International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Binary Grey Wolf Optimizer in Diabetes Prediction

Andi Nugroho*[1], Harco Leslie Hendric Spits Warnars[2], Sani Muhamad Isa[3], Widodo Budiharto[4]

Submitted: 27/01/2024 **Revised:** 07/03/2024 **Accepted:** 14/03/2024

Abstract: Diabetes is a disease that is feared today because it has claimed quite a lot of victims. Thus knowing diabetes early can prevent death from diabetes. However, for diabetes detection classification tools, there is still no one that uses z-score normalization in the classification of the PIMA diabetes dataset, The purpose of this research is to be able to find a classification model for predicting diabetes from an early age. we will take a public dataset from the Pima Indians Diabetes Database (PIDD). The method that will be used in this study uses an optimization approach with the Improve Binary Grey Wolf Optimizer (BGWO) and Z-Score normalization, for testing it will use classification algorithms such as Support Vector Machine (SVM), Decision Tree (DT), k-Nearest Neighbors (KNN). The BGWO optimization algorithm and Z-Score normalization is used to optimize the performance of the SVM, DT, and KNN classification algorithm in producing the best accuracy value. The BGWO optimization algorithm and Z-Score normalization can increase the accuracy value from KNN = 76.5%, DT = 74.7%, and SVM = 77.7% to KNN + BGWO = 78%, DT + BGWO = 76.2% and SVM + BGWO = 79.8% in PIDD dataset.

Keywords: *BGWO, Classification, Diabetes, KNN, SVM*

1. Introduction

Diabetes is a disease caused by the occurrence of blood sugar levels in the body being very high. Diabetes has two types, the first type Diabetes mellitus is a chronic disease characterized by very high blood glucose levels, due to the inability of the pancreas to produce sufficient amounts of insulin [1]. It is estimated that about 10% of cases are in the first type. For the second type, our body cannot produce sufficient insulin threshold, which occurs in almost every diabetic patient. It is estimated that about 90% of all diabetes cases worldwide have cases of type 2 diabetes [2]. People with diabetes, in general, will experience a continuous thirst. Besides, people with diabetes will also feel hungry for glucose in their blood, and the patient will also feel the urge to urinate continuously. If these symptoms are not treated quickly, the patient will experience serious complications, including diabetic ketoacidosis and hyperglycaemic hyperosmolar nonketotic syndrome [3].

The typically ingested dose of glucose in the human body ranges from 70 to 99 milligrams per deciliter. However, if the glucose level in our body exceeds the limit of 126 mg/dl, it can be said that the person is included in the category of diabetic [4]. In this study, the public dataset is used with the following URL address [5] with the dataset name Pima Indians Diabetes (PIDD). In this study, an optimization algorithm approach will be used to maximize the use of features in the PIDD dataset [6]. The optimization algorithm used is the Binary Grey Wolf Optimizer (BGWO) algorithm. The BGWO algorithm comes from the Grey Wolf Optimizer (GWO) algorithm, where this algorithm is included in the metaheuristic algorithm. The GWO algorithm is inspired by a pack of grey wolves working together in search of prey [7]. The classification algorithms that will be used in this study are KNN, DT and SVM.

This paper consists of 5 sections as follows. Section 2 discusses related work. Section 3 discusses about the proposed method. Section 4 processes the data and presents the results of the experiments that have been carried out and Section 5 provides conclusions from the processes that have been carried out and the results obtained.

2. Related Work

Machine learning (ML) is a method that can be used to predict diseases, predict the weather, find the best way to reach a goal, make important decisions, and so on [8]. Another benefit of machine learning is its flexibility in increasing knowledge in making predictions and making better decisions from independent variables [9]. This research will discuss the optimization algorithm to determine feature selection, then the use of supervised learning in classifying diabetes. The classification methods to be used are Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbors (KNN). In addition, the optimization algorithm that will be used for feature search is the Binary Grey Wolf Optimizer (BGWO).

