# Soybean disease detection with feature selection using stepwise regression algorithm: LVQ vs LVQ2

**Nida Muhammad[1], Sukmawati Nur Endah*[2], Eko Adi Sarwoko[3], Priyo Sidik Sasongko[4]**
Universitas Diponegoro, Indonesia[1,2,3,4]

## Article Info

**Keywords:**
Soybean plant disease detection, Learning Vector Quantization, LVQ2, Stepwise Regression Algorithm.

**Article history:**
Received 30 August 2019
Revised 17 January 2020
Accepted 24 March 2020
Available 20 May 2020

**Cite:**
Muhamad, N., Endah, S., Sarwoko, E., & Sasongko, P. (2020). Soybean Disease Detection with Feature Selection Using Stepwise Regression Algorithm: LVQ vs LVQ2. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(2). doi:https://doi.org/10.22219/kinetik.v5i2.919

*Corresponding author.
Sukmawati Nur Endah
E-mail address:
sukmawati020578@gmail.com

## Abstract

According to data from BPS, soybean production in Indonesia is still low. One of the causes of decreased soybean production is the existence of diseases in soybean plants (Glycine max (L.) Merr). These diseases are more quickly overcome if they can be detected earlier, by taking advantage of information technology. Detection can be done by looking at the physical characteristics of the plant. There are 35 characteristics or attributes that can be used to categorize types of soybean plant diseases. However, not all 35 attributes necessarily have a great influence on categorizing the types of diseases, so attribute selection needs to be done. This study proposes the detection of soybean disease by using the regression stepwise attribute selection algorithm. For the detection process, it would compare LVQ and LVQ2, which are easily implemented for existing data types. Data was taken from the University of California Irvine Machine Learning Repository as much as 200 data. The results showed that LVQ2 was better than LVQ in detecting soybean plant diseases which resulted in an accuracy of 90.5%, an error rate of 9.5%, sensitivity of 90.5%, and specificity of 98.94% with 17 attributes.

## 1. Introduction

Secondary crops are part of agricultural crops that play an important role in agriculture in Indonesia. The presence of secondary crops is important for farmers because secondary crops have a faster harvest time than main crops such as rice. Therefore, many farmers in the village depend on secondary crops. One of the most cultivated crops is soybeans, because of its multipurpose nature.

Soybean plants (Glycine max L. Merr) is one of the food crops that have been long cultivated by the people of Indonesia. This plant has an important meaning to fulfill food needs in the context of improving community nutrition, because it is a source of vegetable protein which is relatively inexpensive compared to other protein sources such as meat, milk and fish [1][2].

In Indonesia national demand for soybeans increases from year to year. But according to data on soybean production in each province, the amount of national soybean productivity is still low [3][4]. Because soybean productivity is still low, the fulfillment of soybean needs is done by importing soybeans from several countries such as China, Ukraine, Canada, Malaysia and the United States. Handling of disease attacks is one factor to increase soybean productivity [5].

Diseases in soybean plants include several types, including canker, anthracose, frog eye spots, purple seed stain, bacterial pustule, bacterial blight, brown spot, alternaria leaf spot, powdery mildew, and downy mildew [5][6][7][8]. If the disease attacks soybean plants, and handling the disease is too late, it will result in crop failure. For this reason, it is necessary to detect the attack of disease in soybeans. One effort that can be done is to make a intelligent software that is able to detect disease in soybean plants [9][10][11]. The software can be used by agricultural instructors, farmer group leaders, and also farmers to provide information of the diseases that attack the soybean plants. So that in the end disease attack can be dealt appropriately and soybean production can be rise again. One study of soybean disease, namely the Soy Bean Plant Identification System using the Naive Bayes Method resulted in an accuracy of 82% [12].

On the other hand, the disease detection process can be carried out by various methods of artificial neural network methods. (Joshi and Borse 2017) use Backpropagation to test diabetes which results in 81% accuracy [13]. Budianita & Firdaus (2016) used the LVQ2 Method for Psychological Diagnosis of Rumah Sakit Jiwa Tampan Pekan Baru which produced an accuracy of 90% [14].

LVQ2 is a improvement of LVQ, one of them is shown in research on the types of Attention Deficit Hyperactivity Disorder (ADHD) using the LVQ2 method. In the study (Rahadian, Dewi, and Rahayudi 2017), comparing the evaluation results between the LVQ2 method and the LVQ method to identify the type of Attention Deficit Hyperactivity Disorder (ADHD) disease that produced accuracy of 80% on LVQ2, 10% greater than using LVQ [15].

Detection of the soybean plants diseases with LVQ and LVQ2 can be done by looking at the physical characteristics of the plan which can be observed easily by humans. Therefore, data domains in this research are text which transform to numeric, not image. There are 35 characteristics or attributes that can be used to categorize types of soybean plant diseases. However, not all 35 attributes necessarily have a great influence on categorizing the types of diseases, so attribute selection needs to be done. There are various algorithm for selecting attributes. In a study that compared several attribute selection algorithm such as backward elimination, forward selection and stepwise regression, the stepwise regression algorithm resulted in the best accuracy, sensitivity, and specificity compared to other algorithm [16].

Based on this previous study, this study proposes the detection of soybean disease by using the regression stepwise attribute selection algorithm. For the detection process, it would compare LVQ and LVQ2, which are easily implemented for existing data types.

