Management Strategies and Engineering Sciences 2026; 8(2):1-23

A Proposed Model for Diabetes Diagnosis Using a Hybrid Approach Based on an Improved Harris Hawks Algorithm and Machine Learning Methods

Fatemeh. Khani[1], Saeed. Ayat[2*], Mahdi. Esmaeili[3]

1. Ph.D. Student, Department of Computer, Kas.C., Islamic Azad University, Kashan, Iran
2. Associate Professor, Department of Computer Engineering and Information Technology, Payame Noor University, Tehran, IRAN
3. Associate Professor, Department of Computer, Kas.C., Islamic Azad University, Kashan, Iran

*** Corresponding author email address:** dr.ayat@pnu.ac.ir

Received: 2025-11-12 **Revised:** 2026-03-11 **Accepted:** 2026-03-18 **Published:** 2026-03-30

Abstract

Early detection of diabetes represents a major challenge in healthcare systems, as a considerable proportion of patients are identified only at advanced stages, when opportunities for effective intervention have already diminished. Enhancing the accuracy of machine learning models depends critically on selecting optimal features—a process that becomes particularly important in high dimensional medical datasets. In the present study, an improved version of the Harris Hawks Optimization algorithm, referred to as Improved Harris Hawks Optimization (IHHO), was employed for feature selection. The proposed enhancements include the incorporation of chaotic maps for initialization, modification of the prey energy function, and the use of Lévy flight to strengthen the balance between exploration and exploitation. The performance of IHHO, in combination with eight widely used machine learning classifiers—RF, SVM, ANN, Logistic Regression, Linear Regression, KNN, Naive Bayes, and Decision Tree—was evaluated using the Pima Indians Diabetes dataset. The findings indicate that IHHO substantially outperforms both the Baseline methods and the standard HHO algorithm, achieving its best results in the IHHO+RF combination, with an Accuracy of 97.20 %, Precision of 94.51 %, Recall of 98.33 %, and an F1 Score of 96.38 %. Statistical analyses, including ANOVA, Post hoc tests, and the Friedman test, further confirm the significance of these differences. The results demonstrate that IHHO can serve as a robust framework for feature selection in medical diagnosis and can contribute to the development of reliable intelligent systems for diabetes screening.

Keywords: *Diabetes, Feature Selection, Improved Harris Hawks Optimization, Machine Learning, Disease Diagnosis*

How to cite this article:

Khani, F., Ayat, S., & Esmaeili, M. (2026). A Proposed Model for Diabetes Diagnosis Using a Hybrid Approach Based on an Improved Harris Hawks Algorithm and Machine Learning Methods. Management Strategies and Engineering Sciences, 8(2), 1-23.

1. Introduction

Diabetes mellitus is a chronic metabolic disorder characterized by persistent hyperglycemia resulting from impaired insulin secretion, impaired insulin action, or a combination of both. Its clinical significance extends beyond abnormal blood glucose because prolonged metabolic dysregulation can contribute to cardiovascular, renal, neurological, ocular, and vascular complications that progressively diminish patients' functional capacity and quality of life. Contemporary clinical overviews identify diabetes as a major global health concern whose prevention, early detection, and long-term management require coordinated clinical, behavioral, and technological interventions [1]. The burden of diabetes is also economic, encompassing direct medical expenditures, productivity losses, disability, and the substantial costs of treating complications that might have been prevented or delayed through earlier intervention [2]. Consequently, diabetes management increasingly depends on systematic data analysis, continuous monitoring, and decision-support technologies capable of converting heterogeneous patient information into clinically meaningful risk estimates [3]. Big-data analytics has expanded this potential by enabling the integration of laboratory findings, demographic characteristics, physiological measurements, medical histories, and longitudinal records to improve both

population-level surveillance and individual healthcare decision-making [4].

Early diagnosis is particularly important because the metabolic abnormalities associated with type 2 diabetes may develop gradually and remain undetected until considerable physiological damage has occurred. Conventional screening procedures usually rely on individual clinical thresholds and periodic laboratory assessments, whereas predictive models can identify multivariable patterns associated with disease risk. Machine learning has therefore emerged as an important component of chronic-disease detection by uncovering nonlinear relationships in routinely collected medical information [5]. Reviews of computational intelligence in diabetes show that current prediction and monitoring systems encompass classical classifiers, ensemble methods, deep-learning architectures, optimization procedures, and hybrid decision-support models [6]. Unified disease-prediction systems similarly reflect growing demand for reusable machine-learning frameworks that can support the diagnosis of multiple chronic conditions while retaining disease-specific predictive characteristics [7]. Beyond episodic diagnosis, intelligent monitoring is progressing toward distributed edge- and cloud-based systems, as demonstrated by deep-learning applications for artificial-pancreas environments in which rapid and context-sensitive decision-making is essential [8].

A substantial body of research has established the usefulness of supervised learning for diabetes classification. Data-mining techniques have been applied to discover diagnostic patterns in clinical attributes and compare the performance of commonly used classifiers [9]. Machine-learning analysis of demographic and health-survey information has also shown that diabetes detection can be extended beyond narrowly defined clinical samples to larger population datasets [10]. Comparative diabetes-prediction studies have evaluated logistic regression, support vector machines, decision trees, random forests, k-nearest neighbors, neural networks, and probabilistic models, indicating that their performance depends considerably on data preparation, feature representation, and model configuration [11]. The Pima Indians Diabetes dataset remains one of the most frequently used benchmarks for examining such models under standardized conditions [12]. Investigations using both the Pima Indian and Frankfurt datasets further indicate that dataset composition and distribution can influence predictive behavior, emphasizing the need to evaluate the robustness of proposed methods across different data structures [13].

Recent studies have increasingly moved beyond individual conventional classifiers toward deep and ensemble architectures. Stacking ensembles combine the complementary decision boundaries of multiple base learners and may reduce the limitations associated with dependence on a single model family [14]. Deep-learning approaches seek to construct higher-level representations of clinical variables and identify interactions that may remain hidden to simpler algorithms [15]. Multi-layer perceptron models have likewise been optimized and comprehensively evaluated on the Pima Indians dataset, demonstrating the continuing relevance of neural architectures for diabetes classification [16]. Hybrid convolutional neural networks extend this direction by integrating representation learning with prediction procedures designed for diabetes data [17]. Explainable deep gated networks further attempt to combine high predictive capacity with information about the variables influencing early diabetes-risk estimates [18]. Nevertheless, increasing architectural complexity does not independently resolve problems caused by noisy, redundant, imbalanced, or weakly informative input variables.

Interpretability, trustworthiness, and privacy are also central when machine-learning predictions are intended to support clinical decisions. A self-explainable diabetes-diagnosis interface can help clinicians and patients understand how specific attributes contribute to a prediction, thereby improving usability and reducing the opacity associated with complex models [19]. Trustworthy prediction additionally requires attention to data governance and the confidentiality risks associated with aggregating sensitive medical information. Privacy-conscious frameworks based on deep residual networks and proximity-based data processing demonstrate how diabetes prediction can be designed to address predictive performance and patient privacy simultaneously [20]. A clinically useful system must therefore provide not only high aggregate accuracy but also stable, reproducible, interpretable, and ethically defensible predictions.

One of the most persistent methodological challenges in diabetes prediction is class imbalance. In many datasets, non-diabetic cases outnumber diabetic cases, allowing a classifier to achieve apparently acceptable accuracy while failing to detect a clinically significant proportion of affected patients. Methodologies developed specifically for imbalanced diabetes data show that sampling procedures, loss functions, thresholds, and validation strategies can

substantially alter sensitivity and specificity [21]. Attention-enhanced deep belief networks have been proposed for early diabetes-risk prediction under highly imbalanced conditions, illustrating how imbalance can be addressed within the learning architecture [22]. Synthetic minority oversampling represents another important approach, and novel oversampling techniques continue to be developed for complex imbalance problems [23]. Advanced prediction frameworks have also integrated imbalance-handling procedures with machine learning to reduce majority-class dominance [24]. Because an undetected diabetic patient may lose the opportunity for timely treatment, false-negative reduction is particularly important in screening systems [25]. Research based on patient-perceived symptoms similarly demonstrates that classification performance depends on how heterogeneous clinical and symptom information is represented before model training [26].

Feature selection is consequently a decisive component of reliable diabetes-prediction systems. Medical datasets may contain irrelevant, redundant, correlated, noisy, or weakly informative variables that increase computational cost and impair generalization. Adaptive ensemble feature selection can identify stable predictive attributes across different model families rather than linking the selection procedure to one classifier [27]. Hybrid feature-selection and logistic-regression pipelines have similarly been developed to improve early diabetes detection while preserving the interpretability of the final prediction model [28]. Metaheuristic feature selection provides an alternative to exhaustive search by exploring numerous candidate subsets and balancing predictive performance against the number of retained variables [29]. Federated disease-prediction approaches that integrate feature selection and feature extraction further demonstrate that dimensionality management can be combined with decentralized learning when raw patient information cannot be centrally shared [30]. Improved feature selection through simulation optimization confirms that variable selection should be treated as a structured search problem rather than a sequence of isolated univariate decisions [31]. Systematic examinations of high-dimensional metaheuristic feature selection nevertheless identify scalability, computational expense, convergence stability, and reproducibility as unresolved challenges [32].

The effectiveness of feature selection is closely connected to the characteristics of the final classifier. K-nearest neighbor methods, for example, are sensitive to irrelevant dimensions and distorted distance calculations, motivating numerous modifications to improve their robustness [33]. Naive Bayes is computationally efficient and interpretable but may be restricted by its conditional-independence assumption; ensemble procedures developed to enrich recall and precision show that even comparatively simple probabilistic models can benefit from structured combination strategies [34]. Reviews of machine- and deep-learning methods for diabetes prediction conclude that no classifier is uniformly superior across all datasets and preprocessing configurations [35]. A new feature-selection algorithm should therefore be evaluated with several classifier families to determine whether its benefits generalize across linear, nonlinear, probabilistic, instance-based, tree-based, and neural learning paradigms.

Metaheuristic optimization has become attractive for feature selection because the number of possible subsets grows exponentially with the number of candidate variables. Comprehensive taxonomies describe metaheuristics as flexible search mechanisms for nonlinear, discontinuous, combinatorial, and multiobjective problems, while identifying parameter sensitivity, premature convergence, reproducibility, and limited theoretical guarantees as important concerns [36]. Recent evaluations of newly developed metaheuristics similarly emphasize the diversity of biologically, physically, socially, and mathematically inspired algorithms and the importance of rigorous comparative evaluation rather than reliance on algorithmic novelty alone [37]. Metaheuristic optimization has been employed in chronic-disease multiclassification to tune parameters, select variables, and improve decision boundaries, although maintaining an effective balance between exploration and exploitation remains difficult [38]. Swarm-intelligence-based neural models in medical diagnosis have shown promising performance, but their interpretability, computational complexity, robustness, and generalizability require further investigation [39]. Bibliographic analyses of chaos-enhanced metaheuristic optimizers also suggest that chaotic dynamics are increasingly used to diversify initial populations, reduce search stagnation, and strengthen global exploration [40].

