OPEN ACCESS 110

ISSN 2356-5462
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 10, No. 1, Jun 2024. pp. 110-115
doi: doi.org/10.21108/ijoict.v10i1.941

# Web-based Application for Early Diabetes Diagnosis using Learning Vector Quantization (LVQ)

Juni Wijayanti Puspita [1*], Kevin Jieventius Yanto [1], Andi Moh. Ridho Pettalolo [1], Moh. Ali Akbar A. Dg. Matona [1], Lilies Handayani [2, 3]

[1]*Mathematics Department, Tadulako University, Palu, Indonesia*
[2]*Statistics Department, Tadulako University, Palu, Indonesia*
[3]*Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa, Japan*
[*] juni.wpuspita@yahoo.com

**Abstract**

Diabetes is a chronic disease that causes the most deaths in the world. This disease can cause long-term complications that develop gradually, such as heart attacks, strokes, and problems with the kidneys, eyes, skin, and blood vessels. Therefore, early diagnosis of diabetes is crucial for patients to know their diabetes status. In this study, we designed a web-based application for diabetes diagnosis using Learning Vector Quantization (LVQ), which is an artificial neural network algorithm. The dataset from Kaggle's Diabetes Dataset contains eight attributes (pregnancy, glucose, blood pressure, insulin, skin thickness, BMI, diabetes lineage function, and age) and two classes (negative/healthy and positive/diabetes). The results show that the best accuracy is 73.1% with a learning rate of 0.001. These findings can help patients detect diabetes problems early.

**Keywords:** Artificial neural network, Diabetes, Diagnosis, Learning Vector Quantization, Web-based application

## I. INTRODUCTION

Diabetes is a chronic disease that causes the most deaths in the world. In 2021, approximately 537 million people globally were living with diabetes, and this number is expected to reach 643 million in 2030 and 783 million in 2045. In addition, the number of people in prediabetes in 2021 is estimated at around 541 million [2, 3]. Diabetes will appear gradually and not everyone will notice the symptoms in the early stages. Therefore, early diagnosis of diabetes is very important to prevent serious complications through appropriate treatment. Diabetes can be diagnosed with four types of test, namely, fasting plasma glucose test, plasma glucose test after two hours of administration of 75gr oral glucose or tolerance test, HbA1C test, and random plasma glucose test [4].

The adoption of digital technologies in diabetes diagnostics by utilizing the abundance of data available to clinicians and researchers has been developed to revolutionize diabetes management and improve value in healthcare, such as artificial intelligence [5]. One branch of artificial intelligence is neural networks which work based on an architecture that resembles the human brain's neurons. Artificial neural networks can learn from data, generate a rule or operation and make predictions based on the input data. One of the algorithms used in artificial neural networks is Learning Vector Quantization (LVQ).

Received on 31 May 2024. Revised on 26 June 2024. Accepted and Published on 3 July 2024.

Many studies have been carried out to diagnose diabetes using the LVQ classification method. Sulaksono et al. (2013) used LVQ to classify the types of diabetes mellitus (DM), namely DM-type 1, DM-type 2, and negative DM, using 300 training data and 100 testing data. The constructed classification model was able to recognize DM-type 1, DM-type 2, and negative DM with an accuracy of 87.8%, 93.3%, and 76.4%, respectively. DM classification based on the prognosis of DM-type 2 has also been introduced by Aliyanti et al. (2020), with accuracy over 90%. In this study, a dataset consisting of eight attributes will be classified into two classes, namely negative diabetes (healthy) and positive diabetes, using LVQ. Furthermore, a web application based on this classification model was developed to detect diabetes early which is user-friendly and easy to use. Early diagnosis of diabetes is important to allow treatment can be commenced earlier, helping to slow the risk of complications.

