Journal of Theoretical and Applied Information Technology
31st October 2022. Vol.100. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANALYSIS OF THE EFFECT OF FEATURE SELECTION ON THE HANDWRITING AUTHENTICITY CHECKING SYSTEM THROUGH FISHER SCORE AND LEARNING VECTOR QUANTIZATION

DIAN PRATIWI[1], SYAFUDIN[2], TRUBUS RAHARDIANSYAH[3], MUHAMMAD ICHSAN[4], STEVEN SEN[5], AHMAD FAUZY[6]

[1.2.4.5.6]Department of Informatics Engineering, Universitas Trisakti, Indonesia

[3]Department of Law. Universitas Trisakti, Indonesia

E-mail: [1]dian.pratiwi@trisakti.ac.id, [2]fudin@trisakti.ac.id, [3]trubus@trisakti.ac.id, [4]m.ichsan064001600023@std.trisakti.ac.id, [5]steven064001600022@std.trisakti.ac.id, 064002100031@std.trisakti.ac.id

ABSTRACT

Handwriting is part of the authentic form that everyone has. The authenticity of the owner is often difficult to prove because it is easier to fake with the support of technology. This research was then carried out to develop a method to help verify the authenticity of the owner's handwriting, by applying the LVQ method and selecting the Fisher Score feature to analyze the use of the best number and types of features. The steps taken include processing the grayscale color and data size into 500x200 pixels, then segmentation, taking the features of the writing area based on the mean, variance, entropy, energy and contrast values. The feature is then analyzed by Fisher Score for each use of the feature with the highest score, and tested with LVQ to predict the owner of the article. From the test results, it is found that the use of a large number of features in the order of Fisher's highest score, the quality of accuracy is better than the use of a small number of features

Keywords: *Contrast, Energy, Indexing, Mean, Scoring*

1. INTRODUCTION

The development of the Industry 4.0 era which leads to automation and machine-based systems, such as robotics, Internet of Things (IoT), Big Data, and artificial intelligence. Adaptation in almost every field leads to the transformation of smarter management, by leveraging digital and mobile technologies.

Developments in handwriting analysis are also increasingly having a broad scope of use, including employee recruitment, medical diagnosis in psychology, forensics, psychology, and interactions between humans and computers. Although handwriting is often doubted to be able to play a role in forensic authentication or assessing a person's personality because it is still included in pseudoscience, several studies have shown the opposite result.

Previous research that studies and examines handwriting analysis such as that conducted by Aqab, S. and Tarig, M.U. by applying the Artificial Intelligence Deep Neural Network and Image Processing methods to recognize each letter character written in handwriting from printed documents or digital devices. To improve data quality, they used techniques of reducing noise, increasing contrast, binarization, normalization, and skew correction. Based on this research, it was found that the ability of AI has an accuracy of 83 to 93% in interpreting the writing in question [1].

Research from Samsuryadi et al in 2021 showed a comparison of the use of various clustering methods such as Fuzzy C-Means, ANN, SVM, Rule Base System to analyze a person's characteristics from handwriting. The conclusions obtained from this study include the characteristics of handwriting that are able to represent human characters seen from the baseline, margin, slant, letter size, pen pressure, spacing, speed of writing. In addition, the application of more methods has better results, namely deep learning methods such as LSTM, CNN, and RNN [2].

Then there is Jasril's research which uses the same technique, namely LVQ with the aim of classifying images of beef and pork. Jasril also applies a segmentation procedure first in the form of Fuzzy C-Means to obtain the meat area, then extracts the features in the form of GLCM values. The results of the trial showed that the LVQ was able to distinguish between beef and pork with an accuracy of 91.67% [3].

The researchers themselves have actually conducted research on handwriting analysis since 2016, where its application is for digital handwriting security in the process of validating document owners [4][5][6][7][8] and also to analyze personality and psychological conditions. someone [9]. The method applied is LVQ and its features are statistical-based texture features with a precision value of 75.17% and recall value and accuracy of 67.5% from 80 data [4], while for the Uniform-Quantization method [5] and Inner Product [6][7] [8] got an accuracy of less than 67% From all these studies, there are test results that show some of the extracted features are NaN (Not a Number), which can affect the system's computing process and the final level of accuracy. This value needs to be detected early by making a selection so that it is not used as a feature to recognize handwriting.