## 2. Research Method

This study has several stages of the process such as data collection, preprocessing, data partitioning, training, testing, and evaluation. The flow of each process is displayed in the form of problem solving process in Figure 1.

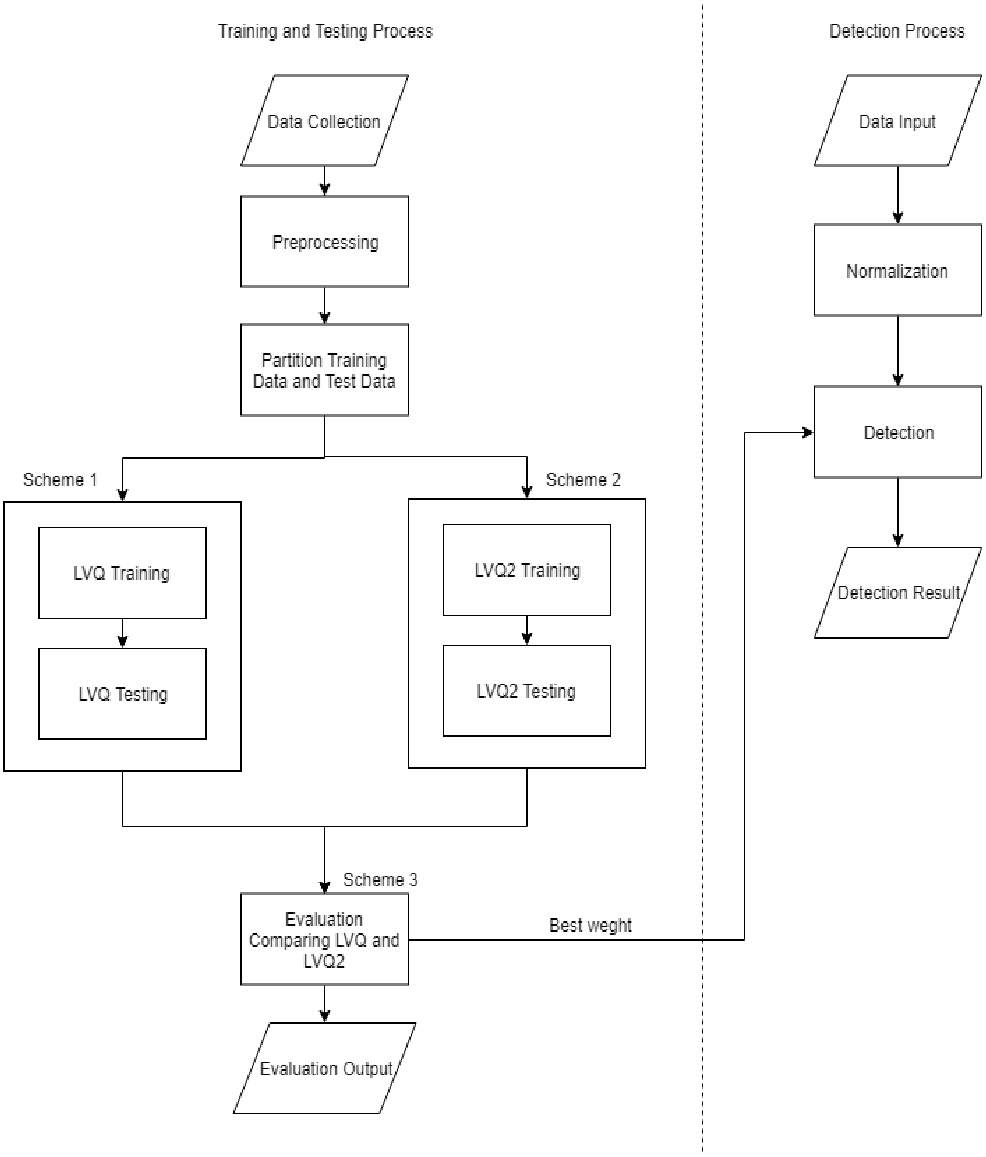

*Figure 1. Research Method*

© 2020 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)

### 2.1 Data Collection

The data used in this study is a soybean dataset taken from the University of California Irvine Machine Learning Repository about the Soybean Data Set [17]. The total data used are 200 data consisting of each 20 data on anthracnose, brown spot, alternaria leaf spot, frog eye leaf spot, purple seed stain, bacterial pustule, powdery mildew, bacterial blight, canker and downy mildew. In this data there are 10 types of diseases. While the number of attributes in each data amounted to 35. These attributes include date, plant stand, precipitation, temperature, hail, crop history, area of damage, severity, seed treatment, seed germination, plant height, leaves, leaf spot halo, leaf spot margin, leaf spot size, leaf shread, leaf malfunction, leaf mildew, stem, lodging, stem canker, canker lesion, fruiting bodies, external decay, mycellium, internal discoloration, sclerotia, fruit pods, fruit spots, seeds, mold growth, seed discolor, size of seeds, shriveling, and roots, can be seen in Table 1.