## II. LITERATURE REVIEW

Learning Vector Quantization (LVQ) are widely used for classification and diagnosis of diabetes. Here are some noTABLE examples. The weight vector optimization approach using a genetic algorithm (GA) to improve LVQ results in classifying diabetes patients has been introduced in [8]. There are seven attributes consist of 268 people affected by DM and 500 people unaffected by DM taken from Pima Indians Database. The attributes are the number of time pregnant, plasma glucose concentrate, diastolic blood pressure, the thickness of triceps skin folds, body weight, DM genealogical history and age. Their results show GA increase the level of LVQ sensitivity. In [7], detection of DM-type 2 was divided into two stages. The first stage is identifying person with DM-type 2 dan non-DM-type 2. If the data is detected as a person with type 2 DM, the classification process continues to predict the prognostic status, whether the person has metabolic syndrome or not. The accuracy of the classification results for stages 1 and 2 was 96.67% and 92.5%, respectively. Putri et al. (2019) proposed diabetes classification with Chi-Square for feature selection using Kaggle database. Using LVQ, the study achieved the highest accuracy on training data of 80% and 90%. Using the same database as [9], Ster and Dobnikar (1996) obtained an accuracy of 75.80%.

## III. RESEARCH METHOD

This study consists of two main stages: first, constructing a diabetes classification model; and second, developing a web-based application to determine whether a patient has diabetes. Here, the diabetes dataset was taken from the Kaggle dataset website [1]. This dataset has 796 data consisting of eight attributes, namely number of times pregnant (pregnancies), plasma glucose concentration a 2-hours in an oral glucose tolerance test (glucose), diastolic blood pressure (blood pressure), triceps skin fold thickness (skin thickness), 2-hour serum insulin (insulin), body mass index (BMI), diabetes pedigree function, and age.

The research flow diagram in the first stage can be seen in Fig. 1. Several steps taken in data preprocessing are as follows:

1. Data transformation by changing categorical data into numerical data.
2. Balancing the dataset, because the amount of data for diabetes suffered and healthy is not balanced.
3. Normalize data using the formula in Eq. (1).

$$v' = \frac{v - min(v)}{max(v) - min(v)} \tag{1}$$

In this study, classification prediction will be done using a Learning Vector Quantization (LVQ) algorithm to classify negative diabetes (healthy) and positive diabetes. Several scenarios are carried out by dividing the percentage of the dataset into training data and test data.

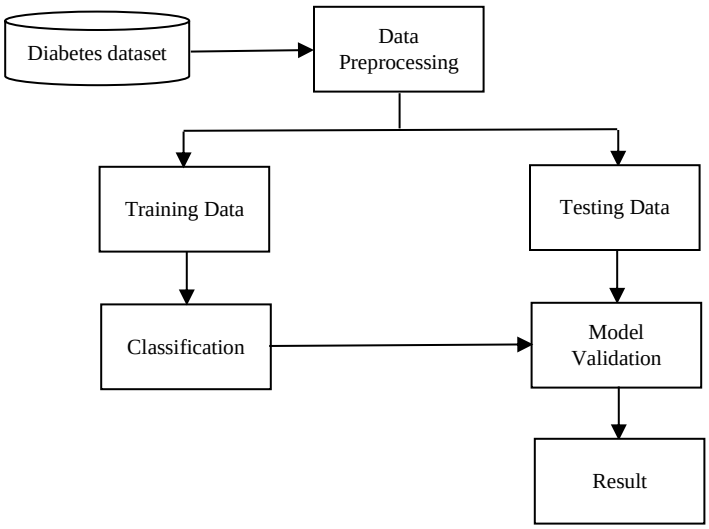



Fig. 1. Flow diagram for diabetes classification using LVQ.

Learning Vector Quantization (LVQ) is an artificial neural network algorithm with a supervised learning method [11]. The LVQ classifier technique uses training data with the desired information class to classify the data [12]. The LVQ algorithm in the training stage is given in [13]. The flowchart of LVQ is shown in Fig. 2.

