Detection of Colletotrichum Gloeosporioides Fungus Isolates Development/Spread for Mango (Mangifera Indica L.) Cultivar from Electronic Nose using Multivariate-Statistical Analysis

N.S. Khalid[1], A.H. Abdullah[1], S.A.A. Shukor[1], N.D Dalila[2], M.J Masnan[3], H. Mansor[1], N.A Rahim[1], Fathinul Syahir. A.S[1]
[1]*School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Pauh Putra, 02600 Arau, Perlis.*
[2]*Malaysian Agricultural Research and Development Institute (MARDI) Sintok, 06050 Bukit Kayu Hitam, Kedah.*
[3]*Institute of Engineering Mathematics, Kampus Pauh Putra, 02600 Arau, Perlis.*
syahirahkhalid@unimap.edu.my

***Abstract*—Agriculture plays a very important role in Asia economic sectors. For Malaysia, it plays a big contribution towards the country's development. *Mangifera Indica L.*, commonly known as Mango, is one of the fruit that has high economic demand and potential in Malaysia export business. However, due to radical climate changes from hot to humid, Mango is exposed towards a number of disease and this will affect its production. *Colletotrichum gloeosporioides* is one of the major diseases that could occur on any types of Mango. This fungus can attack on fruit skin and leaf, therefore a method that able to detect and control it would be much appreciated. Hence, this paper shows that the presence of *Colletotrichum gloeosporioides* type of pathogen can be detected by using Electronic Nose (E-Nose). The E-Nose will detect the Volatile Organic Compound (VOC) that produced from this fungus. Further analysis and justification on its existence are completed by using one of Multivariate-Statistical Analysis method which is Principal Component Analysis (PCA).The analysis results effectively show that the PCA is able to classify the number of isolating days of this type of fungus after cultured. Furthermore the potential of pre-symptomatic detection of the plant diseases was demonstrated.**

***Index Terms*—Colletotrichum Gloeosporiodes; E-Nose; Principal Component Analysis (PCA); Volatile Organic Compound (VOC).**

I. INTRODUCTION

Mangifera Indica L. or Mango is very locally popular and has its origin from the Indo-Burma region. It has been grown in India for the last 4,000 years [1]. From India, the fruit was disseminated by traders, sailors and missionaries to the neo-tropics, Africa, Asia and Australia. Today, Mango is an important commercial crop not only in India, but also in Indonesia, Thailand and Malaysia. Today in Peninsular Malaysia, there are over 300 cultivars with fruits considerably in size, shape color, flavor and fiber content. Harumanis cultivar is special to Perlis and count in the national agenda as a specialty fruit from Perlis for the world. Table 1 shows Perlis exported 3.1 metric tons Harumanis mango to Japan in 2010 and aimed that the export demand will increase to 100 metric tons in 2020. [2]

A good Mango fruit is juicy and sweet. Mangoes are very nutritious and may also have attractive colors and shapes, with its own variation [3]. Because of humid tropics, Mango can be threatened by some fungus. Generally, a disease can be described as a series of damaging physiological processes caused by the consecutive irritation of the plant by a primary agent and pathogen.

There are many types of Mango disease such as powdery mildew, anthracnose, die back, bacterial canker, red rust, phoma blight, sooty mould and scab [4]. However, Anthracnose disease is the most common fungal that can causes great losses to the orchard[5]. Because of this reasons, this paper is focusing on this type of disease.

E-Nose is an intelligent device that able to mimic the role of the human nose. Currently, the E-Nose technology is applied for monitoring, recognition and classification, while in agriculture field, E-Nose achieved great momentum in recent decades in early diagnosis of fungal disease [6][7].

Table 1
The Target and Actual Export of Harumanis Mango to Japan

Year	Target	Export
2010	3.1 metric tan	500 kg
2011	3.1 metric tan	2.4 metric tan
2020	100 metric tan	?