The Harris Hawks Optimization algorithm is relevant to this problem because it uses population-based cooperative search and alternates between exploratory and exploitative behaviors. However, its standard form may experience inadequate initial diversity, unstable transitions between search phases, and entrapment in local optima. Improved Harris Hawks Optimization has been developed for engineering problems by modifying the algorithm's search

dynamics and strengthening its capacity to navigate complex objective landscapes [41]. Related optimization research shows that chaotic mappings and adaptive transformations can improve population distribution and convergence, as demonstrated by an arithmetic optimization algorithm incorporating cosine-transform-based two-dimensional composite chaotic mapping [42]. Vertical and horizontal crossover strategies combined with random-walk mechanisms have likewise been used to improve the global and local search capabilities of population-based optimizers [43]. Within medical imaging, an HHO-optimized vision transformer has been proposed for diabetic-retinopathy detection, providing evidence that Harris-hawk-based search can be integrated effectively with advanced diagnostic architectures [44]. These developments support further examination of modified HHO procedures for tabular diabetes diagnosis, particularly when the objective is to identify a compact and discriminative feature subset.

Hybrid optimization and classification systems have already demonstrated potential in diabetes-related applications. An evolutionary ensemble model combining harmony search and stacking illustrates how metaheuristic search can be integrated with heterogeneous learners to improve diagnostic prediction [45]. Integrated data-mining and metaheuristic techniques have also been applied to predicting diabetic-patient readmission risk, showing that optimization can support clinically relevant outcomes beyond initial disease detection [46]. Optimized hybrid classifiers further indicate that combining complementary computational mechanisms can strengthen analytical models for identifying diabetic patients [47]. However, observed gains may originate from feature optimization, classifier complexity, data balancing, or dataset-specific tuning. Rigorous assessment should therefore compare classifiers without feature selection, classifiers using standard HHO, and classifiers using the improved algorithm under equivalent preprocessing and validation conditions.

Another relevant distinction concerns feature selection and feature extraction. Feature selection preserves the original clinical attributes, whereas feature extraction creates transformed representations that may be more difficult to interpret clinically. Comparative machine-learning research has shown that choosing between these dimensionality-reduction approaches can influence predictive accuracy, efficiency, and computational complexity [48]. In diabetes screening, preserving recognizable variables such as plasma glucose, body mass index, blood pressure, insulin, age, pregnancies, and diabetes pedigree information may facilitate clinical interpretation. A selection method capable of eliminating redundant attributes without obscuring the original variables is therefore particularly appropriate for systems intended to complement professional judgment.

Accordingly, this study aimed to develop and evaluate a hybrid diabetes-diagnosis model that employs an Improved Harris Hawks Optimization algorithm incorporating chaotic initialization, a modified prey-energy function, and Lévy-flight-based search to select optimal features and enhance the performance of eight machine-learning classifiers on the Pima Indians Diabetes dataset.

2. Methodology

In this study, an intelligent framework for diabetes diagnosis is developed based on integrating a metaheuristic optimization algorithm with machine learning models. The primary objective of the framework is to improve predictive accuracy while reducing model complexity through effective feature selection and optimized learning of data patterns. The general structure of the proposed system is designed as a multi-stage process that includes data collection, preprocessing, optimal feature selection using the Improved Harris Hawks Optimization (IHHO) algorithm, classification, and performance evaluation.

In the first stage, patient-related data are extracted from standard and reliable diabetes datasets. These datasets typically contain variables such as age, gender, body mass index (BMI), blood pressure, glucose level, insulin level, and family history of diabetes. In the second stage, preprocessing procedures are applied to ensure the quality and consistency required for machine learning models. This stage involves removing missing values, normalizing numerical attributes, and encoding categorical features. In the third stage, the IHHO algorithm is employed to identify the optimal subset of features. By incorporating three main strategies—using chaotic maps for population initialization, modifying the prey-energy function, and integrating Lévy flight during the exploration phase—the algorithm achieves an effective balance between global and local search and extracts the most influential features associated with diabetes diagnosis. The goal of this step is to eliminate redundant or non-informative features and retain the attributes with the highest diagnostic relevance.

In the fourth stage, the selected features are used as inputs to several machine learning models, including Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), Decision Tree (DT), Logistic

Regression (LR), Linear Regression (LnR), Naïve Bayes (NB), and k-Nearest Neighbors (KNN). These models are trained to predict whether a patient is diabetic or non-diabetic. Finally, the system's performance is assessed using statistical evaluation metrics such as Accuracy, Recall, Precision, and the F1-score.

Overall, the proposed framework is illustrated schematically in Figure 1. As shown, the system consists of a coherent processing flow that spans from raw data acquisition to the final output, where each component plays a critical role in enhancing the overall predictive accuracy. The integration of data-mining techniques with the IHHO algorithm substantially improves the model's ability to extract hidden patterns and enhances decision interpretability, thereby providing an effective tool to support clinical decision-making in the healthcare domain.

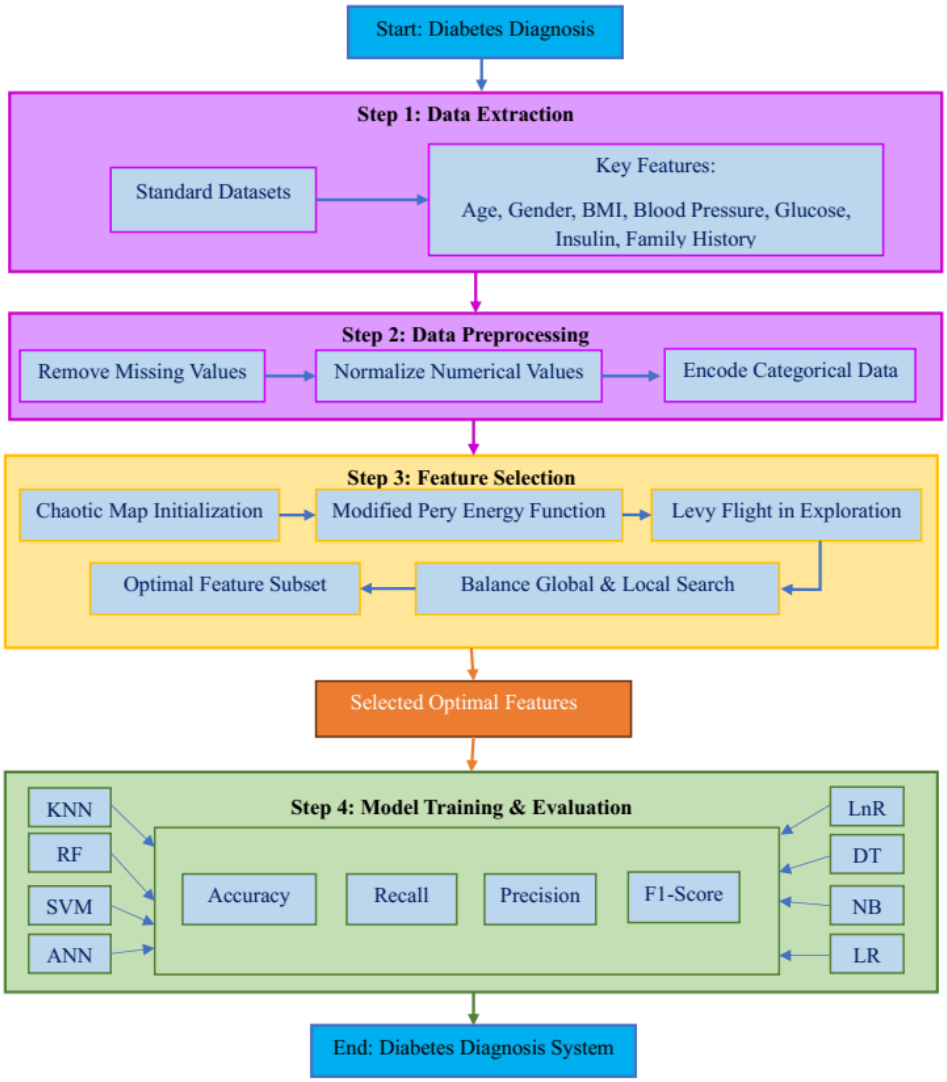

Figure 1. Schematic diagram of the proposed intelligent framework for diabetes diagnosis.

2.1. Dataset Description

To evaluate the effectiveness of the proposed framework, the well-known PIMA Indians Diabetes Dataset—collected by the National Institute of Diabetes and Digestive and Kidney Diseases—was employed. This dataset is widely used in diabetes-related research as a standard benchmark and contains medical information on Native American women aged 21 years and older. The dataset comprises 798 samples and eight independent attributes that represent key clinical indicators associated with the development of type 2 diabetes. These features reflect various aspects of the patients' metabolic status, including glucose level, blood pressure, body mass index (BMI), and family history of

diabetes. Table 1 presents the attributes of the dataset, their data types, and the corresponding value ranges.

Table 1. data set features

Row	Description	Data Type	Data Range
1	Number of pregnancies	Discrete	0-17
2	Plasma glucose concentration 2 hours after OGTT	Discrete	44-199
3	Diastolic blood pressure (mm Hg)	Discrete	24-122
4	Triceps skinfold thickness (mm)	Discrete	7-99
5	2-hour serum insulin (MUU/ml)	Discrete	14-846
6	Body Mass Index (BMI)	Continuous	18.2-67.1
7	Diabetes Pedigree Function (DPF)	Continuous	0.078-2.42
8	Age (years)	Discrete	21-81
9	Class variable (0=non-diabetic, 1=diabetic)	Categorical	—

The output (label) represents the presence or absence of diabetes, defined as a binary variable (1 for diabetic and 0 for non-diabetic). To prevent model bias caused by class imbalance, the SMOTE method was applied for class balancing. As expressed in Equation (1), this approach generates synthetic minority samples based on the Euclidean distance in the feature space, thereby improving class distribution and enhancing the predictive accuracy of learning models.

$$x_{new} = x_i + \lambda(x_{nn} - x_i)$$

(1)

where xi refers to a minority-class sample, xnn denotes one of its nearest neighbors, and lambda is a random number within the interval [0, 1].

