# INTELLIGENT ODOR-PROFILE CLASSIFICATION OF KELULUT HONEY USING CASE-BASED REASONING TECHNIQUE (CBR)

Muhammad Sharfi Najib, Shahrul Hisyam Azih, Nurdiyana Zahed, Muhammad Faruqi Zahari, Wan Muhamad Azmi Mamat, Hadi Manap
Faculty of Electrical and Electronics, Universiti Malaysia Pahang, Pekan, Pahang, Malaysia
E-Mail: sharfi@ump.edu.my

## ABSTRACT

Nowadays, Kelulut is one of a bee species, yet different in size, it produces a honey that is clearer in colour as compared to natural honey bee. However, study on Kelulut honey grade is not yet extensively explored. Hence, this research aim is to classify Kelulut honey based on smell pattern recognition. In order to sense the smell of Kelulut honey, an E-nose that is comprises of sensor array has been used to measure 2000 samples of dataset from two different types of Kelulut honey. The measured data have been normalized and analysed using statistical method. Then, it has been classified using CBR as an intelligent classifier. It is shown that 100% classification rate of accuracy has been achieved for 2 types of Kelulut honey. The performance measures of CBR in terms of accuracy, specificity, sensitivity have been successfully achieved with 100% performance.

**Keywords**: e-nose, odor-profile, kelulut honey, normalization.

## INTRODUCTION

Honey is one of the famous traditional medicines, which are considered important for treating respiratory disease, an infection in the digestive tract and other diseases. Honey can also be used routinely to wrap wounds; burns to reduce pain quickly [1]. Honey is divided into two types which is honey bees and honey Kelulut or stingless bee. In Malaysia, Kelulut honey is becoming an alternative treatment in Malaysia [2]. Several techniques have been found to detect honey such as HS-SPME and GC-MS [3] as a chemical compound marker. However, these techniques are found to be complex for onsite detection. One of the techniques that can detect the properties based on odor-profile.

Detection of an odor-profile by olfaction is the major neurosensory function in human for investigating external chemical environment [4]. Electronic-nose is one a device that has received much attention in the field of sensor technology over the last twenty years, largely due to the discovery of many Applications derived from research in various fields of applied science [5]. E-nose is an instrument that has been developed in order to mimic human's nose. The main sensing components of E-nose is an array of physical sensors [6]. A primary function of E-nose is a systems is detecting odor released in form of gasses or vapor [7] and suitable for honey which produces several volatile compounds [8].

However, E-nose itself is meaningless without employing any intelligent classification system. Several intelligent classification systems that have been reported are k-NN [9], ANN [10] and also case-based reasoning CBR) [11] . Among the listed intelligent system, CBR is found to be one of the intelligent system that require no training [10] and suitable for a weak domain field [12]. This specification of CBR gives the classification process faster as compared to other techniques [13].

Presently, the grading of quality of Kelulut are still being classified based on expertise of human sense and some chemical techniques which require long process. Many sellers who promote Kelulut honey are unable to measure its authenticity. This is a challenge for the entrepreneur, buyers and manufacturers of Kelulut honey and it causes Kelulut market could be affected. Kelulut real worth RM80 to RM100 for a kilogram, thus its range of prices has a possibility can be manipulated.

Hence, this research is a necessity study to classify quality of Kelulut honey. Therefore, this paper proposes a new technique, employing E-nose and CBR, for classifying different types Kelulut from its odor-profile.

## METHODOLOGY

The measurement of Kelulut honey sample is based on two types Kelulut honey samples from Malaysia which are Itama (SK1) and Thoracica (SK2).

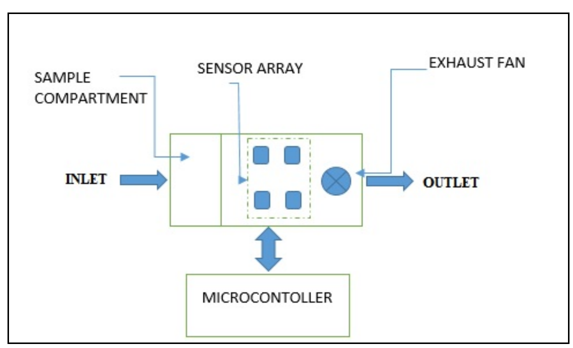


**Figure-1.** A sample of sample data from SK1 before normalization.

Figure-1 shows the E-nose used to measure the odor-profile. The data extracted from Arduino were stored in Microsoft Excel and analyzed using MATLAB. A MATLAB Simulink was explored to verify the graphical representation of the data and for normalization.

