2nd International Conference on Scientific and Academic Research
March 14-16, 2023 : Konya, Turkey
All Sciences Proceedings
http://as-proceeding.com/

Design of a deep neural network for diabetes prediction

FERAS KHALEL [1*], NEHAD T.A RAMAHA[2]

[1] Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey - firaskhalilkhalil@gmail.com
[2] Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey - nehadramaha@karabuk.edu.tr

[*] firaskhalilkhalil@gmail.com Email of the corresponding author

Abstract – Diabetes is a chronic disease with many complications that follow the disease and is one of the causes of death worldwide. The number of people infected with this disease is increasing every day. Therefore, predicting this disease at an early stage helps to avoid many complications that follow the disease. Therefore, many medical sectors have begun to take an interest in using artificial intelligence technologies and benefiting from their services. Data mining and machine learning techniques are used to predict the patient's condition at an early stage. Therefore, this paper uses a neural network containing more than one hidden layer for disease prediction. The designed network gave an accuracy of 95.40%. The accuracy of the Recall scale for infected patients reached 96.59%. It is better than the result of previous studies mentioned in this paper.

Keywords – Diabetes, Pima Dataset, Neural Network, Deep Learning, Machine Learning

I. INTRODUCTION

Diabetes is one of the dangerous diseases and is a source of fear and anxiety for most people all over the world. As it is one of the highest causes of death worldwide along with some other diseases such as cardiovascular diseases, cancer and respiratory diseases[1]. The number of people with diabetes is increasing rapidly in the world. According to the International Diabetes Federation, the number of people with diabetes in 2045 may reach about 700 million people[2]. The World Health Organization says the number of people with diabetes worldwide will reach 380 million in 2025[3].

Diabetes mellitus is a group of symptoms and disorders that occur to the patient, where the blood sugar level of the affected person rises for a long period of time, because the islets of Langerhans do not produce a hormone that regulates glucose[4].

There are three types of diabetes, which are type 1 diabetes, gestational diabetes, and type 2 diabetes[5]. Gestational diabetes is the most prevalent type among the three types. Lack of exercise and weight gain are among the main causes of gestational diabetes[6]. The inability of the pancreas to produce an adequate amount of a hormone that regulates glucose is the main reason for the emergence of type 1 diabetes[7]. Leaving the disease untreated leads to several complications such as visual impairment, foot ulcers, dehydration and encephalitis, and may eventually lead to death[8][9].

And after the science of artificial intelligence and the science of data mining has proven its worth, strength and effectiveness as one of the most important solutions for analyzing large amounts of data. Where the primary goal of data mining is to extract knowledge from huge amounts of data[10]. The healthcare sector has taken advantage of this science and relied on it to help make decisions. With the increasing number of diseases and patients, and thus the number of electronic health records. It has become easy for those interested and researchers in the medical fields to obtain patient data and conduct research studies to discover new information and

knowledge and build predictive models that help in early diagnosis of diseases. Diabetes is one of these diseases, as there is no cure for this disease, and early prediction helps to avoid many complications. This paper presents a method for predicting diabetes by designing a neural network that contains more than one hidden layer. This method focuses on detecting the largest possible number of affected patients. It will also depend on the Pima data set, where it will be analyzed and some of the problems it contains will be presented and addressed. In the rest of the sections of this paper, works related to this research will be presented and compared to this research, and some problems and solutions that will be proposed to solve these problems will be presented. Then the methodology that will be followed in this paper is presented. Then the results will be discussed and analyzed. Finally, the conclusion, where we will briefly present what we have presented in this paper and the results and benefits of this study.

II. RELATED WORK

PREVIOUS STUDIES USED MACHINE LEARNING ALGORITHMS

Suggested by Nagaraj et al Diabetes prediction method and recommendation system to build a machine learning model for prediction, using several algorithms such as XGBoost, support vector machine (SVM), random forest classifier and decision tree. The random forest classifier gives an accuracy of 77%. After predicting the disease, make a recommendation system, include a diet and foods that help to recover from diabetes, and provide some activities to get rid of the disease[11].

