

Cataract Disease Diagnosis System Using Artificial Neural Network *Learning Vector Quantization* Method

Chrisani Waas¹, D. L. Rahakbauw², Y. A. Lesnussa*³

Mathematics Department University of Pattimura, Jln. Ir. M. Putuhena, Poka-Ambon, 97233, Indonesia

E-mail : chrisaniwaas@gmail.com¹, lodewyik@gmail.com², yopi_a_lesnussa@yahoo.com*³

*Corresponding author

Abstract - Artificial Neural Network (ANN) is an information processing system that has certain performance characteristics that are artificial representatives based on human neural networks. ANN method has been widely applied to help human performance, one of which is health. In this research, ANN will be used to diagnose cataracts, especially Congenital Cataracts, Juvenile Cataracts, Senile Cataracts and Traumatic Cataracts based on the symptoms of the disease. The ANN method used is the Learning Vector Quantization (LVQ) method. The data used in this research were 146 data taken from the medical record data of RSUD Dr. M. Haulussy, Ambon. The data consists of 109 data as training data and 37 data as testing data. By using learning rate (α) = 0.1, decrease in learning rate (dec α) = 0.0001 and maximum epoch (max epoch) = 5, the accuracy rate obtained is 100%.

Keywords - Artificial Neural Network, Learning Vector Quantization, Cataracts Disease

1. INTRODUCTION

The eye is a sense of the human body that is constantly adjusting the amount of light entering it, focusing attention on objects near and far, and producing continuous pictures delivered instantly to the brain[1, 2]. When the eye is bothered, the cost for checking and treating is not cheap, moreover the availability of ophthalmologists that are not always available during urgent needs of treatment[3]. Eye diseases can be caused by germs, bacteria, viruses, body organ failures, and also genetically inherited. There are a lot of eye diseases threatening humans including glaucoma, conjunctivitis, keratitis, dry eyes, and cataracts[4]. A cataract is one of the causes of sight disruption where eye lenses that are usually clear are clouded[5, 6]. According to 2001 Survei Kesehatan Rumah Tangga (SKRT) data, cataract is a sight disruption without pain symptoms. Blindness by cataracts are mostly caused by the inability of performing proper surgery and the rest of the cataract sufferers did not know that blindness can be prevented[7, 8].

Doctors often hesitate and even misdiagnose types of cataracts, so to help ensure the validity of the diagnosis taken by a doctor, a system that can identify cataracts can be made. The system is based on artificial intelligence technology that is similar to human intelligence, which can even be better than humans[9]. One of the artificial intelligence technologies created by humans based on the knowledge they learned is Artificial Neural Networks (ANN). ANN is an information processing system that has certain performance characteristics that are artificial representations based on human neural networks[10, 11]. ANN has several methods that have been widely applied to help human performance, one of which is health or medical[12, 13].

Learning Vector Quantization (LVQ) is one of the methods of artificial intelligence in Artificial Neural Networks, where there is a weight vector representing a unit of output to be

used as a reference for the class representing the output by calculating the closest distance between data[14, 15]. The choice of the LVQ method on this problem is because this method depends on several parameters that support the process of identifying cataracts, including the learning rate (α), decrease in learning rate ($\text{Dec}\alpha$), minimum learning rate (α), and maximum epoch (Max Epoch). The purpose of this method is to approximate the distribution of class vectors to minimize errors when classifying[16]. The LVQ network is successfully used in the classification of Diabetic Retinopathy Disease to obtain diagnosis results[17] and also in identifying Dramatic Personality Disorders[18]. So by using LVQ, it is expected to be able to train input data for the process significantly to quickly summarize or reduce large data sets for a small number of vectors[16].

The purpose of this paper is to diagnose the type of cataract based on the symptoms felt by the patient optimally using a decision-making system by applying the artificial neural network LVQ method, to determine the type of treatment as soon as possible. It is expected that this system can further support doctor's diagnosis of cataracts based on the symptoms suffered by the patient.

