The Comparison of Some Version of Linear Vector Quantization (LVQ) for Vitamin and Mineral Deficiency Early Detection

Nina Sevani †
Faculty of Engineering and Computer Science
Krida Wacana Christian University
Jakarta Indonesia

Iwan Aang Soenandi
Faculty of Engineering and Computer Science
Krida Wacana Christian University
Jakarta Indonesia
iwan.as@ukrida.ac.id

Richardo K. Sali
Faculty of Engineering and Computer Science
Krida Wacana Christian University
Jakarta Indonesia
richardo.2016tin007@civitas.ukrida.ac.id

ABSTRACT

Vitamin and mineral deficiency are often ignored because they do not have a direct impact on body health. However, prolonged deficiency can cause various diseases from mild to serious illness. Some previous research in computer science already conducted to make early detection of vitamin and mineral deficiency, but no one has produced an adaptive model to find out the most dominant type of deficiency. Therefore, the goal of this research is to develop an adaptive model using an artificial neural network (ANN) with Linear Vector Quantization (LVQ) as the learning algorithm to make early detection of vitamin and mineral deficiency. LVQ consists of three layers: an input layer that represents the features, output layer that represent the class label, and the competitive layer. The competitive layer will save the distance between the input vector and the codebook vector from each class. The distance will calculate using Euclidean Distance. LVQ also involves some parameters in the training process, like epsilon value, learning rate, codebook vector, epoch, and window size which obtained by trial and error experiment. This research will also compare the performance of some version of LVQ. The experiment results show that the maximum accuracy level obtained by the system is 85.71% by using LVQ3. The dataset used split into data training and data testing with a ratio 84:16 respectively. From our scenario, the optimum model was achieved by using 20 codebook vectors with the number of epochs is 3400 and the value of the learning rate parameter (α) of 0.4, window size (ω) of 0.3, and epsilon (ε) of 0.2.

CCS CONCEPTS

•Computing methodologies ~Machine learning~ Machine learning approaches~ Neural networks

ICONETSI, September 28–29, 2020, Tangerang, Indonesia

ACM ISBN 978-1-4503-8771-2/20/09...$15.00
https://doi.org/10.1145/3429789.3429869

KEYWORDS

Learning, Neural Network, Adaptive Model, Detection

ACM Reference Format:

Nina Sevani, Iwan Aang Soenandi, and Richardo K Sali.2020. The Comparison of Some Version of Linear Vector Quantization (LVQ) for Vitamin and Mineral Deficiency Early Detection. In Proceedings of International Conference on Engineering and Information Technology for Sustainable Industry (ICONETSI 2020), September 28-29, 2020, Tangerang, Indonesia. ACM, New York,NY,USA, 6 pages.

1 INTRODUCTION

Nutritional status has become one of major health problems in Indonesia. Government data revealed a high score of 95.5% as a score for the lack of fruit and vegetable intake [1][2]. Vitamin and mineral deficiency can eventually lead to a variety of health problems from mild to serious problems. Early detection of vitamin and mineral deficiency based on physical symptoms can be done as prevention from prolonged deficiency. Previous research in computer science shows some methods used for early detection of vitamin and mineral deficiency, like conventional rule-based and certainty factor [3][4][5]. However, these methods are considered less adaptive, especially to handle big amount of data with varying complexity. While in fact, the detection of vitamin and mineral deficiency needs many symptoms that related to one another. This condition can increase difficulty in determining the kinds of vitamin and mineral deficiencies. Therefore, the use of ANN that able to handle big and complex data can help in determining the kind of vitamin and mineral deficiency.

One of the important factors in the implementation of ANN is the learning algorithm used to produce an optimum model based on the data training. LVQ is one of the learning algorithms that used in many domains and proven to provide good accuracy [6][7][8][9][10][11][12]. In the time consumption, LVQ also proven can produce the model faster than other learning algorithms [13].

Currently, LVQ already developed until some versions. Previous works show that the variations of LVQ obtain diverse result for some cases such as for tumor brain and nutritional disorders of children [7][14]. Therefore this result will develop an adaptive model using LVQ and compare the performance some of LVQ versions for vitamin and mineral deficiency cases.