Shanjida khan Maliha et al used two diabetes prediction algorithms, the Random Forest and the Support Vector Machine. And it relied on real patient data to determine the disease. It was implemented by the programming language Python and Jupyter. The Support Vector Machine algorithm gave a higher accuracy than Random Forest, where the accuracy reached 86%, while the accuracy of Random Forest reached 78% [3].

Herminiño C. Lagunzad et al relied on a database from Kaggle and applied the ID3 algorithm to this data to predict diabetes. As for the results, it can be for people between 30-40 and 41-50 with diabetes. Delayed recovery and sudden weight loss can also be a sign of diabetes[12] .

PREVIOUS STUDIES USED THE HYBRID MODEL

Biswajit Giri et al. used a hybrid approach to predict diabetes based on the PIMA database. He compared this method to several other algorithms. The proposed method gave an accuracy of 86 percent, while the closest algorithm was 75 percent accurate, which is the Linear Support Vector Classifier algorithm [13].

Chinedu I. Ossai et al proposed a hybrid technique to identify patients with diabetes who are susceptible to URA by relying on MLR, ANOVA and machine learning using the RF algorithm. Those at risk were identified with the following accuracy: - recall: -0.947 ± 0.035, precision: -0.951 ± 0.033 , F1-score: -0.947 ± 0.035, and AUC: -0.994 ± 0.007[14].

THE PIMA DATASET AND THE ALGORITHMS THAT APPLY IT.

Table 1 presents some studies on the Pima data set.

Ref	Algorithm	Accurracy	Recall	Precision
[11]	Random Forest	77%		
[15]	LightGBM + KNN	90.1%	82.1 %	88.9
[16]	Firefly Optimized Neural Network	95.07	88%	88%
[17]	Random Forest	79%	77%	77%
[18]	KNN+SVM+DT	90.62	91%	91
[19]	Naïve Bayes	89.9	84.3	79.3
[20]	Gradient Boosting	92	93	94
[21]	K-Means	86		
[13]	Hybrid classifier	86		
[22]	Genetic Programming Symbolic Regression	79.19%		
[23]	Artificial Neural Network	82		
[24]	Decision Tree	70.80	61.46	76.5
[25]	J48	95.122		
[26]	SVM	79.1	79.1	78.2
[27]	ANN	80.86		