The difference that can be seen from this study is the use of a method in the form of LVQ where the features to be computed are selected first through the Fisher Score. The purpose of this study is to analyze the ability and effect of the application of scoring in its implementation on the recognition of the authenticity of the owner's handwriting

Fisher's score is a selection method or the selection of supervised feature values on the basis of filtering through their weights. This model selection has good feature relevance in supervised learning, which can reduce computational load, achieve higher accuracy and can efficiently reduce space-time complexity [10]. The use of Fisher score is also growing as a feature selection technology in recent years, even by Hamraz, et al. applied to gene classification [11]. Then the ability of the Fisher score can also be implemented for multiclassification and the average value of the scores can be used as a threshold value in the selection of features in the case of identification of malignant breast lesions [12]. While in the research of Sun, L. et al.[13] The use of the Fisher score is also for feature selection in the binary value classification based on the highest score, but does not consider characteristics or correlations between features. Unlike Lim et al. [14], who pay attention to the correlation between features in learning with multilabel data sets, resulting in better classification performance.

n this study, the use of the Fisher score will be applied to the selection of the use of texture features with the highest score scheme selected and the accuracy results analyzed in LVQ learning, to get the best recognition rate..

2. METHODOLOGY

The approach taken by this research starts from the acquisition of raw data in the form of printed documents, then scans and processes it at the pre-processing stage, extracts statistical-based texture feature values, sorts scores and selects features based on the highest fisher value, and finally predicts the owner of the writing by LVQ..

.

2.1 Data Acquisition and Pre-processing

This procedure is an initial stage carried out with the aim of improving or improving the quality of the data. In order for the final results obtained to achieve the best, it is necessary to collect handwritten data with sufficient amount and accuracy in the selection of pre-processing techniques.

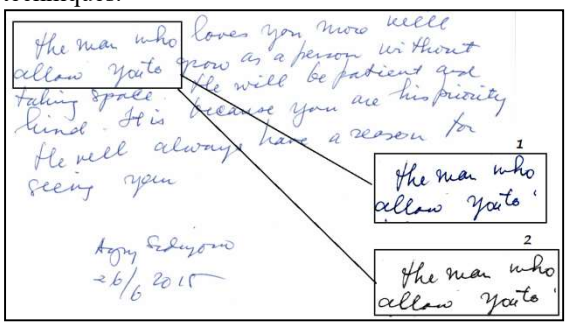

Figure 1: Contrast Enhancement (1) and Greyscale Color Conversion

Figure 1 shows the results of the pre-processing procedures carried out in this research after data acquisition (scanned documents and resized to 500x200 pixels), in the form of contrast enhancement and greyscale color conversion. Contrast enhancement will change the color with the light intensity will be brighter, and the dark intensity will be darker. While in the color conversion, the RGB value (24 bits) of each pixel will be averaged into a greyscale value (8 bits) with a color interval of 0-255. Based on the research of Soe et al. [18], the impact of applying color enhancement, especially on the contrast value, will prevent miss-classification and over-segmentation.

Journal of Theoretical and Applied Information Technology
31st October 2022. Vol.100. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Segmentation is an important step before performing feature extraction or pattern recognition. Segmentation is usually done on data that has areas or shape structures that are difficult to analyze or distinguish.

According to Karthick and Sathiyasekar [24], there are 3 segmentation approaches, namely region based methods, edge based, and Hybrid. The three approaches have their respective advantages and disadvantages. In region based, the theory is simple to implement, only needs a little seed point number to show the property we want, and produces areas with clearer boundaries. On the edge based, it is suitable for separating areas that are oriented in color and shape, but difficult for irregular surfaces (high texture). In this study, segmentation was carried out based on region based where the separation of the handwriting background area was limited to a certain threshold value through the global thresholding method.

Based on research by Bhargavi and Jyothi [25], global thresholding is an appropriate segmentation method used on data that has a fairly different distribution of object intensity and background pixels, and this segmentation increases the system's ability to detect, recognize and measure features.

2.2 Feature Extraction

Feature extraction is a step that is generally carried out to determine the characteristics or important information of a data to be recognized. The patterns that will be studied by a system need to be known in advance for their characteristics so as to facilitate and speed up the computation. Feature extraction is also useful for reducing the number of features in a data set, where the extracted features can represent other feature information so that unused features can be discarded. The new feature set can summarize most of the information contained in the original or original feature.

The existing method of feature extraction is selected based on the type of data used. In image data, extracted features can be in the form of color, shape, and texture values. The approach can be chosen through local extraction or global extraction.

