THE UNIVERSITY of ADELAIDE

Adelaide University

https://www.adelaide.edu.au/

Using Multi-Layer Perceptron Neural Networks for Classification of Diabetes

Nguyen Diem Thanh

*Department of Computer Science, Adelaide University

Abstract: Diabetes mellitus is one of the deadliest and chronic diseases that affect persons who have an increase in their blood glucose levels. Type 1 Diabetes Mellitus *"T1DM"* is considered one of the most dangerous types of diabetes, diabetes specialists face doubts about their diagnosis of the type of diabetes in the patient where the uncertainty about the diagnosis of the disease may lead to delays in controlling the potential complications, especially if they have *T1DM*. It is generally in sleep mode, whenever the patient suffers from any disease, diabetes boosts them which is a common factor of cardiac problems. The prime motto of this work is to classify the diabetes type accurately. The learning-based algorithms play a vital role in supporting decision making in disease diagnosis and prediction. In this paper, compared traditional classification algorithms with multi-layer perceptron neural networks with the assistance of datasets which were collected in Kaggle. The ***"Pima Indians Diabetes"*** is originally from the National Institute of Diabetes and Digestive Kidney Diseases which consists of 9 predictors used to diagnose diabetes types. The dataset includes 500 instances of non-diabetes and 268 instances of diabetes types. Various performance methods with different aspects are evaluated for the Machine Learning algorithm types: K-nearest neighbor, Naïve Bayes, Gaussian NB, and multi-layer perceptron algorithms. This work shows that the multi-layer perceptron algorithm reaches the highest prediction accuracy with the lowest error-rate. Then the performance of the model proposed was evaluated using different kinds of metrics such as overall accuracy, recall, specificity, and others.

Keyworks: Diabetes Prediction, MLP, Perceptron, Classification of diabetes, T1DM

1. Introduction

Diabetes mellitus is a chronic disease and one of the 10 deadliest diseases in countries which is one of the metabolic disorders with inappropriately raised blood glucose level, and it will be released into the bloodstream. Diabetes is one of the main causes of myocardial infarction, diabetic retinopathy, diabetic bone necrosis, kidney failure, amputation, diabetic neuropathy, diabetic coma, and other complications that have not been discovered yet. It is characterized by thickening of the glomerular basement membrane, mesangial expansion, and glomerular sclerosis. These changes cause glomerular hypertension and progressive decline in glomerular filtration rate. At present, diabetes is considered to be one of the lethal diseases across the globe, and people are being affected in a huge number. Around 422 million people are diabetic patients, and about 1.6 million deaths

regarding diabetes every year. Over the past few decades, the number of cases and the prevalence of diabetes are steadily increasing [1]. Diabetes mellitus is classified as type 1 (T1D), type 2 (T2D) and gestational diabetes. The condition where the pancreas will produce little, or no insulin is type 1 diabetes. If the insulin is not absorbed by cells or not produced in enough quantity, it is referred to as type 2 diabetes (T2D). During pregnancy, a high glucose level would increase the risk of complications such as hearing loss, dementia, heart diseases, stroke, depression, vision loss, retinopathy, neuropath, and many more signs regarding diabetes diseases. According to the World Health Organization, diabetes is an epidemic disease that poses a threat to human life, with approximately 80% of diabetes deaths in low-income countries occurring in low and middle-income countries [2]. Diabetes caused 4.2 million deaths [3] but can be detected in the early stages. Also, the correct diagnosis of the type of disease plays a major role in reducing the symptoms of the disease. T1DM is called a silent killer because of its dangerous level which affects people with high blood sugar in the body. There are two different types of diabetes: T1DM and T2DM, which affect people who are unable to produce insulin or are unable to use it to convert the glucose in the blood into energy [2]. T1DM usually affects most children and adolescents who are under 30 which can be seen as common symptoms is high blood sugar and constant thirst [4], unfortunately, this type cannot be prevented and can only be treated with insulin injections. Otherwise, the other type is T2DM is the most common diabetes type and often affects adults. Usually, it is associated with high blood pressure, obesity, atherosclerosis, and other diseases [5].

Commonly, doctors and the specialized medical areas use some of the tests and factors which help them in the diagnosis according to the person's weight, blood pressure, fasting blood glucose (FBS), random blood glucose (RBS), and so on [4,6,7]. The earlier stage can be detected by machine learning mechanisms that have proven their ability to help doctors make an initial diagnosis [8,9]. To make an accurate prediction, all attributes used to predict the diabetes must have to contain the correct values without missing any values, and most suited features must be selected. To make the quality optimal features, different methodologies can be applied to the dataset to select their optimal attributes such as the genetic algorithm [10]. The techniques such as decision tree [11], support vector machines [12]. Naïve Bayes, and many more have been used for classification of the disease.

Neural Network as proposed as Multi-Layer Perceptron which is one of the popular techniques works on the concept of the biological neurons. It consists of an interconnected cluster of neurons as well as process information. It has the capability to classify the disease by identifying the different symptoms of the patient [13]. In this work, MLP is used mainly for detecting and classification of diabetes people compared with others to give the best context overview.

