Improvement of grey wolf optimizer with adaptive middle filter to adjust support vector machine parameters to predict diabetes complications

Fereshteh Jeyafzam¹ · Babak Vaziri¹ · Mohsen Yaghoubi Suraki² · Ali Asghar Rahmani Hosseinabadi³ · Adam Slowik⁴

Received: 7 May 2020 / Accepted: 21 May 2021 / Published online: 12 June 2021
© The Author(s) 2021

Abstract

In medical science, collecting and classifying data from various diseases is a vital task. The confused and large amounts of data are problems that prevent us from achieving acceptable results. One of the major problems for diabetic patients is a failure to properly diagnose the disease. As a result of this mistake in diagnosis or failure in early diagnosis, the patient may suffer from complications such as blindness, kidney failure, and cutting off the toes. Nowadays, doctors diagnose the disease by relying on their experience and knowledge and performing complex and time-consuming tests. One of the problems with current diabetic, diagnostic methods is the lack of appropriate features to diagnose the disease and consequently the weakness in its diagnosis, especially in its early stages. Since diabetes diagnosis relies on large amounts of data with many parameters, it is necessary to use machine learning methods such as support vector machine (SVM) to predict the complications of diabetes. One of the disadvantages of SVM is its parameter adjustment, which can be accomplished using metaheuristic algorithms such as particle swarm optimization algorithm (PSO), genetic algorithm, or grey wolf optimizer (GWO). In this paper, after preprocessing and preparing the dataset for data mining, we use SVM to predict complications of diabetes based on selected parameters of a patient acquired by laboratory test using improved GWO. We improve the selection process of GWO by employing dynamic adaptive middle filter, a nonlinear filter that assigns appropriate weight to each value based on the data value. Comparison of the final results of the proposed algorithm with classification methods such as a multilayer perceptron neural network, decision tree, simple Bayes, and temporal fuzzy min–max neural network (TFMM-PSO) shows the superiority of the proposed method over the comparable ones.

Keywords Diabetes · Support vector machine · Adaptive middle filter · Feature selection · Grey wolf optimizer

1 Introduction

As the twenty-first century progresses, we are witnessing globalization, changes in people's lifestyles and industrialization, one of the consequences of which is a change in the pattern of diseases [1]. Until recently, contagious diseases were considered to be the major health problem in third world countries, but now the increasing role of non-contagious diseases in mortality, especially in developing countries, is a serious threat. Diabetes is one of the most important diseases in this group [2]. Diabetes is a chronic endocrine disorder characterized by a malfunction in glucose metabolism due to problems with the production or

utilization of insulin hormone. The long-term risks of diabetes are extremely serious for health, such as premature death, blindness, loss of organs if gangrene is not controlled, and impotence. Patients that require insulin treatment and whose disease has begun in childhood, adolescence, or early adulthood are at risk for such problems [1].

Self-care behavior, which is a key concept in health promotion, refers to decisions and activities that a person can use to adapt to a health problem or improve his health. Self-care behaviors prevent early and late complications of the disease and guarantee a long life for the patient. In diabetes, self-care is one of the most important factors for controlling the disease. Empowerment and acceptance status are personality factors that affect patients' status and

Extended author information available on the last page of the article

increase their ability to deal with problems such as illnesses. According to existing studies, the most important predictor of mortality in diabetic patients is lack of self-care [3].

Nowadays, it is important in medical science to collect a great deal of data on various diseases. Medical centers collect this data for a variety of goals. Researching these data to obtain useful results and models for diseases is one of the goals of using these data. A large amount of data and confusion resulting from that is a problem that prevents us from achieving acceptable results. Data mining is therefore used to overcome this problem and find useful relationships between risk factors in diseases [1].

The intensity of competition in the scientific, social, economic, political, and military fields has also increased the importance of speed or time of access to information. Therefore, the need to design systems that are capable of quickly discover interest information to users, with a focus on minimal human intervention, on the one hand, and approaching analysis methods proportional to the volume of bulk data, on the other, is well sensed. At present, data mining is the most important technology for the efficient, accurate, and rapid processing of bulk data, and its importance is increasing. Data mining is a bridge between statistics, computer science (CS), artificial intelligence (AI), pattern recognition (PR), and data machine learning. Data mining is a complex process for identifying the correct, new and potentially useful patterns and models in a large amount of data, so that these patterns and models are understandable to humans [4].

Data mining is not a product that can be purchased but is a scientific process that should be implemented as a project. Data are often bulky and cannot be used alone, but the hidden knowledge in the data can be used. Therefore, utilizing the power of data mining processes to identify patterns and models as well as the relationship between different elements in the database to discover the knowledge behind the data and ultimately convert the data into information becomes more and more essential. Data mining usually refers to the discovery of useful patterns among the data. A useful pattern is a model of data that describes the relationship between a subset of data and is valid, simple, understandable, and new [4].

In the information age, data are one of the most important assets of any organization. However, data can become a valuable resource for the organization when used correctly. To transform the potential value of data into usable information and knowledge, many organizations have adopted “data mining”. Because through data mining, it will be possible to discover the relationships, trends, and patterns hidden among data and gaining new knowledge in the field of explicit and latent organizational challenges [5].

In this paper, we try to create a data mining system that can first preprocess the collected data by laboratory tests of 1573 patients in the endocrinology department of Mazandaran University of Medical Sciences. Secondly, we use the one-versus-all method of SVM classifier to predict the type of disease based on the medical data of every patient into seven different diabetic complications, namely eye problem complication, high-blood-pressure complication, dialysis history complication, heart attack complications, stroke complications, diabetes foot ulcer complication, and diabetes coma complication. Thirdly, we improve the accuracy of the SVM method by feeding selected features of a patient using improved grey wolf optimizer (GWO). The improved GWO uses weighted adaptive middle filter (WAMF) at each step of the algorithm implementation, to filter the outliers (wolves far from the target) through a dynamic window. GWO algorithm [6] is a part of swarm intelligence algorithms [7]. These algorithms are widely used in many other practical application [8–10]. In this paper, we show how the GWO algorithm can be used in the medical area.

In brief, the structure of the paper is organized as follows: In Sect. 2, related work is presented. The proposed method is fully described in Sect. 3. The simulation results of the proposed algorithm and conclusion are summarized in Sects. 4 and 5, respectively.

2 Related work

Until now, many classification methods have been proposed for diabetes diagnosis problems that can be broadly classified into four major categories.

Artificial neural network (ANN)-based categories of classification method is the most frequently used method reported in the literature. In 2007, Anbananthan et al. [11] used ANN and DT made of C4.5 algorithm to diagnose diabetes in individuals based on features such as age and blood pressure. In 2008, Chan et al. [12] have studied the microvascular complications of diabetes. To do so, he compared the C5.0 algorithm and the multilayer perceptron neural network (MLP NN). Different factors have been identified for each of these complications, and their effect on each complication has been studied. Patil and Durga [13] have used the a priori algorithm to create turbulence rules for finding hidden relationships between variables. In 2009, Fang [14] has used various data mining techniques to cluster patients with diabetes. Important features considered in this study are age, family history, and weight. The accuracy of the model created using clustering is 80%. In 2014, Ganapathy et al. [15] propose a pattern classification system by combining temporal features with fuzzy min-max (TFMM) neural network-based classifier for effective

decision support in medical diagnosis. In this work, a particle swarm optimization (PSO) algorithm-based rule extractor is proposed for improving the detection accuracy. Accuracy of the proposed TFMM-PSO method is compared with other methods [16–20] using the University of California Irvine (UCI) Machine Learning Repository Dataset [21]. Most of the reviewed methods lack in selecting a proper number of features that make the classifiers slow.

Decision tree-based algorithms can be categorized as the second batch of methods used for diabetic prediction. Breault et al. [22] performed the classification and analysis of regression using the classification and regression tree (CART) system in 2002 and deduced the dependency between a series of features. The classification accuracy was 59.9%. Miyaki et al. [23] also have used the card method to judge the factors influencing the incidence of diabetes in 2002. Rohlfing et al. [24] used linear regression analysis to examine the relationship between type 1 diabetes and HbA1c in 2002. Silverstein et al. [25] performed experiments on three medical databases and produced rules and then compared these rules with predetermined rules.