2. RESEARCH METHOD

2.1. Research Design

The steps in this research design include:

- a. Collecting data.
In this research, the data used are medical records of patients with Cataract Disease (Congenital Cataracts, Juvenile Cataracts, Senile Cataracts, and Traumatic Cataracts) namely data symptoms experienced by each patient and the type of illness suffered based on doctor's diagnosis. The data used are secondary data obtained from 146 medical records of patients with cataracts in RSUD Dr. M. Haulussy Ambon. The data is then shared to conduct training programs as training data (75%) and test data (25%) for the system.
- b. Designing and create a diagnostic system using MATLAB software.
After the data related to the symptoms of each cataract disease is obtained, the next step is to design and create a diagnostic system to recognize the symptoms of existing cataracts. In making the system, the steps taken start from making program design using Graphical User Interface Design (GUIDe), then proceed with making MATLAB coding so that program design can function.
- c. Performing system testing.
In this step, testing is done on the system that has been designed. System testing is done by calculating the level of accuracy in diagnosing the type of disease in the test data.
- d. Drawing conclusions.
Conclusions are based on the accuracy of the system designed with the doctor's diagnosis.

2.2. Data Analysis Method

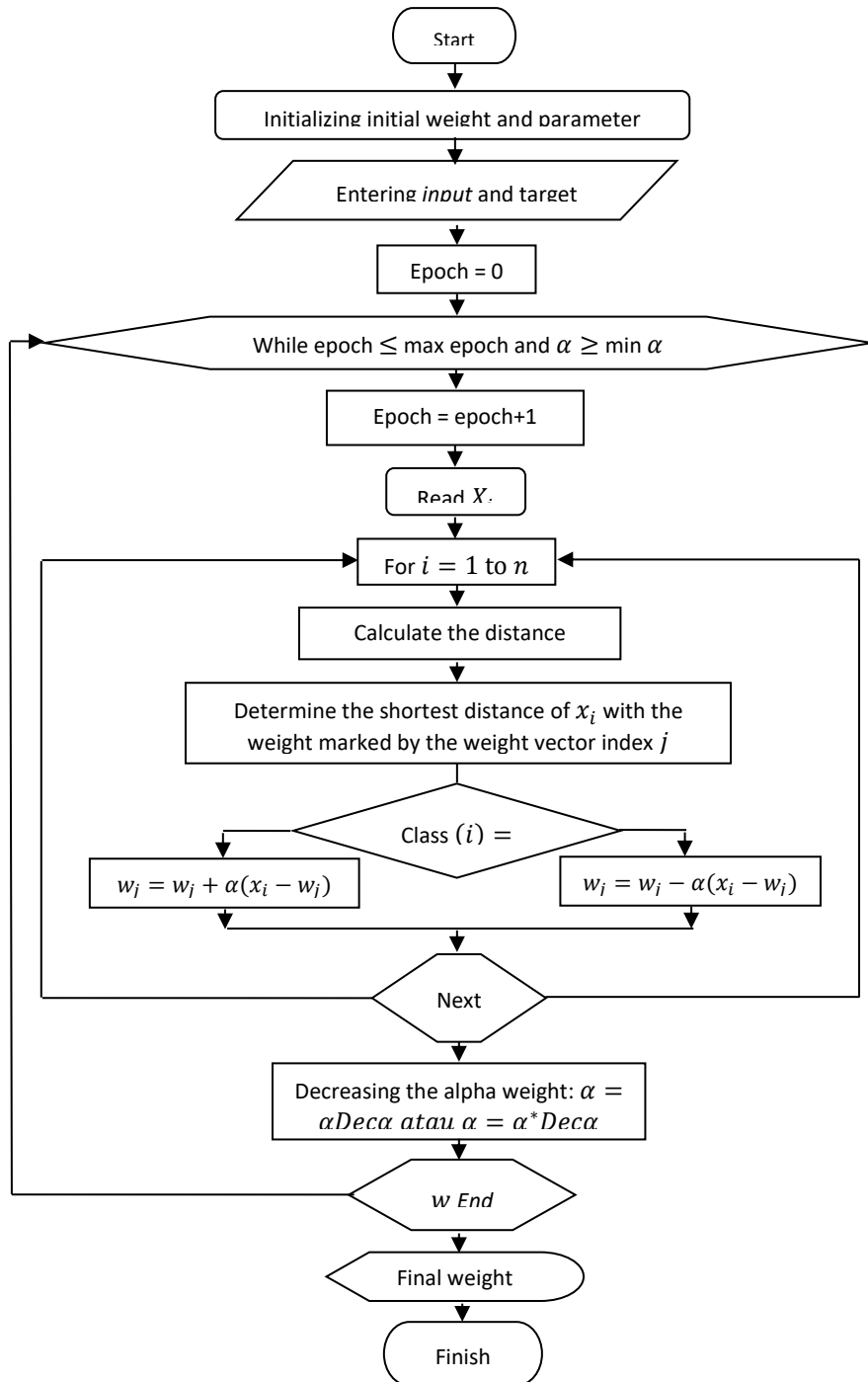


Figure 1. Flowchart of the training process on the LVQ network

The methods used to analyze the data in this research are based on the algorithm of the research process of the LVQ method shown in figure 1 above. After training is completed, final weights are obtained. These weights are then used further for testing. The testing process algorithm is shown in the following figure.