2 METHOD

LVQ has developed into some versions. This experiment using three versions of LVQ: LVQ1, LVQ2.1, and LVQ3. The difference between the versions is in the parameters used in the training process and the time to update the value of the distance in the competitive layer. There are some steps to implement ANN using the LVQ algorithm. It starts with data transformation from the patient's physical symptoms into dataset to be processed by LVQ. Next step is the determination of the optimum parameters value by using an experiment that is designed into several scenarios. After determining the optimum parameter value, the next step is designing the ANN architecture and testing the performance of the resulting model. In the testing step, the experiment will be conducted using some dataset ratio. The assessment will be done using a confusion matrix

2.1 Dataset Preparation

Dataset preparation aims are to transform the data from physical symptoms into input data to be processed by the algorithm. The physical symptoms are the symptoms that can be observed and feel immediately by the patient, as mentioned in Table 1. Each of these symptoms can be related to each other and lead to the same diagnose result. ANN will train the dataset to recognize each input vector using the LVQ algorithm. As supervised learning, LVQ works using dataset that has an input variable (X_i) and target (class). In general, the dataset will be divided into training data to training the model and testing data to test the performance of the model.

The dataset used in this research contains 100 data, the same with that used in the previous study [3][4][5] which obtained from doctors or medical teams. Each data may have the same feature (physical symptoms) with other data in the dataset. These features will be transformed into binary number to be processed by LVQ algorithm. Table 1 and Table 2 show some examples of the physical symptoms as initial data and the result for the transformation.

Table 1 Transformation of The Data

No	Psychical Symptoms	Transformation	Diagnose Result
1	• Dizzy eyes (X_1) • Pale in the inside of the eyelids (X_2) • Panting breath (X_3)	• X_1 = 1 • X_2 = 1 • X_3 = 1 • The last *features* X_4 to X_{107} has 0 value	Irons Mineral
....	....		....
86	• Dry mouth (X_{91})	• X_{91} = 1 • The last *features* X_1 to X_{90} and X_{92} to X_{107} have 0 value	Vitamin C

Tabel 2 Conversion Result

X_1	X_2	X_3	X_4	X_5	.....	X_{107}	Class
1	1	1	0	0	.....	0	Iron
0	0	0	0	0	.....	0	Zinc
.....	.....	.....	.....	.....	.....	.....	.....
1	0	1	0	0	.....	0	A

2.2 Designing Artificial Neural Network Model

Based on the dataset used, the architecture of the ANN consist of 107 nodes for input layer following the features (psychical symptoms), and 17 nodes for output layer following the types of deficiency can be detected. LVQ also consists of one additional layer called a competitive layer to bridge the input and output layer. In the competitive layer, each node will save the distance between input vector to each codebook vector located on the edge that is connected directly to each of these nodes. The distance will be calculated using Euclidean Distance as formulated in Equation (1) [3].

$$D(i) = \sum_{i=1}^{n} (x_i - y_i), \quad \text{.................................(1)}$$

where D(i) is the distance, X_i is the input vector, Y_i is the codebook vector, and n is the amount of the input node. The codebook vector is a representation vector of each class in the dataset. Each value in the codebook vector will dynamically change during the training process. Figure 1 displays the ANN architecture used in the experiment. The input nodes written in the notation X_i to X_{107}, output nodes written in the notation C_1 to C_{17} which represent each class target. The competitive layer written in the notation Y_1C_1 to Y_1C_{17} and Y_2C_1 to Y_2C_{17} which saves the Euclidean distance between input vector to codebook vector. The codebook vector depicted in form of a connecting lines (edges) between input layer (X_1 to X_{17}) to competitive layer (Y_1C_1 to Y_1C_{17}).

