Classification and diagnosis of diabetic with neural network algorithm learning vector quantizatin (LVQ)

by Arnita Arnita

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Classification and diagnosis of diabetic with neural network algorithm learning vector quantizatin (LVQ)

IOP Conf. Series: Journal of Physics: Conf. Series **1188** (2019) 012091 doi:10.1088/1742-6596/1188/1/012091

Classification and diagnosis of diabetic with neural network algorithm learning vector quantizatin (LVQ)

Arnita, M S Sinaga and Elmanani
Universitas Negeri Medan, Jalan Willem Iskandar Pasar V Medan, Indonesia

E-mail: arnita@unimed.ac.id

Abstract. Determining the type of Diabetes Mellitus (DM) is very important to determine what treatment is suitable for a patient. Unfortunately patient information about what type of diabetes is often ignored, so the patient gets a wrong diagnosis. This study aims to build a classification model in determining a DM patient diagnosed with one type of DM, namely type 1 DM, type 2 DM, Gestational DM or special type DM. The indicators used in determining the classification for diagnosing patients are age, sex, blood pressure, levels of blood glucose, weight, and height. The classification method used is the Neural Network method with Learning Vector Quantization (LVQ) algorithm. Algorithm LVQ provides results 96% accuracy for training data with final epoch is 759 and 90% accuracy for testing data.

Introduction

Diabetes Mellitus (DM) is not a new phenomenon for the people of Indonesia. The World Health Organization (WHO) later released data that the number of patients with Diabetes Mellitus (DM) has increased to 422 million [1]. Diabetes is a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both. Chronic hyperglycemia of diabetes is associated with long-term damage, dysfunction and failure of various glands, especially the eyes, kidneys, nerves, heart and blood vessels. It is important to know the signs and symptoms of diabetes as well as the diagnosis is to be treated as soon as possible [2]. However, over the past few decades, research has found that different types of diabetes have different causes may be similar despite the pathology program.

Based on the definition of the National Diabetes Data Group (NDDG) and the World Health Organization (WHO), there are four types of diabetes, namely diabetes mellitus insulin-dependent (type 1 diabetes), diabetes mellitus, non-insulin-dependent (type 2 diabetes), gestational diabetes mellitus (DMG), and diabetes secondary to other conditions. Diabetes can be diagnosed with the classic signs and symptoms of diabetes and blood glucose levels are very high, with a fasting plasma glucose (FPG) ≥ 140 mg / dl, or with glucose venous plasma ≥ 200 mg / dL at 2 hours after 75 g of glucose oral [2-4]. However, to determine a person's entry in the category DM particular type is not easy. Because that it depends on the circumstances at the time a diagnosis is made.So far most doctors and patients feel less important to label certain types of diabetes than understanding the pathogenesis of hyperglycemia and to treat it effectively.

In a case in which a pregnant woman is diagnosed with gestational diabetes mellitus (GDM) as having blood glucose levels are higher than normal after delivery, it turns out after a series of laboratory tested women suffering from type 2 diabetes. Another case where the person is diagnosed with diabetes because of taking steroids exogenous and the blood glucose to normal after stopping taking

IOP Conf. Series: Journal of Physics: Conf. Series **1188** (2019) 012091 doi:10.1088/1742-6596/1188/1/012091

glucocorticoids. Therefore, the determination of a person diagnosed with diabetes mellitus type what's important to do, so that treatment is given according to the condition and type of diabetes disease suffered by patients.

During this time a person suffering from diabetes diagnosis enforcement of certain types go through the stages of laboratory tests and doctor. However, the method of determining the classification in data mining diagnose a patient suffering from a certain type of diabetes can be done through an algorithm.*Learning Vector Quantization*(LVQ) is a classification method in which each unit of output presented a class. LVQ is used for grouping where the number of groups already defined architecture (target / classes are specified). LVQ provides some interesting features that is easy to apply and rationally understood. This resulted in many researchers and practitioners choose to use as a method of classification LVQ robust than other artificial neural network method that relies on a lot of black box [5], According to some peneliian LVQ give better results with a high degree of accuracy and a small error [6], based on research [7] LVQ provide accuracy values closer to actual results with the number of different iterations and the accuracy of 79.31% for iteration = 60 and 90. While in research [8] LVQ able to predict the accuracy of signatures to 98% on the test data. Therefore, it will be developed classification model using LVQ to determine the type of DM diagnosed a patient.