2.1. Binary Grey Wolf Optimizer (BGWO)

BGWO is an algorithm developed from the Grey Wolf Optimizer (GWO) algorithm; Mirjalili first discovered the GWO algorithm in 2014. Initially, this algorithm was created to get the best accuracy value by optimizing global search and existing datasets. This algorithm is based on the behaviour of the grey wolf in search of food sources[10][11]. In the GWO algorithm, there are four types of grey wolves, the highest order in the grey wolf group in hunting. The first type is alpha (α), beta (β), delta (δ), and omega (ω) [12]. In the grey wolf group, there are three

[1,2]*Computer Science Department, BINUS Graduate Program-Doctor of Computer Science, Bina Nusantara University, Jakarta, 11530, Indonesia*
[3]*Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, 11530, Jakarta, Indonesia*
[4] *Computer Science Department, School of Computer Science, Bina Nusantara University, 11530, Jakarta, Indonesia*
** Corresponding Author Email: andi.nugroho002@binus.ac.id*

highest types in the grey wolf pack, including α, β, δ then for ω will be the follower of his leadership, that is, the three leaders of the grey wolf [13]. α is the leader of a herd of grey wolves in search of prey for shelter to nesting tours, β is the assistant to the leader of the grey wolf pack, δ type beta are wolves that will be a substitute for types α, β and usually, he is good at hunting, as well as the strongest team when hunting, and finally type ω is the lowest type in the hierarchy of grey wolves [15]. And the formula for GWO can be seen in the following Eq. [16]:

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A}.\vec{D} \quad (1)$$

Where the value t represents the iteration and the value of $\vec{D}$ is the vector distance between the wolf's current position $\vec{X}(t)$ and the prey position $\overrightarrow{X_p}(t)$ and $\vec{X}$ indicates the position vector of a grey wolf. The formula can be seen in Eq. 2 below.

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_p}(t) - \vec{X}(t)\right| \quad (2)$$

Vector variables $\vec{A}$ and $\vec{C}$ are vector coefficients used in eq. 3 and 4, respectively [16-17].

$$\vec{\mathrm{A}} = 2\vec{\mathrm{a}}.\overrightarrow{\mathrm{r}_1} - \vec{\mathrm{a}} \quad (3)$$

$$\vec{\mathrm{C}} = 2.\overrightarrow{\mathrm{r}_2} \quad (4)$$

For variables $\overrightarrow{r_1}$,$\overrightarrow{r_2}$ are vectors randomized from the value [0,1], then $\vec{a}$ is the primary control of the parameter, this value will continue to decrease from value 2 to 0 as shown in Eq. 5 [19].

$$\vec{\mathrm{a}} = 2.\left(1 - \frac{\mathrm{t}}{\mathrm{T}}\right) \quad (5)$$

The variable t value is the current iteration value, and the T value is the total iteration value for each wolf that will update its respective position (X, Y) randomly based on the prey position(X', Y'). Vector $\vec{A}$ and $\vec{C}$ after the encircling process is complete, the hunting process begins with the following formula [20].

$$\begin{cases} \overrightarrow{\mathrm{D}_\alpha} = \left|\overrightarrow{\mathrm{C}_1}.\overrightarrow{\mathrm{X}_\alpha}(\mathrm{t}) - \vec{\mathrm{X}}(\mathrm{t})\right| \\ \overrightarrow{\mathrm{D}_\beta} = \left|\overrightarrow{\mathrm{C}_2}.\overrightarrow{\mathrm{X}_\beta}(\mathrm{t}) - \vec{\mathrm{X}}(\mathrm{t})\right| \\ \overrightarrow{\mathrm{D}_\delta} = \left|\overrightarrow{\mathrm{C}_3}.\overrightarrow{\mathrm{X}_\delta}(\mathrm{t}) - \vec{\mathrm{X}}(\mathrm{t})\right| \end{cases} \quad (6)$$

$$\begin{cases} \vec{\mathrm{X}}_1(\mathrm{t}+1) = \left|\overrightarrow{\mathrm{X}_\alpha}(\mathrm{t}) - \vec{\mathrm{A}}_1.\vec{\mathrm{D}}_\alpha\right| \\ \vec{\mathrm{X}}_2(\mathrm{t}+1) = \left|\overrightarrow{\mathrm{X}_\beta}(\mathrm{t}) - \vec{\mathrm{A}}_2.\vec{\mathrm{D}}_\beta\right| \\ \vec{\mathrm{X}}_3(\mathrm{t}+1) = \left|\overrightarrow{\mathrm{X}_\delta}(\mathrm{t}) - \vec{\mathrm{A}}_3.\vec{\mathrm{D}}_\delta\right| \end{cases} \quad (7)$$

In the iteration cycle above, the final position vector results are obtained

$$\vec{\mathrm{X}}(\mathrm{t}+1) = \frac{\vec{\mathrm{X}}_1(\mathrm{t}) + \vec{\mathrm{X}}_2(\mathrm{t}) + \vec{\mathrm{X}}_3(\mathrm{t})}{3} \quad (8)$$

