Multivariate Prediction Model for Early Detection and Classification of Bacterial Species in Diabetic Foot Ulcers

Azian Azamimi Abdullah[1,3], Nurlisa Yusuf[1], Mohammad Iqbal Omar[1], Ammar Zakaria[1], Latifah Munirah Kamarudin[1], Ali Yeon Md Shakaff[1], Abdul Hamid Adom[1], Maz Jamilah Masnan[1], Yeap Ewe Juan[2], Amizah Othman[2] and Mohd Sadek Yassin[2]

[1] Center of Excellence for Advanced Sensor Technology (CEASTech) Universiti Malaysia Perlis (UniMAP), 02600 Jejawi, Perlis, Malaysia

[2] Hospital Tuanku Fauziah (HTF), Jalan Kolam, 01000 Kangar, Perlis, Malaysia

[3] Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192 , Japan

Abstract

Many diabetic patients eventually develop foot ulcers are at risk for further infection and subsequent amputation if they are not treated promptly. Hence, this study is focused on identifying wild type strain bacteria and standard ATCC bacteria using e-nose which are PEN3 and Cyranose320. Data collected from both e-nose are processed using multivariate classifier such as LDA, KNN, PNN, SVM and RBF. The results indicate that rapid detection of bacteria using e-nose has increased the effectiveness, efficiency, reliability and reduced diagnosis time in identifying bacterial species on foot ulcer infection.

Keywords: Diabetic, Foot Ulcer, E-Nose, PEN3, Cyranose320, LDA, KNN, PNN, SVM, RBF.

1. Introduction

Over the years, a number of diabetes mellitus patients have been rising rapidly. In Malaysia, an alarming number of up to 600 thousand diabetes patients have been reported in the year of 2012 by the Ministry Of Health Malaysia and the numbers are still increasing [1]. This chronic and dreadful disease has resulted in various complications to the patient where they would have higher chances in getting ulceration of the foot. Diabetic foot ulcers typically found on any foot of diabetic patients due to poor feedback from nerve signalling, improperly fitting shoes, repetitive movement with stress and poor blood circulation, which often caused by harmful pressure on parts of the foot.

2. Literature Review

Bacteria found on diabetic foot ulcers are often polymicrobial where it consists of few types of bacteria in a colony. Severe infected foot usually yield polymicrobial isolates due to bacteria and fungi, whereas mild infections frequently yield monomicrobial [2]. Each bacterium will release Volatile Organic Compound

(VOC) which can act as a 'bio-marker' or odour identification for each bacterium.

E-nose acts as a useful diagnostic tool in providing rapid detection of odours, detection of microbial species, detection of harmful and dangerous chemicals as well as odours that unlikely to be detected by the human olfactory system due to certain threshold concentration. Unlike other analytical instruments, e-nose allows the identification of varieties of organic samples as a whole, where it is identical to a source that released the mixture without having to identify individual chemical species within the sample mixture [3]. An e-nose system typically consists of four main components which are headspace delivery unit, sensor array, an information-processing feature extraction unit, and odours classification unit that sequentially operate.

The multi sensor array contains an array of sensors with broad sensitivities that provide dynamic responses of the interaction between an odour sample and the sensing elements [1],[2],[5]. The first developed sensor array was a Metal Oxide Semiconductor (MOS), which could be used to detect 20 odours [5]. There are few types of transducers available in the e-nose. Each of these sensors has different selectivity and sensitivity threshold patterns where it could yield a unique "odour signature" for the VOC in the headspace of each sample under test [6].

Information-processing feature extraction unit could be used to apply for several purposes such as processing the data obtained from the array sensors, eliminate the interfering environment factor towards the array sensors, and pre-classify the recorded data [4]. As for classification unit, pattern analysis constitutes a very important building block in the development of gas sensor instruments given its ability of detecting, identifying and measuring volatile compounds [7].

