Jurnal Mantik
Volume 4, Number 3, Nov 2020, pp. 1772-1778 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

LVQ Algorithm for The Classification of Hypertension Based on ESH Guideline

Elvandric Lase[1], Wonderson[2], Christensen[3], Dinda Afrianti[4]
Andika Dian Permana[5], Abdi Dharma[6]

[123456]Computer Science Department, Faculty of Technology and Computer Science
[123456]Universitas Prima Indonesia, Medan, Indonesia

E-mail: abdidharma@unprimdn.ac.id[6]

ARTICLE INFO

Article history:
Received: 12/07/2020
Revised: 22/08/2020
Accepted: 30/09/2020

Keywords:
Hypertension, ESH Guideline, Learning Vector Quantization, Machine Learning, Accuracy

ABSTRACT

Hypertension was a global health problem, including Indonesia, that increases mortality, morbidity, and cost. In Indonesia, hypertension kept on increasing due to change in lifestyle, consumptions of food with a high level of fat, cholesterol, less physical activity, and a high level of stress, etc. One of the classifications of hypertension used in some country were European Society of Hypertension (ESH) guideline. Learning Vector Quantization (LVQ) was a method in machine learning for classifying data. LVQ were often used in pattern recognition processes such as images, sounds, etc. The purpose of this study was to see an increase in accuracy of hypertension classification based on ESH guideline as weight data. In this study, hypertension classification based on ESH guideline was used as weight data with LVQ method and the parameters used were 2 features, 100 epochs, 0.05 learning rate, 0.01 reducing factor, train data 70%, validation data 30%, and test data 30% from total data used. The result obtained in this study were 94.6667% in the hypertension classification process based on ESH guideline using LVQ method. The conclusion of this study, there was an increase in the accuracy of hypertension classification based on ESH guideline using the LVQ method.

1. Introduction

Blood pressure diseases is one of the main factors that threatening human health [1]. And as we know, high blood pressure could be referred to as hypertension, which is a global health problem that increases mortality, morbidity, and cost, including in Indonesia. Hypertension is also causing damage to essential organs such as the brain, heart, kidneys, retina, aortic blood vessels, and peripheral blood vessels. According to Basic Health Research (Riskesdas), in 2018, Indonesia's hypertension has been increased by 34.1%, with a population size of 260 million compared to 2013 [2].

In Indonesia, hypertension keeps increasing due to lifestyle changes, food consumption with a high level of fat, cholesterol, less physical activity, high stress levels, and more [3][4]. Hypertension is a form of cardiovascular disease which can be diagnosed positive when the systolic blood pressure ≥ 140 mmHg and/or the diastolic blood pressure ≥ 90 mmHg when measured at the clinic or health facility [2][5]. Usually, hypertension does not have any symptoms like other diseases, so it is challenging to detect hypertension. Patients usually do not know that they have hypertension, they notice it when there is an association with other conditions such as diabetes or stroke, so this disease is often called the silent killer [2][3].

There are several classifications of hypertension types used in each country, The Seventh Joint National Committee (JNC7) on Prevention, Detection, Evaluation and Treatment of High Blood Pressure is used generally, and this study used one of the hypertension classification that is European Society of Hypertension (ESH) guideline because this guide is used by Indonesian Society of Hypertension in 2019 Hypertension Management Consensus [2]. Classification of blood pressure for ESH guideline is divided into 7 types, either for the systolic or diastolic blood pressure, and shown in Table 1 [2].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Table 1.
Classification of Blood Pressure Based on ESH Guideline

Category	Systolic Blood Pressure (mmHg)		Diastolic Blood Pressure (mmHg)
Optimal	< 120	and	< 80
Normal	120 – 129	and/or	80 – 84
Normal-high	130 – 139	and/or	85 – 89
Hypertension grade-1	140 – 159	and/or	90 – 99
Hypertension grade-2	160 – 179	and/or	100 – 109
Hypertension grade-3	≥ 180	and/or	≥ 110
Isolated systolic hypertension	≥ 140	and	< 90

LVQ is a single layer network that includes an input and output layer where there is a value between layers. It was introduced by Kohonen in 1982[3]. LVQ is an algorithm that is carried out at the competitive layer by supervised training. Usually, the competitive layer of LVQ will learn automatically in classifying the input data. If several vectors are close together, the input data will be grouped in the same class [6][7]. LVQ is different from Self-Organizing Map because LVQ classifies data using data that has been decided for learning in the supervised layer [8].