III. METHODOLOGY

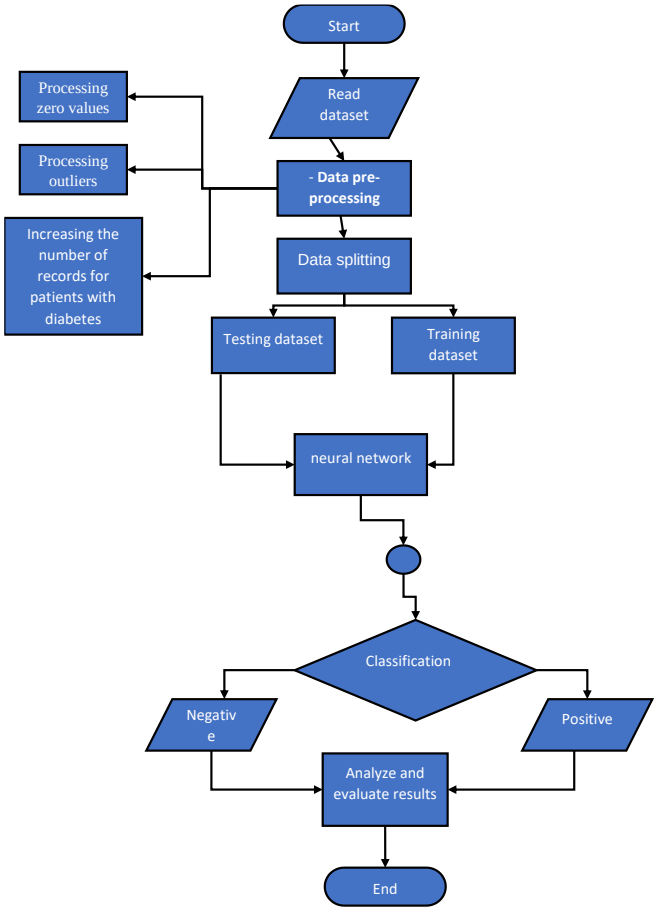

Fig 1 shows the model used.

dataset

The dataset that we used in this paper was obtained from website Kaggle. It is a Pima dataset compiled by the National Institute of Diabetes and consists of 768 records, of which 500 had diabetes and 268 did not. It contains 9 features.

Table 2 *presents the features of the dataset*

feature	**Data type**
Pregnancies	integer
Glucose	integer
BloodPressure	integer
SkinThickness	integer
Insulin	integer
BMI	double
DiabetesPedigreeFunction	double
Age	integer
Outcome	integer

Data exploration

Fig.2 shows some important information about the dataset, such as the arithmetic mean, the highest value, and the lowest value for each feature.

Column	Exclude Column	Minimum	Maximum	Mean	Standard Deviation	Variance
Pregnancies		0	17	3.845	3.370	11.354
Glucose		0	199	120.895	31.973	1022.248
BloodPressure		0	122	69.105	19.356	374.647
SkinThickness		0	99	20.536	15.952	254.473
Insulin		0	846	79.799	115.244	13281.180
BMI		0	67.100	31.993	7.884	62.160
DiabetesPedigreeFunction		0.078	2.420	0.472	0.331	0.110
Age		21	81	33.241	11.760	138.303

Showing 1 to 8 of 8 entries

Fig 2 Statistical information about features.

A bar chart

is a graph that displays categorical data in rectangular columns with heights proportional to the values they represent. Fig.3 displays the bar graph of the Pima data set.

It is observed through Fig.3 that the feature values for diabetic patients are greater than those of healthy people

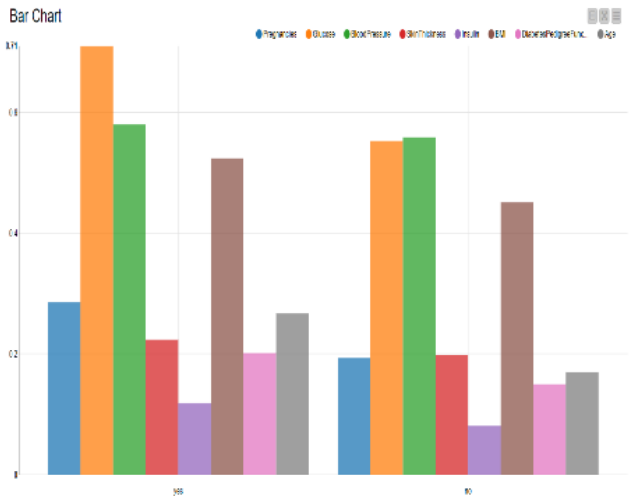

Fig 3 displays the bar chart of the Pima data set

neural network used

The designed network consists of three layers. The number of iterations was set to be 100

IV. Results and discussion

Figure 4 is the confusion matrix of the designed neural network. 166 samples were correctly predicted against 8 samples were incorrectly predicted. The accuracy reached 95.40%. The accuracy of the Recall scale for infected patients reached 96.59%. The result reached is better than the result of the previous studies mentioned in this paper.

