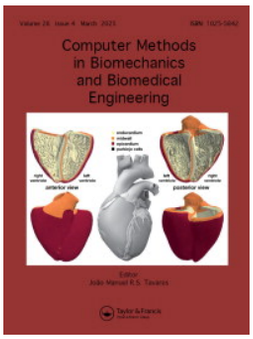


Computer Methods in Biomechanics and Biomedical Engineering

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gcmb20

An improved decision tree model through hyperparameter optimization using a modified gray wolf optimization for diabetes classification

Muhammad Sam'an, Farikhin & Muhammad Munsarif

To cite this article: Muhammad Sam'an, Farikhin & Muhammad Munsarif (13 Feb 2025): An improved decision tree model through hyperparameter optimization using a modified gray wolf optimization for diabetes classification, Computer Methods in Biomechanics and Biomedical Engineering, DOI: 10.1080/10255842.2025.2460178

To link to this article: https://doi.org/10.1080/10255842.2025.2460178

Published online: 13 Feb 2025.

Submit your article to this journal

Article views: 93

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gcmb20

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING
https://doi.org/10.1080/10255842.2025.2460178

Taylor & Francis
Taylor & Francis Group

Check for updates

An improved decision tree model through hyperparameter optimization using a modified gray wolf optimization for diabetes classification

Muhammad Sam'an[a,b], Farikhin[c] and Muhammad Munsarif[a]

[a]Department of Informatics, Universitas Muhammadiyah Semarang, Semarang, Indonesia; [b]Faculty of Business, Management and Information Technology, Universiti Muhammadiyah Malaysia, Malaysia; [c]Department of Mathematics, Universitas Diponegoro, Semarang, Indonesia

ABSTRACT

Diabetes is a chronic condition that affects blood sugar levels and vital organs in the body. Early detection is crucial given the increasing global prevalence of diabetes and the grave risk of complications if not properly managed. Thus, a good prediction system is necessary. Although the Decision Tree (DT) is commonly used for classification, it is less effective with large datasets. We propose hyperparameter optimization of the DT using the Grey Wolf Optimization (GWO), which has exploration and both exploitation capabilities. However, the limited search space of GWO may hinder practical exploration and exploitation, leading to premature optimization. To address this, we propose a modified GWO (MGWO) by adding the Levy distribution function to enhance the movements of alpha, beta, and delta wolves. We also provide GA (Genetic Algorithm) as a comparative algorithm. The fitness value of MGWO is 0.8498, surpassing GWO (0.8373) and GA (0.8492). Evaluation results indicate that MGWO and GA yield similar and superior accuracy compared to GWO. The proposed method outperforms existing ones. Further research is needed to evaluate the impact of varying the number of wolves on optimization performance and classification accuracy.

ARTICLE HISTORY
Received 17 March 2024
Accepted 14 January 2025

KEYWORDS
Diabetes; decision tree; hyperparameter optimization; grey wolf optimization; classification

1. Introduction

Diabetes is a chronic condition that affects blood sugar levels and can harm vital organs in the body. There are two main types, Type 1 and Type 2, which affect adults differently. Type 1 diabetes occurs when the body cannot produce enough insulin, whereas type 2 diabetes arises from the body's resistance to insulin. Effective diabetes management, including insulin usage, is vital for the 422 million people worldwide affected by diabetes, which causes 1.5 million deaths annually. Diabetes manifests with symptoms like extreme thirst, increased appetite, frequent need to urinate, feeling unwell, and slow healing of cuts. If not addressed promptly, diabetes can escalate into a severe illness. In 2017, 451 million people globally had diabetes, and this number is expected to rise to 693 million by 2045. Studies by Habibi et al. (2015); Yaribeygi et al. (2016); Gosh (2017) indicate that health concerns increase when diagnosis is delayed. Guidelines for preventing and reducing the risk of diabetes emphasize the importance of early detection (Zhu et al. 2015; Woldesemayat 2019). Although we focus on preventing diabetes, current treatment methods are insufficient. Therefore, accurate prediction is essential for early detection, representing a significant advancement in medical care.

Machine Learning (ML) models are considered a solution for diabetes prediction. For instance, the diabetes prediction system, based on IoT-edge Artificial Intelligence (AI)-blockchain and using the Random Forest (RF) algorithm, shows an average accuracy that is 4.57% higher than Logistic Regression (LR) and Support Vector Machine (SVM) (Hennebelle et al. 2024). Combining data mining and metaheuristic techniques for predicting the early readmission probability of diabetic patients showed that RF, GA-SVM, SVM, and Neural Network models have accuracies of 74.04%, 73.52%, 72.40%, and 70.44%, respectively, with a 1.12% increase in SVM accuracy through Genetic Algorithms (GA) (Zeinalnezhad and Shishehchi 2024). Hybrid ML and interaction features in the Feature Interaction-based Greedy Sequential Feature Selection to predict type 2 diabetes and prediabetes resulted in a model with high accuracy (98.87% for Type 2 diabetes and 90.12% for prediabetes) and good interpretability. ML models with PCA and ensemble bagging decision tree classification accurately predicted early-stage diabetes

CONTACT Muhammad Sam'an muhammad92sam@unimus.ac.id
© 2025 Informa UK Limited, trading as Taylor & Francis Group

mellitus (Nagpal et al. 2023). AHDHS-Stacking showed effectiveness in early diabetes prediction and feature screening on the PIMA Indians Diabetes dataset (Zhang et al. 2024). The XGBoost model predicted insulin adherence in newly initiated users with Type 2 diabetes mellitus (Chen et al. 2023). Other studies focused on predicting diabetic retinopathy, osteoporosis risk, and gestational diabetes mellitus using various ML models such as XGBoost, NGBoost, EBM, LASSO, and CatBoost, each showing promising results in terms of accuracy and predictive performance (Chan et al. 2023; Li et al. 2023; Wu et al. 2023; Yagin et al. 2023; Zhang et al. 2023; Zhou et al. 2023).

The LR, extra tree classifier, RF, gradient boosting decision tree (GBDT), and XGBoost were evaluated for predicting the risk of end-stage renal disease in newly diagnosed Type 2 diabetes mellitus patients, with XGBoost showing the highest Area Under Curve (AUC) of 0.953 among the models. The study used various sensors, including glucose, electrocardiogram, and accelerometer, using the XGBoost algorithm to achieve the highest prediction accuracy of 98.2% for diabetes, with a 4%–5% increase compared to single sensors (Site et al. 2023). While these models show promising results, consistently achieving accuracy rates above 99% remains challenging due to the complex nature of the disease and the many influencing factors. Ongoing research aims to refine and enhance these ML techniques, emphasizing continuous evaluation to ensure their effectiveness in real-world healthcare scenarios.

Decision Trees (DT), widely used for classification tasks, offer advantages such as computational efficiency, interpretability, and simplicity in understanding variable relationships for predictions (Parhi and Patro 2024). Their ability to provide clear and understandable decision rules makes them highly interpretable compared to more complex models like neural networks or ensemble methods. This interpretability is particularly valuable in fields like healthcare, where understanding the reasoning behind predictions is crucial (Lu et al. 2023). Additionally, The DT is computationally efficient and can handle both numerical and categorical data, making them versatile for various types of datasets (Yaqoob et al. 2023).

Despite their ease of implementation without requiring extensive technical expertise, The DT face limitations such as over-fitting, challenges with imbalanced datasets, susceptibility to noise, and specificity to training data (Talebi et al. 2024), which can potentially lead to decreased performance when confronted with unseen data and lower accuracy compared to some other prediction models (Ul Hassan et al. 2023). Various studies have explored strategies to enhance the DT performance, mainly through hyperparameter optimization. Notable improvements have been observed in C4.5, CTree, and CART with hyperparameter tuning, addressing limitations and achieving better performance (Mantovani et al. 2018). Statistical significance in the sensitivity of J48 has been demonstrated across 102 datasets (Mantovani et al. 2017). The Iterative DT efficiently optimizes hyperparameters, showcasing competitive and stable performance (Saum et al. 2022). Moth-Flame hyperparameter optimization has proven effective in identifying failure modes in steel plate corrosion (Tran and Kim 2024). Additionally, diverse tree-based algorithms, including CTree, C4.5, Bagged CART, and RF, have been employed for hyperparameter optimization to explore relationships between experimental peptide affinities and virtual docking data (Feng et al. 2023). Hyperparameter optimization using RF for predicting solute transport in heterogeneous sandstone has enhanced the accuracy of breakthrough curve predictions (Perez et al. 2023). Hyperparameter optimization strategies for DT involve various optimization algorithms such as GA (Guido et al. 2023; Zeinalnezhad and Shishehchi 2024), Bayesian Optimization (Kadam and Jadhav 2020; Kumar et al. 2022; Albahli 2023), Grid Search (Habib and Khursheed 2022; Almutairi 2023), Response Surface Method (Moradi et al. 2020), Particle Swarm Optimization (Tian et al. 2022), harmony search (Zhang et al. 2024), and Firefly Algorithm (Qi and Tang 2018). The existing works collectively highlight that hyperparameter optimization for DT models is a challenging task, requiring diverse optimization algorithms to make DT superior in classification tasks.

Mirjalili et al. (2014) introduced the Grey Wolf Optimizer (GWO) as a promising algorithm for addressing various standard optimization problems by mimicking grey wolves' social hierarchy and hunting abilities. GWO has been successfully applied to feature selection in different scenarios (Emary et al. 2015; Al-Tashi et al. 2020, Almutairi 2023; Alyasseri et al. 2022; Thaher et al. 2022; Albahli 2023; Almazini et al. 2023; Cinar 2023; Hou et al. 2023; Jain et al. 2023; Lin et al. 2023; Yuvaraja et al. 2023) and hyperparameter optimization in Deep Learning, including Convolutional Neural Networks (Challapalli and Devarakonda 2022; Cuong-Le et al. 2022; Kuyu and Ozekmekci 2022; Mohakud and Dash 2022; Priyanka and Kumar 2022; Xu et al. 2023). Although the GWO algorithm performs well for hyperparameter optimization, it is limited in exploring the solution space,

which is crucial for finding the best solution among a wide range of hyperparameters. This limitation can hinder the algorithm's effectiveness in complex optimization tasks, as it may struggle to escape local optima and fully explore the global search space (Liu et al. 2024). Therefore, we propose a Modified GWO (MGWO) by adding a Levy distribution function to enhance the movements of alpha, beta, and delta wolves, thus improving the algorithm's ability to explore the solution space. We chose the Levy distribution for its qualities that can help handle complex and dynamic search spaces, allowing the algorithm to cover a larger area more effectively and increasing the likelihood of finding a global optimum. This modification aims to address the inherent limitations of the original GWO and enhance its performance in hyperparameter optimization tasks (Lei et al. 2023).