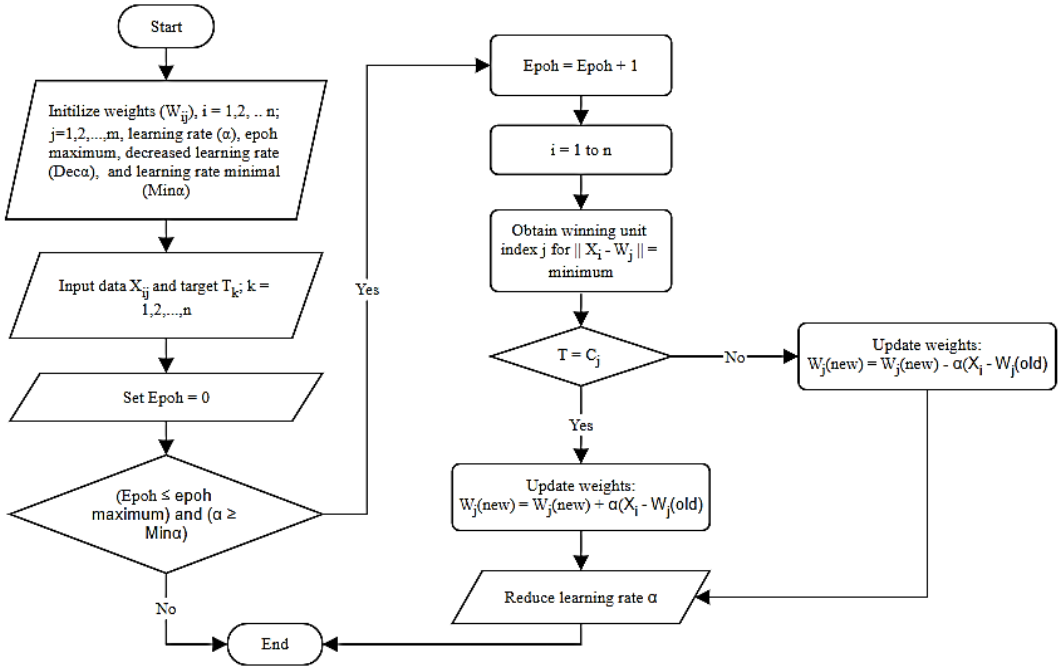



Fig. 2. Flowchart of LVQ algorithm.

In the second stage, the application of diabetes based on website is created using Visual Studio Code (VSC) software with Java script, CSS, and HTML as programming languages. Visual Studio Code is an open sources

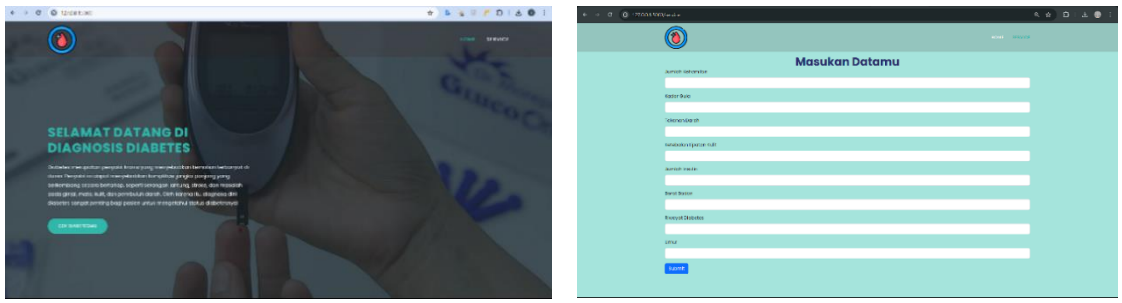
application for editing code designed by Microsoft and compatible with Windows, Limec and MacOS operating systems. This application provides convenience in writing code for various types of programming, such as C++, C, Java, Python, PHP, GO. VSC also has a feature that allows automatic recognition of programming languages and provides different color highlighting according to the function in the code.