MLP LEANER

Confusion Matrix

Rows Number : 174	no (Predicted)	yes (Predicted)
no (Actual)	81	5
yes (Actual)	3	85
	96.43%	94.44%

Fig 4 Displays the confusion matrix of a designed neural network

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
95.40%	4.60%	0.908	166	8

Fig 5 Overall Statistics

V. Conclusion

In this paper, work has been done to design a neural network that contains more than one hidden layer to predict diabetes. The designed network was trained on a Bima data set. The designed network gave good results and is better than the previous studies mentioned in this paper. The accuracy of the designed neural network in this paper reached 95.40%. The accuracy of the Recall scale for infected patients reached 96.59%.

REFERENCES

[1] *2019 International Conference on Intelligent Computing and Control Systems (ICCS).* IEEE.

[2] W. Shen, INESC TEC (Organization), Universidade de Trás-os-Montes e Alto Douro, M. IEEE Systems, International Working Group on Computer Supported Cooperative Work in Design, and Institute of Electrical and Electronics Engineers, *Proceedings of the 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD) : May 6-8, 2019, Porto, Portugal*.

[3] S. Khan Maliha and M. A. Mahmood, "An Efficient Model for Early Prediction of Diabetes Utilizing Classification Algorithm," in *Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022*, 2022, pp. 1607–1611. doi: 10.1109/ICICCS53718.2022.9788441.

[4] M. I. Qrenawi and W. Al Sarraj, "Identification of Cardiovascular Diseases Risk Factors among Diabetes Patients Using Ontological Data Mining Techniques," in *Proceedings - 2018 International Conference on Promising Electronic Technologies, ICPET 2018*, Nov. 2018, pp. 129–134. doi: 10.1109/ICPET.2018.00030.

[5] T. B. Ho, Quỹ phát triển khoa học công nghệ quốc gia (Vietnam), IEEE Vietnam Section., Nhà xuất bản Khoa học và kỹ thuật, and Institute of Electrical and Electronics Engineers., *NICS 2018 : 2018 5th NAFOSTED Conference on Information and Computer Science (NICS) : proceedings : November 23-24, 2018, Ho Chi Minh City, Vietnam*.

[6] P. N. Astya, Galgotias University. School of Computing Science and Engineering, Institute of Electrical and Electronics Engineers. Uttar Pradesh Section, and Institute of Electrical and Electronics Engineers, *IEEE International Conference on Computing, Communication and Automation (ICCCA 2017) : proceeding : on 5th-6th May, 2017*.

[7] SCAD Institute of Technology, IEEE Electron Devices Society, and Institute of Electrical and Electronics Engineers, *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC 2017) : 10-11, February 2017*.

[8] *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC).* IEEE.

[9] IEEE Robotics and Automation Society, Tianjin li gong da xue, Kagawa Daigaku, Beijing li gong da xue, Guo jia zi ran ke xue ji jin wei yuan hui (China), and Institute of Electrical and Electronics Engineers, *2019 IEEE International Conference on Mechatronics and Automation : IEEE ICMA 2019 : August 4-7, 2019, Tianjin, China.*

[10] Institute of Electrical and Electronics Engineers, *2018 4th International Conference on Frontiers of Signal Processing (ICFSP 2018) : September 24-27, 2018, Poitiers, France.*

[11] P. Nagaraj, V. Muneeswaran, A. Dharanidharan, K. Balananthanan, M. Arunkumar, and C. Rajkumar, "A Prediction and Recommendation System for Diabetes Mellitus using XAI-based Lime Explainer," in *International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2022 - Proceedings*, 2022, pp. 1472–1478. doi: 10.1109/ICSCDS53736.2022.9760847.

[12] H. C. Lagunzad, M. A. C. Impang, M. V. Gonzaga, J. F. Lawan, F. C. Pineda, and R. A. A. Tanjente, "Predicting the Early Sign of Diabetes using ID3 as a Data Model," in *2022 14th International Conference on Computer and Automation Engineering, ICCAE 2022*, 2022, pp. 135–139. doi: 10.1109/ICCAE55086.2022.9762442.