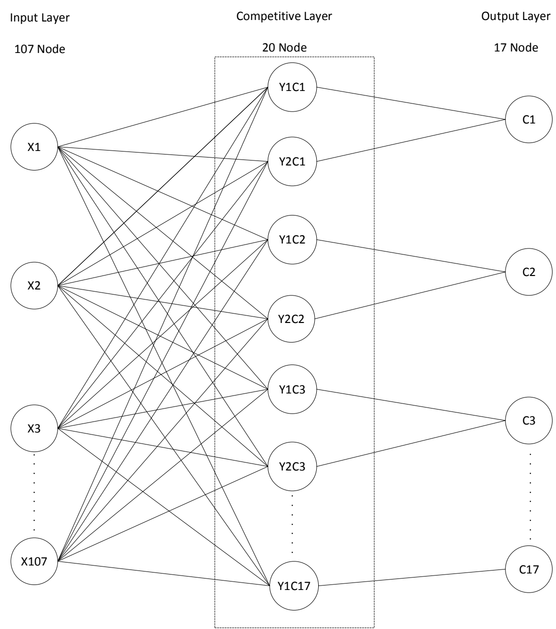

Figure 1: ANN Architecture with LVQ Algorithm

2.3 Determining Parameters Value

There are some parameters needed to be determined before the training process using LVQ algorithm, like epsilon value, learning rate (α), the amount of the codebook vector, the amount of the epoch (ε), and the window (ω). To get the best value for each parameter, this research does several scenarios by trial and error process.

The First Scenario, aims to get the optimal value of the learning rate, window size, and epsilon value. The codebook vector number set to 20 and the total epoch set to 1000. The learning rate value configured on the range of 0.1 to 0.5, the window size configured on the size of 0.2 to 0.3, and the epsilon value will be configured on the range of 0.1 to 0.5. The result of the experiment using first scenario can be seen in Table 3.

From the experiment results, the best accuracy is 77.53% obtained with epsilon value 0.2, α=0.4, and ω=0.3.

The Second Scenario, want to determine the optimum codebook vector number and discover the effect of the codebook vector with the model performance. In this scenario, the best value of epsilon, window size, and learning rate from first scenario will be used. The codebook vector will be configured in the range of 17 to 41. The result of the experiment for second scenario showed in Table 4.

TABEL 3 RESULT OF FIRST SCENARIO

No	Parameter			Accuracy
	Learning rate	Window size	Epsilon	
1	0.3	0.3	0.1	76.40%
2	0.3	0.3	0.2	75.28%
3	0.3	0.3	0.3	74.16%
4	0.3	0.3	0.4	71.91%
5	0.3	0.3	0.5	73.03%
No	Parameter			Accuracy
	Learning rate	Window size	Epsilon	
6	0.2	0.3	0.1	71.91%
7	0.2	0.3	0.2	70.79%
8	0.2	0.3	0.3	73.03%
9	0.2	0.3	0.4	74.16%
10	0.2	0.3	0.5	70.79%
11	0.1	0.3	0.1	62.92%
12	0.1	0.3	0.2	62.92%
13	0.1	0.3	0.3	65.17%
14	0.1	0.3	0.4	64.04%
15	0.1	0.3	0.5	65.17%
16	0.4	0.3	0.1	74.16%
17	0.4	0.3	0.2	77.53%
18	0.4	0.3	0.3	73.03%
19	0.4	0.3	0.4	75.28%
20	0.4	0.3	0.5	70.79%
21	0.5	0.3	0.1	74.16%
22	0.5	0.3	0.2	73.03%
23	0.5	0.3	0.3	73.03%
24	0.5	0.3	0.4	74.16%
25	0.5	0.3	0.5	70.79%
26	0.3	0.2	0.1	64.04%
27	0.3	0.2	0.2	61.80%
28	0.3	0.2	0.3	59.55%
29	0.3	0.2	0.4	62.92%
30	0.3	0.2	0.5	61.80%
31	0.2	0.2	0.1	65.17%
32	0.2	0.2	0.2	61.80%
33	0.2	0.2	0.3	64.04%
34	0.2	0.2	0.4	61.80%
35	0.2	0.2	0.5	61.80%
36	0.1	0.2	0.1	57.30%
37	0.1	0.2	0.2	57.30%
38	0.1	0.2	0.3	59.55%
39	0.1	0.2	0.4	58.43%
40	0.1	0.2	0.5	60.67%
41	0.4	0.2	0.1	69.66%
No	Parameter			Accuracy
	Learning rate	Window size	Epsilon	
42	0.4	0.2	0.2	66.29%
43	0.4	0.2	0.3	67.42%
44	0.4	0.2	0.4	62.92%
45	0.4	0.2	0.5	60.67%
46	0.5	0.2	0.1	67.42%
47	0.5	0.2	0.2	66.29%
48	0.5	0.2	0.3	62.92%
49	0.5	0.2	0.4	58.43%
50	0.5	0.2	0.5	60.67%