$\overrightarrow{X_\alpha}(t)$, $\overrightarrow{X_\beta}(t)$, $\overrightarrow{X_\delta}(t)$are the vector variable represents the best locations of all types of wolves in the current iteration. And for the Binary Grey Wolf Optimizer formula (BGWO) [21].

$$\mathrm{X}_{\mathrm{binary}}(\mathrm{t}+1) = \mathrm{f}(\mathrm{x}) = \begin{cases} 1, \mathrm{sigmoid}\left(\frac{\vec{\mathrm{X}}_1(\mathrm{t}) + \vec{\mathrm{X}}_2(\mathrm{t}) + \vec{\mathrm{X}}_3(\mathrm{t})}{3}\right) \geq \mathrm{rand} \\ 0, \mathrm{otherwise} \end{cases} \quad (9)$$

The variable $X_{binary}(t+1)$ is an additional variable to provide an update on the location of the food source, and after that, the sigmoid function $S(X)$ is created in the formula Eq. 14 with the value of r being a random value of $\epsilon[0,1]$

$$\mathrm{S}(\mathrm{X}) = \frac{1}{1+\mathrm{e}^{(-10(\mathrm{x}-0,5))}} \quad (10)$$

Pseudocode of BGWO [22]

```
Parameter initialization BGWO
Randomly generates gray wolves positions
Calculate the fitness of each gray wolf
Find D_α, D_β, and D_δ
For i = 1: max_it do
   Update a, A, and C by eq. (3),(4),(5)
   Calculate position of each grey wolves by eq. (9),(10)
   Calculate the fitness of each grey wolf
   Update D_α, D_β, and D_δ
End for
Output D_α
```

2.2. Support Vector Machine (SVM)

Traditional Support Vector Machine (SVM) has been successfully used on existing classification problems. SVM has characteristics used in existing properties such as kernel, scarcity, and global solutions [23]. There are two main challenges to the SVM algorithm. First, primal SVM is used to solve quadratic programming problems on a large scale, leading to higher computational complexity [24]. Second, when linear independence cannot be met, it is difficult to achieve satisfactory performance with only one row or two parallel classification hyperplane [25][26]. SVM was created to solve binary classification problems [27][28]. SVM is a very sophisticated machine learning algorithm because the SVM algorithm defines the dataset with a hyperplane and uses an extraordinary level to classify the existing dataset points [29][30]. SVM is an algorithm that relies on hyperplane to separate classification data. The Optimal hyperplane has been determined as shown in eq.11[31].

$$\mathrm{w}^{\mathrm{t}}.\mathrm{x} + \mathrm{b} = 0 \quad (11)$$

w is the coefficient of the vector orthogonal to the hyperplane, and the variable b indicates the distance between the points in the dataset and the point origin. If it is known that the value of x is a point on a two-dimensional graph with the formula $y = ax + b$ then the variable x is changed to x_1, and y is changed to x_2, so the formula $ax_1 - x_2 + b = 0$ is simplified to $x = (x_1, x_2)$ and $w = (a, -1)$ then it will look like in Eq. 12.

$$\mathrm{w} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_\mathrm{i} \mathrm{y}_\mathrm{i} \mathrm{x}_\mathrm{i} \quad (12)$$

α_i is the Lagrange multipliers variable, and the constant n is the value of the vector.

$$\mathrm{y}_\mathrm{i}(\mathrm{w}^{\mathrm{t}}.\mathrm{x} + \mathrm{b}) - 1 = 0 \quad (13)$$

Based on Eq. 13, it can be concluded that the function of the linear discriminant is as follows

$$\hat{y}_\mathrm{i} = \mathrm{sgn}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_\mathrm{i} \mathrm{y}_\mathrm{i} \mathrm{x}^{\mathrm{t}} \mathrm{x}_\mathrm{i} + \mathrm{b}\right) (14)$$

And for non-linear functions are as follows

$$\hat{y}_\mathrm{i} = \mathrm{sgn}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_\mathrm{i} \mathrm{y}_\mathrm{i} \mathrm{K}(\mathrm{x}_\mathrm{i}, \mathrm{x}) + \mathrm{b}\right) \quad (15)$$

Where the value of $K(x_i, x)$ is determined from determining the value of the kernel function. For the Gaussian kernel function, it has a function like $K(x_i, x) = exp(-\gamma\|x - x_i\|^2)$ then for a polynomial function like this$K(x_i, x) = (x^t x_i + 1)^d$.

