CLASSIFICATION OF PREDIABETES AND TYPE 2 DIABETES USING ARTIFICIAL NEURAL NETWORK

Dijana Sejdinović[1], Lejla Gurbeta[1,2], Almir Badnjević[1,2,3,4], Maja Malenica[5], Tanja Dujić[5], Adlija Čaušević[5], Tamer Bego[5], Lejla Divović Mehmedović[6]

[1] International Burch University, Sarajevo, Bosnia and Herzegovina
[2] Medical device inspection laboratory Verlab Ltd. Sarajevo, Bosnia and Herzegovina
[3] Faculty of Electrical Engineering, University of Sarajevo, Bosnia and Herzegovina
[4] Technical Faculty, University of Bihać, Bosnia and Herzegovina
[5] Faculty of Pharmacy, University of Sarajevo, Bosnia and Herzegovina
[6] Clinical Center University of Sarajevo, Bosnia and Herzegovina

Abstract. In this paper development of Artificial Neural Network for classification of prediabetes and type 2 diabetes (T2D) is presented. For development of this system 310 samples consisting of information about Fasting Plasma Glucose (FPG) and blood test called HbA1c were used. All samples were obtained from several healthcare institutions in Bosnia and Herzegovina, and diagnosis of prediabetes, T2D and healthy patients in this dataset were established by medical professionals. Two-layer feedforward backpropagation network with 15 neurons in hidden layer and sigmoid transfer function, used for classification of prediabetes and T2D in this paper, was trained with 190 samples.
Testing of developed neural network was performed with 120 samples for validation also obtained from healthcare institutions in Bosnia and Herzegovina. Out of 120 samples, developed network was accurate in 94.1% cases for the classification of prediabetes and in 93.3% cases for classification of T2D.

Keywords: prediabetes, diabetes type 2, Fasting Plasma Glucose, HbA1c, classification, Artificial Neural Network, Pattern Recognition

1 Introduction

The rapid global rise in the prevalence of type 2 diabetes (T2D) constitutes a health threat to the individual and is a major burden for health economy. Patients with diabetes can suffer from diverse complications that seriously erode quality of life [1]. Rates of T2D have increased markedly since 1960 in parallel with obesity [2]. As of 2013 there were approximately 368 million people diagnosed with the disease compared to around 30 million in 1985 [3]. Diabetes is a chronic and progressive metabolic disorder and according to International Diabetes Federation (IDF) in 2015 there were 415 milion patients diagnosed with diabetes, and it is expected that the rate will increase to 642 milions by 2040 [4]. Prediabetes is the precursor stage before diabetes mellitus in which not all of the symptoms required to diagnose diabetes are present, but blood sugar is abnormally high. This stage is often referred to as the "grey area" [5].

According to National Institute of Diabetes and Digestive and Kidney Diseases T2D, is a long term metabolic disorder that is characterized by high blood sugar, insulin resistance, and relative lack of insulin. Common symptoms include increased thirst, frequent urination, and unexplained weight loss. Long-term complications from high blood sugar include heart disease, strokes, diabetic retinopathy which can result in blindness, kidney failure, and poor blood flow in the limbs which may lead to amputations [6].

T2D primarily occurs as a result of obesity and not enough exercise. Some patients are more genetically at risk than others. According to National Institute of Diabetes and Digestive and Kidney Diseases it makes up about 90% of cases of diabetes, with the other 10% due primarily to diabetes mellitus type 1and gestational diabetes [6]. Diagnosis of diabetes is by blood tests such as fasting plasma glucose, oral glucose tolerance test, or HbA1C. Typically it begins in middle or older age, although rates of T2D are increasing in young people. T2D is associated with a ten-year-shorter life expectancy [7].

Diabetes disease diagnosis via proper interpretation of the diabetes data is an important classification problem. Hasan Tamurtas et al [8] in their study presented comparative pima-diabetes disease diagnosis. For that purpose, a multilayer neural network structure which was trained by Levenberg–Marquardt (LM) algorithm and a probabilistic neural network structure is used [8].