TABLE 4 RESULT OF SECOND SCENARIO

No	Codebook Vector Number	Accuracy
1	17	76.92%
2	20	80.77%
3	23	50%
4	26	46.15%
5	29	46.15%
6	32	50%
7	35	57.69%
8	38	65.38%
9	41	53.85%

The result of the second scenario shows the best accuracy is 80.77% with 20 of codebook vector numbers.

Third Scenario is the scenario to get the optimal value of epoch number. This scenario still using the best parameter value obtained from the first and second scenarios (the epsilon value of 0.2, α=0.4, and ω=0.3, the codebook vector number of 20). The result of the third scenario showed in Table 5, with the number of epoch is 2400 or 3400, which can give accuracy of 80.77%.

TABLE 5 RESULT OF THIRD SCENARIO

No	Number of Epoch	Accuracy
1	1000	57.69%
2	1200	61.54%
3	1400	61.54%
4	1600	57.69%
5	1800	69.23%
6	2000	73.08%
7	2200	76.92%
8	2400	80.77%
9	2600	69.23%
10	2800	73.08%
11	3000	73.08%
12	3200	65.38%
13	3400	80.77%
14	3600	76.92%

Based on the experiment result using three scenarios, Table 6 summarize the optimum value for all the parameter used by LVQ to train the model.

Table 6 The Optimum Value of Parameters In LVQ

α	ω	ε	Epoch Number	Codebook Vector Number
0.4	0.3	0.2	2400	20

2.4 Codebook Initialization

There are some methods to initialize the codebook vector, such as splitting algorithm, random algorithm, and sorting algorithm [18]. This research using random training data even as the algorithm for initializing the codebook vector. This algorithm will randomly select the vector in the dataset to be used as codebook vector. The use of this algorithm is motivated by the better result of the algorithm compared to other algorithms.

2.5 Performance Testing

To measure the performance of the model, we use confusion matrix. There are three values used for the assessment: precision, sensitivity, dan F1-score. These three assessment values will be calculated for each of the class targets. Equation (2) to (4) will be used to assess the performance of the model to predict deficiency. [19].

$$Precision = \frac{TP}{TP+FP}, \qquad (2)$$

$$Sensitivity = \frac{TN}{TN+FP}, \qquad (3)$$

$$F1-score = 2 * \frac{(Precision * Sensitivity)}{(Precision + Sensivity)}, \qquad (4)$$

3 EXPERIMENT RESULT AND DISCUSSION

The testing phase for the model done using two scenarios. The first scenario is using the different ratios of data training and data testing from the same dataset. The second scenario is using different versions of LVQ algorithms.

3.1 Testing With Different Combination Ratio

One group of dataset will be divided into some ratio of data training and data testing. Each pair of ratio will be used in the experiment using parameter value in Table 6. By using 100 data, the ratio of data training and data testing will be set into 80:20, 82:18, 84:16, 86:14, 88:12 dan 90:10, respectively. The experiment result showed in Table 9. From the result can be concluded that the best accuracy is 85.71%, with using 84 data for training and 16 data for testing.

Table 9 Experiment Result Using Different Ratio of Data Training And Data Testing

Ratio Data Train: Test	Accuracy
80:20	82.35%
82:18	73.33%
84:16	85.71%
86:14	66.67%
88:12	80%
90:10	77.78%

3.2 Testing With Different Versions of Linear Vector Quantization

The aim of the experiment using different versions of LVQ algorithms is to know the most appropriate version of LVQ in the case of vitamin and mineral deficiency. There are three versions of LVQ to be tested: LVQ1, LVQ2.1, and LVQ3. The experiment using the best parameter value mentioned in Table 6 and data ratio 84:16.

Table 10 shows the assessment value for each class target using LVQ3 as the training algorithm. The best accuracy for LVQ3 is 85.71%.