2.3. Decision Tree (DT)

Decision Tree (DT) is one of the algorithms included in supervised machine learning which is commonly used on datasets that have been labelled or often referred to as dataset classifications. Then the dataset is divided into tree branches and leaves which are described by giving true and false values [32]. The DT algorithm is very popular as a learning technique in statistics and machine learning [33]. In the DT algorithm, there are naming nodes including root nodes, internal nodes and leaf nodes. The root node is the node at the top of the DT, the internal node is the branch with the leaf node, and the leaf node is the last node. A schematic view of the DT can see in Fig. 1.

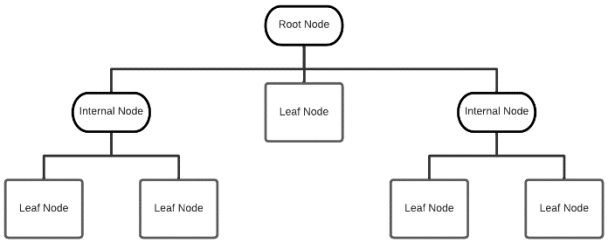

Fig.1. Schematic view of a decision tree [34]

In the DT algorithm there is a calculation formula called the entropy value and the information gain value. Both of these value calculations have their own criteria, because entropy values are usually used to measure impurities in existing datasets. Then the gain information value is usually used to segment the dataset that has been selected for classification. The DT formula can be seen in Eq. 16 [35].

$$\text{Entropy}(S) = \sum_{i=1}^{C} p_i \log 2^{p_i} (16)$$

Where variable S is the entropy of the classification set to the value of variable C, then the value of the variable p_i is the ratio of the number of subsets of the dataset and $i-th$ is the attribute value in the dataset.

$$\text{Gain}\ (S, A) = \sum V \in V(A) \frac{|S_V|}{S} \text{Entropy}(S_V) \quad (17)$$

The value of the Gain variable (S,A) is obtained from equation 16, where the value of attribute A is included in the set $V(A)$ and S_V is part of the value of variable S and the value of variable S is equal to the value of variable V.

2.4. *K-Nearest Neighbors* (KNN)

The KNN algorithm is usually widely used by researchers in the fields of pattern recognition, classification and regression research [36][37]. The KNN algorithm will generally assign a weight value to the K variable which will be used in determining the number of neighbors [38]. KNN has many distance functions that are used to measure the distance between two different data samples, one of which is the Euclidean Distance function [39]. Then the Euclidean Distance formula is like the Eq. 18.

$$d_B(x, y) = \sqrt{\sum_{a \in B} (x_a - y_a)^2} (18)$$

B is a subset of features where $B(B \subseteq C)$ value x, y are the variables of the two sample datasets and x_a consists of the variable x which is the reference value a.

2.5. Z-Score Normalization

Z-Score normalization is a measure of the divergence of an observation in a dataset to find results that are close to the original value taken from the average value [40]. For the formula as shown in eq. 19.

$$Z = (X - \mu)/\sigma \quad (19)$$

The z value is a normalized Z-Score. Where X is the value to be searched, is the mean value of the feature column in the PIDD dataset, and is the standard deviation value obtained through the function.

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \quad (20)$$

The variable N is the total data from the feature, and x_i is the value of the dataset to be changed.

2.6. Dataset

The dataset used in this study is a dataset obtained from the public dataset with the URL: PIDD. This dataset describes people with diabetes, especially women from a minimum age of 21 years to 81 years, by looking at the features of the number of Pregnancies, Glucose content, Blood Pressure, Skin Thickness, and Insulin use. As seen from body weight Body Mass Index (BMI), Diabetes Pedigree Function, Outcome is a classification score of 0 is a woman who does not indicate diabetes, but for a value of 1 it is the opposite. With a total number of lines reaching 768 and there are 9 features in the PIDD dataset. Table 1 shows a brief dataset from rows 1 to 5 and from rows 764 to 768.

Table 1 shows that there are several row with a value of 0, so this 0 data will then be deleted to get the best accuracy value. The process will be explained in the following process in the pre-processing phase of the methodology.

