Prediction of Diabetes with Deep Neural Network

Yaser Issam Hamodi
Department of Computer Engineering, Ministry of Higher Education & Scientific Research Iraq, Baghdad, Iraq. E-mail: yaserissamtc@gmail.com

Received April 26, 2021; Accepted August 23, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18I2/WEB18355

Abstract

Diabetic mellitus is hitting the globe since decades. it often leads to dangerous health issues such as kidney problems, heart strokes, nervous system disturbance and eye problems etc. The prediction and detection of such a deadly disease is pivotal. In the conducted study, we have built a diabetic prediction model which is based on deep neural networks. We performed our experiments with two-fold and four-fold cross validation. Our diabetic prediction model has reported an accuracy of 98.45% which is quite high.

Keywords

Diabetic, Machine Learning, Deep Neural Network, Prediction, Health Monitoring.

Introduction

Diabetes is considered one of the severe and common ailments in all over the world. It leads to high blood glucose and blood sugar levels. Three types of diabetes are centre of study; Type-1, Type-2 and the third one is the Gestational diabetes. In Type-1 diabetes, enough insulin is not produced by the human body. Whereas in type-2, insulin is not created and utilized by the human body (Varma, R., 2014). More people belong to type-2 diabetes as more than 90 t0 95% people have been affected by it worldwide. However, the Gestational diabetes befell during pregnancy.

Diabetic patients' strength is increasing rigorously day by day. Diabetes often lead to dangerous health issues such as kidney problems, heart strokes, nervous system disturbance and eye problems etc. conferring to a survey which was conducted by International Diabetes Federation (IDF) in 2019, almost 463 million people across the globe are affected with diabetes (Singh, A., 2018; Alfian, G., 2020). Moreover, it was also predicted by the researchers that this figure could be inflamed to 642 million. In this evasion, prediction of diabetes in early stage is pivotal which may reduce death rate. For medical diagnosis, in order to predict the diabetes, samples from patients are collected such as triceps skinfold, serum insulin plasma glucose concentration, body mass and systolic blood stress etc. (Goswami, S.K., 2014; Bodapati, J.D., 2021). Massive datasets are being produced. For Processing such large datasets, computer-based approaches are preferred. In various recent scientific papers, reports and studies, deep learning and machine learning approaches have acquired substantial accuracy over existing techniques (Dwivedi, 2018). Deep neural networks learning variants are growing rapidly (Ayon, S.I., 2019). It works similar to human mind (Geirhos, R., 2002). In deep neural networks data is represented in multi levels and they handle selective-invariance dilemma proficiently (LeCun, Y., 2015; Zhou, H., 2020; Alhayani, B., 2020). This study is focusing on the future patients of diabetes or people on high risk of diabetic patients (Alhayani, B., 2020; Alhayani, B., 2021; Hasan H.S., 2021). For the conduct of this study, we have chosen deep neural network. This will predict the high-risk future individuals (Kwekha-Rashid, 2021; Yahya, W., 2021; B. Alhayani, 2021; B. Alhayani, 2017; B. Al Hayani, 2020).

For diabetic prediction, we have used deep learning network. At first, a dataset of significant size was used for training the neural network (Mahajan, H.B., 2021; Mahajan, H.B., 2018; Mahajan, H.B., 2019; Mahajan, H.B., 2020). After training our diabetic prediction model, we tested the model as well. Test set provided accurate result in almost all cases. The result accuracy of our prediction model is reported much better which is discussed in detail in result section. The remaining paper is organized as follow. Section II presents our research methodology. In section III, shows the performance check of the purposed system. Results are discussed in the next section which is section IV. Whereas, the last section, section V concludes the research by discussing findings.

Methodology

This section contains all the details about algorithm. To predict diabetes by using deep neural network technique. Following steps are used to conduct this study: (A) Data set collection, (B) pre-processing step (C) Implementing deep neural network and (D) Performance of the system (E) Experiment and results. We have explained the algorithm which we have used in our experiment.

Data Set

The data set operated were picked up from the health science library of Koç University Hospital - Medical Center Turkey. This dataset contains multiple attributes. We have utilized its 14 attributes. The dataset were again separated into two phases, the first phase of data set contains 7 attributes named as; Age, Sex, gamma-glutamyl transferase (GGT) (Kwekha-Rashid, A.S., 2021; Yahya, W., 2021; Alhayani, B., 2021; Alhayani, B., 2017).