3. Methodology

This experiment started with bacteria sample preparations in Hospital Tuanku Fauziah (HTF), Perlis, Malaysia under full supervisions of microbiologists. The bacteria sample preparations are done in both wild and standard American Type Culture Collection (ATCC) types of bacteria strain of *Escherichia coli* (ECOLI), *Pseudomonas aeruginosa* (PAE) and *Staphylococcus aureus* (SAE). Wild type strain bacteria is the bacteria that is directly found on diabetic foot ulcer in diabetic patients while standard ATCC bacteria is the bacteria that is being cultured and mature in an optimum laboratory environment. Those bacteria were cultured in a blood agar medium and was incubated at the temperature of 37 $^{\circ}$C.

After 6 and 24 hours of bacteria growth in the incubator, the bacteria were taken out and subjected to PEN3 and Cyranose320 data collection. The e-noses are set to specific settings in term of various time durations and repeated for the sniffing purging process during data collection.

Raw data produced from the e-noses are combined for the feature exraction as the dimension reduction approach. The dimension reduction is important to ensure that only important features are applied for further classification. This process is implemented using linear discriminant analysis (LDA) and is implemented using SPSS 17.0.

As for the classification and prediction accuracy LDA, Artificial Neural Network (ANN), Probabilistic Neural Network (PNN), Radial Basis Function (RBF) and K-Nearest Neighbour (KNN) were used in Matlab R2012a. Data is divided into two sets. The first set which include 60% of the data is used for train-

ing, and the other 40% of data is used for testing purposes. In order to validate the findings of the experiments, analyses using Gas Chromatography-Mass Spectrum (GC-MS) was also being done. Figure 1 shows the flow chart for this experimental research.

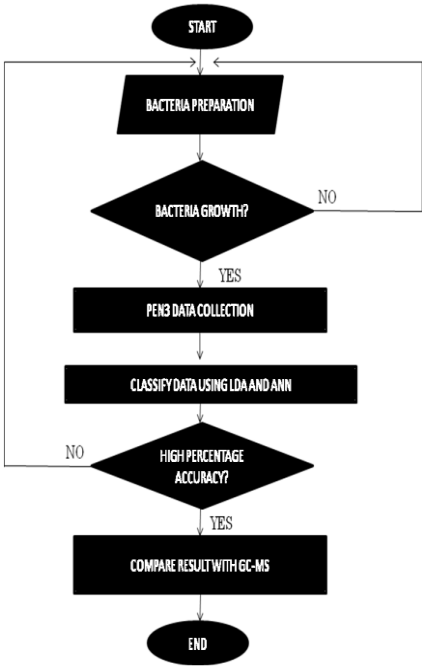

Fig. 1: Flow chart of the experiment

4. Experimental results

The results presented in this section are from 2 different types of E-nose which are PEN3 and Cyranose320. The results are analysed for every 6 hours start from the moment where the bacteria are inserted into an incubator for growth until 36 hours. Since the objective of this experiment is to investigate the ability of both e-noses in classifications of bacteria odours, hence the data analysis will be executed according to time variant. Further description of experimental procedure can be found in [3]-[4].

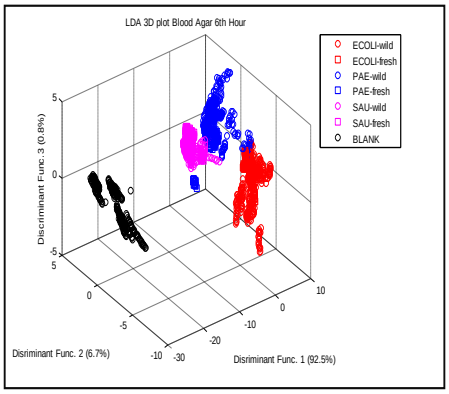

Fig. 2: LDA 3D Plot Bacteria Odours Classifications using PEN3 for 6 Hour of Bacteria Growth

Figure 2 shows the clustering of different types of bacteria using PEN3 for 6 hours of growth from the moment those bacteria is being incubated. Bacteria are fully classified into corresponding groups. Even though graphically there are some small portion of standard PAE and wild type strain Ecoli that plotted away from the main group's classifications; however the classification percentage accuracy is still high. In total, the discriminant functions describe 100% information of the overall bacteria.