2.2. Data Preprocessing

Raw data typically contain noise, missing values, outliers, and irregular observations, all of which can significantly degrade the performance of machine learning models. Therefore, before conducting feature selection and classification, a data preprocessing procedure was implemented to enhance the overall quality and consistency of the dataset. This stage involves several key steps, which are described below.

a) Data Cleaning

Missing values were first handled using appropriate statistical techniques, such as replacing them with the mean value of each feature [50]. To reduce the impact of outliers, the quantile trimming method [51] was applied within the 5th to 95th percentile range to remove abnormal observations. In addition, duplicate records were eliminated to prevent bias during the training process.

b) Feature Normalization

Since the features operate on different numerical scales, normalization was applied to prevent features with larger ranges from dominating the learning process. The Min–Max normalization method was used, which maps each feature value into the interval [0, 1] [52]. Equation (2) presents the formula associated with this method.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(2)

where x represents the original value of the feature, xmin and xmax denote the minimum and maximum values of that feature, respectively, and x’ indicates the normalized value. This approach facilitates the convergence of learning models and enhances the numerical stability of computations.

c) Encoding of Categorical Features

Some features, such as gender or family history of diabetes, are qualitative in nature and must be converted into numerical values before they can be utilized by machine learning models. In this study, Label Encoding was applied to binary features (such as gender), while One-Hot Encoding was used for multi-level categorical attributes. These transformations preserve the categorical structure of the data without imposing artificial ordinal Equation ships among the categories. For binary features, a direct mapping to numerical values is applied. If C = {c1, c2, … , ck} represents the set of categorical values of a given feature, the label encoding process is defined according to Equation (3).

$$f(c_i) = i - 1, i = 1,2,\ldots,k$$

(3)

For multi-level categorical features, each category is mapped to a binary vector in which the component corresponding to the target category is set to 1 and all other components are set to 0. For example, if the feature "physical activity" includes three levels—low, moderate, and high—then the encoding is represented as follows:

• low → (1, 0, 0)
• moderate → (0, 1, 0)
• high → (0, 0, 1)

d) Data Splitting for Training and Testing

Finally, the preprocessed data were divided into training (80%) and testing (20%) subsets to prevent overfitting during the learning process and to evaluate the performance of the model on previously unseen data. The preprocessed dataset was then forwarded to the feature selection module based on the improved Harris Hawks Optimization algorithm.

2.3. Feature Selection Using the Improved Harris Hawks Optimization (IHHO)

Feature selection is one of the most critical steps in the process of medical disease diagnosis through data mining, as it eliminates irrelevant attributes, reduces data dimensionality, enhances classification accuracy, and improves the computational efficiency of the model. In this study, an improved version of the Harris Hawks Optimization algorithm was employed to obtain an optimal subset of features. This algorithm, inspired by the cooperative hunting behavior of Harris hawks, establishes an effective balance between exploration and exploitation.

In the standard HHO algorithm, the hunters (hawks) attempt to identify the optimal solution by searching the feasible space and adapting their movements relative to the position of the prey. However, the original algorithm often suffers from insufficient diversity in the initial population and a limited capability to escape from local optima [55, 56]. To address these limitations, three key strategies were incorporated into the algorithm in this study:

1. Enhancement of the Initialization Phase Using Chaotic Maps

To prevent premature convergence and to increase the diversity of the initial population, this study employed the Tent chaotic map for initializing the population in the improved Harris Hawks Optimization algorithm [57]. Chaotic maps, due to their nonlinear behavior, high sensitivity to initial conditions, and uniform distribution over the search space, serve as an effective alternative to conventional random distributions [58]. These characteristics enable the algorithm to explore different regions of the search space and reduce the likelihood of becoming trapped in local optima. The general formulation of the Tent chaotic map is defined in Equation (4).

$$x_{t+1} = \begin{cases} \frac{x_t}{r}, & 0 \le x_t < r \\ \frac{1-x_t}{1-r}, & r \le x_t < 1 \end{cases} \quad (4)$$

where xt denotes the current value of the chaotic sequence, xt+1 denotes the next value, and r r r is the chaos control parameter. A typical value for r is set to 0.7 to ensure appropriate chaotic behavior. The initial value x0 is randomly selected within the interval [0, 1], and the resulting sequence is employed to generate the initial positions of the hawks within the search space. Subsequently, to produce binary values from this sequence, the following transformation is applied.

$$X_i^0 = \begin{cases} 1, & \text{if } x_t \ge 0.5 \\ 0, & \text{else} \end{cases} \quad (5)$$

Accordingly, the initial population of the algorithm is initialized using a chaotic distribution, which results in a more uniform coverage of the search space and enhances the algorithm's capability to explore optimal regions.

2. Improvement of the Prey Energy Function

In the classical HHO, the prey energy function decreases linearly, which may cause a sudden decline in the learning rate during intermediate stages. In the IHHO, an exponential function with a random decay coefficient (Equation 6) is used to model the prey's energy, enabling a smoother and more realistic learning process. In this Equation , r r r is a random coefficient within the interval [0, 1] that allows the algorithm to maintain a more dynamic balance between local and global search. This modification preserves the algorithm's dynamism throughout the search process and reduces the likelihood of being trapped in local optima.

$$E = 2E_0\, e^{-r(\frac{t}{T})^2} \quad (6)$$

where E0 is the initial energy, T T represents the total number of iterations, and t t is the current iteration.

3. Enhancement of the Exploration Phase Using Lévy Flight

To strengthen the algorithm's ability to navigate the search space, Lévy flight [59] is employed to simulate the random, zigzag movement behavior of the hawks after detecting the prey. This technique enhances the exploration capability by updating the hawks' positions using Lévy flight. The Lévy flight Equation ship is defined as Equation (7).

$$X_{t+1} = X_t + Levy(\beta) \times (X_t - X_m) \quad (7)$$

where Xm denotes the mean position of the hawks and Levy(Brepresents the random step distribution with parameter β∈[1.3,1.9]. Consequently, the hawks are more likely to escape local regions and move to new areas within the search space. Moreover, a dynamic strategy was designed to adaptively adjust the deviation rate and the step size of the Lévy flight, balancing global search during the early stages with more precise local search in later iterations. This adaptive adjustment is described by Equation (8).

$$\alpha(t) = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min}) \times e^{-\gamma t/T_{max}} \quad (8)$$

where a(t) denotes the deviation rate at iteration t, Tmax is the total number of iterations, and a control coefficient for the convergence rate. By this approach, the deviation rate gradually decreases, so that by the end of the search process, the algorithm focuses on optimal regions. At the end of each iteration, the best hawk (i.e., the solution with the lowest fitness value) is considered the temporary optimal position. The process continues until the stopping criteria are met (a predefined number of iterations or no improvement over several consecutive iterations).

Given the above design, the IHHO algorithm is expected to intelligently and efficiently explore the feature space and, by selecting a minimal subset of effective features, improve the accuracy of the diabetes diagnosis model.

2.4. Objective Function

In the feature selection process, each hawk is represented as a binary vector, where each component indicates the inclusion (1) or exclusion (0) of a particular feature. To align with clinical and technical requirements of diagnosis, the objective function is defined as a weighted sum of performance and cost criteria. The selected criteria include the average Area Under the Curve (AUC) and sensitivity calculated via k-fold cross-validation, the ratio of selected features, and computational cost. These criteria are combined as follows and minimized by the IHHO algorithm.

$$\text{Fitness}(S) = w_1(1 - \text{AUC}) + w_2 \frac{|S|}{N} + w_3(1 - \text{Sensitivity}_{CV}) + w_4(1 - \text{Specificity}_{CV}) \quad (9)$$

Where S S S denotes the set of selected features and N N N represents the total number of features. The criteria incorporated in the objective function are:

- **AUCCV:** The average Area Under the Curve calculated via cross-validation.
- **SensitivityCV:** The average sensitivity (Recall) over the same cross-validation.
- **SpecificityCV:** The average specificity over the same cross-validation, included to prevent false positive alarms.
- **SN:** The ratio of selected features.

The weights wi were assigned in accordance with the research requirements as follows: w1=0.40, w2=0.25, w3=0.10, and w4=0.25. The criteria are computed based on the mean results of cross-validation to ensure generalizability and stability of the measurements.

Accordingly, the output of the IHHO algorithm is an optimal subset of features that maximizes discriminative power while minimizing the computational complexity of the model. These selected features are subsequently fed into machine learning classifiers for diabetes diagnosis.

2.5. Classification Using Machine Learning Algorithms

Following feature selection with the IHHO algorithm and determination of the optimal subset of discriminative features for diabetes diagnosis, the classification phase aims to predict the status of patients (diabetic / non-diabetic). To provide a comprehensive evaluation and assess the robustness of the proposed feature selection method, eight diverse machine learning classifiers covering various algorithmic paradigms were employed. This diverse set includes ensemble methods, linear models, instance-based learners, probabilistic classifiers, and neural networks, thereby enabling assessment of the generalizability of the proposed IHHO method across different learning techniques. These models were selected for their high capability in modeling nonlinear Equation ships, good performance on medical data, and favorable generalizability. All algorithms were trained using 10-fold cross-validation to ensure the stability and reliability of results.

2.5.1. *Support Vector Machine (SVM)*

Support Vector Machine [60] is among the most powerful supervised learning algorithms for binary classification problems, aiming to find an optimal hyperplane that separates data into two distinct classes. In this study, the Radial Basis Function (RBF) kernel was used due to its strong ability to model nonlinear decision boundaries. The SVM decision function is defined according to Equation (10). During the training process of SVM, two key hyperparameters— the penalty parameter C and kernel parameter γ— are tuned. The parameter C determines the model's rigidity in handling misclassification errors; higher values of C lead to a harder decision boundary and increased risk of overfitting, while smaller values produce a softer decision boundary promoting better generalization. The parameter γcontrols the width of the RBF kernel and determines the influence extent of training samples on the shape of the decision boundary. Both C (penalty) and γ (kernel width) were optimized via grid search to achieve the best balance between accuracy and overfitting prevention.

$$f(x) = \text{sign}(\sum_{i=1}^{N} \quad \alpha_i y_i K(x_i, x) + b)$$

$$K(x_i, x) = \exp\left(-\gamma \parallel x_i - x \parallel^2\right)$$

(10,11)

where K denotes the RBF kernel function and α_i are the Lagrange multipliers.