Table 10 Result of LVQ3

Class	Precision	Sensitivity	F1-Score
Iron	1	1	1
Fluorine	1	1	1
Iodine	1	1	1
Zinc	1	1	1
A	0	0	0
Class	**Precision**	**Sensitivity**	**F1-Score**
B1	1	1	1
B2	0.667	1	0.8
B3	0.5	1	0.667
B5	0	0	0
B6	0	0	0
B7	0	0	0
B9	0	0	0
B12	0	0	0
C	1	1	1
D	1	1	1
E	1	1	1
K	0	0	0
Average Weight	0.774	0.857	0.805
Accuracy	85.71%		

By using LVQ2.1, the best accuracy of the model can reach 76.47%, is shown in Table 11.

Table 11 Result of LVQ2.1

Class	Precision	Recall	F-Score
Iron	1	1	1
Fluorine	1	1	1
Iodine	1	1	1
Zinc	1	1	1
A	0	0	0
B1	1	1	1
B2	1	1	1
B3	0.5	1	0.667
B5	0	0	0
B6	0	0	0
B7	0	0	0
B9	0	0	0
B12	0	0	0
C	1	1	1
D	0	0	0
E	0.5	1	0.667
K	0	0	0
Average Weight	0.706	0.765	0.725
Accuracy	76.47%		

Next, we using LVQ1 than the best accuracy only reached 70.59%, is shown in Table 12.

From the experiment using three types of LVQ, we can conclude that the best accuracy for vitamin and mineral deficiency obtained using LVQ3. The worst accuracy showed using LVQ1. This result also revealed that for vitamin and mineral deficiency, the newest version of LVQ can give higher accuracy. In other words, the improvement in the newest version of LVQ proven can give higher accuracy.

To ensure the accuracy of the LVQ in the case of early detection of vitamins and minerals, the results also compared with the Backpropagation Network (BPN), another learning algorithm commonly used in ANN. The experiment using BPN show that LVQ3 has better performance than BPN which give only 73% accuracy.

Table 12 Result of LVQ1

Class	Precision	Recall	F1-Score
Iron	1	1	1
Fluorine	0.333	1	0.5
Iodine	1	1	1
Zinc	1	1	1
A	0	0	0
B1	1	1	1
B2	1	1	1
B3	0	0	0
B5	0	0	0
B6	0	0	0
B7	0	0	0
B9	0	0	0
B12	0	0	0
C	0.667	1	0.8
D	0	0	0
E	0.5	1	0.667
K	0	0	0
Average Weight	0.598	0.706	0.633
Accuracy	70.59%		

4 CONCLUSION

From the experiment of comparison performance testing, it was concluded that LVQ as the learning algorithms can be successfully implemented in the case of vitamin and mineral deficiency. The optimum model obtained using a learning rate (α) of 0.4, window size (ω) of 0.2, and epsilon value (ε) of 0.3. The best accuracy gets when using random training data even for the initialization of codebook vector and LVQ3 as the learning algorithm, with the accuracy of 85.71% and using 84:16 ratio for data training and testing.

For further experiments, we can add more data for the training process and do more detailed testing process. For example, we can divide the dataset into training, validation, testing data, and do k-fold cross-validation

REFERENCES

[1] Kemenkes RI.2018. *Hasil* Utama Riset Kesehatan Dasar Tahun 2018

[2] T. A. Purba. 2019. *Pentingnya Mengajarkan Kebiasaan Pola Makan Gizi Seimbang Sejak Dini*, [Online]. Available: https://lifestyle.bisnis.com/read/20190801/236/1131488/pentingnya-mengajarkan-kebiasaan-pola-makan-gizi-seimbang-sejak-dini.

[3] N. Sevani, R. Unwaru, P. Studi, and T. Informatika. 2014. Aplikasi Deteksi Dini Defisiensi Mineral Mikro pada Manusia Berbasis Web, *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 8, no. 2, pp. 213–222. https://doi.org/10.22146/ijccs.6548

[4] N. Sevani and Y. J. Chandra. 2016. Web Based Application for Early Detection of Vitamin and Mineral Deficiency. *CommIT (Communication Inf. Technol. J.*, vol. 10, no. 2, p. 53. DOI: https://doi.org/10.21512/commit.v10i2.1575

[5] N. Sevani and M. Joshua. 2015. Implementasi Forward Chaining Untuk Diagnosa Defisiensi Vitamin Larut Dalam Lemak Berbasiskan Web, *J. Inform.*, vol. 10, no. 2. DOI 10.21460/inf.2014.102.293.

