An Artificial Neural Network Model to Diagnosis of Type II Diabetes

Sareh Mortajez, Amir Jamshidinezhad*

Department of Health Information Technology, Faculty of Para Medical Sciences, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

ABSTRACT

Introduction: Diabetes is a disease caused by an increase in blood glucose levels due to insulin secretion deficiency (type 1 diabetes) or impaired insulin activity (type 2 diabetes). More than 90% of people with this condition are diagnosed with type 2 diabetes. Due to sharply prevalence of type 2 diabetes in recent years, the prognosis and early diagnosis of the disease have become even more important. In this study, a model for diagnosis of type 2 diabetes was developed using Artificial Neural Network (ANN) method.

Objectives: Minimizing the diagnosis faults of diabetes disease, using a hybrid ANN and the Genetic optimization algorithm.

Method: In this study, a hybrid ANN-Genetic Algorithm model was developed for classification of diabetic patients. Therefore, the number of optimal neurons as well as hidden layers was determined to design the architecture of the ANN model. To reduce the mean square error of the MSE network and optimize the accuracy of the diagnostic system a Genetic Algorithm (GA) was combined with the proposed ANN model. For experiment process, the model was considered on a dataset included 768 samples to diagnose the patients with type II diabetes from other cases.

Findings: The results showed a precision of 85% for diagnosing of type-2 diabetic patients. The proposed structure based on the lower mean square error of the MSE, indicated the best performance of the ANN with the MSE rate of 0.155.

Conclusion: The developed intelligent model showed an effective performance in comparison with existing methods with a minimum error and maximum confidence in the diagnosis process of diabetic disease.

Key words: Diagnosis of diabetes, Type-2 diabetes, Neural networks, Genetic optimization algorithms

INTRODUCTION

Diabetes mellitus or diabetes is one of the most commonly diagnosed diseases in the world. In fact, diabetes is a disease in which the production or function of the insulin, or both, is impaired in the body. About 220 million people worldwide are reported to suffer from this illness [1-3]. Diabetes mellitus, if not controlled, causes complications such as metabolic disorders and weight loss, ocular disorder, kidney disease, neurological damage, and cardiovascular disease because of high blood sugar levels [4]. The second type of diabetes is much more dangerous, which is growing in recent years. Data and information provided by health care clinics can be used as a suitable source for research in the field of diagnosis. Diabetes mellitus is one of the most commonly diagnosed illnesses with preconditions such as frequency of pregnancy, age, sex, blood pressure, heart rate, and other factors that can be used to diagnose the disease [5]. Analysis of these data seems to be somewhat unlikely due to the lack of data organization and the existence of hidden patterns of such information without specialized tools. Data mining can be defined as a method of discovering useful information from unorganized data without explicit information. The study of data used in diabetes patients' files in the form of a standardized data set allows researchers to compare their work [6]. Diagnosis of diabetes is possible with medical tests performed on several occasions, but these methods are not well received due to time, cost, and even fear of different clinical trials. When the disease is diagnosed, unfortunately its dangerous symptoms, such as vision damage, kidney and amputation would be occurred. Developing software systems that can provide a minimum of blood tests with high precision and, low cost and time to estimate the amount of diabetes or blood sugar is an interesting method. Diabetes mellitus has a series of intrinsic properties and an output feature, which is characterized by the input characteristics of the out-of-date (healthy-patient) attribute [7]. Today, many efforts have been made to prevent diabetes. Common approaches to using Artificial Intelligence (AI) and data mining techniques and strategies play a crucial role in identifying a variety of diseases. The discovery and analysis of knowledge using intelligent systems improved the quality

of service and reduced the human faults in the diagnosis of diabetic patients [8,9]. In this study, we intend to use ANN and GA to enhance the diabetes diagnostic performance. The proposed system used the concepts of genetic optimization algorithm to improve the effectiveness of prediction and classification of patients into type-2 diabetes and non-diabetes categories. As the optimization algorithms are unable to detect hidden knowledge in various data such as diabetes, it is necessary to combine with the data mining techniques to be able to diagnose diabetes. The goal of using the GA with the ANN was to create a more accurate intelligent system for diagnosing diabetes.

The present research seeks to answer such questions:

- How can I turn existing data into useful information for adopting intelligent clinical decisions?
- How a high-and-low-cost model can be developed to diagnose and treat diseases, and in particular diabetic disease based on artificial intelligence solutions?