2.5.2. *Random Forest (RF)*

Random Forest [61] is an ensemble method based on building a collection of decision trees that determines the final label through majority voting. Due to its high resistance to overfitting, ability to handle noise, and robust performance on medical data, RF is considered a suitable option for this study. In this research, the model was trained with 100 decision trees, and the Gini index was used as the criterion for node splitting. The random selection of a subset of features for each tree fosters structural diversity in the forest and enhances the model's resistance to overfitting.

2.5.3. *Artificial Neural Network (ANN)*

Artificial Neural Network [62] was employed as the third classification model. The network architecture consists of an input layer (matching the number of selected features), a hidden layer with 20 neurons, and a binary output layer. The ReLU activation function was applied in the hidden layer to enhance nonlinear modeling capacity, while the Sigmoid function in the output layer was used to produce the probability of diabetes occurrence. The model was trained using the Adam optimizer with a learning rate of 0.001, and 100 training epochs were set to achieve proper convergence. The simple structure combined with the ability to model complex Equation ships among features motivated the choice of ANN in this study.

2.5.4. *Logistic Regression*

Logistic Regression [63] is a linear model for binary classification that estimates probabilities using a logistic function. This model employs the logistic (sigmoid) function to confine the output values to the range (0,1), interpreting the output as probabilities according to Equation (12).

$$P(y = 1 \mid X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}$$

(12)

where xi are the features and Bi are the model weights. The class assignment decision is typically performed using a threshold of 0.5; that is, if the predicted probability exceeds 0.5, the sample is assigned to the positive class, otherwise to the negative class. In this study, L2 regularization was applied to prevent overfitting. Due to its interpretability, computational efficiency, and well-established statistical properties, logistic regression is considered an important baseline. Its performance using the selected features provides insight into the linear separability of the features.

2.5.5. *K-Nearest Neighbors (KNN)*

K-Nearest Neighbors (KNN) is a non-parametric algorithm [64] used for classification and regression tasks. To predict the class of a new sample, this algorithm selects the dominant class among the k closest training samples in the feature space. The distance between samples is typically computed using metrics such as the Euclidean distance, defined by Equation (13).

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^{n} \quad (x_{im} - x_{jm})^2}$$

(13)

where xi and xj are the feature vectors of two samples, and n n n is the number of features. Due to its simplicity and

the lack of need to estimate model parameters, KNN is one of the fundamental and well-known algorithms in machine learning for classification tasks.

2.5.6. Linear Regression

Linear Regression [65] is a supervised learning algorithm used for predicting continuous values. It models the Equation ship between one or more independent features (x1,x2,…,xn) (and a dependent variable y. The linear regression model determines the coefficients βi by minimizing the sum of squared errors, which is expressed formally in Equation (14).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

(14)

where ε \varepsilon ε is the model's random error term. Linear regression is one of the simplest and most widely used methods for quantitative data analysis and continuous value prediction in scientific and engineering problems.

2.5.7. Naive Bayes

Naive Bayes is a supervised classification algorithm [66] based on Bayes' theorem and the assumption of feature independence. For a sample with features (x1,x2,…,xn) and class Ck C_k Ck is calculated according to Equation (15). The final class assignment is made by selecting the class with the maximum posterior probability.

$$P(C_k|x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} P(x_i|C_k)P(C_k)}{P(x_1,x_2,\ldots,x_n)}$$

(15)

2.5.8. Decision Tree

Decision Tree is a supervised classification and regression algorithm that partitions data into homogeneous subsets through a series of conditional decisions. Each internal node represents a feature and a condition on that feature, while each leaf node specifies the final class label or the predicted value. Feature splitting is usually performed based on criteria such as information gain or the Gini index.

$$IG(S, A) = Entropy\ (S) - \sum_{v\ \in Values(A)} \frac{|S_v|}{|S|} Entropy\ (S_v)$$

(16)

Decision trees, due to their interpretability and capability to model nonlinear Equation ships among features, are among the widely used algorithms in classification tasks.

3. Findings and Results

This section presents a set of quantitative and qualitative evaluations related to the performance of the proposed algorithm. The primary objective is to analyze the efficacy of the Improved Harris Hawks Optimization (IHHO) algorithm as a feature selection method and to examine its impact on enhancing the performance of various classifiers. To achieve this goal, the evaluation process is organized into multiple stages. First, the experimental setup, datasets, and algorithm parameters are described to clarify the framework of the experiments. Subsequently, the models' performances are reported and analyzed in three independent scenarios: the baseline case (without feature selection), the classical HHO version, and the improved IHHO version.

In the first subsection, the experimental settings including implementation details, hyperparameter tuning, evaluation metrics, number of repetitions, and environmental requirements are presented. This information ensures the reproducibility of the experiments and provides a proper foundation for the analysis of results in the following sections.

In the second subsection, the performance results of the classifiers under the three aforementioned approaches are compared using Accuracy, Precision, Recall, and F1-score metrics. This comparison aims to determine the role of feature selection and quantify the improvement achieved by IHHO at the level of learning models. Analysis of this section demonstrates that IHHO consistently enhances the performance of all classifiers and, when combined with the Random Forest model, yields the best results.

In the third subsection, an analytical discussion of the results focuses on the reasons behind the improvement, the convergence behavior of IHHO, dimensionality reduction, and the impact of this reduction on balancing Precision and Recall. Additionally, the performance of IHHO is compared against the original HHO and baseline scenarios using statistical and inferential approaches to rigorously document the effectiveness of the proposed components—including chaotic maps, the modified energy function, and Lévy flight.

3.1. Evaluation Metrics

Accuracy, especially in disease diagnosis tasks such as diabetes detection, is a key metric for assessing the

performance of classification models. This metric indicates the proportion of diabetic and non-diabetic patients correctly identified. In other words, accuracy provides an overall picture of the model's ability to correctly distinguish individuals' health status. Its computation formula is given by Equation (15).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

where:

- **TP (True Positives)** is the number of diabetic patients correctly identified by the model.
- **TN (True Negatives)** is the number of healthy individuals correctly recognized by the model.
- **FP (False Positives)** is the number of individuals incorrectly classified as diabetic by the model while they are actually healthy.
- **FN (False Negatives)** is the number of diabetic patients incorrectly classified as healthy by the model.

In medical research, including diabetes diagnosis, accuracy is an important metric to evaluate the overall predictive power of the model. However, given the critical importance of Type I and Type II errors—especially cases where diabetic patients are not detected—relying solely on accuracy is not recommended. Therefore, accuracy is used alongside other metrics such as Sensitivity (Recall), Precision, and F1-Score [41–43] to provide a comprehensive analysis of the model's performance from the perspective of correctly identifying both patients and healthy individuals.

Sensitivity (Recall) is one of the key metrics in evaluating disease detection systems, including diabetes diagnosis. It measures the model's ability to correctly detect diseased individuals and indicates what proportion of actual patients have been correctly identified by the model. The formula for Sensitivity is given by Equation (16):

$$\text{Recall (\%)} = \frac{TP}{TP+FN} \times 100 \quad (18)$$

Precision indicates how accurate the model's prediction is when it classifies an individual as diabetic. In other words, this metric measures the likelihood that the model's positive predictions are correct. The formula for Precision is expressed as Equation (19):

$$\text{Precision (\%)} = \frac{TP}{TP+FP} \times 100 \quad (19)$$

In diabetes diagnosis models, Precision plays an important role in evaluating the system's ability to avoid false alarms. False Positive (FP) alerts impose additional costs, cause patient anxiety, and lead to unnecessary tests. Therefore, although attention to Recall (minimizing missed patients) is crucial in medical diagnosis, a high Precision ensures that the system's positive predictions are accurate and reliable.

The next metric is the F-Score, which, according to Equation (6), is a type of harmonic mean between Precision and Sensitivity (Recall).

$$F - Score = \frac{2*precision+recall}{Precion+recall} \quad (20)$$

3.2. Experimental Settings

All experiments were conducted in a controlled computational environment using MATLAB R2022a software. Implementations were executed on a system equipped with an Intel Core i7-11800H processor, 16 GB RAM, and Windows 11 operating system. To reduce the impact of fluctuations caused by the stochastic nature of metaheuristic algorithms, each experiment was repeated 10 times, and the average results were reported as the final output.

The dataset, after undergoing preprocessing steps including missing value imputation, variable standardization, and Min-Max normalization, was fed into the models. Classifier performance evaluation was conducted using 10-fold cross-validation to ensure result stability, reduce overfitting, and enhance the generalizability of the models.

In all configurations, the input features were first selected using the IHHO algorithm, and then classifiers were trained based on these selected features.

The key parameters used for the metaheuristic algorithms and machine learning classifiers are presented in Table 2. This table includes all constant values such as population size and number of IHHO iterations, mutation strategies, the type of kernel used in SVM, weight update rates in ANN, the number of neighbors in KNN, and other essential settings.

Table 2. Parameter Settings of the IHHO Algorithm and Machine Learning Classifiers

IHHO	
Population	30
Max iterations	100
β (Levy Flight)	[1.5 – 1.9]
Exponential energy rate	Random value in each repeat
SVM (RBF Kernel)	
C	10
γ	0.1
Random Forest	
Number of trees	100
Splitting criterion	Gini Impurity
Artificial Neural Network	
Hidden layer neuron	20
Activation function	)output (Sigmoid) ،latent (ReLU
Training Algorithm	Adam
Learning rate	0.001
Epoch	300
Naive Bayes	
Feature function	Gaussian
Smoothness	0.001
KNN	
(K) number of neighbors	5
Distance criteria	Euclidean
Weighing strategy	Uniform
Logistic Regression	
Regression type	Binary Logistic
Regularization	L2
(λ) adjustment coefficient	0.01
Linear Regression (LnR)	
Regularization	Ridge L2
(λ) adjustment coefficient	0.001
Decision Tree (DT)	
Branching criterion	Gini
Max depth	20
Min sample for Split	2

The parameters of the IHHO algorithm, including population size, number of iterations, coefficients of the exponential energy function, Lévy flight parameter range, and chaotic map parameters, were selected according to settings reported in the literature and after preliminary testing. The population size was set to 30, and the number of iterations to 100. The Lévy distribution parameter β was assigned within the interval [1.5–1.9], and the initial value for the chaotic map was randomly set within the range (0,1).

3.3. First Experiment: Performance Analysis of Models in Three Scenarios — Baseline, HHO, and IHHO

Next, the performance of the classification algorithms is presented and analyzed in three different scenarios: the baseline case (no feature selection), feature selection using the original Harris Hawks Optimization algorithm (HHO), and feature selection using its improved version (IHHO). These three scenarios allow a precise analysis of the benefit of feature selection and the effect of the proposed improvements. Table 3 presents the complete results of all three scenarios. The aim of this analysis is to evaluate the extent to which the IHHO-based feature selection process improves classification accuracy and performance.