VOL. 11, NO. 10, MAY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Initially, the 10ml of each fresh samples (SK1 and SK2) were filled in a dedicated sample vials. The vial that was filled with fresh sample was preheated in special heater in the constant temperature in the E-nose. In order to get a fine measurement, the samples under heating process were ensured not to be exposed to the external contaminants. Each measurement has been taken in duration of five minutes and another five minutes; the E-nose was left empty. This procedure is designed to ensure that the E-nose has an ample time to have a better sensitivity. Ultimately, 200 measurements were taken from each dataset; hence generating 400 E-nose readings with each reading consisted of an array of 4 sensor values (200 sampled data x 4 sensors (S1, S2, S3 and s4) x 5 measurements x 2 samples) which is equivalent to 8000 overall data. After completing data raw data preparation, pre-processing technique was applied before beginning with CBR computation. All the sample datasets were normalized per measurement. After the normalization of the pattern of Kelulut honey odor-profile, it was then divided into 10 datasets of with average value from each S1, S2, S3 and S3 was taken and shown in Table 1. Each of an average data was set as a case in CBR library. CaseID 1 until CaseID 5 and CaseID 6 until CaseID 10 were stored as SK1 and SK2 cases respectively.

**Table-1.** A sample of case library design from SK1.

| Problem | | | | | Solution |
|---|---|---|---|---|---|
| CaseID | A1 | A2 | A3 | A4 | Type |
| caseid _001 | DC11 | DC12 | DC13 | DC14 | SK1 |
| caseid _002 | DC21 | DC22 | DC23 | DC24 | SK1 |
| caseid _003 | DC31 | DC32 | DC33 | DC34 | SK1 |
| caseid _004 | DC41 | DC42 | DC43 | DC44 | SK1 |
| caseid _005 | DC51 | DC52 | DC53 | DC54 | SK1 |
| caseid _006 | DC61 | DC62 | DC63 | DC64 | SK2 |
| caseid _007 | DC71 | DC72 | DC73 | DC74 | SK2 |
| caseid _008 | DC81 | DC82 | DC83 | DC84 | SK2 |
| caseid _009 | DC91 | DC92 | DC93 | DC94 | SK2 |
| caseid _010 | DC101 | DC102 | DC103 | DC104 | SK2 |

They were then analyzed across the groups using box plot. In this process, one row of data involves four sensors from SK1 database was set as current case, while the rest of the nine sets data (4 cases from SK1 and 5 cases from SK2) were set as stored case. The significant of this process is repeated to determine the similarity results from current cased and stored data. The similarity calculation results were then inserted into voting table to find compute the performance measures of CBR using accuracy, specificity and sensitivity of this instrumentation method.

## RESULT AND DISCUSSION

### Raw data measurement results

Table-2 and Table-3 shows the summarized collection of data from SK1 between measured data 1 to measured data 200 for a sample of dataset.

**Table-2.** A sample of measured raw data for A dataset SK1.

| Measured data (DM) | S1(Ω) | S2(Ω) | S3(Ω) | S4(Ω) |
|---|---|---|---|---|
| $DM_1$ | 285 | 189 | 167 | 113 |
| $DM_2$ | 285 | 190 | 168 | 113 |
| . | . | . | . | . |
| . | . | . | . | . |
| $DM_{200}$ | 285 | 189 | 168 | 113 |

**Table-3.** A sample of measured raw data for A dataset SK2.

| Measured data (DM) | S1(Ω) | S2(Ω) | S3(Ω) | S4(Ω) |
|---|---|---|---|---|
| $DM_1$ | 329 | 241 | 188 | 137 |
| $DM_2$ | 329 | 241 | 188 | 137 |
| . | . | . | . | . |
| . | . | . | . | . |
| $DM_{200}$ | 328 | 241 | 189 | 136 |

Graphical representation of both CaseID 1 (Figure-2) and CaseID 6 (Figure-3) data from SK1 and SK2 respectively, is visualized into MATLAB. These two graphical representations show pattern recognition from different group of Kelulut odor-profile sample. From this 3D visualization, the proof of significant different is difficult to be classified without normalization technique.