II. LITERATURE REVIEW

A. Fruit Anthracnose

Fruit anthracnose are caused by a pathogen called *Colletotrichum gloeosporioides. C.gloeosporioides* facultative parasite belongs to the order of Melanconiales [8]. This disease are the main disease occurred on Mango and be considered as the most destructive disease.

Fruit anthracnose can resulted in a drastic pre- and post-harvest losses at any stages of mango's development. The disease will starts as tiny, necrotic, dark spots on the fruit surface [9]. These spots will enlarge to a bigger, round spots which often coalesce, forming larger, irregular patches on the skin. Figure 1 shows a comparison between a healthy mango which is on the left and mangoes infected with anthracnose which is on the right[3]. The disease causes large losses to young shoots, flowers, fruits and also affects fruits during storage. Fruits infected at mature stage can carry the fungus into storage and cause considerable loss during transit and marketing. *C.gloeosporioides* can be directly controlled by

suitable chemical.

Figure 1: Healthy Mango on the Left and Mangoes infected with Anthracnose on the Right.

B. E-Nose

The field of measurement technology in the sensors domain is speedily changing due to the availability of statistical tools to manipulate many variables simultaneously[10]. Because of this, it is led the way of generating dataset from sensor. Currently, multiple sensors or specifically multi sensor data fusion are more favorable due to rapid and better presentation [10]. Basically, E-Nose is formed by having more than one sensor or a specific array of gas sensors with different purpose, a signal collecting unit and suitable pattern recognition software, all controlled and executed by a computer [11]. This E-nose uses a static headspace technique for odour sampling process. Table 2 shows the list of sensors used. E-Nose have been used strongly to correlate specific physiological changes in plant health condition [12].

Table 2
E-Nose Type of Combination Senso

Sensor	Detection
Sensor 1(S1)	Organic solvent vapour
Sensor 2(S2)	Ammonia
Sensor 3(S3)	Water Vapour
Sensor 4(S4)	Gasoline and diesel engine
Sensor 5(S5)	Carbon dioxide
Sensor 6(S6)	Air contaminant
Sensor 7(S7)	Carbon monoxide
Sensor 8(S8)	Methane

C. Principal Component Analysis (PCA)

The pattern recognition and classification techniques that are mostly applied in E-Nose applications are; Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), k –Nearest Neighbor (k-NN), Artificial Neural Network (ANN) ,Feed Forward Back Propagation Neural Network (FFBPNN), Radial Basis Function(RBFNN), Generalized Regression Neural Network (GRNN)[13].One of the most popular exploratory data analyses in chemical sensors is PCA [10].

PCA is an unsupervised technique used for clustering data according to groups. PCA was first described by Karl Pearson in 1901 [14]. The idea of PCA is to finds a new set of linearly uncorrelated variables from a set of inter-correlated variables through an orthogonal transformation process. This technique will reduce the size of multidimensional data while rejecting redundancies without losing important information[15]. The new set of variables are known as Principal Components(PC) and are less than or equal to the number of original variables. The principal components are ordered in descending values of variances.

III. METHODOLOGY

C.gloeosporioides was isolated from infected fruits with anthracnose symptoms of the Mango trees grown in Malaysia Agricultural Research and Development Institute (MARDI) Sintok, Kedah, in 2016. It was isolated on Acidic Potato Dextrose Agar (PDA) and then maintained at room temperature. For this experiment, 7- day old, 14-day old, 21 - day old and 28-day old of Potato Dextrose Agar plate of *C. gloeosporioides* was used after the media cultured have been prepared. The data is collected by using the same pathogen in different days. The data is collected by using the same pathogen and same petri dish in different days. In comparison, the data without pathogens is taken for reference.

During the data collecting process, the distance between the sample surface and the tips of the E-Nose nozzle was kept constant to maintain the repeatability as shown in Figure 2. Figure 3 shows the E-Nose setup for collecting data. At the end, all the data will be stored in a database. Figures 4 till 7 shows the colony of 7- days old, 14-day old, 21-day old and 28-day old *C. gloeosporioides* cultures on PDA of top surface respectively.

