MEKATRONIKA
e-ISSN: 2637-0883
VOL. 2, ISSUE 2, 28 – 35
DOI: https://doi.org/10.15282/mekatronika.v2i2.6747

ORIGINAL ARTICLE

Intelligent Classification of Cocoa Bean using E-nose

Nur Amanda Nazli[1], Muhammad Sharfi Najib[1,*], Suhaimi Mohd Daud[2], Mujahid Mohammad[2], Zainal Baharum[3] and Mohamed Yusof Ishak[4]

[1]Faculty of Manufacturing & Mechatronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[2]Faculty of Electrical and Electronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[3]Division of Biotechnology, Cocoa Innovation &Technology Centre, Malaysian Cocoa Board, Negeri Sembilan, Malaysia.
[4]Division of Regulatory and Quality Control, Cocoa Innovation &Technology Centre, Malaysian Cocoa Board, Negeri Sembilan, Malaysia.

ABSTRACT – Cocoa bean (Theobrama cacao) is an essential raw material in the manufacture of chocolate, and their classification is crucial for the synthesis of good chocolate flavour. Cocoa beans appear to be very similar to one another when visualised. Hence, an electronic device named the electronic nose (E-Nose) is used to classify the odor of cocoa beans to give the best cocoa bean quality. E-nose is a set of an array of chemical sensors used to sense the gas vapours produced by the cocoa bean and the raw data collected was kept in Microsoft Excel, and the classification took place in Octave. They then underwent normalisation technique to increase classification accuracy, and their features were extracted using mean calculation. The features were classified using CBR, and the similarity value is obtained. The results show that CBR's classification accuracy, specificity and sensitivity are all 100%.

ARTICLE HISTORY
Received: 3rd Nov 2020
Revised: 29th Nov 2020
Accepted: 13th Dec 2020

KEYWORDS
Cocoa bean
Chocolate
E-nose
CBR

INTRODUCTION

The olfactory system is the sensory system used for smelling, which smelling is a vital sensation that allows species with odorant receptors to recognise food, threats and enemy [1]. The smell is one of the most significant ways for most living beings and humans to communicate with the world, as a smell can affect our sense of taste. The cocoa bean has been the world's most significant demand for its product which flavour and odor are the most significant characteristics and critical qualities of cocoa quality [2]. Odor is a mixture of tiny molecules with a very low concentration in inhaled air when it comes in contact with the human sensory system, which experiences the perception of odor [3]. Not all raw cocoa bean gives the smell of chocolate, and human sniffers cannot simply determine the odor of cocoa. Regardless of the significance of olfactory sense to humankind, the human nose only has 1,000,000 fragrance receptors compared to canines that have more than 100 million receptors that separate aromas, at any rate, multiple times [4]. The moderately poor affectability and separation capacities of the human sense have prompted the need for electronic instruments with sensors equipped to conduct rehashed segregations with a high probability of removing human fatigue [5].

Electronic nose (E-nose) is the most effective instrument for identifying good odor and bad odor to replicate a human's olfactory system [6]. E-nose is a set array of sensors that chemically interact with volatile organic compounds (VOCs) [7] that send signals to classify the odor. The sensors are unfixed to the gas molecules, and each sensor has a different sensing medium. The gas mixtures can be acknowledged from the characteristics of the response of the sensor. E-nose has been developed for qualitative classification and recognition of various kinds of environments and various foods [8]. It is easily operated, lower cost and time consuming method for odor classification of cocoa bean. It also detects the smell more accurately than the human sense of smell. It also determines the toxic or poisonous gas that human sniffers are impossible to sniff [9].

Artificial intelligence (AI) is applied in this research to classify the odor profile of cocoa beans to support human decision-making [10] and give the solution that can be mapped to the original case. AI aims to explain intelligence by creating computer programs that use symbolic inference or logic to provide intelligent behaviour [11]. It is not time-independent and requires every device to determine by having time in mind. Case-based reasoning or CBR is acknowledged that it offers a stored case to explain its reasoning by solving the problem by using the previous case that looks similar to the current case [12]. There are four stages of retrieving, reusing, revising, and retaining to classify the odor in the CBR system [13]. This system chooses a subset of cases from the database in retrieve point, closely related to the current case. For the next step, reuse fits the current problem from the cases chosen from the previous stage. The Revise stage is where the solution obtained is tested and checked that it has the right solution. For the final stage, retain will decide whether or not to integrate the newly solved case into the CBR database [14]. This intelligent classification can classify the odorant with a high accuracy rate compared to other intelligent classifications [15].