2. Methodology

The data used in this study is the medical records of patients with diabetes mellitus in Pirngadi Hospital, Medan from years 2015-2018. From the existing data will be established diabetics classification model based on the type of diabetes that is there, namely diabetes type 1, diabetes type 2, and gestational diabetes mellitus type DM special type. However, due to limited availability of data, only 3 DM type class that will be used as the output type 1, type 2 and type special .. There are 57 medical records with details of 12 patients with diabetes mellitus type 1 DM patient data, patient data 42 DM type 2 and 3 DM patient data with special conditions. based on [4] variables used in this classification is x_1 = gender, x_2 = weight, height = x_3, x_4 = blood pressure, x_5 = random blood glucose levels (GDS), and x_6 = age. To form the classification model using LVQ DM type, then the data is divided into two parts, namely the data for training and testing data for a total of 45 randomly selected data is used for training and the remaining 12 are used for testing.

Learning Vector Quantization(LVQ) is a classification method in which each unit of output presented a class. LVQ is used for grouping where the number of groups already defined architecture (target / classes are specified). The working principle of LVQ algorithm is the reduction of its neighboring nodes (neighbor) and illustrated through asitektur LVQ in figure 1 below [9].

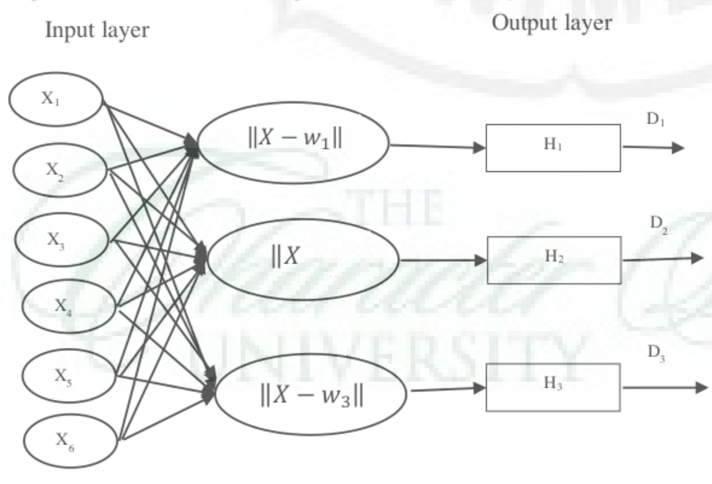


Figure 1. Architecture LVQ

IOP Conf. Series: Journal of Physics: Conf. Series **1188** (2019) 012091 doi:10.1088/1742-6596/1188/1/012091

Information

- x1 - x6 is the input value
- $\|X - w_1\|$ up with is the distance weighting $\|X - w_3\|$
- H1 - H3 is the output layer
- D1 - D3 is the output value

Flowchart LVQ development stage classification model for diagnosing diabetes patients is shown in Figure 2:

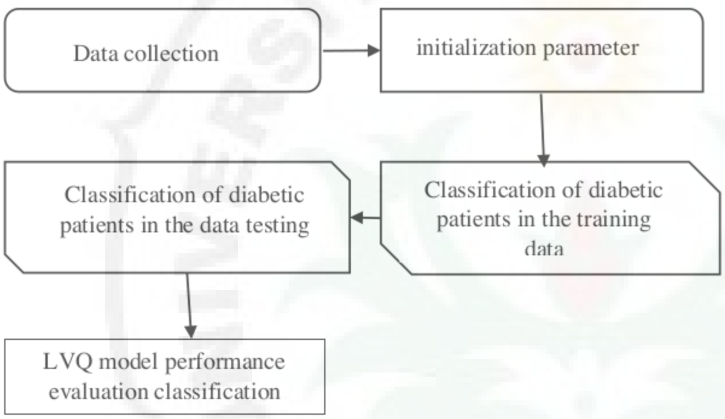


Figure 2. Flowchat Classification Model Development Stage Diagnosis of Diabetes Patients

While the steps of the LVQ algorithm on training data are [9-10]:

1. Initialization of the initial weight (w) and LVQ parameters, namely maxEpoch, α, decα, and minα.
2. Enter the input data (X) and a target class (T).
3. Set initial conditions: the epoch = 0.
4. Do if (*epoch* <*maxEpoch*) and (α ≥ min α)
   a. *epoch* = *Epoch* + 1
   b. *set J such that* $\|X_i - w_j\|$ minimal use of calculation formulas *Euclidian distance*
      D (j) =$\sum(W_{ij} - x_i)^2$ (1)
   c. Fixed Wj with the following provisions:
      If T = Cj then Wj (t + 1) = wj (t) + α (t) [x (t) -wj (t)] (2)
      If the T ≠ Cj then Wj (t + 1) = wj (t) - α (t) [x (t) -wj (t)] (3)
   d. Subtract the value of α by: α = α - α * α dec (4)
5. Test the stop condition with the output of the optimal weight.

IOP Conf. Series: Journal of Physics: Conf. Series **1188** (2019) 012091 doi:10.1088/1742-6596/1188/1/012091

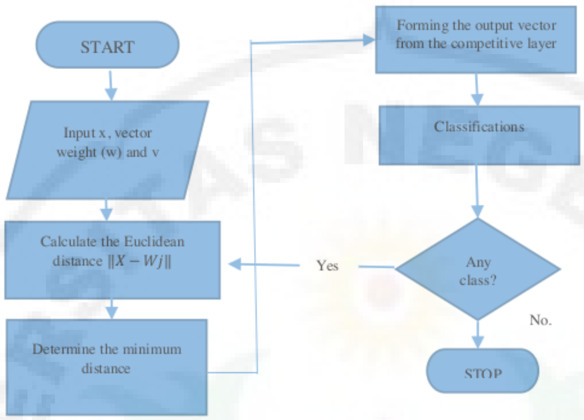


Figure 3. Flow Chart Testing LVQ

After classification models built using training data obtained by the testing involves testing and measuring accuracy of data generated classification model. To measure the performance of model classification parameters used is the level of accuracy. A within the classification system are expected to classify all data sets properly, but there is no doubt that the performance of a system can not be 100% accurate. To measure the level of accuracy can be used the following formula [11]:

$$\text{Accuracy} \frac{= Mount\ Of\ data\ predicted\ correctly}{Number\ of\ predictions\ made}$$

3. Results and Discussion

In the process of network training used some parameters to build a classification model diagnose patients with diabetes mellitus, these parameters are defined in Table 1.

Table 1. Parameters used in the training process LVQ network

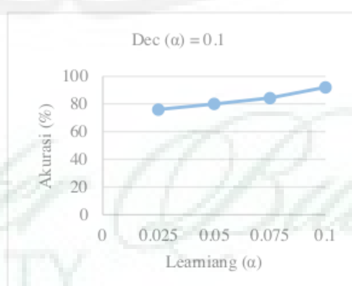
Indicator	Value
Number Patterns Put on Training	45 Data
Number Patterns Feedback on Testing	12 Data
Number Patterns Target	3 class
Variations in the training rate (α)	0.1; 0075; 0.050; 0025
Renewal rate α dec Training	0:01; 0.1; 0:25; 0.5; 0.75
Maximum iterations (epoch)	1000

The value on the parameter α that is used is in the range of 0 <α <1 [12]. The results of the training pross then matched against the target data obtained based on diagnoses that are of medical records. If the result of the training process carried provides high compatibility to the actual data, the accuracy of the training process should reach a high percentage. Evaluate the accuracy of the training process is obtained from the parameters that have been defined in Table 2.

Table 2. Classification of Diabetes Model Training on Various Parameters

Dec (α)	α	final epoch	Level of accuracy (%)
0:01	0.1	759	93
	0075	525	89.5
	0:05	340	89.5
	0025	270	88
0.1	0.1	110	92
	0075	90	84.3
	0:05	85	80
	0025	73	76
0:25	0.1	65	92
	0075	53	83.5
	0:05	44	80
	0025	37	79.9
0.5	0.1	31	88
	0075	28	79.9
	0:05	24	70
	0025	20	68.7
0.75	0.1	17	85
	0075	12	83.8
	0:05	10	80
	0025	7	78.7

In Table 2 shows that the highest accuracy was obtained by 93% in network training, the learning rate parameter α of 0.1, dec α = 0:01 and the last epoch value is 759. Average minimum accuracy of 78.7% with parameter α = 0.025, α dec = 0.75 and last epoch was 7.