[6] D. Fitria, M. A. Ma'sum, E. M. Imah, and A. A. . Gunawan. 2014. Automatic Arrhythmias Detection Using Various Types of Artificial Neural Network, *J. Comput. Sci. Inf.*, vol. 2, pp. 90–100. DOI: https://doi.org/10.21609/jiki.v7i2.262

[7] S. N. H. Sheikh Abdullah et al. 2016. Round Randomized Learning Vector Quantization for Brain Tumor Imaging. *Comput. Math. Methods Med.*, Vol. 2016 , Article ID 8603609, 19 pages http://dx.doi.org/10.1155/2016/8603609.

[8] B. A. Rahadian, C. Dewi, and B. Rahayudi. 2018. The performance of genetic algorithm learning vector quantization 2 neural network on identification of the types of attention deficit hyperactivity disorder. *Int. Conf. Sustain. Inf. Eng. Technol. SIET* 2017, vol. 2018–January, pp. 337–341. DOI: 10.1109/SIET.2017.8304160.

[9] R. Sonavane, P. Sonar, and S. Sutar. 2017. Classification of MRI brain tumor and mammogram images using learning vector quantization neural network. Proc. 2017 3rd IEEE *Int. Conf. Sensing, Signal Process. Secur.* ICSSS 2017, pp. 301–307. DOI: 10.1109/SSPS.2017.8071610

[10] F. M. Putra and F. Syafria. 2018. Penerapan Learning Vector Quantization 3 (LVQ3) untuk Mengidentifikasi Citra Darah Acute Lymphoblastic Leukemia (ALL) dan Acute Myeloid Leukemia (AML), *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, p. 27. DOI: 10.24014/coreit.v4i1.6124.

[11] A. Budiman, P. Utomo, and S. Rahayu. 2019. Android based rice pest detection system using learning vector quantization method, *IOP Conf. Ser. Earth Environ. Sci.*, vol. 293, no. 1. doi:10.1088/1755-1315/293/1/012001.

[12] S. Sanjaya. 2018. Learning Vector Quantization 3 (LVQ3) and Spatial Fuzzy C- Means (SFCM) for Beef and Pork Image Classification. *Indonesian Journal of Artificial Intelligence and Data Mining* vol. 1, no. 2, pp. 60–65. DOI: http://dx.doi.org/10.24014/ijaidm.v1i2.5024

[13] F. Syafria, A. Buono, and B. P. Silalahi. 2014. A comparison of backpropagation and LVQ: A case study of lung sound recognition, *Proc. - ICACSIS 2014 Int. Conf. Adv. Comput. Sci. Inf. Syst.*, pp. 402–407. DOI: 10.1109/ICACSIS.2014.7065873.

[14] E. Budianita and Widodo Prijodiprodjo. 2013. Penerapan Learning Vector Quantization (LVQ) untuk Klasifikasi Status Gizi Anak, *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 2, pp. 155–166. https://doi.org/10.22146/ijccs.3354.

[15] Z. Immadudin and I. Habibie. 2013. *Teknik Biomedis : Teori dan Implementasi* . Penerbit Fakultas Ilmu Komputer Universitas Indonesia.

[16] G. Puşcaşu, A. Stancu, B. Codreş, and A. Odreş. 2018. New learning strategy for prototypes in linear vector quantization, 22nd *Int. Conf. Syst. Theory, Control Comput. ICSTCC 2018* - Proc., pp. 194–200. DOI: 10.1109/ICSTCC.2018.8540745.

[17] H. Li and B. Xu. 2007. A new initial codebook algorithm for learning vector quantization. *IEEE Int. Conf. Commun.*, pp. 2610–2613. DOI: 10.1109/ICC.2007.432.

[18] A. Tharwat. 2018. Classification assessment methods," Appl. Comput. Informatics, 2018. [2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79).* ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774.