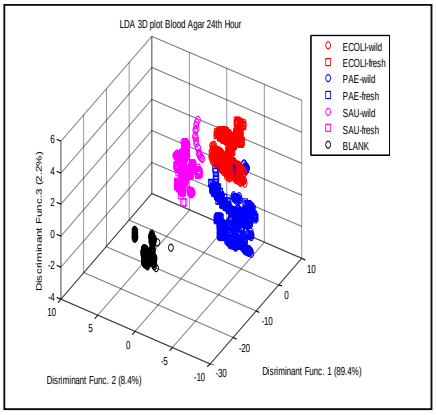

Fig. 3: LDA plot of bacteria odours classifycation using PEN3 for 24 hour of bacteria growth

Figure 3 shows the 3D LDA plot for bacteria classifications using PEN3 data collections on the 24th hour after those bacteria being incubated. The classifications and clustering between Ecoli, PAE and SAU are classified properly where there is obvious and visible clustering shown and could be differentiated according to colours. However, there are slight overlapped between wild Ecoli and wild PAE. There are also some portions of wild SAU that being plotted away from the main clustering of both wild and standard ATCC SAU. The discriminant function 1 is able to describes 89.4% information for the classification followed by 8.4% and 2.2% for discriminant function 2 and 3.

From the 6 and 24 hours of bacteria growth result that obtained, it is clear that even in the 6 hours of bacteria growth, PEN3 could classify all the bacteria as well as the blank blood agar petri dish into corresponding classification groups which indicate the differences of bacteria odours or VOC that released by the respective bacteria.

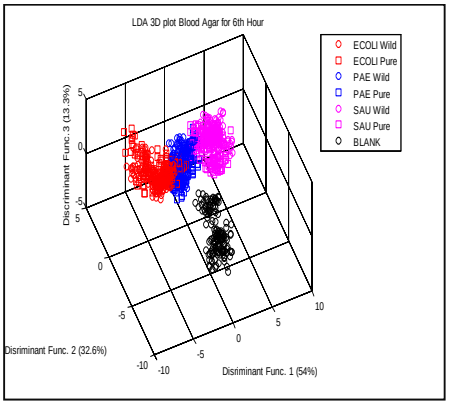

Fig. 4: LDA 3D plot bacteria odours classifications using Cyranose320 for 6 hours of bacteria growth

Figure 4 shows the 3D LDA plot of diabetic foot bacteria for 6 hours of growth by using Cyranose320. The 3 types of bacteria which are Ecoli, PAE and SAU are plotted near to each other and as well as the blank blood agar. Ecoli and PAE show slight overlapping between each other. The discriminant for the output graph is 54%, 32.6% and 13.3% respectively for discriminant function 1, 2, and 3.

Bacteria classification that presented in a form of visual graph with colour indicator for the 24 hour of bacteria growth by using LDA as feature extraction is shown in Figure 5. Ecoli, PAE and SAU show distinct clustering even there is some portion of overlapping between them. There is a small amount of standard ATCC SAU that plotted away from the main classifications of SAU bacteria. The discriminant function 1 for the output graph is 99.3%, for discriminant function 2 is 0.6% and 0.1% for discriminant function 3.

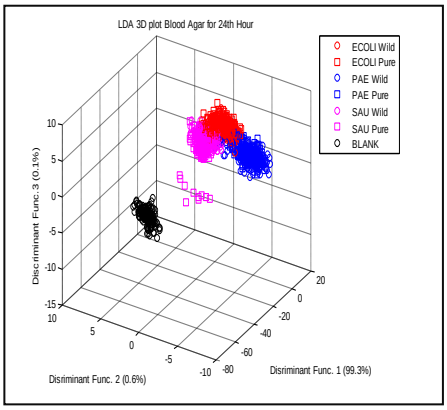

Fig. 5: LDA 3D plot bacteria odours cations using Cyranose320 for 6 hours of bacteria growth

Table 1-4 are divided into two parts. Table 1 and 3 demonstrate the classification findings using raw data without performing any dimension reduction. Whereas Table 2 and 4 illustrate the classification results using extracted features which was performed using LDA and followed by the respective classification approaches.

