Journal of AI and Data Mining
Vol 7, No 4, 2019, 617-630
DOI: 10.22044/JADM.2018.6446.1758

Prediction and Diagnosis of Diabetes Mellitus using a Water Wave Optimization Algorithm

S. Taherian Dehkordi[1], A. Khatibi Bardsiri[2*] and M. H. Zahedi[3]

1. Department of Computer Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran.
2. Department of Computer Engineering, Bardsir Branch, Islamic Azad University, Bardsir, Iran.
3.Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran.

Received 20 November 2017; Revised 06 April 2018; Accepted 26 September 2018
**Corresponding author: a.khatibi@srbiau.ac.ir (A. Khatibi).*

Abstract
Data mining is an appropriate way to discover the information and hidden patterns in large amounts of data, where the hidden patterns cannot be easily discovered in normal ways. One of the most interesting applications of data mining is the discovery of diseases and disease patterns through investigating the patients' records. Early diagnosis of diabetes can reduce the effects of this devastating disease. A common way to diagnose this disease is to perform a blood test, which, despite its high precision, has some disadvantages such as: pain, cost, patient stress, and lack of access to a laboratory. Diabetic patients' information has hidden patterns, which can help you investigate the risk of diabetes in individuals without performing a blood test. The use of neural networks, as powerful data mining tools, is an appropriate method to discover hidden patterns in diabetic patients' information. In this work, in order to discover the hidden patterns and diagnose diabetes, a water wave optimization (WWO) algorithm, as a precise metaheuristic algorithm, was used along with a neural network to increase the precision of diabetes prediction. The results of our implementation in the MATLAB programming environment using the dataset related to diabetes indicated that the proposed method was capable of diagnosing diabetes at a precision of 94.73%, a sensitivity of 94.20%, a specificity of 93.34%, and an accuracy of 95.46%, and was more sensitive than the methods like support vector machines, artificial neural networks, and decision trees.

Keywords: *Diabetes Mellitus, Data Mining, Artificial Neural Networks, Water Wave Optimization (WWO) Algorithm.*

1. Introduction

The advance of technologies and human progress have been very extensive and rapid. Despite the benefits of these technologies, the lack of mobility and use of fast food have spread obesity among individuals in different societies, and obesity itself is also an important factor in the development of many diseases such as diabetes, hyperlipidemia, arteriosclerosis, strokes, and myocardial infarction [1, 2]. Diabetes mellitus can be called the disease of the century, as many people in the world are affected by this disease or are at its risk. Diabetes mellitus has very destructive effects on the health of individuals, and if diagnosed late, it can cause irreparable damages to the sense of sight, kidneys, heart, arteries, and so on [3].

The disease is not dangerous in itself but its complications and secondary diseases that it causes in the body and its various tissues make it known as one of the health threats of the present century.

There are three different types of diabetes mellitus. Type 1 diabetes has hereditary aspects, and only a small amount of insulin hormone, which plays a role in controlling sugar levels in the body, can be produced in this type of diabetes [4].

In type 2 diabetes, the amount of insulin in the body is sufficient but the body cells are not able to absorb blood sugar for various and even unknown reasons, which cause the level of blood sugar or glucose to increase in blood, and be deposited on vascular walls in the form of hazardous compounds, which is considered as one of the factors in the incidence of various strokes or heart attacks and amputation [5]. In the third type, which is known as gestational diabetes, blood sugar levels increase during pregnancy for various reasons such as: consumption of various types of corticosteroids, and if controlled, this type of diabetes disappears at the end of pregnancy [6] The highest incidence of diabetes can be considered to belong to type 2 diabetes, which occurs in the body due to the lifestyle and increased obesity. Unfortunately, diabetes has no visible symptoms until advanced stages of the disease. This disease can exist in a person's body for many years, without the person himself being aware of his illness. The disease is diagnosed when symptoms such as damage to sight and kidneys and failure of the body's wounds to heal, especially in the leg and foot areas, reveal the disease, at which time the therapeutic effects are no longer appropriate for healing the person [7].

The problem with the diagnosis of diabetes is that a person suspected of the disease has to undergo several blood tests, and even follow up and repeat these tests for several days or weeks, and after taking drugs or following diets, the blood tests have to be repeated over and over until they finally find out if the person is diabetic or healthy. Unfortunately, people are not willing to undergo clinical trials for different reasons, and the fact that these tests can be costly and time-consuming discourages them from undergoing the tests. Reducing the number of clinical trials or blood tests down to one or two cases and placing more reliance on diagnostic methods based on non-clinical trials such as weight, height, and gender is an effective way to diagnose diabetes, and a number of studies have been conducted in this regard [8].

One of the ways to diagnose the disease is to use datasets and information gathered by doctors, which can usually be found in the records of patients or those who come for a visit. Identifying disease patterns that are hidden in a large amount of information are possible with the aid of special tools such as data mining. Therefore, with the help of patients' information and the information of those who come to see a doctor, and subsequently exploring this data, one can achieve stunning results with the help of data mining, which is exclusive to knowledge discovery and provision of useful information [9]. In this paper, with the aid of an artificial neural network (ANN) as a data mining technique [10] and its combination with a metaheuristic water wave optimization (WWO) algorithm [11], we presented and described a new method for the diagnosis of diabetes. In the other words, the proposed system is used to increase the accuracy and effectiveness of prediction and classification of individuals into two healthy and patient categories and use the evolutionary algorithms of water waves. Given that evolutionary algorithms alone cannot detect hidden patterns and knowledge in various data such as diabetes, it is necessary to combine with one of the data mining techniques to be able to diagnose diabetes. The purpose of using a WWO algorithm in this study is to minimize the classification error rate for healthy subjects and patients in the ANN technique. Reducing the classification error rate for healthy subjects and patients is in fact one of the optimization problems, which increases the precision of distinguishing diabetic patients from healthy subjects. In this work, our intention was to select a neural network and put the diabetes disease data collection as a test data to this neural network, so that the neural network at test stages can create a proper output with the help of test data and optimization algorithm for water waves. An important innovation in this research work is the improvement of artificial neural network learning in diagnosing diabetes using water wave's meta-heuristic algorithm. In this article, the parameters of neural network such as its weight are determined at the same time as the training by the water-wave algorithm to optimize their values, and finally, the optimal parameters determined by water wave algorithm lead to a reduction in classification error in the patient than in the healthy persons. Evolutionary or meta-exploratory algorithms are used in many studies, because they do not require information about the derivative or gradient of the objective function of the problem. The WWO algorithms are also considered among the strong and precise evolutionary algorithms due to their strong theory and modeling. We discuss the history related to diabetes and other diseases at the beginning of this paper. Next, we will make a review of water wave optimization (WWO) algorithms. Then we will describe the proposed method for the diagnosis of diabetes. Finally, we will evaluate the proposed method, and compare it with other methods.

2. Research background

There are two main reasons for an abnormal increase in the level of the body's blood sugar. In the first type, the pancreas is not able to produce sufficient natural insulin, which leads to type 1 diabetes. In the second type, the pancreas produces sufficient insulin, but for whatever reason, the body cells are not able to absorb insulin, which results in type 2 diabetes [13].