LVQ is often used for feature extraction at the beginning of pattern recognition processes such as images, sounds, and many more. In the learning process, the value will be arranged in a range depending on the input value. The learning objective is to make a group with a similar unit in one place, so it is very suitable for pattern classifications [9]. While in the learning process, a vector will be inserted based on the output unit that has the smallest range between the weight vector and the input vector [10]. The LVQ network architecture shown in Figure 1 [3][11].

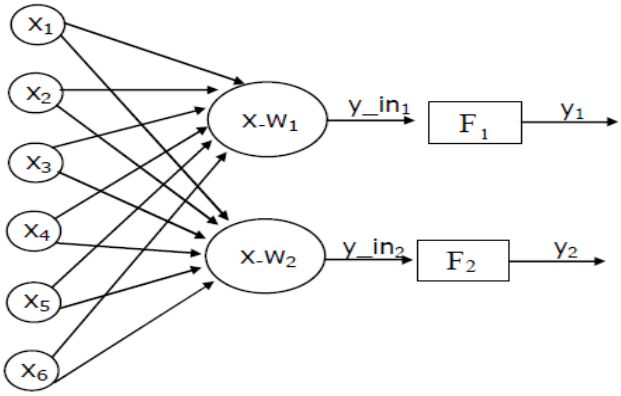

Fig 1. LVQ network architecture

Explanation:
$X_1 \ldots X_6$ = Input/input data
$X.W_1$ = Value that connecting each neuron to the input layer for the first neuron in output layer
$X.W_2$ = Value that connecting each neuron to the input layer for the first neuron in output layer
F_1, F_2 = Function of linear transfer used as an artificial neural network is trained using LVQ and has a value range from 0 to 1
Y = Output/output value

Compared to others algorithms in classification LVQ can increase the category numbers that caused by a change in the model, this algorithms is very flexible when there is a changes in the number of the categories so it is relatively more simple and easy to adjust [12]. LVQ network does not have an outer layer, that maps the typical Artificial Neural Network (ANN) response for the output layer. LVQ operate differently and work to move the nodes, for representing the underlying data through an iterative training [13]. One of the advantages of using LVQ is the ability that is giving the training to the competitive layers so that it can classify the given input automatically [14].

Previous study using LVQ method to identify eyes diseases gain the average accuracy of 82.80% [15]. Another study using LVQ method for classify the status of the volcano obtained the average accuracy of 88% [10]. The other study that use LVQ method to classify the quality of river obtain the average accuracy of 81.13% [14]. In a previous study for classify hypertension types based on JNC7 using LVQ gain the accuracy 93.841% [3]. So in this study the method is still using LVQ to classify the hypertension types based on ESH

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

guideline, things that differentiate it from previous study is the hypertension classification based on JNC7 as weight data. One of the differences between JNC7 and ESH guideline is the classification of the hypertension types where JNC7 has 4 types of hypertension: normal, pre-hypertension, hypertension grade-1, and hypertension grade-2 [16][17]. And ESH guideline has 7 types of hypertension [2]. This study purpose was to see if there was an increase in accuracy of hypertension classification based on ESH guideline as weight data.

2. Research Methodology

The algorithm used in this study is LVQ, hypertension classification based on ESH guideline was used as weight data, and the parameters that used were 2 features, 100 epochs, learning rate 0.05, reducing factor 0.01, training data 70%, validation data 30% and testing data 30% from the total data.

