorithm behavior should adapt to encountered search landscape characteristics. Empirical convergence analysis demonstrates that early search phases benefit from aggressive exploration with large mutation scaling factors to locate promising basins of attraction, while later phases require intensive local exploitation with small scaling factors to refine solutions [65]. Our staged strategy operationalizes these principles through convergence-aware phase transitions monitored by fitness improvement rates and entropy-based diversity metrics [66,67]. The specific process is illustrated in Fig. 1 and Algorithm 1. Here's a detailed description of each step:

Step 1: Generate initial individuals. This involves generating random individuals, each representing a solution characterized by a combination of hyperparameter configuration and feature selection strategies. We use a heterogeneous encoding scheme and initialize parameters for the DE algorithm iterations. These include population size(N), maximum iterations(T), upper and lower bounds, initial scaling factor (F_0), and crossover probability(CR_0). In the t th iteration ($t = 1, 2, \dots, T$, where T is the maximum number of iterations), each solution is represented as X_i^t where i ranges from 1 to N , with N being the population size and D being the dimensionality of the optimization problem. The i th candidate solution is $X_i^t = [X_{i1}^t, X_{i2}^t, \dots, X_{iD}^t]$, and the randomization technique for the i th solution component is as follows:

$$X_i = \begin{cases} \text{RandomInteger from } [X_{\min}, X_{\max}], & \text{if } i \in H_{idx} \\ \text{RandomInteger from } \{0, 1\}, & \text{if } i \in F_{idx} \end{cases} \quad (1)$$

where $H_{idx} = \{1, 2, \dots, n_h\}$ is the index set of hyperparameters, with n_h being the number of hyperparameters to be optimized, and $F_{idx} = \{n_h + 1, n_h + 2, \dots, D\}$ is the index set of features, with $(D - n_h)$ being the total number of features available for selection. When $i \in H_{idx}$, the variable X_i is generated according to its predefined value range and type (e.g., integer or continuous). When $i \in F_{idx}$, the value of the variable X_i is randomly chosen from $\{0, 1\}$, where 0 indicates that the feature is not selected and 1 indicates that the feature is selected.

Step 2: Evaluate the fitness of individuals. An effective fitness function is crucial for enhancing model performance and efficiency. Given the importance of recall in early disease detection, where missing true cases can lead to delayed treatment and serious health consequences, we use recall as the primary metric for evaluation [68]. Additionally, to achieve a more concise and efficient model, we introduce a regularization penalty term in the optimization process to promote feature selection. Therefore, we construct an objective function that considers both recall and feature selection, as shown in Equation (2):

$$\text{Fitness}(x) = \alpha(1 - \text{Recall}(x)) + \beta \cdot (|S(x)| / |F|) \quad (2)$$

where $\text{Recall}(x)$ represents the average recall rate obtained through Stratified five-fold cross-validation. The parameters α and β control the relative importance of recall optimization and feature selection in the

fitness function, respectively, with the constraint $\alpha + \beta = 1$ ensuring a normalized multi-objective optimization framework. Specifically, α determines the weight assigned to recall performance, while β determines the weight assigned to feature parsimony. By adjusting α , which ranges from 0 to 1, the balance between the model's pursuit of high recall and feature parsimony can be controlled. A larger α emphasizes recall performance, making the model more sensitive to minority class detection, whereas a larger β encourages the selection of fewer features, promoting model interpretability and computational efficiency. The term $|S(x)|/|F|$ represents the ratio of selected features to total candidate features, where $|S(x)|$ denotes the number of selected features and $|F|$ is the total number of candidate features.

Step 3: Mutation. In the differential evolution algorithm, mutation is crucial for generating new individuals. The mutation strategy is typically denoted as 'DE/*/n', where '*' indicates the base vector selection method and n represents the number of difference vectors. To address the dual challenges of hyperparameter optimization and feature selection in our heterogeneous optimization problem, we propose a two-phase hybrid mutation strategy that accounts for the distinct characteristics. For hyperparameter optimization, in the early stage ($t < T/3$), we employ DE/rand/1 strategy to promote extensive exploration in complex parameter landscapes, as shown in Equation (3):

$$\text{DE/rand/1 : } Y_i^t = X_{r1}^t + F(X_{r2}^t - X_{r3}^t) \quad (3)$$

where Y_i^t represents the mutant vector of the i th solution in the t th iteration. X_{r1}^t , X_{r2}^t , X_{r3}^t are three distinct randomly selected vectors from the current population, and F is the scaling factor. As the search progresses to the later stage ($t \geq T/3$), we transition to DE/current-to-best/1:

$$\text{DE/current-to-best/1 : } Y_i^t = X_i^t + F(X_{best}^t - X_i^t) + F(X_{r1}^t - X_{r2}^t) \quad (4)$$

Where X_i^t is the current vector and X_{best}^t represents the best solution found so far. This strategy balances exploration and exploitation through current solution participation, enabling fine-tuned hyperparameters search.

For feature selection, in the early stage ($t < T/3$), we employ DE/best/1/bin to establish stable feature combinations in the search early, as shown in Equation (5):

$$\text{DE/best/1/bin : } Y_i^t = X_{best}^t + F(X_{r1}^t - X_{r2}^t) \quad (5)$$

Where X_{best}^t serves as the base vector, and the difference vector is formed by two randomly selected solutions. In the later stage ($t \geq T/3$), we switch to DE/rand-to-best/1/bin:

$$\text{DE/rand-to-best/1/bin : } Y_i^t = X_{best}^t + F(X_{r1}^t - X_{r2}^t) + F(X_{r3}^t - X_{r4}^t) \quad (6)$$

which introduces an additional difference vector ($X_{r3}^t - X_{r4}^t$) to enhance exploration while maintaining the influence of the best solution, effectively balancing between exploration of new feature combinations and exploitation of known good solutions.

Step 4: Crossover. The crossover step involves probabilistically combining the mutant vector and the target vector to generate a new trial vector, which represents the offspring solution. This process promotes population diversity and convergence. In this method, we employ uniform crossover as the recombination operator, where dimensional values are selected from the mutated individual or the current population individual based on a probability criterion. This mechanism ensures that at least one component from the mutant vector is preserved in the trial vector, maintaining effective information exchange. The crossover rate CR controls the proportion of dimensions inherited from the mutant vector, thereby balancing exploration and exploitation. The trial solution generated through uniform crossover can be mathematically defined as follows:

$$Z_{ij}^t = \begin{cases} Y_{ij}^t, & \text{if } \text{rand}_{ij}(0,1) \leq CR \text{ or } j = k \\ X_{ij}^t, & \text{Otherwise} \end{cases} \quad (7)$$

where, $\text{rand}_{ij}(0,1)$ represents a uniformly distributed random number between 0 and 1 for the j th dimension of individual i , and $k \in \{1, 2, \dots, D\}$ is a randomly selected dimension index to ensure at least one component from the mutant vector is inherited. For each dimension j of individual i , if the random number is less than or equal to the crossover rate CR , or if j equals the randomly selected index k , we inherit the value from the mutant vector Y_{ij}^t ; otherwise, we retain the value from the current target vector X_{ij}^t . Through this structured crossover mechanism, the algorithm effectively maintains population diversity, ensures consistent evolution through guaranteed mutation, preserves beneficial traits from both parent vectors, and facilitates efficient information exchange between solutions.

Step 5: Selection. The selection process in DE is crucial for determining which solutions X_i^{t+1} (either the target solutions or the trial solutions) will survive and be included in the next search iteration. This mechanism maintains the population size for each generation. The selection process in DE is mathematically represented as follows:

$$X_i^{t+1} = \begin{cases} Z_i^t, & \text{if } f(Z_i^t) \leq f(X_i^t) \\ X_i^t, & \text{Otherwise} \end{cases} \quad (8)$$

DE selects solutions by evaluating their fitness using the objective function $f(\cdot)$. Since we aim to minimize the fitness function, If the newly generated trial vector Z_i^t yields a better fitness value than the current target vector X_i^t , it replaces the target vector in the next iteration. By preserving and propagating the best solutions to the subsequent generations, the selection process contributes to the algorithm's refinement and convergence.

Step 6: Adaptive Parameter Adjustment. Building on the entropy-based adaptive framework established above, we implement dynamic adjustment of the scaling factor F and crossover rate CR to balance exploration and exploitation throughout the optimization process. We employ Shannon entropy H_s to quantify hyperparameter diversity and binary entropy H_b for feature selection diversity. These measures are combined with weights ω_h and ω_f (where $\omega_h + \omega_f = 1$) to yield the overall diversity metric δ_i at iteration i . The adaptive adjustment mechanism is formulated as follows:

$$F = \begin{cases} F_{i-1}, f_{i_next} \leq f_i \text{ and } \rho < 0.5 \\ F_0 + k^* \rho, f_{i_next} > \delta_i \text{ and } \rho < 0.5 \\ 1 - k^* \rho, \text{ else} \end{cases} \quad (9)$$

$$CR = \begin{cases} CR_{i-1}, f_{i_next} \leq f_i \text{ and } \rho < 0.5 \\ CR_0 + k^* \rho, f_{i_next} > \delta_i \text{ and } \rho < 0.5 \\ 1 - k^* \rho, \text{ else} \end{cases} \quad (10)$$

$$\delta = \omega_h * H_s(\text{hyperparameters}) + \omega_f * H_b(\text{features}) \quad (11)$$

Here, k represents the number of iterations in the inner loop, F_0 is the initial scaling factor, and CR_0 is the initial crossover rate. When diversity δ_i is high (indicating dispersed population), the mechanism reduces F to encourage convergence while increasing CR to facilitate information exchange. Conversely, when diversity is low (indicating convergence), F increases to inject exploration while CR decreases to preserve promising solutions. The weights ω_h and ω_f control the relative influence of hyperparameter diversity versus feature selection diversity in guiding the adaptive process.

