Elaeis Guineensis Nutritional Lacking Identification based on Statistical Analysis and Artificial Neural Network

H. MUHAMMAD ASRAF[1], NOORITAWATI MD TAHIR[1], S.B. SHAH RIZAM[1] & R ABDULLAH[2]
[1]Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450, Shah Alam,
[2]Centre for Research and Innovation Management, Universiti Kebangsaan Malaysia, Bangi Selangor
MALAYSIA
Corresponding author: norita_tahir@yahoo.com or noori425@salam.uitm.edu.my

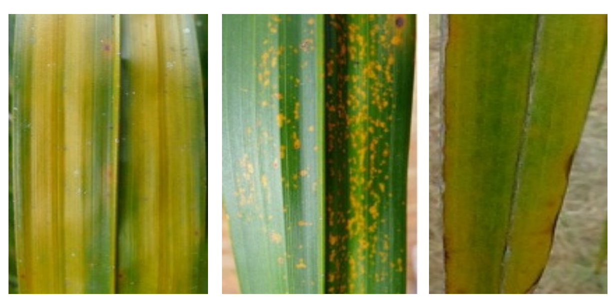
Abstract - In this study, nutritional disease classification of Elaeis Guineensis or widely known as oil palm is discussed. At present, nitrogen, potassium, magnesium are the main category nutrition deficient prevalent in oil palm plantation and these deficiencies can be identified based on the affected leaves surface appearance. Hence in this work, an alternative method based on image processing technique is proposed for identification of nutritional lacking in Elaeis Guineensis. Firstly, twenty seven features are extracted from three main groups that represent the Elaeis Guineensis leaf surface images namely RGB color features, RGB histogram based texture features as well as gray level co-occurrence matrix attributes. Next, feature selection via ANOVA and Multiple Comparison Procedure is conducted. Further, to verify the effectiveness of feature extraction and feature selection done, ANN is chosen as classifier. Initial findings based on classification accuracy attained confirm that the proposed method is capable to categorize nutritional lacking in Elaeis Guineensis with above 83% success rate prior to statistical analysis and over 86% with ANOVA as subset selection.

Keywords – Artificial Neural Network (ANN), Elaeis Guineensis, ANOVA, feature extraction, feature selection

1. Introduction

Nowadays, nutritional loss disease among palm oil crops is a great concern to ensure normal growth that will lead to greater crop yield production. Lack of nutrients or otherwise indirectly will affect plant growth and hence sufficient nutrient intake is vital to ensure optimal plant growth and productivity and this process can never be based on estimation. Essential minerals element can be tackled by adding specific amount of nutrients and accurate nutritional type to the plant through soil [4]. In addition, the availability of conventional approach to determine nutrient status contains in crops possess some disadvantages. For instance, soil analysis and foliar sampling for chemical analysis are both well known involves tedious procedure, cost intensive and time consuming [1], [2], [3] & [4]. At present in oil palm plantation, diagnosing of nutrient deficiency are mostly done manually that is based on judgments possessed by palm oil expert personnel that diagnose the causes and possibilities of oil palm diseases.

Practically, nutrient composition hold by oil palm trees is essential particularly during growth and development stages. The uptake in balance of micronutrient and macronutrient are important to ensure the healthy growth which will yield higher potential in oil palm fresh fruit bunch of production. As stated earlier, insufficient amount of nutrient in oil palm plants will affect its growth, productivity and indirectly reveal the symptoms that can be observed on the leaf surface as indication of one of the three categories nutritional lacking namely nitrogen, potassium, and magnesium. As depicted in Figure 1, nitrogen imbalance or disproportion is expressed by uniform pale green or yellow leaflets and a sharply reduced tree growth rate, with midrib tissues and rachis turn to bright yellow in colour and tapered leaflet that roll inwards [7] & [8].