The methodology is consisting of data preprocessing and machine learning modelling. It will be visualized in Figure 2 as the flowchart in this study. Furthermore, this study is conducted with the Python programming language.

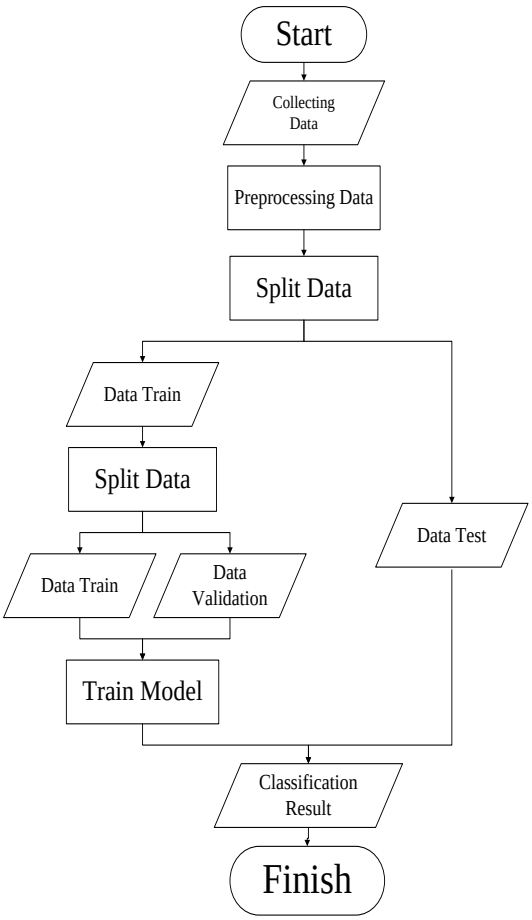

Fig 2. The flowchart in this study

2.1 Preprocessing

This study uses one public dataset obtained from Kaggle, the data used in this study are 250 data from 70000 data that have been provided, and two features used out of 12 features. The dataset shared by Svetlana Ulianova about cardiovascular disease and retrieved from https://www.kaggle.com/sulianova/cardiovascular-disease-dataset

Preprocessing data is a process that is converting the raw data to useable data in the machine learning model [18]. Preprocessing is done before the data inserting it into the algorithm. This process has been carried out on the dataset to be researched and imported to Jupyter Notebook.

Dataset normalization is done manually because the features used in this study are 2 features, and the given features from the dataset are 12, so it must be normalized first. This study used 2 features because the weight data used in this study only require data about systolic and diastolic blood pressure. Also, adding

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

seven labels because the classification uses weight data from hypertension classification based on ESH guideline.

2.2 Machine Learning Modelling

After the data that have been processed, then the next is machine learning modelling, starting with split the data. Train data and test data will be divided by splitting the data, where 70% is the train data, and 30% is the test data. After dividing the train data and test data, the data will be divided once more that produce train data and validation data. Then insert the weight data as the comparison for the train data, validation data, and the test data. After that insert, the weight data and then the algorithm will calculate the distance using Euclidean distance that can be formulated as [3][15]:

$$D = \sqrt{\sum_{i=1}^{n}(qi - pi)^2} \quad (1)$$

Euclidean distance is the most used algorithms in calculating distance in machine learning classifications such as LVQ or K-means algorithm. Euclidean distance can be explained as a distance between data points that belong to a set or two sets of data points [19]. And to get the accuracy can be formulated as:

$$Accuracy = \frac{the\ amount\ of\ correct\ data}{the\ amount\ of\ used\ data} \times 100\% \quad (2)$$

After the distance calculation process is complete, the next step is to use cross-validation to find the result from the train data, validation data, and test data. Cross-validation is a method that resampling data that has an ability to generalization for predicting the model and prevents overfitting [20]. To find out the model that has been created is a good model, a graph is made to see the accuracy of the train data and test data. After the results come out, to know how accurate the data with this modelling, a confusion matrix is created, which is a package from pyplot.