Step 7: Update Population. Update individuals based on operation outcomes. Replace the original individual with the new one if it has a better fitness value.

Step 8: Check Termination Condition. If the predefined termination conditions are met, the algorithm terminates and outputs the optimal hyperparameter configuration and feature subset. If not, return to steps 2–7 and continue the iterative optimization process.

Through these steps, the AHDE algorithm effectively optimizes the hyperparameter configuration and performs feature selection, improving the model's performance. Its adaptive mechanisms and heterogeneous encoding strategies enable it to handle various types of variables flexibly, identifying the optimal solution.

Algorithm 1:Pseudocode of AHDE-ML method

```

1: Input: Population size ( $N$ ), Maximum iterations ( $T$ ), dimension of problem ( $D$ ),
   Feature index set ( $F_{idx}$ )
2: Hyperparameter index set ( $H_{idx}$ ), Initial scaling factor ( $F_0$ ), Initial crossover
   rate ( $CR_0$ ),
3: Weights for diversity measure ( $\omega_H$ ,  $\omega_F$ ), and objective function  $f()$ 
4: Output: Optimal feature subset, Optimal hyperparameter configuration, best
   fitness value
5: Initialization
6: Initiate  $F_0$  and  $CR_0$  Initialize population  $X = \{X_1, X_2, \dots, X_n\}$  using Eq. (1)
7: Evaluate initial population using fitness function Eq. (2)
8: Initialize best solution  $X_{best}$ 
9:  $t = 0$ 
10: while  $t < T$  and not converged do
11:   Calculate population diversity  $\delta$  using Eq. (11)
12:   for  $i = 1$  to  $N$  do
13:     Mutation for hyperparameter optimization
14:     if  $i \in H_{idx}$  then
15:       If  $t < T/3$  then
16:         Generate mutant vector using DE/rand/1 Eq. (3)
17:       else
18:         Generate mutant vector using DE/current-to-best/1 Eq. (4)
19:       end if
20:     Mutation for feature selection
21:     else if  $i \in F_{idx}$  then
22:       If  $t < T/3$  then
23:         Generate mutant vector using DE/best/1/bin Eq. (5)
24:       Else
25:         Generate mutant vector using DE/rand-to-best/1/bin Eq. (6)
26:       end if
27:     end if
28:   Crossover
29:   Generate trial vector using uniform crossover Eq. (7)
30: Selection
31: Evaluate trial vector fitness
32: Apply the selection operator using Eq. (8), consider both the offspring and
   the parent's
33: Adaptive parameter update
34: Compute the adaptive parameters  $F_i$  and  $CR_i$  using Eq. (9) and (10),
   respectively.
35: End for
36: Update best solution  $x_{best}$ 
37:  $t = t + 1$ 
38: end while
39: return optimal feature subset, hyperparameter configuration, and fitness value

```

3. Experiment

This study presents a series of experiments investigating the application of four metaheuristic algorithms in the field of medical diagnosis. These metaheuristic algorithms include GA, PSO, DE, and our proposed AHDE algorithm. We apply these algorithms to five widely used machine learning models in the medical domain: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGB), and Random Forest (RF), Multilayer Perceptron(MLP) as a deep learning baseline for tabular data [69]. MLP is selected because it treats clinical features as independent variables without imposing spatial assumptions, making it particularly suitable for diabetes prediction tasks where feature ordering is arbitrary [70]. The experimental evaluation is conducted using three carefully selected benchmark datasets. The first part will discuss the experimental setup and main procedures, while the second part will compare and analyze the results.

3.1. Experimental setup

Before conducting actual optimization experiments, several important preparatory steps must be completed to ensure a fair comparison between different optimization algorithms and frameworks.

First, data partitioning. We divided the preprocessed datasets into a training set (70%) and a testing set (30%). The training set used Stratified five-fold cross-validation, incorporating the average recall and the number of selected feature subsets into the objective function for iterative optimization. Finally, we evaluated the models using the testing set.

Second, hyperparameter configuration space. When comparing different optimization methods, we ensured that they operated within the same parameter configuration space. Table 5 provides detailed information on the configuration space for each machine learning model. The optimized parameters and their ranges are referenced from the following literature: [71–75].

Third, iteration count and termination conditions. To ensure a fair comparison among optimization techniques, we set the maximum number of iterations to 50 for all models. Furthermore, based on empirical studies and multiple experimental validations, we implemented a convergence criterion where the algorithm terminates if the best fitness value remains constant for 20 consecutive iterations.

Fourth, AHDE parameter configuration. The key parameters of the AHDE framework were configured as follows: In the fitness function (Equation (2), the recall weight α was set to 0.975 and the feature selection weight β was set to 0.025. This configuration prioritizes recall performance while imposing a mild penalty to promote parsimonious feature subsets. For the differential evolution component, the population size was set to 50, with an initial scaling factor F_0 of 0.5 and an initial crossover rate CR_0 of 0.9. These parameters were maintained consistently across all three datasets to ensure a fair comparison.

Fifth, randomness control. To mitigate the effects of randomness on the experimental results, all experiments were repeated 10 times using different random seeds, and the average results were reported.

All experiments were conducted on a machine equipped with a 6-core i7-8700 processor and 16 GB of memory, using Python 3.10 for

Table 5
Configuration space for the hyperparameters of tested models.

Model	Hyperparameter	Types	Search Space
SVM	C	Continuous	[0.1,50]
	Kernel	Categorical	{'linear', 'poly', 'Rbf', 'sigmoid'}
	gamma	Continuous	[0.0001,10], valid only for 'poly', 'Rbf', 'sigmoid'
KNN0000	n_neighbors	Integer	[1,20]
	weights	Categorical	{'uniform', 'distance'}
	p	Categorical	[1,2]
XGB	n_estimators	Integer	[50,100]
	max_depth	Integer	[3,10]
	learning_rate	Continuous	[0.01,0.3]
RF	n_estimators	Integer	[10,100]
	max_depth	Integer	[5,50]
	min_samples_split	Integer	[2,20]
	min_samples_leaf	Integer	[1,20]
	max_features	Integer	[1,8]
MLP	Hidden layer 1 size	Integer	[16, 128]
	Hidden layer 2 size	Integer	[8, 64]
	Hidden layer 3 size	Integer	[4, 32]
	Learning rate	Continuous	[0.0001, 0.01]
	L2 regularization (alpha)	Continuous	[0.0001, 0.01]
	Batch size	Integer	[16, 64]

development and testing.

3.2. Results

3.2.1. Model performance evaluation

This section presents the experimental results of various optimization models for diabetes prediction. For comparison, we categorized the hybrid models into five groups to evaluate the performance of the AHDE algorithm across different machine learning models. In Table 6, the Wilcoxon test with a significance level of 0.05 is used to test whether there is a statistically significant difference between the proposed AHDE method and other algorithms. These differences are marked with an asterisk * The more *, the better the proposed AHDE method.

Table 6 displays evaluation metrics for the Dataset 1 (PIDD) across the various models. The results indicate that the AHDE algorithm achieved the best performance among the four machine learning models. In terms of recall, the rates for AHDE-SVM, AHDE-KNN, AHDE-XGB, AHDE-RF, and AHDE-MLP were 0.8570, 0.9226, 0.8988, 0.9053, and 0.8721 respectively, surpassing the second-best models by 7.77%, 4.18%, 2.73%, 2.5% and 5.30%. Notably, AHDE-KNN exhibited the highest recall rate at 0.9226, while AHDE-RF excelled in accuracy, precision, and F1 score, achieving values of 0.8481, 0.8165, and 0.8574, respectively, demonstrating its overall superior performance.

Based on their superior performance, AHDE-KNN and AHDE-RF were further evaluated on Dataset 2 (LMCH) and Dataset 3 (ESDRPD), as shown in Table 7. In Dataset 2 (LMCH), both optimization models performed well, with all metrics exceeding 0.95. AHDE-RF was stable in recall, accuracy, and F1 score, with values of 0.9894, 0.9888, and 0.9894, respectively. In Dataset 3 (ESDRPD), AHDE-KNN was notable, achieving recall, accuracy, precision, and F1 score of 0.9989, 0.9573, 0.9243, and 0.9602, respectively.