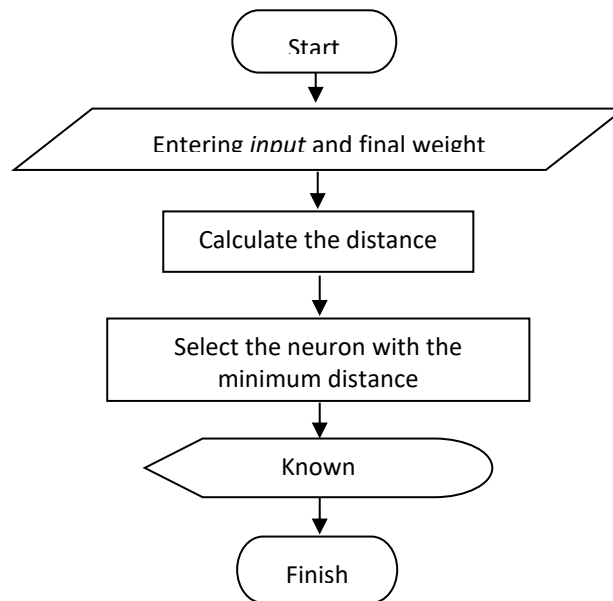


Figure 2. Flowchart of the testing process on the LVQ network

3. RESULTS AND DISCUSSION

3.1 Setting Input Variables

Inputs used in the form of symptoms that cause cataracts. Where these symptoms are used to diagnose cataract based on the data that the patient has obtained. Based on clinical symptoms, namely cataract diseases Congenital Cataract, Juvenile Cataract, Senile Cataract, and Traumatic Cataract has 13 symptoms which then become input variables in the Artificial Neural Network. These symptoms are White/Grayish Pupil, Age under 1 year old, Declining Sharpness, Double Visions, Itchy Eyes, Blurred Vision, Red Eyes, Sensitivity towards Light, Age between 1 to 50 years old, Age over 50 years old, Contrasting Sensitivity, Myopia Translation, Monocular Diplopia.

The symptoms of a cataract are determined, next is to determine the value of each symptom. The value of the variable is determined by 0 and 1, in this research it is determined that if the value of the variable 1, the disease is getting worse. Conversely, if the value of a variable 0, the disease is getting milder. The variables and values of each disease symptom are presented in Table 1.

3.2 Determination of Output Variables

Furthermore, the output to be obtained is congenital, juvenile, senile and traumatic cataracts. The design of output as follows:

- "1" is given for patients diagnosed with Congenital Cataracts.
- "2" is given for patients diagnosed with Juvenile Cataract.
- "3" is given for patients diagnosed with Senile Cataract.
- "4" is given for patients diagnosed with Traumatic Cataract.

Table 1. Symptoms of Diseases and their Value

Symptoms	Variable	Value
White/Grayish Pupil (X_1)	No	0
	Yes	1
Age under 1 year old (X_2)	No	0
	Yes	1
Declining Sharpness (X_3)	No	0
	Yes	1
Double Visions (X_4)	No	0
	Yes	1
Itchy Eyes (X_5)	No	0
	Yes	1
Blurred Vision (X_6)	No	0
	Yes	1
Red Eyes (X_7)	No	0
	Yes	1
Sensitivity towards Light (X_8)	No	0
	Yes	1
Age between 1 to 50 years old (X_9)	No	0
	Yes	1
Age over 50 years old (X_{10})	No	0
	Yes	1
Contrasting Sensitivity (X_{11})	No	0
	Yes	1
Myopia Translation (X_{12})	No	0
	Yes	1
Monocular Diplopia (X_{13})	No	0
	Yes	1

3.3 Learning Vector Quantization Method Manual Calculation

This manual calculation uses 8 data of cataract patients diagnosed by doctors, 13 symptoms influencing cataract diagnosis are also used. Determining the data used for initializing weight vector (w) on the 1st, 2nd, 3rd and 4th training data. The training vector (x) on training data which are the 5th, 6th, 7th, and 8th training data. Determination of these data is under the condition of representing 4 classes/targets which are class/target 1, 2, 3 and 4.