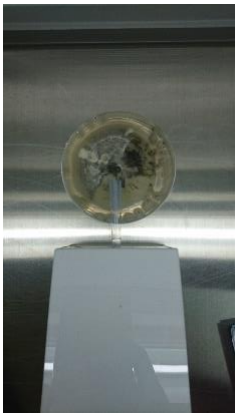

Figure 2: E-Nose Nozzle was Kept Constant to Maintain the Repeatability

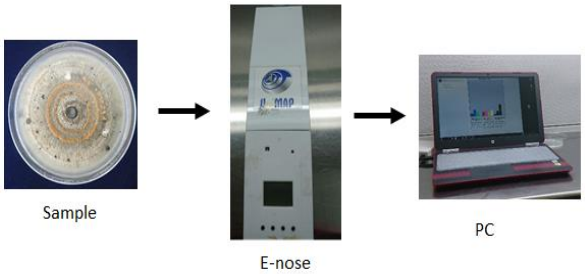

Figure 3: The E-Nose Setup for Collecting Data

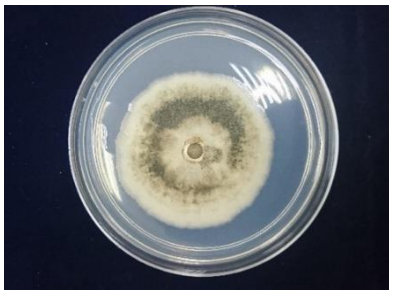

Figure 4: Colony of 7-day old *C. gloeosporioides* Cultures on PDA of Top Surface

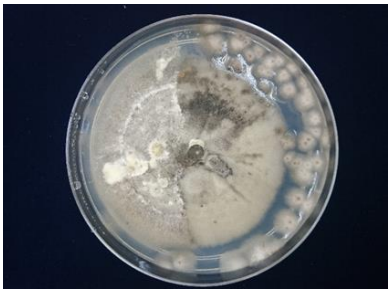

Figure 5: Colony of 14-day old *C. gloeosporioides* Cultures on PDA of Top Surface

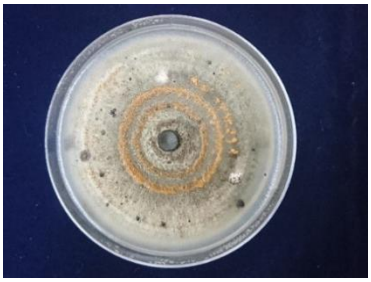

Figure 6: Colony of 21-day old *C. gloeosporioides* Cultures on PDA of Top Surface.

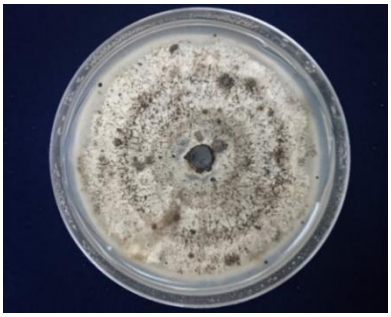

Figure 7: Colony of 28-day old *C. gloeosporioides* Cultures on PDA of Top Surface.

In this study, MATLAB and MINITAB software has been used to analyze the data. These software are widely used in engineering applications. The data used for PCA and loading plot analysis is from the baseline manipulation sample data. Baseline manipulation sample is a pre-processing of data done after the input data was collected. This process is important to remove the outliers and compensate the disturbance effect. Thus, it can enhance the data quality and can improve sensitivity. After that, PCA is used for grouping the data. The formula for baseline manipulation as shown below:

$$X_{bl} = X_m - X_o \quad (1)$$

where: $X_{bl} =$ The baseline manipulation data
$X_m =$ The sample data
$X_o =$ The reference data

IV. Results and Discussions

The results presented in this experiment can be divided into three parts. The first part of data analysis is about data pre-processing by using baseline manipulation technique. For the second part, analysis of data is grouped by using PCA. The last part is about loading plot of the sensor to explore relationship between original variables and subspace dimensions.