IOP Conf. Series: Journal of Physics: Conf. Series **1188** (2019) 012091 doi:10.1088/1742-6596/1188/1/012091

Figure 4. The pattern of increase in the value of the parameter learniang accuracy rate (α) change and constant parameter α dec

Simulation of changes in the value of learning rate and the reduction rate α (dec α) training data shown in Figure 4, shows that the value of the learning rate α dec affect the value of α and accuracy. Wherein if the value pengrurangan pembelajan rate (dec α) is constant and the value of α decreases learniang will result in a decrease in the level of accuracy that is generated. However, if the learning rate (α) remains and reduction in the rate of learning (dec α) decreases, the level of accuracy of the training process provides a level of accuracy that is not stabilized (fluctuating). This is consistent with the results of research conducted [12]. That the greater the value of α used will result in the rate calculation is not stable. Conversely smaller value of α used will result in higher accuracy value calculation but the rate will be slowed down.

In testing process was performed using 12 data randomly selected from a total of 57 data and represent every class there. Parameters used in the testing process is the same as the parameters is done on the training process are shown in Table 1. The results of the testing process classification models using LVQ diabetes patients obtained the results shown in Table 3:

Table 3. Classification of Diabetes Model Tests on Various Parameters

Dec (α)	α	Level of accuracy (%)
0:01	0.1	90
	0075	89
	0:05	89
	0025	88
0.1	0.1	87
	0075	87
	0:05	86

IOP Conf. Series: Journal of Physics: Conf. Series **1188** (2019) 012091 doi:10.1088/1742-6596/1188/1/012091

	0025	85
0:25	0.1	83
	0075	83
	0:05	82
	0025	81
0.5	0.1	80
	0075	80
	0:05	80
	0025	80
0.75	0.1	80
	0075	79
	0:05	78
	0025	78

There are several grades of high accuracy in the testing process. The results of the testing process, obtained the highest accuracy rate was 90% with a parameterdec α = 0.1 and α = 0.001. And the value of the lowest accuracy in the testing process is obtained on the value of learning rate (α) 0.1 and decreasing learning rate (dec α) = 0.01. While the lowest accuracy value for the testing process was obtained at 78% with the value of learning (α) 0,05 dan 0.025 dan dec(α) = 0.75.

Figure 5. The pattern of increasing accuracy values based on dec (α) parameters varies and the learniang rate (α) parameters are constant

It is the same as the training process, where if the pace of learning α is constant and the learning rate decreases (dec α) declining will result in a decrease in the level of accuracy produced. However, if the value of the learning rate (α) is constant and the value of the learning rate reduction (dec α) decreases, then the level of accuracy in the testing process will experience an increasing pattern. So that it can be said that the development of the diabete patient classification model is highly influenced by the learning rate value (α) and dec (α). Where optimal generated under simulated parameters testing process is of the highest accuracy at a value of 90%, the value of the learning rate (α) = 0.1 and a reduction in the rate of learning (dec α) = 0.01, respectively.

4. Conclusions

Highest levels of accuracy obtained in simulas formation LVQ classification model learning process was 96% in the testing process and 90%. Training and testing process provides a pattern the same parameters in generating the accuracy of classification models were established. Wherein, the smaller the value of learning rate will result keakuarsian level models are also getting smaller. While the best parameter values to produce a good classification model in the learning process or test is α = 0.1 and dec (α) = 0.01, respectively.

Acknowledgments

Thanks are conveyed to Universitas Negeri Medan through the Research Institute UNIMED which has funded this research. Thanks also to the Dean of MIPA UNIMED for the support of facilities and infrastructure during this research. Thank you also to the Hospital Dr. Pirngadi Medan for the services provided during data collection.