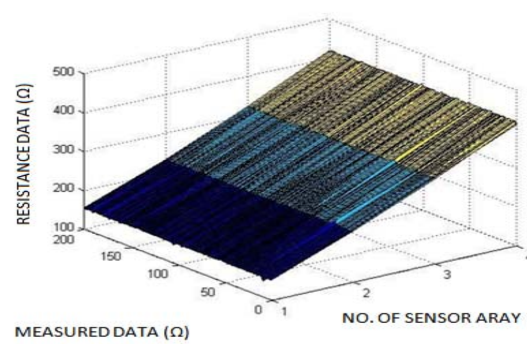


**Figure-2.** A sample of measured data from SK1 before normalization

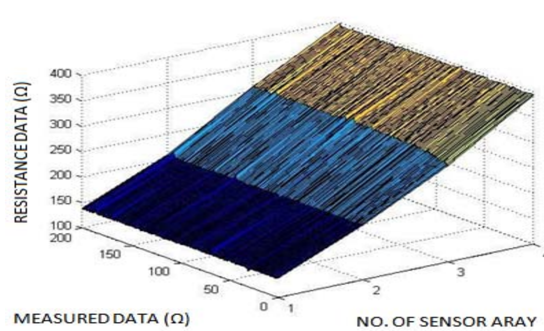


**Figure-3.** A sample of measured data form SK2 before normalization.

ARPN Journal of Engineering and Applied Sciences



From experimental procedure, initial values of potentiometer for sensor S1, S2, S3 and S4 were tuned to 50 ohm, 150 ohm, 250 ohm and 400 ohm respectively. The tuning values procedure was for E-nose measurement without samples for sensor normalization and calibration procedure. After the samples of Kelulut were inserted into E-nose chamber, it can be shown that the response from SK1 and SK2 are 7.51% and 6.32% towards contaminant sensor (S1) respectively. This low sensitivity of air contaminant sensor showed that the air in chamber is almost 100% clean without impurities. For combustible gas sensor (S2), it is clearly can be seen that sensor react most well with both SK1 and SK2 sample. Odor-profile form SK1 and SK2 was recorded 70.55% and 45.5% respectively. It can be said that Kelulut odor-profile has a great presence of combustible-like odor in its volatile compound. Meanwhile, hydrogen gas sensor (S3) has shown a quite similar result of sensitivity with air contaminant sensor which was 8.52% and 8.43% for SK1 and SK2 respectively where both sample SK1 and SK2 are less responsive to this sensor. For carbon monoxide gas, tuning resistance was set to 400 ohm, but both samples SK1 and SK2 has shown a great decline towards its value, where SK1 and SK2 was recorded 41.14% and 40.04% sensitivity respectively. It can be dictated that Kelulut odor-profile does not responsive to carbon monoxide-like odor in its gas.

**Normalized raw data measurement results**

Table-4 and Table-5 shows the collection of the summarized measured data from SK1 and SK2 respectively between measured data 1 to 200. The measured data is clearly filled in the first (1$^{st}$) row until the second last row (200$^{th}$). This measured data are the normalized data using (0-1) normalization mode. The data in the last row of Table-4 and Table 5 are the calculated average data from each sensor S1, S2, S3 and S4. This row data has been set as input to the store case in CBR case library as shown in Table-6 which consists of 10 cases.

**Table-4.** A sample of normalized measured data for A dataset SK1.

| Normalized Measured data (NDM) | S1(Ω) | S2(Ω) | S3(Ω) | S4(Ω) |
|---|---|---|---|---|
| $NDM_1$ | 1.000 | 0.663 | 0.586 | 0.396 |
| $NDM_2$ | 1.000 | 0.667 | 0.589 | 0.396 |
| . | . | . | . | . |
| . | . | . | . | . |
| $NDM_{200}$ | 1.000 | 0.663 | 0.589 | 0.396 |
| $ANDM_{200}$ | 1.000 | 0.665 | 0.588 | 0.395 |

