Iridology-based dyspepsia early detection using linear discriminant analysis and Cascade Correlation Neural Network

Conference Paper · May 2014

DOI: 10.1109/ICoICT.2014.6914054

Mahmud Dwi Sulistiyo
Telkom University

Iridology-Based Dyspepsia Early Detection Using Linear Discriminant Analysis and Cascade Correlation Neural Network

Mahmud Dwi Sulistiyo
School of Computing
Telkom University
Bandung, Indonesia
mahmuddwis@telkomuniversity.ac.id

Retno Novi Dayawati
School of Computing
Telkom University
Bandung, Indonesia
retnonovi@telkomuniversity.ac.id

Martintyas Pahirawan P. A.
School of Computing
Telkom University
Bandung, Indonesia
martintyasppa@gmail.com

Abstract—**Dyspepsia is a condition of indigestion and became one of the diseases with a large number of patients in Indonesia. Early detection of Dyspepsia is done to assist in the prevention of the disease. Iridology is a method of early detection in human organs disorders by analyzing the iris patterns. However, the application of Iridology technique often finds difficulties because it requires a high level of precision in the iris image observation. The low quality of the image also leads to the high possibility of human error. Therefore, research in this paper is aimed to build Iridology-based image processing system to detect early Dyspepsia. Stages of feature extraction and classification is important and determines how the performance of the established system. The method of Linear Discriminant Analysis (LDA) is used in the feature extraction stage to reduce the image feature dimension and obtain the features vector of the image that is being observed. Meanwhile, Cascade Correlation Neural Network (CC-NN) is a classification model that is used to determine whether an observed image shows the symptoms of Dyspepsia or not. Based on previous studies, it was shown that the CC-NN with its learning mechanism capable of producing high accuracy in classification problems. With the combination of these two methods, the system can generate fairly high accuracy in detecting Dyspepsia using Iridology-based techniques. The highest accuracy rate that can be achieved by the system is 95.45% for both in training and testing set.**

Keywords—Dispepsia, Iridology, Linear Discriminant Analysis, Cascade Correlation Neural Network

I. INTRODUCTION

Dyspepsia is a digestive disorder that is defined as a condition with gastrointestinal discomfort, especially in the upper body organs. Dyspepsia was ranked 10^{th} with a percentage of 1.5% for the category of 10 types of disease for outpatients in all hospitals in Indonesia. In 2004, Dyspepsia ranks to 15 from a list of 50 diseases with the highest inpatients in Indonesia with a proportion of 1.3% and peaked at number 35 on the list of causes of death with a 50% mortality rate of 0.6% [1]. This makes Dyspepsia be a quite dangerous disease and needs more serious action in order not to increase the number of patients with the disease.

So far, the examination of the patients with Dyspepsia is done after seeing the symptoms that have appeared with the use of USG, endoscopy, X-ray, and CT-Scan. This large number of patients demanded an early examination of the Dyspepsia sufferer as prevention of the disease. One technique for early detection of the disorder in human organs is Iridology. Iridology is a method that can give an indication of the oragan systems condition in the human body through the characteristics or signs that is visible in the iris of the eye.

According to an Iridologist, Dr. Asdi Yudiono, in applying the science of Iridology, there is difficulty in finding the characteristic details of the iris due to its small size, so that this requires very high precision. Another problem is the occurrence of human errors performed by iridologist, so it needs a system that could reduce such errors. The established system is an image processing system which is capable to learn the image patterns, so that it can recognize whether the iris image is being observed contains characteristics of a disturbance in the human stomach organs or not.

An image pattern recognition-based system requires image processing stage, feature extraction, and classification. Before getting into the classification process, the iris image, which has a very high dimension, needs dimension reduction and feature extraction. Feature extraction method used here is Linear Discriminant Analysis (LDA). This method has the ability to reduce the dimensionality of the data. However, it remains maintaining information as much as possible which is important inside. This method proved to be better than the other methods, namely the Principal Component Analysis (PCA) on previous research to a classification case.

Meanwhile, the classification model used here is Neural Networks with Cascade Correlation techniques in the learning process. A study on the classification of cervical cancer cells by Mc Kenna has successfully achieved a level of accuracy of 95.8% by applying Cascade Correlation Neural Network (CC-NN) [7]. Learning mechanisms in these methods not only make the network learn to adapt to the values of the connection weights between nodes, but also the structure of the network classification model.