According to the statistics of the World Health Organization (WHO), the highest percentage of diabetes is related to type 2 diabetes, which is about 90%, and its main cause is obesity and lack of mobility. People affected by this type of diabetes usually do not show any initial symptoms but after a few years, they will become aware of this hidden disease through other diseases.
In the long term, this disease damages the body and its tissues, which causes disorders in different organs of the body. Among the long-term complications of this disease is the development of diseases such as: cardiovascular diseases, kidney diseases, neurological diseases, and ocular diseases. It should be noted that cardiovascular diseases have the highest mortality rates in diabetic people. Important factors contribute to the development of type 2 diabetes such as: overweight and obesity, inactivity, high fat diets, consumption of low-fiber foods, race, genetics, and age. The risk of diabetes in men and women increases with increased weight [12]. Different studies have so far been conducted to diagnose diabetes through data mining and machine learning techniques, some of which we refer to in what follows.
In order to diagnose diabetes, Sankaranarayanan et al. [14] haveused two algorithms: FP-Growth and Apriori, along with associative laws. Using these two algorithms to discover laws, they discovered a set of laws and relationships in the Pima Indian Diabetes Database (PIDD), and presented it for the estimation of diabetes. In their proposed method, they presented a total of 25 laws for the diagnosis of diabetes based on the features of the PIDD dataset. For instance, it is pointed out in the fourth law that if a person's diastolic blood pressure is 50, the person is not diabetic.
Pangaribuan et al. [15] have used an extreme learning machine (ELM) to diagnose diabetes. The results of their study show that if the number of neurons in the hidden layers is considered zero, the precision of the diagnosis of diabetes increases, and with the number of neurons in the hidden layers increasing, the precision of the diagnosis of diabetes in the training and test datasets decreases. In order to evaluate their proposed method with regard to the diagnosis of diabetes, they compared their method, which was based on extreme learning, with a multi-layer back-propagation artificial neural network. The results of their study show that the precision of the diagnosis of diabetes in the training and test datasets, is higher in the extreme learning neural network than in the multi-layer neural network.
On the other hand, the precision levels of the extreme learning technique in the training and test datasets are 1.12 and 2.14 times those of the multi-layer artificial neural network, respectively. In addition to the higher precision of the extreme neural network compared with that of the multi-layer neural network in the diagnosis of diabetes, the diagnosis duration is also shorter in the extreme neural network than in the artificial neural network. They showed that the speed values of the extreme learning technique in the training and test datasets are 3.102 and 5.136 times those of the multilayer artificial neural network, respectively.
In order to diagnose diabetes, Kayaer et al. [16] have used three types of neural networks; i.e. a GRNN, MLP, and RBF neural network, along with the PIDD dataset. The results of their study show that the RBF neural network has the worst precision in the test dataset. The overall results of their study show that the GRNN neural network has the highest precision in its test dataset, which is about 82.21%. The overall results suggest that the GRNN neural network is an appropriate structure for the diagnosis of medical events.
In order to diagnose diabetes, Amatul Zehra et al. [17] have investigated the effect a pre-processing phase of the normalization type on five data mining techniques using the PIDD dataset. The results related to the precision of the diagnosis of diabetes without data pre-processing and with data pre-processing in five techniques: Bayesian network, multi-layer neural network, decision table, decision tree, and simple chart, show that applying pre-processing phases in the above five techniques, increases the precision of the true classification, and in the same way, reduces the precision of the false classification.
Kandhasamy et al. [18] have used four data mining techniques including decision tree J48, k-nearest neighbors (KNN) classifier, support vector machine, and random forest to diagnose diabetes.
The results of their research work indicate that diagnosis of diabetes with the aid of a decision tree has the highest precision compared with similar techniques, and this precision is about 82.73%. Among the results of their work is increasing the precision of diagnostic techniques through applying pre-processing phases and removing noise data in the PIDD dataset.
Different pattern recognition tools and data mining algorithms have so far been presented, among which Weka, Tanagra, and MATLAB can be mentioned. For instance, Rashedur et al. [19] have used three tools: Weka, Tanagra, and MATLAB along with various data mining techniques such as multi-layer neural network, Bayesian network, decision tree, and fuzzy lattice reasoning (FLR). in order to diagnose diabetes. The results of their

experiments show that the tool Tanagra has a higher precision of recognition than the tools Weka and MATLAB do, especially in the Bayesian network technique. In their work, they investigated the effects of different parameters and criteria for error evaluation, such as: mean square error, and mean error. in Weka. They investigated Weka and Tanagra in the diagnosis of diabetes in terms of runtime with different data mining tools, and the results obtained show that the multi-layer neural network takes a larger amount of time in both tools than the other algorithms do. In general, the processing time for the diagnosis of diabetes is longer in Weka than in Tanagra [20]. Dania and Abhari implemented classification models to diagnose type two diabetes [21]. They proposed Fuzzy Expert System (FES) that used the Fuzzy Inference System (FIS) model for incidence of diabetes. Their proposed system was applied to the Pima Indian Dataset and they used two data mining tools called WEKA and MATLAB. The results obtained showed tangible superiority of their method to both methods Support Vector Machine (SVM) and Regression.

3. Water Wave Optimization (WWO) Algorithm

Water waves and their propagation on the surface of sea and oceans are among the interesting natural phenomena, which can be described through physical laws and mathematical equations.
Various attempts have so far been made to model the water wave behavior using mathematical relations. These attempts date back 400 years. Isaac Newton can be considered the first scientist who conducted important studies on the equations of water waves and their propagation. The water wave optimization (WWO) algorithm is one of the new metaheuristic algorithms, which is modeled and created based on the physical laws governing water waves on the sea’s surface. Metaheuristic algorithms such as WWO algorithms consider the solutions to the problems as water waves, and update them in the search space of the problem in each iteration until finally the optimal solution is extracted in the form of strong waves.
A WWO algorithm has a set of rules, which are created based on the wavelength and sea depth.
Waves are actually energy. Energy, not water, moves across the ocean's surface. Water particles only travel in a small circle as a wave passes. For instance, the height of a water wave created on the sea completely depends on the sea depth. This relationship is proportional inversely. The greater depth of the sea means the lower height of the water wave, and vice versa. Figure 1 shows the relationship between the sea depth and amplitude of the waves.

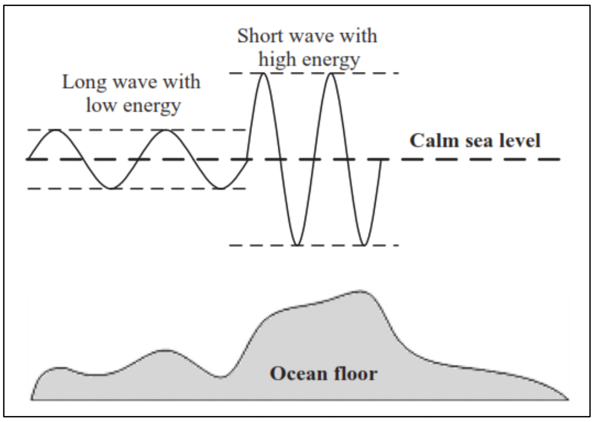

Figure 1. Relationship between the wave height and the seabed.

One of the common phenomena occurring in waves is the refraction phenomenon, which causes the wave to change its direction by 180 degrees, thus its energy is focused on the direction opposite the direction of the wave motion at the moment of collision.
The refraction phenomenon occurs as a result of the collision of water waves with a hard object such as a rock on the sea floor. In the refraction phenomenon, a wave undergoes a phase change, and its propagation turns opposite the initial direction, the main reason for which is its collision with hard obstacles.
A water wave refraction phenomenon, which is created by a seabed hump, can be clearly seen in Figure 2. Finally, the water waves have different shapes during their propagation, as shown in Figure 3, in a way that the wave created in the deep area is free at first, then as time passes and it moves towards the coastline, its amplitude increases and its wavelength decreases in a way that the wave is steepened and ultimately turns into breaking waves.

