Classification of Ammonia Odor-profile Using k-NN Technique

Fathimah Abdul Halim[1], Muhammad Sharfi Najib[1], Kamarul Hawari Ghazali[1], Mohd Razali Daud[1], Muhamad Faruqi Zahari[1], Abd. Aziz Mohd Azoddein[2]

1 Faculty of Electrical & Electronic Engineering, University Malaysia Pahang, 26 600, Pekan, Pahang.

2 Faculty of Chemical and Natural Resources, University Malaysia Pahang, Gambang, 26300, Pahang.

***Abstract*-- This paper presents the application of k-NN in classifying the low and high concentration of ammonia. High concentration of ammonia in water causes serious problems to the water environment and living things in the water. Instruments that can directly detect ammonia concentration without any chemical treatment added are limited. Thus, this paper presents the classification of ammonia odor-profile using E-nose and the classification of ammonia in water using k- nearest neighbor (k-NN) with 97% success rate.**

***Index Terms*-- Ammonia, Classification, k-NN, E-Nose.**

I. INTRODUCTION

Ammonia (NH3) is a combination of one hydrogen and three nitrogen atoms that have characteristics such as easily dissolve in water, colorless, and have a unique pungent odor. Ammonia in the water gives more serious harm because it is very toxic to aquatic organism. Additionally, the concentration of ammonia in the water can increase due to agricultural excess and decomposition of biological waste [1]. The high concentration of ammonia results in the increasing of ammonia toxicity, especially for fish and mammals [2]. Its toxicity effects on the brains of vertebrates that leads to convulsion and death [3, 4]. Nevertheless, monitoring and controlling the ammonia level in the water is very crucial for the water environment, fish and farming industry. The conventional ammonia detection methods that have been used to determine ammonia are Kjeldahl method [5, 6], Nessler's reagent calorimetric method [7], and the most common methods are spectrometers and spectrophotometers [8]. However, these types of detection methods have limitations in labor intensive, expensive, complex operation and not suitable for outdoor use. On the other hand, the samples used for these instruments require samples to be added and treated with toxic chemical reagent. The availability of instrument that can straight away detect the presence of ammonia in water is currently limited.

However, beside these methods, Electronic Nose (E-nose) is one of the electronic-based instrument that is used to detect the ammonia based on the unique profile of an ammonia odor [9]. An E- nose is the intelligent instrument that classifies the chemical odors mimicking a human. It consists of a gas sensor array and various pattern recognition algorithms which are the sensor that are used to produce a unique profile of an odor. This odor will be analyzed using pattern classification methods [10, 11]. The sensor functions as a sniffer of the vapor from a sample and provides a set of measurements data. Gas sensors tend to have very wide-ranging selectivity thus it will respond to many different samples provided. The detection of odors and gases has been applied to many industrial applications such as medical [12], environmental [13], and food product [14, 15] and one of the most important applications is the detection of hazardous gases [16]. Ultimately, the sensor used can be very portable and small for convenience purpose, diverse operations and also have analysis and detecting capabilities [17]. Electronic nose is inexpensive, fast, simple and a convenient method to detect ammonia in the water.

There are several pattern recognition systems that are used for classification such as Principal Component Analysis (PCA) [18], Support Vector Machine (SVM) [19], Discriminant Factor Analysis [20], and Artificial Neural Network [21] and also KNN. In this study, K-nearest neighbor method is used to compare the concentration level of ammonia samples. K-nearest neighbor is known as statistical classification algorithm and widely used for classifying data based on lowest distance training. Previously researchers' results showed that k-NN couple with e-nose produced percentage accuracy above 99% [22-23]. K-nearest neighbor (KNN) characteristics are simple but robust classifier and qualified to produce high performance outcomes even for complex applications [24-26]. This study presents the implementation of k-NN classification technique using sensor reading from the E - nose for Ammonia classification. The E - nose is used to produce a unique profile of ammonia in the water. Using these features and profile of ammonia in k-NN classification will produce the percentage accuracy performance. This accuracy is indicating the reliability of k-NN in the pattern recognition of ammonia.