3. Results and Discussion

To find the best parameters to produce better accuracy, testing the epoch and the learning rate is needed. And the parameters for testing in this study are the learning rate, epoch, train data, validation data, test data, and reducing factor of 0.01.

3.1 Testing

In this process, parameters such as epoch, learning rate, and reducing factor will be tested first on the training data and validation data. The accuracy obtained from this test is 94.2623% for the training data and 96.2264% for validation data. To determining whether the model is a good model, a graph is made to show the accuracy obtained from all the epochs on the training data and the validation data, as in Figure 3. The graph below shows that the model is a good one because there is no overfit or underfit in the model.

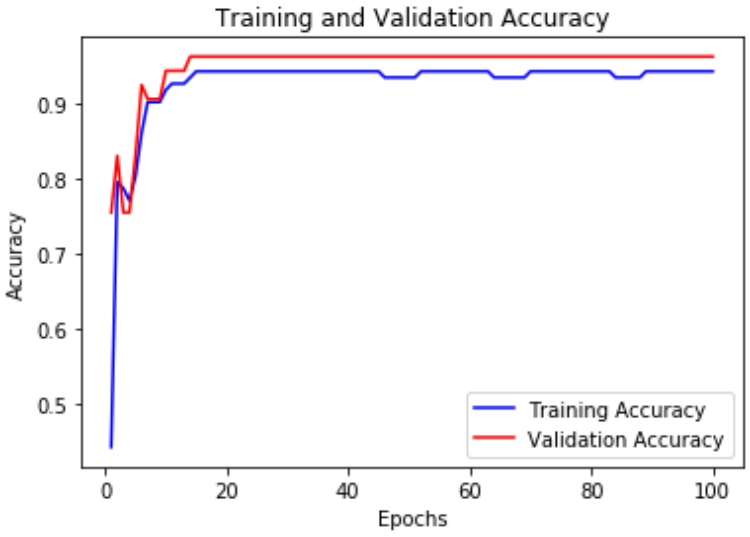

Fig 3. Graph accuracy on training data and validation data

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

3.2 Model Testing on The Test Data

This test aims to find the final result of the model's accuracy using the parameters obtained from the test on the training data and the validation data, namely 100 epochs, 0.05 learning rate, and 0.01 reducing factor. And the accuracy results obtained in this test are 94.6667%.

3.3 Confusion Matrix

A table that aims at describing the classification model in the test data to determine the amount of the correct and incorrect data is the confusion matrix. Confusion matrix is the most classical decision-measure methods in supervised machine learning [21]. Figure 4 is an example of the confusion matrix. From the example, we can understand how to read a confusion matrix. And Figure 5 is the confusion matrix of the model.

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Fig 4. Confusion matrix example

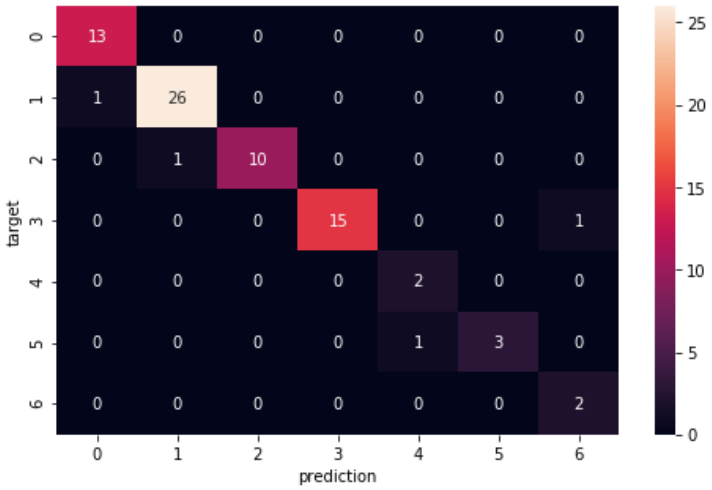

Fig 5. Model confusion matrix

In the confusion matrix, there are only four incorrect classifications on the model from the 75 data used in the test data.