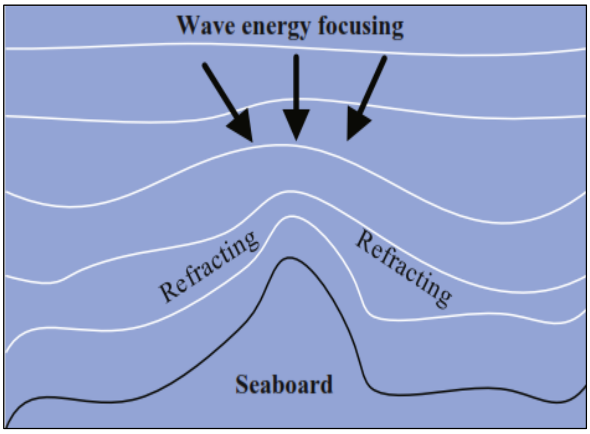

Figure 2. Concept of the wave refraction phenomenon and its relationship with the seabed.

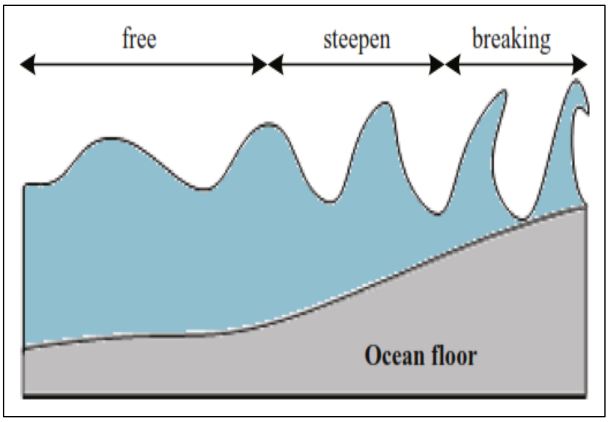

Figure 3. Changes in the wave shape relative to its propagation towards the coastline.

The WWO algorithm is a population-based algorithm. In this algorithm, each solution is considered a water wave, and first, each wave is scattered in the search space of the problem. Then, each water wave tries to discover the optimal solution to the target function based on the following three mechanisms [11]:
the following three mechanisms:

- Wave propagation
- Wave refraction
- Breaking of the wave

In this algorithm, each wave or problem solution has characteristics such as height and wavelength, which are regularly updated during its propagation in the search space of the problem.
In the WWO algorithm, the height and wavelength of a wave can be represented by $h \in \mathbb{Z}^+$ and $\lambda \in \mathbb{R}^+$, respectively, in a way that the wave height is assumed to be a positive integer, and the wavelength a positive decimal. Wave propagation can be considered the first phenomenon on a set of waves, which displaces them in the search space of the function. Propagation causes a problem solution to be updated to a new solution, and in general, it causes the population of water waves to change regularly relative to the algorithm's iteration, thus increasing their fit, which is calculated by the objective function. When modeling the behavior of water waves, shallow points can be considered as optimal points, and points with greater depths as non-optimal points. In each iteration of the algorithm, waves or problem solutions regularly move from deep areas toward the coastline or shallow areas, which are assumed optimal. In the meantime, some of the waves or solutions that reach the coastline or optimal points break, so that ultimately other new solutions are created in the search space of the problem.
In the WWO algorithm, propagation is applied separately to each wave of the population according to relation (1), so that these waves move in the search space of the problem [11].

$$x'(d) = x(d) + rand(-1.+1).\lambda L(d) \qquad (1)$$

where, λ is assumed as the current wavelength, rand(-1,+1) as a random number with uniform distribution in the interval [-1,+1], L(d) the range of changes in the dimension d, and *x(d)* and *x' (d)* the current and new positions of a wave in the search space of the problem or objective function. In wave propagation, a wave can reach a point more optimal or less optimal than the current point. With the propagation of each wave in the search space of the problem, its height and wavelength are required to be updated again. Hence, if a wave propagates to a point that is less optimal than the current point, the wave height, represented by h_max, is reduced by one unit. At the same time, after the propagation of each wave, its wavelength is updated according to relation (2) [11].

$$\lambda = \lambda.\alpha^{-(f(x)-f_{\min}+\varepsilon)/(f_{\max}-f_{\min}+\varepsilon)} \qquad (2)$$

where, f_{max} and f_{min} are considered as fit values for the fittest and unfittest wave of the population, respectively, α as the absorption coefficient of waves that is a constant in the WWO algorithm, and its value is determined approximately, and ϵ a small positive number that prevents the occurrence of division by zero in the fraction.
The propagation phenomenon is not the only phenomenon to be applied to the waves but waves can be refracted after propagation.
In order to refract waves, an arbitrary wave is selected. If its fit value is greater than that of the best wave of the population, then it will be chosen as the best wave of the population, and the refraction phenomenon will occur in it. Relation (3) is used for refraction [11]:

$$x'(d) = N\left(\frac{x^*(d) + x(d)}{2} . \frac{|x^*(d) + x(d)|}{2}\right) \qquad (3)$$

where, $x*$ and N(μ .σ) are defined as the best waves of the population, respectively, in terms of fit and random and normal distribution around the fit wave, with mean and standard deviation equal to μ and σ, respectively.
The refraction phenomenon, like propagation, affects the wavelength of waves, and if the evaluation or objective function is a minimum finding function, then the wavelength changes according to relation (4) [11]:

$$\lambda' = \lambda.\frac{f(x')}{f(x)} \qquad (4)$$

where, x, x^', f(x), and f(x^') are assumed to be the current position of a wave, new position of a wave based on refraction, and fit values for the current and new waves, respectively.

To put it simply, if the wave is refracted to a fitter area, the wavelength will decrease otherwise, the wavelength will increase. Another phenomenon that is applied to waves is the breaking phenomenon, which occurs when the wave height decreases and reaches zero. In this case, we remove the wave from the population and replace it with a new wave in the search space of the problem randomly. This phenomenon is modeled in relation (5) [11].

$$x'(d) = x(d) + N(0.1).\beta L(d) \tag{5}$$

where β, N(0.1), and $L(d)$ are assumed to be the breaking rate of the waves, a random number in the interval [0,1], and the motion limit of a wave in the dimension d, respectively.

The results of the experiments and implementation of the WWO algorithm on a set of evaluation or benchmark functions show that this algorithm is more precise than metaheuristic algorithms such as invasive weed optimization (IWO) algorithms, gravitational search algorithms (GSA), biogeography-based optimization (BBO) algorithms, and bat algorithm (BA). In general, in the optimization algorithm of water waves, as in any meta-heuristic algorithm or evolutionary algorithm, a set of initial solutions is encoded in the form of a population. In this meta-heuristic algorithm, each solution of any problem is identified in the form of a wave, and a set of waves are considered as the primary population of problem. In the optimization algorithm for water waves, each solution of problem or wave is encoded with characteristics such as wave height or wave amplitude or wavelength. In the wave water optimization algorithm, solutions of problem are first encoded in waves and randomly distributed in number of waves in the space of problem search [11].

In general, the mechanism of WWO algorithms can be considered based on the application of three phenomena: propagation, refraction, and breaking on a population of waves, being used as solutions to the problem, as follows: each wave can move in the search space of the problem under the influence of three phenomena: propagation, refraction, and breaking.