II. METHODS

A. Iridology

Iridology (may also be referred to as Iridodiagnosis [8] or Iridiagnosis [9]) is the science and practice that can reveal the presence of inflammation, accumulation of toxins in the tissue, the gland dam, including the location and condition severity (acute, subacute, chronic, and degenerative) through reading patterns, colors, and other iris characteristics of the eye [5].

Support for research and development of this method is still lacking [10] because it has not fully accepted in the medical community, some of which refer to it as pseudoscience [11]. However, in some European countries such as Germany and UK, Iridology technique has been extensively researched and developed. In fact, in the United States, this method has been popular since it was introduced by Dr. Bernard Jenson in the 1950s and is now widely applied to support more accurate analysis for medical practitioners [12]. In Indonesia, the application of Iridology can be seen in some pratice places of alternative medicine.

In the basic concepts of Iridology, iris consists of 28,000 nerve fibers linked to the brain through the nerves that nourish the eye, the optic nerves, and autonomic nerves that pass from the brain to the organs of the body. In case of organ malfunctioning, the information will be sent to the brain as the center of information through the autonomic nervous control. The signal will be transmitted from brain to the optic nerve through the iris of the eye, to form a pattern, color, or certain characteristics in the iris [5]. Thus, a biometric technology based on iris recognition can be used to assist the identification purposes.

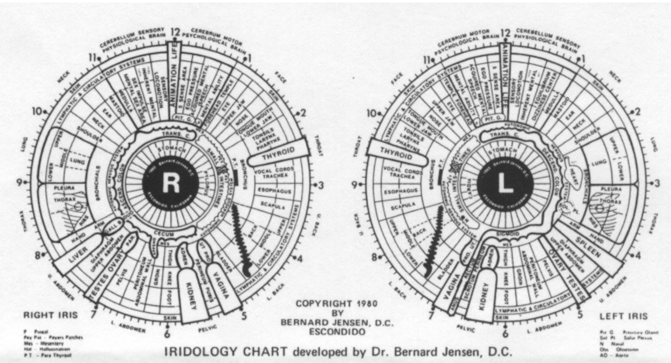

Figure 1. Iridology chart developed by Bernard Jensen [12]

B. Linear Discriminant Analysis [6]

Linear Discriminant Analysis (LDA) is a technique to find the linear combination of features which best separates two or more classes of objects. The objective of LDA is to reduce the dimensionality of the dataset, while still retaining as much information in it. Techniques on LDA attempted to find linear projection matrix to maximize the inter-class distribution (between-class scatter matrix denoted as S_w) by minimizing simultaneously the distribution matrix in the class itself (within-class scatter matrix denoted as S_b).

The matrix of S_w and S_b are then defined respectively as follow.

$$S_w = \sum_{j=1}^{c}\sum_{i=1}^{N_j}(x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{1}$$

$$S_b = \sum_{j=1}^{c}(\mu_j - \mu) - (\mu_j - \mu)^T \tag{2}$$

From the formula above, x_i^j is the i^{th} sample of class *j*, μ_j is the mean class *j*, μ is the global mean, *c* is the number of class, and N_j is the number of the sample in the class *j*.

Calculating the value of S_b will inform distance or separator between average vector values of each class and the global mean. Meanwhile, the result of S_w measures the distance between feature vector and the average vector values of each class. After gaining both values S_w and S_b, LDA finds the optimal projection by maximizing the ratio of $\frac{det|S_b|}{det|S_w|}$.

The maximum value will be obtained if the column vectors of the projection matrix *W*, is an eigenvector of $S_w^{-1}S_b$. Final representation of the LDA is a new matrix which is the product of the projection matrix with initial matrix as the input of the LDA. As the dimension of the projection matrix is *m x n*, where *m* is less than *n*, it will get a new matrix with smaller dimensions than the input matrix.

C. Cascade Correlation

Cascade Correlation is a type of feedforward neural network architecture and supervised learning algorithm. Cascade Correlation combines the two ideas. The first idea is the cascade architecture, which is increasing the number of hidden units in the network, where there is one unit on each additional neuron and never changed after it is added. The second idea is learning algorithm when the addition of hidden units. For each new hidden unit, performed maximizing the correlation between the output units with the residual error will be eliminated.

