

Linear Discriminant Analysis vs. Genetic Algorithm Neural Network with Principal Component Analysis for Hyperdimensional Data Analysis: A study on Ripeness Grading of Oil Palm (*Elaeis guineensis* Jacq.) Fresh Fruit

Divo Dharma Silalahi

SMART Research Institute, PT. SMART Tbk, Riau, Indonesia

Conсорcia E. Reaño and Felino P. Lansigan

Institute of Statistics, University of the Philippines Los Baños

Rolando G. Panopio

*Institute of Mathematical Sciences and Physics,
University of the Philippines Los Baños*

Nathaniel C. Bantayan

*Institute of Renewable Natural Resources,
University of the Philippines Los Baños*

Using Near Infrared Spectroscopy (NIRS) spectral data, the Linear Discriminant Analysis (LDA) performance was compared with the Genetic Algorithm Neural Network (GANN) to solve the classification or assigning problem for ripeness grading of oil palm fresh fruit. The LDA is known as one of the famous classical statistical techniques used in classification problem and dimensionality reduction. The GANN is a modern computational statistical method in terms of soft computing with some adaptive nature in the system. The first four new components variables as result of Principal Component Analysis (PCA) also were used as input variables to increase the efficiency and made the data analysis process faster. Based on the results, both in training and validation phase GANN technique had lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), higher percentage of correct classification and suitable to handle large amount of data compared to the LDA technique. Therefore, the GANN technique is superior in terms of precision and less error rate to handle a hyperdimensional problem for data analysis in ripeness classification of oil palm fresh fruit compared to the LDA.

Keywords: Near Infrared Spectroscopy, Neural Network, Genetic Algorithm, Linear Discriminant Analysis, Principal Component Analysis, Oil Palm, Ripeness

1. Introduction

The NIRS Instrument is an analytical tool with many applications in research areas such as pharmaceuticals, agriculture, medicine, food, etc. This instrument determines chemical or nutrient analysis based on the interaction of physical matter with light in the near infrared spectral region (700-2500nm). Infrared spectrum from this light will give sufficient information about structure compound of samples. Guidetti et al. (2010) said that in particular NIRS is a rapid and non-destructive technique, which is requiring minimal sample processing before analysis. This instrument couples with chemometrics method appears to be one of the most powerful analytical tools for studying food products. In this study, NIRS instrument screens the characteristics of sample parameter on oil palm fresh fruit and generates the spectral data. These spectral data then were used as raw data and coupled with statistical technique to solve the classification or assigning problem for ripeness grading of oil palm fresh fruit.

There are many classical statistical techniques used for classification or assigning an object to an appropriate group. One of the famously used technique is Linear Discriminant Analysis (LDA) or sometimes known as Fisher's linear discriminant. The LDA is not only used in classification problem but also used as a dimensionality reduction method and is widely used in machine learning and computer vision (Belhumeur et al., 1997; Swets and Weng, 1996; Dudoit et al., 2002). Haeb-Umbach and Ney (1992) explained that the LDA is a well-known classical technique in statistical pattern classification particularly for improving the discrimination and compressing the information contents (with respect to classification) of a feature vector by a linear transformation. In the data classification process, LDA generates the linear combinations of observed features that best characterize two or more classes of objects. These linear combinations are related to the linear function which as description of group separation and referred to as canonical variates or discriminant coordinates (Rencher, 2002). For separating the groups LDA finds the best best linear combinations of variables, if there are different groups then it is possible to find separate linear discriminant functions which partition the between groups variance into decreasing order (Fieller, 1982).

Besides the LDA technique, another statistical method that has proved to be useful in assigning an object to an appropriate group or classification is Neural Network (NN). NN is a non linear (NN) modern computational statistical method with some adaptive nature in the system (öztürk, 2003). This computational model consists of several processing units, which assimilate the relationship between input and output through large number of weighted connections (Stepniewski and Keane, 1996; öztürk, 2003). Though NN is a powerful method, several studies such as those by Guyer and Yang (2000), Öztürk (2003), Yalkin and Korkmaz (2012) and Vizitiu and Popescu (2010) have cited problems in the NN process,