2. Proposed Method

The most basic and common method of Neural network is Multi-Layer Perceptron network. It uses a back propagation learning method for the classification of the instances following depicts shown as Fig 1.

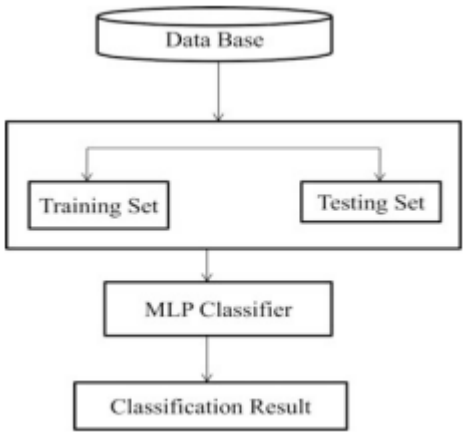

Fig 1: MLP Classification Architecture

As can be seen from the picture, the dataset is processed in the first stage for the suitability of feeding to the model, then the data are divided into the graphs as training and testing data. The purpose of this research is the creation of a model to classify and to provide predictive analysis on the diagnosis of diabetes, which allows organizations that provides service to cover health access information on the diagnosis of diabetes mellitus quickly, through using machine learning method, on this paper, using MLP techniques to classify medical datasets is to help the medical specialist and physicians process the non-linear data automatically and find the correct diagnosis [14]. On the other hand, there is a need to optimize the diagnosis of diabetes through an evaluation process of symptomatic features and risk factors using machine learning techniques. The evolutionary algorithm has been applied to predict medical diagnosis in general and diabetes in particular.

In this research, a proposed model and a set of machine learning techniques will be used to classify diabetes mellitus and its types (i.e T1DM and T2DM). A "***Pima Indians Diabetes"*** dataset has been collected to be used in this work to evaluate applied models in classifying the Diabetes types. The proposed model proves its efficacy in the classification of diabetes.

3. Dataset

The **"*Pima Indians Diabetes"*** dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The datasets consist of several medical predictor variables and one target variable (Outcome). The main objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. In particular, all patients are females at least 21 years old. The following are the attributes used to diagnose diabetes types:

- ***Age***: The age of the patient is one of the most powerful predictors in predicting diabetes types. Where the T1DM is commonly among young people while T2DM is common for older people
- ***Diabetes:*** Having first and second-degree relatives with diabetes have the effect of transmitting the disease from generation to generation.
- ***Pregnancies:*** Number of pregnancies. The number of pregnancies affects the determination of the type of sugar because pregnant women are more likely than others to develop diabetes.
- ***BMI:*** The body fat index according to the height and the weight of patients. BMI has a role in the classification of the type of diabetes. Eq (1) used to calculate the BMI values [15]:

$$\boldsymbol{BMI = weight\ (kg) \div height\ (m^2)}$$

4. Methodology

Multi-Layer Perceptron (MLP) is a class of Neural Network algorithms. It consists of one input layer, one output layer, and one or more hidden layers of nodes that are mapped nonlinearly. The nodes in each layer are connected by weight to all nodes in the next layer. The Perceptron Multi-Layer architecture depends on the choice of the number of layers, the number of nodes in the hidden layer, and the objective functions used [16]. The number of neurons in the input layer is determined based on the number of input variables, while the number of neurons in the output layer is based on the number of output classes [17]. In this study, 16 neurons were used in the input layer and two neurons in the output layer. Meanwhile, the number of hidden layers and the number of neurons in the hidden layer are determined through the optimization process. In this study,

MLP is used to predict the risk of early-stage diabetes. When carrying out the learning process. MLP uses five main parameters, consisting of learning rate, momentum, number of epochs, threshold for errors, and number of neurons in hidden layer. The quality of the solutions offered by each model can be obtained using fitness function. In this case, the fitness function is calculated using accuracy. To calculate accuracy

$$Accuracy = \frac{The\ number\ of\ sample\ correct}{The\ number\ of\ all\ sample}$$

To evaluate, several methods were used to measurement the effective of the model such as precision, recall, f1-score, and many more.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

Recall (Sensitivity) is the ratio of correctly predicted positive observations to all observations in actual class.

F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

Where True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN)

5. Results and discussion

In this article, several machine learning methods compare and discuss the results of the experiments using the applied models.

Table 1. The results of applying different machine learning models

Approach	Performance Metrics		
	Accuracy (%)	Sensitivity (%)	F1 Score (%)
MLP	86.36%	86.24%	86.2%
Gaussian-NB	74.68%	74.68%	74.5%
Perceptron	72.08%	72.08%	72.68%
K-NN	76.62%	76.62%	74.26%

As shown in Table 1, the MLP model outperforms the other known machine learning classification algorithms in classification accuracy, sensitivity, and f1 score. The results show that based on various features used in the literature, such as the number of pregnancy rates, diastolic blood pressure, glucose, age, etc.