*Table 1. Attribute of Soybean Dataset*

| Number | Attribute | Data Type | Range and Attribute Information |
|---|---|---|---|
| 1. | date | Numeric | 0 : april, 1 : may, 2 : june, 3 : july, |
| 2. | plant stand | Numeric | 4 : august 5 : september, 6 : |
| 3. | precipitation | Numeric | october |
| | | | 0 : normal, 1 : less than |
| 4. | temperature | Numeric | normal |
| | | | 0 : less than normal, 1 : |
| 5. | hail | Numeric | normal, |
| 6. | crop history | Numeric | 2 : more than normal |
| | | | 0 : less than normal, 1 : |
| | | | normal, |
| | | | 2 : more than normal |
| 7. | area of damage | Numeric | 0 : yes, 1 : no |
| | | | 0 : different with previous |
| 8. | severity | Numeric | year, |
| 9. | seed treatment | Numeric | 1 : equal than previous year, |
| 10. | seed germination | Numeric | 2 : equal than two years ago, |
| 11. | plant height | Numeric | 3 : equal with several years |
| 12. | leaves | Numeric | ago |
| 13. | leaf spot halo | Numeric | 0 : spread, 1 : bottom area, 2 : top area, 3 : all area |
| 14. | leaf spot margin | Numeric | 0 : small, 1 : rather severe, 2 : severe |
| 15. | leaf spot size | Numeric | 0 : not exist, 1 : fungicide, 2 : other |
| 16. | leaf shread | Numeric | 0 :  90-100%, 1 : 80-89%, 2 : |
| 17. | leaf malfunction | Numeric | < 80% |
| 18. | leaf mildew | Numeric | 0 : normal, 1 : not normal |
| | | | 0 : normal, 1 : not normal |
| 19. | stem | Numeric | 0 : not exist, 1 : yellow circle, |
| 20. | lodging | Numeric | 2 : not yellow circle |
| 21. | stem canker | Numeric | 0 : small edge, 1 : not mall edge,<br>2 : not exist |
| 22. | canker lesion | Numeric | 0 : less than 1/8 inchi,<br>1 : greather than 1/8 inchi, 2 |
| 23. | fruiting bodies | Numeric | : no exist |
| 24. | external decay | Numeric | 0 : not exist, 1 : exist |
| 25. | mycellium | Numeric | 0 : not exist, 1 : exist |
| 26. | internal discoloration | Numeric | 0 : not exist, 1 : top surface, |
| 27. | sclerotia | Numeric | 2 : bottom surface |
| 28. | fruit pods | Numeric | 0 : normal, 1 : not normal |
| | | | 0 : yes, 1 : no |
| 29. | fruit spots | Numeric | |

| Number | Attribute | Data Type | Range and Attribute Information |
|---|---|---|---|
| | | | 0 : not exist, 1 : underground, |
| 30. | seeds | Numeric | 2 : above ground, 3 : above |
| 31. | mold growth | Numeric | the second node |
| 32. | seed discolor | Numeric | 0 : not exist, 1 : brown, 2 : |
| 33. | size of seeds | Numeric | blackish brown, 3 : dark |
| 34. | shriveling | Numeric | brown |
| 35. | roots | Numeric | 0 : not exist, 1 : exist |
| | | | 0 : not exist, 1 : hard and dry, 2 : wet<br>0 : not exist, 1 : exist<br>0 : no, 1 : brown, 2 : black<br>0 : not exist, 1 : exist<br>0 : normal, 1 : sick, 2 : little, 3 : not exist<br>0 : not exist, 1 : colourful, 2 : brown dark spots, 3 : dark brown spots,<br>4 : not exist<br>0 : normal, 1 : not normal<br>0 : not exist, 1 : exist<br>0 : not exist, 1 : exist<br>0 : normal, 1 : less than normal<br>0 : not exist, 1 : exist<br>0 : normal, 1 : rotten, 2 : cyst |

## 2.2 Preprocessing

The steps taken in preprocessing only consist of two stages, that is data selection and data transformation stages. Data Selection is used to retrieve relevant data for analysis. The relevant data are the attributes used. Data selection was carried out by two selections, that is irrelevant attribute selection and attribute selection using the Stepwise Regression Algorithm.

Stepwise regression can be done by the basic steps (algorithms) as follows [18]:

1. Determination of the correlation matrix between the dependent variable Y (results) on the independent variable (X1-Xn).
2. The independent variable that has the largest correlation coefficient with the dependent variable is the first variable that goes into the regression equation.
3. The next variable that enters equations is a variable (other than the one that was previously entered) which has the largest significant contribution to the regression equation. The value of the statistic F that must be surpassed by the independent variable is called F to enter.
4. When additional variables are included in the equation, evaluation of individual contributions for total of regression squares from other variables entered in the equation is calculated using the F test. If the value of F statistic is less than F to remove, then the variable is omitted from the regression equation.
5. Interpretation of the model obtained.

At the stage of data transformation, the process carried out is normalization. Normalization is one strategy in conducting data transformation [19]. Normalization is the process of scaling attribute values from data to a value in a certain range

.

## 2.3 Partition of Training Data and Test Data using K-Fold Cross Validation

All data from normalization, amount of 200 data will be divided into two, that is training data and test data. The division is done by the K-Fold Cross Validation method. The K-Fold cross validation method generally uses K = 10 [20]. Data is divided into 10 partitions, so there are 20 data on each partition. This 20 data consist of 2 data for each type of disease. In fold 1, the first partition or dataset 1 in which there are 20 data, will be used as test data. While the remaining 180 other data are used as training data.