4. Proposed method

Diagnosis of diabetes through an artificial neural network and based on the characteristics of diabetic patients is, in fact, distinguishing patients from healthy subjects. In other words, it is considered a kind of classification problems, whose purpose is to minimize errors in the classification of healthy subjects and patients. In fact, this problem is an optimization problem, whose purpose is minimizing diagnostic errors or errors in the classification of healthy subjects and patients [22, 23]. Each neural network has important features such as weights and thresholds, based on which a neural network can map the input data to appropriate the output data.

The optimal selection of weights and thresholds can greatly reduce the errors of data classification, and increase the quality of classification and the precision of artificial neural networks. Through the process of learning using training data, an artificial neural network can, to a certain extent, select weights and thresholds in a way that they are optimal, and reduce the error rate of classification in the identification of diabetic patients. The quality of the weights and thresholds of an artificial neural network depends on the learning process of the artificial neural network. In most cases, weights and thresholds are selected in a way that they are not completely optimal. Hence, the error rate of classification increases, and diabetic patients are identified with little precision. The optimum weights and thresholds and/or the level of their optimality can be determined based on the minimality of the error rate of classifying diabetic patients. To put it in better words, selecting optimum weights and thresholds in an artificial neural network such as a multi-layer artificial neural network results in the reduced classification error rate. The reduced classification error rate can also be considered as a result of selecting better optimum weights and thresholds in an artificial neural network [24].

Finding optimal weights and thresholds in a multi-layer neural network is a function of the classification error rate.

Hence, in order to improve the quality of an artificial neural network and to reduce its error rate in the identification of diabetic patients, you can start with searching in threshold and weight values, and try to choose the best optimum weights among them, which results in a better minimization of the classification error rate. A multi-layer artificial neural network can be formulated based on relation (6):

$$f(W.x.b) = \sum_{i=1}^{l}\sum_{j=1}^{n}\sum_{k=1}^{m} W_{ij}.x_k + b_i \tag{6}$$

where l, n, and m are considered as the number of hidden layers, number of neurons in each hidden

layer, and number of inputs, respectively, and on the other hand, W_{ij}, x_k, and b_i are defined as weights, inputs, and the threshold of each hidden layer, respectively.
A multi-layer artificial neural network is optimal when its classification error rate is also minimum, in which case, the artificial neural network is shown in the form of relation (7):

$$f^*(W^*.x.b^*) = \sum_{i=1}^{l}\sum_{j=1}^{n}\sum_{k=1}^{m} W^*{}_{ij}.x_k + b^*{}_i \quad (7)$$

where W^* and b^* are optimum values for weights and bias in an artificial neural network, respectively, which minimize the classification error rate in the identification of diabetic patients. The structure of a multi-layer artificial neural network in fact consists of weights and bias used in it, an example of which is shown in figure 4. The coding of a multi-layer artificial neural network can be defined as an array of its weights and thresholds, the main advantage of which is that these components can be used as an initial population for a metaheuristic algorithm such as: a WWO algorithm. The mean classification error rate in a multi-layer artificial neural network depends on the selection of values for the thresholds and weights used in the artificial neural network. The optimal selection of these values causes the mean classification error rate as the objective function of the problem to be minimized as much as possible. In fact, minimizing the mean classification error rate is a hard and difficult problem in the diagnosis of diabetes. Here, we want to reduce this error rate through choosing optimum weights and thresholds using a WWO algorithm, and calculate a vector pair of optimum weights and thresholds or $(w^*.b^*)$ using the proposed algorithm.
In the proposed method, in order to minimize the error rate of distinguishing diabetic patients from healthy subjects, it is necessary to codify a multi-layer artificial neural network in the form of a population member of the WWO algorithm. Therefore, a list, an array or an indicator is used to codify a multi-layer artificial neural network in the form of a population member of the WWO algorithm, as shown in relation (8):

$$NN = \left[w_1.w_2......w_d.b_1.b_2.........b_d \right] \quad (8)$$

where each array in fact represents a multi-layer artificial neural network or a wave in the search space of the diabetes problem. Employing the WWO algorithm in the proposed method, first, the structure of a multi-layer artificial neural network with a certain number of hidden layers and neurons is defined. Then a number of artificial neural networks are created in the form of waves with random weights and thresholds, whose values are in the interval $[-1, +1]$. These artificial neural networks or waves are evaluated based on the test data, and their fit values are determined. In the proposed method, each multi-layer artificial neural network is codified in the form of a water wave with a position vector. At the beginning stages of the proposed method, the position vector of each particle is randomly created in the interval $[-1. +1]$. Multi-layer artificial neural networks or their equivalent waves can diagnose diabetes, when they can accurately determine the error rate of the classification of diabetic patients, and to put it in better words, they can reduce the classification error rate.
The classification error rate called mean square error (MSE) can be used to measure the fit value of each water wave or artificial neural network in the diagnosis of diabetes.
The quality of a multi-layer artificial neural network in the classification and identification of diabetes is typically measured using the mean classification error rate in the identification of diabetic patients, the criterion for which is expressed in relation (9):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (9)$$

where y_i is the actual class number of the i-th sample, $\hat{y}_i$ the estimated class number of the i-th sample, and n the number of test samples used in the proposed method. Through an appropriate codification of each multi-layer artificial neural network (ANN) in the form of a water wave, a WWO algorithm can be implemented on the ANN samples, whose stages are as follow:
First, a multi-layer artificial neural network with a certain number of hidden layers is codified in the form of an array of weights and thresholds. Here, each multi-layer artificial neural network is considered in the form of a water wave with a set of weights and thresholds.
Values of weights and thresholds in multi-layer neural networks or their equivalent waves are selected randomly. At this stage, the initial population of water waves is evaluated by the objective function of the problem.
Each water wave or its equivalent neural network is influenced by three phenomena; propagation, refraction, and breaking; in order to update the set of weights and thresholds used in it.
The fittest wave or multi-layer artificial neural network with a minimum error rate in the

identification of diabetic patients is selected in each iteration of the algorithm. The stages of the WWO algorithm are implemented on a population of multi-layer artificial neural networks until the final iteration in order to constantly reduce the error rate of the diagnosis of diabetes in the population of waves. The optimum neural network, which has been extracted, is evaluated with the aid of the diabetes dataset and test data. The purpose of applying a WWO algorithm to a set of multi-layer artificial neural networks as water waves is to update the values of weights and thresholds having been used in order to reduce the error rate in the diagnosis of diabetes. These stages are shown in the proposed flowchart in figure 5.

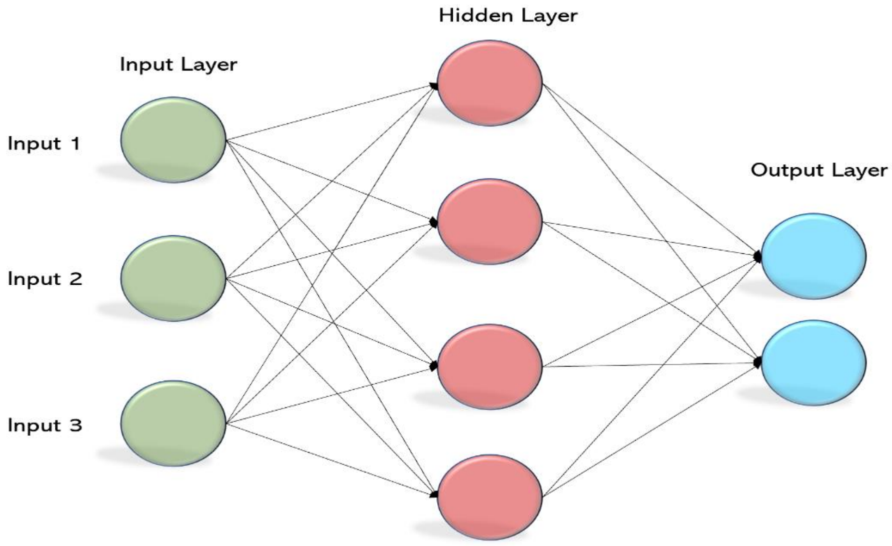

Figure 4. Structure of a multi-layer artificial neural network.