The graph of raw data from ambient, 7-day old, 14-day old, 21-day old and 28 -day old of cultured day from this experiment is shown in Figures 8 to 12 respectively. While the graphs after filtering by using baseline manipulation data are shown in Figures 13 to 16. The purpose of filtering data is to smoothing the data and to increase the Signal-to-Noise-Ratio (SNR). This means the data sets are refined into easily what we needs, without including other data that can be repetitive, inappropriate or even sensitive.

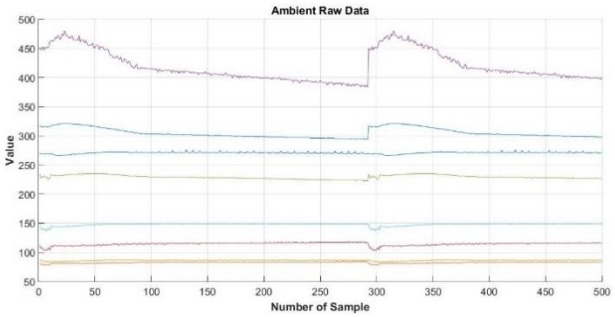

Figure 8: Ambient Raw Data before Filter

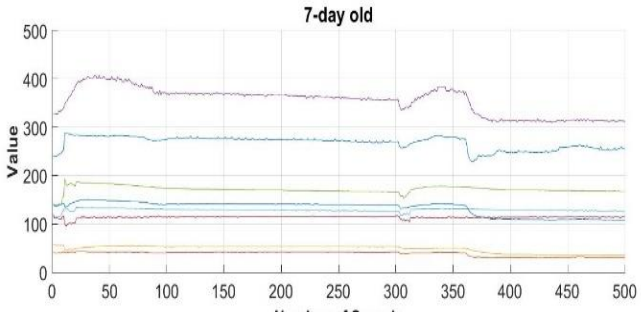

Figure 9:7-day old *C. gloeosporioides* Cultures Raw Data before Filter

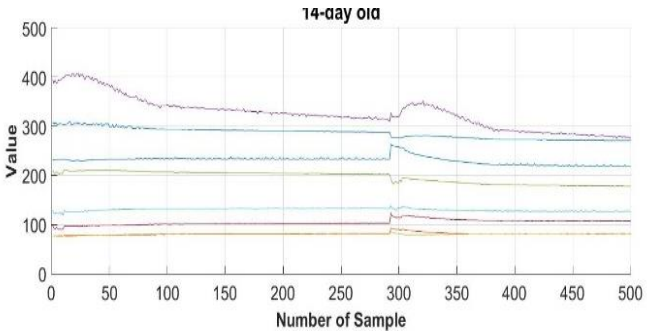

Figure 10: 14-day old *C. gloeosporioides* Cultures Raw Data before Filter

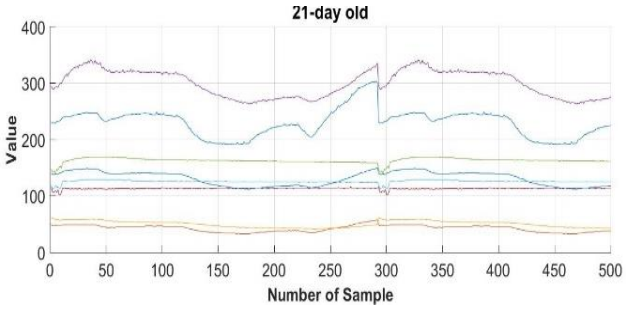

Figure 11: 21-day old *C. gloeosporioides* Cultures Raw Data before Filter

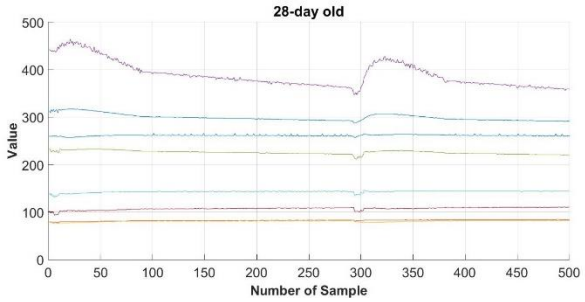

Figure 12: 28-day old *C. gloeosporioides* Cultures Raw Data before Filtering

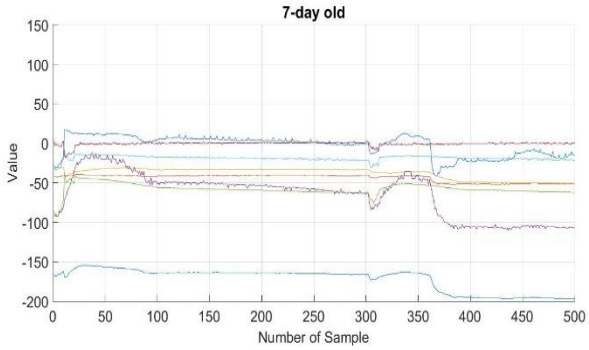