## IV. RESULTS AND DISCUSSION

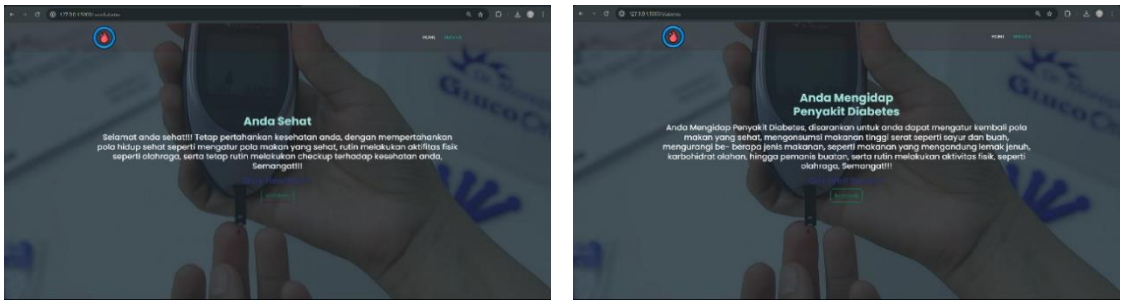
In this section, we present the performance of Learning Vector Quantization (LVQ) over the datasets. Classification results with several scenarios for dividing training and testing data can be seen in TABLE 1. The highest accuracy, that is, 73.1%, can be achieved by using ratio training data 80% and testing data 20% with a learning rate $\alpha = 0.001$. Here, adding more training data tends to improve accuracy. The simulation results show that the smaller the learning rate value, the better the accuracy.

The findings in TABLE 1 indicate that different learning rates do significantly affect the performance of the classification model. However, several studies have discussed that learning rate is critical for achieving good performance in neural network optimization [14-17]. The work of Wilson & Martinez [17] has even discussed how to efficiently select a learning rate that maximizes generalization accuracy and how to decide when the learning rate is small enough to produce maximum generalization accuracy.

Furthermore, the trained LVQ model is used in the web-based application. Here, we designed the user interface of a web-based application for diabetes diagnosis with an easy-to-use interface, as shown in Fig. 3. Users only need to fill in the available input data. This application takes information from eight attributes as its main input. By pressing the "submit" button, the user will get an output that diagnoses whether the user has diabetes or non-diabetes (healthy).




(a)




(b)

Fig. 3. Web-based application for diabetes diagnostic measurements: (a) user interface for providing input and (b) user interface showing output.

TABLE I
CLASSIFICATION RESULTS.

| Training-Testing Ratio (%) | Accuracy (%) | | | | |
|---|---|---|---|---|---|
| | $\alpha = 0.001$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.1$ | $\alpha = 0.5$ |
| 20:80 | 70.1 | 65.5 | 68.1 | 69.5 | 65.5 |
| 20.1:79.9 | 70.3 | 67.7 | 67.7 | 70.3 | 65.8 |
| 30:70 | 70.4 | 67.3 | 67.3 | 67.1 | 62.1 |
| 40:60 | 68.7 | 68.4 | 64.8 | 64.9 | 61.8 |
| 50:50 | 70 | 69.5 | 65.3 | 66.9 | 62.2 |
| 60:40 | 72.2 | 69.6 | 65 | 65.9 | 62 |
| 70:30 | 71.7 | 70.8 | 67.6 | 67.5 | 64.8 |
| 80:20 | 73.1 | 72 | 68.9 | 68.6 | 65.3 |
| 90:10 | 72.8 | 71.6 | 68.3 | 66.8 | 65.7 |

## V. CONCLUSION

This work is done to develop diabetes diagnostic measurement by using Learning Vector Quantization (LVQ) algorithm. We obtained a recognition rate of 73.1% with a learning rate of 0.001. Based on the trained LVQ model's, a web-based application is designed, in a user-friendly way, to make diagnosis whether the user is healthy or diabetes. Early diagnosis of diabetes allows a person to receive effective treatment, as the disease may be in its initial stages, thereby delaying disease progression and reducing the risk of complications.

## REFERENCES

[1] AEMYJUTT, and willian. (2023). diabetesDataAnslysis [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/3789048

[2] S. Damtie *et al.*, "The magnitude of undiagnosed diabetes mellitus, prediabetes, and associated factors among adults living in Debre Tabor town, northcentral Ethiopia: A community-based cross-sectional study," *Heliyon*, vol. 9, no. 7, 2023, doi: 10.1016/j.heliyon.2023.e17729.