especially in network architecture process. The design of network architecture mainly consists of the selection of the training algorithm, the determination of network layers and the numbers of neurons in each layer. Öztürk (2003) noted that NN has its own effective algorithm to solve the neural architecture problems, but the technique is based only on the iterative procedure and takes a long time to learn the whole process. Hence, these often lead to inefficient and complex network architecture in NN. To address these problems, Öztürk (2003) and Yalkin et al. (2012) proposed an optimization technique to improve the convergence speed and reduce the computational complexity of network architecture in NN. This technique is known as genetic algorithm (GA) which is widely used to solve optimization and search problems. GA is an approximate solution to optimization problems based on a heuristic search algorithm of machine learning that mimics the process of natural selection (Mitchell, 1998). In this study, GA was used for the learning phase of NN. This learning phase starts from training the weight of the connection, designing the network architecture adaptation and leading to the evolution of an optimal learning mechanism rule (Abraham, et al., 2008). Hence, in order to reduce the complexity of NN design and make NN become more efficient NN can be combined with GA. This technique thus is known as genetic algorithm neural network (GANN).

In order to handle hyperdimensional problem in data analysis, the Principal Component Analysis (PCA) method were applied on original NIRS spectral raw data. This PCA extracted new component variables as input variables in the LDA and GANN. These new component variables were used to increase the efficiency and made the data analysis process faster. Thus, the performance of LDA was compared with the GANN in minimizing the misclassification for ripeness grading of oil palm fresh fruit.

2. Linear Discriminant Analysis

The LDA was introduced by Fisher (1936) in his famous study about the use of multiple measurements in taxonomic problems. The LDA search for the best linear combinations of objects and seeking the data points directions in each classes which are far or closely different from each other. The LDA is based on assumptions of multivariate normality distributed classes and equality of classes covariance matrix. Relating to this basic assumption, some references such as Duda et al. (2001) stated in fact linear discriminant analysis frequently achieves good performances in the tasks of object recognition, even though the assumption of normality and common covariance matrix among groups are often violated. McLachlan (2004) mentioned that LDA is a popular technique since violation of the statistical assumptions often has a minor effect. In application, LDA is usually used in prediction, pattern recognition, decision making, clustering, identification, and discrimination, among others.

The approach method by Fisher (1936) of obtaining a low-dimensional representation of the data in discriminating several populations for visual inspection and graphical descriptive purposes has given good extension in discriminant method (Johnson and Winchern, 2007). A side for discriminating several populations, these discriminants also has a benefit as a basis of classification purpose. Furthermore, the particular concepts of LDA's classification method as extension of fisher linear discriminant analysis could be adopted from Johnson and Wichern (2007).

Define the linear discriminant

$$Y_k = a_k'X$$

is the k discriminant for $k \leq s$, so that

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{bmatrix} \text{ has mean vector } \mu_{iY} = \begin{bmatrix} \mu_{iY_1} \\ \mu_{iY_2} \\ \vdots \\ \mu_{iY_s} \end{bmatrix} = \begin{bmatrix} a'_1 \mu_i \\ a'_2 \mu_i \\ \vdots \\ a'_s \mu_i \end{bmatrix}$$

where s is number of nonzero eigenvalues or number of discriminants.

3. Genetic Algorithm Neural Network

To set up the GANN model, it is needed to specify the structure of NN in the model design. Feed Forward Neural Network (FFNN) is used as the structure of NN which is one of the popular structures among NN. This FFNN structure would give an efficient network to solve the complex problem by modeling complex input-output connection (Karimi et al., 2010) (Bahman and Ali, 2010). For designing the FFNN for the purpose of nonlinear approximation, architectures one or two hidden layers are typically used (Stepniewski and Keane, 1996). Hence, Multilayer Perceptron (MLP) is used to maps the set of input data onto the set of appropriate output. MLP is a general feed-forward artificial NN, which solves non-separable problems (Flake, 1998). It consists of more than one hidden layer of neurons and the updates in the layers are exhibited starting from the input to the output. Svozil et al. (1997) stated that feed-forward NN means that NN knows the desired output and adjusting of weight coefficients is done in such way, that the calculated and desired outputs are as close as possible.