Figure 13: 7-day old *C. gloeosporioides* Cultures Raw Data after Filtering

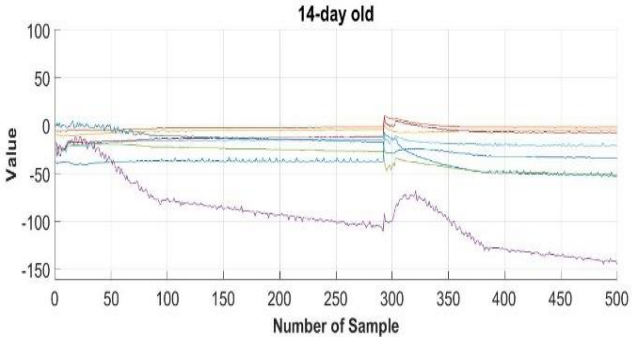

Figure 14: 14-day old *C. gloeosporioides* Cultures Raw Data after Filtering

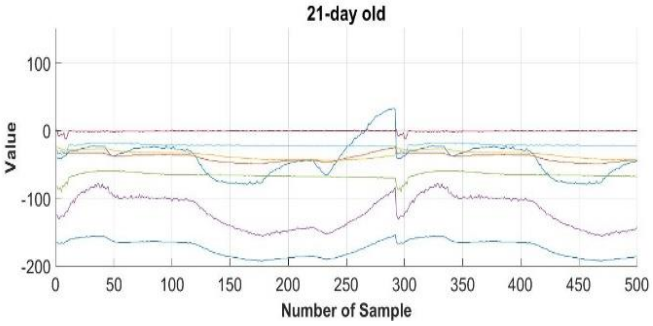

Figure 15: 21-day old *C. gloeosporioides* Cultures Raw Data after Filtering

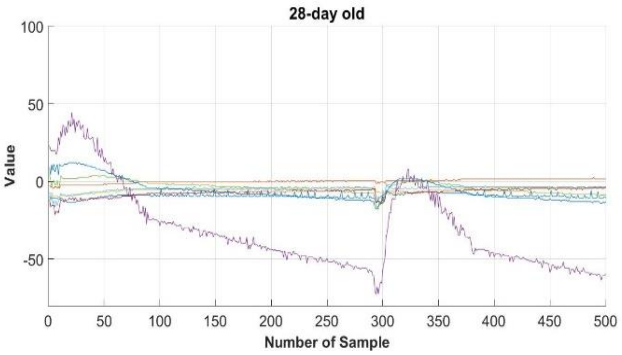

Figure 16: 28-day old *C. gloeosporioides* Cultures Raw Data after Filtering

Figure 17 present the score plot of PCA applied to *C. gloeosporioides* dataset. This scatter plot of the first two PCs that represent about 97.36% of the total variation in different group. The first PC is 80.31% accounted for the largest variance while second PC is 17.05%. From the sample scores on PC1 and PC2, grouping of data has been clearly separated. PCA score plot also displayed a clear separation between day 7 and day 14, day 28. In this assessment, the collected data are used to decide whether the data taken are acceptable or not and can be used for further analysis. Additionally, PCA is able to use for outliers data detection.

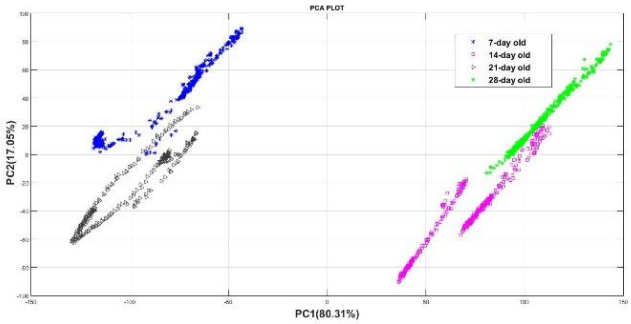