In the baseline scenario, where no feature selection method was applied, the accuracy of various algorithms ranged from 78.35% to 72.55%, with the highest accuracy observed for the Random Forest (RF) classifier at 78.55%. These values indicate that the presence of irrelevant and redundant features led to a reduction in the models' discriminative ability.

Table 3. Classifiers' Performance in Scenarios Without Feature Selection, with HHO-based Feature Selection, and with IHHO-based Feature Selection

Algorithm	Metric	Baseline	HHO	IHHO	Improvment IHHO vs Baseline	Improvment IHHO vs HHO
RF	Accuracy	35.78	83.87	2.97	%85.18	%37.9
	Precision	48.77	54.83	94.51	%03.17	%97.10
	Recall	54.80	89.91	98.33	%79.17	%44.6
	-Score1F	98.78	52.87	96.38	%40.17	%86.8
SVM	Accuracy	84.75	6.82	63.95	%79.19	%03.13
	Precision	52.80	35.80	62.93	%10.13	%27.13
	Recall	72	25.83	09.98	%09.26	%84.14
	-Score1F	02.76	77.81	78.95	%76.19	%01.14
ANN	Accuracy	13.74	7.79	77.94	%64.20	%07.15
	Precision	17.74	34.78	81.91	%64.17	%47.13
	Recall	6.74	75.78	99.97	%39.23	%24.19
	-Score1F	38.74	55.78	84.94	%46.20	%29.16
LnR	Accuracy	55.72	55.78	35.93	%80.20	%80.14
	Precision	09.73	47.78	64.95	%55.22	%17.17
	Recall	12.71	03.75	54.86	%42.15	%51.11
	-Score1F	1.72	71.76	86.90	%76.18	%15.14
LR	Accuracy	01.75	84.78	08.93	%07.18	%24.14
	Precision	62.74	6.78	67.94	%05.20	%07.16
	Recall	29.72	62.75	62.88	%33.16	%00.13
	-Score1F	42.73	08.77	56.91	%14.18	%48.14
KNN	Accuracy	87.74	01.80	54.96	%67.21	%53.16
	Precision	55.72	08.76	01.94	%46.21	%93.17
	Recall	06.76	08.85	24.95	%18.19	%16.10
	-Score1F	26.74	33.80	62.94	%36.20	%29.14
NB	Accuracy	91.71	57.79	96.92	%05.21	%39.13
	Precision	19.75	51.79	2.95	%01.20	%69.15
	Recall	53.72	43.77	91.86	%26.65	%18.63
	-Score1F	82.73	46.78	93.50	%26.65	%19.17

DT	Accuracy	74	89.79	48.96	%48.22	%59.16
	Precision	38.73	05.77	77.94	%39.21	%72.17
	Recall	75.72	23.82	47.89	%72.16	%24.7
	-Score1F	06.73	56.79	04.92	%98.18	%48.12

By applying the feature selection method based on HHO, a significant improvement was observed in all evaluation metrics. The accuracy of classifiers ranged from 78.55% to 87.837%, with the Random Forest (RF) classifier achieving the best performance with an accuracy of 87.33%. These results indicate that removing irrelevant features and reducing data dimensionality can enhance the stability and efficiency of learning models, although the original version of HHO still faces some limitations in local search.

Ultimately, employing the feature selection approach based on the IHHO algorithm significantly improved the performance of all models. The results show that classifier accuracy values ranged between 92.96% and 97.20%, with the highest accuracy belonging to the RF classifier at 97.20%, accompanied by an F-score of 96.38%. Additionally, the K-Nearest Neighbors (KNN) and Decision Tree (DT) classifiers demonstrated very satisfactory performances with accuracies of 96.54% and 96.48%, respectively. This significant difference compared to the classical HHO version and the baseline scenario suggests that the enhancements applied in IHHO, including the use of chaotic maps, Lévy flight, and an optimized energy function, have ensured faster convergence and more effective feature selection.

Figure 1 illustrates the trend of accuracy changes for various classification models across three optimization approaches: IHHO, HHO, and baseline. It can be observed that employing IHHO results in a stable and significant enhancement in accuracy across all models. This improvement is most pronounced in models such as RF and KNN, indicating that IHHO is capable of creating a more optimal overlap between the search space and the model structure. In contrast, HHO delivers moderate performance close to the baseline, although it still outperforms the baseline in most models. The overall trend indicates that IHHO has been able to enhance the quality of machine learning models across a wide range of classifiers and improve performance stability compared to comparative methods. This demonstrates the high efficiency of IHHO in increasing accuracy and preventing performance degradation in complex and parameter-sensitive models.

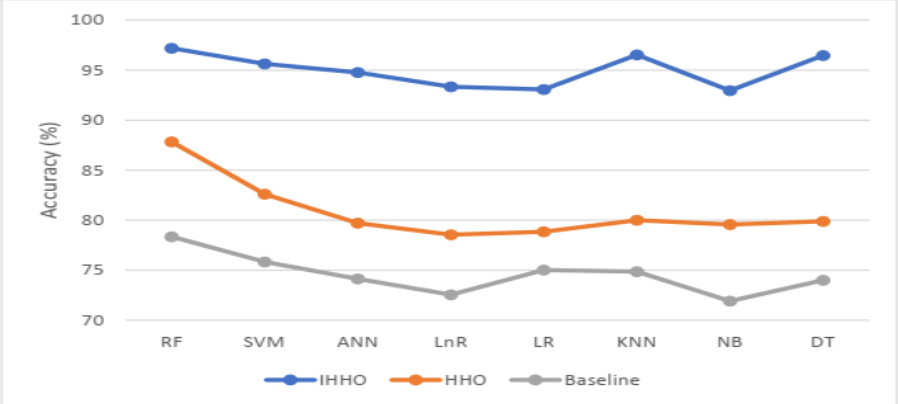

Figure 2. Accuracy Comparison of Baseline, HHO, and IHHO Methods Across All Classifiers

Figure 2 illustrates the comparison of the F1-score for a set of classification models under the three approaches: IHHO, HHO, and Baseline. The observed trend indicates that IHHO consistently delivers significantly better performance than the other two methods across all models, and the stability of F1-score values under this method is noticeably higher. A sharp decline in the F1-score is observed for the K-Nearest Neighbors (KNN) model with both HHO and Baseline methods, whereas IHHO maintains a higher level of performance. This demonstrates that IHHO is capable of maintaining a balance between Precision and Recall even in models that are highly sensitive to parameter variations. In contrast, the HHO and Baseline methods exhibit fluctuations in most models and have limited ability to prevent performance deterioration.

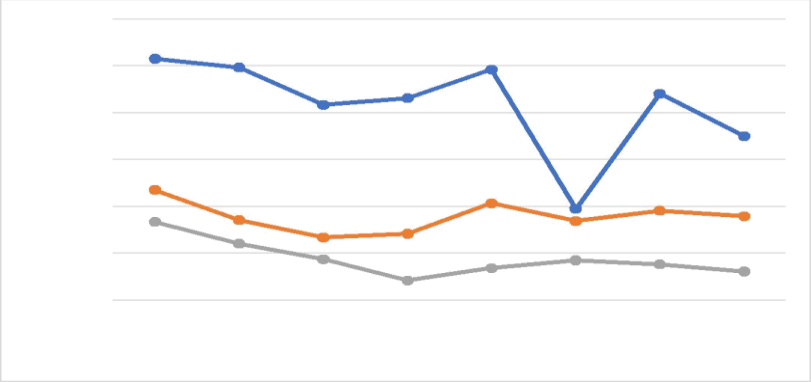

Figure 3. Comparison of the F1-Score Using Baseline, HHO, and IHHO Methods for All Classifiers

Figure 3 compares the performance of the Random Forest model across three approaches using four metrics: Accuracy, Precision, Recall, and F1-score. The results indicate that IHHO exhibits significantly higher performance than the other two methods in all metrics. The notable difference between IHHO and HHO, particularly in Recall and Accuracy, demonstrates IHHO's capability in enhancing the model's decision-making quality and reducing classification errors. Furthermore, HHO's performance is superior to the baseline across all metrics, but the gap between HHO and IHHO remains substantial. The stable performance of IHHO across all four indicators highlights the method's effectiveness in optimizing machine learning models and simultaneously improving accuracy, Generalizability, and balance among them.

The analysis of the presented Figures 1, 2, and 3 reveals that the proposed IHHO method demonstrates considerably higher and more stable performance compared to the classical HHO version and the baseline scenario across all evaluations, including Accuracy, Precision, Recall, and F1-score. The consistent and superior trend of IHHO across different classification models indicates that this method not only improves accuracy but also maintains a balance among the primary evaluation metrics, preventing the performance fluctuations observed with HHO and the baseline. IHHO's superiority in all three figures underscores its strong capability in establishing an optimal search space, selecting effective features, and preventing model performance degradation.

The results obtained from Table 3 clearly indicate that the proposed IHHO method delivers the best performance in the feature selection process for diabetes detection. On average, it has increased model accuracy by more than 10% compared to the classical HHO version and by more than 20% compared to the baseline scenario.

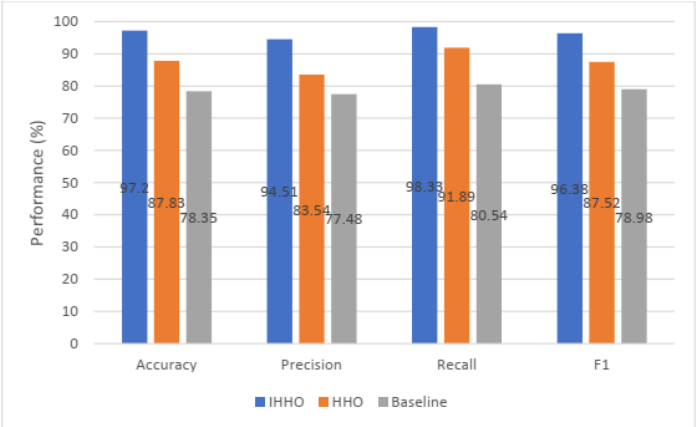

Figure 4. Comparison of RF Model Performance Among Baseline, HHO, and IHHO Methods

3.4. *Second Experiment: Comparison of the Proposed Method with Other Metaheuristic Algorithms*

To evaluate the standing of the proposed method against existing approaches, the performance of the IHHO-RF model was compared with a collection of recent studies published between 2023 and 2024. Table 4 provides a comprehensive overview of this comparison across four key metrics: Accuracy, Precision, Recall, and F1-Score. As the data shows, the proposed method has achieved the highest accuracy among all reported methods, reaching an Accuracy of 97.20%. This value represents a significant superiority compared to the closest competitive work published in 2023, which reported an accuracy of 97.06%.