To further validate the models' performance, we employed Receiver Operating Characteristic (ROC) curves, a widely used method for evaluating models on imbalanced datasets. Curves closer to the upper left corner indicate better model performance. Fig. 3 illustrates ROC curves for the four model categories on the Dataset 1 (PIDD). Clearly, the

Table 6

Comparison of the proposed model against comparative models in terms of Recall, Accuracy, Precision, and F1 Score measures for Dataset 1 (PIDD).[* denotes statistical significance ($p < 0.05$) compared to AHDE model; Bold: best performance within classifier group; Bold underlined: best performance across all models].

Model		Rec	Acc	Prec	F1 score
SVM	SVM	0.7588*	0.7585*	0.7652*	0.7616*
	GA-SVM	0.7481*	0.7766*	0.7983	0.7722*
	PSO-SVM	0.7678*	0.7875*	0.8083	0.7841*
	DE-SVM	0.7952*	0.8000	0.8086*	0.8003*
	AHDE-SVM	0.8570	0.8024	0.7744	0.8128
KNN	KNN	0.8220*	0.7702*	0.7510*	0.7843*
	GA-KNN	0.8836*	0.8145*	0.7788	0.8276*
	PSO-KNN	0.8849*	0.8246	0.7925	0.8358
	DE-KNN	0.8856*	0.8304	0.8009	0.8408
XGB	AHDE-KNN	0.9226	0.8318	0.7832	0.8472
	XGB	0.8018*	0.7900*	0.7890*	0.7951*
	GA-XGB	0.8488*	0.8138*	0.7961*	0.8209*
	PSO-XGB	0.8607*	0.8125*	0.7845*	0.8200*
RF	DE-XGB	0.8749*	0.8166	0.7848	0.8268
	AHDE-XGB	0.8988	0.8221	0.7817	0.8358
	RF	0.7974*	0.7986*	0.8058	0.8013*
	GA-RF	0.8815*	0.8405	0.8137*	0.8481*
MLP	PSO-RF	0.8808*	0.8325*	0.8064*	0.8416 *
	DE-RF	0.8541*	0.8266*	0.8129*	0.8325*
	AHDE-RF	0.9034	0.8481	0.8165	0.8574
	MLP	0.7650*	0.7620*	0.7680	0.7665*
AHDE	GA-MLP	0.8150*	0.7851*	0.7522*	0.7829*
	PSO-MLP	0.8282*	0.8052	0.7580*	0.7921*
	DE-MLP	0.8200*	0.7753*	0.7452	0.7809*
	AHDE-MLP	0.8721	0.8050	0.7800	0.8226

Table 7

Performance Metrics Results Obtained from AHDE-KNN and AHDE-RF on Dataset 2 (LMCH) and Dataset 3 (ESDRPD). [Bold: best performance across all models].

Dataset 2 (LMCH)				
Model	Rec	Acc	Prec	F1 score
AHDE-KNN	0.9780	0.9862	0.9957	0.9868
AHDE-RF	0.9894	0.9888	0.9894	0.9894
Dataset 3 (ESDRPD)				
AHDE-KNN	0.9989	0.9573	0.9243	0.9602
AHDE-RF	0.9870	0.9240	0.8856	0.9313

proposed optimization models are nearer to the upper left corner compared to others, particularly in the KNN, XGB, and RF models, which exhibit smoother curves, indicating stable responses across thresholds. The MLP-based models also show improved ROC curves with AHDE optimization, though the curves are less smooth compared to tree-based models, reflecting the inherent complexity of neural network decision boundaries. Fig. 4 displays the ROC curves for the AHDE-KNN and AHDE-RF models on the Dataset 2 (LMCH) and Dataset 3 (ESDRPD). Both curves are significantly far from the diagonal line and close to the upper left corner. This indicates that these models outperform random guessing.

Furthermore, the area under the ROC curve (AUC) is a key evaluation metric for binary classification problems. It summarizes the performance of the ROC curve, with larger areas indicating better performance. In Dataset 1 (PIDD), the AUC values for the optimized models AHDE-KNN, AHDE-XGB, AHDE-RF and AHDE-MLP were 0.9143, 0.9043, 0.9276 and 0.8914 respectively. All these values are categorized as commendable, indicating that these models exhibit good fitting abilities and high predictive accuracy. In Dataset 2 (LMCH) and Dataset 3 (ESDRPD), the AHDE-KNN model outperformed AHDE-RF in terms of true positive rate (TPR) and AUC values. Both models demonstrated effective classification capabilities suitable for practical applications.

Fig. 5 presents the convergence curves of the optimization algorithms across different machine learning models. Each curve represents the mean fitness value at various iteration counts, allowing for an evaluation of the convergence speed and effectiveness of the algorithms. Among the four models, the AHDE algorithm showed the best convergence performance, achieving the lowest fitness value.

While DE and AHDE exhibited similar convergence characteristics in the early iterations, AHDE demonstrated better global search capabilities in the later stage. During the middle stages of evolution, AHDE's search accuracy gradually improved, leading to more favorable search results. Although the PSO algorithm converged quickly, its overall efficiency was lower, making it susceptible to local optima.

A comprehensive analysis of the convergence curves indicates that the AHDE algorithm performed well across all models, converging effectively to the lowest fitness value. This outcome supports the effectiveness of AHDE in feature selection and hyperparameter optimization, highlighting its ability to find improved solutions efficiently. In comparison, DE performed closely, while PSO and GA had relatively lower performance in this specific task.

Fig. 6 displays the selection frequency of each feature among 16 optimization models across 10 random runs. The results show that the Glucose feature had the highest selection frequency, appearing 149 times, underscoring its importance in the model. This indicates that blood glucose levels are a key indicator for early diabetes diagnosis. Following Glucose, the BMI, Age, and Insulin features were also frequently selected, appearing 100, 95, and 92 times, respectively, highlighting their roles in diabetes prediction. Other features, such as Blood Pressure (52 times) and Skin Thickness (50 times), also showed relatively high selection frequencies, though lower than those of Glucose, BMI, and Age. Combining the data from Table 3 and Fig. 3, it is

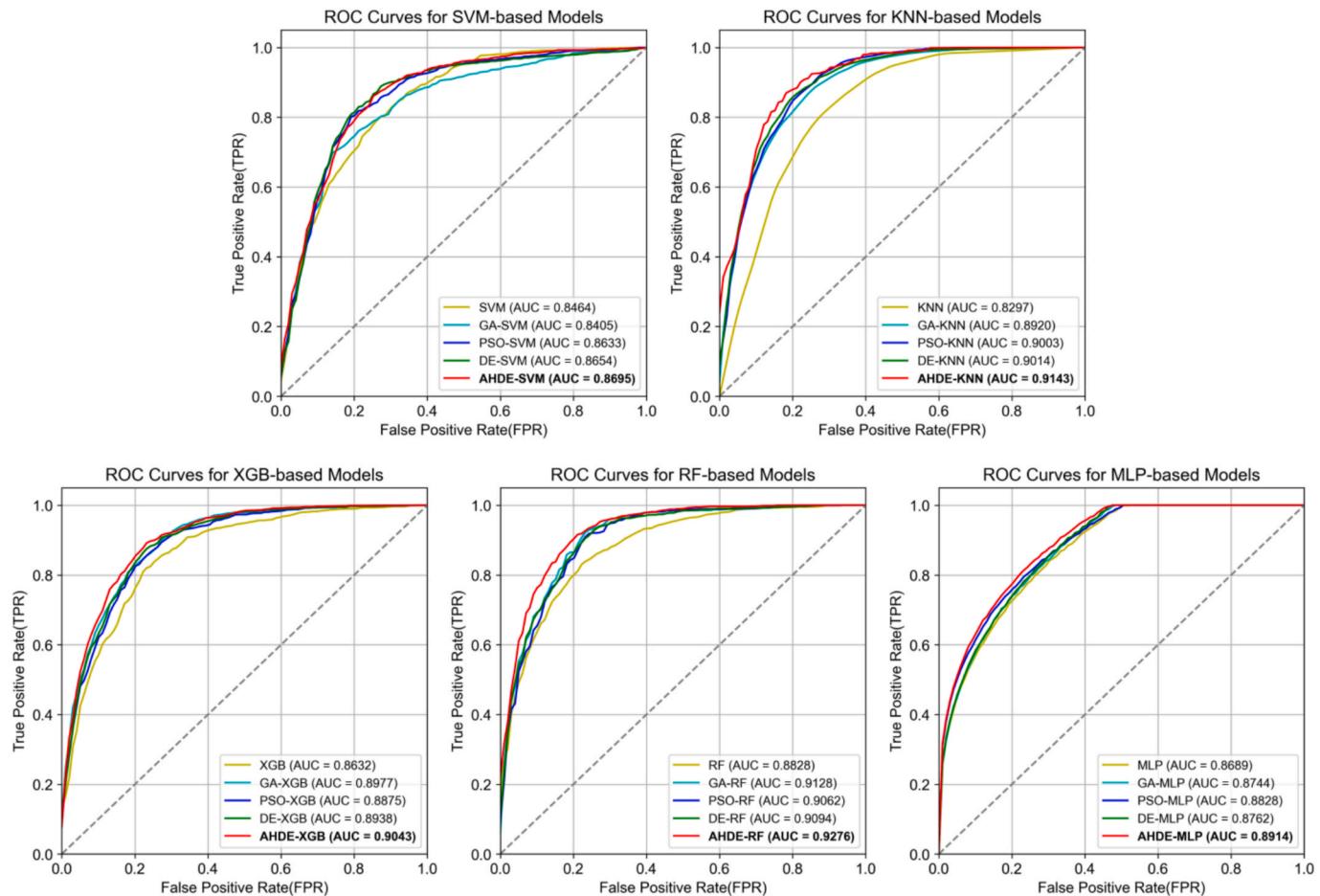


Fig. 3. Performance Comparison of Optimization Models Using ROC Curves on Dataset 1 (PIDD).