3.4 Discussion

The average accuracy obtained from this study is 94.6667% with 2 features, 100 epochs, 0.05 learning rate, 0.01 reducing factor, and 250 data where this is an improvement from the previous study.

In previous studies using the same method, the average accuracy is 93.841%, with the amount of data 100, 6 epochs, 0.1 learning rate, 0.001 reducing factor, and 50% train data [3]. The study using the Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3) method, the accuracy was 80% because of the Mamdani Inference was used as the classification [5]. In a study using the Naïve Bayes method, the accuracy was 83.67%, and the decision tree is 77.55% this was because the data used in this study contains information of 52 patients along with their hypertension diagnosis [22]. While in a study using an ANN, the accuracy was 82% this was due to the usage of the classical ANN alongside with the real time predictive systems [23]. Study that using Decision Tree method obtain the accuracy of 92.6573% while having error 7.3427% this

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

was because of the decision tree making rules in the data mining algorithms that are C4. 5 for creating the category of hypertension factors [24].

LVQ is less efficient when a lot of data was used. After all, it only has one hidden layer, and also, the way these algorithms works is to collect the data first then calculate the distance between data, which results in the accuracy [25].

4. Conclusion

This study concludes, there was an increase in the accuracy of classifying hypertension based on ESH guideline using LVQ model. The more data we insert into the model, the less efficient it will be because of one hidden layer. And the accuracy depends on the initialization of the model and the parameters used, and the data distribution of each class in the training data.

For future study there are various things that can be considered: 1) add more features to classify hypertension, such as age, blood sugar, cholesterol, etc.; 2) add a new algorithm to classify hypertension to increase the accuracy, such as Random Forest, K-Nearest Neighbor, and Logistic Regression.

5. References

[1] B. Zhang, H. Ren, G. Huang, Y. Cheng, and C. Hu, "Predicting blood pressure from physiological index data using the SVR algorithm," *BMC Bioinformatics*, vol. 20, no. 1, p. 109, Dec. 2019.

[2] A. A. Lukito, E. Harmeiwaty, and N. M. Hustrini, *Konsensus Penatalaksanaan Hipertensi 2019*, vol. 36, no. 6. 2019.

[3] I. Agustinus, E. Santoso, and B. Rahayudi, "Klasifikasi Risiko Hipertensi Menggunakan Metode Learning Vector Quantization (LVQ)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 8, pp. 2947–2955, 2018.

[4] S. Sakr *et al.*, "Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project," *PLoS One*, vol. 13, no. 4, p. e0195344, Apr. 2018.

[5] M. R. Andriansyah, E. Santoso, and Sutrisno, "Klasifikasi Risiko Hipertensi Menggunakan Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. Vol 2 No 12 (2018), pp. 7088–7096, 2018.

[6] A. W. Rahmadani, A. I. Jaya, and N. Nacong, "PREDIKSI PENYAKIT TUBERCULOSIS PARU (TB PARU) MENGGUNAKAN METODE LEARNING VEKTOR QUANTIZATION (LVQ)," *J. Ilm. Mat. DAN Terap.*, vol. 15, no. 1, pp. 20–27, May 2018.

[7] W. A. Setyowati and W. F. Mahmudy, "Optimasi Vektor Bobot Pada Learning Vector Quantization Menggunakan Particle Swarm Optimization Untuk Klasifikasi Jenis Attention Deficit," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. May, pp. 4428–4437, 2018.

[8] Z. A. Leleury, Y. A. Lesnussa, and J. Madiuw, "Sistem Diagnosa Penyakit Dalam dengan Menggunakan Jaringan Saraf Tiruan Metode Backpropagation dan Learning Vector Quantization," *J. Mat. Integr.*, vol. 12, no. 2, p. 89, Jul. 2017.

