ofelia

by - -

Submission date: 21-Mar-2024 01:09PM (UTC+0000)
Submission ID: 227404687
File name: 7743-20905-1-RV.docx (117.41K)
Word count: 1960
Character count: 10779

JIKO (Jurnal Informatika dan Komputer)
Vol. x, No. x, April 2024, pp. x-x
DOI: 10.33387/jiko

Accredited KEMENDIKBUD RISTEK, No.105/E/KPT/2022
p-ISSN: 2614-8897
e-ISSN: 2656-1948

ARTIFICIAL NEURAL NETWORK MULTI-LAYER PERCEPTRON FOR DIAGNOSIS OF DIABETES MELLITUS

Ofelia Cizela da Costa Tavares[1], Abdullah Z. Abidin[2]

[1]Department of Computer Sciences, Dili Institute of Technology
[2] Professional IT Consultant Enterprise at Malaysia
*Email: [1]offytasyah1803@gmail.com, [2]abdz.abidin@gmail.com

(Received: dd mmm yyyy, Revised: dd mmm yyyy, Accepted: dd mmm yyyy)

Abstract

Diabetes Mellitus is a disease caused by an unhealthy lifestyle so blood sugar is not controlled, causing complications. This disease is one of the most dangerous diseases in the world. Approximately 422 million people worldwide suffer from diabetes, the majority living in low- and middle-income countries and 1.5 million deaths are caused by diabetes each year. The number of cases and prevalence of diabetes have continued to increase over the last few decades. Artificial Neural Networks are a part of machine learning that can solve various problems. One of them is in terms of disease diagnosis. MLP has the advantage that learning is done repeated so that it can create a system that is durable and consistent and works well. This research aims to implement the Multi-Layer Perceptron Artificial Neural Network method for diagnosing diabetes mellitus, and then evaluating the MLP by analyzing precision, recall, f1 score and calculating accuracy. Next, it is validated with k-fold cross-validation. In the experiment in this study, several scenarios were used, where the best scenario was obtained when using 8 input layers, 7 hidden layers, and 1 output layer and 5000 iterations. The results of the experiment showed that the multi-layer perceptron was successful in classifying diabetics and non-diabetics by percentage. Precision 77.24%, Recall 72.58%, F1 Score 76.86%, accuracy 75% and average accuracy 78.01%.

Keywords: *Backpropagation, Diabetes Mellitus, Multi-Layer Perceptron.*

**Corresponding Author*: *Ofelia Cizela da Costa Tavares*

1. INTRODUCTION

Diabetes mellitus is a chronic disease that causes up to 1.5 million deaths a year (WHO, 2023). This disease is synonymous with an unhealthy lifestyle, the amount of sugar and calorie intake that enters the body is not balanced with sufficient body activity so sugar accumulates in the blood and becomes diabetes or what is more commonly called diabetes mellitus [1].

Various efforts have been made to reduce the rate of diabetes mellitus sufferers. One way is to carry out an initial diagnosis of diabetes mellitus. For initial diagnosis, of course, skilled health workers are needed, but the lack of health workers in various countries is another obstacle. However, the development of artificial intelligence has become a solution for the world of health. Various studies have been carried out for the diagnosis of diabetes mellitus [2-5].

This research will utilize a Neural Network machine learning algorithm. In artificial neural networks, various models can be applied. For this research, a multi-layer perceptron was used to diagnose diabetes mellitus. The use of MLP has been carried out by several other researchers on various types of objects and is considered reliable [6-7]

In the proposed research, the evaluation process uses precision, recall, f1 score, and accuracy calculation analysis. The aim is to determine the reliability of the multi-layer perceptron model in diagnosing diabetes mellitus. Then validation will be carried out using k-fold cross-validation. The validation process is carried out because validating the performance of the model can be improved

2. RESEARCH METHOD

Multi-layer perceptron is part of an Artificial Neural Network (ANN). Consists of several neurons, and there are connections between these neurons. These neurons will transform the information received through their output connections to other neurons. In a multi-layer perceptron, this relationship is known as

weight. This information is stored at a certain value at a weight. In multi-layer perception uses the Backpropagation algorithm. The following are the steps of this research.