Furthermore, Siti Farhanah, Bt Jaffar and Dannawaty Mohd [9] proposed a method for diagnosing diabetes. The diagnosis is accomplished using back propagation neural network algorithm. The inputs to the system are plasma glucose concentration, blood pressure, triceps skin fold,

A. Badnjevic (ed.), *CMBEBIH 2017*,
IFMBE Proceedings 62,
DOI: 10.1007/978-981-10-4166-2_103

serum insulin, Body Mass Index (BMI), diabetes pedigree function, number of times a person was pregnant and age. The biggest challenge to this method was the missing values in the data set [9]. This system was later modified and presented by T.Jayalakshmi and Dr.A.Santhakumaran [10]. They have proposed an idea to overcome the missing values that was not addressed and this included constructing the data sets with reconstructed missing values, thereby increasing the classification accuracy [10]. It has been proven that the benefits of introducing machine learning into medical analysis are to increase diagnostic accuracy, to reduce costs and to reduce human resources. Artificial Neural Networks (ANN) is currently the next promising area of interest. Already it is successfully applied to various areas of medicine such as diagnostic systems, bio chemical analysis, and image analysis and drug development [10].

Due to extent amount of possibilites of machine learning algorithms, a novel method for classification of prediabetes and T2D using artificial neural network is presented in this paper using parameters obtained from 310 patients in Bosnia and Herzegovina.

2 Methods

2.1 Dataset for development of Artificial Neural Network for classification of prediabetes and diabetes type 2

When performing diagnosis of prediabetes and T2D, doctors use glucose test, a type of blood test, to determine the amount of glucose in the blood. There are different kinds of glucose tests such as fasting plasma glucose, oral glucose tolerance test, or HbA1C [4].

For development of Artificial Neural Network for diagnosis of prediabetes and T2D, database was formed using two parameters: fasting plasma glucose (FPG) and HbA1C blood tests. Blood tests were performed by medical professionals following standard protocols.

As it can be seen from Table I, according to the results of FPG test, patient is classified as healthy if the FPG is below 6,1 mmol/l. If the FPG is above 6,1 but below 7,0 mmol/l the patients is classified with prediabetes. However, if the results are above 7,0 mmol/l the patients is diagnosed with T2D.

Other important parameter is HbA1c. Additionally it gives an integrated index of glycaemia over the entire 120-day lifespan of the red blood cell. According to the International Diabetes Federation (IDF) all patients that have HbA1c above 6,5% are classified as diabetic. For prediabetes criteria is between 5,7% and 6,4% , while patients classified as healthy have HbA1c below 5,7%, as it can be seen from Table I [4].

Database consisted of 310 samples of patient's FPG and HbA1c obtained from several healthcare institutions in Bosnia and Herzegovina, out of which 102 were for healthy patients, 81for prediabetes and 127 were for T2D, as it can be seen in Table II.

Data set was divided on training data used for ANN development and validation data for testing of performance of designed ANN. Data division was performed using trial and error method during the process of network training.

2.2 Artificial Neural Network for classification of prediabetes and diabetes type 2

Artificial Neural Network was developed in order to per-

Table 1. Parameters for the classification of prediabetes and diabetes type 2

Criteria	T2DM	Prediabetes
FPG	≥ 7,0 mmol/l	≥ 6,1 mmol/l
HbA1c	> 6,5 %	5,7% - 6,4%

Table 2. Database used for development of Artificial Neural Network

Total number of samples	Healthy	Prediabetes	T2D
310	102	81	127

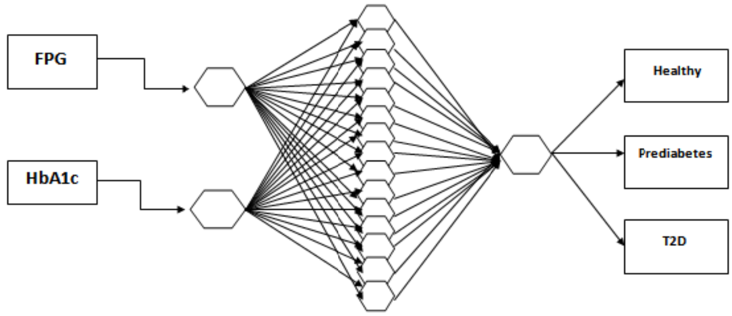

Fig. 1. Block diagram of ANN for classification of prediabetes and T2D

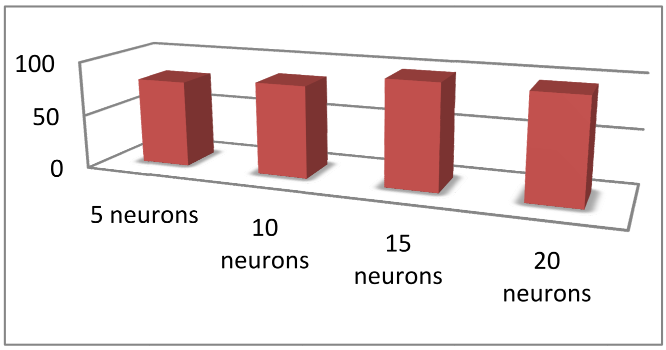

Fig. 2. Graphical display of output accuracy of networks with 5, 10, 15 and 20 neurons in hidden layer

form classification of prediabetes and T2D into three output classes: healthy patients, patients with prediabetes and patients with T2D.