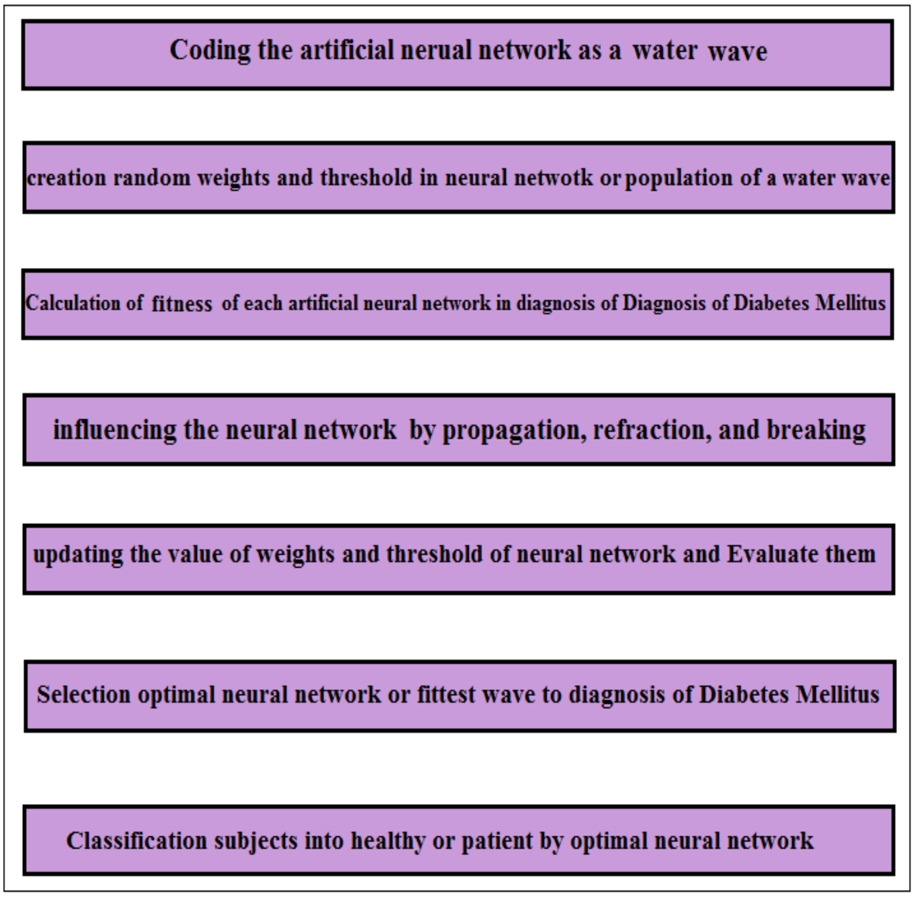

Figure 5. The flowchart proposed for the diagnosis of diabetes.

In figure 6, there is also a pseudo code of the proposed method for diagnosis of diabetes. Under the following code, you can find out the following steps for the proposed method:

- A neural network encoded as wave of water, each wave of water has a weight and a bias.
- Primary neural networks or waves were created and initial weights and bias are initialized randomly.
- In each repetition, the water waves were evaluated and best wave or neural network is selected to diagnose of disease, and suitable wave is a wave that has least error in diagnosing diabetes.
- Emission phases, reflection and dissociation were implemented, and weights and bias of the water waves were updated.
- In each replication, the best water wave, that has least error in diagnosing of diabetes, has been updated.
- In the final repetition of a suitable wave, neural network with least error is used to diagnose of diabetes.

```
Neural network coding in the form of a wave of water
Initial population of water waves randomly
It=1
While It<=IterMax do
 For each wave∈ Pop do
 Evaluate any wave or neural network with MSE
    Propagate wave to new wave
 If  new wave is better  wave
  If new wave is better best wave
Break new wave and set best wave= new wave
Else
Decrease Wave height= Wave height-1
If Wave heigh==0
Refract new wave
Update Wave heigh
End if
Endfor
It=It+1
End while
Decoding best wave to ANN
Evaluating ANN with test data
```

Figure 6. The pseudo code proposed for the diagnosis of diabetes.

5. Analysis of results

The water-wave optimization algorithm is a population-based method, and due to the fact that the problem space is searched by a set of solutions as well as a robust modeling of the propagation, reflection and dissociation is a proper and accurate algorithm in solving problems on optimization. To measure the convergence and accuracy, we can use a set of evaluation functions such as Sphere to determine its accuracy in comparison with the other algorithms. In figure 7, one example of implementation of the water-wave optimization algorithm on the sphere benchmark function and its comparison with valid meta-heuristic algorithms such as particle, bat, and night worm are shown. The evaluation and analysis of the test sample with a population of 30 and a number of repetitions of 100 were performed and the number of tests was also 25, so average error in terms of repetition in the algorithms is displayed.

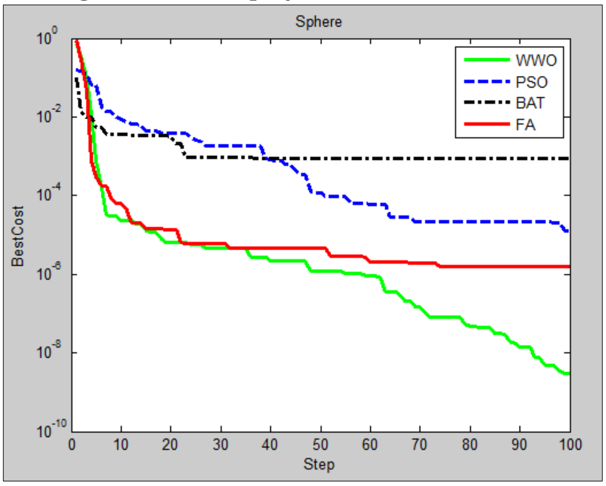

Figure 7. Comparison of convergence and accuracy of water wave algorithm.

The results of the experiment confirm that water wave algorithm with a less error than the other algorithms can find an optimal value.

In this section, the proposed method for the identification of diabetic patients is implemented using an appropriate dataset and programming environment, and its performance is determined based on the data mining criteria.

The MATLAB programming environment was used to measure the performance of the proposed method, and the Weka software environment as a powerful software program in pattern recognition and information classification was used to analyze the dataset and the other data mining and machine learning algorithms. A variety of datasets were presented to measure the performance of machine learning and data mining algorithms used for diabetes, among which the Pima Indian Diabetes Dataset (PIDD) is a global dataset for diabetes, which has 768 records, each of which keeps the information of a person who presents himself at the clinic [17].

