MEKATRONIKA
e-ISSN: 2637-0883
VOL. 4, ISSUE 2, 48 – 57
DOI: https://doi.org/10.15282/mekatronika.v4i2.8943

ORIGINAL ARTICLE

Correlation Investigation of Odour-Profile and Temperature Intensity

Muhammad Naqiuddin Ali Ibrahim[1], Muhammad Sharfi Najib[1,*] and Suhaimi Mohd Daud[2].

[1]Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[2]Pusat Bioaromatik, Universiti Malaysia Pahang, 26300 Pahang, Malaysia.

ABSTRACT – Water is important to all human and economic activity. Despite its abundant water resources, Malaysia's water demand and consumption was increased in recent years. This study had been investigating the condition of Malaysia's water resources at present. This study aims to identify water by examining its odour profile and temperature intensity. E-Anfun E-nose was applied to generate a pattern of the volatile compounds included in the samples. The optimal correlation was discovered between odour feature extractions from many samples produced using the same E-Anfun unit. An intelligent algorithm namely Case-Based Reasoning (CBR) was employed to detect the odour of different national standard tap water samples. CBR has produced measurable performance of 100% classification rate of accuracy.

ARTICLE HISTORY
Received: 20th Oct 2022
Revised: 10TH Nov 2022
Accepted: 23rd Nov 2022

KEYWORDS
Water
Odour-Profile
E-nose
Artificial Intelligence
Case Based Reasoning

1.0 INTRODUCTION

Water is an essential component of practically all forms of production, including agriculture, industry, power, and transportation [1]; water is a source of life, livelihood, and prosperity. It is a vital component of life and essential to the existence of all living things [2]. [3] It had been acknowledged for millennia that human survival and ecological protection depend on the dependable availability of abundant water of adequate quality. The World Health Organization (WHO) [4] states that access to safe drinking water is a fundamental human right. Since 1990, 2,6 billion people have gained access to improved drinking water, with 96% residing in urban areas and 84% living in rural regions (WHO and UNICEF 2015). 663 million people continue to rely on unimproved tap water sources, including as unprotected wells, springs, and surface water, according to [5]. Almost half of the population resides in Sub-Saharan Africa, while one-fifth is in Southern Asia.

Malaysia is a tropical nation that receives abundant rainfall throughout the year, but the nation continues to struggle with water scarcity and water quality issues. Biochemical oxygen demand (BOD) from sewage, agriculture, and manufacturing, ammoniacal nitrogen from animal farms and domestic sewage, and suspended particles from earthwork and land clearing activities are the principal pollutants in rivers [6]. The remaining 1% is sourced from groundwater [7]. In Malaysia, tap water, bottled water, and bottled mineral water are the most common sources of drinking water [8]. The rising costs of water and wastewater treatment, as well as the increasing demand for safe drinking water, have led businesses and governments around the world to study innovative water-saving technology.

E– Nose is a device that compares and identifies gaseous samples (or emissions from liquid or solid samples) using non– specific chemical sensors and statistical pattern recognition algorithms [9]. E-nose offers a wide range of commercial applications in fields such as agriculture, biomedical, cosmetics, the environment, food, industry, the military, regulatory science, and pharmaceuticals [10]. The role of the E-nose is to detect odour created as gases or vapour [11], which is ideal for water because it contains a variety of volatile odour-profiles and temperature intensities [12]. CBR is one of the reported intelligent systems that requires no training and is suitable for a limited domain area [13]

2.0 METHODOLOGY

Figure 1 shows the whole process of analyzing and interpreting the data from the sample collection until the performance measurement. The water samples were taken from National standard tap water and were used in the experiment with five samples with different temperature.