Since MLP is also a class of FFNN, MLP can distinguish itself by the presence of one or more hidden layers, whose computation nodes are correspondingly called hidden neurons or hidden units. Commonly, the standard MLP of feed-forward neural network consist of input layer, hidden layer and output layer. Then, each layer has a weight matrix, bias vector b , and an output vector y . The subscript number will be appended to each layer of the variable of interest as

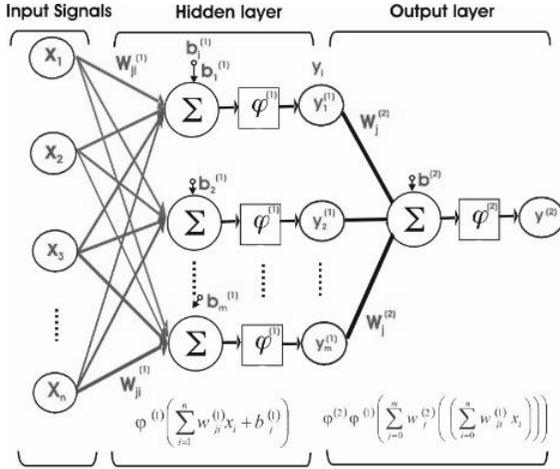


Figure 1. Three Layers of MLP Feed-forward Neural Network

information for differentiating. Suppose, $X^T = [x_1, x_2, \dots, x_n]$ are denoted as input signals; $w_j^T = [w_{j1}, w_{j2}, \dots, w_{jn}]$ are the weights of neurons j ; $\phi(v)$ is the activation function of the neurons; $b^T = [b_1, b_2, \dots, b_m]$ are the bias vector and $Y^T = [y_1, y_2, \dots, y_n]$ are the output signals. Note that, in feed-forward network consists of a set of input neurons which feed the input pattern to a layer or subgroup of neurons (Gopal, 2012). Each layer will makes independent computations by using its own weight matrix w , bias vector b , and an output vector y . This computation then will be received and passed to the others layers in MLP feed-forward network. By forming a linear combinations of n input and adding the bias b_j , the general computation in each layer of NN can be defined as

$$y_j = \phi \left(\sum_{i=1}^n w_{ji} x_i + b_j \right); \quad j = 1, 2, \dots, m$$

In order to reduce the complexity of NN in architectures design and make NN becomes more efficient network, based on references (Öztürk, 2003; Yalkin and Korkmaz, 2012) GA can be used for these purposes which generally are affected by number of neurons in each layer since the model become larger and more complex. In the GANN computational process, GA operates a population of potential solutions by applying the principle of survival of the fittest to produce better optimization solution in training process of MLP feed-forward NN (Chipperfield et al., 1994).

Holland (1975) stated that a GA process starts with a population of individuals, which grows toward optimum solution through genetic operators such as selection,

crossover and mutation that encouraged by the sequence of biological process. In each generation of the population, then a new chromosome is created by process of selecting individuals according to their level of fitness in the problem domain, and reproducing those by using operators such as crossover and mutation borrowed from natural genetics (Chipperfield et al., 1994). Remember, the chromosomes also can be said as a set of parameters and often as a simple string which define the alternative solution of optimization for GA process.