II. Methodology

A. Ammonia Sampling and Preparation

There are two categories of ammonia solution, high concentration (25 ppm) and low concentration (5 ppm) respectively. The concentration range can be referred in Department of Environment (D.O.E) standard [27].The sample of ammonia in the water will be prepared into five different dilutions. The samples were filled into 5 closed vials for each different diluted solution. The solution will be diluted until each vial contains 10ml of solution. Each sample of five different concentrations has ten repeated reading data taken by the e - nose. Before the concentration samples was insert in the e-nose, the samples concentration were validated using chemical laboratory standard method known as spectrophotometer. Spectrophotometer is known as a commercial instrument that has been used by industries to detect ammonia concentration in liquid phase. This commercial instrument is used to make sure the diluted ammonia concentration follow the standard range stated by DOE and act as benchmark data for e-nose.

The confirmed concentration of high and low ppm then were proceed using E-nose detection. Then there are supposed to produce 50 sample data ready. The E-nose consists of 4 array sensors that will detect the odors of ammonia concentration at the same time. Ultimately, the E-nose reading will has a dimension of 4 x 10 for each sample. Each sample will consist of 40 E-nose reading. With five different concentrations, sensor inside the E-nose was activated and sensed the odors of ammonia concentration that produced in total 2000 raw data samples reading. The measurement of the E-nose is based on the values of electric resistance of the sensor when ammonia odor exists. These sensor measurement produces a unique odor profile of ammonia. Therefore, these odor profile is used as an odor mark that have been used for odor classification.

B. Data analysis

After completing the raw data samples, the preprocessing technique was applied. Before beginning with k-NN classification method, all the sample data sets were normalized per sample region to obtain the features and odor-profiles graph. According to B.Tudu et al [28], normalization technique is the vital step in analyzing data in order to increase the accuracy of the classifier performance for odor profile of the e-nose. Normalization technique was carried out by using matlab software by adjusting the values measured data on different scales and produces specific statistical graph profile. The features then were analyzed by using box plot to visualize and summarize the data. The mean of the normalized data sample were also obtained.

These mean of normalized data sets were then divided into k-NN training and sample data. k-NN classification method consist of two data which are training and sample data. Training data is the data that has been stored as the sampling data reference. Testing data is unknown data that is presented for testing based on the algorithm that calculates its K closest neighbor and its class is assigned by voting principle. k-NN operational principle is to classify the whole data into training and sample data point and calculation of the lowest distance between training and sample data is called as the nearest neighbor.

By implementing a k-NN algorithm for classification odor profiles of ammonia is effective because k-NN is suitable for large training data and capable to produce high performance result even for complex operations. k-NN algorithm has been widely used in electronic nose application in determining the classification performance of the samples [29]. In features matching, the distance between training and sample data were calculated using Euclidean distance equation [30].

$$d\ (x,y) = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2} \qquad (1)$$

Where d is the distance between x_i and y_i, D is a number of input features, x_i is the input features (samples data) and y_i is the feature vector (training data).

This technique consists of several steps, the initial step of k-NN is setting the values of the k parameter, and usually k parameter is an odd number where the odd number decreases the prediction error of the samples [30]. The samples data of ammonia will be classified into training and sample data. These sample data is classified according to the majority of parameter k-nearest neighbor in the training data. Then, the sample data will be assigned based on the closest distance in the training group data.

III. Results

The raw data has been normalized and pre-processed in order to obtain a unique profile odor of ammonia with different concentrations. Five different concentrations of ammonia were divided into groups of high and low concentration. All the sample data sets were normalized per sample region.

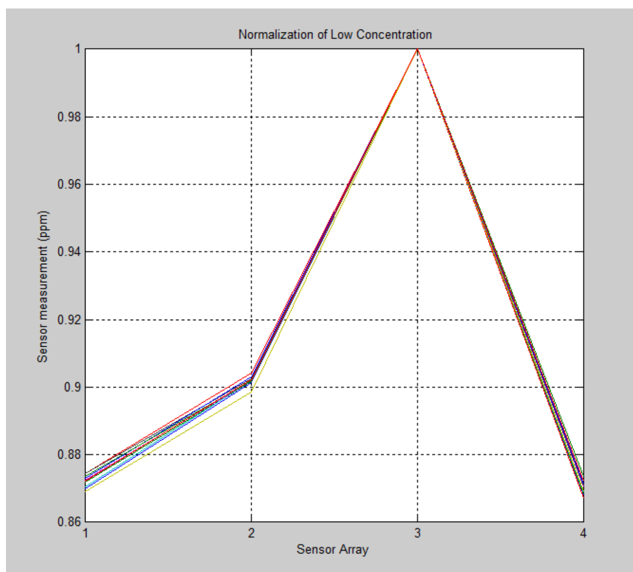

Fig. 1. A sample of Normalized data for low concentration of ammonia.