The key contributions of this paper are follows:

- To improve the accuracy of diabetes prediction through hyperparameter optimization in the DT model using GWO and MGWO
- To enhance the algorithms exploration capabilities by incorporating levy distribution function in improving the steps of alpha, beta and delta wolves

The rest of this paper is organized as follows. Section 2 describes the methods. The detailed proposed model is also discussed in Section 2. Section 3 discusses the experimental results and their analysis. Finally, the conclusion and future scope are outlined in Section 4.

2. Materials and methods

Figure 1 illustrates the workflow of the proposed model for predicting diabetes through hyperparameter optimization of DT using GWO and MGWO.

2.1. Data source

We accessed publicly available data from The Behavioral Risk Factor Surveillance System (BRFSS) for 2015, which includes 21 feature variables for 253,680 subjects. From these records, 160,029 respondents were diagnosed as normal, 19,681 were diagnosed with pre-diabetes, and 2,812 had been diagnosed with diabetes. Most of the variables are related to chronic health conditions other than diabetes, such as cancer and asthma. Diabetes disrupts glucose regulation, either by insufficient insulin production or ineffective use of insulin, leading to complications such as heart disease, vision loss, and kidney disease. While there is no cure, lifestyle changes like weight loss, healthy eating, and physical activity, along with medical treatments, can help manage the disease. Early diagnosis is important, allowing for proactive management and reducing long-term complications. The Diabetes Health Indicators Dataset from Kaggle (Teboul 2021), derived from the BRFSS 2015 data, will be utilized. This dataset has 253,680 survey responses that have been cleaned and balanced, providing a comprehensive overview of various health indicators relevant to diabetes. The characteristics of the BRFSS dataset are presented in Table 1.

2.2. Data preprocessing

We used exploratory data analysis (EDA) in the data pre-processing. EDA involves calculating the extent of missing data, means, variances, medians, quartiles, and other relevant statistical measures for each feature. The goal was to thoroughly understand how these features are distributed throughout the original dataset. As shown in Figure 2, the BRFSS dataset has no missing values for any of its attributes. After that, we identified outliers in each attribute using the Interquartile Range (IQR) value, calculated with the formula in Equation (5). Q_1 represents the lower quartile (25th percentile) and Q_3 represents the upper quartile (75th percentile).

$$IQR = Q_3 - Q_1 \tag{1}$$

As shown in Figure 3, there are many outliers in the 22 continuous attributes within the BRFSS dataset. These values might make the information less accurate and lower the model's performance. To tackle this problem, we remove the values identified as outliers. We do this to reduce the chances of the models struggling to achieve accurate results due to statistical errors. Because each feature has a different scale, the direct adjustment process presents some specific challenges. Next, we standardize the values of each feature from 0 to 1 using the Z-score normalization technique as formulated in Equation (6), where μ represents the mean and σ represents the standard deviation.

$$Z - score = \frac{x - \mu}{\sigma} \tag{2}$$

2.3. Data splitting: k-fold cross validation

K-Fold validation with $k = 10$ is applied to the BRFSS dataset, considering the health status categories 'Normal,' 'Pre-diabetes,' and 'Diabetes.' The dataset is divided into 10 folds, ensuring that each fold maintains the same proportion of the three health status categories

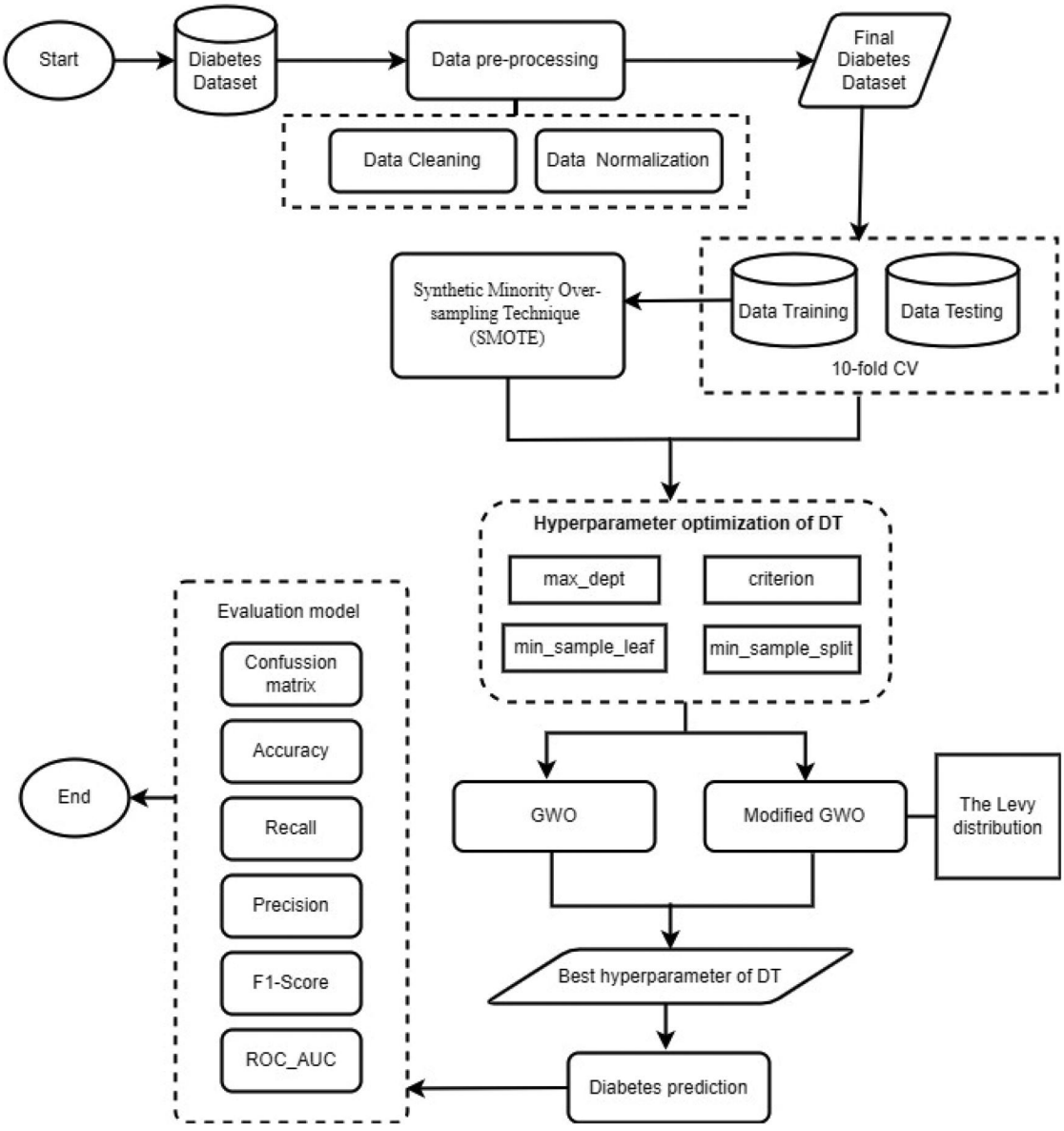


Figure 1. The workflow of the proposed model.

as in the complete dataset. This approach allows the machine learning model to be trained and tested on various subsets of the data, ensuring that the model's performance is comprehensively and robustly evaluated across the entire dataset. By maintaining a balanced distribution of health status categories, the k-fold cross-validation process reduces potential bias and provides a reliable assessment of the model's predictive capabilities. The visualization of the division between the training set and the testing set using k-fold cross-validation is shown in Figure 4.

2.4. Synthetic minority over-sampling technique

Imbalanced data can be disruptive when training ML models (Ray et al. 2020). SMOTE (Synthetic Minority Oversampling Technique) is a popular approach to address this data imbalance by adding more samples from the minority class. However, when using SMOTE to augment samples from the minority class, new samples are generated randomly between two sample points and lack control. This can lead to the emergence of more unusual or abnormal samples (Liu et al. 2023).

The BRFSS dataset has three classes, and the distribution of these classes can be seen in Figure 5. There is an imbalance among the classes in this dataset. As shown in Figure 5(a), around 87.67% of the data falls into the dominant class, namely, normal. Meanwhile, the other classes contribute only about 10.78% for pre-diabetes and 1.54% for diabetes collectively. The problem is the domination of the majority class over the minority classes. The created model tends to favor

Table 1. Characteristics of key attributes in the diabetes health indicators dataset.