This paper shows the E-nose and CBR classifier to demonstrate an effective approach for cocoa bean classification. E-nose gathered data to establish the odor profile of cocoa beans, which was then classified using case-based reasoning.

*CORRESPONDING AUTHOR | Muhammad Sharfi Najib | ✉ sharfi@ump.edu.my

METHODOLOGY

Experiment Design

Figure 1 shows the flowchart on the process to get the classification of the cocoa bean. The data on the odor of raw cocoa bean was collected using e-nose. Then, the collected raw data will undergo pre-processing stage where the features will be derived using the normalisation method and arithmetic mean calculation to determine cocoa bean's odor profile. The data were then classified using a classifier of CBR. Lastly, the performance is measured to get the classification's sensitivity, specificity and accuracy percentages.

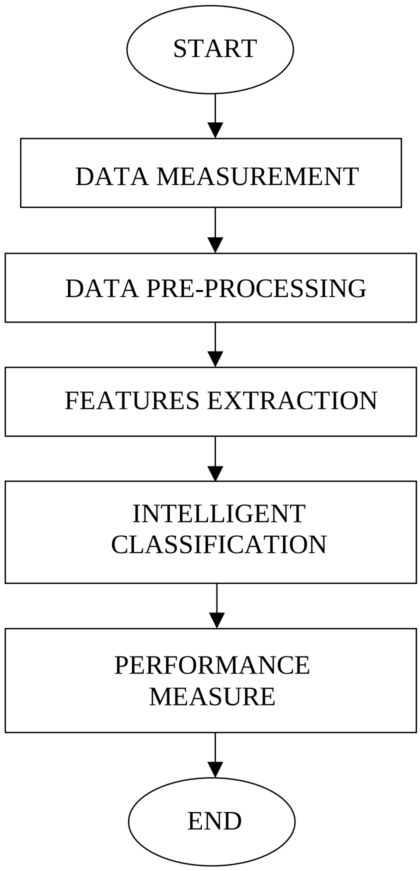

Figure 1. Flowchart for cocoa bean classification

Data Measurement

Figure 2 depicts the flow of getting raw data of cocoa beans using E-nose. There are five types of the smell of cocoa bean: cocoa smell, musty smell, black pepper smell, smoke smell, and copra smell. The samples are weighted 30g each and placed into a dish plate. E-nose operated like an odor inhaler that inhales the odor extracted from a cocoa bean sample. Data collected was then trapped in the E-nose chamber, ready to transmit to the computer via a universal serial bus (USB) cable for data collection.

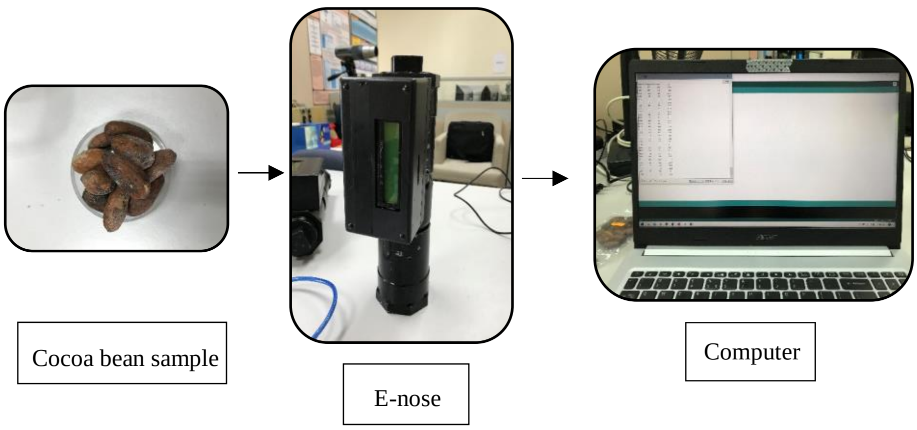

Figure 2. Flow of Collecting Odor of Cocoa Bean Sample

After 5 minutes in one experiment for every sample, the data measurement collected was 200. Then the experiment was repeated five times which gives the total result of 1000 data for every sample, and the data was tabulated on Table 1 in Microsoft Excel.