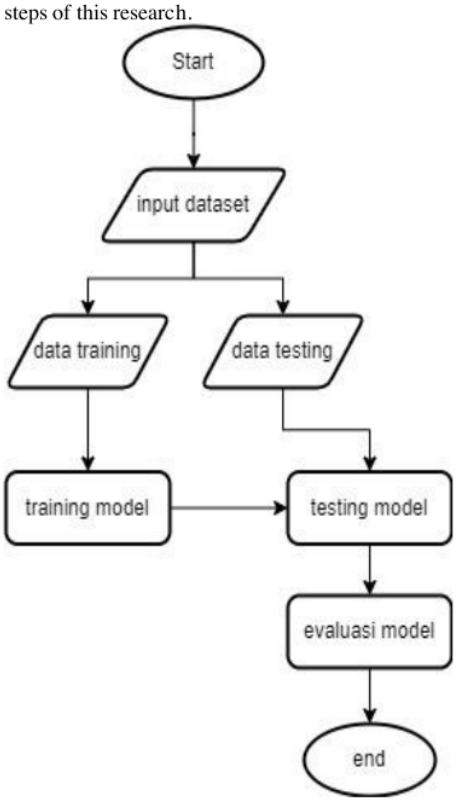

Figure 1. research steps.

2.1 Dataset

The type of data used in this research is secondary data. This data is data that comes from kaggle. This data has information about 768 female patients. The types of data are shown in Table 1.

Tabel 1. Lists of criteria

No	criteria	Unit
1	*Number of times pregnant*	
2	*Plasma glucose concentration*	mg/dl
3	*Diastolic blood pressure*	mmhg
4	*Triceps skin fold thickness*	mm
5	*2-Hour serum insulin*	mu U/ml
6	*Body mass index*	kg/m^2
7	*Diabetes pedigree function*	-
8	*Age*	Years
9	*Class*	

2.2 Architecture backpropagation

In this research, the machine learning model used is the multi-layer perceptron. The steps for this model have been carried out by [8] which applies the backpropagation algorithm. In this research, testing was carried out using several scenarios, including initializing the hidden layer and increasing the number of iterations. Figure 2 is an illustration of the backpropagation architecture with input layer 8, hidden layer 2 and output layer 1.

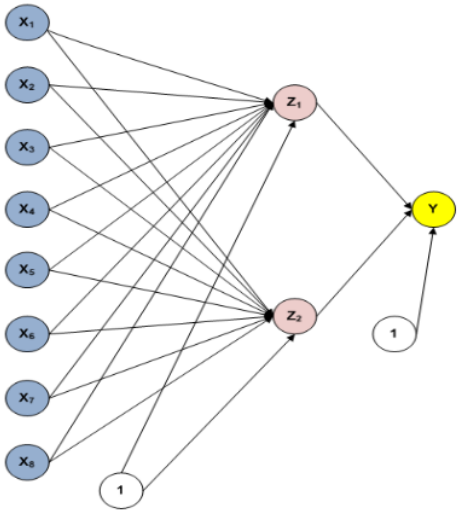

Figure 2. Architecture Backpropagation

2.3 Evaluation Model

In research to evaluate the multi-layer perceptron model, a confusion matrix is used. The confussion matrix is a table that states the number of test data that are correctly classified and the number of test data that are incorrectly classified [9]. The table for the confusion matrix is shown in Table 2.

Table. 2 *Confusion Matrix*

		Class prediction	
		1	0
Real class	1	TP	FN
	0	FP	TN

TP (*True Positive*), FP (*False Positive*), FN (*False Negative*), TN (*True Negative*).

After creating the confusion matrix table, the accuracy value is then calculated using Equation 1, then presisi using Equation 2, Recall using Equation 3, And f1 score using Equation 4 [10].

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (1)$$

$$Precision = \frac{\sum TP}{\sum TP + \sum FP} \times 100\% \quad (2)$$

$$Recall = \frac{\sum TP}{\sum TP + \sum FN} \times 100\% \quad (3)$$

$$F1 - Score = 2\,x\,\frac{precission\ x\ recall}{precisin\ +\ recall} \times 100\% \qquad (4)$$

In the research, model validation will also be carried out using k-fold cross validation. This is done so that all data has the same rights, both as training data and testing data. Apart from that, in validation using k-fold cross-validation the accuracy value can increase. [11].

3. RESULT AND DISCUSSION

The dataset in this study was taken from Kaggle which can be accessed openly via the following link https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset. This data is divided into 2 categories, namely data with classes 0 and 1. The distribution of the amount of data from each class is 500 for class 0 and 268 for class 1, as shown in Figure 3.

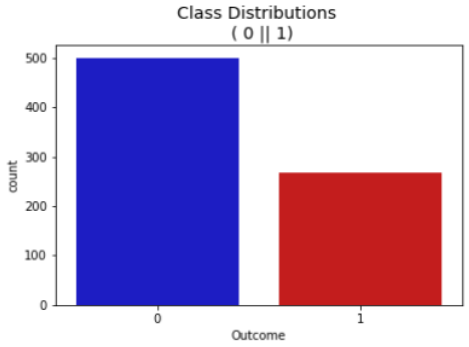

Figure 3. Distribution of class

3. 1 Experimental result

In the research, the experimental process consists of adding hidden layers and the number of iterations, with a total of 6 trials. for dataset division, 80% training data and 20% test data. The testing process has also been validated using k-fold cross validation with 10-fold. The experimental results are shown in Figure 4 for the confusion matrix with the highest accuracy results and all experiments are shown in Table 3.