1) Cascade architecture

Architecture on Cascade Cascade Correlation Neural Network (CC-NN) is illustrated by figure 2. The number of input units and output units adapted to the problems faced. For example, in the illustration of Figure 2, the network starts with a simple architecture, which has three input units and two output units without hidden units. Each input unit and the input bias (which value is +1) associated with each output unit.

Hidden units are added to the network one by one until the stop condition is met, ie the number of hidden units reached the maximum number specified or generated an error value has reached the threshold. The process of adding each hidden unit begins with the creation of candidate units are connected to all input units and hidden units previously. Candidate unit is not connected to the output unit.

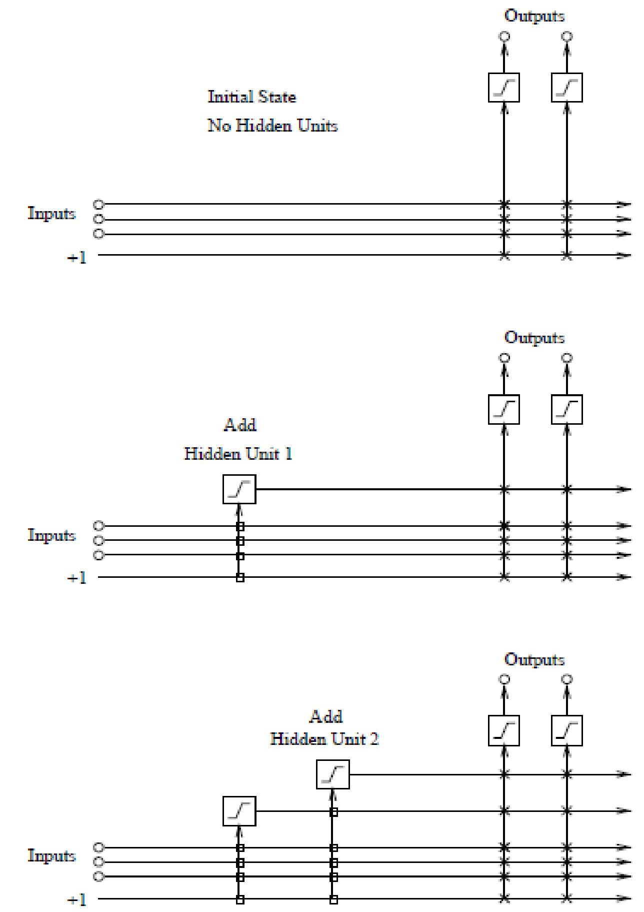

Figure 2. Illustration of Cascade Correlation process [4]

The candidate units (connection box in Figure 2) are trained using the weight training algorithm to maximize the value of the correlation between a candidate unit and the residual error, and then the weights are frozen. The last step is to install the candidate hidden units to the output unit by connecting it to the unit. The entire network (connection X in Figure 2) is trained using the training algorithm. Variations can be done with the use of several candidate units, but the units are installed as hidden units only the best candidate.

The correlation value (*S*) is defined as follows.

$$S = \sum_{o}\left|\sum_{p}\left(V_p - \overline{V}\right)\left(E_{p,o} - \overline{E_o}\right)\right| \tag{3}$$

where V is a candidate unit value, E_o is a residual error value of the output unit *o*, while $\bar{V}$ and $\overline{E_o}$ are the average values of all patterns.

To maximize the value of *S*, it needs a calculation of $\partial S/\partial w_i$ which is the partial derivative of *S* for each weight of the candidate input unit w_i. The formula used is as follows.

$$\partial S/\partial w_i = \sum_{p,o} \sigma_o\left(E_{p,o} - \overline{E_o}\right) f'_p I_{i,p} \tag{4}$$

where σ_o is a correlation sign between candidate unit and the output *o*. f'_p is a derivative function of candidate unit for pattern *p*, ehile $I_{i,p}$ is the input received by candidat unit from unit *i* to pattern *p*.