Nitrogen Potassium Magnesium

Figure 1: Sample of leaf images suffers from deficiencies in nitrogen, potassium and magnesium nutrition respectively

Next, potassium deficient in oil palm is demonstrated by the altering pale green spots on the pinnae of older fronds transpire into orange - yellow rectangular spots and the leaf pinnae tips will further dry up. This is distinguishable based on the appearance of confluent orange spotting. Meanwhile symptoms due to lacked of magnesium are revealed due to colour variation on leaflets that is exposed to light specifically from olive green to ochre patches whereas the shaded pinnae remains green along with changes in fronds from ochre to bright yellow that turned to necrotic [7] & [8]. For this reason, in this study new algorithms are proposed for automated disease identification due to the three types of nutrient lacking through image processing technique and artificial intelligence classification. This paper is organized as follows: Firstly, the proposed methodology will be discussed that will explain in detail algorithms developed for feature extraction based on colour and texture of the leaf images along with statistical analysis as feature selection. Also, Artificial Neural Network (ANN) for disease recognition task will be elaborated. Next, experimental analysis and results attained based on the developed algorithms will be explained. Finally, we will conclude our findings.

2. Proposed Methodology

This section will describe in detail the methodology developed. Initially, 300 samples of Elaeis Guineensis leaf images are acquired from the field as database, with equal distribution in each diseased category; nitrogen, potassium and magnesium. Figure 2 illustrated the proposed methodology.

2.1 Feature extraction

Firstly, feature extraction algorithms to be developed was based on comprehensive interviews, inputs and feedback provided by the palm oil experts based on their vision judgment for detection of nutrient lacking via oil palm leaf surface appearance, as practice currently. Therefore colours and texture are chosen as the significant leaf characteristics to be investigated for feature extraction. From leaf colours, four features are extracted from the leaf images:

- Mean Image Intensity (R,G,B)
- Total mean

Secondly, as for histogram-based texture feature eighteen feature vectors are computed as listed below:

- Skewness (R,G,B)
- Kurtosis (R,G,B)
- Mean (R,G,B)
- Variance (R,G,B)
- Energy (R, G,B)
- Entropy (R,G,B)

Next, five features are extracted from gray level co-occurrence matrices (GCMs) namely:

- Variance
- Entropy
- Contrast
- Energy
- Homogeneity

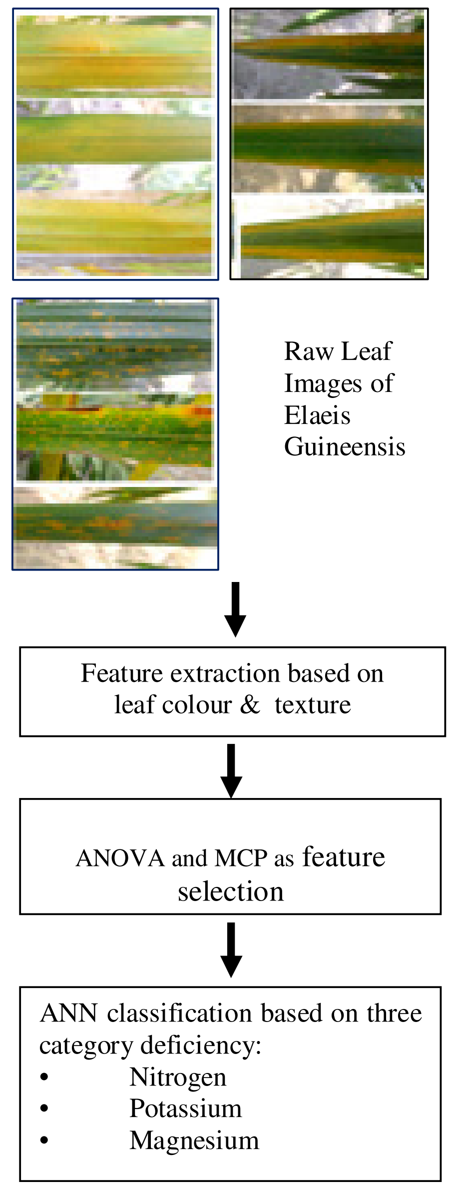


Figure 2: Overview of the overall methodology for Elaeis Guineensis disease classification

On the whole, twenty seven features are chosen as features. Further, prior to identification of disease based on ANN, these extracted features are normalized as to ensure scaling of heterogeneous datasets and to avoid dominant features towards others. Normalization of feature vectors into zero mean μ and unit variance, σ is as in (1):

$$x_i' = \frac{x_i - \mu}{\sigma_x} \tag{1}$$

Upon normalization, Kolmogorov-Smirnov test will be conducted to confirm each feature obey normal distribution criteria.