Figure 17: PCA Plot of *C. gloeosporioides* in Difference Number of Days

Figure 18 shows the graph of loading plot from a PCA. Loading plot is a plot that shows the relation among variables. Lines that go in the same direction and are near to one another demonstrate how the variables may be grouped. Based on loading plot, S2, S3 and S5 have large positive loadings on component 1. While S7 and S1 have large negative loadings on component 2. We believed that, these variables have been grouped together as they are strongly correlated from a statistical point of view. However, variables on opposite side of origin have negative correlation

To verify the presence of C. gloeosporioides, Olympus Compound microscope with 40X magnification was used to examine its morphology characteristics. Figure 19 shows the conidia morphology characteristics of C. gloeosporioides fungus from the microscope.

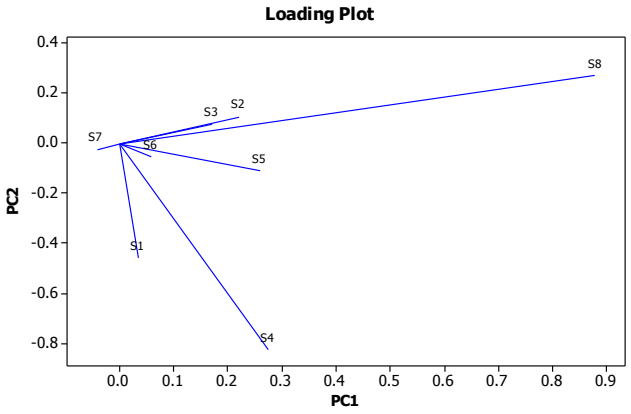

Figure 18: Loading plot of E-Nose detection of *C. gloeosporioides.*

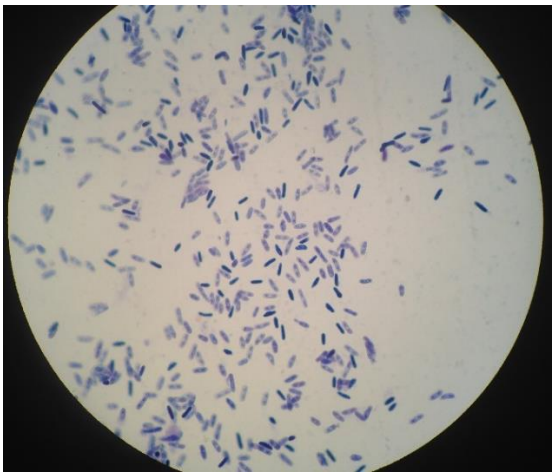

Figure 19: Conidia Morphology of *C. gloeosporoiodes* by using Microscope.