2) *Learning algorithm*

Training algorithm used for CC-NN here is the Quick Propagation, or can be called with QuickProp. The algorithm is a modification of the Back Propagation algorithm with the assumption that the error curve during the training process can be approximated by a parabolic curve which is concave and change the slope of the curve is not influenced by other weights also changed. The slope (*S*) in question is the sum of the partial derivative of the error based on the weight given to all the patterns in each epoch are denoted as follows.

$$S(t) = \sum_{p=1}^{P} \frac{\partial E(p)}{\partial w} \tag{5}$$

QuickProp uses the weight and value of the change in the value of the slope of the previous epoch. The value of the new weight change $\Delta w(t)$ is calculated as follows.

$$\Delta w(t) = \frac{S(t)}{S(t-1) - S(t)} \Delta w(t-1) \tag{6}$$

The formula that is used to initialize the weights changes are as follows.

$$\Delta w(0) = -\alpha S(0) \tag{7}$$

where α refers to learning rate parameter value.

III. SYSTEM DESIGN

A. *The Proposed System and The Proposed One*

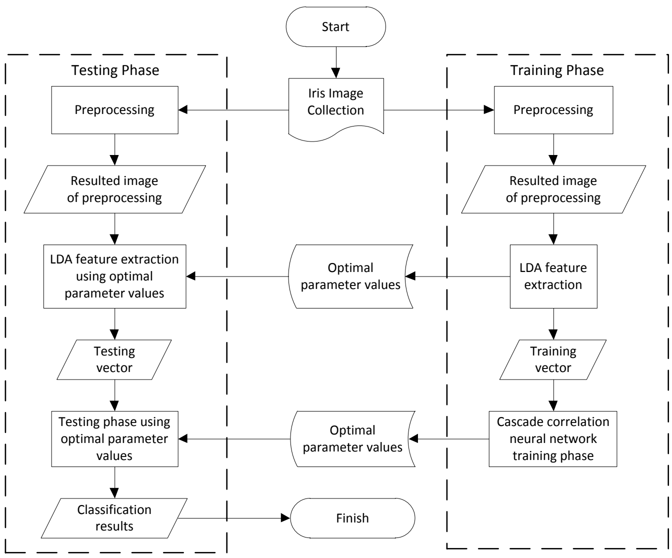

Figure 3. Flowchart of the proposed system

In the proposed system, all the data both training data and test data preprocessing stage prior experience, which consists of grayscaling, image quality improvement, cutting the image (focusing on iris sections), normalized by the polar transformation, and making the stomach area according Iridology charts. Training data entered into the stage of feature extraction using LDA and generate training vectors and also the optimal parameters. The optimal parameters are projection vectors that will be used to extract testing data and generate testing vectors.

Training vectors are used to train the CC-NN to obtain the optimal parameters such as the number of hidden units and weights along the connections between neurons units. The final step is to test the use of parameters and the training results showed the classification of test vectors to analyze the system performance. The followings are more detailed explanation of the processes.

B. Image Preprocessing

The first stage is the Region of Interest (ROI) extraction. Each iris image is cut focusing on the patient's eye. The second stage is converting the RGB iris image to the grayscale one.

The quality of image is then improved with adaptive histogram equalization techniques to increase the color contrast in the image of the iris. The mechanism is performed by changing the darkest point in the image into black, while the brightest point in the image becomes bright white.

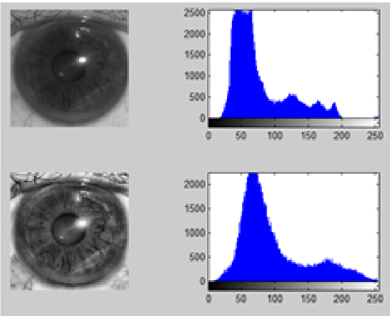

Figure 4. Hystogram equalization

The next stage is a stage of transforming normalized polar image. Basically it changes a circle form into a straight line. This transformation makes the iris region can easily be processed in a matrix. Daugman's Rubber Sheet is the method used in this stage. Each point on the iris area will be mapped into a pair of polar coordinates (r, θ), where r is in the interval [0,1] and θ is the angle is in the interval [0,2π].

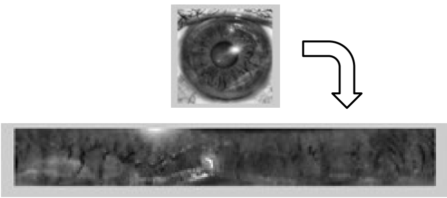

Figure 5. Polar transformation

The last step in the preprocessing is making the stomach area of the image of the polar transformation. The position of the stomach area for identification of disease Dyspepsia taken based on Iridology chart, located between radius of 0.4 and 0.5 in the cros-sectional of the iris.