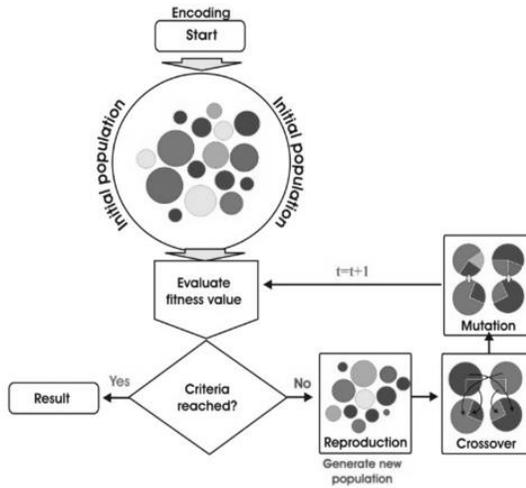


Figure 2. Genetic Algorithm Process

4. Research Methodology

The NIRS spectral data from total of 3870 oil palm fresh fruit samples (observations) with 522 wavelengths variables were used as original raw data for dimensional reduction using PCA. Oil palm fresh fruit samples were obtained from two different progenies of Tenera variety with code number 9005165E and 9008008E planted at Kandista Sari Estate (KNDE), PT. SMART Tbk (Sinar Mas Agribusiness Resources and Technology) Riau Province, Indonesia. These two progenies are known to vary in yield potentials. For each palms, 30 fruits at varying stages of ripeness were randomly taken from the inner (10), middle (10) and outer (10) positions of a bunch over a period of 6 to 27 weeks after anthesis (WAA). Next, total of 90 samples per WAA were sent immediately to the laboratory for NIR scanning and wet chemistry analysis. The oil and moisture content in mesocarp are the two parameters used as reference to determine the ripeness grade (unripe, under ripe, ripe and over ripe) of oil palm fresh fruit (see in Silalahi et al., 2016).

The raw data were divided into two subsets: training and validation. The ratio used in these subsets was 85% of total raw data for training data and 15% for validation data. The first four new components variables (PC 1, PC 2, PC 3 and PC 4) as result of Principal Component Analysis using original raw data were used as dataset for input variables to increase the efficiency and made the data analysis process in LDA and GANN faster. Comparison on these two statistical techniques was done with criteria the model with lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) and the highest percentage (%) of good classification is categorized as the better technique in this ripeness classification problem. The computational statistical software like MATLAB R2011b for windows also has been used to process the LDA and GANN result.

5. Results and Discussion

Table 1. The constant and coefficient for LDA function using training data

	Function (Y_j)			
	1 = Unripe	2 = Under Ripe	3 = Ripe	4 = Over Ripe
a_0 (constant)	-1.5084	0.0702	2.1920	5.7940
a_{j1}	-607.193	33.103	370.137	684.287
a_{j2}	-10.218	90.787	75.194	-409.754
a_{j3}	-626.254	305.155	714.404	-770.011
a_{j4}	1000.221	-210.573	-592.965	-626.782

Based on the coefficient variable for LDA in Table 1, it can be seen how the discriminant variables affect the ripeness class of the samples. In the first linear function (unripe) class, the PC 4 has greater magnitude than the coefficients for the other three variables. While in the others three linear functions (under ripe, ripe and over ripe), the PC 3 has the greatest and dominant impact of the four variables to determine the discriminant score.

By adding a_0 as the constant term of function y_j and defining a_{j1} , a_{j2} , a_{j3} , and a_{j4} as coefficient for variables PC 1, PC 2, PC 3 and PC 4, respectively. So that, the linear discriminant function above can be used to compute the linear discriminant score for each class in all samples. To determine the ripeness class, the highest linear score of the fourth functions the more likely that class (grade). This linear discriminant score are not probabilities, hence may assume negative values. The last process in LDA is to determine the classification phase for training and validation data. The percent of good classification results for training and validation data was simply processed the cases are those that were correctly and incorrectly classified based on the analysis. Figure 2 clearly figures out the classification results for each ripeness class both in training and validation. Total 2,735 of 3,288 samples from training data were correctly predicted as their

original classes, hence it can be stated that the classification results using the LDA function on training data has successfully predicted 83.2% of samples into their original ripeness classes. In validation phase, total 476 of 582 samples from validation data were correctly predicted means LDA function has successfully predicted about 81.7% of samples into their original ripeness classes.