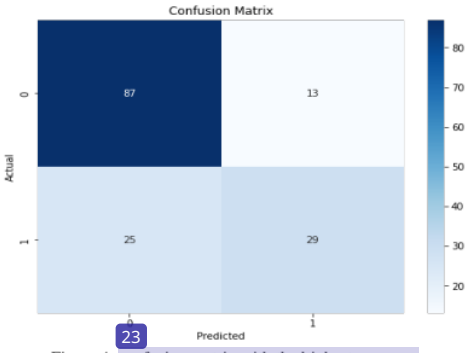

Figure 4. confusion matrix with the highest accuracy

Table 3. Experimental results

No	Hidden layer	iteration	accuracy
1	7	5000	75%
2	2	5000	73%
3	5	5000	71%
4	5	10000	71%
5	2	10000	73%
6	7	10000	75%

Determination of hidden layers

In experiments on research into the classification of diabetes in women using a multi-layer perceptron, the process of determining hidden layers was carried out in 3 types, namely 2, 5, and 7. The experimental results show that initializing the hidden layer with the number 7 has very good model performance. Where an accuracy value of 75% is obtained. Then, when validation was carried out using k-fold cross validation, the average accuracy value for hidden layer 7 obtained an accuracy value of 78.01%.

Determination of the number of iterations

Apart from determining hidden layers, this research also tested by determining the number of iterations. There are two initializations for the number of iterations carried out in the classification of diabetes in women, namely 5000 and 10.000. Based on experimental results, especially for the dataset used in this research, determining the number of iterations does not affect the performance of the model for diabetes classification.

Results of precision, recall, and f1 score

Apart from the testing process by calculating accuracy values, this research also calculated precision, recall and f1-score values. The same is true for calculating accuracy. In this study, the precision, recall and f1 score were highest when hidden layer 7 was used. Precision was 77.24%, recall was 72.58% and f1 score was 76.86%. All experimental results are shown in Table 4.

Table 4. precision, recall and f1 score evaluation results

hidden layer	presisi	recall	f1 score
7	77.24%	72.58%	76.86%
2	70.55%	68.62%	72.39%
5	75.88%	71.92%	76.02%

Validation results with k-fold cross-validation

There is a difference between calculating accuracy without validation and using validation. Where when validating using 10-fold, the average accuracy experienced a very significant increase, especially in the hidden layer 5 scenario. Before carrying out validation the resulting accuracy was 71%

but after being validated with 10-fold the average accuracy was 77.04%. The accuracy value increased by 6%, especially in hidden layer 5. The difference in results without validation and using k-fold cross validation is shown in Figure 6.

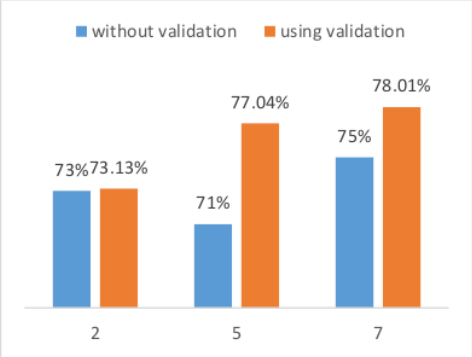

Figure 6. Comparison results using validation and without validation.

4. CONCLUSION

Based on the results of experiments that have been carried out, show that determining the number of hidden layers can influence the performance of the multi-layer perceptron model in the classification of diabetes mellitus in women, where the highest accuracy value was obtained when using hidden layer 7 with an accuracy value of 75%, precision 77.24%, recall 72.58 and f1 score 76.86%. Meanwhile, determining the number of iterations cannot affect the performance of the multi-layer perceptron model, especially on the dataset used in the research. Then the data validation process is also able to increase the accuracy of diabetes mellitus classification. Especially in the scenario with the initialization of hidden layer 5, accuracy increased by more than 6%.

5. REFERENCE

[1] C. Catherine *et al*., *Prevalence of diabetes and high risk for diabetes using A1C criteria in the US population in 1988–2006*.. 2010. Diabetes Care, 33. 562-568

[2] A. Ameer *et al*., "DIABETES CLASSIFICATION BASED ON KNN: *IIUM Engineering Journal*., vol. 21, no. 1, pp. 175–181, 2020, doi: 1 0.31436/iiumej.v21i1.1206.

[3] V. Anuja Kumari, and R.Chitra. "Classification Of Diabetes Disease Using Support Vector Machine," *International Journal of Engineering Research and Applications,*, vol. 3, no. 1, pp, 1797-1801. 2013.