Table 1. Information Dataset PIDD[41]

No	*Feature*	*Description of the attribute*
1	Pregnancies	Number of Times
2	Age	Years
3	Glucose	Plasma glucose concentration (mg/dL)
4	Insulin	2-hour serum insulin (mu U/ml)
5	BMI	Body mass index(weight in kg/(height in m).^{2}
6	Blood Pressure	Diastolic blood pressure (mm Hg)
7	Skin Thickness	Triceps skin fold thickness(mm)
8	Pedi	Diabetes Pedigree Function
9	Outcome	1: Diabetic, 0: Non Diabetic

3. Methodology

The methodology in this study uses an optimization algorithm approach and a classification algorithm. This study will use three classification algorithm comparisons: SVM, Decision Tree (DT), and K-Nearest Neighbors (KNN). The optimization algorithm that will be used in this research is BGWO. This algorithm will select the best features to produce good accuracy and minimize excessive memory and processor usage when carrying out the classification process. The stage of this research begins with a dataset search. The dataset used is the PIDD dataset. Then enter the next stage, namely pre-processing, wherein in the pre-processing stage, there are processes such as finding missing values in the dataset, normalizing with the Z-Score, and using the feature selection algorithm with BGWO. The next step is to use a

classification algorithm that will be used to obtain accuracy, precision, recall, and F1-Score values. A classification algorithm is used to determine the outcome of those with diabetes with results that are not included in people with diabetes. The algorithm will conduct training and testing of the dataset to produce the best accuracy value. The PIDD dataset is a public dataset that anyone can use. In this research, all processes from feature selection and classification algorithms use Matlab 2018 tools.

The research methodology used is a machine learning approach, where the process begins with data pre-processing, Z-Score normalization, feature selection, and classification algorithms (SVM, KNN, DT), which can be seen in figure 2. The dataset obtained through the public dataset with the dataset name PIDD is then processed to remove data that has a value of 0 with an example, as shown in table 2 below.

Table 2. Dataset PIDD Missing Value

No	*Name Of Feature*	*Total Delete row*
1	Pregnancies	111
2	Glucose	5
3	Blood Pressure	25
4	Skin Thickness	172
5	Insulin	119

Number of Iterations

Table 2 is part of the pre-processing process to find the missing values in the dataset. In this study, the missing values are searched for a value of 0, and the row with a value of 0 for each feature will be deleted. The value 0 that is not deleted is only in the outcome feature because the outcome is a classification value from the PIDD dataset. The classification algorithm will use this outcome value to obtain accuracy, precision, recall, and F1 score. After obtaining a dataset value without a value of 0, the next step is normalization. For the normalization process, the Z-Score normalization formula is used. Why does the dataset have to be normalized? It equates the value of each feature so that the gap is not too big. If there is a large difference in gap values in the dataset, it will result in poor accuracy results in the classification algorithm; therefore, a normalization process is needed in the dataset preprocessing section. Next is the process of finding feature selection that will be used in the three classification algorithms, for feature selection algorithms using BGWO to get the best features for use by classification algorithms such as SVM, KNN, and DT.

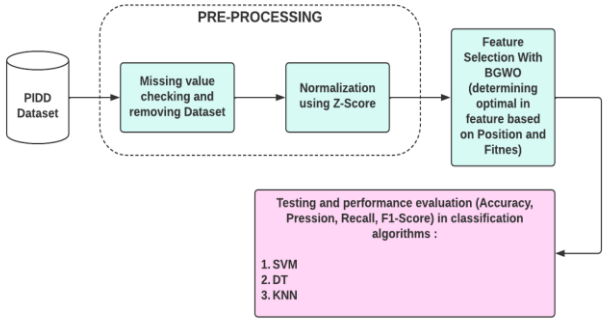

Fig 2. Research Methodology

4. Result

The PIDD dataset is obtained from the public dataset and consists of 768 rows and 9 features. Before the dataset is used, the first thing to do is to remove the value 0 in each existing row.

4.1. Feature Selection with BGWO

After the normalization process is complete, the next step is to enter the feature selection stage. This stage is the final stage of data pre-processing. At this stage, the parameters of the BGWO algorithm will be set as follows.

Table 3. Set Parameter BGWO

BGWO Parameter	Value
N	14
Max Iteration	100

The value of N is the number of wolves used in the BGWO algorithm, and the parameter value is the value tested previously to get the best value in the PIDD dataset. Then Max Iteration is the maximum repetition in the BGWO algorithm. The PIDD data set has 9 features where each feature is related to each other. The BGWO algorithm is used on the PIDD data set to find relationships between these features. The relationship between features will be used in the classification algorithm. The feature selection results from the BGWO algorithm will be used with the aim of increasing the accuracy, precision, recall and F1-score values. The BGWO algorithm used in the PIDD data set consists of 9 features, but with the BGWO algorithm there are only 5 features as seen in table 4.