2.2 Statistical Analysis as Feature selection

The aim of feature selection is to select a subset of features relevant to the target concept. With feature selection, the sufficient optimize feature to the target concept will be identified. In addition, most researches have proven that feature selection process will increase or improve classifiers prediction. A good subset of features are features that is able to resemble original features the closest possible without compromising accuracy rate. Hence in this work, statistical analysis is selected as feature selection. One of the standard methods for testing statistical significance between a set of independent variables is Analysis of Variance (ANOVA). This is done by taking a single feature and the class associated with the data samples and measuring the significance of the class features by predicting the features mean value or the p-value of the feature set. Alternatively, Multiple Comparison Procedure (MCP) test can be used to determine which means amongst a set of means differ from one another. With ANOVA, evidence that the group means differ will be revealed but the main purpose is to determine which means are significantly different. Therefore, subsequent to ANOVA, MCP will be applied for this task. The MCP will compare the difference between pair of means and the groups that differ significantly will be revealed.

2.3Artificial Neural Network (ANN) as Classifier

The effectiveness of extracted features before and after statistical analysis will be verified using ANN. The ability of ANNs to express highly nonlinear decisions make them apposite for recognition of complex pattern since this classifier is capable to maintain accuracy although some input features are unsuitable or inadequate. In this study, the feedforward MLP neural network is used as classifier to test the efficacy of the feature vectors in classification of nutritional disease. MLP is widely used in numerous application for instance agricultural, military and medical sectors [11] [12] & [13] and proven to be able to solve real life pattern recognition problems.

3 Results And Discussion

In this study, the aim is to test the validity of feature vectors for identification of nutritional lacking in oil palm. As mentioned earlier, 300 leaf images of each deficiency category constitutes the database to generate the feature vectors in this study. Firstly, all twenty seven features listed earlier undergo normalization phase. Upon completion of normalization process, each feature went through normality test and it was found that nine features are not normally distributed and therefore these features are eliminated. One of the features is variance value extracted from GCM features whilst another eight features were from histogram-based texture features:

- Skewness (R,G,B)
- Kurtosis (R,G,B)
- Entropy (R, G)

The remainder eighteen features that have surpass both normalization process and normality test are known as Type I feature vectors and further experienced statistical analysis prior to classification. Accordingly, the statistical significance of all Type I feature vectors is verified based on statistical analysis.

In this analysis, null hypothesis will be discarded for p-value near zero and at least one sample mean is significantly different from the other sample means. Upon verification of statistically significant via ANOVA the selected feature vectors will undergo the MCP test. Thus, from the ANOVA test conducted with significant level of $\alpha = 0.05$, the outcome for p-values of feature vectors numerically indistinguishable from zero are identified. Further, MCP test is performed to determine the number of optimized feature vectors that will act as input to ANN classifier. Based on MCP and homogeneous subset test, the feature vectors listed below are verified to be significantly different between groups. Three features are from the leaf colour, five features are from histogram-based texture features whilst another three features are from GCMs respectively as outline below:

- Mean Intensity B
- Mean Intensity G
- Total mean

ISBN: 978-1-61804-141-8

- Mean R
- Variance G
- Energy G
- Variance B
- Energy B

- Entropy
- Contrast
- Homogeneity

As a result, the statistical analysis test has usefully lessened the feature vectors to eleven or 58% reduction from the initial features quantity and these feature vectors are known as Type II feature vectors. As aforementioned, Multilayer Perceptron (MLP) is chosen as classifier. A three-layer NN with weights adjusted using the Levenberg-Marquardt was trained to determine the relationship between the extracted features and the oil palm nutritional type.