Tabel 2. Data Calculation Training Manual

No	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	y	Data
1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	INITIAL WEIGHT
2	1	0	0	0	1	1	1	0	1	0	0	0	0	2	
3	1	0	0	1	0	0	0	1	0	1	0	0	0	3	
4	0	0	1	0	0	1	0	0	1	0	1	0	1	4	
5	0	0	1	1	0	1	0	1	0	1	0	0	0	3	TR AI NI NG
6	1	1	0	0	0	0	0	0	0	0	0	0	0	1	
7	1	0	1	0	1	1	0	1	1	0	0	0	0	2	
8	0	0	0	0	0	1	0	1	1	0	0	1	1	4	

Initialize Max Epoch value = 5, *learning rate* (α) = 0.1 and decrease of *learning rate* (Dec α) = 0.0001

To manually calculate the training data to get the final weight, the following process is done:

- EPOCH 1.

Step 0. Initial weight initialization:

$$\begin{aligned} w_1 &= (1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) \\ w_2 &= (1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0) \\ w_3 &= (1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0) \\ w_4 &= (0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0) \end{aligned}$$

Step 1. Start the calculation.

Step 2. For *input* vector $x_5 = (0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0)$ with $T = 3$, do step 3-4.

Step 3. $J = \|x_5 - w_{ij}\|$

For distance to:

a. 1st weight (w_1)

$$j = \sqrt{\frac{(0-1)^2 + (0-1)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2}{(1-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}}$$

$$j = 2.646$$

b. 2nd weight (w_2)

$$j = \sqrt{\frac{(0-1)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (0-1)^2 + (1-1)^2 + (0-1)^2}{(1-0)^2 + (0-1)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}}$$

$$j = 2.828$$

c. 3rd weight (w_3)

$$j = \sqrt{\frac{(0-1)^2 + (0-0)^2 + (1-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2}{(1-1)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}}$$

$$j = 1.732$$

d. 4th weight (w_4)

$$j = \sqrt{\frac{(0-0)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2}{(1-0)^2 + (0-1)^2 + (1-0)^2 + (0-1)^2 + (0-0)^2 + (0-1)^2}}$$

$$j = 2.449$$

The smallest / minimum distance is at 3rd weight (w_3), then $j = 3$.

Step 4. Because $T = 3$ and $C_j = C_3 = 3$ then $C_j = T$, *update* w_3 as follows:

$$\begin{aligned} w_3(\text{baru}) &= w_{3j}(\text{lama}) + \alpha [x_{5j} - w_{3j}(\text{lama})] \\ w_3(\text{baru}) &= (1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0) + 0.1 \begin{bmatrix} (0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0) \\ -(1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0) \end{bmatrix} \\ w_3(\text{baru}) &= (0.9, 0, 0.1, 1, 0, 0.1, 0, 1, 0, 1, 0, 0) \end{aligned}$$

Then this (new) w_3 weight will be used as the 3rd weight in calculating the distance to the next training data. This manual calculation process will continue until the 109th training data because out of the 146 existing data, 109 data is used as training data and 37 data is used as testing data.

3.4 Data Processing Design

The number of data from cataract patients was 146 data, which used 105 data for training, 4 data for initial weight and 37 data for testing system accuracy in recognizing input patterns.

Table 3. The Distance Between Data Testing with the Final Weight of Training Results