Attribute	Description
Diabetes_012	0 = no diabetes, 1 = prediabetes, 2 = diabetes
HighBP	High blood pressure (0 = no, 1 = yes)
HighChol	High cholesterol (0 = no, 1 = yes)
CholCheck	Cholesterol check in last 5 years (0 = no, 1 = yes)
BMI	Body Mass Index
Smoker	Smoked 100+ cigarettes (0 = no, 1 = yes)
Stroke	Ever had a stroke (0 = no, 1 = yes)
HeartDiseaseorAttack	Coronary heart disease or heart attack (0 = no, 1 = yes)
PhysActivity	Physical activity in last 30 days (0 = no, 1 = yes)
Fruits	Consume fruit daily (0 = no, 1 = yes)
Veggies	Consume vegetables daily (0 = no, 1 = yes)
HvyAlcoholConsump	Heavy alcohol consumption (0 = no, 1 = yes)
AnyHealthcare	Any healthcare coverage (0 = no, 1 = yes)
NoDocbcCost	Could not see doctor due to cost (0 = no, 1 = yes)
GenHlth	General health (1 = excellent, 5 = poor)
MentalHlth	Days mental health not good (1–30 days)
PhysHlth	Days physical health not good (1–30 days)
DiffWalk	Difficulty walking/climbing stairs (0 = no, 1 = yes)
Sex	Gender (0 = female, 1 = male)
Age	Age
Education	Education level (1 = less than HS, 6 = college grad)
Income	Income level (1 = ¡10, 000, 8 =75,000+)

	count	mean	median	std	min	25%	50%	75%	max	skewness	kurtosis
Diabetes_012	253680	0.297	0	0.698	0	0	0	0	2	1.976379	1.980085
HighBP	253680	0.429	0	0.495	0	0	0	1	1	0.286903	-1.917687
HighChol	253680	0.424	0	0.494	0	0	0	1	1	0.307073	-1.905706
CholCheck	253680	0.963	1	0.19	0	1	1	1	1	-4.881243	21.826529
BMI	253680	28.382	27	6.609	12	24	27	31	98	2.121991	10.997233
Smoker	253680	0.443	0	0.497	0	0	0	1	1	0.228809	-1.947647
Stroke	253680	0.041	0	0.197	0	0	0	0	1	4.657312	19.690557
HeartDiseaseorAttack	253680	0.094	0	0.292	0	0	0	0	1	2.778725	5.721315
PhysActivity	253680	0.757	1	0.429	0	1	1	1	1	-1.195539	-0.570686
Fruits	253680	0.634	1	0.482	0	0	1	1	1	-0.557496	-1.689198
Veggies	253680	0.811	1	0.391	0	1	1	1	1	-1.592229	0.535194
HvyAlcoholConsump	253680	0.056	0	0.23	0	0	0	0	1	3.854109	12.854156
AnyHealthcare	253680	0.951	1	0.216	0	1	1	1	1	-4.181091	15.481522
NoDocbcCost	253680	0.084	0	0.278	0	0	0	0	1	2.995272	6.971655
GenHlth	253680	2.511	2	1.068	1	2	2	3	5	0.422864	-0.383277
MentHlth	253680	3.185	0	7.413	0	0	0	2	30	2.721132	6.441534
PhysHlth	253680	4.242	0	8.718	0	0	0	3	30	2.207382	3.496149
DiffWalk	253680	0.168	0	0.374	0	0	0	0	1	1.773897	1.146710
Sex	253680	0.44	0	0.496	0	0	0	1	1	0.240348	-1.942233
Age	253680	8.032	8	3.054	1	6	8	10	13	-0.359901	-0.581235
Education	253680	5.05	5	0.986	1	4	5	6	6	-0.777251	0.039429
Income	253680	6.054	7	2.071	1	5	7	8	8	-0.891340	-0.280347

Figure 2. The statistics of BRFSS dataset.

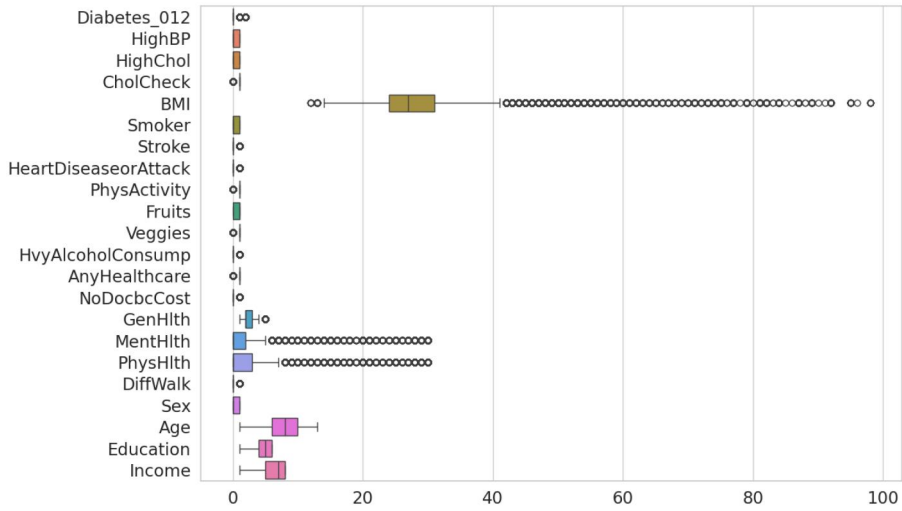


Figure 3. The BRFSS dataset distribution box plot.

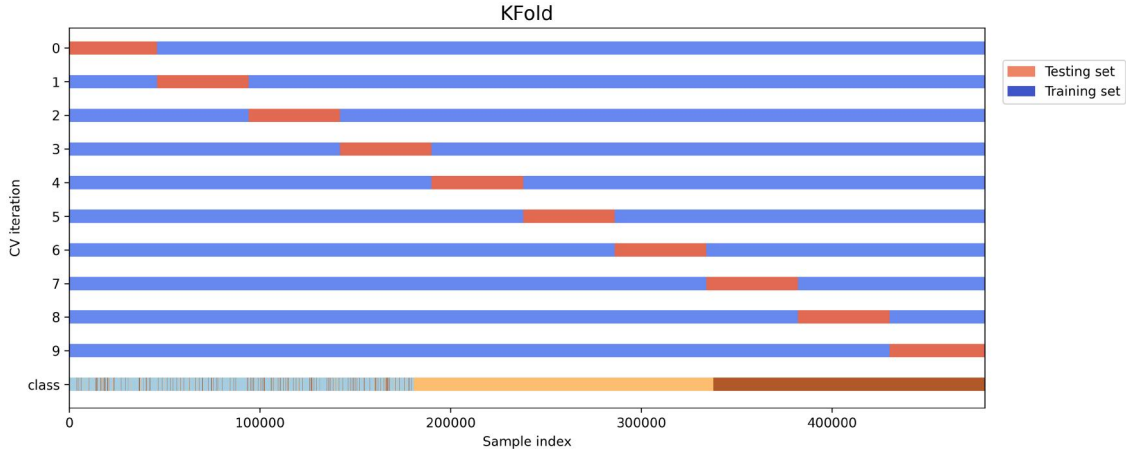
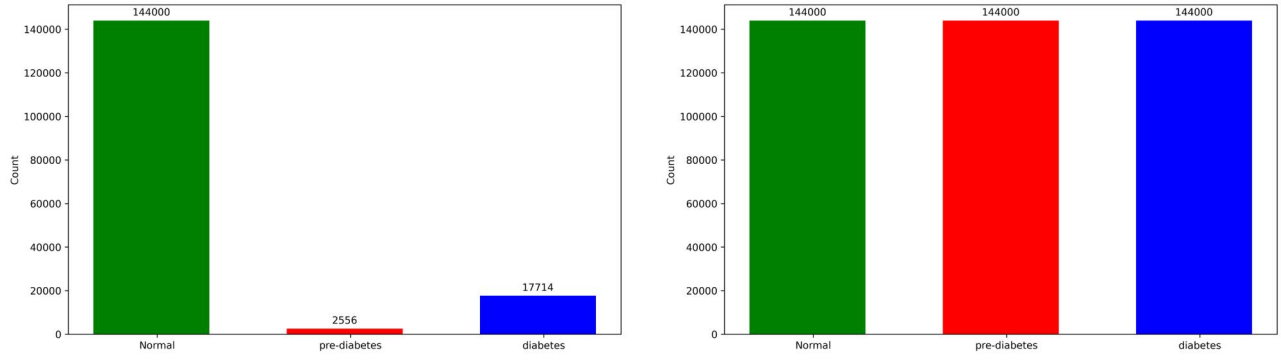


Figure 4. Visualizing k-Fold validation with $k = 5$.



(a) Before SMOTE (b) After SMOTE

Figure 5. The class distribution of the BRFSS dataset.

the majority class, resulting in less accurate classification for the minority classes. SMOTE is used to address this problem. Applying SMOTE resolves the issue of imbalanced distribution, and the classes are spread more evenly throughout the dataset, as illustrated in Figure 5(b). An experimental study was conducted to compare and discuss the results before and after the implementation of SMOTE.

2.5. Basic theory of decision tree

Decision Tree (DT) is a well-known classification algorithm used in classification tasks. This algorithm employs simple decision rules derived from data features to build a prediction model. Its key advantages lie in its interpretability and its ability to handle both numerical and categorical data. In the DT algorithm, commonly used objective functions are Gini impurity and Entropy. Gini impurity measures the uniformity of the target label distribution in the dataset, while Entropy measures the uncertainty or disorder in the dataset. Both metrics are used to ensure that the resulting DT is optimal and can classify data accurately and efficiently. The following formulas calculate Gini impurity and Entropy:

$$GINI(\mathrm{D}) = 1 - \sum_{i=1}^{\mathrm{J}} (p_i)^2 \tag{3}$$

$$Entropi(\mathrm{D}) = -\sum_{i=1}^{\mathrm{J}} p_i \log_2 p_i \tag{4}$$

where J is the number of target classes D and p_i is the proportion of the sample that belongs to a particular class

These hyperparameters play an important role in balancing the model's complexity and preventing overfitting, ultimately influencing the DT effectiveness in capturing patterns within the data. More details are presented in Table 2.

2.6. Grey wolf optimization

In 2014, Mirjalili et al. (2014) created the Grey Wolf Optimizer (GWO) algorithm, inspired by how grey wolves behave in groups. In the wild, grey wolves usually live together in groups of 5 to 12 members. They have a clear social order, with the most dominant wolves, males and females, being called 'alpha (α)' and holding the top position. These wolves make decisions about food, sleep, hunting, and where the group lives. The other wolves follow the lead of the α wolves. Additionally, 'beta (β)' wolves follow α decisions and oversee the lower-ranking wolves. Another group, called 'delta (δ)' wolves, helps α and β hunt and find prey.