Table 4. Feature Selection BGWO

No	Feature	Feature Choice
1	Pregnancies	•
2	Age	•
3	Glucose	•
4	Insulin	
5	BMI	
6	Blood Pressure	
7	Skin Thickness	•
8	Diabetes Pedigree Function	
9	Outcome	•

Table 4 shows that the features that will be used in the classification algorithm consist of five features, including pregnancy, age, glucose, insulin, BMI, blood pressure, skin thickness, diabetes pedigree function, and outcome. The feature selection results show that the relationship between features in the PIDD data set consists of five features. These five features are then used in the classification algorithm to obtain accuracy, precision, recall and F1 score results. Figure 3 below shows the fitness value which is close to 0, namely 0.114286. If the fineness value is close to 0, the accuracy value will be better.

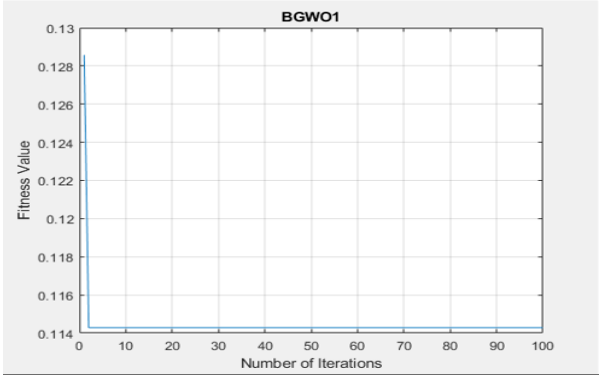

Fig 3. Fitness Value BGWO PIDD

4.2. Classification Algorithm

The feature selection process in the BGWO algorithm has carried

out the PIDD dataset. Next, checking the data pattern first uses a scatter plot with the appearance as shown in Figure 3.

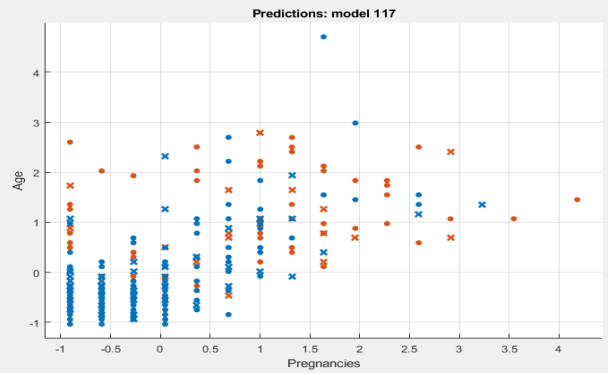

Fig 3. Scatter Plot dataset PIDD

From these data, it can be concluded that the data distribution looks uneven, which causes the process of identifying the correctness of the prediction error to be relatively high. However, choosing a good classification algorithm will also provide the best accuracy. In addition, determining parameters for each classification algorithm increases the current accuracy value. As for determining the parameters of each classification algorithm can be seen in table 5.

Table 5. Parameter Classification Algorithm

Parameter	*SVM*	*KNN*	*DT*
Kernel Function	Quadratic		
Kernel Scale	3		
Box Constraint Level	5		
Multiclass method	One-vs-one		
Preset		Cosine KNN	Fine Tree
Number of Neighbors		7	
Distance Metric		Minkowski	
Distance Weight		Equal	
Maximum number of splits			6
Split Criterion			Gini's diversity index
Surrogate Decision Splits			Off

This parameter has been tested on the PIDD dataset. For testing the BGWO feature selection algorithm and Z-score normalization using the cross-validation method. Where this method is used to determine the training and testing of the PIDD dataset with a cross-validation value of 5. Then the results of the classification algorithm testing are shown in table 6.

Table 6. Comparison Between SVM, KNN, and Decision Tree Algorithms

No	*Pima Indians Diabetes Dataset*				
	Algorithm	Accuracy	Precision	Recall	F1-Score
1	KNN	76.5%	88%	65.19%	74.89%
2	DECISION TREE	74.7%	81%	68.07%	73.97%
3	SVM	77.7%	89%	66.42%	76.07%
4	KNN+BGWO	78%	72.97%	81%	76.78%
5	DECISION TREE+BGWO	76.2%	69.49%	82%	75.23%
6	SVM+BGWO	79.8%	75.93%	82%	78.85%

Table 6 shows that the results of the normalization process, feature selection, and classification algorithm obtained the highest result using the SVM classification algorithm with an accuracy value of 79.8%. This is because the classification process has gone through several stages above. This value is generated because of the feature selection algorithm BGWO which increases the current accuracy value. The SVM precision value is also high, which means errors occur with truth predictions above 75.93%. In addition, using Z-Score normalization also increases the accuracy of the SVM algorithm.