Table 1 Table of raw data of cocoa bean

Data measurement (DM)	Sensor Reading			
	Sensor 1	Sensor 2	Sensor 3	Sensor 4
DM_1	DM_{11}	DM_{12}	DM_{13}	DM_{14}
DM_2	DM_{21}	DM_{22}	DM_{23}	DM_{24}
.	.	.	.	.
.	.	.	.	.
DM_{1000}	DM_{1001}	DM_{1002}	DM_{1003}	DM_{1004}

Data Pre-Processing

The DM of cocoa beans is normalised using the normalisation method in the pre-processing stage to eliminate any data anomalies that may influence the results using Equation 1. The maximum value separates each row in raw data calculation from its row to obtain a normalised value. This approach reduces data fluctuation and eliminates classification errors.

$$R' = \frac{R}{R_{max}} \tag{1}$$

From Equation 1, R' is the normalisation value, R is the sensor's reading (ohm (Ω)), and R_{max} is the highest value of sensor reading. Hence, the normalisation method was used in which the data measurement was minimised from 1000 normalised data (5 experiments x 200 data samples) into 200 normalised data for every sample.

Features Extraction

The amount of data was minimised in a dataset by generating new features from the original feature set. The mean normalised data were clustered from 200 mean normalised data times 5 for each type of cocoa beans sample to 10 cases. So in 1 case, there were 20 normalised data. After that, mean calculation is used for all experiments for every case, the mean normalised data for every cases were reduced from 20 dataset to 1 dataset. As a result, the data collected is ten times with four data measurements per sample.

Intelligent Classification

CBR is then used to classify the derived features. This technique is a way of solving problems by referring to past cases and experiences that are applicable to the present situation. The CBR cycle of the cocoa bean has four phases, which are depicted in Figure 3.

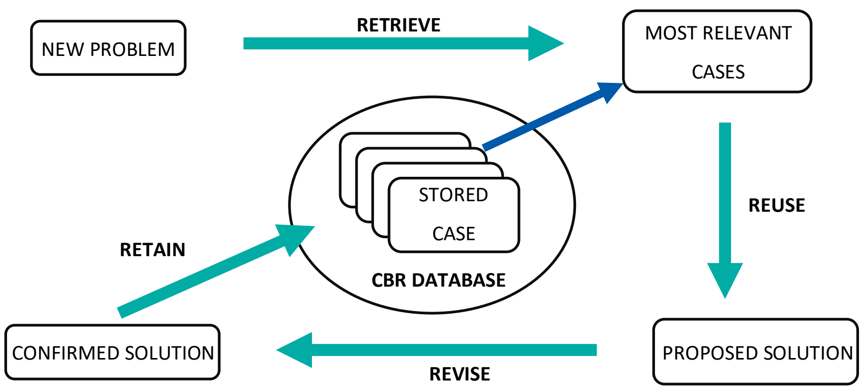

Figure 3. Case-Based Reasoning cycle process

In Retrieve point, this system chooses a subset of cases from the database closely related to the current case. For the next step, reuse fits the solution to the current problem from the cases chosen from the previous stage. The Revise stage is where the solution obtained is tested and checked that it has the right solution. Retain will decide whether or not to integrate the newly solved case into the CBR database for the final stage. The similarity value must be calculated before

CBR computation can be performed. In each scenario, the calculation was repeated. The formula represents in Equation (2)

$$Similarity\ (T, S) = \frac{\sum_{i=1}^{n} f(T_i, S_i) \times w_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

From Equation (2), *T* and *S* denote as target and source cases, respectively. n is denoted as the number of attributes in each case, *i* denotes individual or single attributes from 1 to *n*, *f* denotes the similarity function for attribute *i* in cases *T* and *S*, and *w* denotes the value of attribute *i* for weighing.

Performance Measure

The final step in cocoa bean classification is to determine the performance measure, which was accomplished using the uncertainty matrix process. The Equations 3 to 5 were used to evaluate the sensitivity, precision, and accuracy of the cocoa bean samples classification.

$$sn = \frac{TP}{TP + FN} \tag{3}$$

$$sp = \frac{TN}{TN + FP} \tag{4}$$

$$acc = \frac{TP + TN}{P + N} \tag{5}$$

The notation of sn, sp and acc are the sensitivity, specificity and accuracy sequentially. TP is the true positive, which is number that correctly recognised class results, FP is the false positive which the results that eithere were correctly assigned to the class, TN is the number of correctly organised that do not belong to the class known as true negative and FP is false positive that were not recognised results. P and N are the positive and negative same as true and false results related corresponding. The confusion matrix is a table that describes the performance of a classification model. It contains information about the actual and predicted case. CBR voting table results were used to determine the number of true positives, false positives, true negatives, and false negatives.