**Table-5.** A sample of normalized measured data for A dataset SK2.

| Normalized Measured data (NDM) | S1(Ω) | S2(Ω) | S3(Ω) | S4(Ω) |
|---|---|---|---|---|
| $NDM_1$ | 1.000 | 0.733 | 0.571 | 0.416 |
| $NDM_2$ | 1.000 | 0.733 | 0.571 | 0.416 |
| . | . | . | . | . |
| . | . | . | . | . |
| $NDM_{200}$ | 1.000 | 0.728 | 0.578 | 0.416 |
| **$ANDM_{200}$** | **1.000** | **0.729** | **0.574** | **0.416** |

Figure-4 shows the comparison of pattern recognition between normalized measured data of odor-profile from SK1 and SK2. The X-axis is the normalized measured data while Y-axis is the number of sensor array. Two measured data are represented in Figure-3; which are SK1 (blue line) and SK2 (green line). The general representation across the type of Kelulut does not show any significant differences and variance. However, even the variance is not significant from direct visualization; it can be shown in the next statistical representation in boxplot.

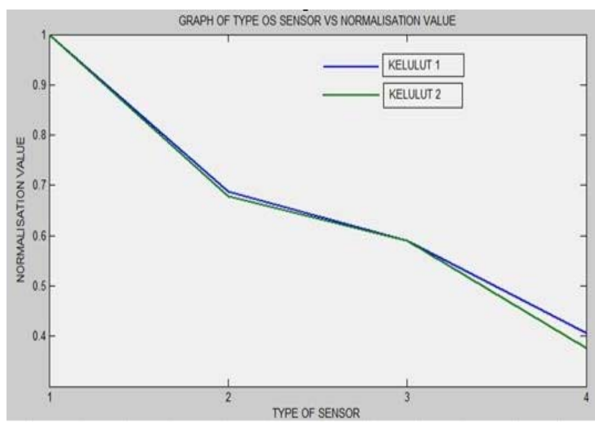


**Figure-4.** A pattern recognition comparison of SK1 and SK2.

**Boxplot for SK1 and SK2 samples**

This section shows the normalized measured data which was analyzed using boxplot as shown in Figure-4 and Figure-4. Each of the boxplot depicts additional pattern of odor-profile of sample SK1 and SK2 respectively. X-axis shows the number of associated sensor and Y-axis portrays the normalizing data of the resistance response. Generally, the pattern recognition from Figure-5 (SK1) and Figure-6 (SK2) has a very less differences. However, if the both figures are analyzed statistically in terms of variance, they have shown slightly different at sensor S2.

ARPN Journal of Engineering and Applied Sciences



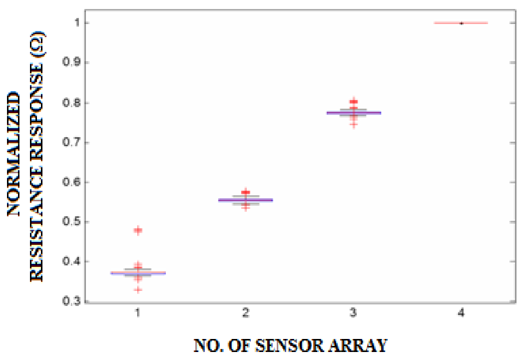


**Figure-5.** A Boxplot of SK1.

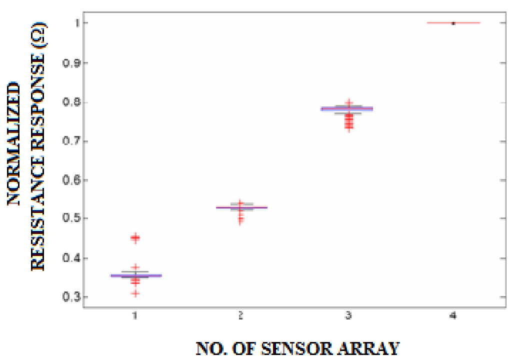


**Figure-6.** A Boxplot of SK2.

**CBR computation and voting**

Table-6 shows the average normalised measured data from SK1 and SK2 samples while Table-7 shows the CBR case library for SK1 and SK2 samples. The average normalized measured data from S1, S2, S3 and S4 were set as attributes (A1, A2, A3 and A4) shown in Table-5. Hence, there are 10 cases are tabulated in Table-5.