In the GA process, the computational method has been used to generate the GA as solutions to optimize the convergence speed and reduce the computational complexity of network architecture in NN. These computational consists the whole process of training NN model using GA. In the GA, the processes generally start from initialization of population (random number), evaluation of fitness function and finally generation of new population. While in the NN, these generations of new population were used to construct the network architecture training and finally testing the network to make sure that the network is trained properly. The individual evaluation from 300 set of parameters as alternative solution of optimization was generated from 1 to 100 generations in order to find the best fitness in each class. The running of GA system was successfully finding the best fitness in the unripe and over ripe classes with the best fitness in these two classes were 5.009. While, in the under ripe and ripe classes the running of GA system were not successfully finding the best fitness. Even the sets of parameters in under ripe and ripe classes were not reached the best fitness; the maximum fitness value obtained can be still used as fitness value for the fittest individual set of solution in their classes. Remember, the fitness value is categorized as the best fitness function if it is equals or greater than 5 (or with RMSE ≤ 0.2). Thus, the best fitness in each class contains 25 parameters as solutions to the network parameter were used to optimize the GANN network. These network parameters can be shown in Table 2. In input layer consist of four unit signals x_i , hidden layer consist of 16 unit weights $w_{ji}^{(1)}$ in each class, four unit bias $b_{j(i)}^{(1)}$ in each class and four unit outputs $y_{j(i)}^{(1)}$ in each class, while output layer consist of 16 unit weights , 4four unit bias , and 4 unit . Total GANN parameters used in each class were 25 genes; hence overall 100 genes were used to solve this multi-class pattern classification problem. Thus, the three layers of MLP feed-forward neural network in this multi-class (4-classes) pattern classification problem can be illustrated as in Figure 2.

Table 2. The summary of optimal set of network parameters in training GANN

Parameters	Unripe	Under Ripe	Ripe	Over Ripe
	$y_1^{(2)}$	$y_2^{(2)}$	$y_3^{(2)}$	$y_4^{(2)}$
$w_{11}^{(1)}$	2.9048	-5.0818	-2.5060	-7.6292
$w_{21}^{(1)}$	-8.9439	0.4699	-0.7561	-2.8492
$w_{31}^{(1)}$	8.8628	9.9471	-4.7717	9.8940
$w_{41}^{(1)}$	1.7055	-7.6912	9.0676	7.1349
$w_{12}^{(1)}$	-3.8530	-4.5701	1.7640	-5.1909
$w_{22}^{(1)}$	4.5700	6.5251	0.5603	2.7665
$w_{32}^{(1)}$	-5.3198	-4.4291	7.0886	5.1373
$w_{42}^{(1)}$	4.4041	-8.7630	7.6763	7.9320
$w_{13}^{(1)}$	-3.5060	5.4655	-5.1674	9.7431
$w_{23}^{(1)}$	-7.0996	-8.3323	2.0830	-4.1642
$w_{33}^{(1)}$	0.5733	-3.1149	7.2410	5.6531
$w_{43}^{(1)}$	6.5063	7.7332	6.4338	8.6611
$w_{14}^{(1)}$	3.4836	-3.4818	3.0538	6.2382
$w_{24}^{(1)}$	-2.7219	-7.9729	4.6655	8.4887
$w_{34}^{(1)}$	2.0470	-7.9248	8.2921	8.1235
$w_{44}^{(1)}$	-0.3450	-6.8347	1.7138	-9.3008
$b_1^{(1)}$	0.8058	-9.1016	-4.0739	1.0687
$b_2^{(1)}$	0.0858	5.0860	9.5944	-0.3701
$b_3^{(1)}$	6.6772	-9.2167	-8.3169	7.2942
$b_4^{(1)}$	1.2192	-3.0741	9.4016	8.2151
$w_1^{(2)}$	-2.8688	-2.6191	6.6165	2.1800
$w_2^{(2)}$	-4.0163	-4.0557	-1.6409	7.7640
$w_3^{(2)}$	8.4077	-8.4750	-0.5740	7.1125
$w_4^{(2)}$	5.9733	-5.3552	2.2961	-3.8651
$b^{(2)}$	-8.9464	4.2649	-0.7758	-8.2350

These optimal set of MLP network parameters then were applied to GANN network in order to approximate the output value of activation function of all classes. These activation functions were used to classify the ripeness grade for each sample; the process is similar to LDA where the highest activation function values from the output layer of the fourth GANN model the more likely that class.