The GWO algorithm uses these ideas to solve problems in computer algorithms more effectively. The grey wolves perform several essential tasks within their group. They guard their territory, alert other wolves to dangers, and care for injured or weak members. Omega (ω) wolves, positioned at the lowest level, obey the orders of all other wolves. The success of wolf hunting relies heavily on their social hierarchy. The social behavior of grey wolves can be mathematically modeled by considering the prey's location as the optimal solution and representing the wolves' positions as solutions in the search space. α Wolves are the best solution as they are closest to the prey. β and δ Wolves are the next best solutions according to their positions in the group hierarchy. In the search space, omega wolves follow the updates in the positions of α, β, and δ wolves. Suppose their positions in the search space are denoted as X_α, X_β, X_δ, and X_ω, respectively. The GWO algorithm main steps involve encircling prey, hunting, attacking, and searching. Encircling prey is when wolves move around the prey during hunting, and this can be mathematically modeled using Equations (5)–(8).

$$\overrightarrow{D} = \left| \overrightarrow{C} \cdot \overrightarrow{X_p}(t) - \overrightarrow{X}(t) \right| \tag{5}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X_p}(t) - \overrightarrow{C} \cdot \overrightarrow{D} \tag{6}$$

$$\overrightarrow{C} = 2 \cdot \overrightarrow{a} \cdot \overrightarrow{rand_1} - \overrightarrow{a} \tag{7}$$

$$\overrightarrow{A} = 2 \cdot \overrightarrow{rand_2} \tag{8}$$

where $\overrightarrow{X}(t)$ and $\overrightarrow{X_p}(t)$ represent the current iteration's position vectors of the grey wolf and the prey, respectively. $\overrightarrow{C}$ and $\overrightarrow{A}$ are coefficient vectors, $\overrightarrow{rand_1}$ and $\overrightarrow{rand_2}$ are two random vectors in [0, 1], and $\overrightarrow{a}$ is a vector whose value decreases from 2 to 0 during iterations. The prey-hunting process is led by the α wolf, with the participation of β and δ wolves. These three wolves are

Table 2. List of DT parameters.

Parameter	Description	Type data
max_depth	The maximum depth of the decision tree; if None, nodes expand until leaves have samples less than min_samples_split.	[integer, integer]
min_samples_split	The minimum number of samples required to split a node. Suppose an integer is the minimum absolute number of samples required to split. If a float, it represents the fraction of the total number of samples.	[integer, integer]
min_samples_leaf	The minimum number of samples required to be a leaf node. Similar to Min Samples Split, it prevents the creation of nodes with too few samples if an integer is the minimum absolute number of samples required. If a float, it represents the fraction of the total number of samples.	[integer, integer]
Criterion	The criterion to measure the quality of the split 'gini' for Gini impurity and 'entropy' for Entropy	['gini', 'entropy', 'log_loss']

assumed to know the region where the prey might be located, assisting in obtaining the three best search agents. These agents further help update other wolves' positions, as seen in Equations (9)–(15).

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1}\,\overrightarrow{X_\alpha} - \overrightarrow{X}\right| \tag{9}$$

$$\overrightarrow{D_\beta} = \left|\overrightarrow{C_2}\,\overrightarrow{X_\beta} - \overrightarrow{X}\right| \tag{10}$$

$$\overrightarrow{D_\delta} = \left|\overrightarrow{C_3}\,\overrightarrow{X_\delta} - \overrightarrow{X}\right| \tag{11}$$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \overrightarrow{D_\alpha} \tag{12}$$

$$\overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \overrightarrow{D_\beta} \tag{13}$$

$$\overrightarrow{X_3} = \overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot \overrightarrow{D_\delta} \tag{14}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{15}$$

The attacking step is like the exploitation step, done using the factor '$\overrightarrow{a}$'. When the prey stops, the wolves in motion initiate an attack on the prey. The value of $\overrightarrow{A}$ is a random number within the range of $[-2r, 2r]$, and the value of $r2$ is within the range of $[-1, 1]$. The new position of the search agent can be any position between its current location and the prey's location. Therefore, the attacking condition is applicable when $\left|\overrightarrow{A}\right| < 1$.

The search for the best solution is modeled after the wolves' searching behavior. Wolves spread out to search for prey and come together when they find it. They spread out to find better prey if $\left|\overrightarrow{A}\right| > 1$ and converge towards the prey if $\left|\overrightarrow{A}\right| < 1$. The random value $\overrightarrow{C}$ is utilized to prevent local optima and promote exploration, introducing randomness from the algorithm's start to finish, enhancing the exploration concept without bias.

2.7. Hyperparameter optimization of DT model using GWO and modified GWO

This research aims to enhance the GWO algorithm's performance in optimizing DT model hyperparameters. The range of hyperparameter values to be optimized has been specified based on Table 3. The objective function is formulated in Equation (16). $\overrightarrow{params_i} \in \mathbb{R}^k$ is a k-dimensional hyperparameter vector. TX_i represents some data selected from the respective training data. The $i_{\max}$ value is used to manage the number of iterations the DT requires to optimize the hyperparameters. Selecting a higher value for $i_{\max}$ will consume more time to optimize the model than a smaller value. Therefore, the value of $i_{\max}$ should be determined by the data conditions used so that the GWO optimization process does not incur excessive computational costs. The DT hyperparameter optimization using the GWO algorithm is explained in Algorithm 1.

Table 3. Range parameters value of DT.

Parameter	Range	Symbol
max_*depth*	[1,100]	md
min_*samples_splith*	[2,20]	mss
min_*samples_leaf*	[1,10]	msl
criterion	[0,1]	cr

criter_mapping = 0: 'gini', 0.5: 'entropy', 1: 'log_loss'.

$$\max(acc = DecisionTreeClassifier(\overrightarrow{params_i}, TX_i)), i < i_{\max} \tag{16}$$

Algorithm 1: DT Hyperparameter optimization using GWO algorithm

```
procedure GWO_DECISION_TREE_OPTIMIZATION (Data, maxiter)
Initialize a population of grey wolves (X_i (1,2, ..., n); a, A, C) randomly.
Set α, β, δ positions to the initial positions of three random wolves.
Define md, mss, msl, cr.
for each iteration from 1 to maxiter do
    Calculate the fitness of each wolf by Equation (16).
    Update the position of current wolf by Equation (15).
    for each wolf in the population do
        Update a, A, C.
        Calculate the fitness of all agents.
        Update the position of the current wolf by using Equations (12)–(14)
        Clamp the wolves' positions to stay within the search range.
        Evaluate the fitness of the new position.
        if the new position has better fitness than the current position then
            Update the position.
        end if
    end for
end for
Return the best hyperparameters found.
end procedure
```

The shortcomings of the GWO search space for DT hyperparameter optimization with a wide range may result in ineffective early optimization. Therefore, we propose a modified GWO by incorporating the Lévy distribution function. The Lévy distribution function is a type of probability distribution used in optimization algorithms to generate random steps with varying distances. The Lévy flight mechanism enhances the search

process by allowing for larger, more explorative moves early in the optimization process, which helps in escaping local optima. As the optimization progresses, the step sizes decrease, enabling fine-tuning of the hyperparameters. This balance between exploration and exploitation improves the effectiveness and efficiency of the DT hyperparameter optimization. The general formula for the Lévy distribution is:

$$L(x;\mu,c) = \frac{1}{\pi c}\exp\left(-\frac{1}{2c}(x-\mu)^2\right) \tag{17}$$

where μ is the position parameter (usually the mean value), and c is the scale parameter (which controls the variance). By integrating the Lévy distribution function into a modified GWO, we aim to improve the movement of α, β, and δ, thereby enhancing the algorithm's ability to explore solution spaces more effectively. The DT hyperparameter optimization using a modified GWO algorithm is explained in Algorithm 2.

Algorithm 2: DT Hyperparameter optimization using modified GWO algorithm

procedure MODIFIED_GWO_DECISION_TREE_OPTIMIZATION(Data, max*iter*)
Initialize a population of grey wolves (X_i (1,2, ..., n); $\vec{a}$, $\vec{A}$, $\vec{C}$) randomly.
Set α, β, δ positions to the initial positions of three random wolves.
Define md, mss, msl, cr.
for each iteration from 1 to max*iter* **do**
  Calculate the fitness of each wolf by Equation (16).
  for each wolf in the population **do**
    Update the position of current wolf by Equation (15).
    Update $\vec{a}$, $\vec{A}$, $\vec{C}$ positions using Equation (17)
    Calculate the fitness of all agents.
    Update the position of the current wolf by using Equations (12)–(14)
    Clamp the wolves' positions to stay within the search range.
    Evaluate the fitness of the new position.
    if the new position has better fitness than the current position **then**
      Update the position.
    end if
  end for
end for
Return the best hyperparameters found.
end procedure

2.8. Evaluation performance

The performance evaluation of the proposed method uses a confusion matrix for a three-class classification problem. The confusion matrix includes true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP). This matrix is employed to calculate accuracy, recall, precision, and F1-score, formulated as follows.

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TN+FP+TP+FN} \\ Recall &= \frac{TP}{TP+FN} \\ Precison &= \frac{TP}{TP+FP} \\ F1-score &= 2\times\frac{Precison\times Recall}{Precison+Recall} \end{aligned} \tag{18}$$

Additionally, we use the ROC AUC (Receiver Operating Characteristic - Area Under the Curve) metric to evaluate the overall performance of the model, providing insight into its ability to distinguish between classes.

3. Result and discussion

Decision Tree (DT) is a popular classification model due to its computational efficiency, interpretability, and ability to intuitively understand variable relationships. However, DT often faces overfitting issues that can affect classification performance. To overcome this problem, we automatically optimize DT hyperparameters using the GWO and modified GWO algorithms. In both algorithms, we use 5 wolves and run the algorithm for 30 iterations. We chose this number of wolves to provide a sufficiently large population for adequate exploration in the hyperparameter search space. This approach allows us to obtain a good representation of various possible solutions and increases the chances of finding optimal parameters. Additionally, 30 iterations give the algorithm enough time to thoroughly explore the search space and achieve convergence to a stable solution. With this setup, we achieve an optimal balance between exploration and exploitation in hyperparameter search while ensuring time efficiency in the optimization process.