The standard network that is used for pattern recognition is a two-layer feedforward network, with a sigmoid transfer function in the hidden layer, and a sigmoid transfer function in the output layer. The linear feed forward (FF) neural network is designed since according to the application experts, that kind of ANN is sufficient to properly perform the classification of input data [11].

The input layer is made of network inputs. Input layer is followed by a hidden layer which consists of parallely connected neurons. The network was tested on training performance with 5, 10, 15 and 20 neurons. Each neuron performs a weighted summation of the inputs, which is then passed to some non-linear activation function. Since, this classification is basically highly non-linear data classification, we choose sigmoid function for hidden layer.[12].

During the design of the network, different number of neurons was used in order to measure the performance. ANN was tested with 5, 10, 15 and 20 neurons. According to the Mean Square Error (MSE), appropriate number of neurons was choosen.

3 Results

According to previous research and expert recommendation discussed in previous section of this paper, for classification of prediabetes and diabetes type 2 in this paper, two-layer feedforward artificial neural network, with sigmoid function was developed. In order to determine the number of neurons in hidden layer which will yield minimum performance error, during development network was trained with 5, 10, 15 and 20 neurons in hidden layer.

3.1 Training performance of developed ANN

Performance of training is dependent on the number of neurons in hidden layer which influences network complexity. There are many methods for choosing appropriate num-

Table 2. Artificial Neural Network performance in classification to class „Prediabetes“

n=120	Classified as Prediabetes	Classified as NO Prediabetes	
ACTUAL Prediabetes	TP=27	FN=3	30
ACTUAL NO Prediabetes	FP=4	TN=86	90
	31	89	

Table 3. Artificial Neural Network performance in classification to class „Diabetes type 2“

n=120	Classified as T2D	Classified as NO T2D	
ACTUAL T2D	TP=48	FN=2	50
ACTUAL NO T2D	FP=6	TN=64	70
	54	66	

ber of neurons in hidden layer, but there is not strictly defined one. In this study trial and error method was used.

As it can be seen from Fig. 2, the ANN was trained with 5, 10, 15 and 20 neurons and the best performance during training of ANN was obtained using 15 neurons in the hidden layer.

3.2 Testing performance of developed ANN

For testing of developed ANN, data set containing 120 samples from patients from Clinical Center University of Sarajevo was used. Out of 120 samples 40 samples are from healthy patients, 30 samples are from patients diagnosed with prediabetes and 50 samples are from patients with T2D. Tables below represent validation results where Table 2 shows ANN performance for classification of prediabetic patients and Table 3 shows ANN performance for classification of patients with T2D.

According to these results, specificity, sensitivity and accuracy of the network are calculated according to equations (1), (2) (3):

$$sensitivity = \frac{true\ positive}{(true\ positive + false\ negative)} \quad (1)$$

$$specificity = \frac{true\ negative}{(true\ negative + false\ positive)} \quad (2)$$

$$accuracy = \frac{(true\ negative + true\ positive)}{(true\ negative + true\ positive + false\ negative + false\ positive)} \quad (3)$$

These results indicate that there is 90% of chance that by using this artificial neural network patient with prediabetes will be diagnosed with this condition (Table 4). When this network is used for the patients with T2D there is 96% chance that patients will be correctly diagnosed with the disease (Table 5). Percentage of specificity indicates that

Table 4. Specificity, sensitivity and accuracy of the ANN

	Sensitivity	Specificity	Accuracy
Prediabetes	90,0 %	95,5 %	94,1 %
T2D	96,0%	91,4%	93,3%
System Mean Performance	93%	93,5%	93,7%

when healthy patients are tested with this network, there is 95,5% of chance that they will be classified as negative to prediabetes, and 91,4% of chance that they will be classified as negative to T2D. Overall accuracy of the network is different for prediabetes and T2D classification. When used for classifying prediabetes network is 94,1% accurate, while for classifying diabetes type 2 network is 93,3% accurate.