© 2020 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)

## 2.4 Training

Training is carried out to recognize input patterns that lead to a class / target. Training is running on training data that generated from the data partition process. Training process flowcharts with the LVQ algorithm are shown in Figure 2.

The difference in LVQ training and LVQ2 training lies in the weight update process. In LVQ, the weight that is updated is the weight with the smallest Euclidean Distance (winner). Whereas in LVQ2 training, there are two weights to be updated, that is the smallest weight with Euclidean Distance (winner) and the second smallest (runner up) if it meets the requirements. The training flowchart with the LVQ2 algorithm is shown in Figure 3.

## 2.5 Testing

LVQ and LVQ2 testing process is done by finding the smallest distance between the weight of the training results to the test data using Euclidean Distance. All training result weights LVQ and LVQ2 in each class will find the Euclidean Distance value from the test data.

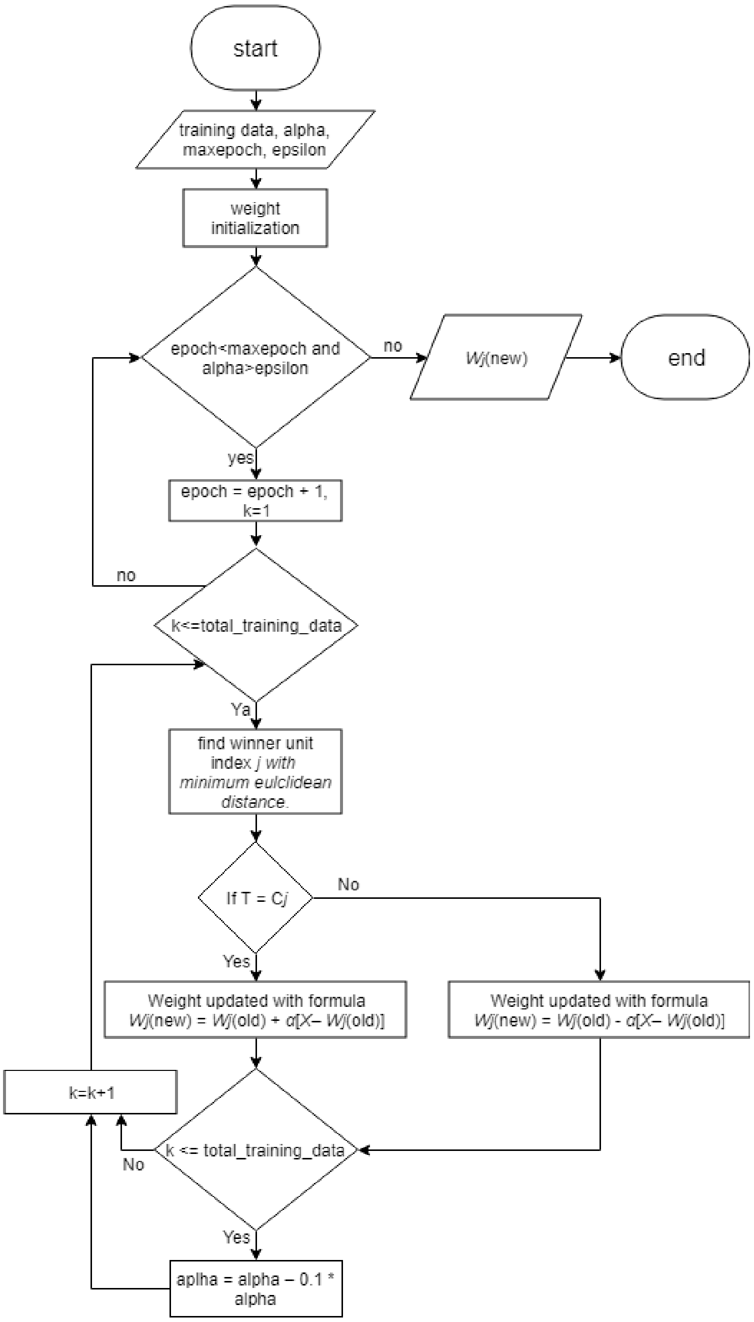
start
training data, alpha, maxepoch, epsilon
weight initialization
epoch<maxepoch and alpha>epsilon
no
Wj(new)
end
yes
epoch = epoch + 1, k=1
no
k<=total_training_data
Ya
find winner unit index j with minimum eulclidean distance.
If T = Cj
No
Yes
Weight updated with formula Wj(new) = Wj(old) + α[X– Wj(old)]
Weight updated with formula Wj(new) = Wj(old) - α[X– Wj(old)]
k=k+1
k <= total_training_data
No
Yes
aplha = alpha – 0.1 * alpha

*Figure 2. Training Process Flowchart in LVQ*

*Cite: Muhamad, N., Endah, S., Sarwoko, E., & Sasongko, P. (2020). Soybean Disease Detection with Feature Selection Using Stepwise Regression Algorithm: LVQ vs LVQ2. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(2). doi:https://doi.org/10.22219/kinetik.v5i2.919*

The value of all Euclidean Distance is then compared to getting the smallest value which means it is the class/target closest to the test data.

### 2.6 Evaluation

Evaluation was conducted to find out information about accuracy, error rate, sensitivity, and specificity of the test results from confusion matrix. The evaluation of both LVQ and LVQ2 methods was calculated all to be compared. The best evaluation method is used in the detection process.