Trautvetter et al. [26] have used the association rule and decision tree (DT) to extract knowledge from the medical database. Juan et al. [27] have developed a type 2 diabetes data processing system (DDPS) using a combination of C4.5 and EM (maximum expectation) algorithms in 2007. Jarullah [28] has used the DT to diagnose type 2 diabetes. DT is generated using J48 decision tree classification algorithm (DTCA) in Weka software. Aljumah et al. [29] have used regression to analyze the prediction of diabetes treatment in two groups of young and old ages based on drug treatment and side effects. Antonelli et al. [30] have proposed a multi-level clustering-based analysis framework for identifying treatment pathways and examining patients for specific diseases. The proposed method has worked well in identifying groups of patients with similar disease history and increasing the severity of their complications. All decision tree-based algorithms need prior knowledge about different classes that require many annotated samples by experts to design the tree.

SVM-based algorithms are the third type of methods that we discovered in our literature review. In 2007, Huang et al. [31] conducted a study on identifying the major factors affecting diabetes controlling by using feature selection in the patient management system. In 2008, Han et al. [32] predicted diabetes in the patient database using Rapid Miner software and ID3 decision tree algorithm (DTA). In 2007, Cho et al. [33] predicted the presence of neuropathy in diabetic patients using SVM classification, feature selection, and visualization. In [34], authors have

attempted to diagnose diabetes using data mining algorithms that are very important in diagnosis and prediction. In this study, SVM, k-nearest neighbor, Bayes network (BN), ID3, C4.5, C5.0, and CART are used for diabetes detection. In this study, 768 diabetic patients from the PID dataset with 8 important features are used to train and test the data, 80% of which are used as training data and 20% are used as test data. The results show that the SVM model is more accurate than other algorithms, and has an accuracy of 81.77%. Han et al. [35] have developed a batch system for the diagnosis of diabetes. They specifically used the SVM to diagnose diabetes. In this study, SVM is used to screen diabetes, while at the same time a group learning module is added to make the black box related to SVM decisions more comprehensive and transparent. In addition, this scheme is a useful and appropriate method to solve the imbalance problem. Radha and Srinivasan [36] have used three classification methods to predict diabetes. This study compares the results of five supervised data mining algorithms using five performance criteria. The three algorithms are C4.5, SVM, and k-nearest neighbor. The performance of data mining algorithms is compared based on accuracy, computation time, and bootstrap accuracy. This study describes the algorithmic discussion of the UCI dataset for this disease in the large dataset repository. In [37, 38], authors have used hybrid methods for feature selection and SVM for classification. In the existing databases, there are some not-so distinct and redundant features. These features are major contributing factors to the success of the classification tool and system processing time. The system developed in this study has attempted to increase system speed and success by eliminating these redundant features. Therefore, the purpose of this study is to investigate the effect of removing unnecessary and obsolete features from the dataset on classification success by using an SVM classifier. The feature selection algorithm based on the Bee Colony Optimization Algorithm (BCOA) developed in this study is the first sample of the BCOA used in feature selection. We also choose to use SVM in order to classify diabetic complications. However, we find that using SVM alone is not very accurate, so that we improve the method by selecting relevant features using an improved GWO method of optimization.

3 The proposed data mining system

In this section, we discuss the preprocessing method and the improved GWO method and also feature selection part of SVM classifier, totally called a complete data mining system.

3.1 Data aggregation

Required data are collected from the endocrinology department of Mazandaran University of Medical Sciences. The file information is from the second half of the year 2015. There are 1573 initial records of patients, 53 of which lack of complete information. The average age of patients is 53 years, and 30% are male and the rest are female. 70% of patients have a family history of diabetes. The laboratory features of the patients are evaluated and identified at this stage. For each patient, 23 features including name, family, file number, address, height, weight, age, body mass index, gender, heredity, maximum blood pressure, minimum blood pressure, education, fasting blood sugar, 2-h blood sugar, cholesterol, harmful fat, useful fat, triglyceride, blood urea, creatinine, activity rate, tobacco use, and 8 complications including high blood lipids, eye complication, high blood pressure, dialysis history, cardiac problems, stroke, diabetic foot ulcer, diabetic coma have been registered.

3.2 Preprocessing

Preprocessing in data mining usually involves in data cleaning (DC), data integration (DI), data reduction (DR), and data transformation (DT). In the real world, data are not always perfect, and as for medical information, this is always true. Therefore, if the quality of the data is not good enough, some steps of the preprocessing should be performed on data to improve the quality of data and deliver high-quality data to the data mining algorithm to minimize the impact of data weakness. Usually, data preprocessing and preparation consumes more than 70% of the time required for data mining and 75–90% of the success of data mining projects depends on that. In this study, Naim software (<https://www.knime.com/>) is used for data preprocessing and preparation. Table 1 describes details of the dataset.

In this work, we employ DC and DT of preprocessing steps that are discussed in the following subsections.

3.2.1 Data cleaning (DC)

DC that is sometimes called data cleansing is the process of detecting, deleting, or correcting corrections in a database that have some errors and focuses on quantifying, or removing null attributes, balancing noisy values, detecting and deleting out-of-bounds values.

Table 1 Field description of the Mazandaran University of Medical Sciences dataset

Field number	Description
1	Age
2	Sex
3	BMI
4	Inherit
5	SBP
6	DBP
7	Education
8	FBS
9	2hpp-BS
10	TC
11	LDL
12	HDL
13	TG
14	BUN/CR
15	Activity
16	Dispeplimi
17	Eye problem
18	High pressure
19	Dialyze
20	Heart problem
21	Stroke
22	Foot ulcers
23	DKA
24	Smoke

In this paper, some information such as name, family, file number, and address is removed from the file. Next, we excluded the records of patients who have incomplete test information such as cases that had zero values for blood pressure, fasting blood glucose, blood glucose 2 h after the meal, and triglyceride because of the impact of these features on the final result. Basically, incomplete test results occurred in two cases. In the first case, the patient does not cooperate and does not complete the test process. In the second case, the patients are not recognized as diabetic patients based on their first visit. Chen and Astebro [39] proved that rational deletion is an efficient way of replacing important features with techniques such as mean, random assignment, regression assignment, and Bayesian models. Samples with several missing features are also deleted, and other missing values are initialized using common and probable values. Some features, such as blood urea and creatinine, are not important alone; if the ratio is between 10 mg/dL (milligrams per deciliter) and 20 mg/dL, the normal condition is reported and more than 20 mg/dL

means gastrointestinal bleeding or urinary tract obstruction. The ratio of these features indicates the likelihood of a kidney complication. Height and weight are not important lonely, but their body mass index is effective. As a result, these features have been removed and related indicators have been used instead.

Classification is a data preprocessing technique that minimizes the impact of minor errors that occur when receiving data. So classification is used to solve the noise problem in data. Data can be categorized in various ways, and then the data of each category can be represented in a more general sense. Based on reputable scientific and medical resources and sites and with the approval of a specialist physician, features such as body mass index, systolic blood pressure, diastolic blood pressure, fasting blood sugar, 2-h postprandial blood sugar, cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglyceride are classified as follows.

- Body mass index classification

Body mass index is a statistical measure to compare a person's weight and height. In fact, this measurement does not measure obesity but is a useful tool for estimating one's weight according to height. This index was developed between 1830 and 1850. It is very simple to calculate and is used in many applications to determine overweight and weight loss. Body mass index is obtained by dividing a person's weight in kilograms by his/her square height in meters as shown in Eq. 1. Table 2 shows the body mass index classification.

$$\text{Body mass index} = \frac{\text{weight in kilograms}}{(\text{height in meters})^2} \quad (1)$$

- High-blood-pressure classification

High blood pressure (hypertension) is a chronic disease in which blood pressure in the arteries rises. Following this increase in pressure, the heart must work more intensely than normal to maintain blood circulation in the blood vessels. Blood pressure consists of two systolic and diastolic scales that are dependent on the contraction (systolic) or relaxation (diastolic) of the heart muscle between beats. Nearly 50% of patients with high blood pressure are unaware of their disease, and many patients are accidentally informed of their blood pressure. Table 3 shows the

classification of systolic and diastolic blood pressure. The units are based on the millimeter of mercury (mmHg).

- Blood sugar classification

High blood glucose (sugar) is one of the risk factors that increase the risk of complications of diabetes. This dataset used two types of blood sugar tests (fasting and 2 h after meals). Table 4 shows the blood sugar classification. The units in this table are measured by milligrams per deciliter (mg/dL).

- Cholesterol classification

Cholesterol is a fatty, wax-like substance that is made in the liver and other cells. Cholesterol that moves through the blood attaches to proteins and forms a package called a lipoprotein. Lipoproteins are divided into high-density and low-density groups. Tables 5, 6, and 7 show the classification of cholesterol, HDL, and LDL.