RESULTS AND DISCUSSION

Figure 4 shows the mean normalised value against sensor attribution of cocoa smell, musty smell, black pepper smell, smoke smell and copra smell of cocoa bean samples. The X-axis indicates sensor attribution, which is the amount of sensors installed in e-nose to recognise the volatile molecules of the cocoa bean, and Y-axis indicates the mean noramlized value of the sensor resistance after using normalisation technique mean calculation.

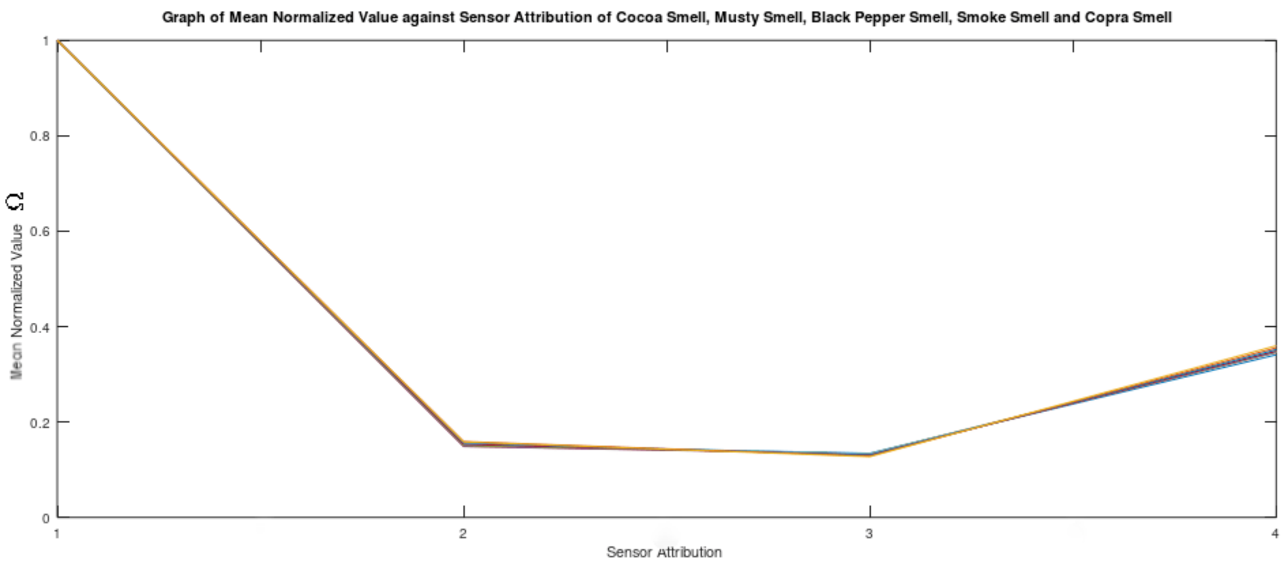

Figure 4. Graph of Mean Normalized Value against Sensor Attribution of Cocoa Smell, Musty Smell, Black Pepper Smell, Musty Smell and Copra Smell

In Figure 4, all samples were plotted in one graph to see the differences between each sample. Though the graphs plotted look similar, the characteristic of the smell is still different to each other. All the clustered normalised data were then tabulated in Table 2 on Microsoft Excel. The clustered data was assigned according to its sample and case. The table made up of 50 cases, with 10 cases for every sample. The first ten cases (from case 1 to case 10) represent a cocoa scent examination. The following ten cases (cases 11 through 20), (cases 21 through 30), (cases 31 through 40), and (cases 41 through 50) reflect musty, black pepper, smoke, and copra smell of cocoa bean samples, respectively. Sensors (S1, S2, S3, and S4) are located in the e-nose and are labelled 1, 2, 3, and 4 in Figure 4. To execute the classification procedure, these instances, also known as saved cases, will be stored in CBR memory.