## 3. Results and Discussion

### 3.1 Test Scenario

The test scenario aims to test the level of accuracy, error, sensitivity, and specificity of LVQ and LVQ2 neural network methods along with the applied stepwise regression algorithm, to detect soybean disease. This study uses 4 test scenarios with each parameter can be seen in Table 2.

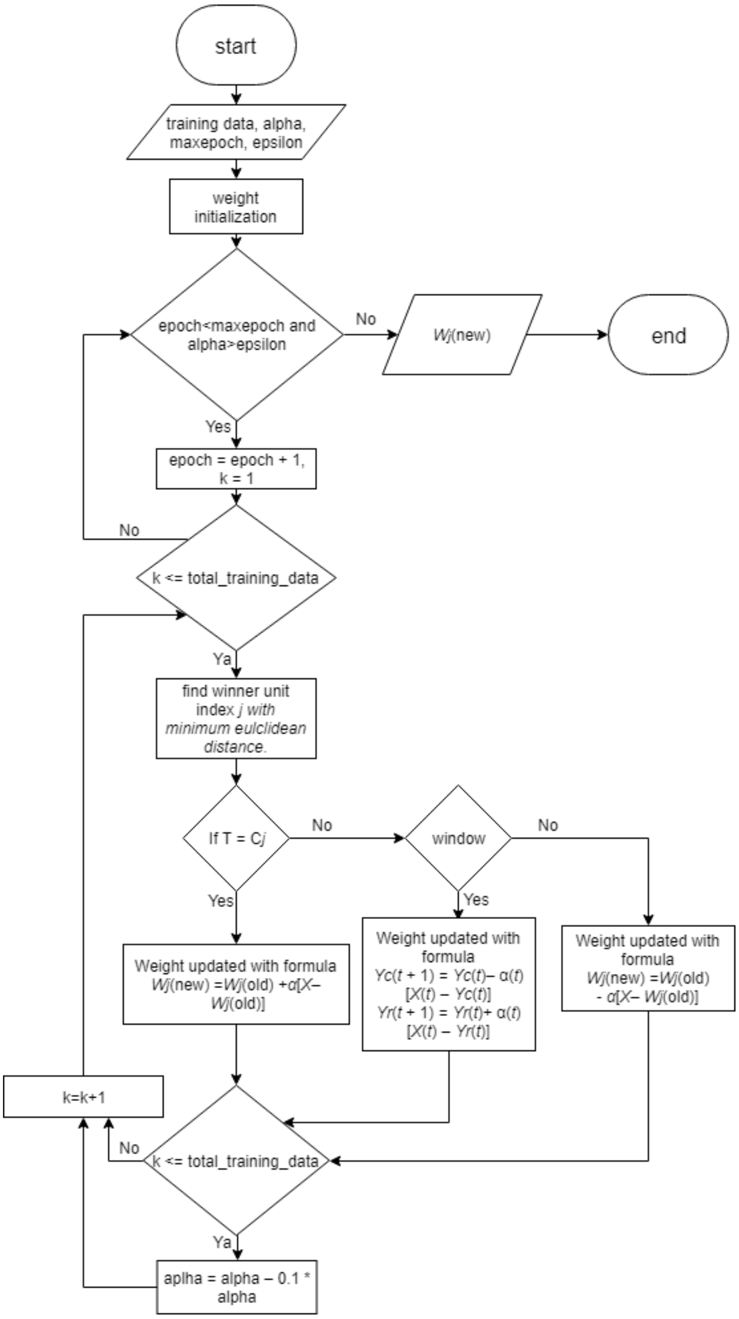
start
training data, alpha, maxepoch, epsilon
weight initialization
epoch<maxepoch and alpha>epsilon
No
Wj(new)
end
Yes
epoch = epoch + 1, k = 1
No
k <= total_training_data
Ya
find winner unit index j with minimum eulclidean distance.
If T = Cj
No
window
No
Yes
Yes
Weight updated with formula Wj(new) =Wj(old) +α[X– Wj(old)]
Weight updated with formula Yc(t + 1) = Yc(t)– α(t) [X(t) – Yc(t)] Yr(t + 1) = Yr(t)+ α(t) [X(t) – Yr(t)]
Weight updated with formula Wj(new) =Wj(old) - α[X– Wj(old)]
k=k+1
No
k <= total_training_data
Ya
aplha = alpha – 0.1 * alpha

*Figure 3. Training Process Flowchart in LVQ2*

© 2020 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)

*Table 2. Test Scenario*

| Scenario | Stepwise Regression | Learning rate (α) | Epsilon | Method |
|---|---|---|---|---|
| 1 | none | 0.1, 0.15, 0.2, 0.25, 0.3, 0.35 | 0.04, 0.01, 0.001 | LVQ, LVQ2 |
| 2 | p to enter = 0.15<br>p to remove = 0.15 | 0.1, 0.15, 0.2, 0.25, 0.3, 0.35 | 0.04, 0.01, 0.001 | LVQ, LVQ2 |
| 3 | p to enter = 0.15<br>p to remove = 0.15 | 0.1, 0.15, 0.2, 0.25, 0.3, 0.35 | 0.04, 0.01, 0.001 | LVQ, LVQ2 |
| 4 | p to enter = 0.15<br>p to remove = 0.15 | 0.1, 0.15, 0.2, 0.25, 0.3, 0.35 | 0.04, 0.01, 0.001 | LVQ, LVQ2 |