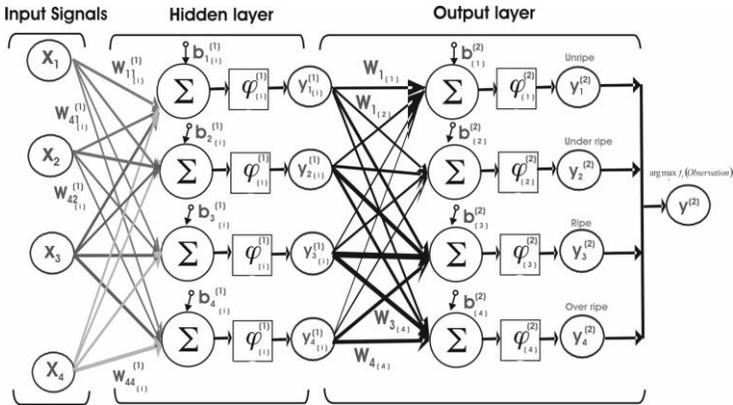


Figure 3. Three Layers of MLP Feed-forward Neural Network for Multiclass (4-classes)

The calculation for activation function as value in second output layer can be illustrated as below:

$$\begin{aligned}
 y_{(1)}^{(2)} &= \phi^{(2)} \phi^{(1)} \left(\sum_{j=1}^m w_j^{(2)} \left(\sum_{i=1}^n w_{ji}^{(1)} x_i + b_j^{(1)} \right) + b^{(2)} \right) \\
 &= \phi^{(2)} \phi^{(1)} \left(\sum_{j=1}^4 w_j^{(2)} \left(w_{j1}^{(1)} PC_1 + b_1^{(1)} + w_{j2}^{(1)} PC_2 + b_2^{(1)} + w_{j3}^{(1)} PC_3 + b_3^{(1)} + \right. \right. \\
 &\quad \left. \left. w_{j4}^{(1)} PC_4 + b_4^{(1)} \right) + b^{(2)} \right) \\
 &= \phi^{(2)} \phi^{(1)} \left(\sum_{j=1}^4 w_j^{(2)} \left(w_{j1}^{(1)} PC_1 + b_1^{(1)} + w_{j2}^{(1)} PC_2 + b_2^{(1)} + w_{j3}^{(1)} PC_3 + b_3^{(1)} + \right. \right. \\
 &\quad \left. \left. w_{j4}^{(1)} PC_4 + b_4^{(1)} \right) + b^{(2)} \right)
 \end{aligned}$$

Figure 3 visually shows the original ripeness and their model prediction using training and validation data, respectively. In training phase, the LDA model still had high correct and incorrect prediction where high correct prediction were placed in unripe and over ripe grade about 100% and 79% respectively; while high incorrect prediction were placed in under ripe and ripe grade about 81% and 50%, respectively. Similar to the GANN model, instead of having correct prediction, the GANN model also still had incorrect prediction where high correct prediction were placed in unripe and over ripe grade about 98% and 80%, respectively; while high incorrect prediction was only placed in under ripe grade about 54%. Generally, based on these correctness predictions in training phase, the GANN model performed better than LDA model. Since almost in the four ripeness grades

the GANN model had lower percentage of incorrect prediction compared to LDA model prediction.

In validation phase, the LDA model also still had high correct and incorrect prediction as in training phase where high correct prediction were placed in unripe and over ripe grades about 100% and 78%, respectively; while high incorrect prediction were placed in under ripe and ripe grade about 87% and 61%, respectively. These incorrect percentages obtained were higher than in training phase. In the GANN model, the high correct prediction were placed in unripe and over ripe grades about 98% and 71%, respectively; while high incorrect prediction were placed in under ripe and ripe grade about 51% and 52%, respectively. In this validation phase, the GANN model again performed better than LDA model. Since almost in the four ripeness grades the GANN model had lower percentage of incorrect prediction compared to LDA model prediction.