Figure 6 compares the hyperparame valuesand fitness scores at each iteration based on the updates of α, β, and δ positions in both GWO and MGWO. From Figure 6(a), it can be seen that the max_depth value stabilizes at 37 starting from iteration 14 to the end for the α position, from iteration 11 to the end for the β

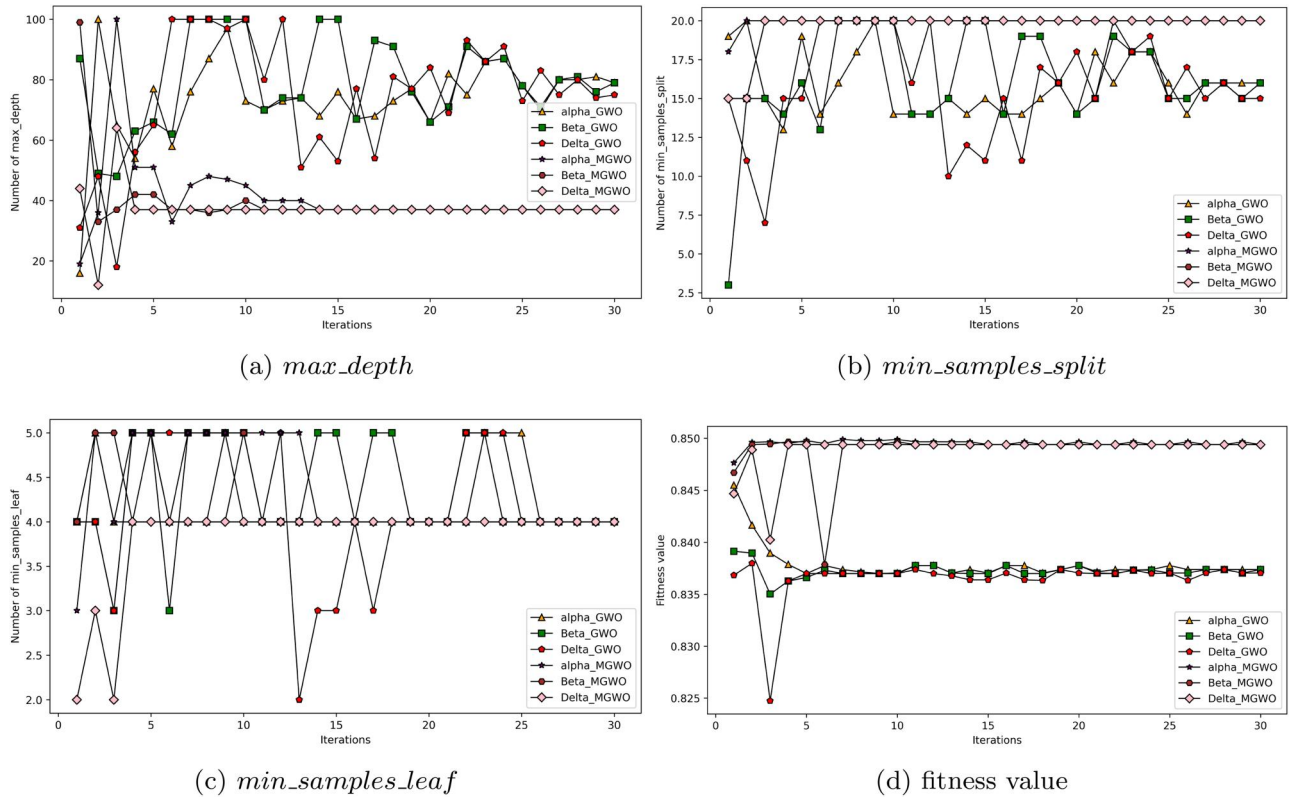


Figure 6. The comparison of hyperparameter and fitness values based on the positions of α, β, and δ in GWO and MGWO.

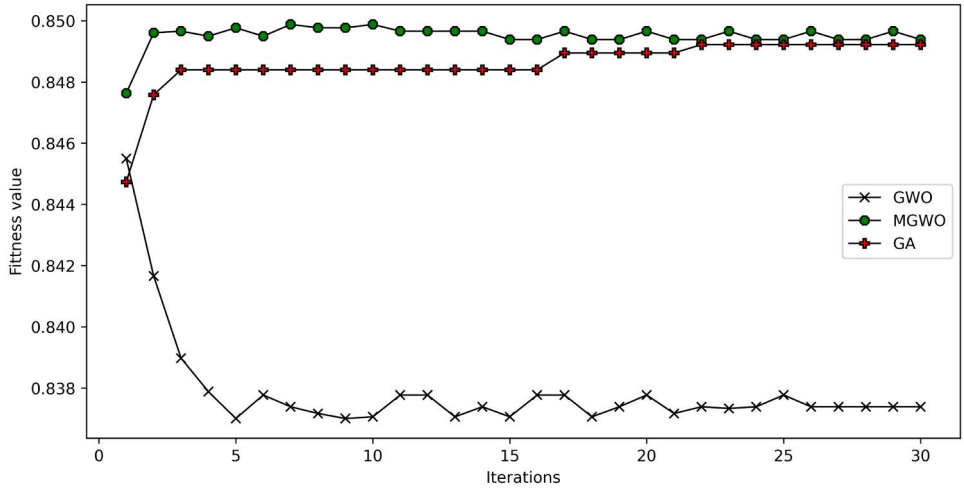


Figure 7. The comparison of fitness value in GWO and MGWO with GA.

position, and from iteration 4 to the end for the δ position. Conversely, the max_depth value is unstable in GWO for all α, β, and δ positions. This stability of max_depth indicates that MGWO is more effective in determining the optimal maximum depth of DT, which is important for avoiding overfitting and ensuring the model is neither too complex nor too simple.

Figure 6(b) shows that the min_samples_split value stabilizes at 20, generated by MGWO starting from iteration 2 to the end for the α and β positions, and from iteration 3 to the end for the δ position.

Table 4. The comparison of best hyperparameter values from GWO and MGWO with GA.

Parameter	Optimization algorithm		
	GWO	MGWO	GA
max_*depth*	79	37	29
min_*samples_splith*	16	20	19
min_*samples_leaf*	4	4	4
criterion	gini	entrophy	entrophy

Conversely, the min_samples_split value is unstable in GWO for all α, β, and δ positions. This stability of $min_samples_split$ indicates that MGWO is better at

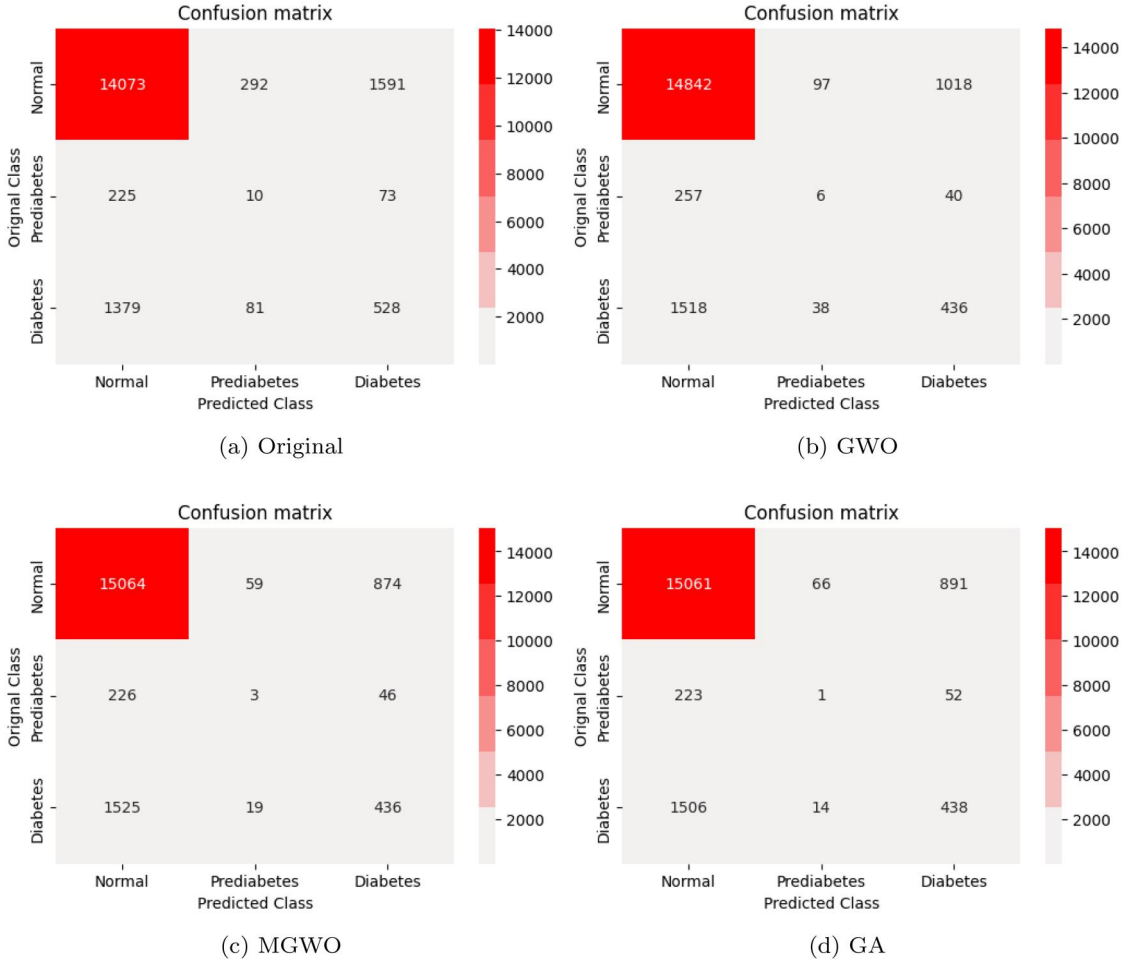
determining the minimum number of samples required to split an internal node, which is important for preventing overfitting and ensuring each node split is significant. Figure 6(c) shows that MGWO also achieves stability for the min_samples_leaf value at 4 starting from iteration 14 to the end for the α position, and from iteration 4 to the end for the β and δ positions. This stability of min_samples_leaf indicates that MGWO is more effective in determining the minimum number of samples that must be present in a leaf node, which is crucial for reducing noise and avoiding overfitting. For the criterion parameter, MGWO uses 'entropy' while GWO uses 'Gini.' This criterion type is generated from the initial iteration to the end for both algorithms. Entropy as a criterion is better in some cases because it is more sensitive to uncertainty in the splits, while Gini is often faster to compute and effective in many practical situations.

Table 5. The comparison of evaluations from hyperparameter optimization with GWO and MGWO with GA.

Evaluation	Original	Optimization algorithm		
		GWO	MGWO	GA
Precision				
Normal	90	89	90	90
Prediabetes	3	4	4	1
Diabetes	24	29	32	32
Recall				
Normal	88	93	94	94
Prediabetes	3	2	1	0
Diabetes	27	22	22	22
f1-Score				
Normal	89	91	92	92
Prediabetes	3	3	2	1
Diabetes	25	25	26	26
Accuracy	80	84	85	85

The optimal values of all these DT parameters align with the fitness values produced by both algorithms as shown in Figure 6(d). MGWO achieves a fitness value of 0.849, better than GWO's fitness value of 0.837. These results indicate that MGWO is more effective in exploring and exploiting the hyperparameter search space, resulting in a model that is better in terms of generalization and overall performance.