1. Scenario 1: Scenario 1 is done to determine the performance of the LVQ and LVQ2 algorithms on the soybean crop dataset with all attributes. In this scenario the Stepwise Regression algorithm is not applied
2. Scenario 2: Scenario 2 is done to look for the performance of the LVQ and LVQ2 algorithms on soybean crop datasets with the Stepwise Regression algorithm selection attribute at p to enter = 0.15, p to remove = 0.15. This selection produces 17 selected attributes.
3. Scenario 3: Scenario 3 is done to determine the performance of the LVQ and LVQ2 algorithms on soybean dataset with the Stepwise Regression algorithm selected attributes at p to enter = 0.2, p to remove = 0.2. This selection produces 19 selected attributes.
4. Scenario 4: Scenario 4 is done to determine the performance of the LVQ and LVQ2 algorithms on soybean crop datasets with the selection attributes of the Stepwise Regression algorithm at p to enter = 0.05, p to remove = 0.05. This selection produces 15 selected attributes.

### 3.2 Experiment Result

1. Scenario 1: Based on Figure 4 Scenario 1 without attribute selection get the best accuracy results of each method with amount of 88.5% on LVQ and 88% for LVQ2 method.

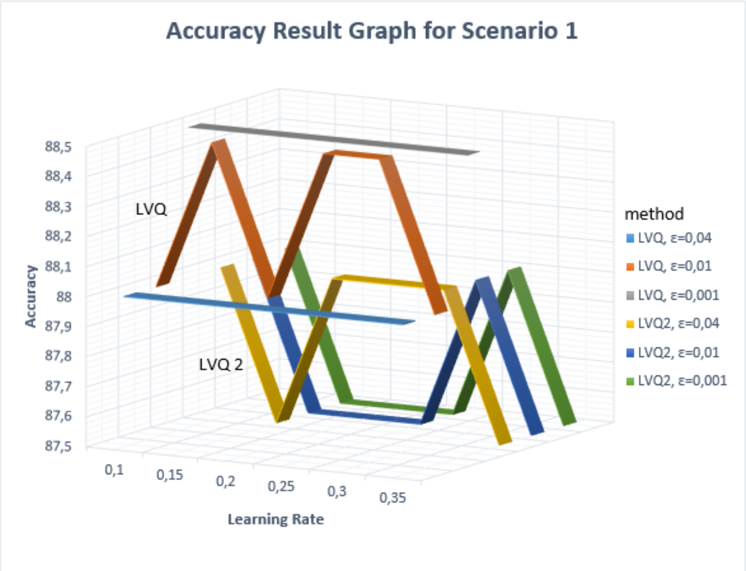

*Figure 4. Accuracy Result Graph for Scenario 1*

2. Scenario 2: Based on Figure 5 Scenario 2 with 17 selected attributes getting the best accuracy results of each method with amount of 86.5% on LVQ and 90.5% for LVQ2 method.

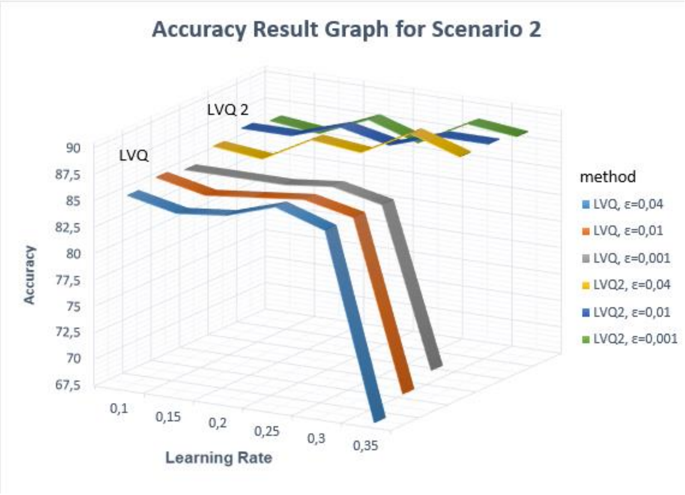

*Figure 5. Accuracy Result Graph for Scenario 2*

3. Scenario 3: Based on Figure 6 Scenario 3 with 19 selected attributes get the best accuracy results of each method with amount of 86.5% on LVQ and 89.5% for LVQ2 method.

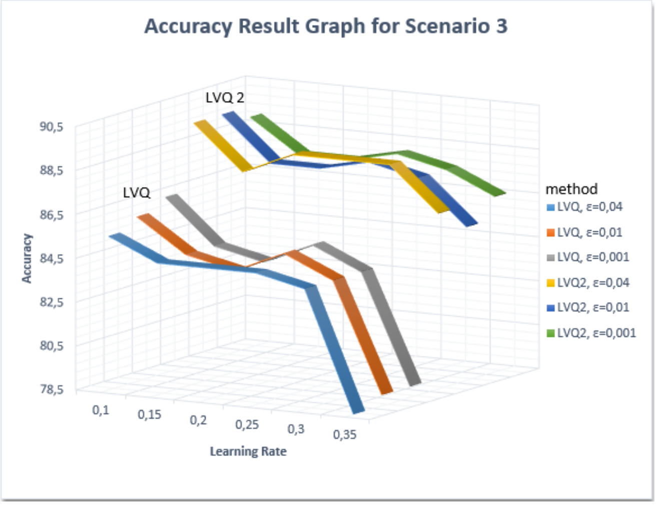

*Figure 6. Accuracy Result Graph for Scenario 3*

4. Scenario 4: Based on Figure 7 Scenario 4 with 15 selected attributes get the best accuracy results of each method with amount of 86% on LVQ and 89% for LVQ2 method.