Difference in the results between prediabetes and T2D classification is due to uneven number of samples in the data set. For more precise result larger data set with equal number of samples in all three classes will be needed.

In the future work, the intention of authors is to develop Graphical User Interface for this artificial neural network, and make it applicable for end users, patients and doctors.

4 Conclusion

In this paper an Artificial Neural Network for classification of prediabetes and T2D is presented. Prediabetes is precursor stage of T2D, but not obligatory. It is defined as an elevation of plasma glucose above the normal range but below that of clinical diabetes. On the other hand, diabetes is a long term metabolic disorder that is characterized by high blood sugar, insulin resistance, and relative lack of insulin. T2D primarily occurs as a result of obesity and not enough exercise. It is a major health problem in both industrial and developing countries, and its incidence is rising. Although detection of diabetes is improving, about half of the patients with T2D are undiagnosed. Because of this it is of crucial importance to improve diagnostic methods for diabetes. Artificial neural networks as a powerful tool in many fields have proved to be very useful in medical diagnostic.

In this paper ANN is trained with 190 samples from several healthcare institutions in Bosnia and Herzegovina. Samples are classified based on FPG and HbA1c, into three classes: healthy patients, prediabetes and T2D patients. FPG and HbA1c are important parameters serving as criteria for diagnosis of these diseases according to the International Diabetes Federation (IDF).

In the case of prediabetes classification sensitivity is calculated to be 90%, specificity 95,5% and accuracy 94,1%. For the classification of T2D networks sensitivity is 96%, specificity 91,4% and accuracy 93,3%. Even though these results are satisfying, they can be improved by training network with larger data set that has equal number of samples in all three classes.

ANN as a tool in diagnostic of prediabetes and T2D can be further improved to expand its application. It can be used as a background in developing a device that could help doctors in more rapid diagnostics, monitoring of patients condition and possibly giving adequate therapy.

References

1. Boyle, J. P., Thompson, T. J., Gregg, E. W., Barker, L. E., & Williamson, D. F: Projection of the year 2050 burden of diabetes in the US adult population: Dynamic modeling of incidence, mortality, and prediabetes prevalence (2010)
2. Ganguly, S; Tan, HC; Lee, PC; Tham, KW: Metabolic bariatric surgery and type 2 diabetes mellitus: an endocrinologist's perspective. Journal of biomedical research. 29 (2): 105–11 (2015)
3. Smyth, S; Heron, A.: Diabetes and obesity: the twin epidemics. Nature Medicine. 12 (1): 75–80. (2010)
4. International Diabetes Federation (2016). http://www.idf.org/WDD15-guide/facts-and-figures.html. Accessed July 2016.
5. Disse, E; et al.: Heterogeneity of pregnancy outcomes and risk of LGA neonates in Caucasian females according to IADPSG criteria for gestational diabetes mellitus. Diabetes Metab, 39 (2): 132–138 (2013)
6. National Institute of Diabetes and Digestive and Kidney Diseases (2014)
7. Tfayli, H; Arslanian, S.: Pathophysiology of type 2 diabetes mellitus in youth: the evolving chameleon. Arquivos brasileiros de endocrinologia e metabologia. 53 (2): 165–74. (2009)
8. Temurtas, H., Yumusak, N., & Temurtas, F.: A comparative study on diabetes disease diagnosis using neural networks. Expert Systems with Applications, 36(4), 8610-8615. (2009)
9. Farhanah S et al.: Diabetes mellitus forecast using artificial neural networks, Asian conference of paramedical research proceedings, Kuala Lumpur, Malaysia (2005)
10. Jayalakshmi T. and Santhakumaran A.: A novel classification method for classification of diabetes mellitus using artificial neural networks. International Conference on Data Storage and Data Engineering (2010)
11. Aljovic, A., Badnjevic, A., Gurbeta, L.: Artificial Neural Networks in the Discrimination of Alzheimer's disease Using Biomarkers Data, IEEE 5th Mediterranean Conference on Embedded Computing (MECO). Bar, Montenegro (2016)
12. Badnjevic, A., Gurbeta, L., Cifrek, M., Marjanovic, D.: Classification of Asthma Using Artificial Neural Network, IEEE 39th International convention on information and communication technology, electronics and microelectronics (MIPRO). Opatija, Croatia (2016)