*Diabetes Prediction Using Machine Learning Algorithms*

Bishad Deb Roy Utsha  
*Department of Computer Science and Engineering*  
*East West University*Dhaka, Bangladesh  
2016-3-60-017@std.ewubd.edu

Md. Mahfuzul Islam Mazumder  
*Department of Computer Science and Engineering*  
*East West University*Dhaka, Bangladesh  
2016-3-60-048@std.ewubd.edu

Pranto Biswas  
*Department of Computer Science and Engineering*  
*East West University*Dhaka, Bangladesh  
2017-1-60-007@std.ewubd.edu

*Abstract*—

Keywords—

# Introduction

Diabetes is a long-lasting disease that happens when the pancreas fails to create enough insulin, or when the body cannot use the insulin produced efficiently. Insulin is a hormone that controls the level of sugar in the blood. Hyperglycemia or hyperglycemia is a common result of uncontrolled diabetes and, over time, causes severe damage to many organs, particularly nerves and blood vessels.

With the development of living standards, diabetes is increasingly common in people’s daily life. Therefore, how to quickly and accurately diagnose and analyze diabetes is a topic worthy studying. In medicine, the diagnosis of diabetes is according to fasting blood glucose, glucose tolerance, and random blood glucose levels [2,3]. The earlier diagnosis is obtained, the much easier we can control it. Machine learning can help people make a preliminary judgment about diabetes mellitus according to their daily physical examination data, and it can serve as a reference for doctors [4]. For machine learning method, how to select the valid features and the correct classifier are the most important problems.

Recently, numerous algorithms are used to predict diabetes, including the traditional machine learning method [1], such as support vector machine (SVM), decision tree (DT), logistic regression and so on.

Machine learning methods are widely used in predicting diabetes, and they get preferable results. Logistic regression and Decision tree are one of popular machine learning methods in medical field, which has grateful classification power. Neural network is a recently popular machine learning method, which has a better performance in many aspects. So, in this study, we used decision tree, random forest (RF) and neural network to predict the diabetes.

# Related work

Medical applications of data mining include prediction of the effectiveness of surgical procedures, medical tests and medications, and discovery of relationships among clinical and pathological data [5]. Moustakas and Charisse’s’ work [7] surveyed the role of machine learning in medical decision making and provided an extensive literature review on various ML applications in medicine that could be useful to practitioners interested in applying ML methods to improve the efficiency and quality of decision-making systems in medical applications.

Apart from the works mentioned above, a lot of research has been done specifically using ANN in diagnosing diabetes mellitus and some approaches are discussed below.

Siti Farhanah, Bt Jaffar and Dannawaty Mohd [6] 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, 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. This system was later modified and presented by T.Jayalakshmi and Dr.A.Santhakumaran[8].They have proposed an idea to overcome the missing values that was not addressed by Siti Farhanah Bt Jaafar [6] and this included constructing the data sets with reconstructed missing values, thereby increasing the classification accuracy[8]. They have also proposed an alternate method to overcome missing value by performing data pre-processing, which also speeds up the training process by reducing the actual learning time. Various missing value techniques and pre-processing methods were analyzed. By adopting these modifications, the results improved and achieved a classification accuracy of 99% [6].

**Machine learning Methods:**

Data Requirements:

**Dataset**: We have required data set of Diabetes for the simulations. Its data collected from the Kaggle. In this data set we have 768 row and 9 columns.

* **Pregnancies:**Number of times pregnant
* **Glucose:** Plasma glucose concentration over 2 hours in an oral glucose tolerance test
* **Blood Pressure:**Diastolic blood pressure (mm Hg)
* **Skin Thickness:** Triceps skin fold thickness (mm)
* **Insulin:** 2-Hour serum insulin (mu U/ml)
* **BMI:** Body mass index (weight in kg/(height in m)2)
* **DiabetesPedigreeFunction:** Diabetes pedigree function (a function which scores likelihood of diabetes based on family history)
* **Age:** Age (years)
* **Outcome:** Class variable (0 if non-diabetic, 1 if diabetic)

REFERENCES

[1].Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., and Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Comput. Struct. Biotechnol. J. 15, 104–116. doi: 10.1016/j.csbj.2016.12.005

[2].American Diabetes Association (2012). Diagnosis and classification of diabetes mellitus. Diabetes Care 35(Suppl. 1), S64–S71. doi: 10.2337/dc12-s064

[3].Cox, M. E., and Edelman, D. (2009). Tests for screening and diagnosis of type 2 diabetes. Clin. Diabetes 27, 132–138. doi: 10.2337/diaclin.27.4.132

[4].Lee, B. J., and Kim, J. Y. (2016). Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning. IEEE J. Biomed. Health Inform. 20, 39–46. doi: 10.1109/JBHI.2015.2396520

[5] Ranjit Abraham, Jay B.Simha, Iyengar S. “Medical datamining with a new algorithm for feature selection and Naïve Bayesian classifier”, 10th International Conference on Information Technology.

[7] Moustakis, V. and Charissis, G. (1999). Machine learning and medical decision making, In Proceedings of Workshop on Machine Learning in Medical Applications, Advance Course in Artificial Intelligence- ACAI99, Chania, Greece, 1-19.

[6] Siti Farhanah Bt Jaafar and Dannawaty Mohd Ali, “Diabetes mellitus forecast using artificial neural networks”, Asian conference of paramedical research proceedings, 5-7, September, 2005, Kuala Lumpur, MALAYSIA

[8] T.Jayalakshmi and Dr.A.Santhakumaran, “A novel classification method for classification of diabetes mellitus using artificial neural networks”. 2010 International Conference on Data Storage and Data Engineering.