Even in the 6 hour of bacteria growth, the classification percentage accuracy of VOC that released from bacteria is high. Of all the methods of percentage accuracy, generally RBF shows the highest percentage of accuracy (100%) for bacteria classifications for both training and testing. The results obtained from this experiment are promising where both e-nose PEN3 and Cyranose320 could classify different types of bacteria even before the optimum time of growth for bacteria which is 24 hours. This would greatly helpful in the diabetic foot ulcer detection in a shorter time compared to conventional method that consumed up to 3 days long due to some compulsory medical ethic processes that could not be avoided. PEN3 data collections for 6 hours of bacteria growth shows classifications of bacteria where it is visible that each bacteria has own clustering. Wild and standard ATCC conditions of those bacteria are plotted in a same clustering too. This has greatly proven that even in 6 hours, PEN3 is able to classify the bacteria by sniffing the volatile organic compound that released.

Cyranose320 data collections show good results too even after 6 hour of bacteria growth. LDA plot from the Cyranose320 input data results show significant clustering and were plotted away from blank blood agar without overlapping. These findings show that accurate classification of bacteria species can be obtained even in 6 hour of bacteria growth. This somewhat proved that there is particularly volatile organic compound that already produced by the respective bacteria.

Table 1. Percentage Accuracy uses Raw PEN3 Sensor Data as Input

Method	6th Hour		12th Hour		18th Hour		24th Hour		30th Hour		36th Hour	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
LDA	98.6	98.2	98.7	98.6	99.6	99.8	97.6	97.1	99.5	99.4	99.4	99.3
KNN	100	100	99.9	100	100	100	100	99.9	100	100	100	100
PNN	100	100	100	99.9	100	99.3	100	99.8	100	100	100	100
SVM	98.1	99.9	99.7	100	100	100	99.7	99.9	100	100	100	100
RBF	100	100	100	100	100	100	100	100	100	100	100	100

Table 2. Percentage Accuracy for Bacteria Classifications by using the LDA Discriminant Value from PEN3 as Input

Method	6th Hour		12th Hour		18th Hour		24th Hour		30th Hour		36th Hour	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
LDA+KNN	99.9	100	99.9	100	100	100	100	99.9	100	100	100	100
LDA+PNN	100	98.7	100	98.4	100	99.2	100	99.6	100	99.5	100	99.4
LDA+SVM	98.2	99.4	99.8	100	100	100	98.2	98.4	100	100	100	100
LDA+RBF	100	100	100	100	100	100	100	100	100	100	100	100

Table 3. Percentage Accuracy for Bacteria Classifications by using Raw Cyranose 320 Sensor Data as Input

Method	6th Hour		12th Hour		18th Hour		24th Hour		30th Hour		36th Hour	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
LDA	98.3	96.7	99.5	98.9	93.4	91.9	96.9	95.7	99.5	96.5	98.3	97.8
KNN	96.5	87.3	95.7	83.5	93.2	80.5	93.7	80	96.2	86.8	92.8	83.3
PNN	100	89.2	100	83.6	99.7	81	100	78.5	100	87.5	99.5	85
SVM	99.7	99.6	100	100	99.4	99.2	95.4	97.8	98.6	100	98.8	99.8
RBF	100	100	100	100	100	100	100	100	100	100	100	100

Table 4. Percentage Accuracy for Bacteria Classifications by using the LDA Discriminant Value from Cyranose320 as Input

Method	6th Hour		12th Hour		18th Hour		24th Hour		30th Hour		36th Hour	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
LDA+KNN	99.2	97.8	99.6	98.9	96.5	94.3	98.2	97.6	99.4	99.1	99.1	98.9
LDA+PNN	100	84.6	100	84.7	100	80.7	100	84.8	100	84.6	100	82.4
LDA+SVM	97	97.5	98.6	98.9	92.4	95.2	94.2	97.6	96.4	98.9	98.3	98.4
LDA+RBF	100	100	100	100	100	100	100	100	100	100	100	100