- Triglyceride classification

Triglyceride is a type of fat in the body. Triglycerides act as a source of energy for the body. When you need a lot of energy, the body breaks down these fats and converts them into energy so that cells can use it. But increased levels of triglycerides in the blood can block arteries and damage the pancreas. Table 8 shows the triglyceride classification.

In statistics, outlier data are data that are far from the rest of the data. Different methods such as regression and clustering are used to deal with outliers and smooth them. This database uses box plot in Naim software to solve the outlier's problem.

3.2.2 Data transformation (DT)

Data transformation also called data conversion helps to convert and consolidate data into a form suitable for data mining. There are several methods for converting data such as minimum–maximum normalization. Normalization is a way of putting data in a similar domain. In other words, a data miner may encounter situations where the features contain values that are in a different range or domain. These large-value features may have a much greater impact on the cost function than low-value features. This problem will be solved by normalizing features so that their values are in the same domain. Normal values allow for more accurate comparisons of different datasets. It also reduces the impact of the sharp difference between the values of the different features. To build the data model before starting the model training, the data are subdivided into its largest corresponding value to be normalized to values between zero and one. This will minimize the effect of the actual scale, and all entries will be in the same domain. Normalization makes it possible to compare data with different measurement criteria.

Table 2 Body mass index classification

Category	Body mass index (kg/m ²)
Underweight	< 18.5
Normal weight	[18.5; 25)
Overweight	[25; 30]
Obese	> 30

Table 3 The classification of systolic and diastolic blood pressure

Category	Systolic blood pressure (mmHg)	Diastolic blood pressure (mmHg)
Low blood pressure	< 90	< 60
Normal	[90; 120)	[60; 80)
The risk of high blood pressure	[120; 140]	[80; 90]
High blood pressure	> 140	> 90

Table 4 Blood sugar classification

Blood sugar test	Normal (mg/dL)	Pre-diabetic (mg/dL)	Diabetic (mg/dL)
Fasting	[70; 100]	[100; 130]	> 127
2 h after meal	< 140	[140; 200]	> 200

Table 5 Cholesterol classification

Category	Cholesterol level (mg/dL)
Normal	< 200
At risk	[200; 240]
High cholesterol	> 240

Table 6 HDL classification

Category	HDL (mg/dL)
Low values	< 40
Average values	[40; 60]
Normal	> 60

Table 7 LDL classification

Category	LDL (mg/dL)
Normal	< 100
Values close to normal	[100; 130)
At the beginning of the danger	[130; 160)
At risk	[160; 190]
Very high risk level	> 190

Table 8 Triglyceride classification

Category	Triglyceride (mg/dL)
Normal	< 150
Values close to normal	[150; 200)
Dangerous levels	[200; 500)
Very high risk level	> 500

Equation 2 shows details of the min–max normalization used in our data conversion phase:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

3.3 Proposed classification method

There are many data mining methods for modeling. In this paper, the SVM classification is used to find the optimal model and pattern. Modeling is done using Naim software. The main method used here is predictive data mining. A ten-step validation method is used to determine the training and experimental data and to evaluate the performance of the proposed method which is a common technique for estimating the efficiency of classifiers.

In short, training is the process of providing feedback to the algorithm to regulate the power of classification prediction. And testing is the process of determining the true accuracy of the classification produced by the algorithm. During testing, data that have never participated in the training are classified. Usually, after each training step, validation is done to determine classification. The validation step does not provide any feedback to the algorithm for the classification adjustment but only specifies when the training algorithm should be terminated. Then, the error and mean error are calculated at each stage. To determine the category label (the type of complication) after consulting with diabetes specialists, it is concluded that each complication should be studied separately for greater accuracy rather than splitting the complication into microvascular and macrovascular groups. Accordingly, the category label (the type of complication) in the created model is shown in Table 9.

Table 9 Model category label

Dispeplimi	An increase in blood lipids
Eye problem	Eye problem
High pressure	High blood pressure
Dialysis	Dialysis history
Heart problem	Heart problems
Stroke	Stroke
Foot ulcers	Diabetic foot ulcer
Dka	Diabetic coma

3.3.1 Grey wolf optimizer (GWO)

GWO is one of the latest optimization methods designed and implemented based on social behavior and hunting grey wolf. For some problems, this algorithm can provide better results than other algorithms, such as the PSO algorithm and multi-objective decomposition-based evolutionary algorithm [40].

Grey wolves are considered as the highest level of hunters because there is no natural hunter for this type of animal. Grey wolves usually live in groups of 5–20 wolves. Leaders (the first solution), known as alpha, have the duty of deciding on hunting. The second group of grey wolves belongs to the beta class (second-best solutions). Beta wolves help alpha wolves in decision making and other activities in the group. The lowest level in the hierarchy of grey wolves is omega wolves that play the role of goat (third-best solutions). Omega wolves need to join higher classes if needed. Wolves that are not in any of the alpha, beta, or omega categories belong to the delta category. Delta wolves (rest of the candidate solutions) follow the alpha and beta classes, but omega wolves are dominated.

In brief, the common steps of the GWO algorithm are as follows:

- Generate initial population of wolves based on a set of random solutions,
- Calculate the corresponding objective value for each wolf,
- Choose the first three best wolves and save them as alpha, beta, and omega,
- Update the position of the rest of the population (delta wolves) using equations given in [40],
- Update parameters a , A , and C ,
- Go to the second step if the criterion is not satisfied,
- Position and score of the alpha solution is returned as the best solutions.

3.3.2 Improved GWO using weighted adaptive middle filter

The most important factor that controls the performance and accuracy of an optimization algorithm is the compromise between exploration (efficiency) and exploitation. Exploration means the ability of the search algorithm to search different areas of the search space to locate the appropriate optimum. On the other hand, efficiency is the ability to focus the search in the desired range to scrutinize the solution. A good optimization algorithm balances these two contradictory goals. In any algorithm or in the complementary version, it is attempted to improve the performance of the method by controlling these two parameters. The experience shows that in the early iterations, the exploration power needs to be increased and the efficiency becomes more pronounced over time. This means that in the initial iterations, the algorithm performs a variety of searches in space and in the last iterations, it searches the found areas more accurately.

In order to increase the efficiency and accuracy of the GWO to reach optimal values, the results of each step of the GWO are filtered using WAMF. In other words, the value of the search criteria is adjusted more precisely to increase the optimization accuracy. At each step of the algorithm implementation, outlier solutions (wolves far from the target) are filtered through a WAMF with dynamic window in order to increase the accuracy of the algorithm. Algorithm 1 shows the pseudo-code of the improved GWO.

3.3.3 Applying filter at each step of the GWO implementation

As shown in Algorithm 1, a temperature parameter is defined with an initial value of zero and the final value of 1000 at the beginning of the algorithm. The number of wolves, or agents, is considered $n = 25$. The GWO starts by creating a random population of grey wolves (candidate solutions). After assigning random values to parameters C , a , A , the fitness of each individual is defined based on its non-dominated sorting GA (NSGA-II) [41] which is the most popular solution Eq. 5. The calculated fitness of each factor is then put into one of the α , ω , β , or δ categories based on their value.

Once the set of agents has been specified, the position of each agent is updated at each iteration using Eqs. 3–5 [6].

$$\begin{aligned}\overrightarrow{D_\alpha} &= |\overrightarrow{C_1} \cdot \overrightarrow{X_\alpha} - \overrightarrow{X}|, \overrightarrow{D_\beta} = |\overrightarrow{C_2} \cdot \overrightarrow{X_\beta} - \overrightarrow{X}|, \\ \overrightarrow{D_\delta} &= |\overrightarrow{C_3} \cdot \overrightarrow{X_\delta} - \overrightarrow{X}| \end{aligned} \quad (3)$$

$$\begin{aligned}\overrightarrow{X_1} &= \overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \overrightarrow{D_\alpha}, \overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \overrightarrow{D_\beta}, \\ \overrightarrow{X_3} &= \overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot \overrightarrow{D_\delta} \end{aligned} \quad (4)$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \quad (5)$$

As seen in pseudo-code (see Algorithm 1), the algorithm enters the filtering stage before updating parameters C , a , A . In this step, we first define parameter $temp$ (which is the current value of the temperature divided by the final value), variable $Rand$ (which is a random number between zero and one) and variable K (which is the size of the filter window). In the filtering phase, depending on which category the wolf belongs to, a probability is identified for filtering it. That is, the wolves farther from the target are more likely to be chosen. The probability of wolves being selected is as follows:

- $P = 0.1$ when $agent(i)$ is X_α
- $P = 0.2$ when $agent(i)$ is X_β
- $P = 0.3$ when $agent(i)$ is X_δ
- $P = 0.4$ when $agent(i)$ is X_ω

If $P \cdot Rand \leq temp$, the selected wolf is eligible for filtration and enters the final step of applying the filter. Otherwise the next wolf will be chosen. In the final step of applying the filter, a window with k nearest neighbor is formed for the selected wolf. The initial value of the window is 3 because, it is the most populated type of solutions. Depending on which neighbor belongs to each category, a weight is assigned to them. The weights of each category of wolves based on their priority are as follows:

- $weight(j) = 4$ when $window(j)$ is X_α
- $weight(j) = 3$ when $window(j)$ is X_β
- $weight(j) = 2$ when $window(j)$ is X_δ
- $weight(j) = 1$ when $window(j)$ is X_ω

In a descending order. That implies we give higher weight to alpha wolves because they are our first expected solutions. The wolves in the window are then sorted, and their mean is calculated after they are weighed.