V. CONCLUSION

This study proved that E-nose are able to detect the Volatile Organic Compound (VOC) produced from infected Mangoes. The data taken from the samples are tested using PCA. Based on the comparative observations of the development of pathogen experiment that has been conducted, it has been proven that the combination of E-Nose and PCA is able to detect the presence of plant pathogenic bacteria. Previous studies on PCA have proven that this method is strongly useful to be applied before performing any classification techniques. The morphological of the presence pathogen have been proved by using microscope visualization and it's clearly shows the conidia morphological of *C. gloeosporioides.*

ACKNOWLEDGMENT

The authors are grateful to the researcher Malaysia Agricultural Research and Development Institute(MARDI) ,Sintok, Kedah for their advices , collaboration in order to acquire basic knowledge on certain plant pathogens,and providing the facilities throughout the preparation of this project. This project is funded by the Short-Term Grant. (9001-00560), Universiti Malaysia Perlis (UniMAP).

REFERENCES

[1] B. E. Method, "Chapter 1 Introduction 1.1," vol. 65, no. Ma 224, pp. 1–12, 2011.
[2] R. S. M. Farook et al., "Agent-based decision support system for harumanis mango flower initiation," Proc. - CIMSim 2011 3rd Int. Conf. Comput. Intell. Model. Simul., pp. 68–73, 2011.
[3] H. Pierre, "Mangiferin as a Biomarker for Mango Anthracnose Resistance," 2015.
[4] "Ploetz, R.C.. Diseases of Tropical Fruit Crops. Wallingford, GB: CAB International, 2003. ProQuest ebrary. Web. 19 October 2016. Copyright © 2003. CAB International. All rights reserved.," no. October, 2016.
[5] P. Chowdappa, C. S. Chethana, R. Bharghavi, H. Sandhya, and R. P. Pant, "Morphological and molecular characterization of Colletotrichum gloeosporioides (Penz) Sac . isolates causing anthracnose of orchids in India," Biotechnol. Bioinf. Bioeng., vol. 2, no. 1, pp. 567–572, 2012.
[6] S. Li, A. L. Simonian, and B. a. Chin, "Sensors for agriculture and the food industry," ECS Interface, vol. 19, pp. 41–46, 2010.
[7] L. Pan, W. Zhang, N. Zhu, S. Mao, and K. Tu, "Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography-mass spectrometry," Food Res. Int., vol. 62, pp. 162–168, 2014.
[8] A. Uaciquete, "Characterization, epidemiology and control strategies for the anthracnose pathogen (," 2013.
[9] Randy C. Ploetz, "Anthracnose of mango: Management of the most important pre- and post- harvest disease Randy," pp. 1–11, 2009.
[10] Maz Jamilah Masnan, Ammar Zakaria, Ali Yeon Md. Shakaff, Nor Idayu Mahat, Hashibah Hamid, Norazian Subari and Junita Mohamad Saleh, "PRINCIPAL COMPONENT ANALYSIS – ENGINEERING APPLICATIONS"
[11] K. Arshak, E. Moore, G. M. Lyons, J. Harris, and S. Clifford, "A review of gas sensors employed in electronic nose applications," Sens. Rev., vol. 24, no. 2, pp. 181–198, 2004.
[12] F. Martinelli et al., "Advanced methods of plant disease detection. A review," Agron. Sustain. Dev., vol. 35, no. 1, pp. 1–25, 2015.
[13] C. Engineering and P. Thani, "Classification and pattern recognition algorithms applied to E-Nose," no. Eict, pp. 44–48, 2015.
[14] R. Reris and J. P. Brooks, "Principal Component Analysis and Optimization : A Tutorial," pp. 212–225, 2015.
[15] P. Bickel et al., "Springer Series in Statistics.". Springer, 1997.