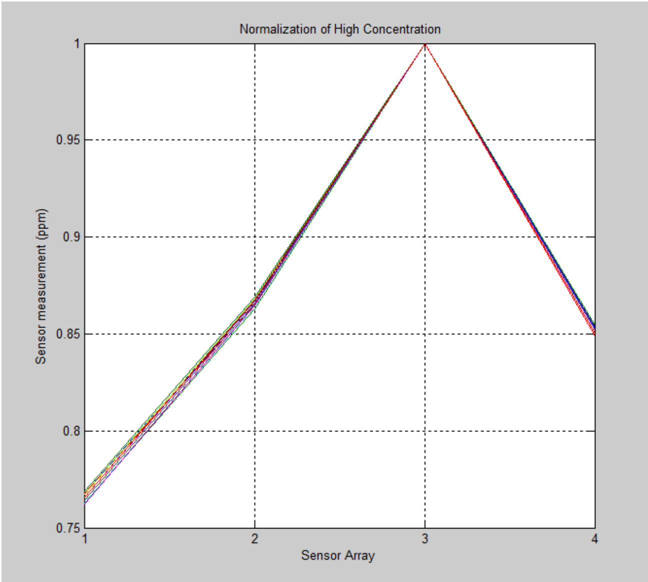

Fig. 2. A sample of Normalized data for High concentration of ammonia.

Fig. 1 and Fig. 2 show a sample measurement of normalized data for high and low concentration of ammonia respectively. Based on the figures above, both graphs show the data collected from sensor 1, 2 and 4 have noticeable differences in patterns and values. These graphs show that low and high concentration of ammonia has dissimilar features and profile. After normalization technique was applied, these features were improved by using boxplot. Boxplot is used to summarize and visualize the data set in a graph.

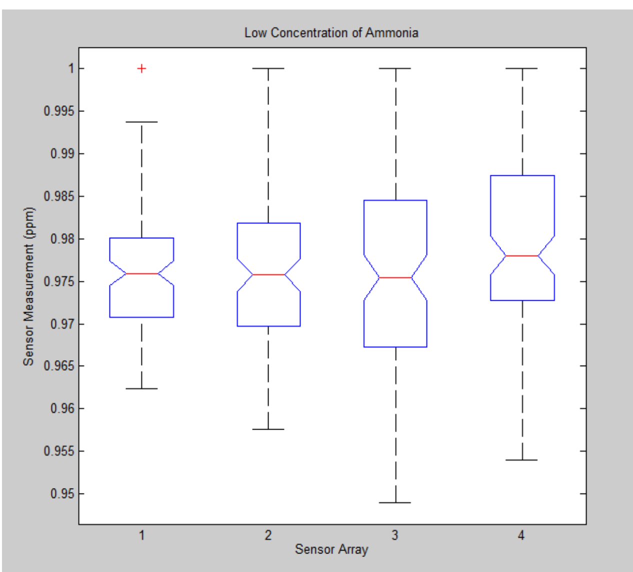

Fig.3. Box plot for low concentration of ammonia

Box plot consists of median, the approximate quartiles, and the lowest and highest data point. Based on Figure 3 and 4, the median for each dataset is indicated by the red center line while the lower and upper quartiles are the edges of the blue line, which is known as the inter-quartile range (IQR). The maximum and minimum values are indicated by the end of the black line and known as a whisker. In figure 4, the four boxplots have nearly identical median values. Figure 3 shows the notches in the boxplot are not overlapped therefore it is concluded that the median of the four sensors is significantly different. It is because notches display the variability of the median between sensors. The sample data sets then were proceeding to the mean of normalized data pattern.