**Table-6.** Average normalized measured data.

| Measurement No. | S1(Ω) | S2(Ω) | S3(Ω) | S4(Ω) |
|---|---|---|---|---|
| $DM_1$ | 1.000 | 0.665 | 0.588 | 0.395 |
| $DM_2$ | 1.000 | 0.814 | 0.508 | 0.351 |
| $DM_3$ | 1.000 | 0.679 | 0.580 | 0.390 |
| $DM_4$ | 1.000 | 0.808 | 0.509 | 0.351 |
| $DM_5$ | 1.000 | 0.695 | 0.571 | 0.385 |
| $DM_6$ | 1.000 | 0.729 | 0.574 | 0.416 |
| $DM_7$ | 1.000 | 0.743 | 0.613 | 0.483 |
| $DM_8$ | 1.000 | 0.730 | 0.578 | 0.423 |
| $DM_9$ | 1.000 | 0.743 | 0.611 | 0.479 |
| $DM_{10}$ | 1.000 | 0.731 | 0.583 | 0.430 |

**Table-7.** CBR case library for SK1 and SK2 sample.

| Problem | | | | | Proposed Solution |
|---|---|---|---|---|---|
| CaseID | A1 | A2 | A3 | A4 | Type |
| caseid _001 | 1.000 | 0.665 | 0.588 | 0.395 | SK1 |
| caseid _002 | 1.000 | 0.814 | 0.508 | 0.351 | SK1 |
| caseid _003 | 1.000 | 0.679 | 0.580 | 0.390 | SK1 |
| caseid _004 | 1.000 | 0.808 | 0.509 | 0.351 | SK1 |
| caseid _005 | 1.000 | 0.695 | 0.571 | 0.385 | SK1 |
| caseid _006 | 1.000 | 0.729 | 0.574 | 0.416 | SK2 |
| caseid _007 | 1.000 | 0.743 | 0.613 | 0.483 | SK2 |
| caseid _008 | 1.000 | 0.730 | 0.578 | 0.423 | SK2 |
| caseid _009 | 1.000 | 0.743 | 0.611 | 0.479 | SK2 |
| caseid _010 | 1.000 | 0.731 | 0.583 | 0.430 | SK2 |

Table-8 shows the results of cases in the CBR library that is converted into voting table. There are 10 cases listed in the figure; caseid_001 until caseid_010. CASE 1 until caseid_005 are belong to odor-profile sample of SK1 while caseid_006 until caseid_010 are belong to odor-profile of odor-profile sample extracted from SK2. From Figure-5, it can be shown that similarity distance value can be observed via horizontal or vertical. The boxes that have been colored by black indicates 100% of similarity as stored case data installed is similar to current case data. Then, similarity distance value of voting ranking into highest (k=1), medium (k=2) and lowest (k=3) in order to observe wether the top three highest results came from same sample or vice versa.

From classification accuracy result of voting tables at Column 1-5, it indicates result for SK1, whereas column 6-10 indicates results for SK2 for all level of voting ranking (k=1, k=2, k=3).

**Table-8.** CBR case library.

| Voting | | | | Classification Accuracy (%) |
|---|---|---|---|---|
| CaseID | K=1 | K=2 | K=3 | Type |
| caseid _001 | SK1 | SK1 | SK1 | 100 |
| caseid _002 | SK1 | SK1 | SK1 | 100 |
| caseid _003 | SK1 | SK1 | SK1 | 100 |
| caseid _004 | SK1 | SK1 | SK1 | 100 |
| caseid _005 | SK1 | SK1 | SK1 | 100 |
| caseid _006 | SK2 | SK2 | SK2 | 100 |
| caseid _007 | SK2 | SK2 | SK2 | 100 |
| caseid _008 | SK2 | SK2 | SK2 | 100 |
| caseid _009 | SK2 | SK2 | SK2 | 100 |
| caseid _010 | SK2 | SK2 | SK2 | 100 |

After similarity function, weight similarity, voting case percentages are calculated, the calculation of the classification (CBR retrieval stage) performance data is applied as shown in Table 6, to calculate accuracy, sensitivity and specificity of data. Accuracy, specificity and sensitivity percentage are used as statistical approaches of the performance of Kelulut odor-profile in SK1 and SK2 samples. The accuracy is the sum of true positive and false negative of SK1 and SK2 were divided by the sum of number of the samples. The second performance measure is a specificity which is the number of true negative results divided by the sum of false positive and true negative results. The last one is sensitivity is the number of true positive results divided by the sum of true positive false negative result.