The following is a formula for some statistical-based texture features (global features) that can be used based on Haralick's reference [19]:

Table 1: Statistical Features

Feature	Equation
Variance	$\sigma^2 = \sum_n (f_n - \mu)^2 P(f_n)$
Entropy	$H = -\sum_n p(f_n).Log_2 P(f_n)$
Energy	$\mu = \left(\frac{1}{MN}\right) * \sum_{i=1} \sum_{j=1} \mathrm{p(i,j)}$
Contrast	$f_2 = \sum_{n=0}^{N_{g-1}} n^2 \left\{ \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} P_{d,\theta}(i,j) \right\}$, where n $= \lvert i - j \rvert$
Greyscale	(R+G+B)/3

Contrast is a measure of the local variations present in an image. If there is a large amount of variation in an image, the contrast will be high. In this study, the variance value is used to analyze how many variations of words there are in a document. Entropy feature is a feature that will show the irregularity of a pattern. Patterns that are increasingly irregular or have many deviations will have high entropy values [4]. The energy feature will provide an overview of the maximum points of the pattern that occur at the edges. In the study of Mangala and Cyril [20], the energy feature is used to detect emotions along with the use of a neural network which produces an optimal accuracy rate of 86%. In this study, the energy feature is used to analyze how much pressure is in writing, where high-value energy has a strong or deep pattern of writing emphasis.

2.3 Fisher Scoring

Fisher Score (FS or FScore) is a technique for selecting monitored features that is quite widely used, because it has reliable performance. However, FS selects independently according to the score obtained.

Broadly speaking, feature selection can be done with three approaches, namely filter-based, wrapper-based, embedded [15]. The filter-based method ranks features as a pre-processing stage prior to learning, and the feature is selected with the highest score (eg. FScore). While in the wrapper-based method, features are selected based on the learning outcomes used. Then the embedded method, specifically feature selection is combined with learning algorithms which are usually to limit the learning procedures. The Fisher score equation according to Niyas and Thiyagarajan is as follows [16]:

$$BCD_i = \sum_{n=1}^{L} (Overallmean - Meanlabel_i)^2$$

$$WCD_i = \sum_{n=1}^{L} \text{Variancelabel}_i$$

$$\text{FS}_i = \frac{BCD_i}{WCD_i}$$

Where the overallmean is the average value of all the features used, then the mean label is the average feature of the feature being scored, while the variance label is the variance value of the features that correspond to each label or feature, and L is the number of classes. Broadly speaking, the Fisher score metric depends on the mean and variance of the features that fit into the label. The features will be ranked in the order of increasing FS (highest score ranked first, feature with second highest score ranked second, and so on). According to the research of Gu, et al. [17] because the score of each feature is calculated independently through the FS, the selected feature automatically follows a heuristic approach which means it is suboptimal. However, if these low-scoring features are combined, they will tend to have a higher accuracy rate.

2.4 LVQ

Learning Vector Quantization (LVQ) is a method that implements supervised learning and is part of the branch of artificial neural network science. The basic LVQ (LVQ1) has several variations, namely LVQ2, LVQ2.1, and LVQ3. The characteristic that LVQ1 has is that only the closest reference vector (winning vector) to the input vector is updated. The direction of the vector displacement depends on whether the reference vector is of the same class as the input vector. The LVQ network has a target to achieve. The competitive layer learns to recognize and classify input vectors. If there are nearly identical vectors, then the competitive layer will place them both in the same class. So LVQ learns to classify the input vector into the specified target class.

Here is the algorithm of the LVQ training stage [21]:

1. Determine the initial weight (W) and LVQ parameters, namely Epoch (maxEpoch), learning rate (α), and the minimum expected error value (min α)
2. Enter the input data (X) and target class (T).
3. Set initial condition: epoch = 0
4. Do it if: (epoch < maxEpoch) and (α≥ min α).
   a. epoch = epoch+1.
   b. Determine J such that ||Xi-Wj|| is minimal using the Euclidean distance formula. D(j) = (Wij - xi)^2
   c. Justify Wj with the following conditions:
      If T = Cj then
      Wj(t + 1) = wj (t) (t)[x(t) – wj]
      If T Cj then
      Wj(t + 1) = wj(t) + (t)[x(t)– wj(t)]
   d. Subtract the value of α
5. Stop the condition test with optimal weight output.

After training, the final weights (W) will be obtained which will later be used in the testing process. The testing algorithm is as follows:

1. Enter the data to be tested, for example Xi,j with i= 1., 2,…. np, and j=1, 2,…m
2. Do it for i=1 to np
   a. Determine J such that ||Xij – Wij|| minimum
   b. J is a class for Xi.