The existence of approaches to diagnosing the diabetic disease caused the present study provided an ANN based GA for diabetic disease diagnosis.

Artificial neural networks

An ANN is a data processing system that is thought of as a human brain. Processing data is carried out by the interconnected processors that work in parallel to solve a complex problem. In these networks, with the help of programming knowledge, a data structure is designed that can act as a neuron. This is called neuronal structure. Next, they learn the network by creating a network between these neurons and applying a learning algorithm [10].

In this memory or network, neurons have two active modes and each edge, synapse, or the relationship between the nodes has a weight. The edges with positive weights stimulate or activate the next passive node, and negatively weighted edges make the next connected node inactive or inert (if it is active). A neural network including the component Layers and weights. The behavior of the network also depends on the relationship between the members [11].

In general, there are three types of neuronal layers in the neural networks. An input layer that receives raw data that is fed to the network, Hidden layers whose performance is determined by the inputs and the weight of the relation between them, and the hidden layers. The weights between inbound and hidden units determine when a hidden unit should be activated. The third layer of output, the output unit function is hidden depending on the activity of the unit and the weight of the connection between the hidden and output unit [11]. There are also single-layer and multi-layer networks that use the single-layer organization in which all units are connected to a layer. And has more computational potential than multi-layer organization. In multi-layer networks, units are numbered by layers. Both layers of a network are interconnected by weights and, in fact, connections [12].

There are several types of relationships or weighted relationship in the neural networks. In the feed forward relation networks, the connections move the signals in one direction only. There is no feedback from the input to the output (the loop), the output of each layer does not affect the same layer. In the backward relation networks, connections are continuous. Therefore, data from the top nodes to the lower layer nodes are feed-backed. The third type of relations connects each layer output to the input of next layer for the same nodes [10,13]. Figure 1 shows the structure of the ANN.
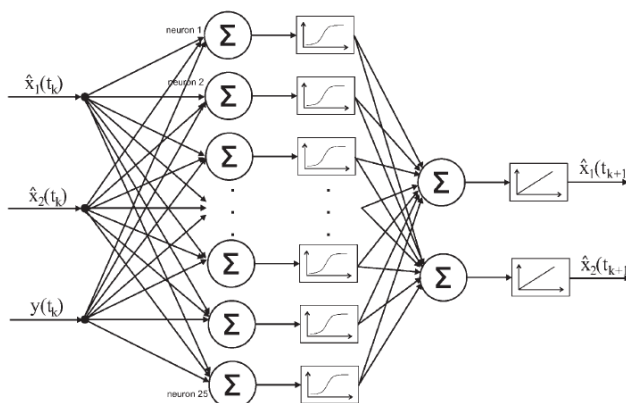

Figure 1: Architecture of an artificial neural network

Genetic algorithms

Genetic Algorithms (GAs) are inspired by nature as an optimization algorithm in many sciences to explore optimal problem solving [14]. The basis of this algorithm is Darwin's natural selection theory. It is an appropriate tool for optimizing discrete and non-linear functions, and also an influential search method in large spaces [15].

GAs begins its Global search with a primary population of zero and one, with the initial population size dependent on the features of the problem's optimization. In the genetic algorithm, the initial population is randomly generated. Each strand is called the chromosome population, and each binary of each strand is called a gene. We must build a new population of the evolution of the initial population. There are three basic functions of selection, composition and mutation. Figure 2 shows the structure of the GAs.

At the selection stage, a batch of chromosomes is selected based on their fitness from the previous population. The fittest chromosome has more chances to choose in the next generation. In the crossover stage, the crossover operator is used to exchange information between two chromosomes. Typically, the crossover operates on a pair of chromosomes, and two children are produced for each pair. The mutation phase is implemented for preventing of local optimal results and searches for more areas of the answer. The termination conditions of the GAs are the achievement of a consistent number of generations or the discovery of a solution with the optimal amount, or the highest degree of fitness for the offspring [16].
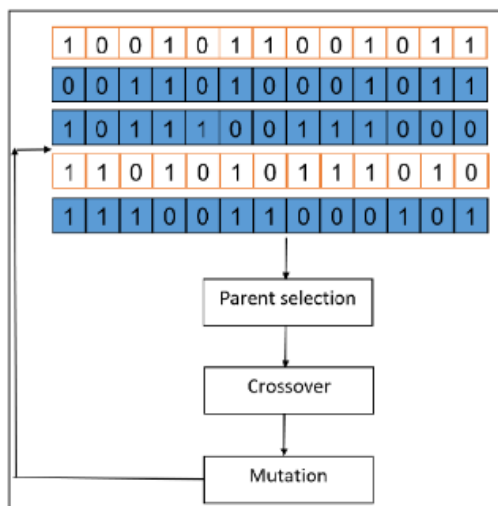
Figure 2: Process of the genetic algorithms