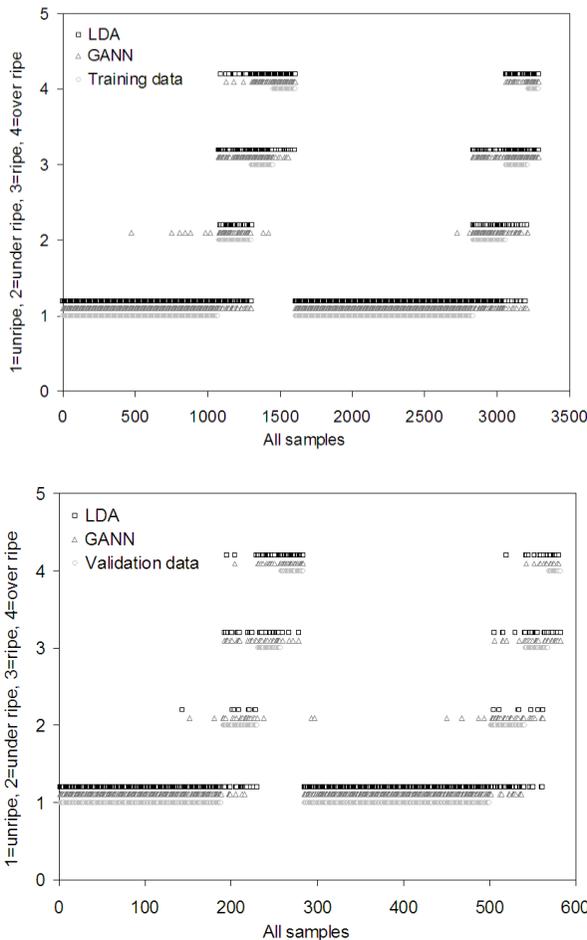


Figure 4. Original ripeness classes vs model prediction (LDA and GANN) using training and validation data

Table 3. Performance measure of LDA and GANN in training and validation phase

Performance Measures	Training		Validation	
	LDA	GANN	LDA	GANN
MAE	0.181	0.138	0.211	0.156
RMSE	0.443	0.375	0.492	0.399
% Good Classification	83.2%	86.3%	81.7%	84.5%

According to the performance measure indicated in Table 3, it obtained that the GANN was indeed slightly superior to the LDA. The value of MAE shown by GANN both in training and validation phase were 0.138 and 0.156, respectively where these were lower than LDA about 0.181 and 0.211, respectively. The value of RMSE released by GANN both in training and validation phase were 0.375 and 0.492, respectively; where again these were lower than LDA about 0.443 and 0.492. The GANN also performed higher percentage of good classification than LDA in training and validation phase, which were 86.3% and 84.5%, while LDA was 82.7% and 80.4%, respectively. Thus, the comparison between the features of LDA and GANN using this simulation data can be seen in Table 4. The GANN has no restriction in basic data assumption; need large sample consumption to train the model; good method to handle a large amount of data sets with less error rate compared to the LDA. A major drawback of GANN is the learning time where it was obtained that GANN still had longer time to train the feed-forward NN model even GA has been used in NN system in order to reduce the learning time to finding the optimal solutions of fittest parameters. Contrary, the LDA technique was performed faster time than GANN in order to finding its optimal solution.

Table 4. Comparison between the features of LDA and GANN

Features	LDA	GANN
Basic data assumptions	Data should be independent, nor-mally distributed and covariance matrix follow homogeneity assumption	No restriction on the data
Learning and prediction time	Fast in learning and prediction	less slowly than LDA in learning but fast as like LDA in prediction
Minimum sample	Small sample consumption	Large sample consumption to train the model
Amount of data handle	Often LDA lose some discriminant information and limited software are available when deal with large amount of data sets	GANN can handle large amount of data sets without any problem
Accuracy	Less than GANN	Good in accuracy
Error rate	Still has high error rate	Less error rate compared to LDA

6. Conclusions

The comparisons between GANN and LDA techniques with PCA have been used to handle hyperdimensional problem in ripeness classification of oil palm fresh fruit. The comparisons showed that in learning process to finding the optimal solution for fittest parameters GANN is less slowly than LDA, but as fast as LDA in prediction process. This less speed performance is due to the individual evaluation of set of parameters in order to find the best fitness in each class. In practice, the less speed in learning process is not much getting attention since the final best parameter obtained then would be used as fix parameter in the prediction. In the result, the GANN technique as proposed method is much better in terms of % good classification and less error rates in performance measure such MAE and RMSE. In data handle, the proposed GANN computation technique is suitable for large amount of data; since it was also known recently only limited some commercial statistical software are available to use simple LDA for large amount of data. Therefore, with referring to the accuracy, error rate and data handle, the GANN technique is considered to be superior used on NIRS spectral data as alternative technique for classifying ripeness grades of oil palm fresh fruit.