This dataset has eight input features and an output feature, which determine whether the intended subject is healthy or patient. One of the two numbers 0 or 1 can be seen in the output of each record, which represent a healthy subject and a patient, respectively. This dataset is sometimes called PIMA as well. Since it is collected based on the Indo-Asian race. Features in this dataset included age of people, gender, number of pregnancies for women, weight, waist circumference, skin thickness, two hours sugar of body, minimum and maximum blood pressure,

genetic history of individual on disease, and ratio of body weight or volume [17].
This dataset is gathered and prepared from the UCI database, which has hundreds and thousands of valid datasets in different fields.
The reason for using the Indo-Asian race in the preparation of this dataset can be considered the higher rate of diabetes in this race than in other races.
Therefore, gathering information about diabetes required less time in this race than in other races.
Diagnosis of diabetes is in fact a type of problem for the classification of individuals into two categories: healthy subjects and patient. Therefore, in order to measure the performance of methods for the diagnosis of diabetes, the criteria for the evaluation of classification methods can be used in data mining. Various criteria such as precision, sensitivity, specificity, and accuracy can be used in order to measure the performance of an information classification algorithm.
We used leave-one-out cross-validation to validate the results. Leave-one-out cross-validation is K-fold cross-validation taken to its logical extreme, with K equal to N, the number of data points in the set. That means that N separate times, the function is trained on all the data except for one point, and a prediction is made for that point. As usual, the average error is computed and used to evaluate the model.
In order to measure the performance of the proposed method and other algorithms used for the diagnosis of diabetes, it is necessary to be familiar with the following concepts: the number of true positive (TP) samples, the number of false positive (FP) samples, the number of true negative (TN) samples, and the number of false negative (FN) samples. Therefore, in what follows, first, these concepts are explained. Then the criteria precision, sensitivity, specificity, and accuracy are also defined based on these concepts:
1. The number of true positive samples: Subjects who are diabetic patients and the proposed method has correctly diagnosed them as diabetic patients.
2. The number of false positive samples: Subjects who are healthy but the proposed method has wrongly diagnosed them as patients.
3. The number of true negative samples: Subjects who are healthy and the proposed method has correctly diagnosed them as healthy subjects.
4. The number of false negative samples: Subjects who are patients but the proposed method has wrongly diagnosed them as healthy subjects.
The decreased number of FP and FN samples and increased number of TP and TN samples indicate that the error rate in the classification and identification of diabetic patients is low, thus improving indicators such as precision, sensitivity, specificity, and accuracy. Relations (10), (11), (12), and (13), respectively show the evaluation criteria: precision, specificity, sensitivity, and accuracy for measuring the performance of the proposed method in the diagnosis of diabetes [23, 25, 26].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Specifity = \frac{TN}{TN + FP} \quad (12)$$

$$Sensivity = \frac{TP}{TP + FN} \quad (13)$$

$$\Pr ecision = \frac{TP}{TP + FP} \quad (14)$$

Each one of the indicators precision, specificity, sensitivity, and accuracy has a certain value in the interval [0. 1] in a way that number 1 represents the highest possible performance and number 0 the lowest possible performance. The above values can also be defined in percent in the interval [0, 1]. To this end, it is sufficient to multiply the above relations by 100. Figure 8 shows the output environment of the method proposed for the diagnosis of diabetes in the MATLAB programming environment.
The simulation environment in MATLAB was implemented in a way that the mean error rate of identifying diabetic patients versus the iteration of the WWO algorithm is shown as an output model.
In the said output, the mean error rate of the diagnosis of diabetes constantly decreases relative to the iteration, which indicates the weights and thresholds converge toward the optimum solution relative to the iteration of the WWO algorithm.
In the above picture, the mean error rate of identifying diabetic patients with a population of 10 subjects for 20 iterations ultimately reaches 0.209 in the proposed method, in a way that its downward trend constantly decreases from the first iteration until the last iteration. The increased size of the initial population is an important indicator that shows the increased precision of metaheuristic algorithms for solving different problems. For example, figure 9 shows the mean error rate of identifying diabetic patients with a population of 20 subjects for 20 iterations. In the output of the above example, the error rate for the initial population of 20 subjects and 20 iterations approximately reaches 0.196, which is less than that for the initial population of 10 subjects. To put it in better words, the increased size of the initial

population of water waves or their equivalent artificial neural networks increases the diversity of creating artificial neural networks (weights and thresholds used in them) in the search space of the problem of diabetes, which causes the problem space to be accurately searched, which, in turn causes the extracted weights and thresholds to be close to the global optimum value. In the graph of figure 10, the mean error rate of distinguishing diabetic patients from healthy subjects is shown for initial populations equal to 10, 20, 30, 40, and 50 in 50 different tests in order to determine the effect of reducing the error rate of classifying and distinguishing diabetic patients from healthy people relative to the size of the initial population of water waves.

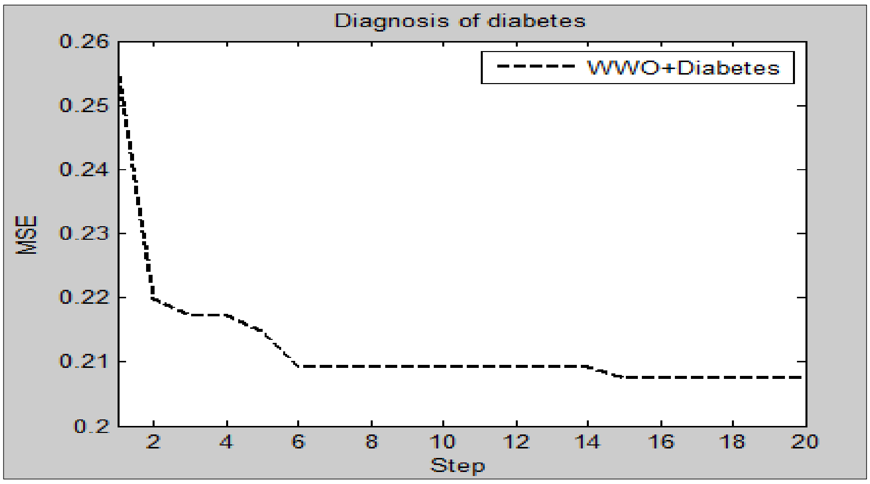

Figure 8. Reducing the mean error rate of identifying diabetic patients with a population of 10 subjects for 20 iterations in the proposed method.

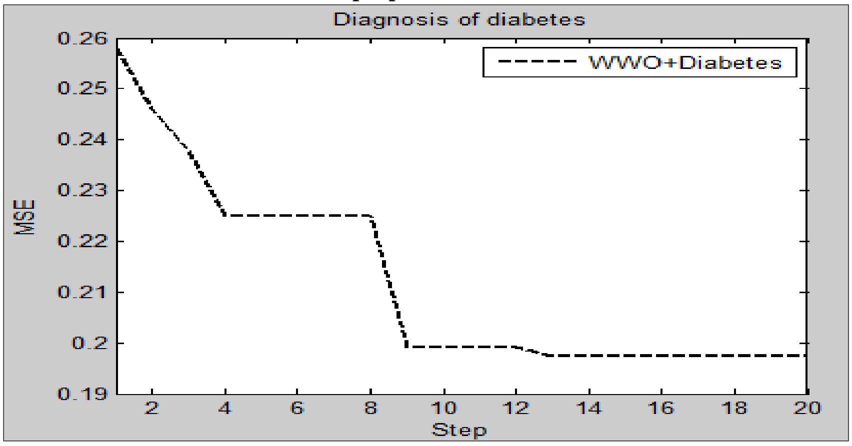

Figure 9. Reducing the mean error rate of identifying diabetic patients with a population of 20 subjects for 20 iterations in the proposed method.

