ESTIMATION OF THE FACTORS ASSOCIATED WITH DIABETES MELLITUS BY MULTILAYER PERCEPTRON ARTIFICIAL NEURAL NETWORK MODEL

Seyhan Polat[1], Hakan Parlakpinar[1], Cemil Colak[2]

1 Department of Medical Pharmacology, Faculty of Medicine, Inonu University, Malatya, TURKEY
2 Department of Biostatistics and Medical Informatics, Inonu University, Malatya, TURKEY

AIM

Diabetes mellitus (DM) is a chronic and endocrine disease characterized by high levels of blood glucose (BG), which causes harmful effects on the heart, blood vessels, eyes, kidneys, nerves, etc [1]. Around the world, nearly 1.6 million people died from DM which is seventh leading cause of the death in 2016. In the past three decades, the prevalence of DM has risen dramatically, and approximately 422 million people have DM worldwide [2]. Identification of factors associated with DM is essential for the management of DM. This study aimed to estimate the factors associated with DM by multilayer perceptron (MLP) artificial neural networks (ANNs).

METHODS

Open access DM data of this study was taken from the website (https://www.kaggle.com/saurabh00007/diabetescsv#diabetes.csv) [3]. The target/output variable was the presence or absence of DM, and the input/estimator variables were selected as body mass index (BMI), BG (mg/dL), diabetes pedigree function (DPF), blood pressure (BP; mmHg), age, pregnancies, insulin (mIU/L), and skin thickness (ST; mm). The MLP ANNs were used to estimate the factors associated with DM. The performance of the model was determined by the accuracy, cross entropy error and area under receiver characteristic curve.

RESULTS

In the current study, 69.3% (532) of data were used in training and 30.7% (236) of data were used in testing procedures. The accuracy rates of DM for the designed model were calculated as 79.1% in the training dataset and 80.9% in the testing dataset. Importance values of the factors were determined as 0.244 for BMI, 0.233 for BG, 0.132 for DPF, 0.128 for BP, 0.082 for age, 0.074 for pregnancies, 0.057 for insulin and 0.050 for ST.

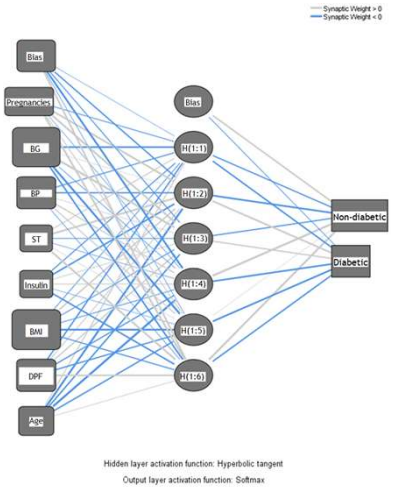


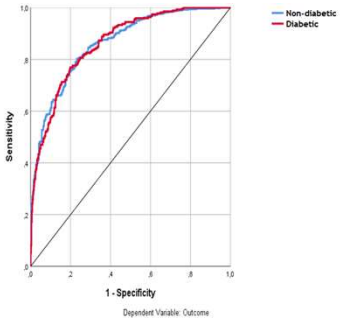
Figure 1. Artificial neural networks scheme of the factors associated with DM.

Table1. Classification of data were used in training and testing procedures, and accuracy rates of DM for the designed model.

Sample		Predicted		
		Non-diabetic	Diabetic	Percent Correct
Training	Non-diabetic	303	45	87,1%
	Diabetic	66	118	64,1%
	Overall Percent	69,4%	30,6%	79,1%
Testing	Non-diabetic	132	20	86,8%
	Diabetic	25	59	70,2%
	Overall Percent	66,5%	33,5%	80,9%

Dependent Variable: Outcome

*Correspondence: hakan.parlakpinar@inonu.edu.tr



Area Under the Curve

		Area
Outcome	Non-diabetic	0,863
	Diabetic	0,863

Figure 2. The area under the receiver operating characteristic curve for non-diabetic and diabetic individuals in the test and training groups.

Table 2. The importance values of the independent variables related with DM.

Independent Variables	Importance	Normalized Importance
Pregnancies	0,074	30,4%
BG	0,233	95,6%
BP	0,128	52,4%
ST	0,050	20,6%
Insulin	0,057	23,6%
BMI	0,244	100,0%
DPF	0,132	54,0%
Age	0,082	33,7%

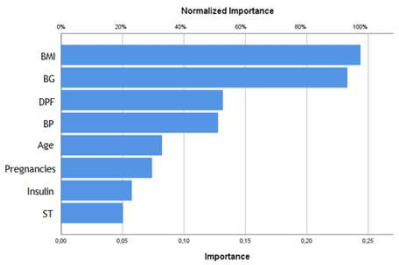


Figure 2. The scheme of the importance levels of the independent variables associated with DM.

DISCUSSION

Understanding the importance levels of the factors associated with DM is very valuable in terms of reducing the incidence of DM and preventing its complications during the follow-up of the disease. In this study, BMI, BG and DPF were the three most important factors associated with DM. In a similar study that has been conducted on a sample of 234 individuals, the variables with the highest level of importance for prediction of type II DM were waist circumference and age [4].

CONCLUSION

According to the results, the three most important factors associated with DM were BMI, BG and DPF. The results of the proposed model in this study can be used in preventive medicine and management of factors associated with DM. Further investigations with detailed study design are needed to test the our results.

References:

1.Diagnosis and classification of diabetes mellitus. Diabetes Care, 2009. 32 Suppl 1(Suppl 1): p. S62-7.
2.Diabetes. 21 November]; Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes#:~:text=Diabetes%20is%20a%20major%20cause,high%20blood%20glucose%20in%202012.
3.diabetes.csv, dataset. 21 November 2020]; Available from: https://www.kaggle.com/saurabh00007/diabetescsv#diabetes.csv.
4.Borzouei, S. and A.R. Soltanian, Application of an artificial neural network model for diagnosing type 2 diabetes mellitus and determining the relative importance of risk factors. Epidemiol Health, 2018. 40: p. e2018007.