[13] Özkaya, U., Öztürk, Ş., & Barstugan, M. (2020). Coronavirus (COVID-19) classification using deep features fusion and ranking technique. *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*, 281-295.

[14] C. I. Ossai and N. Wickramasinghe, "A hybrid approach for risk stratification and predictive modelling of 30-days unplanned readmission of comorbid patients with diabetes," *J Diabetes Complications*, vol. 36, no. 6, Jun. 2022, doi: 10.1016/j.jdiacomp.2022.108200.

[15] N. Dunbray, R. Rane, S. Nimje, J. Katade, and S. Mavale, "A Novel Prediction Model for Diabetes Detection Using Gridsearch and A Voting Classifier between Lightgbm and KNN," in *2021 2nd Global Conference for Advancement in Technology, GCAT 2021*, Oct. 2021. doi: 10.1109/GCAT52182.2021.9587551.

[16] G. S. Tomar and Institute of Electrical and Electronics Engineers, *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT) proceedings*.

[17] D. Kaur Bhullar *et al.*, "Developing a Predictive Supervised Machine Learning Models for Diabetes," in *7th International Conference on Computing, Engineering and Design, ICCED 2021*, 2021. doi: 10.1109/ICCED53389.2021.9664833.

[18] S. Samet, M. R. Laouar, and I. Bendib, "Diabetes mellitus early stage risk prediction using machine learning algorithms," in *5th International Conference on Networking and Advanced Systems, ICNAS 2021*, 2021. doi: 10.1109/ICNAS53565.2021.9628955.

[19] A. Prakash, R. Anand, S. S. Abinayaa, and N. S. Kalyan Chakravarthy, "Normalized Naïve Bayes Model to predict Type-2 Diabetes Mellitus," in *2021 IEEE International Conference on Emerging Trends in Industry 4.0, ETI 4.0 2021*, 2021. doi: 10.1109/ETI4.051663.2021.9619332.

[20] A. A. Khan, H. Qayyum, R. Liaqat, F. Ahmad, A. Nawaz, and B. Younis, "Optimized Prediction Model for Type 2 Diabetes Mellitus Using Gradient Boosting Algorithm," in *Proceedings of the 2021 Mohammad Ali Jinnah University International Conference on Computing, MAJICC 2021*, Jul. 2021. doi: 10.1109/MAJICC53071.2021.9526257.

[21] Sri Shakthi Institute of Engineering and Technology, Institute of Electrical and Electronics Engineers. Madras Section, All-India Council for Technical Education, and Institute of Electrical and Electronics Engineers, *2020 International Conference on Computer Communication and Informatics : January 22-24, 2020, Coimbatore, India*.

[22] China Research Council of Computer Education in Colleges & Universities, Ontario Tech University, IEEE Education Society, and Institute of Electrical and Electronics Engineers, *The 14th International Conference on Computer Science and Education (ICCSE 2019) : August 19 -21, Toronto, Canada*.

[23] Surya Engineering College and Institute of Electrical and Electronics Engineers, *Proceedings of the 3rd International Conference on Computing Methodologies and Communication (ICCMC 2019) : 27-29, March 2019*.

[24] A. S. Sunge *et al.*, "Prediction Diabetes Mellitus Using Decision Tree Models; Prediction Diabetes Mellitus Using Decision Tree Models," 2019.

[25] Sri Sairam Engineering College. Department of Information Technology and Institute of Electrical and Electronics Engineers, *2019 proceedings of the 3rd International Conference on Computing and Communications Technologies (ICCCT'19) : February 21-22, 2019, Chennai, India.*

[26] *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*.

[27] O. Bayat, S. Aljawarneh, H. F. Carlak, International Association of Researchers, Institute of Electrical and Electronics Engineers, and Akdeniz Üniversitesi, *Proceedings of 2017 International Conference on Engineering & Technology (ICET'2017) : Akdeniz University, Antalya, Turkey, 21-23 August, 2017*.