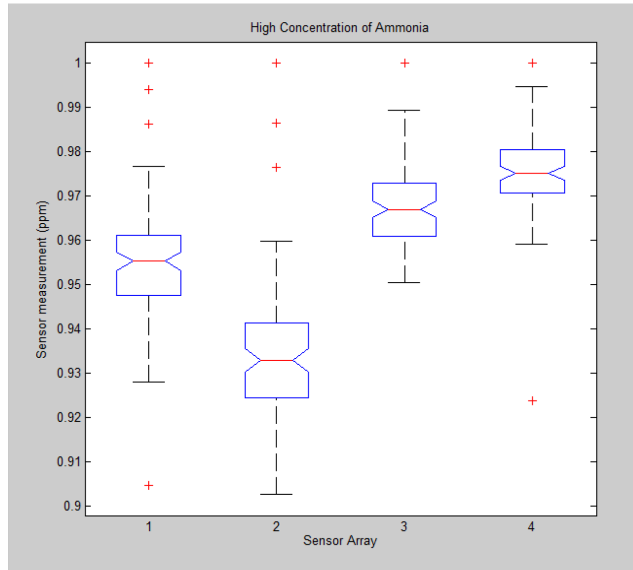

Fig.4. Box plot for high concentration of ammonia

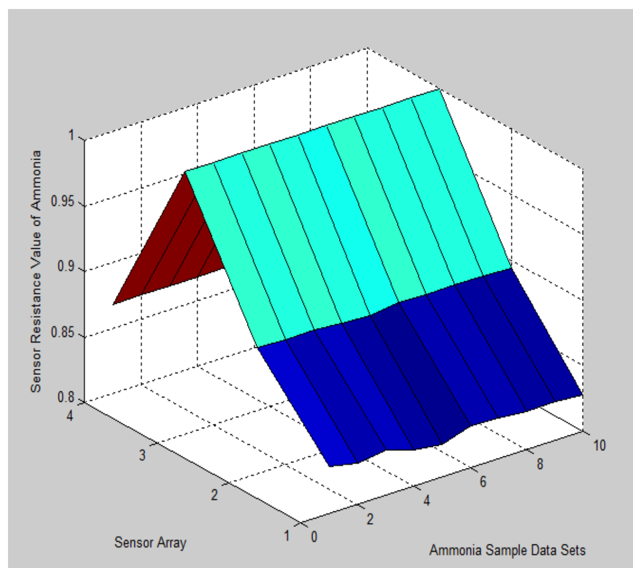

Fig.5. Mean of Normalized Data Samples from Low Ammonia Concentration

Figures 5 and 6 are three dimensional parametric surfaces graph of mean by x, y and z axes. X-axis is ammonia sample data set, y-axis is sensor array, and z- axis is sensor resistance value from the E-nose. These 3-D graphs pattern shows the different patterns of the mean respectively. From the pattern of the boxplot, normalization and mean of normalized data sets, there are significant difference in features and graph pattern of high and low level concentration. The significant sensors data were selected as an input data for training. These mean of normalized data sets were divided into two groups for training and testing. The classifier success rate for the sample data test is 97%. The data were classified by using different values of k which are 1, 2 and 3 nearest neighbor and were found that the different distance did not change the classifier rate.

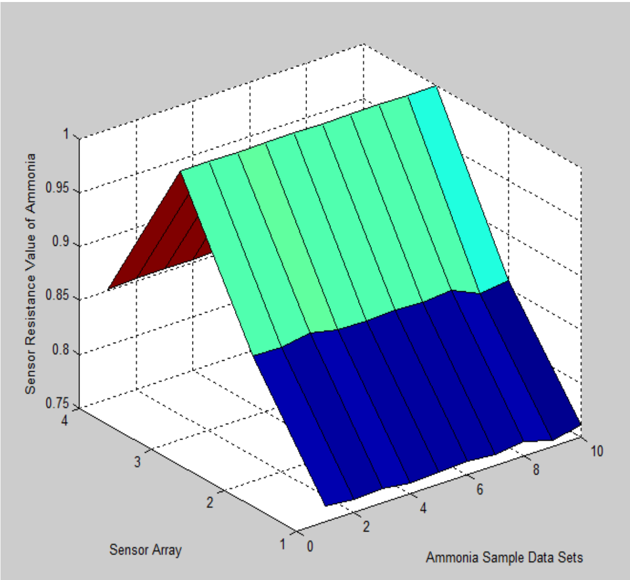

Fig.6. Mean of Normalized Data Samples from High Ammonia Concentration

Table 1 shows the classifier performance of ammonia concentration for the whole data sets. Character correctly classified samples without counting Inconclusive result. Inconclusive Rate is not classified samples. Correct Rate that count in inconclusive Rate sample is correct, which leads to the accuracy performance of ammonia up to 100%. The performance measures using specificity and sensitivity are 97% and 100% respectively.