[9] S. Agustin, B. D. Setiawan, and M. A. Fauzi, "Klasifikasi Berat Badan Lahir Rendah (BBLR) Pada Bayi Dengan Metode Learning Vector Quantization (LVQ)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, no. 3, p. 964X, 2018.

[10] C. F. Virkhansa, B. D. Setiawan, and C. Dewi, "Klasifikasi Status Gunung Berapi dengan Metode Learning Vector Quantization (LVQ)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 7119–7126, 2019.

[11] S. Ramzini, D. E. Ratnawati, and S. Anam, "Penerapan Metode Learning Vector Quantization (LVQ) untuk Klasifikasi Fungsi Senyawa Aktif Menggunakan Notasi Simplified Molecular Input Line System (SMILES)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 12, pp. 6160–6168, 2018.

[12] J.-H. Wu *et al.*, "Risk Assessment of Hypertension in Steel Workers Based on LVQ and Fisher-SVM Deep Excavation," *IEEE Access*, vol. 7, pp. 23109–23119, 2019.

[13] T. J. Bihl, T. J. Paciencia, K. W. Bauer, and M. A. Temple, "Cyber-Physical Security with RF Fingerprint Classification through Distance Measure Extensions of Generalized Relevance Learning Vector Quantization," *Secur. Commun. Networks*, vol. 2020, no. 1, pp. 1–12, Feb. 2020.

[14] R. Hamidi, M. T. Furqon, and B. Rahayudi, "Implementasi Learning Vector Quantization (LVQ) untuk Klasifikasi Kualitas Air Sungai," *J-Ptiik*, vol. 1, no. 12, pp. 1758–1763, 2017.

[15] E. B. Ladauw, D. E. Ratnawati, and A. A. Supianto, "Identifikasi Penyakit Mata Menggunakan Metode Learning Vector Quantization (LVQ)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6989–6996, 2018.

[16] A. C. Joseph, M. S. Karthik, R. Sivasakthi, R. Venkatanarayanan, and J. Sam Johnson Udaya Chander, "JNC 8 versus JNC 7 – Understanding the evidences," *Int. J. Pharm. Sci. Rev. Res.*, vol. 36, no. 1, pp. 38–43, 2016.

[17] D. Irawan, I. Muhimmah, and T. Yuwono, "PROTOTYPE SMART INSTRUMENT UNTUK KLASIFIKASI PENYAKIT HIPERTENSI BERDASARKAN JNC-7," *J. Teknol. Inf. dan Terap.*, vol. 4, no. 2, pp. 111–118, Apr. 2019.

[18] N. Azwanti and E. Elisa, "Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4.5," *InfoTekJar (Jurnal*

Nas. Inform. dan Teknol. Jaringan), vol. 3, no. 2, pp. 116–123, Feb. 2019.
[19] T. Rechkalov and M. Zymbler, "A Study of Euclidean Distance Matrix Computation on Intel Many-Core Processors," in *Communications in Computer and Information Science*, vol. 910, no. August, 2018, pp. 200–215.
[20] D. Berrar, "Cross-Validation," in *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1–3, no. April, Elsevier, 2019, pp. 542–545.
[21] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny).*, vol. 507, no. July, pp. 772–794, Jan. 2020.
[22] B. Afeni, T. Aruleba, and I. Oloyede, "Hypertension Prediction System Using Naive Bayes Classifier," *J. Adv. Math. Comput. Sci.*, vol. 24, no. 2, pp. 1–11, Jan. 2017.
[23] D. LaFreniere, F. Zulkernine, D. Barber, and K. Martin, "Using machine learning to predict hypertension from a clinical dataset," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1–7.
[24] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, Jun. 2016.
[25] E. B. Budianita, "Penerapan Metode Learning Vector Quantization2 (LVQ 2) Untuk Menentukan Gangguan Kehamilan Trimester I," *J. Sains dan Teknol. Ind.*, vol. 15, no. 2, p. 144, Jun. 2018.