Reference

[1] Choudhury H, *et al* 2018 An update on natural compounds in the remedy of diabetes mellitus: A systematic review *J Tradit Complement Med* **8** (3) 361
[2] Diabetes D O F, 2007 Diagnosis and classification of diabetes mellitus *Diabetes Care* vol 30 no SUPPL 1
[3] Barrera F L, Velázquez A and Pérez L M 2010 Explorando los determinantes del buen manejo forestal comunitario *Interciencia* **35** (8) 560
[4] Harris M I 1995 Classification, Diagnostic Criteria, and Screening for Diabetes *Diabetes Am 2nd* p 18
[5] Comrie B and Polinsky M 1993 Causatives and transitivity *Stud Lang companion Ser. v23* **8** 399
[6] Hamidi R, Furqon M T and Rahayudi B, 2017 Implementasi Learning Vector Quantization (LVQ) untuk Klasifikasi Kualitas Air Sungai *J Pengemb Teknol Inf dan Ilmu Komput* **1** (12) 1758
[7] Meliawati R, Soesanto O and Kartini D 2016 Penerapan Metode Learning Vector Quantization (LVQ) Pada Prediksi Jurusan Di SMA PGRI 1 Banjarbaru *Kumpul J Ilmu Komput* **04** (01) 11
[8] Qur'ani D Y and Rosmalinda S 2010 Jaringan Syaraf Tiruan Learning Vector Quantization untuk Aplikasi Pengenalan Tanda Tangan *Semin Nas Apl Teknol Inf 2010 (SNATI 2010)* **1** 1
[9] Ranadhi D, Indarto W and Hidayat T 2006 Implementasi Learning Vector Quantization (LVQ) Untuk Pengenal Pola Sidik Jari Pada Sistem Informasi Narapidana LP Wirogunan *Media Inform* **4** (1) 51
[10] Lukman A 2015 Implementasi Pengolahan Citra dan Algoritma LVQ untuk Pengenalan Pola Buku no 20
[11] Sudibyo U, Kusumaningrum D P, Rachmawanto E H and Sari C A 2018 Optimasi Algoritma Learning Vector Quantization (LVQ) Dalam Pengklasifikasian Citra Daging Sapi dan Daging Babi Berbasis GLCM dan HSV *J SIMETRIS ISSN 2252-4983* **9** (1) 1
[12] Eliyen K, Tolle H and Muslim M A 2017 Learning Vector Quantization Untuk Klasifikasi Penilaian Pada Virtual Patient Case **2** (2) 98

Classification and diagnosis of diabetic with neural network algorithm learning vector quantizatin (LVQ)

ORIGINALITY REPORT

16%	12%	9%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	procceding.unram.ac.id Internet Source	3%
2	Anak Agung Putri Ratna, F. Astha Ekadiyanto, Mardiyah, Prima Dewi Purnamasari, Muhammad Salman. "Analysis on the Effect of Term-Document's Matrix to the Accuracy of Latent-Semantic-Analysis-Based Cross-Language Plagiarism Detection", Proceedings of the Fifth International Conference on Network, Communication and Computing - ICNCC '16, 2016 Publication	2%
3	dspace.brunel.ac.uk Internet Source	2%
4	filer.case.edu Internet Source	2%
5	Mohammad Anemangely, Ahmad Ramezanzadeh, Behzad Tokhmechi, Abdollah Molaghab, Aram Mohammadian. "Drilling rate prediction from petrophysical logs and mud	1%

logging data using an optimized multilayer perceptron neural network", Journal of Geophysics and Engineering, 2018
Publication

6	Submitted to Regis University Student Paper	1 %
7	iopscience.iop.org Internet Source	1 %
8	Submitted to LaSalle University Student Paper	1 %
9	Submitted to University of Paisley Student Paper	1 %
10	watermark.silverchair.com Internet Source	1 %
11	core.ac.uk Internet Source	1 %
12	Submitted to University of Surrey Student Paper	<1 %
13	"Abstracts of the EASD, Stockholm 2010", Diabetologia, 2010 Publication	<1 %
14	www.ncbi.nlm.nih.gov Internet Source	<1 %
15	docplayer.net Internet Source	<1 %

16	howshealth.com Internet Source	<1 %
17	Rina Ayu Wulan Sari, Ivan Jaya, Marischa Elveny, Sri Melvani Hardi. "Classification of Beef Freshnesh With Color and Texture Feature Extraction Using Learning Vector Quantization Algorithm", 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), 2021 Publication	<1 %
18	www.jatit.org Internet Source	<1 %
19	Submitted to University of Hull Student Paper	<1 %
20	rune.une.edu.au Internet Source	<1 %
21	"Diagnosis and Classification of Diabetes Mellitus", Diabetes Care, 2009 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		