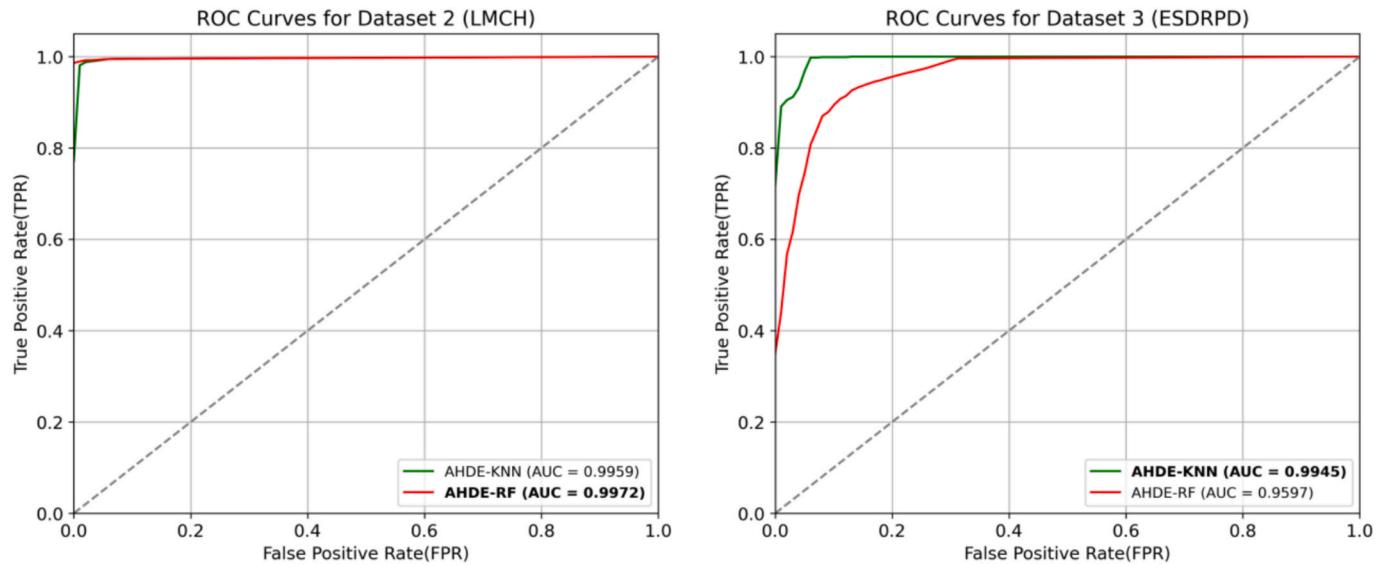


Fig. 4. ROC Curves of AHDE-KNN and AHDE-RF on Dataset 2 (LMCH) and 3 (ESDRPD).

clear that the Glucose feature is central to model construction, with BMI and Age also being essential for early diabetes diagnosis.

After selecting the optimal hyperparameter configuration and feature subsets for AHDE-RF, we ranked feature importance across the three datasets. Fig. 7(a) reinforces the significant role of Glucose in early diabetes diagnosis, while BMI, Age, and Insulin are also notable features.

These findings align with previous studies on PIDD that consistently identified Glucose and BMI as crucial predictors [10,76,77]. Fig. 7(b) presents the ranking of feature importance for Dataset 2 (LMCH), where HbA1c has an importance value of 0.45, which is consistent with clinical practice using HbA1c as a key indicator for diabetes diagnosis and blood glucose control [78]. BMI follows with an importance of 0.38, indicating

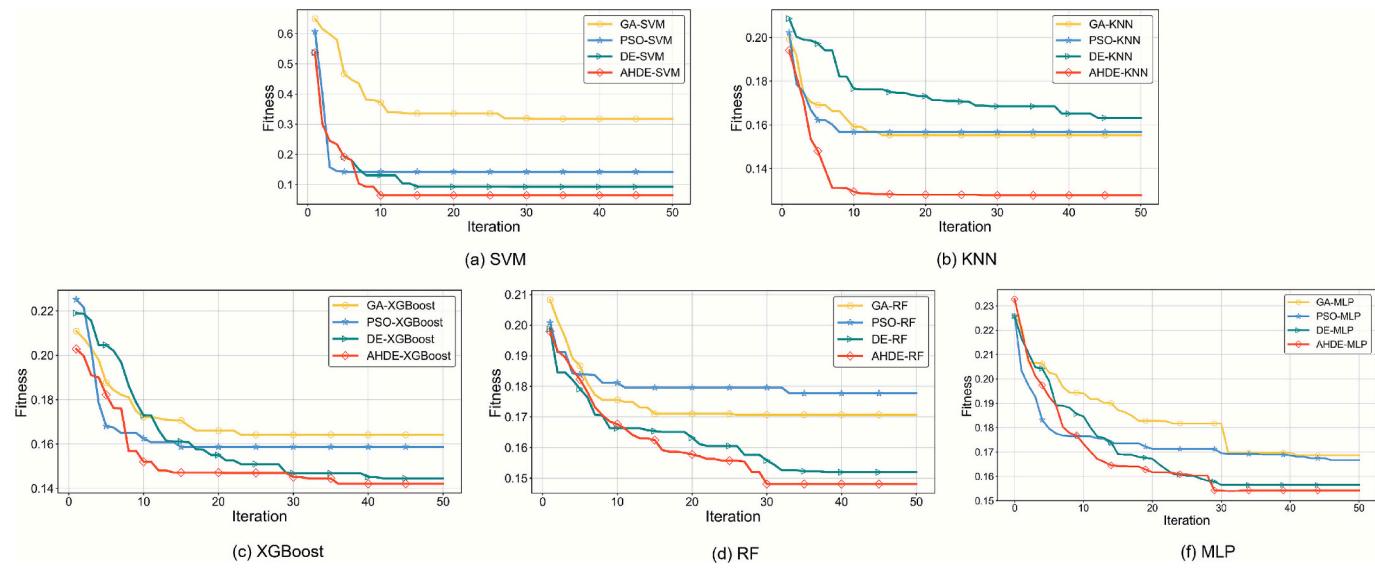


Fig. 5. Performance Convergence of Various Optimization Models on Dataset 1 (PIDD).

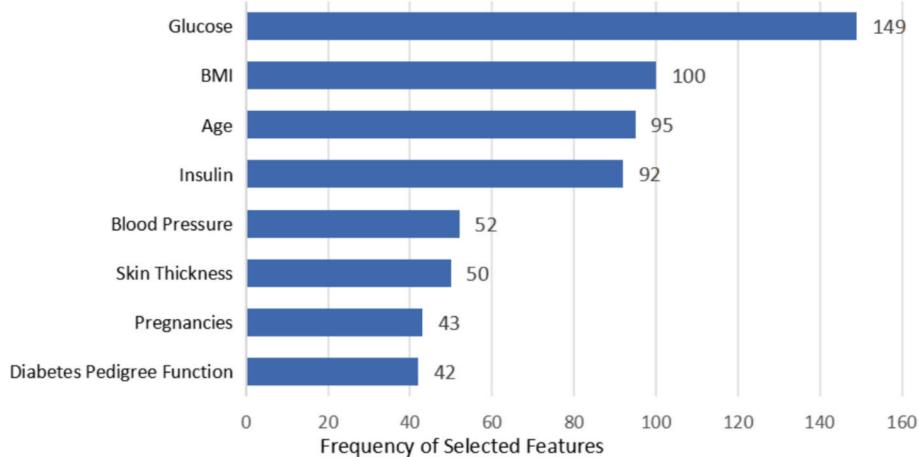


Fig. 6. Frequency of Selected Features by Optimization Models on Dataset 1 (PIDD).