REFERENCES

- ABRAHAM, A., GROSAN, C., PEDRYCZ, W., 2008, *Engineering Evolutionary Intelligent Systems*, (Vol. 82), Springer Science & Business Media.
- BELHUMEUR, P. N., HESPANHA, J. P., and KRIEGMAN, D. J., 1996, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, In *Computer Vision—ECCV'96*, Springer Berlin Heidelberg: 43-58.
- CHIPPERFIELD, A., FLEMING, P., POHLHEIM, H., and FONSECA, C. 1994. Genetic algorithm toolbox user's guide. Department of Automatic Control and Systems Engineering, University of Sheffield, Version, 1.
- DUDA, R. O., HART, P. E., and STORK, D. G., 2001, *Pattern Classification* (Vol. 492), 2nd Ed., John Willey & Sons.
- DUDOIT, S., FRIDLAND, J., SPEED, T. P., 2002, Comparison of discrimination methods for the classification of tumors using gene expression data, *Journal of the American Statistical Association*, 97(457): 77-87.
- IELLER, N., 1982, *Multivariate Data Analysis: Practical and theoretical aspects of analyzing multivariate data with R*, NRJF, University of Sheffield.
- FISHER, R. A., 1936, The use of multiple measurements in taxonomic problems, *Annals of eugenics*, 7(2): 179-188.
- GOPAL, M., 2012, *Digital Cont & State Var Method*, Tata McGraw-Hill Education.
- GUYER, D., and YANG, X. 2000. Use of genetic artificial neural networks and spectral imaging for defect detection on cherries, *Computers and Electronics in Agriculture*, 29(3), 179-194.
- GUIDETTI, R., BEGHI, R., and BODRIA, L., 2010, Evaluation of grape quality parameters by a simple Vis/NIR system, *Transactions of the ASABE* 53(2): 477-484.

- HAEB-UMBACH, R., and NEY, H., 1992, Linear discriminant analysis for improved large vocabulary continuous speech recognition, In *Acoustics, Speech, and Signal Processing*, 1992. ICASSP-92., 1992 IEEE International Conference on (Vol. 1), IEEE: 13-16.
- HOLLAND, J. H. 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. U Michigan Press.
- JOHNSON, R. A., and WICHERN, D. W. 2007. *Applied Multivariate Statistical Analysis*. Six Edition, Prentice Hall, New York.
- KARIMI, B., MENHAJ, M. B., and SABOORI, I., 2010, Multilayer feed forward neural networks for controlling decentralized large-scale non-affine nonlinear systems with guaranteed stability, *Int. J. Innov. Comput., Inf. Control*, 6(11): 4825-4841.
- MCLACHLAN, G., 2004, *Discriminant Analysis and Statistical Pattern Recognition* (Vol. 544), John Wiley & Sons.
- MITCHELL, M., 1998, *An Introduction to Genetic Algorithms*, MIT PRESS.
- ÖZTÜRK, N., 2003, Use of genetic algorithm to design optimal neural network structure, *Engineering Computations*, 20: 8: 979 – 997.
- RENCHER, A. C., 2002, *Methods of Multivariate Analysis*, John Wiley & Sons.
- SILALAH D.D, REAÑO C.E., LANSIGAN F.P., PANOPIO R.G., BANTAYAN N.C., DAVRIEUX F., CALIMAN, J.P., YUAN Y.Y., and SUDARNO. 2016. Near infrared spectroscopy: a rapid and non-destructive technique to assess the ripeness of oil palm (*Elaeis guineensis* Jacq.) fresh fruit. *Journal of Near Infrared Spectroscopy* 24(2): 179-190.
- STEPNIEWSKI, S. W., and KEANE, A. J., 1996, Topology design of feedforward neural networks by genetic algorithms, In *Parallel Problem Solving from Nature-PPSN IV*, Springer Berlin Heidelberg: 771-780
- SVOZIL, D., KVASNICKA, V., and POSPICHAL, J. 1997, Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1), 43-62.
- SWETS, D. L., and WENG, J. J., 1996, Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8): 831-836.
- YALKIN, C., and KORKMAZ, E. E., 2012, A Neural Network Based Selection Method for Genetic Algorithms, Neural Network World - *International Journal on Neural and Mass-Parallel Computing and Information Systems* 22(6): 495-510.