To make confirmations regarding the classifications that made of various types of bacteria that commonly found on diabetic foot ulcer, GC-MS is used to detect the VOC released by each bacterium. Figure 6 shows the GC-MS result for wild and standard ATCC condition of E. coli. The upper graph of Figure 6 indicates the VOC compound that presence in wild type E. coli while the bottom part shows the present compound in standard ATCC E. coli. There are two obvious similar compounds for both wild and standard ATCC E. coli which are the 1-Hexanol, 2-Ethyl, and 3-Trifloroacetoxydodecane. However the percentage of 1-Hexanol, 2-Ethyl are lower in wild E-Coli compared to standard ATCC E. coli. Besides that, there is also one obvious VOC that presence in wild E-Coli that is the Benzaldehyde.

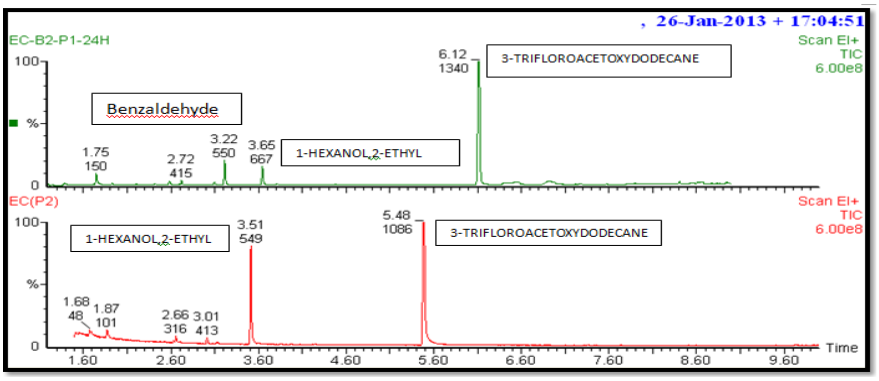

Fig. 6: GC-MS Result for Wild and Standard ATCC Condition of E.Coli

From the GC-MS results it can be concluded that both e-noses could classify the bacteria into specific groups. Since the compound for both wild and standard ATCC bacteria are almost the same, hence both wild and standard ATCC conditions of specific bacteria are classified in a similar grouping in 3D LDA plot. There are differences of compounds found for different species of bacteria. However, for the standard and wild bacteria, the VOC match each other's bacteria species. Normally optimum time of growth for bacteria is within 24 hours. However, this study only need six hours to culture the bacteria species and further classification can be implemented. This quick process is useful for diabetic foot ulcer detection compared to the conventional method which used up to 3 days long due to some compulsory medical ethic processes that could not be avoided.

5. Conclusion

Early detection of bacteria species in diabetic foot ulcer is very critical in an effort to reduce the number of imputation of diabetic patient. An attempt to resolve the problem is to conduct an experiment to investigate the nature of bacteria morphology from the perspective of sensor technology. From the study it can be concluded that both e-noses i.e. PEN3 and Cyranose320 are able to mimic the human sense of smell and usefull for diabetic foot ulcer bacteria detection. Human sense of smell is closely related to function of nose that can specify different odorants by detecting the smell, sent the information to the brain and interpret the specific odorant detected. The same processes are adopted by both e-noses.

In this experiment, the most concerned is that both e-noses could detect the bacteria even before the optimum time of growth within 6 hours time. This would greatly improve the detections compared to conventional methods that would last for 3 days. This alternative method for diabetic foot ulcer detections could possibly reduce the number of patient's to undergo amputations due to improper time of treatment and antibiotic. In conventional approach, before the clinical testing result is obtained, physician would provide general medication to the patient that may not be respond certain antibiotic, and this would caused worse infection. Then only after the clinical testing result is available, suitable treatment and medication can be provided to the patient. For classification purposes, LDA, KNN, PNN, SVM and RBF were applied. Basically, these methods are partitioned into two; parametric method (such as LDA) is assumed to follow the Gaussian conditional densities, and the nonparametric methods (like KNN, PNN, SVM, RBF) with an assumption of unknown group conditional densities. Therefore, in this experiment we wish to employ both approaches to see which method is suitable.