[3] H. Sun *et al.*, "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Res Clin Pract*, vol. 183, Jan. 2022, doi: 10.1016/j.diabres.2021.109119.

[4] D. Hardianto, "Telaah Komprehensif Diabetes Melitus: Klasifikasi, Gejala, Diagnosis, Pencegahan, dan Pengobatan: A Comprehensive Review of Diabetes Mellitus: Classification, Symptoms, Diagnosis, Prevention, and Treatment," *Bioteknologi &Amp; Biosains Indonesia (JBBI)*, vol. 7, no. 2, 2021.

[5] A. N. Klonoff, W. A. Lee, N. Y. Xu, K. T. Nguyen, A. DuBord, and D. Kerr, "Six Digital Health Technologies That Will Transform Diabetes," *Journal of Diabetes Science and Technology*, vol. 17, no. 1. 2023. doi: 10.1177/19322968211043498.

[6] J. Sulaksono, Moch. H. Jauhari, and F. R. Hariri, "Sistem Pendukung Keputusan Penentuan Penyakit Diabetes Mellitus menggunakan Metode Learning Vector Quantization," *Semnasteknomedia Online*, vol. 2, no. 1, 2014.

[7] N. Aliyanti, R. Ratianingsih, and J. W. Puspita, "Sistem Pendukung Keputusan Untuk Mendeteksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Metode Learning Vector Quantization (LVQ)," *Jurnal Ilmiah Matematika Dan Terapan*, vol. 17, no. 2, 2020, doi: 10.22487/2540766x.2020.v17.i2.15336.

[8] I. Permana, N. E. Rozanda, F. Syafria, and F. N. Salisah, "Optimization learning vector quantization using genetic algorithm for detection of diabetics," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, 2018, doi: 10.11591/ijeecs.v12.i3.pp1111-1116.

[9] N. K. Putri, Z. Rustam, and D. Sarwinda, "Learning Vector Quantization for Diabetes Data Classification with Chi-Square Feature Selection," in *IOP Conference Series: Materials Science and Engineering*, 2019. doi: 10.1088/1757-899X/546/5/052059.

[10] B Ster and A Dobnikar, "Neural Networks in Medical Diagnosis: Comparison with Other Methods," *Proceedings of the International Conference EANN96*, vol. 1, no. September 2015, 1996.

[11] A. Sato and K. Yamada, "Generalized Learning Vector Quantization," in *NIPS 1995: Proceedings of the 8th International Conference on Neural Information Processing Systems*, 1995.

[12] O. Abualghanam, O. Adwan, M. A. Al Shariah, and M. Qatawneh, "Enhancing the Speed of the Learning Vector Quantization (LVQ) Algorithm by Adding Partial Distance Computation," *Cybernetics and Information Technologies*, vol. 22, no. 2, 2022, doi: 10.2478/cait-2022-0015.

[13] R. Devita, R. H. Zain, H. Syahputra, E. Afri, and I. Maulina, "Implementation and Development of Learning Vector Quantization Supervised Neural Network," in *Journal of Physics: Conference Series*, 2022. doi: 10.1088/1742-6596/2394/1/012009.

[14] Y. Li, C. Wei, and T. Ma, "Towards explaining the regularization effect of initial large learning rate in training neural networks," in *Advances in Neural Information Processing Systems*, 2019.

[15] L. Blier, P. Wolinski, and Y. Ollivier, "Learning with Random Learning Rates," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020. doi: 10.1007/978-3-030-46147-8_27.

[16] M. Fadli, M. Ifan, R. Cipta, S. Hariyono, and N. M. Saraswati, "Penerapan Metode Learning Vector Quantization (LVQ) untuk Menentukan Irigasi Lahan Pertanian di Desa Penggarutan," 2021.

[17] D. R. Wilson and T. R. Martinez, "The need for small learning rates on large problems," in *Proceedings of the International Joint Conference on Neural Networks*, 2001. doi: 10.1109/ijcnn.2001.939002.