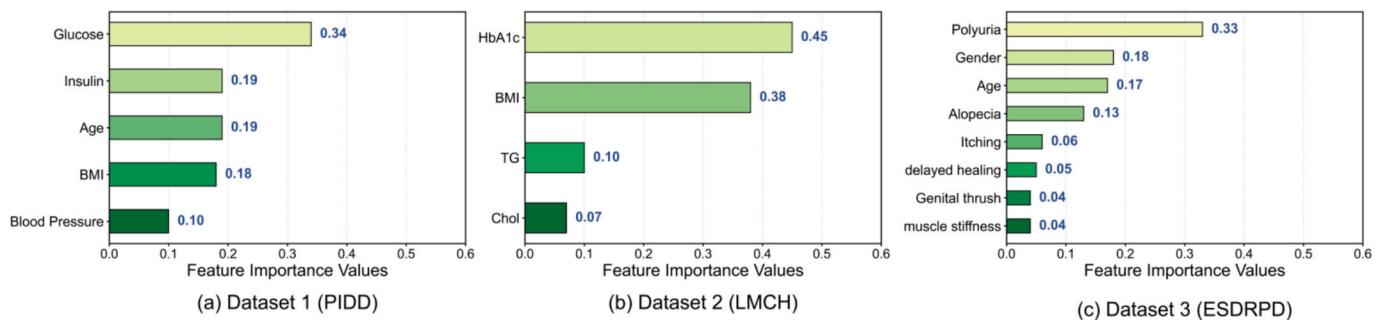


Fig. 7. Feature Importance Ranking for Datasets.

that weight management and metabolic health are relevant in the study population. Although TG, an important component of blood lipids, has a lower importance value than the previous two, it still reflects considerable influence, suggesting that triglycerides should not be overlooked in health assessments. Fig. 7(c) displays the ranking of feature importance for Dataset 3 (ESDRPD), where Polyuria, Gender, and Age are the top three features, with importance values of 0.33, 0.18, and 0.17, respectively, highlighting their relevance in this dataset. This feature

importance ranking aligns with previous ESDRPD analyses that highlighted the significance of these clinical indicators [31,79].

3.2.2. Ablation study

To systematically evaluate the contribution of each key component in the AHDE framework, we conducted an ablation study by independently removing the adaptive parameter mechanism and the dual-phase mutation strategy. We used AHDE-RF as the base model due to its

superior overall performance on Dataset 1. **Table 8** presents the performance comparison across three datasets, where * denotes statistically significant differences ($p < 0.05$, Wilcoxon signed-rank test) compared to the full AHDE model. **Fig. 8** illustrates the convergence behavior of different model variants.

Removing the adaptive parameter mechanism (w/o Adaptive) led to substantial performance degradation across all datasets. On Dataset 1 (PIDD), recall, accuracy, and F1 score decreased by 4.24%, 4.23%, and 4.24%, respectively. Similar trends were observed on Dataset 2 (LMCH) and Dataset 3 (ESDRPD), with accuracy declining by 1.61% and 1.78%, respectively. The convergence curves reveal that w/o Adaptive consistently converged to higher fitness values with slower convergence rates, particularly evident after iteration 20, indicating that the adaptive parameter mechanism is crucial for dynamic exploration-exploitation balance.

Removing the dual-phase mutation strategy (w/o Dual Mutation) resulted in moderate performance reductions, though generally less severe than removing the adaptive mechanism. On Dataset 1 (PIDD), recall and F1 score decreased by 2.81% and 2.73%, respectively. On Dataset 2 (LMCH), the variant achieved competitive performance with only marginal declines in recall (0.95%) and precision (0.69%), suggesting that highly balanced datasets are less sensitive to mutation strategies. However, on Dataset 3 (ESDRPD), the w/o Dual Mutation variant showed lower precision (2.38% decrease) and F1 score (2.02% decrease) despite achieving higher accuracy. The convergence curves indicate that w/o Dual Mutation reached slightly higher final fitness values than AHDE (Full), confirming its contribution to optimization efficiency.

3.2.3. Comparison with state-of-the-art methods

After comprehensive internal evaluations, we further compared our AHDE framework with recent competitive methods on Dataset 1 (PIDD), which is widely used in existing studies and enables direct comparison with state-of-the-art approaches. Unlike existing methods that treat feature selection and hyperparameter optimization as separate processes, AHDE's unified optimization approach demonstrated improved performance across multiple metrics.

As shown in **Table 9**. The proposed AHDE framework was evaluated against diabetes prediction methods published between 2020 and 2024, demonstrating competitive performance across multiple metrics. The framework achieved an ROC AUC of 0.9276, showing significant improvements over other methods including Saxena et al.'s approach (0.8460) [80] and Choubey et al.'s method (0.8440) [81]. The recall rate reached 0.9226, surpassing A-HMDE method [39] (0.8821) by 4.05%. While Li et al.'s HS-Kmeans approach [76] achieved the highest accuracy of 0.9165, our model maintained competitive performance with an accuracy of 0.8481. In terms of precision, our score of 0.8165 was competitive with most methods, though lower than Ramesh et al.'s 0.873 [82]. The F1 score of 0.8574 outperformed other methods where

this metric was reported. These results indicate that the AHDE framework, by integrating feature selection and hyperparameter optimization, achieves balanced and competitive performance across all metrics, providing a reliable solution for diabetes prediction tasks. Particularly, our method shows strong performance in ROC AUC and recall metrics, which are crucial indicators for medical diagnosis applications.

4. Discussion

4.1. Comprehensive performance analysis

Experimental results demonstrate that AHDE achieves significant improvements across key performance metrics, particularly in recall and ROC AUC. As shown in **Table 5**, AHDE optimization elevates recall rates above 0.85 across all base models (SVM, KNN, XGB, RF), with AHDE-KNN and AHDE-RF reaching 0.9226 and 0.9034, respectively. The ROC curve analysis in **Fig. 3** further confirms that AHDE maintains high true positive rates while effectively controlling false positives, with AUC values consistently exceeding 0.90, notably achieving exceptional performance above 0.99 on the LMCH dataset. These improvements result from the combination of three synergistic mechanisms: a heterogeneous encoding scheme, an entropy-based adaptive strategy for diversity control, and a dual-phase mutation approach.

Analysis of three datasets reveals unique performance patterns between AHDE-KNN and AHDE-RF models. Both achieve comparable high performance on Dataset 1 (continuous medical measurements) and Dataset 2 (diverse biochemical indicators), but significant differences emerge on Dataset 3 (binary diagnostic features). AHDE-KNN shows superior performance on ESDRPD, with higher AUC (0.9945 vs. 0.9597), recall (0.9989 vs. 0.9870), and F1-score (0.9602 vs. 0.9313). This stems from KNN's distance-based calculations being more effective for binary feature spaces, where symptom-based Euclidean distances provide meaningful similarity measures, while RF's tree-splitting mechanism excels with continuous variables but faces limitations with binary symptomatic data. This suggests AHDE-KNN as the preferred choice for symptom-dominant diagnostic scenarios.

The ablation study systematically validated the contribution of AHDE's key components. Removing the adaptive parameter mechanism resulted in substantial performance degradation, with recall and F1 score decreasing by over 4% on Dataset 1 (PIDD), demonstrating its critical role in balancing exploration and exploitation. The dual-phase mutation strategy showed moderate but consistent improvements, particularly on imbalanced datasets like ESDRPD where precision gains reached 2.38%. Convergence analysis further demonstrated that both components accelerate optimization, with the full AHDE model achieving superior solutions in fewer iterations, validating their synergistic contribution to navigating the complex heterogeneous search space.

The feature importance analysis identified a hierarchical structure of

Table 8

Ablation study of AHDE components in terms of Recall, Accuracy, Precision, F1 Score, and AUC measures across three datasets. [* denotes statistical significance ($p < 0.05$) compared to AHDE (Full); Bold: best performance].

Dataset	Metric	Rec	Acc	Prec	F1 score	AUC
Dataset 1 (PIDD)	w/o Adaptive	0.8650*	0.7820*	0.8122	0.8210*	0.8850*
	w/o Dual Mutation	0.8780*	0.7951	0.8253	0.8340*	0.9020
	AHDE (Full)	0.9034	0.8165	0.8481	0.8574	0.9276
Dataset 2 (LMCH)	w/o Adaptive	0.9720*	0.9735*	0.9750*	0.9735*	0.9920
	w/o Dual Mutation	0.981	0.9915	0.9820*	0.9815	0.9955
	AHDE (Full)	0.9894	0.9894	0.9888	0.9894	0.9972
Dataset 2 (ESDRPD)	w/o Adaptive	0.9720*	0.8920	0.8850*	0.8980*	0.9380*
	w/o Dual Mutation	0.9805	0.9085	0.9020*	0.9125*	0.9485*
	AHDE (Full)	0.9870	0.8856	0.9240	0.9313	0.9597

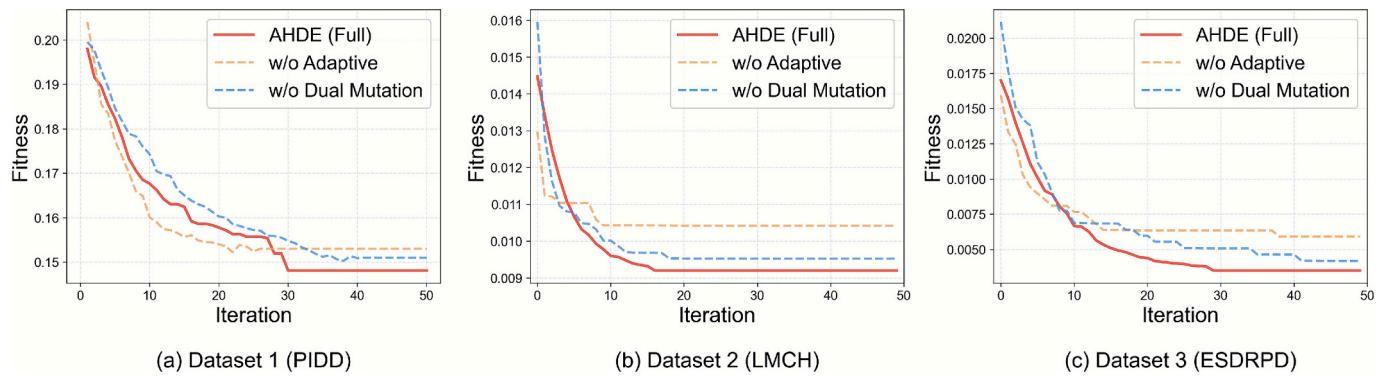


Fig. 8. Performance Convergence Curves of AHDE Variants in Ablation Study.