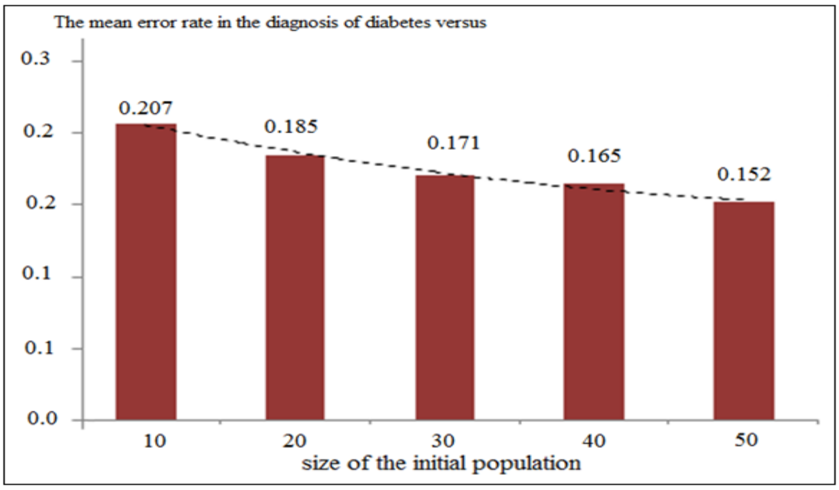

Figure 10. The mean error rate in the diagnosis of diabetes versus the size of the initial population of water waves.

According to the above graph, increasing the size of the initial population of water waves or their equivalent artificial neural networks causes the error rate in the diagnosis of diabetes to decrease constantly in a way that in populations equal to 10 and 50, the mean diagnostic error rates are 0.207 and 0.152, respectively, which shows a decrease of 26%. In order to measure the performance of the proposed method, we used the criteria: precision, sensitivity, specificity, and accuracy. We considered the size of the initial population equal to 70, the number of iterations equal to 10, and the number of tests equal to 50. Then we calculated the criteria precision, sensitivity, specificity, and accuracy in each test. Finally, we took their mean values into consideration. The mean values in the results of our tests show that the proposed method diagnoses diabetes at a precision of 94.73%, sensitivity of 94.20%, specificity of 93.34%, and accuracy of 95.46%. On the other hand, our statistical analysis in the WEKA environment also shows that the proposed method has a higher sensitivity and precision in the diagnosis of diabetes than methods such as: support vector machine, artificial neural network, and decision tree do. In figures 11, 12 and 13, respectively, the implementation of the back-up vector machine technique, artificial neural network, and decision tree in the diagnosis of diabetes are presented.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall
	0.898	0.459	0.785	0.898
	0.541	0.102	0.74	0.541
Weighted Avg.	0.773	0.334	0.769	0.773

Figure 11. Diagnosis of diabetes using support vector machine.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall
	0.832	0.392	0.798	0.832
	0.608	0.168	0.66	0.608
Weighted Avg.	0.754	0.314	0.75	0.754

Figure 12. Diagnosis of diabetes using artificial neural network.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall
	0.814	0.403	0.79	0.814
	0.597	0.186	0.632	0.597
Weighted Avg.	0.738	0.327	0.735	0.738

Figure 13. Diagnosis of diabetes using decision tree.

Comparison of the above output with our proposed method and our statistical analysis in the WEKA environment also suggests that the proposed method is more sensitivity to diagnosis of diabetes than supporting vector machines, artificial neural network, and decision tree. One of the indicators for assessing the knowledge discovery methods in diagnosing diabetes is the use of learning time and diagnosis of illness. On the other hand, the best time for implementation of algorithm can be used as a benchmark. Here, the proposed method uses back-up vector machine, artificial neural network, and decision tree on the diabetes data set and calculate the average learning time for them for 30 different tests, and specifies average of run time for comparing their average for comparison criterion. It is shown in figure 14.

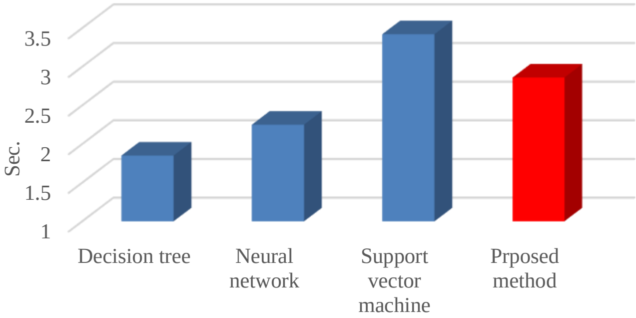

Figure 14. Comparing the runtime of the proposed algorithm with other methods.

According to the above diagram, it can be concluded that run time of the tree technique is less than other methods, and the vector machine method is more efficient than the other methods. On the other hand, run time of the proposed method is higher than artificial neural network, and this is quite predictable because proposed algorithm is a combination of neural network and water-wave algorithm.

6. Conclusion

Diabetes, as one of the widespread diseases in the current era, endangers the lives of many people every year. Physicians carry out different tests in order to diagnose diabetes, such as oral glucose tolerance test (OGTT), fasting plasma glucose (FPG) test, and random glucose test. The problem with the diagnosis of diabetes through clinical trials is that each one of these trials has to be performed many times in several sessions so that the physician can diagnose diabetes in a person. Performing this number of trials in several sessions causes a delay of several days in the diagnosis of diabetes. On the other hand, in addition to their costs, this number of tests wastes people's time for the diagnosis of diabetes and causes problems with their willingness to follow-up the disease. In the future research works, we try to use the system to detect other diseases such as cardiovascular disease. In this research work, a multi-layer artificial neural network was created by a diabetes dataset. From this artificial neural network and using the water-wave optimization algorithm in the form of an array of primary population members, a new method was proposed to increase the accuracy of diagnosis of diabetes. The results of the implementation in the MATLAB programming environment indicate that the proposed method with 94.73% accuracy, 94.20% sensitivity, 93.34% diagnosis, and 95.46% accuracy diagnose the diabetes, which show diabetes as accurate and with a less error than methods such as decision tree and diabetes vector machine. In this paper, based on the records of diabetic patients and the valuable information contained in these records, we presented a system based on an artificial neural network learned with the aid of a WWO algorithm so that based on the training data, it can provide an estimation of diabetic patients among the candidates only through two clinical trials. The advantage of the proposed system is its simplicity and almost negligible cost for the diagnosis of diabetes. In the future study, we will try to use this system to diagnose other diseases such as cardiovascular diseases.

References

[1] Richards, G., Rayward-Smith, V. J., Sönksen, P. H., Carey, S., & Weng, C. (2001). Data mining for indicators of early mortality in a database of clinical records. Artificial intelligence in medicine, vol. 22, no. 3, pp. 215-231.

[2] Woods, C. P., Hazlehurst, J. M., & Tomlinson, J. W. (2015). Glucocorticoids and non-alcoholic fatty liver disease. The Journal of steroid biochemistry and molecular biology, vol. 154, pp. 94-103.

[3] Tuttle, K. R., Bakris, G. L., Bilous, R. W., Chiang, J. L., De Boer, I. H., Goldstein-Fuchs, J., & Neumiller, J. J. (2014). Diabetic kidney disease: a report from an ADA Consensus Conference. American journal of kidney diseases, vol. 64, no. 4, pp. 510-533.

[4] Sperling, M., Grzelak, T., Pelczyńska, M., Jasinska, P., Bogdanski, P., Pupek-Musialik, D., & Czyzewska, K. (2016). Concentrations of omentin and vaspin versus insulin resistance in obese individuals. Biomedicine & Pharmacotherapy, vol. 83, pp. 542-547.