Med is the middle of the positions of the K nearest neighbors of the selected wolf.

At final stage, mean agent fitness is calculated. If this value is less than the fitness of the selected wolf, then the new position of the selected wolf is calculated using Eq. 6 which is equal to the average of the old position and the Med value.

$$\text{New position} = \frac{(\text{Med} + \text{Old position of the current search agent})}{2} \quad (6)$$

Then, the parameters C , a , A are updated, agents' fitness is calculated and put into the α , β , and δ categories, and finally algorithm starts the next iteration. Otherwise, the filter window size will be increased by one unit. Then, weighing to the neighbor wolves, middle selection and fitness calculation operations are performed. If the fitness of the mean factor is still less than the selected wolf, $K = K + 1$. This operation will continue for each wolf selected until $K = 10$.

3.3.4 Avoid improved GWO from stocking in local optimum

The filtering operation is controlled by a temperature parameter. Initially, the temperature is 0, which is very low pressure for filtering operation. While the algorithm is running, temperature increases, and as the temperature increases, the filtering pressure increases. This way, a different amount of filter pressure can be realized during the algorithm is running. In other words, the algorithm first starts the filtering operation at very low pressure (almost zero) and increases the filter pressure upward. To prevent the algorithm stuck in local optimum at the beginning of the algorithm, we increase the exploration operation by preserving the diversity and changing the wolf category. The improved GWO is used to adjust parameters C and α . The range of search space for parameter C (penalty parameter) is considered between 0.01 and 3500 and for parameter α is considered between 0.01 and 32 [42]. The objective function of the improved GWO is defined as follows:

$$\text{Objective function} = \text{Minimize}(\text{Error rate}) \quad (7)$$

Algorithm 1: Pseudo-code of the improved GWO by WAMF

```

1: Initialize the grey wolf population  $X_i$  ( $i = 1, 2, \dots, n$ )
2: Initialize  $a$ ,  $A$ , and  $C$ 
3:  $t=0$ 
4: Max_number_of_iteration=1000
5:  $n$ =numbber of agents
6: Calculate the fitness of each search agent
7:  $X_\alpha$  = the best search agent
8:  $X_\beta$  = the second best search agent
9:  $X_\delta$  = the third best search agent
10:  $X_\omega$  = the others
11: while  $t < \text{Max\_number\_of\_iterations}$  do
12:    $\text{temp} = t / \text{Max\_number\_of\_iterations}$ 
13:   for each search agent do
14:     Update the position of the current search agent
15:   end for
16:    $K=3$ 
17:   for  $i=1$  to  $n$  do
18:     if agent( $i$ ) is  $X_\alpha$  then
19:        $P=0.1$ 
20:     end if
21:     if agent( $i$ ) is  $X_\beta$  then
22:        $P=0.2$ 
23:     end if
24:     if agent( $i$ ) is  $X_\delta$  then
25:        $P=0.3$ 
26:     end if
27:     if agent( $i$ ) is  $X_\omega$  then
28:        $P=0.4$ 
29:     end if
30:     if  $P \cdot \text{Rand} \leq \text{temp}$  then
31:       Proposed_Method_Start:
32:       window= $K$ -Nearest Neighborhoods of current agent (agent( $i$ ))
33:       for  $j=1$  to  $K$  do
34:         if window( $j$ ) is  $X_\alpha$  then
35:           weight( $j$ )=4
36:         end if
37:         if window( $j$ ) is  $X_\beta$  then
38:           weight( $j$ )=3
39:         end if
40:         if window( $j$ ) is  $X_\delta$  then
41:           weight( $j$ )=2
42:         end if
43:         if window( $j$ ) is  $X_\omega$  then
44:           weight( $j$ )=1
45:         end if
46:       end for
47:       Med=Calculate median of weighted positions in the window
48:       if fitness(current agent) < fitness(Med) then
49:         New position of the current search agent is: (Med+Old position of current
search agent)/2

```

3.3.5 Features selection

The purpose of feature selection techniques is to remove irrelevant and ineffective features in the data. Unrelated features do not provide useful information to the classifier. Feature selection techniques are a subset of feature extraction methods. In feature extraction, new features are created as a function of all problem features, while in

feature selection, a subset of all features is selected. Using the feature selection method reduces training time and computational time and increases classifier generalization capability. A feature selection algorithm uses a search method to select a subset of features and an evaluation criterion to rank this subset. In the simplest algorithm, all subsets of possible features are investigated and a subset with the lowest classification error rate is selected. Full

search in feature space has a high computational burden. So GWO is used for feature selection in order to solve this problem. Each of these features is important in the diagnosis and prediction of diabetes complications. In other words, not all features are of equal value. For example, in the diagnosis of diabetes, two features (1) body mass index and (2) family history are of different importance. What is the value of each feature and how much does it play in diagnosing the disease is an important issue. In this paper, the value and role of each of them in identifying the various complications is carefully determined by weighing each of the features. The feature selection process in the proposed method involves the following steps:

- A. Producing function

This function generates candidate sets in the initial population of the GWO to select and weight the features.

- B. Fitness function

This function evaluates the set of candidate solutions for feature selection and weighting at each stage of the GWO and returns the prediction accuracy as the fitness of each factor.

- C. Update agent position

Based on the GWO, the position of the agents is updated at each stage.

- D. Using adaptive middle filter

In order to improve the efficiency of GWO in achieving optimal accuracy in prediction, the results of each step are filtered by WAMF. In other words, the value of the exploration criteria is adjusted more precisely to increase optimization accuracy.

- E. Termination condition

- a. Reaching an optimal accuracy in predicting diabetes complications.
- b. Complete the number of iterations determined to run the GWO.

In the proposed method, to determine the value and role of each feature in the diagnosis and prediction of diabetes complications, a random number between 0 and 1 is assigned to each feature indicating the degree of importance of the features and optimized by the GWO. The weighted values of the features are given to SVM as input and features are selected based on the final weight of each feature. In the validation section, first, the error percentage of the proposed method is compared with that of the GWO-SVM, GA-SVM, and PSO-SVM in 500 iterations. Then, the results of the proposed method are compared with that of machine learning algorithms such as DT, SB, and multilayer perceptron neural network (MLP NN).

4 Experimental results and discussion

Data of 1573 patients are collected from the endocrinology department of Mazandaran University of Medical Sciences. After preprocessing, they are described, simulated, and analyzed in MATLAB 2016 software that is used on an Intel Core i7 processor, 2.60 GHz CPU and 16 GB RAM and the running OS platform is Microsoft Windows 8.1. In the proposed method, to determine the value and role of each feature in predicting complications, a weight in the interval $[0, 1]$ is assigned to each of them indicating the degree of importance of a feature using the proposed optimization method after 500 iterations. Table 10 shows an example of weighted features for all 8 diabetic complications.

The weight of each feature given as input to the SVM is optimized at 500 iterations. In all listed complications, data samples divided into ten subsets were nine of them used for training, and the remaining one used for testing. This procedure was repeated ten times until each of the ten subsets was evaluated. Since GWO is classified as meta-heuristic methods, it needs to run multiple times to get the best result. Hence, we repeated the whole process of choosing test and train data ten times (each with randomized data sequences). After optimization for each complication, the error percentage obtained at each stage of the proposed algorithm is compared with that of the GWO-SVM, GA-SVM, and PSO-SVM. Latter, we compare the accuracy of the proposed method with other machine learning algorithms using two different datasets and present results in the following subsections.