Table 2. Data of Odor Profile of Cocoa bean sample

SAMPLE	CASE ID	S1	S2	S3	S4
COCOA SMELL	CASE 1	1.0000	0.228049	0.157147	0.384785
	CASE 2	1.0000	0.221301	0.150734	0.385856
	CASE 3	1.0000	0.218008	0.149035	0.387133
	CASE 4	1.0000	0.214192	0.144719	0.389100
	CASE 5	1.0000	0.213344	0.141388	0.392772
	CASE 6	1.0000	0.210119	0.140911	0.398511
	CASE 7	1.0000	0.212586	0.141419	0.402968
	CASE 8	1.0000	0.213860	0.138898	0.406596
	CASE 9	1.0000	0.214027	0.137868	0.409534
	CASE 10	1.0000	0.215952	0.137879	0.411046
MUSTY SMELL	CASE 11	1.0000	0.170986	0.137891	0.349392
	CASE 12	1.0000	0.172631	0.137650	0.351745
	CASE 13	1.0000	0.173635	0.137129	0.352596
	CASE 14	1.0000	0.173105	0.136103	0.356095
	CASE 15	1.0000	0.174900	0.135591	0.357245
	CASE 16	1.0000	0.176942	0.135075	0.357712
	CASE 17	1.0000	0.179704	0.134827	0.363163
	CASE 18	1.0000	0.178898	0.133366	0.366276
	CASE 19	1.0000	0.178664	0.132044	0.366575
	CASE 20	1.0000	0.179875	0.131473	0.369316
BLACK PEPPER SMELL	CASE 21	1.0000	0.177648	0.154066	0.400234
	CASE 22	1.0000	0.178130	0.152880	0.402738
	CASE 23	1.0000	0.177103	0.151283	0.406872
	CASE 24	1.0000	0.178930	0.150535	0.409902
	CASE 25	1.0000	0.178527	0.148633	0.408920
	CASE 26	1.0000	0.179319	0.147688	0.410664
	CASE 27	1.0000	0.179839	0.148163	0.410733
	CASE 28	1.0000	0.180081	0.147039	0.413421
	CASE 29	1.0000	0.182264	0.146720	0.413965
	CASE 30	1.0000	0.179893	0.147647	0.413783
SMOKE SMELL	CASE 31	1.0000	0.203237	0.180994	0.435661
	CASE 32	1.0000	0.202805	0.178592	0.438655
	CASE 33	1.0000	0.204851	0.177611	0.442249
	CASE 34	1.0000	0.203823	0.173792	0.446153
	CASE 35	1.0000	0.203329	0.174673	0.445670
	CASE 36	1.0000	0.204904	0.172032	0.448276
	CASE 37	1.0000	0.203430	0.171720	0.448985

	CASE 38	1.0000	0.206341	0.171483	0.451189
	CASE 39	1.0000	0.204985	0.168265	0.455183
	CASE 40	1.0000	0.203086	0.166430	0.455134
COPRA SMELL	CASE 41	1.0000	0.151085	0.134554	0.340603
	CASE 42	1.0000	0.150615	0.133087	0.345271
	CASE 43	1.0000	0.149083	0.133688	0.348022
	CASE 44	1.0000	0.148646	0.133164	0.346432
	CASE 45	1.0000	0.151179	0.133083	0.346852
	CASE 46	1.0000	0.153135	0.133425	0.347227
	CASE 47	1.0000	0.153267	0.131165	0.348426
	CASE 48	1.0000	0.156393	0.130159	0.351785
	CASE 49	1.0000	0.158863	0.128616	0.355499
	CASE 50	1.0000	0.158081	0.127293	0.359839

According to Table 2, the highest normalised value is 1.0000, which can be found in column S1, and the lowest normalised value is 0.127293 to 0.180994, which can be found in column S3. It is because on S1, the raw data was separated by the highest value in each row, and since then, every row of raw data has to be divided by the value of S1. The sensors in the e-nose have varying degrees of responsiveness.

Intelligent Classification

Sensor attribution (S1 through S4) refers to the sensors that were used in E-nose. The sum of weights is called a normalised weight. The stored case of the cocoa bean application, also known as the previous case, is the source case, while the target case is the new case or case that needs to be checked. The similarity was calculated using Equation 2 and tabulated in Table 3 below shows the example similarity percentage.