[5] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences, vol. 29, no. 2, pp. 93-99.

[6] Gotz, D., Wang, F., & Perer, A. (2014). A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. Journal of biomedical informatics, vol. 48, pp. 148-159.

[7] Shah, B. R., & Lipscombe, L. L. (2015). Clinical diabetes research using data mining: a Canadian perspective. Canadian journal of diabetes, vol. 39, no. 3, pp. 235-238.

[8] Lee, Y. C., Lee, W. J., & Liew, P. L. (2013). Predictors of remission of type 2 diabetes mellitus in obese patients after gastrointestinal surgery. Obesity research & clinical practice, vol. 7, no. 6, pp.494-500.

[9] Zhuo, X., Zhang, P., & Hoerger, T. J. (2013). Lifetime direct medical costs of treating type 2 diabetes and diabetic complications. American journal of preventive medicine, vol. 45, no. 3, pp. 253-261.

[10] Zolbanin, H. M., Delen, D., & Zadeh, A. H. (2015). Predicting overall survivability in comorbidity of cancers: A data mining approach. Decision Support Systems, vol. 74, pp. 150-161.

[11] Munro, K., Miller, T. H., Martins, C. P., Edge, A. M., Cowan, D. A., & Barron, L. P. (2015). Artificial neural network modelling of pharmaceutical residue retention times in wastewater extracts using gradient liquid chromatography-high resolution mass spectrometry data. Journal of Chromatography A 1396, pp.34-44.

[12] Zheng, Y. J. (2015).Waterwave optimization: a new nature-inspired metaheuristic.Computers & Operations Research, vol. 55, pp. 1-11.

[13] Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: a survey of the literature. Journal of medical systems, vol. 36, no. 9, pp. 2431-2448.

[14] Knight, K., Badamgarav, E., Henning, J. M., Hasselblad, V., Gano Jr, A. D., Ofman, J. J., & Weingarten, S. R. (2005). A systematic review of diabetes disease management programs. Am J Manag Care, vol. 11, no. 4, pp. 242-50.

[15] Sankaranarayanan, S., & Perumal, T. P. (2014). Diabetic Prognosis through Data Mining Methods and Techniques. International Conference on Intelligent Computing Applications (ICICA), pp.162-166, IEEE, Coimbatore, India, 2014.

[16] Pangaribuan, J. J. (2014). Diagnosis of diabetes mellitus using extreme learning machine. International Conference on Information Technology Systems and Innovation (ICITSI), pp. 33-38, IEEE, Bandung, Indonesia, 2014.

[17] Kayaer, K., & Yıldırım, T. (2003). Medical diagnosis on Pima Indian diabetes using general regression neural networks. International conference on artificial neural networks and neural information processing (ICANN/ICONIP), vol. 181, pp. 181-184, Istanbul, Turkey, 2003.

[18] Amatul, Z., Asmawaty, T., Kadir, A., & MAM, A. (2013). A Comparative Study on the Pre-Processing and Mining of Pima Indian Diabetes Dataset. ICSEC 2014 FSKKP, pp.1-10.

[19] Kandhasamy, J. P., & Balamurali, S. (2015). Performance Analysis of Classifier Models to Predict Diabetes Mellitus. Procedia Computer Science, vol. 47, pp. 45-51.

[20] Rahman, R. M., & Afroz, F. (2013). Comparison of various classification techniques using different data mining tools for diabetes diagnosis. Journal of Software Engineering and Applications, vol. 6 no. 3, pp. 85-97.

[21] Almadni, D., & Abhari, A. (2016). Comparative analysis of classification models in diagnosis of type 2 diabetes. In Proceedings of the Modeling and Simulation in Medicine Symposium (p. 7). Society for Computer Simulation International.

[22] Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M. & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics, vol. 97, pp. 120-127.

[23] Fatima, M., & Pasha, M. (2017). Survey of Machine Learning Algorithms for Disease Diagnostic. Journal of Intelligent Learning Systems and Applications, vol. 9, no. 1, pp. 1-16.

[24] Samant, P., & Agarwal, R. (2018). Machine learning techniques for medical diagnosis of diabetes using iris images. Computer Methods and Programs in Biomedicine, vol. 157, pp. 121-128.

[25] Moslehi, F., Haeri, A., Moini, A. (2018). Analyzing and Investigating the Use of Electronic Payment Tools in Iran using Data Mining Techniques. Journal of AI and Data Mining, vol. 6, no. 2, pp. 417-437.

[26] Karimian, F., Babamir, S. (2017). Evaluation of Classifiers in Software Fault-Proneness Prediction. Journal of AI and Data Mining, vol. 5, no. 2, pp. 149-167.

نشریه هوش مصنوعی و داده کاوی

پیش‌بینی و تشخیص بیماری دیابت با استفاده از الگوریتم بهینه‌سازی امواج آب

سمیه طاهریان دهکردی[۱]، عمید خطیبی بردسیری[۲،*] و محمد هادی زاهدی[۳]

[۱] گروه مهندسی کامپیوتر، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران.

[۲] گروه مهندسی کامپیوتر، واحد بردسیر، دانشگاه آزاد اسلامی، بردسیر، ایران.

[۳] دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی خواجه نصیر طوسی، تهران، ایران.

ارسال ۲۰۱۷/۱۱/۲۰؛ بازنگری ۲۰۱۸/۰۴/۰۶؛ پذیرش ۲۰۱۸/۰۹/۲۶

چکیده:

داده‌کاوی یک روش مناسب برای کشف اطلاعات و الگوهای پنهان در میان حجم عظیمی از داده‌هاست؛ الگوهای پنهانی که با روش‌های مرسوم و عادی قابل دستیابی نیستند. یکی از جالب‌ترین کاربردهای داده‌کاوی کشف بیماری و الگوهای آن با استفاده از رکوردهای ذخیره شده بیماران است. کشف زودهنگام بیماری می‌تواند به میزان قابل توجهی از اثرات مخرب آن کم کند. یکی از راه‌های معمول برای تشخیص بیماری دیابت استفاده از آزمایش خون است که علی‌رغم دقت بالا دارای نقاط ضعفی چون درد، هزینه، استرس بیمار و نبود آزمایشگاهی است. اطلاعات بیماران دیابتی حاوی الگوهایی است که می‌تواند فرآیند تشخیص دقیق بیماری را مستقل از انجام آزمایش‌های پزشکی تضمین کند. استفاده از شبکه‌های عصبی به عنوان یک ابزار داده‌کاوی قوی یک روش مناسب برای کشف الگوهای پنهان بیماری دیابت است. در این مقاله از الگوریتم بهینه‌سازی امواج آب به عنوان یک الگویتم فرااکتشافی مناسب برای بالا بردن دقت تشخیص بیماری دیابت از طریق شبکه‌های عصبی مصنوعی استفاده شده است. نتایج به دست آمده در محیط نرم‌افزار متلب و دیتاست‌های بیماری دیابت، دقت بالای مدل پیشنهادی را نسبت به روش‌های مشابه دیگر مثل ماشین بردار پشتیبان، درخت تصمیم و شبکه عصبی نشان می‌دهد. مقادیر به دست آمده برای چهار معیار ارزیابی صحت، حساسیت، ویژگی و دقت به ترتیب عبارتند از ۹۴,۷۳، ۹۴,۲۰، ۹۳,۳۴ و ۹۵,۴۶ درصد.

کلمات کلیدی: بیماری دیابت، داده‌کاوی، شبکه‌های عصبی مصنوعی، الگوریتم بهینه‌سازی امواج آب.