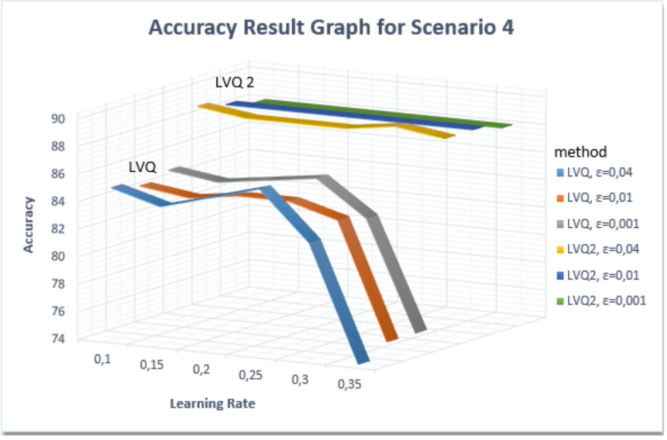

*Figure 7. Accuracy Result Graph for Scenario 4*

© 2020 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)

### 3.3 Result analysis

All of the best accuracy from each scenarios are compared. This best accuracy includes the best accuracy results in each method. The results can be presented in graphical form in Figure 8.

From the graph, it can be seen that the LVQ2 method is a better method than the LVQ method. This can be seen from the overall results of test scenarios which shows increased accuracy when the LVQ2 method is used.

In addition, LVQ2 also shows better performance when used together with the Stepwise Regression Algorithm attribute selection. This is very important because usually the attribute selection process will cause a decrease in performance, as shown in Figure 8, where the performance of the LVQ method decreases when the stepwise regression attribute selection method is applied in scenarios 2, 3 and 4. So LVQ2 is a better method, because it can improve classification performance with the number of attributes that have been reduced or selected. This is very important because the fewer attributes are used, the less resources must be spent.

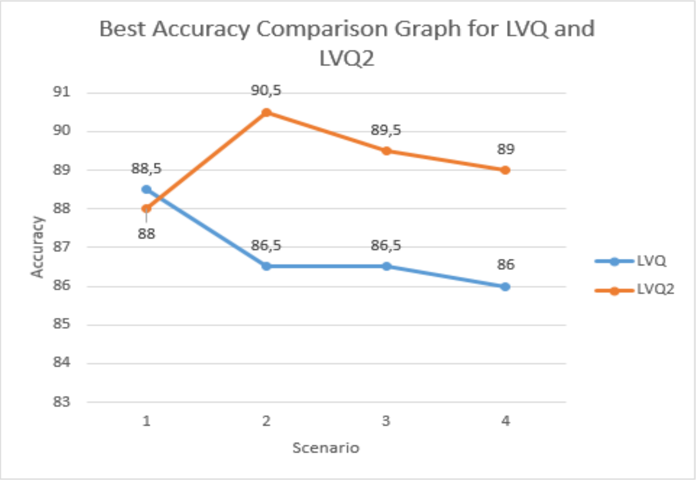

*Figure 8. Best Accuracy Comparison Graph for LVQ and LVQ2*

In this study, the best results lie in scenario 2. This scenario has the input parameters of the LVQ2 method as follows:

a. p to enter, p to remove : 0.15
b. Learning rate (α) : 0.3
c. Epsilon : 0.04
d. Maximum epoch : 100

These parameters produce an accuracy of 90.5%, 9.5% error rate, sensitivity 90.5% and specificity 98.94%. This shows that the system has good capability to detect soybean disease.

## 4. Conclusion

Changes in the value of the learning rate (α) and epsilon affect the results of accuracy, error rate, sensitivity, and specificity in the detection of soybean disease. In this study the two values are interconnected so that they must be combine together. The combination of the two parameters when the value of α = 0.3 and epsilon = 0.04 is a combination with the best accuracy of 90.5%.

Without the Stepwise Regression Algorithm attribute selection, the accuracy, sensitivity, and specificity of the LVQ method are still slightly better than the LVQ2 method. This can be seen in scenario 1 which produces the best accuracy of the LVQ method of 88.5% and LVQ2 of 88%.

Improved accuracy, sensitivity and specificity occur when the attribute selection Stepwise Regression Algorithm is used together with the LVQ2 method. This can be seen in scenarios 2, 3 and 4 which produce the best LVQ2 accuracy of 90.5%, 89.5%, 89%.

Scenario 2 produced best accuracy results using Stepwise Regression Algorithm attribute selection (p to enter = 0.15, p to remove = 0.15) which obtain 17 selected attributes, as follows: date, plant stand, precipitation, leaves, leaf spot leaf spot halo, leaf spot margin, leaf spots size, leaf mildew, stem canker, fruiting bodies, external decay, fruit pods, fruit spot, seeds, mold growth, seed discolor, and roots.