Figure 6. The extraction of the stomach area

C. Feature Extraction Using LDA

Data are entered into the LDA is a vector image preprocessing results. The first step is to perform calculations between scatter matrix and within scatter matrix. Then, the eigenvalues and eigenvectors of each image are calculated. The next step is to determine the projection matrix with size *m x n*. *m* is the size of a long feature vector, which is 3600, while *n* is the number of new designated dimensions. The study here uses 3 different sizes of *n* which are 10, 50, and 100.

Projection matrix is obtained by the eigenvectors with eigenvalues in descending sort. To get the feature vectors that represent the end of the LDA, multiply the input matrix with the projection matrix. In this study, the three sizes of feature vectors used are 33x10, 33x50, and 33x100 with real values.

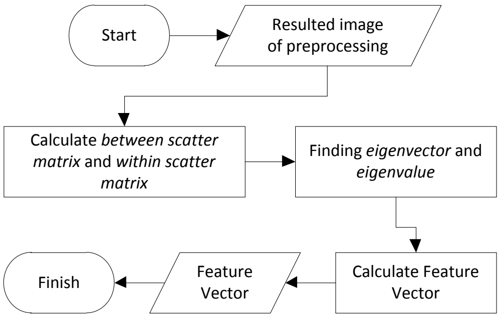

Figure 7. Flowchart of LDA

D. CC-NN Training Process

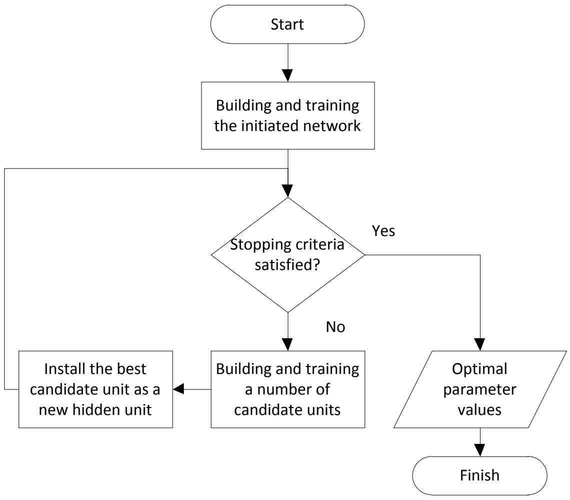

Figure 8. Flowchart of Cascade Correlation training process

Network training procedures used by the CC-NN is supervised learning, the training mechanism that is accompanied with the target. There are two values in the target data, which are 0 and 1. A value of 0 represents no symptoms of Dyspepsia, whereas a value of 1 represents the presence of Dyspepsia symptoms.

The end result of this CC-NN training is the number of hidden units and the weights that connect each unit, including weights from the input units to the hidden units, from the input units to the output units, from the hidden units to the hidden units, and from the hidden units to the output units.

E. System Performance Measurement

Testing is done by comparing the output of classification process using of testing data with target data. Classification performance was assessed through the level of accuracy that is measured from the ratio of the number of correct outputs for a total amount of test data. The calculation can mathematically be written as follows.

$$\text{Accuracy} = \frac{\text{Number of truth}}{\text{Number of total record}} \times 100\ \% \qquad (8)$$

IV. EXPERIMENTAL RESULT

A. Experiment Purposes

In this study, experiments are aimed at two testing purposes, namely testing the LDA as a method of feature extraction and testing of the CC-NN as a method of classification.

1) LDA Experiment

The purpose of this experiment is to find the dimensions for the size reduction, performing LDA results that give the best recognition accuracy on the CC-NN. The ability of CC-NN to perform the recognition stage is not only because of the importance of the information contained by the input feature vector, but also the number of input units used. The scheme used in this experiment is as follow.

a. Input data to the LDA are the image preprocessing results, including 22 training set and 11 testing set

b. New dimension parameters are 10, 50, and 100, added with non-reducted vector, which is sized 3600

2) CC-NN Experiment

The purpose of this experiment was to determine the appropriate parameter settings that influence the system performance. The parameter in question is the number of candidate units. The scheme used here is as follow.

a. The number of input units is the optimal number of dimensions produced on the LDA experiment

b. The number of candidate unit, including 1, 8, and 16

B. LDA Testing and Analysis

The training process for each LDA dimension reduction is illustrated by the graph of MSE changes on picture 12. Learning rate parameter value used is 0.1, the number of candidate units 8, and the maximum number of epoch 1000.