Table 1: ANN Accuracy Rate (%) attained for identification of nutritional lacking (nitrogen, potassium and magnesium) in oil palm crops

Feature Category	**Type I feature vectors**	**Accuracy Rate**	**Type II feature vectors**	**Accuracy Rate**
A. Colors of leaf image	Mean Intensity (R,G,B) Total mean	71.11	Mean Intensity (G,B) Total mean	71.67
B. Histogram-based texture features of leaf image (based on RGB color)	Mean (R,G,B) Variance (R,G,B) Energy (R,G,B) Entropy (B)	62.22	Mean (R) Variance (G,B) Energy (G,B)	65
C. Leaf image Gray level co-occurrence matrices (GCMs)	Entropy Contrast Energy Homogeneity	**75**	Entropy Contrast Homogeneity	**75.56**
$A \cup B$	Mean Intensity(R,G,B) Total mean Mean (R,G,B) Variance(R,G,B) Energy (R,G,B) Entropy (B)	73.89	Mean Intensity (G) Total mean Mean (R) Variance (B) Energy (G,B)	75
$A \cup C$	Mean Intensity(R,G,B) Total mean Entropy Contrast Energy Homogeneity	81.11	Mean Intensity (G,B) Total mean Entropy Homogeneity	85.56
$B \cup C$	Mean (R,G,B) Variance(R,G,B) Energy (R,G,B) B Entropy Entropy Contrast Energy Homogeneity	74.44	Mean (R) Variance (G,B) Energy (B) Entropy Contrast Homogeneity	77.22
$A \cup B \cup C$	Mean Intensity (R,G,B) Total mean Mean (R,G,B) Variance(R,G,B) Energy (R,G,B) B Entropy Entropy Contrast Energy Homogeneity	**83.33**	Mean Intensity (G) Total Mean Mean (R) Variance (B) Energy (B) Entropy Contrast Homogeneity	**86.11**

R= Red; G = Green; B = Blue

ISBN: 978-1-61804-141-8

The NN was trained with 120 dataset of the three category disease type feature vectors and the remaining unseen data is used for testing. The MLP has an input layer that will correspond accordingly to the input features, one hidden layer and an output layer with three neurons to represent the three category nutritional lacking. Firstly, each category of Type I & II feature vectors are evaluated solely. Further, both group feature vectors are fused (combined) to evaluate and verify the most optimize feature vectors for identification of nutritional lacking in oil palm. Next, the classification results of Type I & Type II feature vectors will be elaborated as tabulated in Table 1.

It is observed that GCM leaf images features contributed to highest recognition rate for both categories of feature vectors specifically 75% and accuracy rate for this category is enhanced to 75.56% upon completion of statistical analysis. The worst classification rate is observed from histogram-based texture features leaf images that is below 70% either prior to statistical analysis or otherwise. Additionally, all accuracy rate upon implementation of both ANOVA and MCP or also known as Type II feature vectors attained higher recognition rate irrespective of fusion of group features or otherwise. This indicated that statistical analysis has succeeded the feature selection task by identifying only significant features to be chosen as inputs for identification of nutritional lacking based on ANOVA and MCP test. Further, with regard to fusion, for accuracy results, it is again observed that overall accuracy rate due to fusion of feature vectors has improved.

Hence, it can be concluded that data fusion has effectively increased classification rate. The best accuracy rate was based on fusion of all three categories of features specifically $A \cup B \cup C$ that has achieved success recognition rate of 83.33% prior to statistical analysis and 86.11% with ANOVA and MCP. This result again supported that fusion of features enhanced performance of the proposed method. On the whole, ANN as classifier is capable to perform nutritional lacking classification based on the proposed features.

4 Conclusion

In conclusion, this study has proven the ability of statistical analysis to identify significant features based on three feature groups that is color features of oil palm leaf images, histogram-based texture features of leaf images as well as GCMs of the leaf images to be for recognition of three main nutritional lacking in oil palm crops namely due to nitrogen, potassium and magnesium. As can be seen from the experimental results, ANOVA and MCP tests have successfully identified the dominant feature vectors based on improvement in all accuracy rate performance. This suggests that the proposed method along with statistical data analysis can be put into practice for nutritional lacking identification which will optimize plantation process and assist new planters. Future work includes validating and verifying the proposed method in field scenario or real oil palm plantation. Additional, several other classifiers namely Support Vector Machine (SVM) and Naïve Bayesian will also be opted for recognition of nutritional lacking in the next stage of research.