Table 9

Comparison of the proposed methods with the state-of-the-art [Bold: highest performance value in each column].

Author	Year	Feature Selection(FS) and Hyperparameter Optimization (HPO)	Classifier	Accuracy	Precision	Recall (SN)	F1 Score	ROC AUC
Choubey et al. [81]	2020	FS:PSO	Naïve Bayes	0.7869	0.7800	0.7900	—	0.8440
Singh et al. [10]	2020	FS:Correlation Based	Ensemble Methods	—	0.8420	0.7890	—	—
Roy et al. [83]	2021	—	RF	—	0.8500	0.8500	0.8500	—
Khanam et al. [84]	2021	FS:Pearson Correlation HPO:Manual tuning	RF,KNN	0.7942	0.8040	0.7940	—	—
Ramesh et al. [82]	2021	FS: Extra trees, LASSO HPO:GridSearch	SVM-RBF	0.8320	0.873	0.79	—	—
Naseem et al. [85]	2022	—	LG, ANN	0.8100	0.7500	0.5600	0.6500	—
Fitriyani et al. [86]	2022	—	LR	—	—	—	—	0.7900
Saxena et al [80]	2022	FS:CorrelationHPO:Grid Search (GS)	KNN,RF,DT,MLP	0.7980	—	0.7980	0.7980	0.8460
Tasin et al. [87]	2023	FS:Mutual Information(MI)HPO:Grid Search (GS)	XGB	0.8100	0.8100	—	—	0.84
Li et al. [76]	2023	FS:Harmony Search(HR)-Kmeans	KNN	0.9165	0.5000	0.9111	—	—
Gupta et al. [88]	2023	HPO:Grid Search(GS)	SVM,KNN,DT,RF	0.8052	74.4700	0.7273	72.730	—
Mostafa et al. [39]	2024	FS:Adaptive Hybrid-Mutated Differential Evolution(A-HMDE)	KNN	0.7914	—	0.8821	—	—
Mostafa et al. [39]	2024	FS:Particle Swarm Optimization(PSO)	KNN	0.7881	—	0.8691	—	—
Fatahi et al. [89]	2024	FS:Improved Binary Quantum-based Avian Navigation Optimizer Algorithm(IBQANA)	KNN	0.7746	—	—	—	—
Idris et al. [90]	2024	FS:Recursive Feature Elimination	RF	0.7908	0.7606	0.7047	0.7186	—
Our proposed (AHDE + SVM,KNN,XGB,RF)				0.8481	0.8165	0.9226	0.8574	0.9276

predictive indicators across datasets. Blood glucose consistently emerged as the primary predictor, followed by HbA1c as a crucial long-term glycemic control indicator. Secondary indicators including BMI, age, and clinical symptoms formed a comprehensive evaluation framework. Notably, AHDE's feature selection identified dataset-specific optimal subsets: Dataset 1 (PIDD) benefited from glucose-centric features, while Dataset 3 (ESDRPD) relied more heavily on symptomatic indicators. This adaptive feature selection demonstrates AHDE's capability to identify context-appropriate biomarker combinations rather than applying a universal feature set across different clinical scenarios.

4.2. Clinical implications

Based on our findings, the AHDE framework demonstrates significant clinical application potential through a two-stage implementation pathway.

Stage 1: Primary Screening in Community and Primary Care Settings. The framework serves as an efficient screening tool by processing routine clinical measurements to identify high-risk individuals

warranting further evaluation. High sensitivity is crucial in opportunistic screening scenarios where missing potential cases is more costly than additional confirmatory testing. The feature selection capability enhances practical applicability—requiring only several indicators rather than comprehensive laboratory panels reduces screening costs and enables large-scale population screening in resource-limited settings.

Stage 2: Supporting Clinical Decision-Making. For high-risk individuals, the framework provides interpretable feature importance rankings that facilitate personalized intervention planning. When integrated into electronic health record systems, it generates automated risk alerts and highlights key parameters requiring attention, supporting physicians in prioritizing patients for intensive follow-up or preventive interventions. This decision support is particularly critical for managing prediabetic or early diabetic patients, when timely interventions can effectively delay disease progression.

The model can be embedded as a lightweight module within health information systems, automatically calculating risk scores from routine clinical data. The reduced feature requirements enable deployment in

settings with limited laboratory capabilities, using only basic point-of-care measurements. This accessibility extends diabetes screening to underserved populations and rural areas, ultimately enhancing early detection reach and impact.

4.3. Limitations and future directions

Despite demonstrating improved performance in diabetes prediction, theoretical analysis reveals several critical limitations of AHDE in practical applications.

From a methodological perspective, our comparison with contemporary deep learning architectures remains limited. While we included a multilayer perceptron as a representative baseline, advanced models such as attention-based networks or transformers may offer complementary advantages for complex pattern recognition. However, integrating AHDE with these frameworks presents challenges regarding computational efficiency and hyperparameter complexity. Additionally, AHDE encounters computational challenges in high-dimensional feature spaces due to heterogeneous encoding and entropy calculations, while showing susceptibility to local optima.

Regarding data, our framework currently operates exclusively on tabular clinical data without incorporating multi-modal information sources. Integration of medical imaging, clinical text notes, genetic data, and time-series physiological signals could enhance predictive performance. Furthermore, the framework shows sensitivity to data distribution characteristics, with potential performance variations in imbalanced datasets and those containing missing values.

Concerning generalizability, while AHDE demonstrated consistent performance across tested datasets, validation across diverse populations and healthcare settings remains limited. The datasets would benefit from expansion in scale and diversity to evaluate effectiveness across different demographic groups and healthcare systems.

These findings direct future improvements. We plan to develop efficient encoding schemes for high-dimensional spaces and mechanisms for handling class imbalance and missing data. We aim to extend the framework to cardiovascular diseases, cancer risk assessment, and neurological disorders, conducting multi-center validation studies. We intend to develop real-time clinical monitoring capabilities through algorithmic optimization and parallel computing. Finally, we seek to explore integration with wireless medical sensors and mobile health platforms, incorporating multi-modal data sources to support precision medicine and intelligent healthcare systems.

5. Conclusion

This study presents AHDE (Adaptive Heterogeneous Differential Evolution), an improved optimization framework addressing a fundamental challenge in machine learning-based diabetes prediction: the inherent interdependency between feature selection and hyperparameter optimization. Unlike conventional sequential approaches that optimize these components separately, AHDE achieves simultaneous optimization through three innovations: (1) heterogeneous encoding with type-specific operators handling binary feature selection via tournament selection while processing continuous and discrete hyperparameters through differential mutation; (2) entropy-based adaptive diversity control independently regulating exploration intensity in discrete and continuous subspaces; and (3) convergence-aware staged strategy dynamically balancing exploration and exploitation based on fitness improvement and diversity metrics.

Comprehensive evaluations across three public diabetes datasets demonstrate AHDE's effectiveness. The optimized models achieved exceptional recall rates of 0.9226, 0.9894, and 0.9989, with ROC AUC values of 0.9276, 0.9972, and 0.9945, substantially outperforming sequential baselines (recall: 0.7108–0.8341). AHDE-KNN excelled on symptom-dominant binary datasets while AHDE-RF performed better with continuous measurements and biochemical indicators, revealing

important algorithm-data compatibility insights. Feature importance analysis identified blood glucose and HbA1c as primary predictors, followed by BMI, age, and clinical symptoms, aligning with clinical diagnostic criteria and validating medical interpretability.

The significance extends across multiple dimensions. Methodologically, AHDE overcomes sequential optimization's suboptimality by jointly optimizing interdependent components, providing a generalizable solution for mixed-variable medical diagnosis problems. Clinically, it serves as an efficient screening tool while providing interpretable decision support through transparent feature rankings. By prioritizing recall to minimize missed diagnoses, AHDE aligns with screening priorities. Practically, computational efficiency and reduced feature requirements enable deployment in resource-limited settings via point-of-care devices, extending screening to underserved populations.