3. RESULT AND DISCUSSION

The data used in this research is the same data in previous studies, namely 45 written documents from 15 authors used for the training stage and 25 documents from 5 authors used for the final testing stage. The total number of documents tested were 75 documents, of which the documents were collected from 15 authors, each of whom had 5 documents.

In its implementation, this system is built on a mobile basis with added libraries using python. For the training stage, the parameters used in LVQ are: learning rate 0.005, epoch 200, 15 classes, input as much as 45, initialization of random weight. As for the testing phase, the only different parameter is the number of inputs used, which is 25.

The following is an example of the results of feature values 5 from a total of 75 extracted documents, before selecting the FScore feature:

Table 2: Document Extraction Feature Value

Feature	Document				
	1	2	3	4	5
Energy	0.536	0.103	0.175	0.935	0.131
Variance	7.27	3.13	6.57	6.58	3.36
Contrast	0.211	0.240	0.238	0.235	0.243
Greyscale	211	240	238	235	243
Entropy	0	0.92	0.1	0.8	0.95

Journal of Theoretical and Applied Information Technology
31st October 2022. Vol.100. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Meanwhile, the following is a display of the mobile system that applies FScore and LVQ in this study:

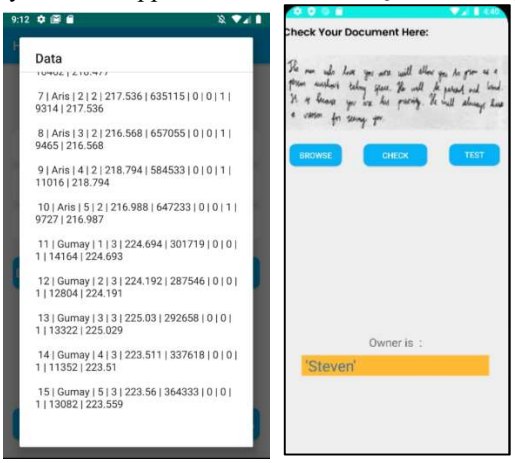

Figure 1: Feature Extraction (left) and Handwritten Analysis Result (right)

The training and LVQ testing stages are carried out with the python framework, after the feature extraction process is processed using java android mobile. As for the data transfer procedure, researchers used JSON, the same as in previous studies [4]. The difference is that at the stage before data transfer, there is a feature selection that is applied to the values of entropy, contrast, grayscale, energy, and variance through the Fisher score. The selected feature trial scheme also consists of 2 approaches, namely reducing the number of features used one by one based on the smallest feature score and using a threshold value based on the calculation of the average score to determine the number of selected features. If the score owned by the feature is below the score threshold, then the feature is not used.

At the training stage, the weights stored from the LVQ learning process have the same variation following the number of features used after selection through Fisher scoring.

The following is the Fisher Score value at the training stage and the results of the learning accuracy:

Table 3: Training Result of LVQ

Feature	Rank	Score	Accuracy(%)				
Energy	1	17.314	68	62	42	44	64
Variance	2	6.683					
Contrast	3	4.9051					
Greyscale	4	4.905					
Entropy	5	3.467					

From table 3 above, the average Fisher score is 7.4548. The percentage of training accuracy obtained when using the second selection approach that uses an average Fisher score of 44%, where there is only one feature that meets the requirements above the threshold value. Then when analyzed further, the use of the Fisher score can also reduce the use of the number of features that need to be used, thus automatically reducing the computational burden when learning writing patterns. However, this reduction does not always result in better accuracy, because more and more features are removed, it will reduce accuracy. This can be seen when using 4 features (68%), 3 features (62%), 2 features (42%), and 1 feature (44%). However, if we look back at the training results, the single energy feature produces better accuracy than the use of variance along with energy. It can be analyzed that the variance feature in this case does not represent much important information in the handwritten data, in contrast to the energy feature. The variance value in this case may contain NaN values, which in previous studies [4] appeared so that it reduced accuracy when used. From this 45 data training scheme, it can be concluded that the maximum accuracy achieved is 68% when the features used are grayscale, contrast, variance, and energy values (selecting 1 feature). The weights generated from training the pattern of these four features are then used in the testing phase. The following is a table of test results from 25 documents:

Table 4: Testing Result of LVQ

Feature	Rank	Score	Accuracy(%)				
Energy	1	17.314	60	44	28	32	52
Variance	2	6.683					
Contrast	3	4.9051					
Greyscale	4	4.905					
Entropy	5	3.467					

Based on table 4, the best percentage value is shown when the LVQ network uses 4 features with the highest Fisher score sequence value of 60%, the same as the training stage. Selection of too many features will further reduce the accuracy of the system because it is possible that important information is also lost. This can be seen in the accuracy when using 1 feature (Energy) reaches 32%, 2 features (energy and variance) reaches 28%, 3 features (energy, variance contrast, and grayscale) reaches 44%. The removal of these features in pattern recognition does not make the system work better, because the accuracy obtained is still greater when using all of the features (accuracy reaches 52%). However, if observed again, this accuracy can be further improved if the weights used are

generated from training with a higher level of accuracy. This of course requires trial and error with different schemes, such as using a larger amount of training data so that the LVQ network can be smarter in recognizing patterns [22], trial and error using different LVQ parameters to find the most optimal results, or combining LVQ methods. with Backpropagation weighting as done by Tse, et al. [23]. In addition, it can also try to use other learning methods such as SVM conducted by Saragih and Rustam [26] in processing the features of the Fisher score to predict disease resistant gene in rice selection, which is able to produce 90% accuracy at low data dimensions.

4. CONCLUSION

The score value generated by FScore from the results of this study is able to reduce the presence of NaN in the selected features to be used. This does not affect the increase in accuracy, but the computational process that is passed after using FScore can prevent errors from appearing during pattern recognition through LVQ because values that cannot be defined are not counted. In addition, the application of FScore is able to show the ability of each feature to represent information on pattern recognition. So in the selection of features that are not used, it can refer to the results of the score. However, this capability must also be supported in the selection of feature extraction techniques and the accuracy of the pre-processing method, so that the features used are appropriate and able to represent the overall information contained in the data. Because the percentage shown between using FScore with more features is still better than using fewer features. So that in making the decision to reduce features from the FScore results, it also requires an analysis of the use of features one by one to ensure the accuracy of using the most optimal number of features.

ACKNOWLEDGMENT

The author would like to thank the Directorate of Resources-Directorate General of Higher Education-Ministry of Education, Culture, Research and Technology, also to Trisakti University.

REFERENCES:

[1] Aqab, S. and Tariq, M.U., "Handwriting Recognition using Artificial Intelligence Neural Network", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol.11, No.7, 2020, pp. 137-146

[2] Samsuryadi, Kurniawan, R., Mohamad, F.S., "Automated Handwriting Analysis based on Pattern Recognition: A Survey", *Indonesian Journal of Electrical Engineering and Computer Science,* Vol. 22, No. 1, April 2021, pp. 196-206.

[3] Jasril and Sanjaya, S., "Learning Vector Quantization 3 and Spasial Fuzzy C-Means for Beef and Pork Image Classification", *Indonesian Journal of Artificial Intelligence and Data Mining*, Vol.1, No. 2, 2018, pp. 60-65.

[4] Pratiwi, D., Syaifudin, Rahardiansyah, T., Gunawan, M. I., Sen, S., Pratama, D. A., "Handwriting Prediction with Learning Vector Quantization Method in Mobile Application", *Journal of Theoretical and Applied Information Technology*, Vol. 99, No. 24, 2021, pp. 5850-5856.

[5] Pratiwi, D., Syaifudin, Rahardiansyah, T., Anggraini, W., Hilman, A., Chairunnisa, N., "Texture Traits with Uniform-Quantization in Handwriting Documents for Digital Forensics Investigation". I*OP Conference Series: Material Science and Engineering*, 645(1), 2019.

[6] Davin, R.P., Pratiwi, D., Syaifudin, Rahardiansyah, T., Rizky, D.L.P, "Implementation of Inner Product to Analyze Digital Handwriting based on Texture Traits", *ACM International Conference Proceeding Series*, 2017, pp. 114-118.

[7] Syaifudin, Pratiwi, D., Trubus, "Handwritten Security Modeling based on Cosine and Inner Product Method", *Technology Reports of Kansai University,* Vol. 62, No. 2, 2020, pp. 107-114.

[8] Syaifudin, and Pratiwi, D., "Security Handwritten Documents using Inner Product", *Lecture Notes in Electrical Engineering,* Vol. 365, 2016, pp. 501-509.