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

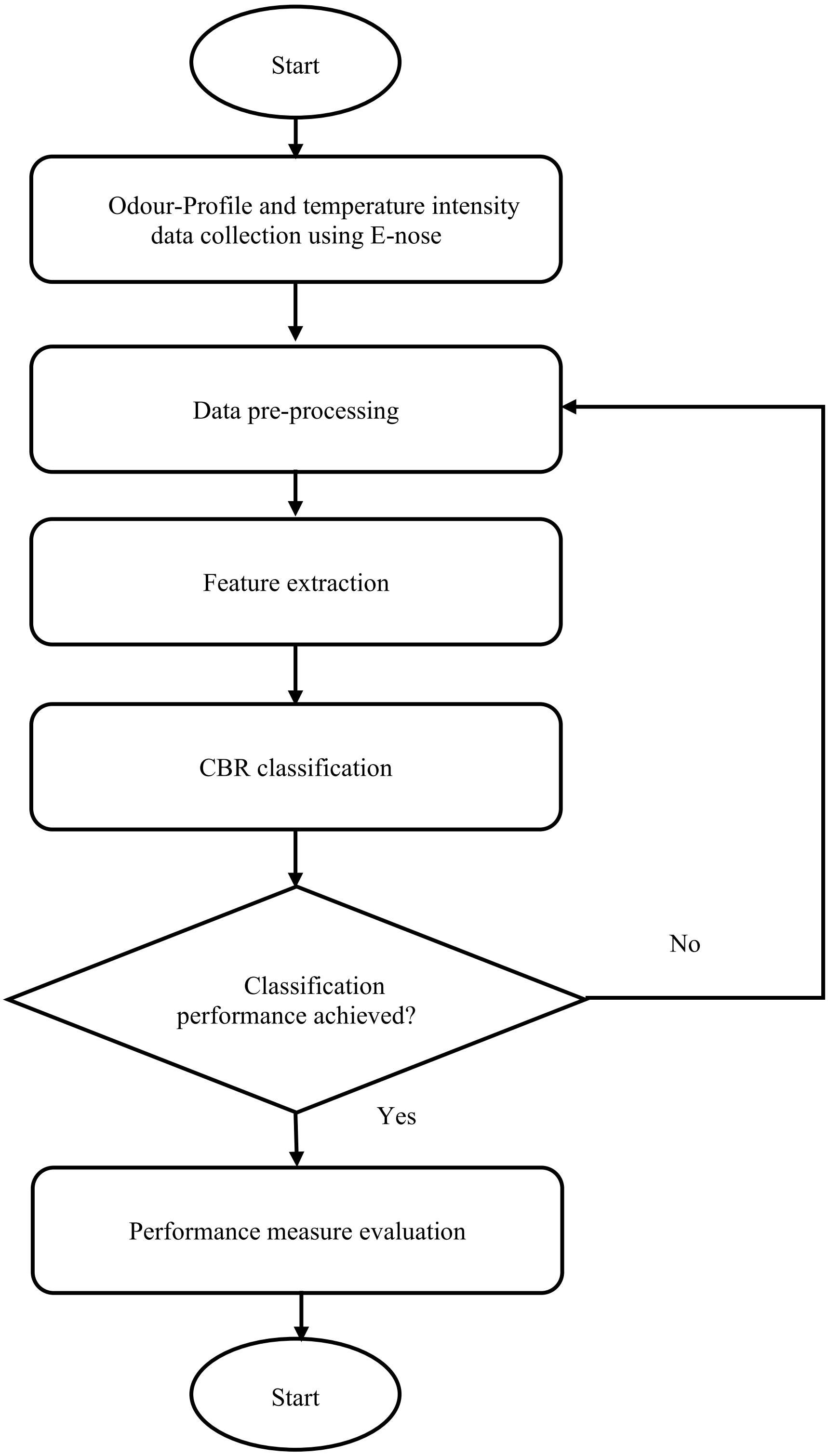

Figure 1. Methodology Flowchart.

2.1 Experimental Setup

E-Anfun were employed to measure tap water's odour and temperature. This tap water had five samples with different temperatures, which is volatile and suitable for the sample beaker's size. The overall classification strategy for tap water odour profiles is depicted in Figure 2. The odour profile data were acquired by interconnecting E-Anfun directly to computer with specified data sampling.