This study acknowledges limitations including limited deep learning comparison, computational challenges in extremely high-dimensional spaces, tabular-data-only applicability, and limited cross-population validation. Future work will improve computational efficiency, integrate multi-modal data, extend to deep learning architectures, and conduct multi-center validation studies. The framework will be applied to cardiovascular diseases, cancer risk assessment, and neurological disorders, contributing to trustworthy, interpretable AI systems for precision medicine.

Authors contribution

L.C. and J.C. conceived the study. L.C. performed data analysis and drafted the manuscript. J.C. supervised the study and provided critical feedback. G.W. and Y.L. assisted with data collection and analysis. All authors contributed to figure and table preparation, reviewed and edited the manuscript, and approved the final version.

CRediT authorship contribution statement

Lin Chen: Writing – original draft, Methodology, Data curation. **Jinzhou Cao:** Writing – review & editing, Supervision, Conceptualization. **Guoqiang Wu:** Visualization, Data curation. **Yuanqi Li:** Visualization, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I shared the link to my data in the Attach File step. The datasets used in this study are publicly accessible from three different sources: the Pima Indians Diabetes Database (PIDD) is available from Kaggle (<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>); the Logical Medical Center Hospital (LMCH) dataset can be accessed through Mendeley Data (<https://data.mendeley.com/datasets/wj9rwkp9c2/1>); and the Early Stage Diabetes Risk Prediction Dataset (ESDRPD) is accessible from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset>).

References

- [1] B. Patil, R.C. Joshi, D. Toshniwal, Hybrid prediction model for type-2 diabetic patients, *Expert Syst. Appl.* (2010), <https://doi.org/10.1016/j.eswa.2010.05.078>.
- [2] H.T. Abbas, et al., Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test, *PLoS One* 14 (2019) e0219636.
- [3] R.D. Joshi, C.K. Dhakal, Predicting type 2 diabetes using logistic regression and machine learning approaches, *Int. J. Environ. Res. Public Health* 18 (2021) 7346.

- [4] S. Oviedo, et al., Risk-based postprandial hypoglycemia forecasting using supervised learning, *Int. J. Med. Inf.* 126 (2019) 1–8.
- [5] D.Y. Shin, B. Lee, W.S. Yoo, J.W. Park, J.K. Hyun, Prediction of diabetic sensorimotor polyneuropathy using machine learning techniques, *J. Clin. Med.* 10 (2021) 4576.
- [6] G. Aguilera-Venegas, A. López-Molina, G. Rojo-Martinez, J.L. Galán-García, Comparing and tuning machine learning algorithms to predict type 2 diabetes mellitus, *J. Comput. Appl. Math.* 427 (2023) 115115.
- [7] K. De Silva, et al., Nutritional markers of undiagnosed type 2 diabetes in adults: Findings of a machine learning analysis with external validation and benchmarking, *PLoS One* 16 (2021) e0250832.
- [8] I. Gnanadass, Prediction of gestational diabetes by machine learning algorithms, *IEEE Potentials* 39 (2020) 32–37.
- [9] F.A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, S.A.C. Bukhari, Detection and prediction of diabetes using data mining: a comprehensive review, *IEEE Access* 9 (2021) 43711–43735.
- [10] N. Singh, P. Singh, Stacking-based multi-objective evolutionary ensemble framework for prediction of diabetes mellitus, *Biocybern. Biomed. Eng.* 40 (2020) 1–22.
- [11] Q. Wang, et al., DMP-MI: an effective diabetes mellitus classification algorithm on imbalanced data with missing values, *IEEE Access* 7 (2019) 102232–102238.
- [12] Q. Zou, et al., Predicting diabetes mellitus with machine learning techniques, *Front. Genet.* 9 (2018).
- [13] A. Yaqoob, N.K. Verma, R.M. Aziz, M.A. Shah, Optimizing cancer classification: a hybrid RDO-XGBoost approach for feature selection and predictive insights, *Cancer Immunol. Immunother.* 73 (2024) 261.
- [14] J. Li, et al., Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques, *Int. J. Med. Inf.* 149 (2021) 104429.
- [15] Y.L. Muller, et al., A missense variant Arg611Cys in Lipe which encodes hormone sensitive lipase decreases lipolysis and increases risk of type 2 diabetes in American Indians, *Diabetes Metab. Res. Rev.* 38 (2022) e3504.
- [16] M. Munoz-Organero, Deep physiological model for blood glucose prediction in T1DM patients, *Sensors* 20 (2020) 3896.
- [17] S.A. Solodskikh, A.S. Velikorondy, V.N. Popov, Predictive estimates of risks associated with type 2 diabetes mellitus on the basis of biochemical biomarkers and derived time-dependent parameters, *J. Comput. Biol.* 26 (2019) 1041–1049.
- [18] A. Yaqoob, N.K. Verma, R.M. Aziz, M.A. Shah, RNA-seq analysis for breast cancer detection: a study on paired tissue samples using hybrid optimization and deep learning techniques, *J. Cancer Res. Clin. Oncol.* 150 (2024) 455.
- [19] Y. Fan, Diabetes diagnosis using a hybrid CNN LSTM MLP ensemble, *Sci. Rep.* 15 (2025) 26765.
- [20] Z. Dong, et al., A multimodal transformer system for noninvasive diabetic nephropathy diagnosis via retinal imaging, *Npj Digit. Med.* 8 (2025) 50.
- [21] Sharifani, K. & Amini, M. Machine learning and deep learning: A review of methods and applications. SSRN Scholarly Paper at <https://papers.ssrn.com/abstract=4458723> (2023).
- [22] Dhanka, S., Sharma, A., Kumar, A., Maini, S. & Vundavilli, H. Advancements in hybrid machine learning models for biomedical disease classification using integration of hyperparameter-tuning and feature selection methodologies: A comprehensive review, *Arch. Comput. Methods Eng.* 10.1007/s11831-025-10309-5 (2025) doi:10.1007/s11831-025-10309-5.
- [23] P. Probst, M.N. Wright, A.-L. Boulesteix, Hyperparameters and tuning strategies for random forest, *Wires Data Min. Knowl. Discov.* 9 (2019) e1301.
- [24] E. Hancer, B. Xue, M. Zhang, Differential evolution for filter feature selection based on information theory and feature ranking, *Knowl.-Based Syst.* 140 (2018) 103–119.
- [25] M. Labani, P. Moradi, F. Ahmadizar, M. Jalili, A novel multivariate filter method for feature selection in text classification problems, *Eng. Appl. Artif. Intell.* 70 (2018) 25–37.
- [26] Yaqoob, A., Verma, N. K., Aziz, R. M. & Saxena, A. Enhancing feature selection through metaheuristic hybrid cuckoo search and harris hawks optimization for cancer classification. In *Metaheuristics for Machine Learning* (eds Kalita, K., Ganesh, N. & Balamurugan, S.) 95–134 (Wiley, 2024). doi:10.1002/9781394233953.ch4.
- [27] B. Nouri-Moghadam, M. Ghazanfari, M. Fathian, A novel multi-objective forest optimization algorithm for wrapper feature selection, *Expert Syst. Appl.* 175 (2021) 114737.
- [28] P. Wang, B. Xue, J. Liang, M. Zhang, Feature clustering-Assisted feature selection with differential evolution, *Pattern Recognit.* 140 (2023) 109523.
- [29] S. Maldonado, J. López, Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification, *Appl. Soft Comput.* 67 (2018) 94–105.
- [30] H. Liu, M. Zhou, Q. Liu, An embedded feature selection method for imbalanced data classification, *IEEECAA J. Autom. Sin.* 6 (2019) 703–715.
- [31] L.P. Joseph, E.A. Joseph, R. Prasad, Explainable diabetes classification using hybrid bayesian-optimized TabNet architecture, *Comput. Biol. Med.* 151 (2022) 106178.
- [32] J. Mens Health 20 (2024) 38.
- [33] H. Acikgoz, A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting, *Appl. Energy* 305 (2022) 117912.
- [34] M. Daviran, A. Maghsoudi, R. Ghezelbash, B. Pradhan, A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach, *Comput. Geosci.* 148 (2021) 104688.
- [35] S. Chitnis, R. Hosseini, P. Xie, Brain tumor classification based on neural architecture search, *Sci. Rep.* 12 (2022) 19206.
- [36] B. Mohan, J. Badra, A novel automated SuperLearner using a genetic algorithm-based hyperparameter optimization, *Adv. Eng. Softw.* 175 (2023) 103358.
- [37] Q. Liu, L. Wu, F. Wang, W. Xiao, A novel support vector machine based on hybrid bat algorithm and its application to identification of low velocity impact areas, *IEEE Access* 8 (2020) 8286–8299.
- [38] K. Balasubramanian, A. Williams, I. Butun, Feature importance guided random forest learning with simulated annealing based hyperparameter tuning, *Preprint at 10.48550/arXiv. 2511 (2025) 00133.*
- [39] R.R. Mostafa, et al., An adaptive hybrid mutated differential evolution feature selection method for low and high-dimensional medical datasets, *Knowl.-Based Syst.* 283 (2024) 111218.
- [40] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* 13 (2012) 281–305.
- [41] M. Türkoğlu, H. Polat, C. Koçak, O. Polat, Recognition of DDoS attacks on SD-VANET based on combination of hyperparameter optimization and feature selection, *Expert Syst. Appl.* 203 (2022) 117500.
- [42] F.J. Martínez-de-Pison, R. González-Sendino, A. Aldama, J. Ferreiro-Cabello, E. Fraile-García, Hybrid methodology based on Bayesian optimization and GA-PARSIMONY to search for parsimony models by combining hyperparameter optimization and feature selection, *Neurocomputing* 354 (2019) 20–26.
- [43] Parameters optimization of support vector machines for imbalanced data using social ski driver algorithm | neural computing and applications. <https://link.springer.com/article/10.1007/s00521-019-04159-z>.
- [44] Q. Zhang, H. Gao, Z.-H. Zhan, J. Li, H. Zhang, Growth optimizer: a powerful metaheuristic algorithm for solving continuous and discrete global optimization problems, *Knowl.-Based Syst.* 261 (2023) 110206.
- [45] A. Zandvakili, M.M. Javidi, N. Mansouri, Simultaneous feature selection and SVM optimization based on fuzzy signature and chaos GOA, *Evol. Syst.* 15 (2024) 1907–1937.
- [46] F. Konak, M.A. Bülbül, D. Türkoğlu, Feature selection and Hyperparameters Optimization Employing a Hybrid Model based on Genetic Algorithm and Artificial Neural Network: forecasting dividend payout Ratio, *Comput. Econ.* 63 (2024) 1673–1693.
- [47] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *J. Glob. Optim.* 11 (1997) 341–359.
- [48] Genetic algorithms | ACM computing surveys. <https://dl.acm.org/doi/abs/10.1145/234313.234350>.
- [49] Kennedy, J. & Eberhart, R. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks* vol. 4 1942–1948 vol.4 (1995).
- [50] Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C. & Johannes, R. S. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. in 261–265 (1988).
- [51] A. Rashid, *Diabetes Dataset*. 1 (2020).
- [52] Unknown. Early stage diabetes risk prediction. UCI Machine Learning Repository 10.24432/C5VG8H (2020).
- [53] *Diabetes Care* 41 (2018) 917–928.
- [54] R. Kahn, et al., Age at initiation and frequency of screening to detect type 2 diabetes: a cost-effectiveness analysis, *Lancet* 375 (2010) 1365–1374.
- [55] S.G. Pauker, J.P. Kassirer, The threshold approach to clinical decision making, *N. Engl. J. Med.* 302 (1980) 1109–1117.
- [56] P. Zhang, et al., Costs of screening for pre-diabetes among U.S. adults: a comparison of different screening strategies, *Diabetes Care* 26 (2003) 2536–2542.
- [57] American Diabetes Association, *Standards of care in diabetes—2023 abridged for primary care providers*, *Clin. Diabetes* 41 (2023) 4–31.
- [58] M. Lozano, D. Molina, F. Herrera, Editorial scalability of evolutionary algorithms and other metaheuristics for large-scale continuous optimization problems, *Soft. Comput.* 15 (2011) 2085–2087.
- [59] P. Belotti, et al., Mixed-integer nonlinear optimization, *Acta Numer.* 22 (2013) 1–131.
- [60] Črepinské, M., Liu, S.-H. & Mernik, M. Exploration and exploitation in evolutionary algorithms: A survey. *ACM Comput Surv* 45, 35:1–35:33 (2013).
- [61] Li, J. et al. Feature selection: a data perspective. *ACM Comput Surv* 50, 94:1–94:45 (2017).
- [62] C.E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* 27 (1948) 379–423.
- [63] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, *IEEE Trans. Evol. Comput.* 1 (1997) 67–82.
- [64] G. Rudolph, Convergence analysis of canonical genetic algorithms, *IEEE Trans. Neural Netw.* 5 (1994) 96–101.
- [65] A.E. Eiben, S.K. Smit, Parameter tuning for configuring and analyzing evolutionary algorithms, *Swarm Evol. Comput.* 1 (2011) 19–31.
- [66] J. Brest, S. Greiner, B. Boskovic, M. Mernik, V. Zumer, Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems, *IEEE Trans. Evol. Comput.* 10 (2006) 646–657.
- [67] S. Das, P.N. Suganthan, Differential evolution: a survey of the state-of-the-art, *IEEE Trans. Evol. Comput.* 15 (2011) 4–31.
- [68] H. Lai, H. Huang, K. Keshavjee, A. Guergachi, X. Gao, Predictive models for diabetes mellitus using machine learning techniques, *BMC Endocr. Disord.* 19 (2019) 101.
- [69] R. Shwartz-Ziv, A. Armon, Tabular data: deep learning is not all you need, *Inf. Fusion* 81 (2022) 84–90.
- [70] L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* 35 (2022) 507–520.
- [71] N. Tran, J.-G. Schneider, I. Weber, A.K. Qin, Hyper-parameter optimization in classification: To-do or not-to-do, *Pattern Recognit.* 103 (2020) 107245.