Diagnosis systems

Classifying and predicting are two types of operations for analyzing data and extracting the model into important categories of data, to understanding and predicting their future behavior. Classification models are used in the analysis of discrete and classical data. Classification known as supervised technique, is the process of finding a model that can detect unknown categories of data objects by identifying categories or data concepts [17]. This learning function maps a data item to one of the predefined categories. In this study, available data was divided into two parts: training and test. The training data was used to learn the rules by the system, and test dataset was used to check the accuracy of the model.

Diagnosis of diabetes is considered as an effective and practical way to diagnose diabetes using data from medical centers and data mining tools. And it is able to diagnose diabetes at a short time and accurately. An artificial neural network is one of the most important tools for diagnosis of diabetes, which minimizes classifying errors and increases diagnostic accuracy of diabetes. Diagnosis of diabetes can be defined as a two-class classification problem .The purpose is to specify the class type based on the characteristics of the patient. In this research, the intention is to perform a diagnosis and classification using the neural network, and optimize this diagnosis using the genetic optimization algorithm to achieve the highest possible accuracy in the diagnosis.

MATERIALS AND METHODS

In the proposed model, a multi-layer artificial neural network was developed. To find the best architecture of ANN model the training outcomes from several experimental processes were considered. The proposed ANN system included one hidden layer architecture with several nodes or neurons in each layer. The proposed model was developed on the MATLAB (R2015b) environment. The diagnosing system was trained a by training dataset to diagnose the diabetes patients. By training the artificial neural network, the weight values of

the multi-layer neural network were adjusted. To optimize the neurons, GA was used to find the best structure of the weights in the proposed diagnosing system.

Data set

The population of this study was a data collection of the Asian diabetic patients called Pima [16]. The data set included 768 people, of which 500 healthy women and 268 women suffered from type 2 diabetes. In this study, 8 factors were used to diagnose of diabetes. Table 1 shows the diabetes risk factors.

Table 1: Diagnostic factors for diabetes

Diabetes Factors
Frequency of childbirth for women
Blood glucose concentration
Diastolic blood pressure
Thickness of the musculoskeletal muscle
Insulin in a two-hour serum
Body Mass Index (BMI)
Having a history of diabetes
Age

To consider the validity of the model, the diabetes data samples were normalized then the samples divided to the training and testing data sets by 70% to 30%, respectively. Therefore, training dataset was used to find the optimum structure of the diagnosing system. Moreover, testing data set was used to experiment the model with the unseen samples to evaluate the validity of diagnosing system.

RESULTS

In this study, we implemented the simulation process with one hidden layer in several experiments including the number of neurons 2, 3, 4, and 5. Table 2 shows the results of the neural network implementation with several neurons for the hidden layer. The proposed structure based on the lower mean square error of the MSE, indicated the best performance of the ANN.

Table 2: Evaluation of network performance based on the number of hidden layer neurons

MSE	Number of neurons
0.171	5
0.155	4
0.176	3
0.173	2

As shown in Table 3, the proposed ANN-GA based diagnostic system compared with ANN for the Pima Diabetes Data Collection. The simulation results showed that the proposed model obtained the performance of

84.5% for diabetes diagnosis. The previous study showed the accuracy of less than 80% to diagnose of diabetes using similar dataset [18].