Fig 2: Performance metrics of each model

Fig.2 show the ability of MLP model in classifying diabetes up to above 80%.

6. Conclusion

The test result show that the use of Multi-Layer Perceptron with optimal parameters can be used to predict the risk of early diabetes with high accuracy. The proposed algorithm is proven to provide the best accuracy compared to others such as K-nearest neighbor, Naïve Bayes, Gaussian NB. It is known that the accuracy provided by Multi-Layer Perceptron gives an average accuracy of 86% which is the highest average score compares to another model.

The hidden layers on MLP perform computations on input and use activation function for generating output. However, learning with more layers will be easier but more training time is required. The weights of each hidden layer also have a very important role to the output produced by MLP. Weights are used to scale input separately. Given the results obtained, the MLP classifier to identify diabetic people is proposed as a useful tool to help the diabetes specialist in the early detection of the disease and to confirm their diagnosis which will help people with a treat as soon as possible. In the future, providing a tool to detect disease based on existing valuable parameters will help people aware of the illness and give preventive treatments automatically. Also, developing a prediction and classification system with real-time analysis remote monitoring applications, measure the effective cost of implementing them, and the impact they can have on people's lives.

References

[1] Mokdad, A. H., Ford, E. S., Bowman, B. A., Nelson, D. E., Engelgau, M. M., Vinicor, F., & Marks, J. S. (2000). Diabetes trends in the US: 1990-1998. Diabetes care, 23(9), 1278-1283.

[2] D. Mauricio, N. Alonso, and M. Gratacòs, "Chronic Diabetes Complications: The Need to Move beyond Classical Concepts", Trends in Endocrinology & Metabolism Journal, Vo1. 31, No. 4, pp. 287-295, 2020.

[3] A. Wako, S. Belay, Y. Feleke, and T. Kebede, "Assessment of risk for severe hypoglycemia among adults with IDDM: validation of the low blood glucose index", Journal of Diabetes and Metabolism, Vol. 21, No. 11, 2017

[4] Bruzda-Zwiech, J. Szczepańska, and R. Zwiech, "Xerostomia, thirst, sodium gradient and interdialytic weight gain in hemodialysis diabetic vs. non-diabetic patients", Medicina Oral Patología Oral y Cirugia Bucal, Vol. 23, No. 4, pp. 406- 412, 2018.

[5] M. A. Said, N. Verweij, and P. van der Harst, "Associations of Combined Genetic and Lifestyle Risks with Incident Cardiovascular Disease and Diabetes in the UK Biobank Study", JAMA Cardiology, Vol. 3, No. 8, pp. 693-702, 2018.

[6] E. Barry, S. Roberts, J. Oke, S. Vijayaraghavan., R. Normansell, and T. Greenhalgh, "Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: systematic review and meta-analysis of screening tests and interventions", BMJ, 2017.

[7] M. J. L. Verhulst, B. G. Loos, V. E. A. Gerdes, and W. J. Teeuw, "Evaluating All Potential Oral Complications of Diabetes Mellitus", Front Endocrinol (Lausanne), Vol. 10, No. 56, p. 56, 2019.

[8] Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal, Vol. 15, pp. 104-116, 2017.

[9] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project", Plos One, Vol. 12, No. 7, 2017.

[10] D. K. Choubey, S. Paul, and S. Kumar, "Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection", In: Proc. of International Conf. on Communication and Computing System, London, pp. 451–455, 2017

[11] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance analysis of data mining classification techniques to predict diabetes", Procedia Computer Science, vo. l82, pp. 115–121, 2016.

[12] N. Barakat, A. P. Bradley, and M. N. H. Barakat, "Intelligible support vector machines for diagnosis of diabetes mellitus", IEEE transactions on information technology in biomedicine, vol. 14, pp. 1114–1120, 2010.

[13] L. Sarangi, M. N. Mohanty, and S. Pattanayak, "Design of MLP Based Model for Analysis of Patient Suffering from Influenza", Procedia Computer Science, Vol. 92, pp. 396–403, 2016.

[14] Saru, S., and S. Subashree. "Analysis and prediction of diabetes using machine learning", International Journal of Emerging Technology and Innovative Engineering 5.4 2019

[15] M. Ji, S. Zhang, and R. An, "Effectiveness of A Body Shape Index (ABSI) in predicting chronic diseases and mortality: a systematic review and meta-analysis", Obesity Reviews, Vol. 19, No. 5, pp. 737–759, 2018.

[16] Ramchoun H, Idrissi M A J, Ghanou Y and Ettaouil M 2017 New Modeling of Multilayer Perceptron Architecture Optimization with Regularization: An Application to Pattern Classification IAENG Int. J. Comput. Sci. 44 261–9

[17] Taşkıran M, Çam Z G and Kahraman N 2015 An Efficient Method to Optimize Multi-Layer Perceptron for Classification of Human Activities Int. J. Comput. Commun. Instrum. Eng. 2 191–5