Table 3. CBR Formulation Table

SENSOR ATTRIBUTION	WEIGHT	NORMALIZED WEIGHT	SOURCE CASE	TARGET CASE	SIMILARITY FUNCTION	WEIGHTED SIMILARITY
S1	1	0.25	1.0000	1.0000	1.0000	0.25
S2	1	0.25	0.228049	0.221301	0.993252	0.248313
S3	1	0.25	0.157147	0.15074	0.993587	0.248397
S4	1	0.25	0.384785	0.385856	0.998929	0.249732
PERCENTAGE SIMILARITY BETWEEN 2 CASES (%)						99.64

Performance Measure

The confusion matrix from CBR voting results is depicted in Table 4. This table can have a clearer picture of cocoa bean odour classification. There are actual and predicted results in this confusion matrix table. The actual case refers to the sample's actual situation, while the forecasting case is based on the formulation of CBR and voting outcomes. It showed that every sample consists of 30 cases that were predicted to be in their group. The data then is evaluated in performance evaluation to find its sensitivity, specificity and accuracy. Each category has 30 true positives, with a total of 150 true positives for all groups. The evaluation is tabulated in Table 5, where the sensitivity, specificity, and accuracy were calculated with a multiplication of 100%.

Table 4. Confusion Matrix Table

TOTAL CASE = 150		PREDICTED RESULT				
		COCOA SMELL	MUSTY SMELL	BLACK PEPPER SMELL	SMOKE SMELL	COPRA SMELL
ACTUAL RESULT	COCOA SMELL	30	0	0	0	0
	MUSTY SMELL	0	30	0	0	0
	BLACK PEPPER SMELL	0	0	30	0	0
	SMOKE SMELL	0	0	0	30	0
	COPRA SMELL	0	0	0	0	30

Table 5. CBR Performance Evaluation

PERFORMANCE EVALUATION	COCOA SMELL	MUSTY SMELL	BLACK PEPPER SMELL	SMOKE SMELL	COPRA SMELL
TOTAL CASE	30	30	30	30	30
TRUE COCOA SMELL	30	0	0	0	0
FALSE COCOA SMELL	0	0	0	0	0
TRUE MUSTY SMELL	0	30	0	0	0
FALSE MUSTY SMELL	0	0	0	0	0
TRUE BLACK PEPPER SMELL	0	0	30	0	0
FALSE BLACK PEPPER SMELL	0	0	0	0	0
TRUE SMOKE SMELL	0	0	0	30	0
FALSE SMOKE SMELL	0	0	0	0	0
TRUE COPRA SMELL	0	0	0	0	30
FALSE COPRA SMELL	0	0	0	0	0
SENSITIVITY COCOA SMELL (%)	100	100	100	100	100
SENSITIVITY MUSTY SMELL (%)	100	100	100	100	100
SENSITIVITY BLACK PEPPER SMELL (%)	100	100	100	100	100
SENSITIVITY SMOKE SMELL (%)	100	100	100	100	100
SENSITIVITY COPRA SMELL (%)	100	100	100	100	100
SPECIFICITY COCOA SMELL (%)	100	100	100	100	100
SPECIFICITY MUSTY SMELL (%)	100	100	100	100	100
SPECIFICITY BLACK PEPPER SMELL (%)	100	100	100	100	100
SPECIFICITY MUSTY SMELL (%)	100	100	100	100	100
SPECIFICITY COPRA SMELL (%)	100	100	100	100	100
ACCURACY COCOA SMELL (%)	100	100	100	100	100
ACCURACY MUSTY SMELL (%)	100	100	100	100	100
ACCURACY BLACK PEPPER SMELL (%)	100	100	100	100	100

ACCURACY SMOKE SMELL (%)	100	100	100	100	100
ACCURACY COPRA SMELL (%)	100	100	100	100	100
OVERALL SENSITIVITY (%)	100				
OVERALL SPECIFITY (%)	100				
OVERALL ACCURACY (%)	100				

The sensitivity, specificity and accuracy of every sample gave results of 100% for each of the cocoa bean samples. The overall sensitivity, specificity and accuracy also showed the result of 100% for cocoa bean sample classification using the CBR classification technique.

CONCLUSION

The results obtained from this research noted that cocoa bean samples, cocoa smell, musty smell, black pepper smell, smoke smell and copra smell have their odor profile. Changes in cocoa bean contents affected the changes in fragrance and odor-profile of the samples, resulting in variations in degradation levels between samples. Although the cocoa bean odor patterns plotted to seem identical, the cocoa bean sample classified 100% correctly to each smell. By applying CBR as intelligent classification, the distinct type of cocoa bean smell can be analysed with a rate of 100% in sensitivity, specificity and accuracy. It is suggested to conduct more experiments on other types of foods and things the humans can consume and threats that will harm the environment and living beings. Moreover, we will improvise how the data is determined after being collected by the E-nose to show the research's effectiveness and efficiency further.