From the findings, if the raw data from the e-noses is used for classification, LDA appeared to perform better than other method. However, if classifications using extracted features are performed, RBF seem outperformed the rest of the methods. Two conclusions can be drawn from the findings. First, if the actual data i.e. without any modification is used (all information are being used) it seems that method follows the Gaussian distribution is suitable for classification of bacteria species. Second, by using the extracted features, where only important features are applied (with certain loss of information), nonparametric approach i.e. RBF best classify the bacteria species.

6. Acknowledgement

This work is supported by the ERGS grant 9010-00010 and 9010-00015 by the

Ministry of Higher Education, Malaysia. The authors would like to thank Hospital Tuanku Fauziah, Kangar, Perlis for their collaboration in this project.

7. References

[1] Zakaria, A., Shakaff, A. Y. M., Adom, A. H., Ahmad, M. N., Masnan, M. J., Aziz, A. H. A., Fikri, N. A., Abdullah, A. H., and Kamarudin, L. M., "Improve classification of Orthosiphon stamineus by data fusion of e-nose and tongue sensors.," *Sensors (Basel, Switzerland)*, vol. 10, no. 10, pp. 8782–96, Jan. 2010.

[2] A. Zakaria, A. Y. M. Shakaff, M. J. Masnan, M. N. Ahmad, A. H. Adom, M. N. Jaafar, S. a Ghani, A. H. Abdullah, A. H. A. Aziz, L. M. Kamarudin, N. Subari, and N. A. Fikri, "A biomimetic sensor for the classification of honeys of different floral origin and the detection of adulteration.," *Sensors (Basel, Switzerland)*, vol. 11, no. 8, pp. 7799–822, Jan. 2011.

[3] A. A. Abdullah, N. Yusuf, A. Zakaria, M. I. Omar, A. Y. Shakaff, A. H. Adom, L. M. Kamarudin, Y. E. Juan, A. Othman, and M. S. Yassin, "Bacteria Classification Using E-nose for Diabetic Wound Monitoring," *Applied Mechanics and Materials*, vol. 339, pp. 167–172, 2013.

[4] A. A. Abdullah, T. W. Jing, C. A. Sie, N. Yusuf, A. Zakaria, M. I. Omar, and A. Yeon, "Rapid Identification Method of Aerobic Bacteria in Diabetic Foot Ulcers using E-nose," *Advanced Science Letters*, vol. 20, no. 5, pp. 37–41, 2014.

[5] Alphus D. Wilson, and Manuela B. Applications and Advances in Electronic-Nose Technologies. (2009) 2-30.

[6] Hassan A.R. Future of Diabetes in Malaysia, Seminar Talk, Ministry of Health Malaysia. (2012)

[7] Shalbha T., Daliparthy D.P., Awanindra D., Sanjiv K.G., Madhukar R., Surya K.S. Microbiological and Clinical Characteristic of Diabetic Foot Infections in Northen India, African Journal of Clinical and Experimental Microbiology. (2010) 330-332.

[8] K. Arshak, E. Moore, G.M. Lyons, F. Harris, and S. Clifford. *A review of gas sensors employed in e-nose application*, Sensor Review, vol. 24, no. 2 (2004) 181-198.

[9] Persaud, K., and Dodd G. *Analysis of Discrimination Mechanisms in the Mammalian Olfactory System Using a Model Nose*, Nature, 299 (1982) 352-355.

[10] Geoffrey C. Green, Adrian D.C. Chan, and Rafik A. Goubran. An Investigation into the Suitability of Using Three E-nose Instruments for the Detection and Discrimination of Bacteria Types, 4 (2006) 1-2.

[11] Teodoro A., Jesus L., Jose A. Paredes, Fernando J. Alvarez, and Jose I. Suarez. E-nose Based on Independent Component Analysis Combined with Partial Least Squares and Artificial Neural Networks for Wine Prediction, 18 (2012) 2-3.