Schistosomiasis (Shisto), Alkaline phosphatase (ALP), An aspartate aminotransferase (AST), and Alanine Aminotransferase (ALT). The second phase contains another 7 attributes. These attributes are named as; Height, Weight, Waist Circumference (WC), History of Diabetes, History of Hypertension, Fasting Blood Sugar (FBS) and Hemoglobin A1C (HBA1C).

Pre-processing of the Data

First, we need to clean and filter the data set for better results. After data filtration then we will estimate the values missing in the data and will normalize the data after dealing with the imbalanced data. Then doing a last verification step to check data is ready and accurate for use.

Preprocessing steps start with the reading of dataset file. Preprocessing steps will be as follows:

i. Removing unwanted and noise data
ii. Estimating the missing values from the data: linear interpolation estimates for one-d are estimated by using seven column vise and seven row wise estimates.
iii. Dealing with imbalanced data: K-fold technique is used to handle imbalanced data. Data is randomly sorted by using cross validation of k-fold. The data is divided into k-folds where K=6. Where 2 folds are used for validation purpose and 4 folds are kept for the training process. When K1 and K2 are used for training then all other folds are used for validation. In iteration when K3 and K4 are used for training then K1, K2, K5 and K6 are used for validation purpose. With next round of iteration 2 folds are used for training and 4 are used for validation purpose. In every round k-folds are switched but 2:4 is kept. More accurate and proofread data will result in improved results.

Implementation of Deep Neural Network

Our purposed methodology depends on the 4 hidden layers of neural network. Neural networks have total of 6 layers; first layer is the input layer then 4 hidden layers to predict diabetes and an output layer on the end. Hidden layers have multiple levels of neurons. Numbers of neurons are 20, 14, 10, and 5 respectively. For the prediction of diabetes, we have tried and tested various hidden layers and various numbers of neurons. Results also vary if we change the number of layers in the model or if we change number of neurons on any layer. Every layer and neuron affect the result of the system. Best results are attained when we have used 4 hidden layers of neural network and our numbers of neurons on each layer (respectively 1 to 4) are 20, 14, 10, and 5. Structure of the developed system with respect to used layers and neurons is shown in Fig 1 below.

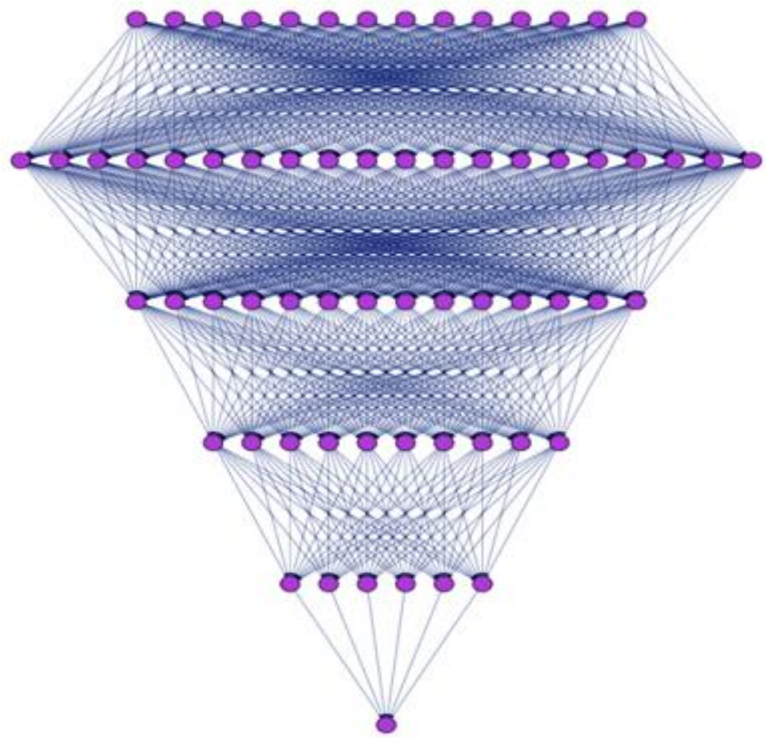

Figure 1 Proposed layered architecture

There are 14 inputs on the input layers, these inputs are the attributes of the data set used. And Produced output is 1. This output tells us whether the individual is diabetic, Pre diabetic or no risk of diabetes. According to the typical working of neural network, this system also works as weighted sum of neurons is calculated according to the one of its entered input. Then bias is added into it and it is decided whether this input must be kept or fired.