Figure 2. Experimental setup of an E-Anfun

As shown in Figure 2 and Figure 3, it began with data collection of sample data from a tap water odour-profile using E-Anfun. There are five samples of tap water from Malaysia. Figure 5 depicts how each sample was taken into five experiment and placed in a separated 400 ml sample beaker. As a result, the total number of measured sample vials of tap water from National standard tap water is five, consisting of 1000 raw measured data for each sample (200 measures x 5 trials). The measurements of raw data were then pre-processed using a normalization technique and a statistical method. Extraction of statistically significant characteristics was exercised in order to acquire significant variance input attributes. The major input characteristics retrieved and saved in CBR library cases in which a total of fifty samples were established (50 cases from tap water with different temperature). Vector-weighting was then applied to the input features to increase variance. After successfully extracted significant feature using advance normalization statistical method, the CBR classification continued with CBR computation and voting, followed by the evaluation of the obtained CBR performance.

Figure 3. Experimental setup of E-Anfun

Figure 4. 400ml beaker for each sample

2.2 Data Measurement

The raw data from each sensor is displayed in Table 1 below. Each sample consists of 200 rows of data. And each sample must undergo five experiments, culminating in the collection of with N samples, where N=1000 data measurements each sample.

Table 1. Data measurement of tap water odour-profile.

Data Measurement	S1	S2	S3	S4
1	DM_{11}	DM_{12}	DM_{13}	DM_{14}
2	$DM_{2}1$	DM_{22}	DM_{23}	DM_{24}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
N	DM_{N1}	DM_{N2}	DM_{N3}	DM_{N4}

Sensors 1, Sensors 2, Sensors 3, and Sensors 4 are denoted by S1, S2, S3, and S4 respecively for each sample in the table above. DM represents the data measurement of tap water samples.

2.3 Data Pre-Processing

By applying Equation 1, the previously gathered raw data were adjusted (1). To acquire the normalized values, divide each row of raw data measurement by its own row's highest value. Consequently, the value was rescaled to a smaller value between zero and one (0-1). R' refers to the normalized value, R to the measurement of raw data, and R_{max} to the highest value from each raw. The minimum and maximum values for normalized data are therefore 0 and 1, respectively. The normalized value is particularly useful for odour profile extraction. By tabulating the values, table 2 was formed.

$$\boldsymbol{R'} \frac{\boldsymbol{R}}{\boldsymbol{Rmax}} \quad (1)$$

Table 2. Normalised value of tap water odour-profile.

Data Measurement	S1	S2	S3	S4
1	ND_{11}	ND_{12}	ND_{13}	ND_{14}
2	$ND_{2}1$	ND_{22}	ND_{23}	ND_{24}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
N	ND_{N1}	ND_{N2}	ND_{N3}	ND_{N4}

Table 2 shows the data normalization table for the tap water sample. The normalized data consists of 1000x4 data points. S1, S2, S3, and S4 correspond to Sensors 1, 2, 3, and 4, respectively. ND represents the normalized data of the tap water sample.

2.4 Feature Extraction

Using the normalization technique, the distinguished characteristic of the odour of each tap water sample were retrieved. Before the data became displayed graphically, each data (1000 rows of data) was subjected to a statistical method known as the mean calculation technique to reduce the raw data measurement to 200 rows. Consequently, the data was clustered into 10 cases per sample to continue the classification of the odour pattern. As a result, each tap water sample had been 10 clusters or data cases. Thus, 50 data cases were been generated (5 sample with different temperature that are 10°C until 50°C).

2.5 Intelligent Classification

In this study, Case Based Reasoning (CBR) is applied to classify the smell pattern of 5 samples containing acquired raw data. Figure 5 depicts the CBR cycles for tap water classification. The cycle began with unidentified tap water, which serves as the sample for testing. The odour profile from the unidentified water sample was retrieved in order to obtain the cases kept in memory. The system compares the unknown odour-profile water sample to the previously stored odour-profile since CBR learns from past circumstances. If the unknown sample had a high percentage of resemblance with a previously stored water sample, the system used the data from the previously stored case to come to a decision.