[4] N. Zaman, *et al*., "Classification of Diabetes using Machine Learning," *Proc. IEEE Comput, 2022 International Conference on Futuristic Technologies (INCOFT)*, vol. 2022-November 2022, doi: 10.1109/INCOFT55651.2022.10094335. Belgaum. India

[5] A. F. Limas, *et al*., "A Comparative Analysis on the Evaluation of KNN and SVM Algorithms in the Classification of Diabetes," *Scientific Journal of Informatics* ., vol. 10, no. 3. pp, 251-26, 2023, doi: 10.15294/sji.v10i3.44269.

[6] S. S. Chai, *et al*., "A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in Sarawak, Malaysia," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 2, pp. 1–11, 2021, doi: 10.1155/2021/2794888

[7] N. F. S. Neto, *et al*., "A Study of Multilayer Perceptron Networks Applied to Classification of Ceramic Insulators Using Ultrasound," *Applied Sciences*, vol. 11, pp. 1-19, 2021, doi: 0.3390/app11041592.

[8] F. Martinez, H. Montiel, and F. Martinez, "A Machine Learning Model for the Diagnosis of Coffee Diseases," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022, doi: 10.14569/IJACSA.2022.01304110

[9] F. Gorunescu, Data Mining Concepts, Models and Techniques. Chennai, India: Scientific Publishing Services Pvt. Ltd, 2011. doi: 10.1007/978-3-642-19721-5.

[10] H. L. Gope and H. Fukai, "Peaberry and normal coffee bean classification using CNN, SVM, and KNN: Their implementation in and the limitations of Raspberry Pi 3," *AIMS Agriculture and Food*, vol. 7, no. 1, pp. 149–167, 2022, doi: 10.3934/agrfood.2022010.

[11]. Zhao et al., "Real-time recognition system of soybean seed full-surface defects based on deep learning," *Computers and Electronics in Agriculture*, vol. 187, p. 106230, Aug. 2021, doi: 10.1016/j.compag.2021.106230.

ofelia

ORIGINALITY REPORT

20%	15%	15%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mid-day.com Internet Source	2%
2	Submitted to University of Sydney Student Paper	2%
3	Submitted to Universitas Amikom Student Paper	2%
4	dspace.umkt.ac.id Internet Source	2%
5	www.jatit.org Internet Source	1%
6	dokumen.pub Internet Source	1%
7	Nadisa Karina Putri, Zuherman Rustam, Devvi Sarwinda. "Learning Vector Quantization for Diabetes Data Classification with Chi-Square Feature Selection", IOP Conference Series: Materials Science and Engineering, 2019 Publication	1%
8	www.researchgate.net Internet Source	

		1 %
9	Nurfadila Utami, Mustakim. "Data Distribution Modelling in Supervised Learning Algorithm is for The classification of Prospective Recipient Candidate", 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2021 Publication	1 %
10	B C Pratama, Ismoyowati, M N Innayah. "Livestock and Animal Specialities Company in ASEAN: Intellectual Capitals and Performances", IOP Conference Series: Earth and Environmental Science, 2019 Publication	1 %
11	www.mdpi.com Internet Source	1 %
12	www.nature.com Internet Source	1 %
13	Takahiro Kitajima, Takashi Yasuno, Hitoshi Sori. "Study on output prediction system of wind power generation using complex-valued neural network with multipoint GPV data", IEEJ Transactions on Electrical and Electronic Engineering, 2013 Publication	1 %

14	Y B N D Artissa, I Asror, S A Faraby. "Personality Classification based on Facebook status text using Multinomial Naïve Bayes method", Journal of Physics: Conference Series, 2019 Publication	1 %
15	ebin.pub Internet Source	1 %
16	talenta.usu.ac.id Internet Source	1 %
17	Ahmad A. M. Abushariah, Mohammad A. M. Abushariah, Teddy Surya Gunawan, J. Chebil et al. "Fusion of speech and handwritten signatures biometrics for person identification", International Journal of Speech Technology, 2023 Publication	1 %
18	link.springer.com Internet Source	1 %
19	researchinventy.com Internet Source	1 %
20	www.hindawi.com Internet Source	1 %
21	Ahmad Al-Khasawneh. "chapter 30 A Method for Classification Using Data Mining Technique for Diabetes", IGI Global, 2016	<1 %

Publication

22	www.beei.org Internet Source	<1%
23	Shuang Yang, Anqi Jin, Xiangyang Zeng, Haitao Wang, Xi Hong, Menghui Lei. "Underwater acoustic target recognition based on knowledge distillation under working conditions mismatching", Multimedia Systems, 2024 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		