The involvement of the Lévy distribution in MGWO can broaden the exploration scope in the search space, contributing to the algorithm's ability to find better solutions. Integrating the Lévy distribution



(a) Original (b) GWO

(c) MGWO (d) GA

Figure 8. The comparison of confusion matrices between GWO and MGWO with GA.

has proven effective in enhancing the algorithm's performance, as reflected in the increased success in achieving maximum and minimum values for certain hyperparameters in MGWO. Therefore, the combination of GWO and the Lévy distribution in MGWO presents opportunities to enhance exploration in the search space and improve the overall algorithm performance in solving complex optimization problems

The fitness comparison of hyperparameter optimization between GWO and MGWO with GA is presented in Figure 7. It can be observed that the fitness values generated by GA show a more stable improvement from the beginning to the end of iterations compared to the fitness values produced by GWO and MGWO. This could be attributed to the nature of the GA algorithm, which can explore the search space widely and diversely, thus providing more consistent solutions over time. Although the fitness value from GA is better than that from GWO, it is still inferior to MGWO. This result comparison indicates that MGWO has an advantage in conducting a more efficient and accurate search space exploration, leading to better optimal solutions than GA for DT hyperparameter optimization.

Table 4 shows the best results from the hyperparameter optimization process for the DT model. The results indicate that MGWO performs better than GWO and GA in hyperparameter optimization, as

Table 6. The comparison of evaluations from hyperparameter optimization with GWO and MGWO using PIMA Indian datasets.

Evaluation	Optimization algorithm	
	GWO	MGWO
Precision		
Normal	92	96
Diabetes	96	85
Recall		
Normal	98	92
Diabetes	87	82
f1-Score		
Normal	95	94
Diabetes	92	88
Accuracy	93	93

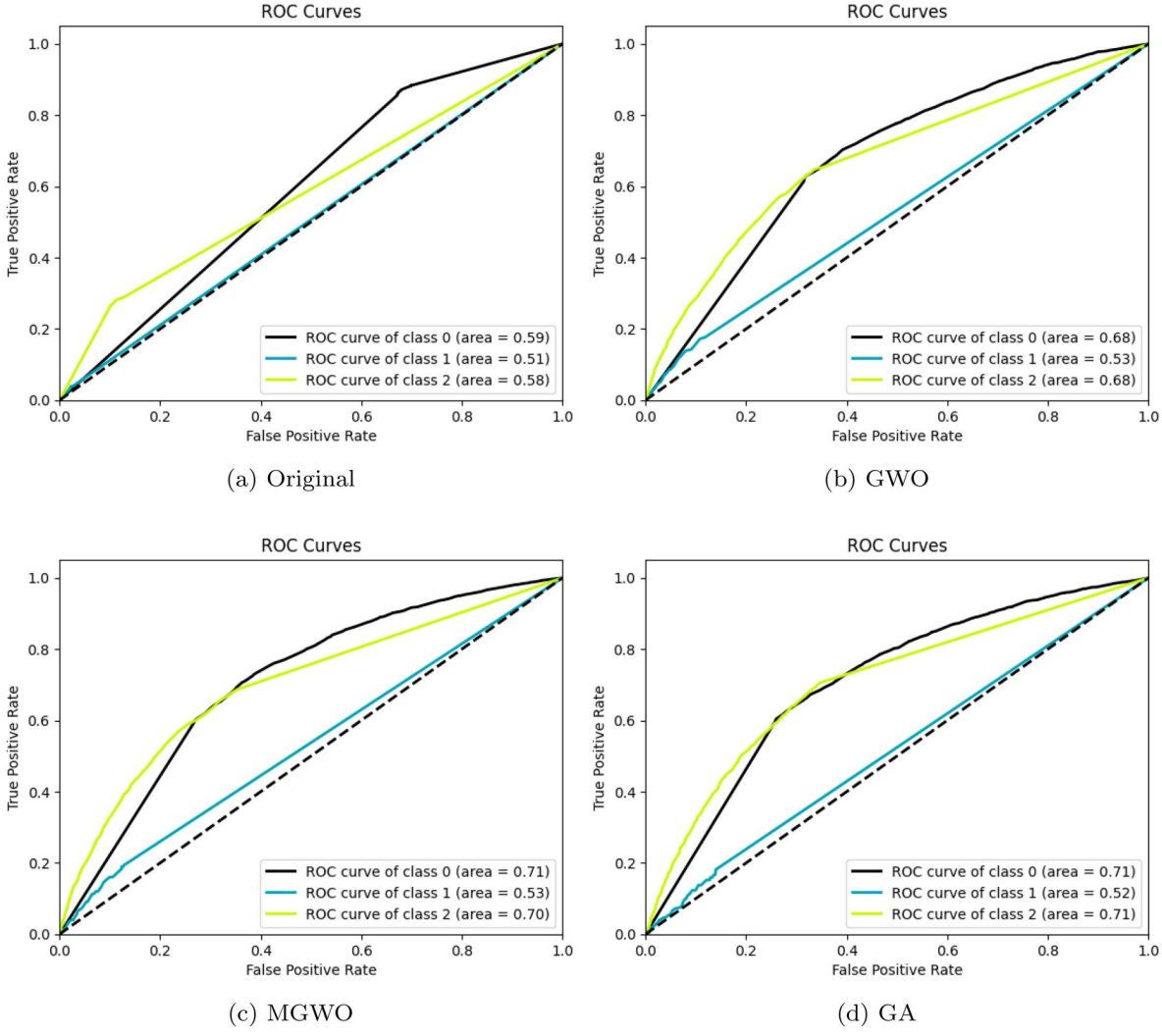


Figure 9. The comparison of ROC between GWO and MGWO with GA.

seen from the fitness scores. MGWO can produce more stable parameter values compared to GWO and GA. A max_depth value that is too large can negatively impact accuracy, as trees that are too deep tend to overfit. On the other hand, higher min_samples_split and min_samples_leaf values can improve accuracy by preventing overly small splits and too few leaves. The consistent use of the 'entropy' criterion shows a preference for this method in measuring node split impurity during DT construction, as it is often more sensitive to uncertainty in splits.

The best hyperparameter values are used to test diabetes classification using imbalanced testing data. The comparison of evaluations from hyperparameter optimization with GWO and MGWO against GA is presented in Table 5, the Confusion Matrix is shown in Figure 8, and the ROC Curve is provided in Figure 9. Based on these evaluation comparisons, hyperparameter optimization using GWO and MGWO successfully improved DT performance. MGWO outperformed GWO in all performance evaluations. MGWO also performed better than GA in terms of precision, recall, and f1-score for class 1 (Prediabetes), and was comparable to GA for precision, recall, and f1-score for class 0 (Normal) and class 2 (Diabetes). These evaluation comparison results are also consistent with the Confusion Matrix and ROC values produced by each algorithm.

We also applied the proposed algorithm for diabetes classification using the Pima Indian Diabetes dataset. This dataset was chosen as a benchmark that is frequently used by researchers to test new algorithms. Table 6 presents a comparison of model performance evaluations, while Figure 10 presents the comparison of Confusion Matrix and ROC values. The results show that GWO outperforms MGWO in all performance evaluations, Confusion Matrix, and ROC values. The superiority of GWO can be explained by the small number of attributes in the PIMA Indian dataset, which has only 9

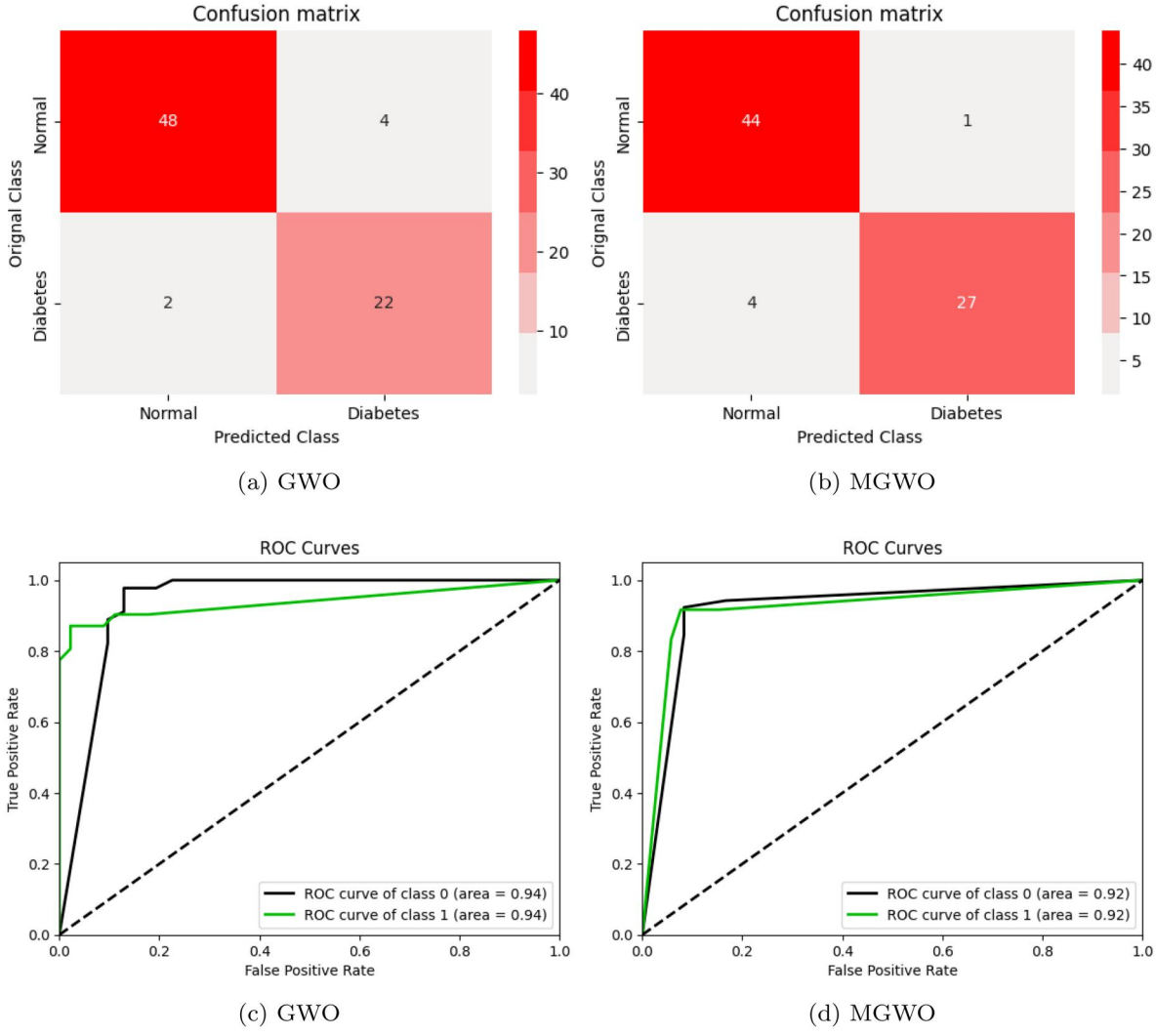


Figure 10. The comparison of confusion matrix with GWO and MGWO using PIMA Indian datasets.