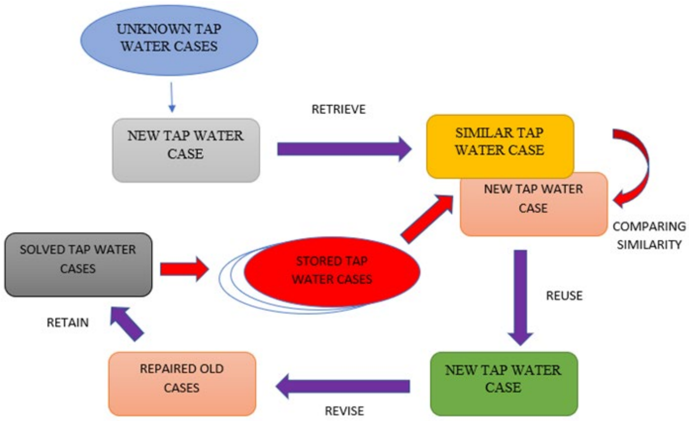

Figure 5. CBR cycle for tap water classification

In the CBR retrieval cycle, the percentage of similarity between two examples was determined by selecting one case from 50 stored examples. The remaining 49 cases are left as the leftover cases in storage. The percentage of similarity was computed using Equation (2). If the percentage of similarity between two cases is large, it indicates that the distance between them is small and that they are members of the same group.

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

In this equation, T and S represent the target and source cases, respectively, n is the number of attributions for the tap water sample, "i" is the single attribution for each case, f is the formulation of the similarity function for the tap water sample, and w is the weight of each attribution.

2.6 Performance Measure

The result of the CBR Classification was evaluated using a confusion matrix. Equations (2)-(4) were utilized to calculate the overall sensitivity, specificity, and accuracy of the tap water sample in the categorization process. The classification's sensitivity was calculated by dividing the true positive value of the classification result by the sum of the classification's true positive (TP) and false negative (FN). The classification's specificity was calculated by diving true negative (TN) by the sum of false positive (FP) and true negative (TN). While the classification accuracy was calculated by dividing the sum of true positive and true negative with the total case (P+N).

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

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

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

3.0 RESULT AND DISCUSSION

3.1 Raw Data Measurement

Figure 6 shows the graph of raw data national tap water for sample 1 until 5. Each sample was subjected to five repeated experiments with different temperatures, each with 200 data measurements, for a total of 1000 data measurements per sample. As a consequence, 5000 data measures were gathered from all samples and are summarized here