Table 4. Comparative Performance of the Proposed IHHO-RF Model with State-of-the-Art Studies

Reference	Accuracy%	Precision%	Recall%	F1-Score%
IHHO + RF (Proposed)	97.20	94.51	98.33	96.38
[49]	97.06	53.96	97.56	--
[19]	80.00	62.00	76.00	68.00
[50]	91.00	93.00	99.00	96.00
[51]	91.65	---	---	---
[14]	91.50	91.50	91.06	91.49
[15]	93.09	93.22	91.06	92.25
[36]	95.42	---	95.6	---
[45]	89.81	---	89.29	---

Specifically, the accuracy of the IHHO+RF model shows a 0.14% improvement compared to the best reported methods in the literature (e.g., 97.06% in study [26]). Meanwhile, compared to weaker approaches such as study [41], which reported an accuracy of 80%, an absolute improvement of over 21.5% is observed. Furthermore, relative to works [42] and [36], which reported accuracies in the range of 91% to 93%, the proposed method achieves improvements of 6.8% and 4.11%, respectively.

Besides accuracy, the sensitivity and F1-score metrics also indicate the superiority of the proposed method. For example, the recall value of the IHHO+RF model is 98.33%, which demonstrates an improvement of about 7.27% compared to study [35] and about 7.54% compared to study [36]. The F1-score likewise shows at least a 3% to 5% better performance compared to other competitive methods. This performance gap can be attributed to the high exploratory power of IHHO, its ability to select optimal feature combinations, and the robustness of the model against noise and feature corEquation . The use of chaotic mapping, Lévy flight, and the improved energy function in IHHO has enabled a more efficient search space traversal, providing classifiers with a richer and more effective feature set. Overall, the comparative analysis indicates that the proposed method not only surpasses most reference methods but also delivers outstanding performance in terms of stability, accuracy, and balance between precision and recall.

3.5. *Statistical Analysis of Methods' Performance*

To evaluate the statistical significance of performance differences among the three methods: Baseline, HHO, and IHHO, a set of parametric and non-parametric statistical tests were conducted. The aim of this analysis is to determine whether the observed improvements in the evaluation metrics in section 4.4 are merely due to natural algorithmic variability or whether the performance increase of IHHO is statistically supported.

3.5.1. *One-way ANOVA Analysis*

The ANOVA results are presented in Table 5. The computed F-value is 89.42, and the p-value is less than 0.0001, indicating that the differences in mean performance among the three methods are statistically highly significant. This result confirms that the performance improvement of IHHO is not due to chance and shows a statistically significant difference compared to the other two methods.

Table 5. One-way ANOVA Results for Performance Comparison.

Source of Variation	SS	df	MS	F	p-value
Between Methods	1245.67	2	622.84	89.42	< 0.0001*
Within method	167.32	24	6.97		
Total	1412.99	26			

3.5.2. Tukey HSD Post-hoc Test

To determine which method significantly outperforms the others, the Tukey HSD test was conducted as a post-hoc analysis. As shown in Table 6, all three pairwise comparisons—IHHO vs. Baseline, IHHO vs. HHO, and HHO vs. Baseline—exhibit statistically significant differences with p-values less than 0.0001. The greatest mean difference is observed between IHHO and Baseline (19.21%), indicating the decisive superiority of the proposed method.

Table 6. Results of Tukey HSD Post-hoc Test for Pairwise Comparisons

	Difference	Lower	Upper		
IHHO vs Baseline	+19.21	+17.84	+20.58	0.0001 <	✓
	+11.47	+10.10	+12.84	0.0001 <	✓
HHO vs Baseline	+7.74	+6.37	+9.11	0.0001 <	✓

3.5.3. Friedman Non-Parametric Test

Given that the performance data of the classifiers across the three methods have a repeated measures and dependent structure, the Friedman test was also conducted as a complementary non-parametric analysis. The results presented in Table 7 show that for all evaluation metrics, the chi-square (χ^2) values are very high and the p-values are less than 0.0001. This indicates that the ANOVA results remain robust and valid even without the assumption of data normality.

Table 7. Results of Friedman Non-Parametric Test Across All Metrics.

Metric	χ^2	df	p-value	Result
Accuracy	16.00	2	<0.0001	Significant
Precision	14.86	2	<0.0001	Significant
Recall	15.43	2	<0.0001	Significant
F-score	16.00	2	<0.0001	Significant

3.5.4. Win-Tie-Loss Analysis

To provide a ranking evaluation independent of data distribution, the Win-Tie-Loss analysis was conducted. The results in Table 8 indicate that the IHHO method outperformed both Baseline and HHO in all 32 comparison scenarios (8 classifiers × 4 evaluation metrics) with no ties or losses recorded. This outcome demonstrates the stability and absolute superiority of IHHO across all evaluation conditions.

Table 8. Win-Tie-Loss Analysis

Comparison	Win	Tie	Loss	Win Rate
IHHO vs Baseline	32	0	0	100%
IHHO vs HHO	32	0	0	100%
HHO vs Baseline	32	0	0	100%

4. Discussion and Conclusion

The present study evaluated an Improved Harris Hawks Optimization algorithm as a feature-selection mechanism for diabetes diagnosis and compared its performance with both the standard Harris Hawks Optimization algorithm and a baseline condition without feature selection. The findings consistently demonstrated the superiority of the proposed IHHO framework across all eight classifiers and all four evaluation criteria. In the baseline condition, classification accuracy was approximately 72.55%–78.55%, indicating that direct use of the complete feature set provided only moderate discrimination between diabetic and non-diabetic cases. Applying standard HHO-based feature selection increased classifier accuracy to approximately 78.55%–87.37%, showing that eliminating less informative attributes improved model performance. The IHHO procedure produced a substantially greater increase, with classifier accuracies ranging from 92.96% to 97.20%. On average, IHHO increased accuracy by approximately 18%–22% relative to the baseline and by approximately 9%–17% relative to standard HHO. This general pattern is consistent with reviews concluding that diabetes-prediction performance depends not only on classifier selection but also on preprocessing, feature optimization, and the capacity of the computational framework to identify stable diagnostic patterns [6, 35].

The strongest performance was obtained through the combination of IHHO and Random Forest, which achieved an accuracy of 97.20%, precision of 94.51%, recall of 98.33%, and F1-score of 96.38%. These values indicate that the proposed model correctly classified most observations while simultaneously maintaining a strong balance between identifying diabetic patients and avoiding incorrect positive predictions. The particularly high recall is clinically important because it suggests that very few diabetic cases were incorrectly classified as non-diabetic. Random Forest may have benefited more than the other classifiers from IHHO feature selection because it can model nonlinear interactions, threshold effects, and conditional relationships among metabolic indicators through ensembles of diverse decision trees. Previous diabetes studies have similarly reported favorable performance for Random Forest, particularly after preprocessing and optimization of clinical variables [11, 13]. The present result is also consistent with evidence that ensemble architectures can improve diagnostic stability by combining complementary patterns from multiple learners [14]. Evolutionary stacking models have likewise demonstrated that the integration of optimization and ensemble learning can produce more balanced diabetes-diagnosis performance than the isolated use of individual classifiers [45]. The effectiveness of optimized hybrid classification reported by Shimpi et al. further supports the finding that diagnostic performance can be strengthened when feature optimization and robust classifiers are incorporated into a unified framework [47].

Although the IHHO–RF combination produced the highest accuracy, the benefit of IHHO was not limited to a single classifier. KNN and Decision Tree achieved accuracies of 96.54% and 96.48%, respectively, while the lowest accuracy among the IHHO-based models was still 92.96%. The SVM and ANN models showed particularly large increases in recall, improving by approximately 26% and 23%, respectively, relative to their baseline performance. The F1-score also improved consistently, demonstrating that the increase in sensitivity was not achieved by indiscriminately labeling observations as diabetic. Instead, IHHO appears to have selected a feature subset that supported a more effective balance between precision and recall across heterogeneous learning paradigms. This classifier-independent improvement supports the model-agnostic feature-selection perspective advanced by Natarajan et al., in which a useful selection strategy should retain its value across different predictive algorithms [27]. The result also aligns with hybrid feature-selection and logistic-regression research showing that carefully selected clinical attributes can increase accuracy without requiring an excessively complex final classifier [28]. Metaheuristic feature selection has similarly been found to improve diabetes prediction by searching for combinations of features that cannot be identified adequately through simple ranking procedures [29]. The observed gains are also compatible with simulation-optimization approaches to feature selection, which treat variable

selection as an interdependent search problem rather than evaluating each variable separately [31]. More broadly, systematic evidence indicates that effective feature selection can reduce noise, redundancy, computational burden, and overfitting in high-dimensional classification tasks [32]. Although feature extraction may also reduce dimensionality, retaining original variables through feature selection can preserve their clinical meaning and facilitate interpretation [48].

The superiority of IHHO over standard HHO can be explained by the complementary effects of chaotic initialization, modification of the prey-energy function, and adaptive Lévy-flight search. In the standard HHO algorithm, random initialization may produce uneven coverage of the search space, causing candidate solutions to concentrate prematurely within limited regions. The chaotic initialization used in IHHO was intended to distribute initial hawk positions more broadly and thereby increase population diversity. This interpretation is consistent with research showing that improved HHO variants can overcome premature convergence and increase optimization accuracy by modifying population dynamics and search behavior [41]. The successful use of HHO to optimize vision-transformer models for diabetic-retinopathy detection also demonstrates the suitability of Harris-hawk-based mechanisms for complex medical optimization tasks [44]. Similarly, composite chaotic mappings have been shown to increase the search diversity and convergence properties of arithmetic optimization algorithms [42], while crossover and random-walk strategies have strengthened the exploration and exploitation capacities of other population-based optimizers [43]. Bibliographic evidence concerning chaotic metaheuristics further indicates that chaotic sequences can provide more structured and diverse search trajectories than conventional pseudorandom initialization [40].