5. Conclusion and Future Work

From the results of the research that has been done, it can be concluded that the use of the SVM+BGWO classification algorithm is the best combination of algorithms from the KNN+BGWO and DT+BGWO algorithms. In future research, the BGWO algorithm will be developed with the Sine Cosine Algorithm (SCA), (SCA-BGWO) algorithm to be able to increase the fitness value of the BGWO algorithm.

References

[1] N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis, and K. Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction," *IEEE Access*, vol. 9, pp. 103737–103757, 2021.

[2] R. Rajni and A. Amandeep, "RB-Bayes algorithm for the prediction of diabetic in Pima Indian dataset," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 6, p. 4866, Dec. 2019.

[3] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Comput. Sci.*, 2018.

[4] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021.

[5] @data-society, "Pima Indians Diabetes Database," 2016. [Online]. Available: https://data.world/data-society/pima-indians-diabetes-database/workspace/file?filename=diabetes.csv. [Accessed: 22-Nov-2022].

[6] M. H. Nadimi-Shahraki, S. Taghian, and S. Mirjalili, "An improved grey wolf optimizer for solving engineering problems," *Expert Syst. Appl.*, vol. 166, p. 113917, 2021.

[7] M. Shariati *et al.*, "A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement," *Eng. Comput.*, vol. 38, no. 1, pp. 757–779, 2022.

[8] R. Liu and C. S. Pun, "Machine-Learning-enhanced systemic risk measure: A Two-Step supervised learning approach," *J. Bank. Financ.*, vol. 136, 2022.

[9] N. Abellán, E. Baquedano, and M. Domínguez-rodrigo, "marks using machine learning methods and computer vision algorithms □," *Geobios*, 2022.

[10] M. Abdel-Basset, K. M. Sallam, R. Mohamed, I. Elgendi, K. Munasinghe, and O. M. Elkomy, "An Improved Binary Grey-Wolf Optimizer with Simulated Annealing for Feature Selection," *IEEE Access*, vol. 9, pp. 139792–139822, 2021.

[11] A. M. Helmi, M. A. A. Al-Qaness, A. Dahou, R. Damaševičius, T. Krilavičius, and M. A. Elaziz, "A novel hybrid gradient-based optimizer and grey wolf optimizer feature selection method for human activity recognition using smartphone sensors," *Entropy*, vol. 23, no. 8, 2021.

[12] M. Roayaei, "On the binarization of Grey Wolf optimizer: a novel binary optimizer algorithm," *Soft Comput.*, vol. 25, no. 23, pp. 14715–14728, 2021.

[13] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili,

and H. Alhussian, "Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019.

[14] T. Thaher, M. Awad, M. Aldasht, A. Sheta, H. Turabieh, and H. Chantar, *An Enhanced Evolutionary Based Feature Selection Approach Using Grey Wolf Optimizer for the Classification of High-dimensional Biological Data*, vol. 28, no. 5. 2022.

[15] X. Sun, Z. Jin, Y. Cai, Z. Yang, and L. Chen, "Grey Wolf Optimization Algorithm Based State Feedback Control for a Bearingless Permanent Magnet Synchronous Machine," *IEEE Trans. Power Electron.*, vol. 35, no. 12, pp. 13631–13640, 2020.

[16] E. Momanyi and D. Segera, "A Master-Slave Binary Grey Wolf Optimizer for Optimal Feature Selection in Biomedical Data Classification," *Biomed Res. Int.*, vol. 2021, 2021.

[17] Q. Al-Tashi *et al.*, "Binary Multi-Objective Grey Wolf Optimizer for Feature Selection in Classification," *IEEE Access*, vol. 8, pp. 106247–106263, 2020.

[18] D. Segera, M. Mbuthia, and A. Nyete, "An excited binary grey Wolf optimizer for feature selection in highly dimensional datasets," *ICINCO 2020 - Proc. 17th Int. Conf. Informatics Control. Autom. Robot.*, no. Icinco, pp. 125–133, 2020.

[19] X. Li and K. M. Luk, "The Grey Wolf Optimizer and Its Applications in Electromagnetics," *IEEE Trans. Antennas Propag.*, vol. 68, no. 3, pp. 2186–2197, 2020.