Figure 9 tells about what happened in training phase using different dimension parameters. By using the dimensions of 10, 50, and 100, the MSE changes are not much different. The MSE produced in each is progressively getting smaller, indicating that the system leads to convergence condition. This is different from the testing vector with dimension of 3600, where there was no change in the condition, indicating that the system does not lead to the convergence condition. This is because the dimension 3600 stuck at a local optimum value and can not recognize the training patterns.

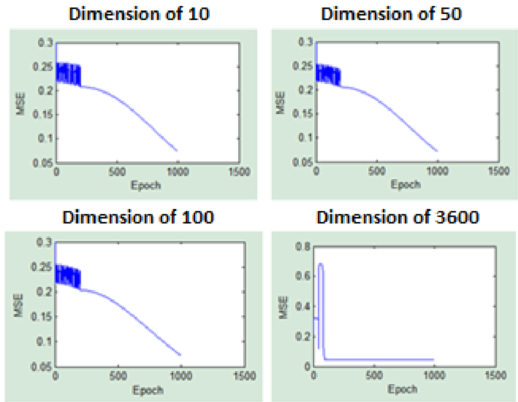

Figure 9. Graph of MSE changes in each epoch on training phase

In 3600 dimensions, almost all values in the vector are considered as feature, even though it is not distinguishing the characteristic between the normal eyes with eyes that have symptoms of dyspepsia. Meanwhile, by using the dimensional reduction result, the system is getting easier to recognize the iris image as it has made the selection of vector values that are considered important to distinguish one class to another class. In addition, the complexity of the calculation and introduction process is reduced.

Here are the results of testing the effect of reduced dimensions LDA with parameter values for the learning rate is 0.1, the number of candidate units is 8, and maximum number of the epoch is 1000.

Table 1. LDA Tesing Results on the Effect of Dimension Size

Dimen-sion	Hidden Unit	Epoch	MSE	Training Accuracy	Testing Accuracy
10	1	1000	0.070604	95.45%	90.91%
50	2	1000	0.068718	95.45%	90.91%
100	35	338	0.328788	67.12%	71.21%
3600	24	558	0.310606	68.93%	72.42%

The table above shows that the best accuracy rate is generated by the dimension of 50 for training set and the dimensions of 10 for testing set. This suggests that the dimensions 10 and 50 can recognize patterns better than the 100 and 3600 dimensions. That is because the LDA is able to reduce the dimensionality of the feature vector, while still retaining as much information in it. Enforcing dimensions to be 10 or 50 is appropriate because the dimension is not disposed of importance on vector containing the iris image. When these values are discarded, the calculation complexity is reduced and the recognition accuracy level tends to be better.

In dimension 100 and 3600, there are vector values of less importance to distinguish one class to another. The magnitude of these dimensions also makes it difficult for the network to find the right class pattern. This suggests that the LDA has a strong influence on the level of recognition accuracy is conducted by CC-NN.

C. CC-NN Testing and Analysis

In the testing and analysis of CC-NN, the used input has a dimension of 10, based on the previous LDA experiment. The learning rate is 0.5, and the maximum number of epoch is

1000. The training process for each candidate value of the unit number is illustrated by the graph in figure 14.

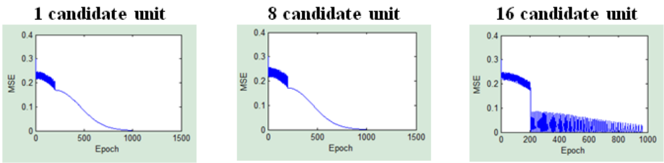

Figure 10. The chart of MSE changes in each epoch on training phase

The graphs illustrate the comparison of training process through each number of candidate units, which are 1, 8, and 16. There is no significant difference in 1 & 8 candidate units. Accordingly, MSE value has fluctuated in the beginning then step down fairly stable before the end of training process.

Significant differences ilustrated by 16 candidate unit, where the error value goes down significantly when hidden units are added to the network. However, its fluctuative condition has been seen almost throughout the training process till the end of the process. This indicates that the value of 16 remains the unstable candidate units to use.