## References

[1] Mapegau, “Pengaruh Cekaman Air Terhadap Pertumbuhan Dan Hasil Tanaman Kedelai (Glycine Max L. Merr),” No. 2001, Pp. 43–51, 2006.
[2] F. Y. Wicaksono, A. W. Irwan, and R. Fitriani, “Response of soybean (Glycine max) var. Wilis due to application of N, P, K and guano fertilizer dosages on Inceptisols Jatinangor,” *J. Kultiv.*, Vol. 16, No. 2, Pp. 333–339, 2017. https://doi.org/10.24198/kultivasi.v16i2.13223

[3] BPS, “Produksi Kedelai Menurut Provinsi Tahun 2014-2018,” 2018.

[4] S. Sulandari, S. Hartono, Y. M. S. Maryudani, and Y. B. Paradisa, “Detection and Distribution of Soybean Mosaic Virus (SMV) and Soybean Stunt Virus (SSV) at Soybean Production Centers in Indonesia,” J. Perlindungan Tanam. Indones., Vol. 18, No. 2, Pp. 71–78, 2014.

[5] H. Semangun, “Penyakit-Penyakit Tanaman Pangan Di Indonesia,” *Gadjah Mada Univ. Press*, 2004.

[6] P. Dongre and T. Verma, “A Survey of Identification of Soybean Crop Diseases,” *Int. J. Adv. Res. Comput. Eng. Technol.*, Vol. 1, No. 8, Pp. 361–364, 2012.

[7] A. Susanti, M. Faizah, M. Lutfi, and S. Khamid, “Penekanan Penyakit Karat Daun Pada Kedelai Akibat Phakopsora pachyrhizi Syd . Menggunakan Mikoriza Indigenous Pada Tanah Litosol,” Vol. 2, No. 1, Pp. 23–31, 2018.

[8] D. Apriyanto, O. Hendra, and A. Mulyadi, “Penampilan Penggerek Polong Kedelai , Etiella zinckenella Treitschke ( Lepidoptera : Pyralidae ), dan Pemilihan Inang pada Kedelai dan Kacang Tanah Performance of Soybean Pod Borer , Etiella zinckenella Treitschke ( Lepidoptera : Pyralidae ), and Host Preference on Soybean and Groundnut,” Vol. 12, No. 1, Pp. 62–67, 2009.

[9] W. Batchelor, “Development of a Neural Network for Soybean Rust Epidemics,” *Trans. ASAE*, Vol. 40, No. 1, Pp. 247–252, 1997. https://doi.org/10.13031/2013.21237

[10] K. S. Kim, T. C. Wang, and X. B. Yang, “Simulation of Apparent Infection Rate to Predict Severity of Soybean Rust Using a Fuzzy Logic System,” *Phytopathology*, Pp. 1122–1131, 2005. https://doi.org/10.1094/PHYTO-95-1122

[11] I. Rahmon, Adebola_Akinsanya, and M. . Eze, “A Neuro-Fuzzy System For Diagnosis of Soya-Beans Diseases,” *Res. J. Math. Comput. Sci.*, 2018. https://doi.org/10.28933/rjcms-2018-04-0501

[12] Nopiyanti, E. Tita, and A. Maesya, “Sistem Identifikasi Penyakit Tanaman Kacang Kedelai Menggunakan Metode Naive Bayes.,” 2015.

[13] Joshi, Sneha, and M. Borse, “Detection and Prediction of Diabetes Mellitus Using Back-Propagation Neural Network,” in *International Conference on Micro-Electronics and Telecommunication Engineering, ICMETE 2016*, Pp. 110–113, 2017. https://doi.org/10.1109/ICMETE.2016.11

[14] Budianita, Elvia, and M. Firdaus, “Diagnosis Penyakit Kejiwaan Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantization2 (LVQ 2) (Studi Kasus : Rumah Sakit Jiwa Tampan Pekanbaru),” Vol. 13, No. 2, Pp. 146–50, 2016.

[15] Rahadian, B. Aristyo, C. Dewi, and B. Rahayudi, “The Performance of Genetic Algorithm Learning Vector Quantization 2 Neural Network on Identification of the Types of Attention Deficit Hyperactivity Disorder,” in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, Pp. 337–41, 2017. https://doi.org/10.1109/SIET.2017.8304160

[16] Maulana, Fadil, and S. N. Endah, “Comparison Selection of Attributes in Preprocessing Data for Diagnosis of Diabetes,” in *1st International Conference on Informatics and Computational Sciences, ICICoS 2017*, Pp. 141–46, 2018. https://doi.org/10.1109/ICICOS.2017.8276352

[17] Michalski, Ryszard S, and R. L Chilausky, “Learning by Being Told and Learning from Examples: An Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis,” Int. J. Policy Anal. Inf. Syst., Vol. 4, No. 2, Pp. 125–161, 1980.

[18] Hanke, John E., and D. W. Wichern, Business Forecasting Ninth Edition. New Jersey: Pearson, 2008.

[19] Han, Jiawei, M. Kamber, and J. Pei, Data Mining Concept and Techniques. Waltham: Morgan Kaufmann, 2011.

[20] Refaeilzadeh, Payam, L. Tang, and H. Lu, “Cross-Validation.” 2008.

© 2020 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)