Table 3: Comparison of diabetes diagnostic accuracy- Pima dataset

Accuracy (%)	Diagnostic Method
78.11	Artificial Immune System [18]
79.37	Fuzzy System [19]
76	Distributed Time Delay Networks [20]
68	Feed forward Neural Networks [21]
84.5	Proposed System

DISCUSSION

With the increasing growth of diabetes (type-1 or type-2) among the human, the need for developing the artificial intelligence models to assist the disease diagnosis is raised. Artificial intelligence models with low computational costs, low limitations and high accuracy are the proper techniques to use for the differentiate diagnosis [17]. Immune system algorithm was an artificial intelligent technique used to diagnose the diabetic patients with the similar dataset of this study [18]. A fuzzy diagnosing method also used for similar diabetic patients' dataset with the accuracy of 79.37% which was lower than the accuracy obtained in this study [19]. Moreover, a feed forward neural network and Distributed Time Delay Networks were proposed to classify the diabetic patients and non-patients categories with the accuracy of 68% and 76%, respectively [20,21]. Achieved results in this research illustrated that the proposed hybrid ANN-GA model classified the type-2 diabetic patients from others with the accuracy of 84.5% with lower diagnostic risk in comparison with the current techniques.

CONCLUSION

We proposed a clinical decision support model based on the ANN to diagnose the diabetic patients. Results obtained from the developed model showed improvement of diagnosing methods for classifying the type II diabetic patients from other categories. The proposed hybrid neural network based on the optimization algorithm had better performance than existing neural networks, fuzzy model and artificial immune system considered in the this study. Therefore, Genetic algorithm as an optimizing technique adjusted the weights of the neurons, considerably that used in the developed diagnostic model. As a result, this intelligent system can assists the medical experts to reach the better understanding of the patients with suspicious diabetes to diagnose type II diabetes.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

1. Vijayan V, Ravikumar A. Study of data mining algorithms for prediction and diagnosis of diabetes mellitus. IJCA 2014; 95.

2. Karthikeyan T, Vembandasamy K. A novel algorithm to diagnosis type II diabetes mellitus based on association rule mining using MPSO-LSSVM with outlier detection method. Indian J Sci Technol 2015; 8:310-20.

3. Karthikeyan T, Vembandasamy K, Raghavan B. An intelligent type-ii diabetes mellitus diagnosis approach using improved FP-growth with hybrid classifier based arm. Res J Appl Sci Eng Technol 2015; 11:549-58.

4. Tudor I. Association rule mining as a data mining technique. Seria Matematică-Informatică–Fizică-Bul 2008; 1:49-56.

5. Kuo CH, Chen SC, Fang CT, et al. Screening gestational diabetes mellitus: The role of maternal age. PloS One 2017; 12:e0173049.

6. Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. Artif Intell Med 2002; 26:37-54.

7. American Diabetes Association. 2. Classification and diagnosis of diabetes. Diabetes Care 2017; 40:S11-24.

8. Huang Y, McCullagh P, Black N, et al. Feature selection and classification model construction on type 2 diabetic patients' data. Artif Intell Med 2007; 41:251-62.

9. Temurtas H, Yumusak N, Temurtas F. A comparative study on diabetes disease diagnosis using neural networks. Expert Syst Appl 2009; 36:8610-5.

10. Atkinson PM, Tatnall AR. Introduction neural networks in remote sensing. Int J Remote Sens 1997; 18:699-709.

11. Yegnanarayana B. Artificial neural networks. PHI Learning Pvt Ltd 2009.

12. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006; 313:504-7.

13. Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. IJAE 2011; 1:111-22.

14. Jamshidnezhad A, Nordin MJ. Bee royalty offspring algorithm for improvement of facial

expressions classification model. Int J Bio-Inspir Com 2013; 5:175-91.

15. Lucasius CB, Kateman G. Understanding and using genetic algorithms Part 1. Concepts, properties and context. Chemometr Intell Lab Syst. 1993; 19:1-33.

16. Pond SLK, Posada D, Gravenor MB, et al. GARD: A genetic algorithm for recombination detection. Bioinformatics 2006; 22:3096-8.

17. Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. JILSA 2017; 9:1.

18. odds.cs.stonybrook.edu/pima-indians-diabetes-dataset/

19. Ghasemi B, Koohestani B, Sarbaz Y. Diagnosing of type 2 diabetes using artificial immune system. 2nd International congress in computer engineering and information technology, Tehran, Allameh Majlesi University 2017.

20. Lekkas S, Mikhailov L. Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases. Artif Intell Med 2010;50:117-26.

21. Bozkurt MR, Yurtay N, Yilmaz Z, et al. Comparison of different methods for determining diabetes. Turk J Elec Eng Comp Sci 2014; 22:1044-55.