4.1 Prediction of health complication

4.1.1 Increased blood lipids complication

Based on the proposed objective function, to predict the complication of increased blood lipids (hyperlipidemia), the proposed method, GWO, PSO, and GA have been used to improve the performance of the SVM algorithm. In Figure 1, the vertical axis shows the error percentage in predicting increased blood lipids, and the horizontal axis represents the number of iterations. In the first iterations, as the initial population is random, error reduction is tangible. But in subsequent iterations the error reduction rate decreases and eventually, the proposed method achieves a better error reduction at the end of the simulation.

4.1.2 Eye problem complication

Figure 2 shows the error percentage of the proposed method, GWO, PSO, and GA algorithm to improve the

Table 10 An example of weighted features for all 8 diabetic complications

Feature	Diabetic complication							
	Increase blood lipid	Eye problem	High blood pressure	Dialysis history	Heart problems	Stroke	Diabetic foot ulcer	Diabetic coma
Age	0.3476	0.4090	0.5183	0.3525	0.6142	0.1574	0.6354	0.1962
Sex	0.2946	0.2068	0.1653	0.9535	0.8275	0.2466	0.1311	0.5397
BMI	0.9583	0.7772	0.0095	0.9988	0.7025	0.7926	0.1762	0.8739
Inherit	0.6989	0.9995	0.3814	0.1835	0.5176	0.3391	0.9725	0.4856
SBP	0.8754	0.6344	0.1338	0.7600	0.4164	0.7215	0.4866	0.1218
DBP	0.1390	0.1888	0.0891	0.6073	0.1953	0.8372	0.9278	0.7392
Education	0.0920	0.0582	0.0373	0.0033	0.1344	0.2010	0.9968	0.5065
FBS	0.4739	0.1898	0.3180	0.3105	0.7651	0.6029	0.2356	0.8864
2hpp-BS	0.8580	0.9092	0.5889	0.6691	0.2965	0.1517	0.5593	0.5515
TC	0.2877	0.7195	0.7506	0.4329	0.9505	0.0547	0.2578	0.9427
LDL	0.2854	0.7660	0.6323	0.2741	0.8186	0.7312	0.0073	0.6617
HDL	0.2668	0.6144	0.3132	0.2197	0.9569	0.2779	0.1230	0.7680
TG	0.5718	0.6771	0.0153	0.5283	0.7770	0.0930	0.8608	0.7347
BUN/CR	0.6923	0.4011	0.2842	0.0802	0.3140	0.4216	0.7524	0.6533
Activity	0.5928	0.8015	0.1400	0.8081	0.0898	0.9237	0.4764	0.1142
Dispeplimi	0.9952	0.0124	0.0127	0.0366	0.0953	0.0926	0.0442	0.2836
Eye problem	0.0945	0.9869	0.0241	0.0855	0.1660	0.0843	0.2006	0.2713
High pressure	0.0810	0.0089	0.9651	0.0348	0.0865	0.0867	0.0026	0.0700
Dialyze	0.0397	0.2932	0.2262	0.9874	0.1794	0.0763	0.2470	0.0349
Heart problem	0.0171	0.0162	0.1728	0.0506	0.9715	0.0265	0.0818	0.0924
Stroke	0.1676	0.1267	0.0882	0.0131	0.0535	0.9770	0.0134	0.0049
Foot ulcers	0.0904	0.0099	0.0148	0.0517	0.0721	0.0395	0.9499	0.0956
DKA	0.0579	0.0902	0.0811	0.0678	0.0628	0.1652	0.1525	0.9754
Smoke	0.2796	0.5537	0.7837	0.5844	0.1163	0.4912	0.3710	0.6156

Higher value then given feature is more significant

Less value then given feature is less significant

performance of the SVM in predicting eye problems. In this figure, the vertical axis represents the error percentage of predicting eye problems and the horizontal axis represents the number of iterations. As can be seen, in the first iterations, the proposed method has a higher error percentage than other methods, and after the filtering phase starts, the error percentage of the proposed method is lower than that of the PSO and GA. Then, in iteration 140, it improves over the GWO and this trend continues until iteration 500. As a result, the proposed method reaches higher accuracy in predicting eye problems.

4.1.3 High blood pressure complication

As can be seen in Fig. 3, the error percentage of predicting high blood pressure in the proposed method is compared to the GWO, PSO, and GA. In this figure, the vertical axis represents the percentage error of predicting high blood

pressure complication and the horizontal axis represents the number of iterations. In the early iterations, the error percentage of the PSO and GA is higher than the proposed method. Then, the error percentage of the proposed algorithm is approximately equal to that of the GWO and this trend continues until the end of the simulation. At the end of the simulation, the proposed method has a lower error than other ones which improved the error percentage.

4.1.4 Dialysis history complication

The error percentage for predicting the dialysis complication in the proposed algorithm is compared to that of the GWO, GA, and PSO and shown in Fig. 4. The vertical axis represents the error percentage in predicting dialysis history, and the horizontal axis indicates the number of iterations. In early iterations, the error percentage of the PSO and GWO is higher than that of the proposed method and

Fig. 1 Error percentage in predicting increased blood lipids

Algorithm 1: Pseudo-code of the improved GWO by WAMF

```

1: Initialize the grey wolf population  $X_i$  ( $i = 1, 2, \dots, n$ )
2: Initialize  $a$ ,  $A$ , and  $C$ 
3:  $t=0$ 
4: Max_number_of_iteration=1000
5:  $n$ =numbber of agents
6: Calculate the fitness of each search agent
7:  $X_\alpha$  = the best search agent
8:  $X_\beta$  = the second best search agent
9:  $X_\delta$  = the third best search agent
10:  $X_\omega$  = the others
11: while  $t < \text{Max\_number\_of\_iterations}$  do
12:    $\text{temp} = t / \text{Max\_number\_of\_iterations}$ 
13:   for each search agent do
14:     Update the position of the current search agent
15:   end for
16:    $K=3$ 
17:   for  $i=1$  to  $n$  do
18:     if agent( $i$ ) is  $X_\alpha$  then
19:        $P=0.1$ 
20:     end if
21:     if agent( $i$ ) is  $X_\beta$  then
22:        $P=0.2$ 
23:     end if
24:     if agent( $i$ ) is  $X_\delta$  then
25:        $P=0.3$ 
26:     end if
27:     if agent( $i$ ) is  $X_\omega$  then
28:        $P=0.4$ 
29:     end if
30:     if  $P \cdot \text{Rand} \leq \text{temp}$  then
31:       Proposed_Method_Start:
32:       window= $K$ -Nearest Neighborhoods of current agent (agent( $i$ ))
33:       for  $j=1$  to  $K$  do
34:         if window( $j$ ) is  $X_\alpha$  then
35:           weight( $j$ )=4
36:         end if
37:         if window( $j$ ) is  $X_\beta$  then
38:           weight( $j$ )=3
39:         end if
40:         if window( $j$ ) is  $X_\delta$  then
41:           weight( $j$ )=2
42:         end if
43:         if window( $j$ ) is  $X_\omega$  then
44:           weight( $j$ )=1
45:         end if
46:       end for
47:       Med=Calculate median of weighted positions in the window
48:       if fitness(current agent) < fitness(Med) then
49:         New position of the current search agent is: (Med+Old position of current
search agent)/2
50:       else
51:         Goto Proposed_Method_Start:
52:       end if
53:     end if
54:   end for
55:   Update  $a$ ,  $A$ , and  $C$ 
56:   Calculate the fitness of all search agents
57:   Update  $X_\alpha$ ,  $X_\beta$ ,  $X_\delta$ , and  $X_\omega$ 
58:    $t=t+1$ 
59: end while
60: Return  $X_\alpha$ 

```

Fig. 2 Error percentage in predicting eye problem

Fig. 3 Error percentage of predicting high blood pressure complication

Fig. 4 Error percentage in predicting dialysis history complication

the error percentage of the GA is lower than that of the proposed method. Then, the error percentage of the proposed method is approximately equal to that of the GWO and is lower than that of the PSO and GA. This trend continues until the end of the simulation. Until iteration 100, PSO has a lower error percentage than the proposed method. Then, by increasing iterations, the error percentage becomes higher than the proposed algorithm but is still

lower than the GA. This trend is constant until the end of the simulation. At the end of the simulation, the proposed method has less error than other ones which indicates improved error percentage.

4.1.5 Heart attack complications

Figure 5 shows the error percentage in predicting heart problems using the proposed algorithm than GWO, GA, and PSO. The vertical axis represents the error percentage in predicting heart problems, and the horizontal axis indicates the number of iterations. As can be seen, at the beginning of the simulation, the proposed method has the highest error percentage compared to the other algorithms. But by increasing iterations and executing the filtering phase, the error percentage of the proposed method reduces. At the end of the simulation, the proposed method has a lower error than other ones, which indicates an improved error percentage.