The modified prey-energy function may also have contributed to the observed performance by creating a smoother transition from global exploration to local exploitation. A linear reduction in prey energy can cause the standard algorithm to shift too rapidly toward local search, thereby increasing the risk of convergence around a suboptimal feature subset. The exponential energy mechanism employed in IHHO maintained greater search flexibility during intermediate iterations, while adaptive Lévy flights enabled occasional long-distance movements that facilitated escape from local optima. These mechanisms address central challenges identified in comprehensive reviews of metaheuristic optimization, including population stagnation, insufficient diversity, parameter sensitivity, and an unstable balance between exploration and exploitation [36]. Recent evaluations of newly developed metaheuristics similarly emphasize that performance gains require meaningful search modifications rather than merely introducing new biological metaphors [37]. The present findings are also compatible with evidence supporting metaheuristic optimization for chronic-disease classification, particularly where feature selection and classifier performance must be optimized simultaneously [38]. Studies of swarm-intelligence-based neural models have further shown that medical classification can benefit from optimization mechanisms capable of navigating nonlinear and multimodal search spaces [39].

The high recall achieved by the proposed framework may additionally reflect the combined effects of class balancing and optimized feature selection. Class imbalance is a major problem in diabetes datasets because a model can produce acceptable overall accuracy by favoring the majority non-diabetic class while failing to identify a clinically unacceptable proportion of diabetic patients. The use of SMOTE in the preprocessing phase increased the representation of minority-class observations, while IHHO subsequently selected variables that maximized diagnostic discrimination. This interpretation agrees with research demonstrating that appropriate imbalance-handling methodologies can improve diabetes-prediction sensitivity and prevent majority-class bias [21]. Novel synthetic oversampling methods have likewise been developed because the quality and distribution of synthetic observations can materially influence classifier performance [23]. Attention-enhanced deep belief networks have produced improved early diabetes-risk predictions under highly imbalanced conditions, further confirming the importance of explicitly addressing minority-class detection [22]. Enhanced machine-learning systems that combine imbalance-handling procedures with predictive modeling have also reported improved diabetes classification [24]. The present model's recall of 98.33% is particularly aligned with the objective of reducing false negatives, which has been identified as a critical requirement in diabetes-detection systems [25]. Research using symptoms extracted from medical records similarly suggests that preprocessing and representation of heterogeneous patient information are essential for reliable disease classification [26].

The proposed IHHO–RF model also compared favorably with the methods included in the manuscript's state-of-the-art analysis. Its accuracy of 97.20% was slightly higher than

the closest reported accuracy of 97.06%, while its precision was considerably more balanced than that of some comparison models. Compared with approaches reporting accuracies between approximately 80% and 93%, the proposed framework achieved substantial absolute gains. Its F1-score of 96.38% further indicated that performance was not driven solely by one evaluation metric. These findings are consistent with diabetes-prediction studies using the Pima Indians dataset, which have demonstrated that machine-learning performance improves when data preparation and model optimization are incorporated systematically [12]. Earlier data-mining models also established that meaningful diagnostic patterns could be extracted from standard diabetes attributes, although their performance was dependent on the selected algorithms and preprocessing decisions [9]. Machine-learning classification using broader demographic and health-survey information has similarly confirmed the capacity of computational models to detect diabetes from combinations of patient-level variables [10]. The results therefore extend previous work by indicating that optimization of the feature space can be at least as consequential as increasing classifier complexity. They also complement research on diabetic-patient readmission, where integrating data-mining algorithms with metaheuristic optimization improved clinically relevant risk prediction [46].

The statistical analyses provide further support for the substantive importance of the observed improvements. The one-way ANOVA yielded an F-value of 89.42 with a p-value below 0.0001, confirming that the mean performance differences among the baseline, HHO, and IHHO conditions were highly significant. The Tukey HSD analysis demonstrated significant differences for all pairwise comparisons. The mean difference between IHHO and baseline was 19.21 percentage points, while the mean difference between IHHO and HHO was 11.47 points; standard HHO also exceeded the baseline by 7.74 points. The Friedman tests remained significant for accuracy, precision, recall, and F1-score, with chi-square values ranging from 14.86 to 16.00 and p-values below 0.0001. Finally, the win–tie–loss analysis showed that IHHO achieved 32 wins, no ties, and no losses against both baseline and HHO across eight classifiers and four metrics. Thus, the advantage of IHHO was not confined to an isolated classifier, metric, or experimental comparison. Rather, the convergence of parametric, non-parametric, and distribution-independent analyses indicates a stable and systematic performance improvement.

The results also have broader implications for intelligent diabetes-management systems. Machine learning is increasingly being used for early chronic-disease detection, personalized risk evaluation, and data-driven clinical decision-making [5]. Unified prediction systems demonstrate the growing interest in frameworks that can apply computational learning across multiple diseases and healthcare settings [7]. At the same time, diabetes-monitoring systems are moving toward edge- and cloud-based environments that require accurate models with manageable computational demands [8]. Because diabetes imposes substantial clinical and economic burdens, even modest improvements in screening sensitivity and timely intervention may have meaningful consequences for patient outcomes and healthcare expenditure [1, 2]. Data analysis and technology have therefore become central to diabetes management [3], while big-data analytics has expanded the possibilities for integrating clinical information into healthcare decisions [4].

Nevertheless, predictive performance alone is not sufficient for clinical adoption. The use of eight classifiers in the present study helped demonstrate the generality of the selected feature subset. The strong KNN performance is noteworthy because distance-based models are particularly sensitive to irrelevant dimensions, and reducing such dimensions can substantially improve neighborhood quality [33]. Improvements in Naive Bayes performance are also compatible with evidence that probabilistic classifiers can benefit from procedures designed to enrich both recall and precision [34]. The favorable neural-network findings align with studies applying multi-layer perceptrons, hybrid convolutional networks, and deep architectures to diabetes prediction [15-17]. However, practical systems should also explain their predictions. Explainable gated networks and self-explainable diagnostic interfaces demonstrate the importance of showing clinicians how individual variables influence diabetes-risk estimates [18, 19]. Patient confidentiality must also be protected, as emphasized by privacy-conscious diabetes-prediction frameworks [20]. Federated learning may provide a suitable path for extending IHHO-based prediction across institutions without centralizing raw patient data [30]. Overall, the findings indicate that the proposed IHHO framework can improve diabetes classification by identifying a diagnostically effective feature space, but its transition from benchmark evaluation to clinical decision support will require external validation, interpretability, privacy protection, and workflow integration.

This study has several limitations that should be considered when interpreting its findings. First, the evaluation was conducted using a single benchmark dataset composed of a relatively restricted demographic population; therefore, the results cannot yet be generalized to men, younger individuals, different ethnic groups, or patients treated in diverse healthcare systems. Second, although cross-validation and statistical comparisons strengthened the internal evaluation, independent external validation was not performed. Third, the reported analysis focused primarily on accuracy, precision, recall, and F1-score, while other clinically relevant measures, including specificity, negative predictive value, area under the receiver-operating-characteristic curve, calibration, and decision-curve utility, were not fully examined. The study also did not report the stability with which individual features were selected across repeated optimization runs. Finally, the additional chaotic, energy-control, and Lévy-flight mechanisms increase computational complexity compared with the baseline and standard HHO procedures.

Future studies should validate the proposed framework using larger, multicenter, longitudinal, and demographically heterogeneous datasets obtained from real clinical environments. External validation should include direct comparisons across healthcare institutions and prospective assessment of newly screened patients. Further research should examine feature-selection stability, convergence behavior, execution time, memory requirements, and sensitivity to optimization parameters through comprehensive ablation studies. The framework could also be extended to distinguish among non-diabetic, prediabetic, and diabetic cases and to predict complications, treatment responses, or hospital readmission. Additional evaluations should include calibration analysis, specificity, negative predictive value, receiver-operating-characteristic curves, decision-curve analysis, and cost-sensitive clinical outcomes. Privacy-preserving federated versions, explainable prediction modules, adaptive online-learning models, and integration with deep or transformer-based architectures also merit investigation.

In practical settings, the proposed framework should initially be implemented as a clinical decision-support or screening aid rather than an autonomous diagnostic system. Healthcare organizations should test the model with local patient data, establish clinically appropriate probability thresholds, and verify its calibration before routine use. Predictions should be presented together with the selected clinical variables and an interpretable explanation so that clinicians can review the basis of each risk estimate. Institutions should also establish procedures for data-quality assessment, missing-value management, model monitoring, privacy protection, and periodic retraining. Particular attention should be paid to false-negative cases because missed diabetes can delay intervention, while false-positive predictions should trigger confirmatory laboratory testing rather than immediate diagnostic labeling.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

References

[1] A. A. Yameny, "Diabetes Mellitus Overview 2024," *Journal of Bioscience and Applied Research,* vol. 10, no. 3, pp. 641-645, 2024.

[2] E. D. Parker *et al.*, "Economic Costs of Diabetes in the United States in 2022," *Diabetes Care,* vol. 47, no. 1, pp. 26-43, 2024.

[3] U. K. Rangaswamy, C. Ratan, A. A. Anil, D. Rajesh, and A. Laxmi Krishnan, "Data Analysis in Diabetes and the Role of Technology in Diabetes Management," in *Diabetes Mellitus*: Academic Press, 2025, pp. 239-272.

[4] M. Nauman, A. S. Almadhor, M. Albekairi, A. R. Ansari, M. A. Fayyaz, and R. Nawaz, "The Role of Big Data Analytics in Revolutionizing Diabetes Management and Healthcare Decision-Making," *IEEE Access,* 2025.

[5] S. Padmalal, P. Arumugam, M. B. Anusha, K. D. Bamane, N. Yamsani, and K. Muppavaram, "Machine Learning for Early Detection of Chronic Diseases: A Case Study in Diabetes Prediction," in *Machine Learning in Healthcare: Data-Driven Decisions, Predictive Modelling, Personalized Medicine*, 2026, p. 171.

[6] P. Nayak, J. S. N. Jyothi, V. Harika, K. Swaraja, and A. S. Hanuman, "Diabetes Monitoring and Prediction Using Computational Intelligence Techniques: A Systematic Review," *SN Computer Science,* vol. 6, no. 3, p. 267, 2025.