Performance of the Purposed System

Performance of this supervised learning algorithm model is visualized by the usage of confusion matrix. Some terms which are important to know the confusion matrix are as follows:

TP = True Positive (system identified the case correctly e.g., system correctly identified that this person is diabetic).
TN = True Negative (system identified the case incorrectly).

FP = False Positive (system rejected the case correctly e.g. system correctly identified this person have no risk of diabetes and rejects that individual).
FN = False Negative (system rejected the case incorrectly).

Performance of the purposed system is measured and calculated by the usage of confusion matrix. It is easy to calculate the performance of the purpose system with the help of this matrix. Following things are calculated with the help of the confusion matrix: Accuracy of the system, Sensitivity of the system and F1 score.

Accuracy of the system is calculated as:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

Sensitivity of the system is calculated as:

$$\text{Sensitivity} = \frac{\text{TP}}{(TP+FN)} \tag{2}$$

F1 score of the system is calculated as:

$$\text{F1 score} = \frac{\text{2TP}}{(2TP+FP+FN)} \tag{3}$$

Results and Discussion

For prediction of diabetes as we divided our dataset into two and four folds. Where two folds are used for validation purpose and four folds are kept for the training process. Results reported were quite satisfactory for all experiments. For processing purpose, we used Intel i7 system with 8GB RAM. An open-source software named Scikit library in python has been used. Moreover, we also integrated Spyder in our experiments. We figured out confusion matrix results using neural networks for our each of our cross validation. Table 1 shows the confusion matrix for two-fold whereas Table 2 presents for four-fold; cross validation.

Table 1 Confusion matrix for two-fold cross validation.

		Predicted Class		Actual Total
		Absence	Present	
Actual class	Absent	**649**	**51**	700
	Present	**6**	**154**	160
Total Predicted		655	205	860

Table 2 Confusion matrix for four-fold cross validation

		Predicted Class		Actual Total
		Absence	Present	
Actual class	Absent	**661**	**39**	700
	Present	**23**	**137**	160
Total Predicted		684	176	860

For our prediction model we considered evaluation matrices of four factors: accuracy (%), sensitivity (%), specificity (%) and F1 Score. Following Table 3 shows the evaluation matrices for two and four folds of our neural prediction model.

Table 3 Evaluation matrices of the diabetes prediction system

Evaluation Metrices	Two-fold	Four-fold
Accuracy (%)	97.05	98.45
Specificity (%)	97.03	98.01
Sensitivity (%)	96.45	98.09
F1 score	0.97	0.99

Figure 2 and Figure 3 graphically shows Table 3 results. It is evident that Four-fold cross validation performed best with 98.45% accuracy whereas for two-fold it was 97.05%. However, accuracy for two-fold was quite good. By considering remaining evaluation metrices specificity, sensitivity and f1-score for four-fold were also high instead of two-fold. F1-score for five-fold is 0.99 on the contrary it is 0.97 for two-fold. Similarly, specificity and sensitivity for five-fold are 98.01% and 98.09% respectively whereas for two-fold there are 97.03 and 96.45% respectively. Conclusively, evaluation metrices for four-fold are high than two-fold.