4.1.6 Stroke complications

As can be seen in Fig. 6, the error percentage in predicting stroke complications in the proposed method is compared to that of the GWO, PSO, and GA. In this figure, the vertical axis represents the error percentage of predicting stroke complications and the horizontal axis represents the number of iterations. At first iterations, the error percentage of the proposed algorithm is higher than that of PSO and GA. Then, it becomes approximately equal to that of the GWO and this trend continues until the 70th iteration; after that, the error percentage of the proposed algorithm reduces. At the end of the simulation, the proposed method has less error than other methods, which indicates improved error percentage by the proposed method.

Fig. 5 Error percentage in predicting heart attack problems

Fig. 6 Error percentage for predicting stroke complication

Fig. 7 Error percentage for predicting diabetic foot ulcer complication

4.1.7 Diabetes foot ulcer complication

As can be seen in Fig. 7, the error percentage of predicting diabetic foot ulcer complication in the proposed method is compared to that of the GWO, PSO, and GA. In this figure, the vertical axis represents the error percentage of predicting diabetic foot ulcer and the horizontal axis represents the number of iterations. At first iterations, the error percentage of the PSO and GA is higher than that of the

Fig. 8 Error percentage for predicting diabetic coma complication

proposed algorithm. Then, until iteration 270, the error percentage of the GWO is lower than that of the proposed algorithm. Then, until the last iteration, the proposed algorithm has the least error percentage.

4.1.8 Diabetes coma complication

Figure 8 shows the error percentage of predicting diabetic coma complication in the proposed method, GWO, and PSO. At the beginning of the simulation, the proposed method has a lower error percentage than the GWO and a higher error percentage than the GA and PSO. Then, it is observed that by applying filtering in iteration 280, the error percentage of the proposed algorithm becomes lower than other ones. This trend continues until the last iteration of the algorithm.

4.2 Evaluation and comparison of proposed method based on machine learning algorithms

In this section, the proposed model is compared with three machine learning algorithms: DT, SB, and MLP NN. The relationship between the actual classes and the predicted classes can be calculated using the confusion matrix; in the following its required parameters are also described. According to Eqs. 8–13, criteria such as accuracy, sensitivity, F-measure, precision, sensitivity, recall criteria are used to compare the proposed model with other ones.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{All}} \quad (8)$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (9)$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})} \quad (10)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (11)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (12)$$

$$\text{F-Measure} = 2 \cdot \text{Precision} \cdot \text{Recall} \quad (13)$$

Tables 11 and 12 show the results of predicting diabetes complications with DT techniques, SB and MLP NN in terms of Accuracy. Complications of diabetes include increase blood lipids, eye problems, high blood pressure, dialysis history, heart problems, stroke, diabetic foot ulcer, and diabetic coma. In all the mentioned complications, the accuracy of the proposed method is higher than that of the DT, SB, and MLP NN (Fig. 9).

Table 11 Predicting diabetes complications by accuracy criterion

Diabetes complications	Proposed method	Decision tree	Simple Bayes	MLP NN
Increase blood lipid	96.0	94.0	81.0	92.0
Eye problem	94.0	91.0	73.0	85.0
High blood pressure	92.0	89.0	69.0	78.0
Dialysis history	97.0	95.0	93.0	94.0
Heart problems	95.0	89.0	72.0	86.0
Stroke	96.0	93.0	90.0	94.0
Diabetic foot ulcer	96.0	94.0	82.0	93.0
Diabetic coma	97.0	93.0	89.0	94.0

Table 12 Comparison of predicting increased blood lipids complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1160	66	261	12	0.99	0.96	0.99	0.80	0.97	Complication	Proposed method
261	12	1160	66	0.80	0.96	0.80	0.99	0.87	No complication	
1228	66	185	20	0.98	0.95	0.98	0.74	0.96	Complication	Decision tree
185	20	1228	66	0.74	0.90	0.74	0.98	0.81	No complication	
1076	111	140	172	0.86	0.91	0.86	0.56	0.88	Complication	Simple Bayes
140	172	1076	111	0.56	0.45	0.56	0.86	0.50	No complication	
1217	92	159	31	0.98	0.93	0.98	0.63	0.95	Complication	MLP NN
159	31	1217	92	0.63	0.84	0.63	0.98	0.72	No complication

Fig. 9 Comparison of predicting diabetes complications with accuracy criterion

4.2.1 Increased blood lipids complication

In Table 12 and Fig. 10, a comparison between the proposed method, MLP NN, SB, and DT in predicting

increased blood lipids complication is shown. Evaluation of values based on sensitivity, specificity, precision, recall, and F-measure criteria illustrates the superiority of the

Fig. 10 Comparison of predicting increased blood lipids complication

Table 13 Comparison of predicting eye problem complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1161	71	256	11	0.97	0.94	0.98	0.78	0.97	Complication	Proposed method
256	11	1161	71	0.78	0.96	0.78	0.99	0.86	No complication	
1080	85	282	52	0.95	0.93	0.95	0.77	0.94	Complication	Decision tree
282	52	1080	85	0.77	0.84	0.77	0.95	0.80	No complication	
946	220	147	186	0.84	0.81	0.84	0.40	0.82	Complication	Simple Bayes
147	186	946	220	0.40	0.44	0.40	0.84	0.42	No complication	
1082	172	195	50	0.96	0.86	0.96	0.53	0.91	Complication	MLP NN
195	50	1082	172	0.53	0.80	0.53	0.96	0.64	No complication	

Fig. 11 Comparison of predicting eye problem complication

weighted features-based classification than the same weighted features.

4.2.2 Eye problem complication

Table 13 and Fig. 11 show the comparison of the proposed method, MLP NN, SB, and DT in predicting eye problem complications. Comparison of the results of the proposed method based on F-measure, sensitivity, specificity, precision, and recall criteria indicates the superiority of the proposed algorithm than compared ones.

4.2.3 High blood pressure complication

A comparison of predicting complications in the proposed algorithm, DT, SB, and MLP NN based on sensitivity, specificity, precision, recall, and F-measure criteria is shown in Table 14 and Fig. 12. A comparison of the results shows the improvement of classification with the weighted features than classification with the same weight features.

4.2.4 Dialysis history complication

The performance of weighted features-based classification compared to modeling the same weighted features in predicting dialysis history complication is evaluated in Table 15 and Fig. 13. The superiority of the proposed method compared to DT, SB, and MLP NN in terms of sensitivity, specificity, precision, recall, and F-measure criteria is approved by comparing results.

4.2.5 Heart attack complications

The performance of weighted features-based classification compared to modeling the same weighted features in predicting heart attack complications based on sensitivity, specificity, precision, recall, and F-measure criteria is evaluated in Table 16 and Fig. 14. The proposed method has better results in the diagnosis of heart attack complications compared to DT methods, SB, and MLP NN.

Table 14 Comparison of predicting high blood pressure complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
863	104	516	16	0.98	0.89	0.98	0.86	0.93	Complication	Proposed method
516	16	863	104	0.83	0.97	0.83	0.98	0.90	No complication	
787	86	541	85	0.90	0.89	0.90	0.83	0.92	Complication	Decision tree
541	85	787	86	0.86	0.86	0.86	0.90	0.86	No complication	
659	260	367	213	0.76	0.72	0.76	0.59	0.74	Complication	Simple Bayes
367	213	659	260	0.59	0.63	0.59	0.76	0.61	No complication	
739	201	426	133	0.85	0.79	0.85	0.68	0.82	Complication	MLP NN
426	133	739	201	0.68	0.76	0.68	0.85	0.72	No complication	

Fig. 12 Comparison of predicting high-blood-pressure complication

Table 15 Comparison of predicting dialysis history complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1441	29	27	2	0.96	0.97	0.98	0.48	0.99	Complication	Proposed method
27	2	1441	29	0.48	0.93	0.48	1.00	0.64	No complication	
1386	65	29	19	0.95	0.96	0.97	0.31	0.97	Complication	Decision tree
29	19	1386	65	0.31	0.60	0.31	0.99	0.41	No complication	
1390	93	1	15	0.92	0.92	0.97	0.19	0.96	Complication	Simple Bayes
1	15	1390	93	0.01	0.06	0.01	0.99	0.02	No complication	
1390	72	22	15	0.78	0.95	0.94	0.23	0.97	Complication	MLP NN
22	15	1390	72	0.23	0.59	0.23	0.99	0.34	No complication	

Fig. 13 Comparison of predicting dialysis history complication **Table 16** Comparison of predicting heart attack complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1166	69	258	6	0.98	0.94	0.95	0.83	0.97	Complication	Proposed method
258	6	1166	69	0.79	0.98	0.79	0.99	0.87	No complication	
787	86	541	85	0.90	0.90	0.90	0.79	0.90	Complication	Decision tree
541	85	787	86	0.86	0.86	0.86	0.90	0.86	No complication	
958	237	111	193	0.83	0.80	0.83	0.32	0.82	Complication	Simple Bayes
111	193	958	237	0.32	0.37	0.32	0.83	0.34	No complication	
1075	147	201	76	0.93	0.88	0.93	0.58	0.91	Complication	MLP NN
201	76	1075	147	0.58	0.73	0.58	0.93	0.64	No complication	

4.2.6 Stroke complications

The performance of weighted features-based classification compared to modeling the same weighted features in predicting stroke complications is evaluated in Table 17 and Fig. 15. The proposed method has better results in the diagnosis of stroke complications compared to DT

methods, SB, and MLP NN based on sensitivity, specificity, precision, recall, and F-measure criteria.