TABLE I
CLASSIFIER PERFORMANCE OF AMMONIA CONCENTRATION.

Characteristic	**Value**
Error Rate	0
Last Correct Rate	1
Last Error Rate	0
Inconclusive Rate	0.0300
Classified Rate	0.9700
Sensitivity	1
Specificity	0.9697
Positive Predictive Value	0.2500
Negative Predictive Value	1
Positive Likelihood	1
Negative Likelihood	1

IV. CONCLUSION

This paper presents the reliability of k-NN classification technique in classifying the features and odor profiles of ammonia. The analysis and results show the differences between the high and low concentration of ammonia in term of odor profile graph pattern. These odor profile results then have been analyzed by using intelligent classification technique k-NN which has rate of success of 97%.

V. REFERENCES

[1] D. J. Randall and T. K. N. Tsui, "Ammonia toxicity in fish," *Mar. Pollut. Bull*, vol. 45, pp. 17–23, 2002.

[2] J. Sashaw, M. Nawata, S. Thompson, C. M. Wood, and P. a Wright, "Rhesus glycoprotein and urea transporter genes in rainbow trout embryos are upregulated in response to alkaline water (pH 9.7) but not elevated water ammonia," *Aquat. Toxicol*, vol. 96, no. 4, pp. 308–13, Mar. 2010.

[3] N. L. J. Wee, Y. Y. M. Tng, H. T. Cheng, S. M. L. Lee, S. F. Chew, and Y. K. Ip, "Ammonia toxicity and tolerance in the brain of the African sharptooth catfish, Clarias gariepinus," *Aquat. Toxicol.* vol. 82, no. 3, pp. 204–13, May 2007.

[4] M. M. Hegazi and S. S. Hasanein, "Effects of chronic exposure to ammonia concentrations on brain monoamines and ATPases of Nile tilapia (Oreochromis niloticus)," *Comp. Biochem. Physiol. C. Toxicol. Pharmacol*, vol. 151, no. 4, pp. 420–5, May 2010.

[5] P. Campins-Falco, S. Meseguer-Lloret, T. Climent-Santamaria, and C. Molins-Legua, "A microscale Kjeldahl nitrogen determination for environmental waters," *Talanta*, vol. 75, no. 4, pp. 1123–6, May 2008.

[6] S. M. Lloret, J. V. Andrés, C. M. Legua, and P. C. Falcó, "Determination of ammonia and primary amine compounds and Kjeldahl nitrogen in water samples with a modified Roth's fluorimetric method," *Talanta*, vol. 65, no. 4, pp. 869–75, Feb. 2005.

[7] H. Jeong, J. Park, and H. Kim, "Determination of NH 4 + in Environmental Water with Interfering Substances Using the Modified Nessler Method," *J. Chem.*, vol. 2013, pp. 1–9, 2013.

[8] A. Baker, D. Ward, S. H. Lieten, R. Periera, E. C. Simpson, and M. Slater, "Measurement of protein-like fluorescence in river and waste water using a handheld spectrophotometer," *Water Res.*, vol. 38, pp. 2934–2938, 2004.

[9] X. F. X. Fang, X. G. X. Guo, H. S. H. Shi, and Q. C. Q. Cai, "Determination of Ammonia Nitrogen in Wastewater Using Electronic Nose," *Bioinforma. Biomed. Eng. iCBBE 2010 4th Int. Conf.*, 2010.

[10] F. Tian, S. Yang, and K. Dong, "Circuit and noise analysis of odorant gas sensors in an E-nose," *Sensors*, no. December 2004, pp. 85–96, 2005.

[11] L. Zhang and F.-C. Tian, "A new kernel discriminant analysis framework for electronic nose recognition." *Anal. Chim. Acta*, vol. 816, pp. 8–17, Mar. 2014.

[12] A. D'Amico, G. Pennazza, M. Santonico, E. Martinelli, C. Roscioni, G. Galluccio, R. Paolesse, and C. Di Natale, "An investigation on electronic nose diagnosis of lung cancer.," *Lung Cancer*, vol. 68, pp. 170–176, 2010.

[13] B. Rajbansi, U. Sarkar, and S. E. Hobbs, "Hazardous odor markers from sewage wastewater: A step towards simultaneous assessment, dearomatization and removal," *J. Taiwan Inst. Chem. Eng.*, vol. 45, no. 4, pp. 1549–1557, Jul. 2014.