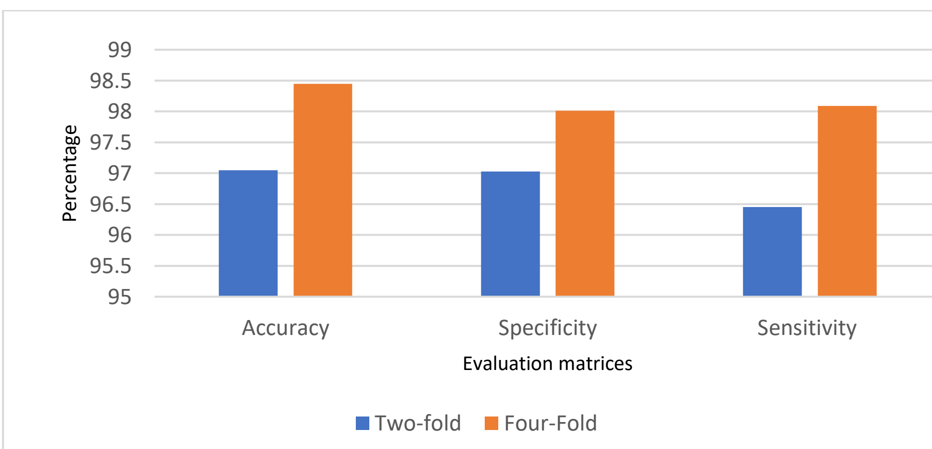

Figure 2 Evaluation matrices of the diabetes prediction system including accuracy, specificity, and sensitivity

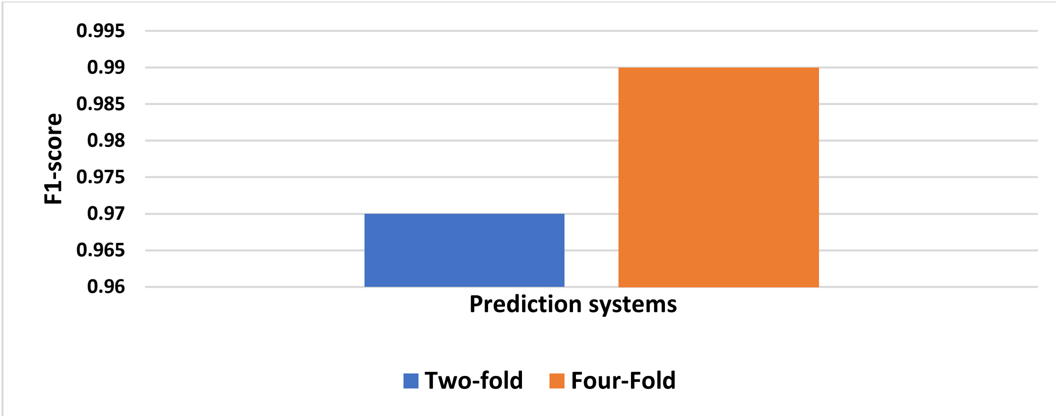

Figure 3 Evaluation matrices of the diabetes prediction system presenting F1-score

Conclusion

Diabetic is a severe ailment therefore proper measures should be taken for it. It also increases mortality rate per year worldwide. Therefore, the prediction and detection of such a deadly disease should be undertaken. In this study, we have built a diabetic prediction model based on deep neural networks. We performed our experiments with two-fold and four-fold cross validation. Our diabetic prediction model has reported an accuracy of 98.45% which is quite high. In future, our proposed model will be helpful for general human beings and specifically for medical staffs.

References

Varma, R., Bressler, N.M., Doan, Q.V., Gleeson, M., Danese, M., Bower, J.K., & Turpcu, A. (2014). Prevalence of and risk factors for diabetic macular edema in the United States. *JAMA ophthalmology*, *132*(11), 1334-1340.

Singh, A., & Lakshmiganthan, R. (2018). Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms.

Alfian, G., Syafrudin, M., Fitriyani, N.L., Anshari, M., Stasa, P., Svub, J., & Rhee, J. (2020). Deep neural network for predicting diabetic retinopathy from risk factors. *Mathematics*, *8*(9).

Goswami, S.K., Vishwanath, M., Gangadarappa, S.K., Razdan, R., & Inamdar, M.N. (2014). Efficacy of ellagic acid and sildenafil in diabetes-induced sexual dysfunction. *Pharmacognosy magazine*, *10*(Suppl 3).

Bodapati, J.D., Shaik, N.S., & Naralasetti, V. (2021). Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. *Journal of Ambient Intelligence and Humanized Computing*, 1-15.

Dwivedi, A.K. (2018). Analysis of computational intelligence techniques for diabetes mellitus prediction. *Neural Computing and Applications*, *30*(12), 3837-3845.

Ayon, S.I., & Islam, M. (2019). Diabetes Prediction: A Deep Learning Approach. International *Journal of Information Engineering & Electronic Business*, *11*(2), 21-27.

Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F.A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence, 2*(11), 665-673.

Le Cun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Zhou, H., Myrzashova, R., & Zheng, R. (2020). Diabetes prediction model based on an enhanced deep neural network. *EURASIP Journal on Wireless Communications and Networking, 2020*(1), 1-13.

Alhayani, B., & Abdallah, A.A. (2020). Manufacturing intelligent Corvus corone module for a secured two way image transmission under WSN. *Engineering Computations.*
https://doi.org/10.1108/EC-02-2020-0107

Alhayani, B.S., & Llhan, H. (2021). Visual sensor intelligent module based image transmission in industrial manufacturing for monitoring and manipulation problems. *Journal of Intelligent Manufacturing, 32,* 597-610. https://doi.org/10.1007/s10845-020-01590-1

Al-Hayani, B., & Ilhan, H. (2020). Efficient cooperative image transmission in one-way multi-hop sensor network. *The International Journal of Electrical Engineering & Education, 57*(4), 321-339.

Alhayani, B., Abbas, S.T., Mohammed, H.J., & Mahajan, H.B. (2021). Intelligent secured two-way image transmission using corvus corone module over WSN. *Wireless Personal Communications,* 1-36. https://doi.org/10.1007/s11277-021-08484-2

Hasan, H.S., Abdallah, A.A., Khan, I., Alosman, H.S., Kolemen, A., & Alhayani, B. (2021). Novel unilateral dental expander appliance (udex): a compound innovative materials. *Computers, Materials and Continua, 68*(3), 3499-3511.
https://doi:10.32604/cmc.2021.015968

Kwekha-Rashid, A.S., Abduljabbar, H.N., & Alhayani, B. (2021). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience,* 1-13. https://doi.org/10.1007/s13204-021-01868-7

Yahya, W., Ziming, K., Juan, W., Al-Nehari, M., Tengyu, L., Qichao, R., & Alhayani, B. (2021). Study the influence of using guide vanes blades on the performance of cross-flow wind turbine. *Applied Nanoscience,* 1-10. https://doi.org/10.1007/s13204-021-01918-0

Alhayani, B., Mohammed, H J., Chaloob, I.Z., & Ahmed, J.S. (2021). Effectiveness of artificial intelligence techniques against cyber security risks apply of IT industry. *Materials Today: Proceedings.*

Alhayani, B., Abbas, S.T., Khutar, D.Z., & Mohammed, H.J. (2021). Best ways computation intelligent of face cyber attacks. *Materials Today: Proceedings.*

Alhayani, B.., & Ilhan, H. (2017). Hyper spectral image classification using dimensionality reduction techniques. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 5*(4), 71–74.

Al Hayani, B., & Ilhan, H. (2020). Image transmission over decode and forward based cooperative wireless multimedia sensor networks for Rayleigh fading channels in medical internet of things (MIoT) for remote health-care and health communication monitoring. *Journal of Medical Imaging and Health Informatics, 10*(1), 160-168.

Mahajan, H.B., Badarla, A., & Junnarkar, A.A. (2021). CL-IoT: cross-layer Internet of Things protocol for intelligent manufacturing of smart farming. *Journal of Ambient Intelligence and Humanized Computing, 12*(7), 7777-7791. https://doi.org/10.1007/s12652-020-02502-0

Mahajan, H.B., & Badarla, A. (2018). Application of internet of things for smart precision farming: solutions and challenges. *International Journal of Advanced Science and Technology,* 37-45.

Mahajan, H.B., & Badarla, A. (2019). Experimental analysis of recent clustering algorithms for wireless sensor network: Application of IoT based smart precision farming. *Journal of Advance Research in Dynamical & Control Systems, 11*(9), 10-5373. http://doi.org/10.5373/JARDCS/V11I9/20193162

Mahajan, H.B., & Badarla, A. (2020). Detecting HTTP vulnerabilities in IoT-based precision farming connected with cloud environment using artificial intelligence. *International Journal of Advanced Science and Technology, 29*(3), 214-226.

Aruna, S., & Abinaya, P. (2020). A Web service recommendation system based on QoS attribute based collaborative filtering. *Webology*, *17*(1), 246-254.