Acknowledgment

The authors would like to thank Universiti Teknologi MARA (UiTM) for the financial support to present this research. Also, this research work was funded by UiTM Excellence Fund Grant No: 600-RMI/ST/DANA 5/3/Dst (398/2011).

References

[1] Jacquemard, J.C., Oil Palm. *The Tropical Agriculturalist.* Macmillan, 1998.

[2] Hartley, C.W.S. (1988). The Oil Palm. *Tropical Agriculture Series.* Longman

[3] Goh KJ, Härdter R, General oil palm nutrition. *In Oil Palm: Management for Large and Sustainable Yields*. Eds T Fairhurst and R Hardter, pp. 191–230, Potash and Phosphate Institute and International Potash Institute, Singapore, 2000

[4] Rajaratnam, J.A., Chan, K.W. and Goh, K.H., The foundation for selecting leaf 17 for nutrient requirements of mature oil palms. *In: Proc. Conf. on Classification and Management of Tropical Soils (Joseph, K.T., ed.). Malaysian Society of Soil Science*, Kuala Lumpur: 340-348, 1997

[5] Foster, H.L. and Chang, K.C., The diagnosis of nutrient status of oil palms in Malaysia. *In: Proc. Int. Developments in Oil Palm (Earp, D.A. and Newall, W.,eds.).* Inc. Soc. of Planters, Kuala Lumpur: 290-312, 1997

[6] Foster, H.L., Tarmizi, A.M. and Zin, Z.Z., Foliar diagnosis of oil palm in Peninsular Malaysia. *In: Proc. 1987 Int. Oil Palm Conf.* (Halim, H.A.H., Chew, P.S., Wood, B.J. and Pushparajah, E., eds.). PORIM and Inc. Soc. of Planters,Kuala Lumpur: 244-261, 1998.

[7] Fairhurst, T., Nutrient Deficiency Symptoms and Disorders in Oil Palm (Elaeis guineensis Jacq.).

Pocket Guide. Potash & Phosphate Institute, Singapore, 31, 1998

[8] T.H. Fairhurst, J.-P. Caliman, R. Härdter and C. Witt., Oil Palm Series Volume 7, *Pocket Guide: Nutrient Disorders and Nutrient Management*. PPI/PPIC: Singapore and AgriSoft Systems Pte Ltd: Australia. 125, 2005

[9] Duda, R.O., Hart, P.E., and Stork D.G., *Pattern Classification. (2nd ed.)*. New York: Wiley-Interscience Publication, 2001.

[10] Dash, M. & Liu, H., Dimensionality Reduction. http://www.public.asu.edu/~huanliu/papers/dm07.pdf [12 April 2009].

[11] Chabaa, S.; Zeroual, A.; Antari, J.; , Application of the MLP neural networks for analyzing non Gaussian signal, *Multimedia Computing and Systems (ICMCS), 2011 International Conference on* , pp1-4, 2011

[12] Achkar, R.; Owayjan, M.; Mrad, C.;, Landmine Detection and Classification Using MLP, *Computational Intelligence, Modelling and Simulation (CIMSiM), 2011 Third International Conference on* , pp.1-6, 2011.

[13] Ozerdem, M.S.; Akpolat, V.; Classification of Bone Density with using Neural Networks, *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th* , pp.1-5, 2007

[14] Bishop, C. M., *Neural Networks for Pattern Recognition*, 3rd ed. Oxford: Oxford University Press, 1995.

[15] Haykin, S., *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York: Macmillan College Publishing, 1998

[16] Xinghuo Yu; Efe, M.O.; Kaynak, O.; A general backpropagation algorithm for feedforward neural networks learning, *Neural Networks, IEEE Transactions on*, 13 (1) pp.251-254, 2002

[17] Zhang, P.; Bui, T.D.; Suen, C.Y.;, Hybrid feature extraction and feature selection for improving recognition accuracy of handwritten numerals, *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on* , pp. 136- 140,2005

[18] Asraf Hairuddin, M.; Nooritawati Md Tahir, Shah Baki, S.R.; , Overview of image processing approach for nutrient deficiencies detection in Elaeis Guineensis, *System Engineering and Technology (ICSET), 2011 IEEE International Conference on* , pp.116-120, 2011

ISBN: 978-1-61804-141-8