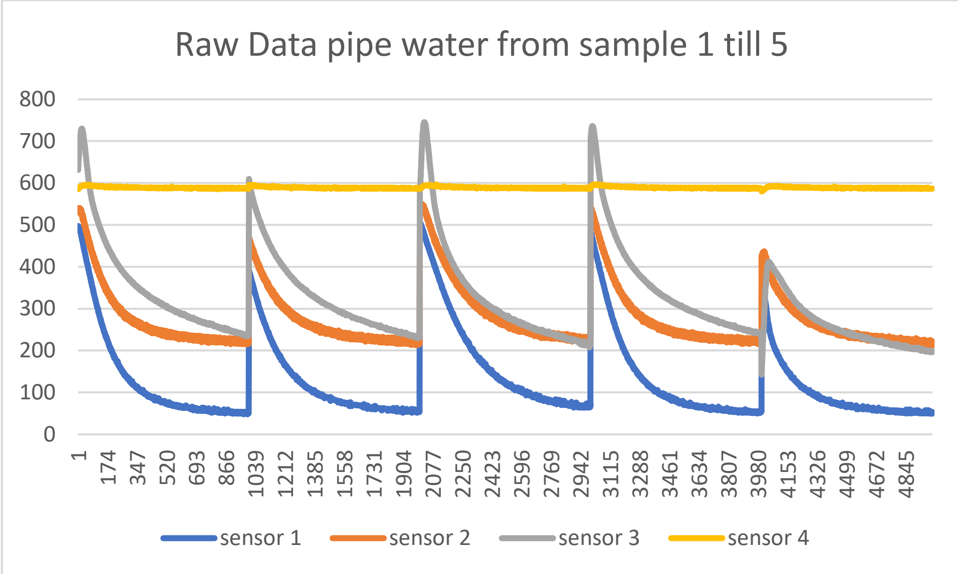

Figure 6. Graph of raw data tap water from sample 1 till 5

E-Anfun was designed to sense four unique chemical sensor arrays incorporated by S1, S2, S3, and S4. Sensor S1 reads the highest for all samples, while S3 reads the lowest for all tap waters samples. Each sample, from sample 1 to sample 5, had a slightly similar pattern, as illustrated in the graphic below. Despite the nearly identical patterns, there are significant differences between the samples that were quantified and used in the categorizing process. A data pre-processing stage is necessary to enhance the significance of the pattern.

3.2 Data Pre-Processing

In order to normalize 5000 previously obtained data measurements, each value in each row of data measurements was divided by the highest value in its own row. The 5000 normalized data points were then separated into five groups, one for each sample. One 1000 normalized data set were then grouped into 10 examples per category. Table 5 displays the CBR case library for tap water samples. The table comprises 50 significant examples, with 10 cases for each temperature sample. These cases were known as stored cases and it was stored in the CBR case library for odour-profile classification. Table 5 in S4 column had the highest value of 1. Because the normalized value of S4 is the same in all situations, the data are not particularly significant for categorization due to the identical value. Value 1 in the entire column S4 identifies the characteristics of the odour-profile of tap water. In this table, the highest normalized value is contained in the full column S4, which includes the value '1', while the lowest normalized value is always contained in the entire column S1. Due to the fact that the previous raw data were split based on the highest value in each row, only column S4 contains the same value. Previously, column S4 had only raw data readings from sensor 4 for tap water. To normalize data, divide each measurement row by the highest value of S1, S2, S3, and S4. Every measurement data row in the data collection had a maximum value in column S4. As a result, when the data measurement was normalized, value 1 was created in column S4 for each normalized data row. Variable sensitivity is exhibited by E-Anfun sensor S4, which represents sensor 4, had been a very high sensitivity on samples of tap water.

Table 5. CBR case library for tap water sample

Case ID	S1	S2	S3	S4
Case_01	0.09147	0.373716	0.365775	1
Case_02	0.091538	0.37375	0.365706	1
Case_03	0.09147	0.373375	0.36557	1
Case_04	0.091298	0.373641	0.365478	1
Case_05	0.091448	0.373373	0.365313	1
Case_06	0.091605	0.373522	0.365154	1
Case_07	0.091641	0.373273	0.364905	1
Case_08	0.091486	0.373301	0.364848	1
Case_09	0.091638	0.373107	0.364672	1
Case_10	0.091536	0.372988	0.364518	1
Case_11	0.085033	0.362743	0.311807	1
Case_12	0.085	0.362783	0.311676	1