[7] S. Verma, R. Rallan, T. Singh, and N. Kumar, "Unified Disease Prediction System Using Machine Learning Techniques," presented at the 2025 International Conference on Automation and Computation, 2025/03, 2025.
[8] D. Kalita, H. Sharma, and K. B. Mirza, "Deep Learning on Edge and Cloud for Intelligent Monitoring and Decision Making in Artificial Pancreas," *Expert Systems with Applications,* vol. 295, p. 128833, 2026.
[9] R. Rastogi and M. Bansal, "Diabetes Prediction Model Using Data Mining Techniques," *Measurement: Sensors,* vol. 25, p. 100605, 2023.
[10] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetes Disease Detection and Classification on Indian Demographic and Health Survey Data Using Machine Learning Methods," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews,* vol. 17, no. 1, p. 102690, 2023.
[11] S. K. S. Modak and V. K. Jha, "Diabetes Prediction Model Using Machine Learning Techniques," *Multimedia Tools and Applications,* vol. 83, no. 13, pp. 38523-38549, 2024.
[12] M. S. Salih, R. K. Ibrahim, S. R. Zeebaree, D. Asaad, L. M. Zebari, and N. M. Abdulkareem, "Diabetic Prediction Based on Machine Learning Using PIMA Indian Dataset," *Communications on Applied Nonlinear Analysis,* vol. 31, no. 5s, pp. 138-156, 2024.
[13] M. R. Hossain, M. J. Hossain, M. M. Rahman, and M. M. Alam, "Machine Learning-Based Prediction and Insights of Diabetes Disease: Pima Indian and Frankfurt Datasets," *Journal of Mechanics of Continua and Mathematical Sciences,* vol. 20, no. 1, pp. 99-114, 2025.
[14] D. Alfredo, C. S. P. Fidel, and A. Gonzalo, "Stacking Ensemble Approach to Diagnosing the Disease of Diabetes," *Informatics in Medicine Unlocked,* vol. 44, p. 101427, 2024.
[15] Z. Zhang *et al.*, "A Novel Evolutionary Ensemble Prediction Model Using Harmony Search and Stacking for Diabetes Diagnosis," *Journal of King Saud University—Computer and Information Sciences,* vol. 36, no. 1, p. 101873, 2024.
[16] M. Scholar, A. S. Singh, and K. K. Agrawal, "Optimizing Diabetes Prediction with Multi-Layer Perceptron: A Comprehensive Analysis on the Pima Indians Dataset," presented at the 2025 3rd International Conference on Communication, Security, and Artificial Intelligence, 2025/04, 2025.
[17] K. S. Farsana and A. Poulose, "Hybrid Convolutional Neural Networks for Pima Indians Diabetes Prediction," presented at the 2024 Fifteenth International Conference on Ubiquitous and Future Networks, 2024/07, 2024.
[18] H. Khan, N. Javaid, T. Bashir, Z. Ali, F. A. Khan, and D. Pamucar, "A Novel Deep Gated Network Model for Explainable Diabetes Mellitus Prediction at Early Stages," *Knowledge-Based Systems,* p. 114178, 2025.
[19] G. Dharmarathne, T. N. Jayasinghe, M. Bogahawaththa, D. P. P. Meddage, and U. Rathnayake, "A Novel Machine Learning Approach for Diagnosing Diabetes with a Self-Explainable Interface," *Healthcare Analytics,* vol. 5, p. 100301, 2024.
[20] I. Shaheen, N. Javaid, Z. Ali, I. Ahmed, F. A. Khan, and D. Pamucar, "A Trustworthy and Patient Privacy-Conscious Framework for Early Diabetes Prediction Using Deep Residual Networks and Proximity-Based Data," *Biomedical Signal Processing and Control,* vol. 112, p. 108361, 2026.
[21] B. Toleva, I. Atanasov, I. Ivanov, and V. Hooper, "An Effective Methodology for Diabetes Prediction in the Case of Class Imbalance," *Bioengineering,* vol. 12, no. 1, p. 35, 2025.
[22] O. Olabanjo, A. Wusu, O. Olabanjo, M. Asokere, O. Afisi, and B. Akinnuwesi, "A Novel Deep Learning Model for Early Diabetes Risk Prediction Using Attention-Enhanced Deep Belief Networks with Highly Imbalanced Data," *International Journal of Information Technology,* pp. 1-23, 2025.
[23] J. Wang and N. Awang, "A Novel Synthetic Minority Oversampling Technique for Multiclass Imbalance Problems," *IEEE Access,* 2025.
[24] I. Abousaber, "Enhanced Diabetes Prediction through Advanced Machine Learning and Imbalance Handling Techniques," presented at the 2025 4th International Conference on Computing and Information Technology, 2025/04, 2025.
[25] M. A. Uddin *et al.*, "Machine Learning-Based Diabetes Detection Model for False Negative Reduction," *Biomedical Materials & Devices,* vol. 2, no. 1, pp. 427-443, 2024.
[26] N. Mumtazah, I. Waspada, K. D. Mutiara, and S. Adhy, "A Combination of Data Mining Methods for Disease Classification Using Patient-Perceived Symptoms from Medical Records," presented at the 2024 7th International Conference on Informatics and Computational Sciences, 2024/07, 2024.
[27] K. Natarajan, D. Baskaran, and S. Kamalanathan, "An Adaptive Ensemble Feature Selection Technique for Model-Agnostic Diabetes Prediction," *Scientific Reports,* vol. 15, no. 1, p. 6907, 2025.
[28] S. P. M. Kodete, S. H. Palli, J. Kola, S. P. Pasupuleti, G. Ravi, and M. H. Prasad, "A Hybrid Feature Selection and Logistic Regression Approach for High-Accuracy Early Diabetes Detection," presented at the 2025 5th International Conference on Soft Computing for Security Applications, 2025/08, 2025.
[29] M. Karuppasamy and K. Poorani, "Metaheuristic Feature Selection for Diabetes Prediction with PGS Approach," *Procedia Computer Science,* vol. 252, pp. 165-171, 2025.
[30] R. Kapila and S. Saleti, "Federated Learning-Based Disease Prediction: A Fusion Approach with Feature Selection and Extraction," *Biomedical Signal Processing and Control,* vol. 100, p. 106961, 2025.
[31] S. Shashaani and K. Vahdat, "Improved Feature Selection with Simulation Optimization," *Optimization and Engineering,* vol. 24, no. 2, pp. 1183-1223, 2023.
[32] A. Das, S. Mollick, H. Mondal, T. J. Bala, A. Nag, and A. Guho, "Meta-Heuristic Algorithms for High-Dimensional Feature Selection: A Systematic Review of Methodologies, Applications, and Emerging Challenges with Future Research Directions," in *Feature Fusion for Next-Generation AI*, 2026, pp. 51-61.
[33] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review and Performance Analysis of Modifications," *Journal of Big Data,* vol. 11, no. 1, p. 113, 2024.
[34] O. Peretz, M. Koren, and O. Koren, "Naive Bayes Classifier—An Ensemble Procedure for Recall and Precision Enrichment," *Engineering Applications of Artificial Intelligence,* vol. 136, p. 108972, 2024.
[35] S. K. S. Modak and V. K. Jha, "Machine and Deep Learning Techniques for the Prediction of Diabetics: A Review," *Multimedia Tools and Applications,* vol. 84, no. 18, pp. 19425-19549, 2025.
[36] K. Rajwar, K. Deep, and S. Das, "An Exhaustive Review of the Metaheuristic Algorithms for Search and Optimization: Taxonomy, Applications, and Open Challenges," *Artificial Intelligence Review,* vol. 56, no. 11, pp. 13187-13257, 2023.
[37] M. S. Shaikh *et al.*, "Applications, Classifications, and Challenges: A Comprehensive Evaluation of Recently Developed Metaheuristics for Search and Analysis," *Artificial Intelligence Review,* vol. 58, no. 12, pp. 1-110, 2025.

[38] A. Singh, N. Prakash, and A. Jain, "Meta-Heuristic Optimization for the Multi-Classification of Chronic Disease: A Review with Machine Learning Perspectives," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 15, no. 3, p. e70030, 2025.
[39] K. Veeranjaneyulu, M. Lakshmi, and S. Janakiraman, "Swarm Intelligent Metaheuristic Optimization Algorithms-Based Artificial Neural Network Models for Breast Cancer Diagnosis: Emerging Trends, Challenges and Future Research Directions," *Archives of Computational Methods in Engineering,* vol. 32, no. 1, pp. 381-398, 2025.
[40] M. Pluhacek, A. Kazikova, A. Viktorin, T. Kadavy, and R. Senkerik, "Chaos in Popular Metaheuristic Optimizers—A Bibliographic Analysis," in *Recent Improvements in the Theory of Chaotic Attractors*: CRC Press, 2025, pp. 106-121.
[41] D. T. Akl, M. M. Saafan, A. Y. Haikal, and E. M. El-Gendy, "IHHO: An Improved Harris Hawks Optimization Algorithm for Solving Engineering Problems," *Neural Computing and Applications,* vol. 36, no. 20, pp. 12185-12298, 2024.
[42] Y. X. Li, J. S. Wang, S. W. Zhang, S. H. Zhang, X. Y. Guan, and X. R. Ma, "Arithmetic Optimization Algorithm with Cosine Transform-Based Two-Dimensional Composite Chaotic Mapping," *Soft Computing,* pp. 1-41, 2025.
[43] Y. L. Qi, C. Xing, J. S. Wang, Y. W. Song, and X. Y. Guan, "Improved Goose Algorithm Based on Vertical and Horizontal Crossover Strategy and Random Walk to Solve Engineering Optimization Problems," *Engineering Letters,* vol. 33, no. 5, 2025.
[44] V. Awasthi *et al.*, "ViT-HHO: Optimized Vision Transformer for Diabetic Retinopathy Detection Using Harris Hawk Optimization," *MethodsX,* vol. 13, p. 103018, 2024.
[45] Z. Zhang, K. A. Ahmed, M. R. Hasan, T. Gedeon, and M. Z. Hossain, "A Deep Learning Approach to Diabetes Diagnosis," presented at the Asian Conference on Intelligent Information and Database Systems, 2024/04, 2024.
[46] M. Zeinalnezhad and S. Shishehchi, "An Integrated Data Mining Algorithms and Metaheuristic Technique to Predict the Readmission Risk of Diabetic Patients," *Healthcare Analytics,* vol. 5, p. 100292, 2024.
[47] J. K. Shimpi, P. Shanmugam, and A. A. Stonier, "Analytical Model to Predict Diabetic Patients Using an Optimized Hybrid Classifier," *Soft Computing,* vol. 28, no. 3, pp. 1883-1892, 2024.
[48] J. Li, M. S. Othman, H. Chen, and L. M. Yusuf, "Optimizing IoT Intrusion Detection System: Feature Selection versus Feature Extraction in Machine Learning," *Journal of Big Data,* vol. 11, no. 1, p. 36, 2024.
[49] V. Ambikavathi, P. Arumugam, and P. Jose, "Diabetes Detection by Data Mining Methods," *Wireless Personal Communications,* vol. 133, no. 4, pp. 2087-2104, 2023.
[50] F. A. Jaber and J. W. James, "Early Prediction of Diabetic Using Data Mining," *SN Computer Science,* vol. 4, no. 2, p. 169, 2023.
[51] X. Li, J. Zhang, and F. Safara, "Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm," *Neural Processing Letters,* vol. 55, no. 1, pp. 153-169, 2023.