Table 7. The accuracy comparison with existing methods for diabetes classification.

Ref.	Datasets	Models	Accuracy (%)
Iyer and Sumbaly (2015)	PIMA Indian	J48	74.86
		Naïve Bayes	79.56
Chang et al. (2022)	PIMA Indian	J48	75.65
		Naïve Bayes	77.83
		Random forest	73.91
Patro et al. (2023)	PIMA Indian	Random Forest	82.82
Reza et al. (2024)	PIMA Indian	Staking Ensemble	77.10
Zhang et al. (2024)	PIMA Indian	AHDHS-Stacking	90.89
Daza et al. (2024)	PIMA Indian	Random Forest	88.5
		Stacking 1 A	91.5
		Stacking 2 A	88.5
Chowdhury et al. (2024)	BRFSS	Logistic regression	83
		Random forest	81.7
		Gradient boosting	83
		AdaBoosting	83
Proposed method	BRFSS	GWO-DT	84
		MGWO-DT	85
	PIMA Indian	GWO-DT	92
		MGWO-DT	93

attributes. With a more limited search space, GWO can find optimal solutions more quickly and efficiently compared to MGWO, which has a more complex exploration mechanism. Additionally, the two classes in this dataset make it easier for the algorithm to achieve better performance without the need for the in-depth exploration performed by MGWO. Next, Table 7 presents a comparison of the accuracy of the proposed method with existing methods for diabetes classification. The comparison results show that the proposed algorithm outperforms the existing algorithms.

4. Conclusion

Diabetes is a chronic condition that affects blood sugar levels and can damage vital organs in the body. Early detection is crucial given the increasing global prevalence of diabetes and the serious risk of complications if not properly managed, making a good prediction system necessary. This study utilizes the BRFSS dataset with 22 variables and 253,680 records with 3 labels, which after pre-processing resulted in 21 variables with 182,522 records ready for diabetes classification. DT is a commonly used model for classification, but its performance decreases with large dataset sizes. Therefore, we propose DT hyperparameter optimization using the GWO algorithm, which, although rarely used, has great potential due to its exploration and exploitation capabilities. However, GWO has a limited search space, so we propose a Modified GWO (MGWO) incorporating the Lévy distribution function to enhance the movements of alpha, beta, and delta wolves. Additionally, we use GA as a comparative optimization algorithm. The results show that DT hyperparameter optimization using MGWO outperforms GWO and GA in terms of fitness values. Evaluation with test data shows that MGWO and GA achieve similar and better accuracy compared to GWO. The next step is to apply MGWO for hyperparameter optimization on other machine learning models and optimization problems. The potential applications include various fields such as image recognition, natural language processing, and anomaly detection. Further research is also needed to configure the number of wolves to gain detailed insights into their influence on hyperparameter optimization performance.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Albahli S. 2023. Efficient hyperparameter tuning for predicting student performance with Bayesian optimization. Multimed Tools Appl. 83(17):52711–52735. doi: 10.1007/s11042-023-17525-w.

Almazini HF, Ku-Mahamud KR, Almazini H. 2023. Heuristic initialization using grey wolf optimizer algorithm for feature selection in intrusion detection. International Journal of Intel Engng Sys. 16(1):410–418. doi: 10.22266/ijies2023.0228.36.

Almutairi SA. 2023. An optimized feature selection and hyperparameter tuning framework for automated heart disease diagnosis. Comp Sys Sci Engng. 47(2):2599–2623. doi: 10.32604/csse.2023.041609.

Al-Tashi Q, Rais HM, Abdulkadir SJ, Mirjalili S. 2020. Feature selection based on grey wolf optimizer for oil gas reservoir classification. 2020 International Conference on Computational Intelligence, ICCI 2020, 211–216. doi: 10.1109/ICCI51257.2020.9247827.

Alyasseri ZAA, Alomari OA, Makhadmeh SN, Mirjalili S, Al-Betar MA, Abdullah S, Ali NS, Papa JP, Rodrigues D, Abasi AK. 2022. EEG channel selection for person identification using binary grey wolf optimizer. IEEE Access. 10:10500–10513. doi: 10.1109/ACCESS.2021.3135805.

Challapalli JR, Devarakonda N. 2022. A novel approach for optimization of convolution neural network with hybrid particle swarm and grey wolf algorithm for classification of Indian classical dances. Knowl Inf Syst. 64(9):2411–2434. doi: 10.1007/s10115-022-01707-3.

Chan YN, Wang P, Chun KH, Lum JTS, Wang H, Zhang Y, Leung KSY. 2023. A machine learning approach for early prediction of gestational diabetes mellitus using elemental contents in fingernails. Sci Rep. 13(1):4184. doi: 10.1038/s41598-023-31270-y.

Chang V, Bailey J, Xu QA, Sun Z. 2022. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Comput Appl. 35(22):1–17. doi: 10.1007/s00521-022-07049-z.

Chen P, Wang B, Zhao L, Ma S, Wang Y, Zhu Y, Zeng X, Bai Z, Shi B. 2023. Machine learning for predicting intrahospital mortality in ST-elevation myocardial infarction patients with type 2 diabetes mellitus. BMC Cardiovasc Disord. 23(1):585. doi: 10.1186/s12872-023-03626-9.

Chowdhury MM, Ayon RS, Hossain MS. 2024. An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset. Healthcare Analytic. 5: 100297. doi: 10.1016/j.health.2023.100297.

Cinar AC. 2023. A novel adaptive memetic binary optimization algorithm for feature selection. Artif Intell Rev. 56(11):13463–13520. doi: 10.1007/s10462-023-10482-8.

Cuong-Le T, Minh HL, Sang-To T, Khatir S, Mirjalili S, Abdel Wahab M. 2022. A novel version of grey wolf optimizer based on a balance function and its application for hyperparameters optimization in deep neural network (DNN) for structural damage identification. Eng Fail Anal. 142:106829. doi: 10.1016/j.engfailanal.2022.106829.

Daza A, Ponce Sánchez CF, Apaza-Perez G, Pinto J, Zavaleta Ramos K. 2024. Stacking ensemble approach to diagnosing the disease of diabetes. Inf Med Unlocked. 44:101427. doi: 10.1016/j.imu.2023.101427.

Emary E, Zawbaa HM, Grosan C, Hassenian AE. 2015. Feature subset selection approach by gray-wolf optimization. In Advances in Intelligent Systems and Computing. Vol. 334; p. 1–13. doi: 10.1007/978-3-319-13572-4_1.

Feng H, Wang F, Li N, Xu Q, Zheng G, Sun X, Hu M, Li X, Xing G, Zhang G. 2023. Use of tree-based machine learning methods to screen affinitive peptides based on docking data. Mol Inform. 42(12):e202300143. doi: 10.1002/minf.202300143.

Gosh S. 2017. Implementation of proper nutrition in managing diabetes: a review article. J Food Sci Res.

Guido R, Groccia MC, Conforti D. 2023. A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers. Soft Comput. 27(18):12863–12881. doi: 10.1007/s00500-022-06768-8.

Habib B, Khursheed F. 2022. Performance evaluation of machine learning models for distributed denial of service attack detection using improved feature selection and hyper-parameter optimization techniques. Concurr Comput. 34(26). doi: 10.1002/cpe.7299.

Habibi S, Ahmadi M, Alizadeh S. 2015. Type 2 diabetes mellitus screening and risk factors using decision Tree: results of data mining. Glob J Health Sci. 7(5):304–310. doi: 10.5539/gjhs.v7n5p304.

Hang OY, Wiwied V, Rosaida R. 2024. Diabetes prediction using machine learning. J Advan Res Appl Sci Engng Tech. 37(1):82–98. doi: 10.1007/978-981-16-0733-2_50.

Hennebelle A, Ismail L, Materwala H, Al Kaabi J, Ranjan P, Janardhanan R. 2024. Secure and privacy-preserving automated machine learning operations into end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction. Comput Struct Biotechnol J. 23:212–233. doi: 10.1016/j.csbj.2023.11.038.

Hou Y, Li S, Gong S, Huang J, Zhang J, Lu J. 2023. Hybrid algorithm of filter and improved gray wolf optimization for fault feature selection of rolling bearing. Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS. 29(5):1452–1461. doi: 10.13196/j.cims.2023.05.004.

Iyer ASJ, Sumbaly R. 2015. Diagnosis of diabetes using classification mining techniques. IJDKP. 5(1):01–14. doi: 10.5121/ijdkp.2015.5101.

Jain R, Joseph T, Saxena A, Gupta D, Khanna A, Sagar K, Ahlawat AK. 2023. Feature selection algorithm for usability engineering: a nature inspired approach. Complex Intell Syst. 9(4):3487–3497. doi: 10.1007/s40747-021-00384-z.

Kadam VJ, Jadhav SM. 2020. Performance analysis of hyperparameter optimization methods for ensemble learning with small and medium sized medical datasets. J Discrete Math Sci Cryptograph. 23(1):115–123. doi: 10.1080/09720529.2020.1721871.

Kumar M, Namrata K, Kumari N. 2022. Hyper-parametric improved machine learning models for solar radiation forecasting. Concurr Comput. 34(23). doi: 10.1002/cpe.7190.