Patient	Class 1	Class 2	Class 3	Class 4	Target	Results
1	0.000	2.030	1.830	2.211	1	1
2	0.000	2.030	1.830	2.211	1	1
3	0.000	2.030	1.830	2.211	1	1
4	2.000	0.841	1.966	1.848	2	2
5	2.236	1.047	2.026	1.820	2	2
6	2.236	1.047	2.026	1.820	2	2
7	2.000	0.896	1.966	1.848	2	2
8	2.449	1.225	2.062	1.902	2	2
9	2.236	1.000	2.026	1.820	2	2
10	2.236	0.625	2.206	2.101	2	2
11	2.000	1.847	1.051	2.077	3	3
12	2.449	2.216	1.357	1.902	3	3
13	2.000	2.370	1.096	2.489	3	3
14	2.000	2.216	1.319	2.256	3	3
15	2.000	2.370	1.096	2.489	3	3
16	2.000	2.216	1.319	2.256	3	3
17	2.449	2.216	1.357	1.902	3	3
18	2.000	2.370	1.096	2.489	3	3
19	1.732	2.101	1.261	2.188	3	3
20	1.732	2.101	1.261	2.188	3	3
21	2.449	2.326	1.182	2.077	3	3
22	2.449	2.216	1.357	1.902	3	3
23	1.732	2.263	1.026	2.426	3	3
24	2.236	2.263	1.075	2.101	3	3
25	2.000	2.216	1.319	2.256	3	3
26	2.236	2.101	1.301	1.820	3	3
27	2.000	2.370	1.096	2.489	3	3
28	1.732	2.263	1.026	2.426	3	3
29	2.646	2.431	2.555	1.284	4	4
30	2.449	1.902	2.145	1.189	4	4
31	2.828	2.326	2.625	1.021	4	4
32	2.646	2.532	1.973	1.694	4	4
33	2.449	2.030	2.378	0.985	4	4
34	2.449	2.326	2.526	1.350	4	4
35	2.449	1.847	2.166	1.004	4	4
36	2.646	2.149	2.367	1.260	4	4
37	2.449	2.326	2.526	1.350	4	4

In the LVQ method, the initial weights use data taken at random from existing patient data. Then the weight will be changed depending on the input vector class according to the class declared as the winning neuron. Initial weights are taken from existing data which must be taken into vector form. After obtaining the final weight of the training data training results, then the output weight of the training results is used in the testing phase with new data input. At this testing phase, the distance between the testing data and the final weight will be calculated using the Euclidean distance. Then the data cluster is determined by selecting the minimum vector distance as the winner.

Based on Table 3. above, it can be seen that after conducting the testing process by calculating the Euclidean distance and selecting the minimum distance as the winning neuron, the 37 testing data obtained results of 37 patients having identified characters or diagnoses using the same program as the results diagnosed by doctor. So it can be concluded that after testing the 37 data to 5 epochs, the success rate of the LVQ network to recognize patterns correctly is 100%.

3.5 Comparison of Test Results

The following table presents test results using several alpha values, reduced alpha values and a maximum of 5 epochs, to find out the best level of accuracy from the LVQ method.

Table 6. Comparison of Test Results

No	α	Dec α	Testing Results		
			Known Data	Unknown Data	Accuracy Rate
1	0.9	0.0001	18	19	48.6%
2	0.8	0.0001	20	17	54.1%
3	0.7	0.001	26	11	70.2%
4	0.6	0.001	27	10	75.7%
5	0.5	0.002	28	9	89.2%
6	0.4	0.0002	31	6	83.8%
7	0.1	0.0001	37	0	100%
8	0.01	0.001	21	16	56.7%
9	0.001	0.0001	36	1	97.3%

Based on the test results as presented in the above table, it can be seen that the accuracy of the diagnosis results is very much influenced by the choice of training rate parameter (α) and decreased training rate (*Dec α*), In addition, based on the training process and data testing it is also known that the selection of initial weights also greatly influences the accuracy of the diagnosis results so that the selection of initial weights must be taken from existing data. From the test results on testing data obtained that the best test results are 100%, located at the value of the training rate (α) = 0,1 for decreased alpha (*Dec α*) = 0,0001 and *epoch* 5.

3.6 Diagnostic Form

To diagnose cataracts in this research, a program has been designed which was built using a toolbox that has been provided by MATLAB so that the process is easy, namely by using Graphical User Interface Design (GUIDe). The following is the design of the interface for the diagnostic form designed.



Figure 3. Display the Diagnosis Interface

The following table is a description of the key functions found in the main view of the program.

Table 7. Function of Keys in Programs

Button	Function
Age Choices	There are 3 buttons in this section which are the age of the patient that must be selected.
Patient Symptoms	In this section there are 10 buttons which are symptoms experienced by patients that must be selected.
Diagnosis	To diagnose the disease and display the patient's disease in the "PATIENT DISEASE" box by reading the input results from X1 to X13, this input pattern is obtained from filling in the patient's age and symptoms of the disease carried out by the patient.
Reset	To reset the program form, so that the form can be filled with new data.