4.2.7 Diabetes foot ulcer complication

Sensitivity, specificity, precision, recall, and F-measure criteria for evaluating the prediction of diabetic foot ulcer

Fig. 14 Comparison of predicting heart attack complications **Table 17** Comparison of predicting stroke complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1431	43	19	6	0.98	0.97	0.97	0.31	0.98	Complication	Proposed method
19	6	1431	43	0.31	0.76	0.31	1.00	0.44	No complication	
1370	77	25	27	0.91	0.95	0.94	0.25	0.96	Complication	Decision tree
25	27	1370	77	0.25	0.48	0.25	0.98	0.32	No complication	
1343	95	7	54	0.96	0.93	0.96	0.17	0.95	Complication	Simple Bayes
7	54	1343	95	0.07	0.11	0.07	0.96	0.09	No complication	
1384	82	20	13	0.93	0.94	0.93	0.20	0.97	Complication	MLP NN
20	13	1384	82	0.20	0.61	0.20	0.99	0.30	No complication	

Fig. 15 Comparison of predicting stroke complication

complications indicate the superiority of the proposed classification method over MLP NN, SB, and DT which is shown in Table 18 and Fig. 16.

4.2.8 Diabetes coma complication

As shown in Table 19 and Fig. 17, sensitivity, specificity, precision, recall, and F-measure criteria have been evaluated to compare diabetes coma complication predicting.

Table 18 Comparison of predicting diabetes foot ulcer complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1377	42	76	4	0.98	0.97	0.99	0.64	0.98	Complication	Proposed method
76	4	1377	42	0.64	0.95	0.64	0.97	0.77	No complication	
1313	74	83	29	0.98	0.95	0.98	0.53	0.96	Complication	Decision tree
83	29	1313	74	0.53	0.74	0.53	0.98	0.62	No complication	
1174	106	51	168	0.87	0.92	0.87	0.32	0.90	Complication	Simple Bayes
51	168	1174	106	0.32	0.23	0.32	0.87	0.27	No complication	
1321	93	64	21	0.98	0.93	0.98	0.41	0.96	Complication	MLP NN
64	21	1321	93	0.41	0.75	0.41	0.98	0.53	No complication	

Fig. 16 Comparison of predicting diabetic foot ulcer complication **Table 19** Comparison of predicting diabetes coma complication

TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure		
1451	31	10	7	0.97	0.95	0.98	0.24	0.98	Complication	Proposed method
10	7	1451	31	0.24	0.59	0.24	1.00	0.34	No complication	
1378	75	13	33	0.94	0.91	0.97	0.15	0.96	Complication	Decision tree
13	33	1378	75	0.15	0.28	0.15	0.98	0.19	No complication	
1288	70	18	123	0.91	0.93	0.91	0.20	0.93	Complication	Simple Bayes
18	123	1288	70	0.20	0.13	0.20	0.91	0.16	No complication	
1387	79	9	24	0.96	0.94	0.93	0.10	0.96	Complication	MLP NN
9	24	1387	79	0.10	0.27	0.10	0.98	0.15	No complication	

Results indicate the superiority of the proposed classification method over MLP NN, SB, and DT methods.

Predicting and correctly diagnosing diabetes complications using AI and machine learning increases the chances of successful treatment. In this study, the middle filter is used to optimize the GWO algorithm and introducing a new model to predict and diagnose diabetes complications. The simulation results show that the proposed model is

more accurate than SB, DT, and MLP NN. The high accuracy in the diagnosis of complications of diabetes indicates the superiority of the proposed method. Complexity and time-consuming implementation are the weaknesses of this method.

Fig. 17 Comparison of predicting diabetes coma complication **Table 20** Description of the code field in UCI dataset [21]

33 = Regular insulin dose
34 = NPH insulin dose
35 = Ultra Lente insulin dose
48 = Unspecified blood glucose measurement
57 = Unspecified blood glucose measurement
58 = Pre-breakfast blood glucose measurement
59 = Post-breakfast blood glucose measurement
60 = Pre-lunch blood glucose measurement
61 = Post-lunch blood glucose measurement
62 = Pre-supper blood glucose measurement
63 = Post-supper blood glucose measurement
64 = Pre-snack blood glucose measurement
65 = Hypoglycemic symptoms
66 = Typical meal ingestion
67 = More-than-usual meal ingestion
68 = Less-than-usual meal ingestion
69 = Typical exercise activity
70 = More-than-usual exercise activity
71 = Less-than-usual exercise activity
72 = Unspecified special event

4.3 Experimental evaluation on UCI dataset

To compare the proposed method with related methods in this domain, we chose to use UCI Machine Learning Repository Dataset [21]. Diabetes files in this dataset consist of four fields per record such as date, time, code, and value. As shown in Table 20, code field of the dataset is described by an integer number showing the statues of a patient. There are 70 text files in the UCI dataset; each file contains one patient's disease history. The patients are insulin deficient. This disease is manifested by many so-called metabolic effects, the main one being high blood glucose, which can be detected by measurements.

In this experiment, we compare the accuracy of previous works with our proposed method on UCI Machine Learning Repository Datasets. Previous works [15, 16, 21] used fuzzy methods to classify diabetic patients. Paper [15] also used a fuzzy method combined with PSO and reported their result on the UCI dataset. In all listed methods, data samples were divided into ten subsets, where nine of them were used for training, and the remaining one was used for testing. This procedure was repeated ten times until each of the ten subsets was evaluated. Since PSO and GWO are classified as metaheuristic methods, they need to run multiple times to get the best result. Hence, we repeated the whole process of choosing test and train data ten times

Table 21 Comparison of classification accuracy of different methods on UCI dataset [21]

Method	FMM [16]	FMM-GA [17]	TFMM-PSO [15]	Proposed method
Accuracy (%)	71.2	76.4	81.3	89.6

(each with randomized data sequences) and listed the averaged results in Table 21.

The result clearly shows boosting the accuracy of the classification system because we use SVM combined with GWO which shows better optimization results compared with PSO.