ACKNOWLEDGEMENT

The authors want to express our gratitude to all staff and hired hands of the Faculty of Manufacturing & Mechatronic Engineering Technology and Faculty of Electrical & Electronic Engineering Technology, Universiti Malaysia Pahang (UMP) involved in this research and to Malaysia Cocoa Board for providing the samples. This research and development are supported by UMP Research Grant (UIC160301 and UIC160904).

REFERENCES

[1] C. A. Wallace, W. H. Sperber, and S. E. Mortimore, "Food Safety for the 21st Century Managing HACCP and Food Safety Throughout the Global Supply Chain," 2018.

[2] A. C. Aprotosoaie, S. V. Luca, and A. Miron, "Flavor Chemistry of Cocoa and Cocoa Products-An Overview," *Compr. Rev. Food Sci. Food Saf.*, vol. 15, no. 1, pp. 73–91, Jan. 2016, doi: 10.1111/1541-4337.12180.

[3] E. Poivet *et al.*, "Functional odor classification through a medicinal chemistry approach," *Sci. Adv.*, vol. 4, no. 2, p. eaao6086, Feb. 2018, doi: 10.1126/sciadv.aao6086.

[4] E. K. Jenkins, M. T. DeChant, and E. B. Perry, "When the nose doesn't know: Canine olfactory function associated with health, management, and potential links to microbiota," *Frontiers in Veterinary Science*, vol. 5, no. MAR. Frontiers Media S.A., p. 1, Mar. 29, 2018, doi: 10.3389/fvets.2018.00056.

[5] Y. H. Jung, B. Park, J. U. Kim, and T. il Kim, "Bioinspired Electronics for Artificial Sensory Systems," *Adv. Mater.*, vol. 31, no. 34, Aug. 2019, doi: 10.1002/adma.201803637.

[6] J. Yan *et al.*, "Electronic nose feature extraction methods: A review," *Sensors (Switzerland)*, vol. 15, no. 11, pp. 27804–27831, 2015, doi: 10.3390/s151127804.

[7] Y. Chang *et al.*, "Detection of Volatile Organic Compounds by Self-assembled Monolayer Coated Sensor Array with Concentration-independent Fingerprints," *Sci. Rep.*, vol. 6, no. 1, pp. 1–12, Apr. 2016, doi: 10.1038/srep23970.

[8] J. Tan and J. Xu, "Applications of electronic nose (e-nose) and electronic tongue (e-tongue) in food quality-related properties determination: A review," *Artif. Intell. Agric.*, vol. 4, pp. 104–115, 2020, doi: 10.1016/j.aiia.2020.06.003.

[9] P. Dhar, P. Kashyap, N. Jindal, and R. Rani, "Role of electronic nose technology in food industry," *Emerg. Sustain. Technol. Food Process.*, no. March, pp. 1–2, 2018.

[10] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," *Int. J. Inf. Manage.*, vol. 48, pp. 63–71, Oct. 2019, doi: 10.1016/j.ijinfomgt.2019.01.021.

[11] J. Bieger, K. R. Thórisson, B. R. Steunebrink, T. Thorarensen, and J. S. Sigurðardóttir, "Evaluation of General-Purpose Artificial Intelligence: Why, What & How."

[12] M. F. Zahari *et al.*, "Gaharu Sensor: Classification Using Case Based Reasoning (CBR)," *Jeecie*, vol. 1, no. 8, pp. 38–41, 2016.

[13] S. M. Daud *et al.*, "Classification of Lubricant Oil Geometrical Odor-Profile Using Cased-Based Reasoning," in *Lecture Notes in Electrical Engineering*, 2020, vol. 632, pp. 139–151, doi: 10.1007/978-981-15-2317-5_13.

[14] Zhai, Ortega, Castillejo, and Beltran, "A Triangular Similarity Measure for Case Retrieval in CBR and Its Application to an Agricultural Decision Support System," *Sensors*, vol. 19, no. 21, p. 4605, Oct. 2019, doi: 10.3390/s19214605.

[15] R. M. Mawardzi, A. Japper-Jaafar, M. S. Najib, S. M. Daud, and S. T. Ya, "Intelligent classification of waxy crude oil odor-profile at different temperature," 2019, doi: 10.1088/1757-899X/469/1/012071.