[9] Pratiwi, D., Santoso, G.B., Saputri, F.H., "The Application of Graphology and Enneagram Techniques in Determining Personality Type based on Handwriting Features", Jurnal Ilmu Komputer dan Informasi, Vol. 10, No. 1, 2017, pp. 11-18.

[10] Sun, L., Wang, T., Ding, W., Xu, J., Lin, Y., "Feature Selection using Fisher Score and Multilabel Neighborhood Rough Sets for Multilabel Classification", *Information Sciences*, Elsevier, Vol. 578, November 2021, pp. 887-912.

[11] Hamraz, M., Khan. Z., Khan, D.M., Gul, N., Ali, A., Aldahmani, S., "Gene Selection in Binary Classification Problems within Functional Genomics Experiments via Robust Fisher Score", *IEEE Access,* Vol.10, May 2022, pp. 51682-51692.

[12] Al-Khawari, H., Athyal, R., Kovacs, A., Al-Saleh, M., Madda, J.P., "Accuracy of the Fisher Scoring System and The Breast Imaging Reporting and Data System in Identification of Malignant Breast Lesions", *Hematol Oncol Stem Cell Ther*, Vol. 2, No. 3, 2009, pp. 403-410.

[13] Sun, L., Zhang, X., Qian, Y., Xu, J., Zhang, S., "Feature Selection using Neighborhood Entropy-based Uncertainty Measures for Gene Expression Data Classification", *Information Sciences*, Elsevier, Vol. 502, Oct 2019, pp. 18-41.

[14] Lim, H., Lee, J., Kim, D.W, "Optimization Approach for Feature Selection in Multi-Label Classification", *Pattern Recognition Letters*, Elsevier, Vol. 89, April 2017, pp. 25-30.

[15] Guyon, I., Elisseeff, A., "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research,* Vol.3, 2003, pp. 1157-1182.

[16] Niyas, M.K.P., Thiyagarajan, P., "Feature Selection using Efficient Fusion of Fisher Score and Greedy Searching for Alzheimer's Classification", *Journal of King Saud University – Computer and Information Sciences*, January 2021.

[17] Gu, Q., Li, Z., Han, J., "Generalized Fisher Score for Feature Selection", *Proceedings of The Twenty-Seventh Conference on Uncertainty in Artificial Intelligence,* July 2011, pp. 266-273.

[18] Soe, T., Kyu, K.K., Htwe, T.T., "Comparative Study the Effect of Color Image Enhancement in Image Segmentation", *International Journal of Research and Innovation in Applied Science (IJRIAS)*, Vol. 5, No. 9, September 2020, pp. 7-12.

[19] Anuradha, K., Sankaranarayanan, Dr. K., "Statistical Feature Extraction to Classify Oral Cancers", *Journal of Global Research in Computer Science,* Vol. 4, No. 2, February 2013, pp. 8-12.

[20] Mangala, G.S.G., Cyril, P.R.P., "Energy Density Feature Extraction using Different Wavelets for Emotion Detection", *International Journal ofApplied Engineering Research*, Vol. 13, No. 1, 2018, pp. 520-527.

[21] Hamidi, R., Furqon, M.T., Rahayudi, B., "Implementasi Learning Vector Quantization (LVQ) untuk Klasifikasi Kualitas Air Sungai. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Vol. 1, No. 12, 2017, pp. 1758-1763.

[22] Laleye, F.A.A., Ezin, E.C., Motamed, C., "Weighted Combination of Naïve Bayes and LVQ Classifier for Fongbe Phoneme Classification", *Tenth International Conference on Signal-Image Technology and Internet-based System*, IEEE Xplore, November 2014, Marocco.

[23] Tse, P., Wang, D.D., Atherton, D., "Improving Learning Vector Quantization Classifier in Machine Fault Diagnosis by Adding Consistency", *International Conference on Neural Networks*, November 1995, Perth, Australia.

[24] Karthick, S. and Sathiyasekar, Dr. K., "A Survey Based on Region based Segmentation", *International Journal of Engineering Trends and Technology (IJETT*), Vol. 7, No. 3, January 2014, pp. 143-147.

[25] Bhargavi, K. and Jyothi, S. "A Survey on Threshold based Segmentation Technique in Image Processing", *International Journal of Innovative Research & Development*, Vol. 3, No. 12, November 2014, pp. 234-239.

[26] Saragih, G.S. and Rustam, Z., "Support Vector Machine with Fisher Score Feature Selection to Predict Disease-Resistant Gene in Rice", *Journal of Physics: Conference Series 1108*, 2018.