- [72] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, *Neurocomputing* 415 (2020) 295–316.
- [73] P. Zhang, Y.F. Jin, Z.Y. Yin, Y. Yang, Random forest based artificial intelligent model for predicting failure envelopes of caisson foundations in sand, *Appl. Ocean Res.* 101 (2020).
- [74] R. Valarmathi, T. Sheela, Heart disease prediction using hyper parameter optimization (HPO) tuning, *Biomed. Signal Process. Control* 70 (2021) 103033.
- [75] F.Z. El-Hassani, M. Amri, N.-E. Joudar, K. Haddouch, A new optimization model for MLP hyperparameter tuning: Modeling and resolution by real-coded genetic algorithm, *Neural Process. Lett.* 56 (2024) 105.
- [76] X. Li, J. Zhang, F. Safara, Improving the Accuracy of Diabetes Diagnosis applications through a Hybrid Feature selection Algorithm, *Neural Process. Lett.* 55 (2023) 153–169.
- [77] M.Y. Shams, Z. Tarek, A.M. Elshewey, A novel RFE-GRU model for diabetes classification using PIMA Indian dataset, *Sci. Rep.* 15 (2025) 982.
- [78] P. Colman, et al., New classification and criteria for diagnosis of diabetes mellitus. the australasian working party on diagnostic criteria for diabetes mellitus, *N. Z. Med. J.* (1999). <https://www.semanticscholar.org/paper/New-classification-and-criteria-for-diagnosis-of-on-Colman-Thomas/ae33f8d4f4c2808389b9226f6a3d95dfc3f3e841>.
- [79] E. García-Escobar, et al., Fatty liver index as a predictor for type 2 diabetes in subjects with normoglycemia in a nationwide cohort study, *Sci. Rep.* 11 (2021) 16453.
- [80] R. Saxena, S.K. Sharma, M. Gupta, G.C. Sampada, A novel approach for feature selection and classification of diabetes mellitus: machine learning methods, *Comput. Intell. Neurosci.* 2022 (2022) 3820360.
- [81] D.K. Choubey, P. Kumar, S. Tripathi, S. Kumar, Performance evaluation of classification methods with PCA and PSO for diabetes, *Netw. Model. Anal. Health Inform. Bioinforma.* 9 (2020) 5.
- [82] J. Ramesh, R. Aburukba, A. Sagahyroon, A remote healthcare monitoring framework for diabetes prediction using machine learning, *Healthc. Technol. Lett.* 8 (2021) 45–57.
- [83] K. Roy, et al., An enhanced machine learning framework for type 2 diabetes classification using imbalanced data with missing values, *Complexity* 2021 (2021) e9953314.
- [84] J.J. Khanam, S.Y. Foo, A comparison of machine learning algorithms for diabetes prediction, *ICT Express* 7 (2021) 432–439.
- [85] A. Naseem, et al., Novel internet of things based approach toward diabetes prediction using deep learning models, *Front. Public Health* 10 (2022).
- [86] N.L. Fitriyani, et al., A comprehensive analysis of Chinese, Japanese, Korean, US-PIMA Indian, and trinidadian screening scores for diabetes risk assessment and prediction, *Mathematics* 10 (2022) 4027.
- [87] Tasin, I., Nabil, T. U., Islam, S. & Khan, R. Diabetes prediction using machine learning and explainable AI techniques. 10.1049/htl2.12039 doi:10.1049/htl2.12039.
- [88] S.C. Gupta, N. Goel, Predictive modeling and analytics for diabetes using hyperparameter tuned machine learning techniques, *Procedia Comput. Sci.* 218 (2023) 1257–1269.
- [89] A. Fatahi, M.H. Nadimi-Shahraki, H. Zamani, An improved binary quantum-based avian navigation optimizer algorithm to select effective feature subset from medical data: a COVID-19 case study, *J. Bionic Eng.* 21 (2024) 426–446.
- [90] N.F. Idris, et al., Stacking with recursive feature elimination-isolation forest for classification of diabetes mellitus, *PLoS One* 19 (2024) e0302595.