From Table-9 shows the summary of the classification performance measures of Kelulut honey odor-profile classification using CBR have achieved 100% in accuracy, specificity and sensitivity.

**Table-9.** Performance measure of dataset SK1.

| Characteristic | Value |
|---|---|
| Accuracy Percentage | 100% |
| Specificity | 100% |
| Sensitivity | 100% |

## CONCLUSIONS

In conclusion, this works has shown that classification of Kelulut odor-profile is feasible and successful using E-nose as a tool of volatile compound sensor dedicated for Kelulut stingless honey in collecting sample of measurements. The data was successfully normalized and the feature of the two sample of Kelulut has been extracted and become input feature to CBR. The CBR classification has classified two different Kelulut sample with 100% accuracy, sensibility and specificity. Shortly, the classification using CBR as an intelligent classifiers give a new field of artificial intelligence for a weak domain field has been developed. It is suggested that the sample data dimensional size is increased in avoiding over fitting and to secure robustness of the CBR classifier.

## REFERENCES

[1] A. Jull, N. Walker, and S. Deshpande. 2014. Honey as a topical treatment for wounds (Review).Cochrane Libr. pp. 1–97.

[2] M. I. Zainol, K. Mohd Yusoff, and M. Y. Mohd Yusof. 2013. Antibacterial activity of selected Malaysian honey.BMC Complement. Altern. Med. pp. 129.

[3] L. F. Cuevas-Glory, J. A. Pino, L. S. Santiago, and E. Sauri-Duch. 2007. A review of volatile analytical methods for determining the botanical origin of honey. Food Chemistry. pp. 1032–1043,

[4] T. Speed. 2014. Chronicle / Lab Book Untangling a Viral Infection A host cell enzyme Keeping Tabs on Development New software.pp. 38–40.

[5] A. D. Wilson and M. Baietto. 2011. Advances in electronic-nose technologies developed for biomedical applications. Sensors. pp. 1105–1176.

[6] M. S. Najib, N. A. M. Ali, M. N. M. Arip, A. M. Jalil, and M. N. Taib. 2010. Classification of Agarwood region using ANN. In Proceedings - ICSGRC 2010: IEEE Control and System Graduate Research Colloquium. pp. 7–13.

[7] M. S. Najib, N. A. M. Ali, M. N. M. Arip, M. A. Jalil, and M. N. Taib. 2012. Classification of Agarwood using ANN. Int. J. Electr. Electron. Syst.

[8] J. A. Pino, R. Marbot, A. Delgado, C. Zumarraga, and E. Sauri. 2006. Volatile constituents of propels from honey bees and stingless bees from Yucatan. J. Essent. Oil Res.pp. 53–56.

[9] S. Güney and A. Atasoy. 2011. Classification of n-butanol concentrations with k-NN algorithm and ANN in electronic nose. INISTA 2011 - 2011 International Symposium on Innovations in Intelligent SysTems and Applications. pp. 138–142.

[10] M. S. Najib, M. U. Ahmad, P. Funk, M. N. Taib, and N. A. M. Ali. 2012. Agarwood classification: A case-based reasoning approach based on E-nose. IEEE 8th International Colloquium on Signal Processing and Its Applications, CSPA. pp. 120–126.

[11] M. F. Zahari, J. Thommy Anak, F. A. Halim, M. S. Najib, K. H. Ghazali, and A. A. M. Azoddein. 2015. Intelligent Classification Hazardous Gas Using Sensor Array. J. Adv. Inf. Technology. pp. 233–237.

[12] J. L. Kolodner. 1992. An introduction to case-based reasoning. Artif. Intell. pp. 3–34.

[13] F. A. Halim, M. S. Najib, K. H. Ghazali, and M. F. Zahari. 2015. ISERD - 2nd International Conference on Modern Trends in Engineering and Technology (ICMTET).