[20] J. Hu *et al.*, "Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection," *Knowledge-Based Syst.*, vol. 213, p. 106684, 2021.

[21] R. Al-Wajih, S. J. Abdulkadir, N. Aziz, Q. Al-Tashi, and N. Talpur, "Hybrid binary grey Wolf with Harris hawks optimizer for feature selection," *IEEE Access*, vol. 9, pp. 31662–31677, 2021.

[22] P. Hu, J. S. Pan, and S. C. Chu, "Improved Binary Grey Wolf Optimizer and Its application for feature selection," *Knowledge-Based Syst.*, vol. 195, p. 105746, 2020.

[23] P. Vijayaragavan, R. Ponnusamy, and M. Aramudhan, "An optimal support vector machine based classification model for sentimental analysis of online product reviews," *Futur. Gener. Comput. Syst.*, vol. 111, pp. 234–240, 2020.

[24] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, no. xxxx, 2020.

[25] C. Yuan, L. Yang, and P. Sun, "Correntropy-based metric for robust twin support vector machine," *Inf. Sci. (Ny).*, vol. 545, pp. 82–101, 2021.

[26] C. Jimenez-Castaño, A. Alvarez-Meza, and A. Orozco-Gutierrez, "Enhanced automatic twin support vector machine for imbalanced data classification," *Pattern Recognit.*, vol. 107, 2020.

[27] Z. Che, B. Liu, Y. Xiao, and H. Cai, "Twin support vector machines with privileged information," *Inf. Sci. (Ny).*, vol. 573, pp. 141–153, 2021.

[28] R. B. Lukmanto, Suharjito, A. Nugroho, and H. Akbar, "Early detection of diabetes mellitus using feature selection and fuzzy support vector machine," *Procedia Comput. Sci.*, vol. 157, pp. 46–54, 2019.

[29] I. G. Lee, Q. Zhang, S. W. Yoon, and D. Won, "A mixed integer linear programming support vector machine for cost-effective feature selection," *Knowledge-Based Syst.*, vol. 203, p. 106145, 2020.

[30] A. Gejji, S. Shukla, S. Pimparkar, T. Pattharwala, and A. Bewoor, "Using a support vector machine for building a quality prediction model for center-less honing process," *Procedia Manuf.*, vol. 46, no. 2019, pp. 600–607, 2020.

[31] C. Mallika and S. Selvamuthukumaran, "A Hybrid Crow Search and Grey Wolf Optimization Technique for Enhanced Medical Data Classification in Diabetes Diagnosis System," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, 2021.

[32] I. Stancic, L. Veic, J. Music, and T. Grujic, "Classification of Low-Resolution Flying Objects in Videos Using the Machine Learning Approach," *Adv. Electr. Comput. Eng.*, vol. 22, no. 2, pp. 45–52, 2022.

[33] A. Plaia, S. Buscemi, J. Fürnkranz, and E. L. Mencía, "Comparing Boosting and Bagging for Decision Trees of Rankings," *J. Classif.*, vol. 39, no. 1, pp. 78–99, 2022.

[34] L. Vanfretti and V. S. N. Arava, "Decision tree-based classification of multiple operating conditions for power system voltage stability assessment," *Int. J. Electr. Power Energy Syst.*, vol. 123, no. October 2019, p. 106251, 2020.

[35] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology …*. jastt.org, 2021.

[36] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Syst. Appl.*, vol. 164, no. July 2019, p. 113981, 2021.

[37] S. Li, K. Zhang, Q. Chen, S. Wang, and S. Zhang, "Feature Selection for High Dimensional Data Using Weighted K-Nearest Neighbors and Genetic Algorithm," *IEEE Access*, vol. 8, pp. 139512–139528, 2020.

[38] S. Zhang, "Challenges in KNN Classification," *IEEE Trans. Knowl. Data Eng.*, pp. 1–13, 2021.

[39] S. Li, K. Zhang, Y. Li, S. Wang, and S. Zhang, "Online streaming feature selection based on neighborhood rough set," *Appl. Soft Comput.*, vol. 113, p. 108025, 2021.

[40] V. Aggarwal, V. Gupta, P. Singh, K. Sharma, and N. Sharma, "Detection of spatial outlier by using improved Z-score test," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, vol. 2019-April, no. Icoei, pp. 788–790, 2019.

[41] R. Patra and B. khuntia, "Analysis and Prediction Of Pima Indian Diabetes Dataset Using SDKNN Classifier Technique," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1070, no. 1, p. 012059, 2021.