[14] M. L. Rodriguez-Mendez, C. Apetrei, M. Gay, C. Medina-Plaza, J. A. De Saja, S. Vidal, O. Aagaard, M. Ugliano, J. Wirth, and V. Cheynier, "Evaluation of oxygen exposure levels and polyphenolic content of red

wines using an electronic panel formed by an electronic nose and an electronic tongue," *Food Chem.*, vol. 155, pp. 91–97, 2014.

[15] X. Hong and J. Wang, "Detection of adulteration in cherry tomato juices based on electronic nose and tongue: Comparison of different data fusion approaches," *J. Food Eng.*, vol. 126, pp. 89–97, 2014.

[16] P. E. Keller, "Electronic noses and their applications," *IEEE Tech. Appl. Conf. Work. Northcon95 Conf. Rec.*, 1995.

[17] Z. T. Z. Tao, W. L. W. Lei, and J. T. J. Teng, "Pattern Recognition of the Universal Electronic Nose," *2008 Second Int. Symp. Intell. Inf. Technol. Appl.*, vol. 3, pp. 249–253, 2008.

[18] B. Tudu, A. Metla, B. Das, N. Bhattacharyya, A. Jana, D. Ghosh, and R. Bandyopadhyay, "Towards versatile electronic nose pattern classifier for black tea quality evaluation: An incremental fuzzy approach," *IEEE Trans. Instrum. Meas.*, vol. 58, pp. 3069–3078, 2009.

[19] K.-T. Tang, S.-W. Chiu, C.-H. Pan, H.-Y. Hsieh, Y.-S. Liang, and S.-C. Liu, "Development of a Portable Electronic Nose System for the Detection and Classification of Fruity Odors," *Sensors (Peterboroug*, vol. 10, pp. 9179–9193, 2010.

[20] M. Zhang, X. Wang, Y. Liu, X. Xu, and G. Zhou, "Species Discrimination among Three Kinds of Puffer Fish Using an Electronic Nose Combined with Olfactory Sensory Evaluation," *Sensors*, vol. 12. pp. 12562–12571, 2012.

[21] M. S. Najib, N. A. M. Ali, M. N. M. Arip, A. M. Jalil, and M. N. Taib, "Classification of Agarwood region using ANN," *2010 IEEE Control Syst. Grad. Res. Colloq. (ICSGRC 2010)*, pp. 7–13, Jun. 2010.

[22] S. Guney and A. Atasoy, "Classification of n-butanol concentrations with k-NN algorithm and ANN in electronic nose," *Innov. Intell. Syst.*, pp. 138–142, 2011.

[23] A. M. Andrew, K. Kamarudin, S. M. Mamduh, A. Y. M, A. Zakaria, A. H. Adom, and D. L. Ndzi, "Classification of Domestic Burning Smell using Covariance k- Nearest Neighbour Algorithm for Early Fire Detection Application," vol. 40, pp. 271–276, 2014.

[24] N. Bhatia, "Survey of nearest neighbor techniques," *arXiv Prepr. arXiv1007.0085*, vol. 8, no. 2, pp. 302–305, 2010.

[25] I. Saini, D. Singh, and A. Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases," *J. Adv. Res.*, vol. 4, no. 4, pp. 331–344, Jul. 2013.

[26] M. Mustafa, M. Taib, Z. Murat, and N. Sulaiman, "Comparison between KNN and ANN classification in brain balancing application via spectrogram image," *J. Comput. Sci.*, vol. 2, no. 4, pp. 17–22, 2012.

[27] D.O.E (2009), "Department of Environment | Ministry of Natural Resources & Environment." Available: http: / / www. doe. gov. my / portal1/en/.

[28] B. Tudu and B. Kow, "Normalization techniques for gas sensor array as applied to classification for black tea," *J. Smart.*, vol. 2, no. 1, pp. 176–189, 2009.

[29] Q. Chen, J. Zhao, Z. Chen, H. Lin, and D.-A. Zhao, "Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools," *Sensors Actuators B Chem.*, vol. 159, no. 1, pp. 294–300, Nov. 2011.

[30] R. Entezari-Maleki, "Comparison of Classification Methods Based on the Type of Attributes and Sample Size," *J. Converg.*, 2009.