5 Conclusion

Diabetes is one of the most common chronic diseases and a major public health problem in the world. It is a rapidly growing and serious chronic disease, and its prevalence has been increasing in Asian countries. The increasing prevalence of diabetes mellitus (DM), the emergence of its complications as a cause of death, the early disability, and the burden on healthcare systems have made it a health priority. Diabetes has a profound effect on the quality of life physically, socially, and mentally. Studies have shown that diabetes can have negative effects on public health and a sense of well-being, in other words quality of life. Diabetes is not treatable, but it can be controlled. Thus, controlling diabetes means preventing and delaying its complications. Poor controlling leads to long-term blood sugar levels, which are strongly linked to late complications such as retinopathy, nephropathy, and cardiovascular disease. These complications are associated with high healthcare costs and reduced quality of life. In this paper, grey wolf optimizer (GWO) is used to solve the diabetes diagnosis problem. The important point of the proposed algorithm is the accuracy of that compared to other classification algorithms, which if used with the improved GWO along with the SVM will increase the accuracy of this algorithm to an acceptable level compared to other classification algorithms. The proposed method is superior over DT, SB, and MLP NN in predicting increased blood lipid with the accuracy of 0.96, eye problem with the accuracy of 0.94, high blood pressure with the accuracy of 0.92, dialysis history with the accuracy of 0.97, heart problems with the accuracy of 0.95, stroke with the accuracy of 0.96, diabetic foot ulcer with the accuracy of 0.96, and diabetic coma with the accuracy of 0.97. The high accuracy in diagnosing diabetes complications indicates the superiority of the proposed method in improving the results of the GWO. We also compared the proposed method with other classifiers on the UCI dataset that also shows its advantage over fuzzy-based classifiers. Complexity and time-consuming execution are the weaknesses of this method. We are trying to reduce time complexity by making changes in the GWO method in our future research.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Maniruzzaman Md, Kumar N, Md Menhazul A, Md Shaykhul I, Suri HS, El-Baz AS, Suri JS (2017) Comparative approaches for classification of diabetes mellitus data: machine learning paradigm. *Comput Methods Progr Biomed* 152:23–34
2. Franz MJ (2003) A core curriculum for diabetes education: diabetes management therapies. *Am Assoc Diabet Educ* 2:1–341
3. Dewangan AK, Agrawal P (2015) Classification of diabetes mellitus using machine learning techniques. *Int J Eng Appl Sci* 2:145–148
4. Tao Z, Shi A, Zhao J (2015) Epidemiological perspectives of diabetes. *Cell Biochem Biophys* 73:181–185
5. Acharya UR, Faust O, Kadri NA, Suri JS, Yu W (2013) Automated identification of normal and diabetes heart rate signals using nonlinear measures. *Comput Biol Med* 43:1523–1529
6. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
7. Slowik A, Kwasnicka H (2018) Nature inspired methods and their industry applications-swarm intelligence algorithms. *IEEE Trans Ind Inf* 14(3):1004–1015
8. Tirkolaee EB, Alinaghian M, Hosseinabadi AAR, Sasi MB, Sangaiah AK (2019) An improved ant colony optimization for the multi-trip capacitated arc routing problem. *Comput Electr Eng* 77:457–470
9. Sangaiah AK, Sadeghilalimi M, Hosseinabadi AAR, Zhang W (2019) Energy consumption in point-coverage wireless sensor networks via bat algorithm. *IEEE Access* 7:180258–180269
10. Kashikolaei SMG, Hosseinabadi AAR, Saemi B, Shareh MB, Sangaiah AK, Bian G-B (2020) An enhancement of task scheduling in cloud computing based on imperialist competitive algorithm and firefly algorithm. *J Supercomput* 76(8):6302–6329
11. Anbananthan KSM, Sainarayanan G, Chekima A, Teo J (2007) Artificial neural network tree approach in data mining. *Malays J Comput Sci* 20:51–62
12. Chan CL, Liu YC, Luo SH (2008) Investigation of diabetic microvascular complications using data mining techniques In: *IEEE international joint conference on neural networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*, pp 830–834
13. Patil BMR, Durga T (2010) Association rule for classification of type-2 diabetic patients In: *2010 Second international conference on machine learning and computing (ICMLC)*, pp 330–334

14. Fang X (2009) Are you becoming a diabetic? A data mining approach. In: Sixth international conference on fuzzy systems and knowledge discovery. FSKD'09, vol 5, pp 18–22
15. Ganapathy S, Sethukkarasi R, Yogesh P (2014) An intelligent temporal pattern classification system using fuzzy temporal rules and particle swarm optimization. *Sadhana* 39:283–302
16. Simpson PK (1992) Fuzzy min–max neural networks-part I: classification. *IEEE Trans Neural Netw* 3(5):776–786
17. Quteishat A, Lim CP, Tan KS (2010) A modified fuzzy min-max neural network with a genetic algorithm-based rule extractor for pattern classification. *IEEE Trans Syst Man Cybern Part A Syst Hum* 40(3):641–650
18. Kanimozhi U, Ganapathy S, Manjula D, Kannan A (2019) An intelligent risk prediction system for breast cancer using fuzzy temporal rules. *Natl Acad Sci Lett* 42:227–232
19. Selvi M, Thangaramya K, Saranya MS, Kulothungan K, Satish RG, Kannan A (2019) Classification of medical dataset along with topic modeling using LDA. In: Proceedings of nanoelectronics, circuits and communication systems
20. Ganapathy S, Kulothungan K, Muthurajkumar S, Vijayalakshmi M, Yogesh P, Kannan A (2013) Intelligent feature selection and classification techniques for intrusion detection in networks: a survey. *EURASIP J Wirel Commun Netw*, Article number: 271
21. Murphy PM, Aha DW (1995) UCI Repository of Machine Learning Databases (Machine-Readab Data Repository). Dept. Inf. Comput. Sci., University of California, Irvine, CA
22. Breault L, Colin R, Fos Peter J (2002) Data mining a diabetic data warehouse. *Artif Intell Med* 26(1):37–54
23. Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K (2002) Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol* 12:243–248
24. Rohlfing CL, Wiedmeyer HM, Little RR, England JD, Tennill A, Goldstein DE (2002) Defining the relationship between plasma glucose and HbA1c analysis of glucose profiles and HbA1c in the Diabetes Control and Complications Trial. *Diabetes Care* 25:275–278
25. Silverstein C, Brin S, Motwani R, Ullman J (2003) Scalable techniques for mining causal structures. *Data Min Knowl Disc* 4:163–192
26. Trautvetter J, Devos P, Duhamel A, Beuscart R (2002) Assessing association rules and decision trees on analysis of diabetes data from the DiabCare program in France. *Stud Health Technol Inform* 90:557–561
27. Juan G, Luo S, Zhang H, Han Y (2007) Type 2 diabetes data processing with EM and C4. 5 algorithm In: Complex medical engineering, 2007. CME 2007. IEEE/ICME International Conference, pp 371–377
28. Jarullah A (2011) Decision tree discovery for the diagnosis of type II diabetes In: 2011 International conference on innovations in information technology (IIT), 2011, pp 303–307
29. Aljumah AA, Ahamad MG, Siddiqui MK (2012) Application of data mining: diabetes health care in young and old patients. *J King Saud Univ Comput Inf Sci* 25:127–136
30. Antonelli D, Baralis E, Bruno G, Cerquitelli T, Chiusano S, Mahoto N (2013) Analysis of diabetic patients through their examination history. *Expert Syst Appl* 40:4672–4678
31. Huang Y, McCullagh P, Black N, Harper R (2007) Feature selection and classification model construction on type 2 diabetic patients' data. *Artif Intell Med* 41:251–262
32. Han J, Rodriguez J, Beheshti M (2008) Diabetes data analysis and prediction model discovery using rapidminer. In: Future generation communication and networking, 2008. FGCN'08. Second international conference, vol 3, pp 96–99
33. Cho BH, Yu H, Kim TH, Kim SI (2008) Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods. *Artif Intell Med* 42:37–53
34. Farahmandian M, Lotfi Y, Maleki I (2015) Data mining algorithms application in diabetes diseases diagnosis: a case study". *MAGNT Research Report*
35. Han L, Luo S, Yu J, Pan L, Chen S (2015) Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. *IEEE J Biomed Health Inf* 19:728–734
36. Radha P, Srinivasan B (2014) Predicting diabetes by cosequencing the various data mining classification techniques. *Int J Innov Sci Eng Technol* 1:334–339
37. Uzer MS, Yilmaz N, Inan O (2013) Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification. *Sci World J* 2013:1–10
38. Ganapathy S, Yogesh P, Kannan A (2012) Intelligent agent based intrusion detection system using enhanced multiclass SVM. *Comput Intell Neurosci* 2012, Article ID 850259
39. Chen G, Astebro T (2003) How to deal with missing categorical data: test of a simple Bayesian method. *Organ Res Methods* 6:309–327
40. Mirjalili S, Saremi S, Mirjalili SM, Coelho LS (2016) Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. *Expert Syst Appl* 47:106–119
41. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197
42. Lin HT, Lin CJ (2003) A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Neural Comput* 3:1–32

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Fereshteh Jeyafzam¹ · Babak Vaziri¹ · Mohsen Yaghoubi Suraki² · Ali Asghar Rahmani Hosseinabadi³ · Adam Slowik⁴

✉ Adam Slowik
aslowik@ie.tu.koszalin.pl

Fereshteh Jeyafzam
f.jeyafzam@iauctb.ac.ir

Babak Vaziri
b.vaziribabak@iauctb.ac.ir

Mohsen Yaghoubi Suraki
M.Yaghoubi@qiau.ac.ir

Ali Asghar Rahmani Hosseinabadi
ark838@uregina.ca

¹ Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

² Department of IT and Computer Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

³ Department of Computer Science, University of Regina, Regina, Canada

⁴ Department of Electronics and Computer Science, Koszalin University of Technology, Koszalin, Poland