Kuyu YC, Ozekmekci N. 2022. Grey wolf optimizer to the hyperparameters optimization of convolutional neural network with several activation functions. ISMSIT 2022—6th International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings, 13–17. doi: 10.1109/ISMSIT56059.2022.9932838.

Lei W, Jiawei W, Zezhou M. 2023. Enhancing grey wolf optimizer with levy flight for engineering applications. IEEE Access. 11:74865–74897. doi: 10.1109/ACCESS.2023.3295242.

Li L, Cheng Y, Ji W, Liu M, Hu Z, Yang Y, Wang Y, Zhou Y. 2023. Machine learning for predicting diabetes risk in western China adults. Diabetol Metab Syndr. 15(1):165. doi: 10.1186/s13098-023-01112-y.

Lin H, Wang C, Hao Q. 2023. A novel personality detection method based on high-dimensional psycholinguistic features and improved distributed gray wolf optimizer for feature selection. Inf Process Manage. 60(2):103217. doi: 10.1016/j.ipm.2022.103217.

Liu Q, Xue Y, Li G, Qiu D, Zhang W, Guo Z, Li Z. 2023. Application of KM-SMOTE for rockburst intelligent prediction. Tunnelling Underground Space Technol. 138: 105180. doi: 10.1016/j.tust.2023.105180.

Liu Y, As'arry A, Hassan MK, Hairuddin AA, Mohamad H. 2024. Review of the grey wolf optimization algorithm: variants and applications. Neural Comput Applic. 36(6): 2713–2735. doi: 10.1007/s00521-023-09202-8.

Lu SC, Swisher CL, Chung C, Jaffray D, Sidey-Gibbons C. 2023. On the importance of interpretable machine learning predictions to inform clinical decision making in oncology. Front Oncol. 13:1129380. doi: 10.3389/fonc.2023.1129380.

Mantovani RG, Horvath T, Cerri R, Vanschoren J, De Carvalho ACPLF. 2017. Hyper-Parameter tuning of a decision tree induction algorithm. Proceedings—2016 5th

Brazilian Conference on Intelligent Systems, BRACIS 2016, 37––42. doi: 10.1109/BRACIS.2016.018.
Mantovani RG, Horváth T, Rossi ALD, Cerri R, Junior SB, Vanschoren J, de Carvalho ACP, de LF. 2018. Better Trees: an empirical study on hyperparameter tuning of classification decision tree induction algorithms. http://arxiv.org/abs/1812.02207.
Mirjalili S, Mirjalili SM, Lewis A. 2014. Grey wolf optimizer. Adv Eng Softw. 69:46–61. doi: 10.1016/j.advengsoft.2013.12.007.
Mohakud R, Dash R. 2022. Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection. J King Saud Univy—Comput Inform Sci. 34(8):6280–6291. doi: 10.1016/j.jksuci.2021.05.012.
Moradi M, Modarres M, Sepehri MM. 2020. Detecting factors associated with polypharmacy in general practitioners' prescriptions: a data mining approach. Scientia Iranica. doi: 10.24200/sci.2020.55569.4283.
Nagpal A, Sabharwal M, Tripathi R. 2023. A novel ensemble machine learning framework for early stage diabetes mellitus prediction. Multidiscip Sci J. 6(3):2024031. doi: 10.31893/multiscience.2024031.
Parhi SK, Patro SK. 2024. Compressive strength prediction of PET fiber-reinforced concrete using Dolphin echolocation optimized decision tree-based machine learning algorithms. Asian J Civ Eng. 25(1):977–996. doi: 10.1007/s42107-023-00826-8.
Patro KK, Allam JP, Sanapala U, Marpu CK, Samee NA, Alabdulhafith M, Plawiak P. 2023. An effective correlation-based data modeling framework for automatic diabetes prediction using machine and deep learning techniques. BMC Bioinform. 24(1):372. doi: 10.1186/s12859-023-05488-6.
Perez LJ, Bebis G, McKenna SA, Parashar R. 2023. Solute transport prediction in heterogeneous porous media using random walks and machine learning. Int J Geomath. 14(1). doi: 10.1007/s13137-023-00240-x.
Priyanka Kumar D. 2022. Grey wolf optimization based hyper-parameter optimized convolution neural network for kidney image classification. Intern J Intel Engng Sys. 15(3):319–334. doi: 10.22266/ijies2022.0630.27.
Qi C, Tang X. 2018. Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study. Comput Ind Eng. 118:112–122. doi: 10.1016/j.cie.2018.02.028.
Ray S, Alshouiliy K, Roy A, Alghamdi A, Agrawal DP. 2020. Chi-squared based feature selection for stroke prediction using AzureML. Intermountain Engineering, Technology and Computing, IETC 2020. doi: 10.1109/IETC47856.2020.9249117.
Reza MS, Amin R, Yasmin R, Kulsum W, Ruhi S. 2024. Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data. Heliyon. 10(2):e24536. doi: 10.1016/j.heliyon.2024.e24536.
Saum N, Sugiura S, Piantanakulchai M. 2022. Hyperparameter optimization using iterative decision tree (IDT). IEEE Access. 10:106812–106827. doi: 10.1109/ACCESS.2022.3212387.
Site A, Nurmi J, Lohan ES. 2023. Machine-learning-based diabetes prediction using multisensor data. IEEE Sensors J. 23(22):28370–28377. doi: 10.1109/JSEN.2023.3319360.
Talebi H, Bahrami B, Daneshfar M, Bagherifard S, Ayatollahi MR. 2024. Data-driven based fracture prediction of notched components. Philos Trans A Math Phys Eng Sci. 382(2264):20220397. doi: 10.1098/rsta.2022.0397.
Teboul A. 2021. *Diabetes Health Indicators Dataset*. Kaggle. https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset.
Thaher T, Awad M, Aldasht M, Sheta A, Turabieh H, Chantar H. 2022. An enhanced evolutionary based feature selection approach using grey wolf optimizer for the classification of high-dimensional biological data. JUCS. 28(5):499–539. doi: 10.3897/jucs.78218.
Tian JW, Qi C, Peng K, Sun Y, Mundher Yaseen Z. 2022. Improved permeability prediction of porous media by feature selection and machine learning methods comparison. J Comput Civ Eng. 36(2). doi: 10.1061/(asce)cp.1943-5487.0000983.
Tran VL, Kim JK. 2024. Hybrid machine learning models for classifying failure modes of unstiffened steel plate girders subjected to patch loading. Structures. 59:105742. doi: 10.1016/j.istruc.2023.105742.
Ul Hassan I, Ali RH, Abideen Z. u, Ijaz AZ, Khan TA. 2023. Towards effective emotion detection: a comprehensive machine learning approach on EEG signals. BioMed Informatics. 3(4):1083–1100. doi: 10.3390/biomedinformatics3040065.
Woldesemayat B. 2019. Prevalence of diabetes mellitus and associated risk factors among adults attending at Feres Meda Health Centre, Addis Ababa, Ethiopia, 2017. IJBMR. 7(1):8. doi: 10.11648/j.ijbmr.20190701.12.
Wu X, Zhai F, Chang A, Wei J, Guo Y, Zhang J. 2023. Application of machine learning algorithms to predict osteoporosis in postmenopausal women with type 2 diabetes mellitus. J Endocrinol Invest. 46(12):2535–2546. doi: 10.1007/s40618-023-02109-0.
Xu Q, Yang H, Yuan X, Rong J, Zhao J. 2023. Enhanced grey wolf algorithm for automatic tuning of an ensemble neural network in predicting PCF optical properties. Opt Express. 31(26):43790–43803. doi: 10.1364/oe.503978.
Yagin FH, Yasar S, Gormez Y, Yagin B, Pinar A, Alkhateeb A, Ardigò LP. 2023. Explainable artificial intelligence paves the way in precision diagnostics and biomarker discovery for the subclass of diabetic retinopathy in Type 2 diabetics. Metabolites. 13(12):1204. doi: 10.3390/metabo13121204.
Yaqoob A, Musheer Aziz R, Verma NK. 2023. Applications and techniques of machine learning in cancer classification: a systematic review. Hum-Cent Intell Syst. 3(4):588–615. doi: 10.1007/s44230-023-00041-3.
Yaribeygi H, Farrokhi FR, Butler AE, Sahebkar A, Mather KJ, Lteif A, Steinberg HO, Baron AD, Ünal D, da Rocha Fernandes JD, et al. 2016. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Research and Clinical Practice. 39(11):1987–1995.
Yuvaraja M, Arunkumar S, Vinodh Kumar P, Mary Immaculate Sheela L. 2023. Improved grey wolf optimization- (IGWO-) based feature selection on multiview features and enhanced multimodal-sequential network intrusion detection approach. Wireless Commun Mobile Comput. 2023:1–13. doi: 10.1155/2023/8478457.

Zeinalnezhad M, Shishehchi S. 2024. An integrated data mining algorithms and meta-heuristic technique to predict the readmission risk of diabetic patients. Healthcare Analyt. 5:100292. doi: 10.1016/j.health.2023.100292.

Zhang J, Lv Y, Hou J, Zhang C, Yua X, Wang Y, Yang T, Su X, Ye Z, Li L. 2023. Machine learning for post-acute pancreatitis diabetes mellitus prediction and personalized treatment recommendations. Sci Rep. 13(1):4857. doi: 10.1038/s41598-023-31947-4.

Zhang Z, Lu Y, Ye M, Huang W, Jin L, Zhang G, Ge Y, Baghban A, Zhang Q, Wang H, et al. 2024. A novel evolutionary ensemble prediction model using harmony search and stacking for diabetes diagnosis. J King Saud Univ—Comp Inform Sci. 36(1):101873. doi: 10.1016/j.jksuci.2023.101873.

Zhou H, Xin Y, Li S. 2023. A diabetes prediction model based on Boruta feature selection and ensemble learning. BMC Bioinform. 24(1):224. doi: 10.1186/s12859-023-05300-5.

Zhu M, Li J, Li Z, Luo W, Dai D, Weaver SR, Stauber C, Luo R, Fu H. 2015. Mortality rates and the causes of death related to diabetes mellitus in Shanghai Songjiang District: An 11-year retrospective analysis of death certificates. BMC Endocr Disord. 15(1):45. doi: 10.1186/s12902-015-0042-1.