Based on patient examination data using the symptoms experienced following are examples of the use of the program in one of the patients. After entering the symptoms experienced by the patient based on the data obtained then click the DIAGNOSA button on the application, then the type of disease suffered by the patient will be shown.



Figure 4. Patient Diagnosis Results

The program diagnoses patients suffering from Congenital Cataracts. It turns out that the diagnosis results from the program are the same as the original data. Furthermore, the same process is carried out in other patients, so this program can be used to diagnose or classify cataracts.

4. CONCLUSION

Based on the research it can be concluded that the system created is able to diagnose cataract disease optimally with *learning rate* (α) = 0,1 and decreased *learning rate* (*Dec* α) = 0,0001 and 100% accuracy of diagnosis. As the learning rate (α) used in the testing process goes smaller, the diagnosis accuracy rate tends to increase.

REFERENCES

- [1] Lubis R. R, Megawati E. R, Lubis D. L. Identifikasi Kelainan Mata dan Koreksi Tajam Penglihatan Presbiopia. *ABDIMAS Talenta Universitas Sumatera Utara*. 2016; 1 (1): 13–19.
- [2] de-Wit L, Wagemans J. Visual Perception. *Encyclopedia of Human Behavior. Second Edition*. 2012: 665-671.
- [3] Faizal E. Case Based Reasoning Diagnosis Penyakit Mata. *Jurnal Teknologi Informasi dan Ilmu Komputer*. 2012; 10 (2): 26–37.
- [4] Zulfikar W. B, Lukman N. Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata. *Jurnal Online Informatika*. Dec 2016; 1 (2): 82–86.
- [5] Lambert S. R, Drack A. V. Infantile Cataracts. *Survey of Ophthalmology*. 1996; 40 (6): 427-458.
- [6] Rasyid1 H. A. Z. A. R, Nawi R. Summary for Policymakers. *Climate Change 2013 - The Physical Science Basis*. Intergovernmental Panel on Climate Change. Edition Cambridge: Cambridge University Press. 2010: 1–30.
- [7] Lee C. M, Afshari N. A. The Global State of Cataract Blindness. *Current Opinion in*

- Ophthalmology*. 2017; 28 (1): 98-103.
- [8] Martono G. H, Eng M, Agrippina S, Yusuf A. Diagnosa Penyakit Katarak Senilis Dengan Menggunakan Metode Case Based Reasoning (Cbr) Berbasis Web. 2016: 6–7.
- [9] Nasri. Kecerdasan Buatan (Artificial Intelligence). *Artif. Intell.* 2014.
- [10] Fausett L. Fundamentals Of Neural Network Architectures, Algorithms, and Applications. *Inc., New Jersey*. 2017.
- [11] Itchhaporia D. Artificial Neural Networks: Current Status in Cardiovascular Medicine. *Journal of the American College of Cardiology*. 1996; 28 (2): 515-521.
- [12] Kumar K, Abhishek A. Artificial Neural Networks for Diagnosis of Kidney Stones Disease. *International Journal of Information Technology and Computer Science*. 2012; 4 (7): 20-25.
- [13] Ganesan N, Venkatesh K, Rama M. A, Palani A. M. Application of Neural Networks in Diagnosing Cancer Disease using Demographic Data. *International Journal of Computer Applications*. 2010; 1 (26): 81-97.
- [14] Widyawati K, Setiawan B. D, Adikara P. P. Optimasi Vektor Bobot Learning Vector Quantization Menggunakan Algoritme Genetika untuk Penentuan Kualitas Susu Sapi. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*. 2017.
- [15] Kohonen T. Learning Vector Quantization. *Neural Networks*, Jan 1988; 1: 303.
- [16] Ladauw E. B, Ratnawati D. E, Supianto A. A. Identifikasi Penyakit Mata Menggunakan Metode Learning Vector Quantization (LVQ). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 2018; 2 (12): 6989–6996.
- [17] Sabrina E. Klasifikasi Penyakit Diabetic Retinopathy menggunakan Metode Learning Vector Quantization (LVQ). *Jurnal Teknik Elektro*. 2017; 6 (2).
- [18] Pahlevi M. K, Setiawan B. D, Afirianto T. Identifikasi Gangguan Kepribadian Dramatis Menggunakan Metode Learning Vector Quantization (LVQ). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 2018; 2 (9): 3103–3111.