Here are the results of tests on the effect of the number of candidate units with dimensions 10, learning rate 0.5, and the maximum number of epoch 1000.

Table 2. Testing Results on the Effect of Candidate Unit Number

Candidate Unit	Hidden Unit	Epoch	MSE	Training Accuracy	Testing Accuracy
1	1	860	0.000939	90.91%	90.91%
8	1	860	0.000928	90.91%	95.45%
16	1	861	0.000942	90.91%	90.91%

The table above test results show that the number of candidate units between 1, 8, and 16 do not significantly affect the training accuracy. This is because in the last epoch, which is about the epoch-860, each of the resulting networks equally leads to convergent condition. However, the number of units of 8 candidates selected as the most optimal because it is more stable during the training process. It also yields the highest testing accuracy rate and the smallest final MSE compared to two other number of unit candidates.

The different test results may be seen if the maximum number of epoch is minimized. So, it will obtain the optimal conditions, in which the training process can be shortened, while the results obtained remains the best.

V. CONCLUSION

Based on the results of the implementation, testing, and analysis has been done in the present study, the combination of the two methods, LDA and CC-NN can be applied to the image processing system for the early detection of the symptoms of dyspepsia. The results were quite promising, with the highest accuracy rate of 95.45% by the given data. Aside from that, by the observation, the optimal dimensions of LDA are 10 and 50; the optimal number of candidate unit is 8.

In this Iridology case, LDA as a method to extract the image characteristics of the iris image can obtain optimal characteristic vector, ie by dimensions much smaller, but still contain important information. CC-NN applied as learning algorithm in this study can produce optimal classifier model.

VI. FUTURE WORK

This on going study still has many shortcomings. One is a few of data involved in the process of training and testing, so the results are not able enough to generalize to real-world problems of Iridology. Therefore, in future studies, the involved data may concern to a number of conditions of taken images that may occur.

In addition, the system constructed in this present study is still offline. This means that the iris image acquisition process is separated from the process of the image tabulation and identification. It is expected that further development of the system is able to combine both of them, so it may result early detection system of Dyspepsia disease in real time processing.

REFERENCES

[1] Bere, Devi Andrayani, et al, 2012. Relationship between Eating Disorder and Smoking with Dyspepsia in Genesis Outpatient Age 20-44 Years in Depok, Sleman, Yogyakarta. Yogyakarta. Universitas Respati.

[2] Daugman, J. 2002. How Iris Recognition Works. IEEE Transactions on Circuits and Systems for Video Technology. Vol 14.

[3] Fahlman, S. E., & C. Lebiere. 1991. The Cascade-Correlation Learning Architecture. Pittsburgh. School of Computer Science Carnegie Mellon University.

[4] Fausett, Laurene V. 1994. Fundamental of Neural Networks: Architectures, Algorithms, and Applications. Prentice Hall. New Jersey.

[5] Manjoer, A, et al, 2000, Medical capita selecta, 3rd edition, Jakarta, Medika aeusculapeus.

[6] Martinez A.M and Kak A.C. 2001. PCA versus LDA. IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233.

[7] McKenna SJ, Ricketts IW, Cairns AY, et al. 1992. Cascade Correlation Neural Networks for the Classification of Cervical Cells. Neural Networks for Image Processing Applications, IEE Colloquium on, pp 5/1-5/4.

[8] Cline D; Hofstetter HW; Griffin JR. Dictionary of Visual Science. 4th ed. Butterworth-Heinemann, Boston 1997. ISBN 0-7506-9895-0.

[9] LindlahrTake, Henry (2010) [1919]. Iridiagnosis and other diagnostic methods. Whitefish, Montana: Kessinger. ISBN 978-1-161-41232-1. Retrieved 12 September 2011.

[10] Ernst, E. (2000). Iridology: Not Useful and Potentially Harmful. Archives of Ophthalmology 118 (1): 120–1. doi:10.1001/archopht.118.1.120 (inactive November 13, 2012). PMID 10636425.

[11] Hockenbury, Don H.; Hockenbury, Sandra E. (2003). Psychology. Worth Publishers. p. 96. ISBN 0-7167-5129-1.

[12] Unknown. 2010. Iridology Study of Eyes to Diagnose Health Problems. http://www.crystalinks.com/iridology.html.