Case_13	0.085069	0.362715	0.311455	1
Case_14	0.084964	0.362788	0.311258	1
Case_15	0.084928	0.362765	0.311185	1
Case_16	0.084862	0.362772	0.311105	1
Case_17	0.084843	0.362936	0.310981	1
Case_18	0.084791	0.362793	0.310889	1
Case_19	0.084755	0.362855	0.310868	1
Case_20	0.084687	0.362668	0.310698	1
Case_21	0.097478	0.38763	0.306964	1
Case_22	0.097552	0.387514	0.306823	1
Case_23	0.097618	0.387388	0.306579	1
Case_24	0.097569	0.387326	0.306431	1
Case_25	0.0975	0.38719	0.30636	1
Case_26	0.097688	0.387207	0.306276	1
Case_27	0.09774	0.387282	0.306247	1
Case_28	0.097586	0.387309	0.306054	1
Case_29	0.09748	0.387126	0.30591	1
Case_30	0.097443	0.386943	0.305663	1
Case_31	0.110999	0.397821	0.348511	1
Case_32	0.110982	0.397685	0.348324	1
Case_33	0.111051	0.397845	0.348192	1
Case_34	0.110958	0.397703	0.348045	1
Case_35	0.11079	0.398015	0.347916	1
Case_36	0.110792	0.397902	0.347855	1
Case_37	0.110626	0.397865	0.347645	1
Case_38	0.110662	0.397888	0.347515	1
Case_39	0.110678	0.397752	0.347379	1
Case_40	0.11068	0.39781	0.347198	1
Case_41	0.134276	0.421729	0.384172	1
Case_42	0.134225	0.421248	0.38415	1
Case_43	0.13423	0.421687	0.384147	1
Case_44	0.134205	0.421456	0.384236	1
Case_45	0.134223	0.421439	0.384254	1
Case_46	0.134176	0.421515	0.38443	1
Case_47	0.134058	0.42143	0.384447	1
Case_48	0.133963	0.42184	0.384602	1
Case_49	0.133982	0.421609	0.384811	1
Case_50	0.134088	0.422149	0.384994	1

As depicted in Figure 7, the normalized data were clustered into 10 cases for each sample in order to extract the profile features for each of them. The features between each sample are slightly different by using normal visualization. In order to find significant differences, normalized data statistically analyzed.

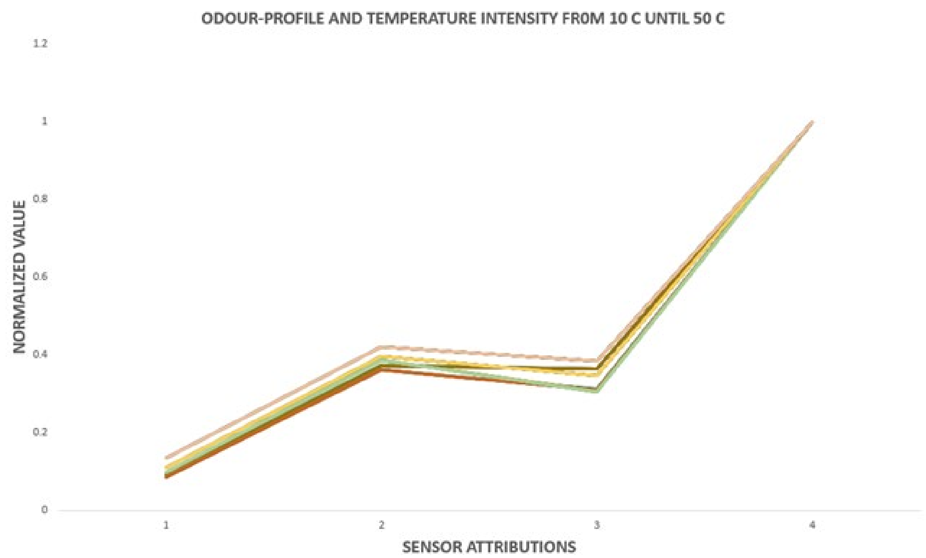

Figure 7. Graph shows the data of tap water temperature 10C until 50C in same graph

Table 6 displays the CBR similarity formulation for samples 1 and 2. The sensor array is represented by S1, S2, S3, and S4 properties. The source is a case containing a sample of tap water. The target is the present case. Equation (2) is utilized to determine the degree of resemblance between the two cases below. By dividing each individual weight by the total weight, the normalized weight was calculated. To determine the percentage of similarity between two cases, the similarity computation for each sensor was summed.

Table 6. CBR formulation for one case of tap water samples

	Stored Case	Target Case	Similarity Function	Weight	Norm_weight	Norm_w*sim
S1	0.091537	0.999931	0.999931	1.000	0.2500	0.249982
S2	0.373749	0.999965	0.999965	1.000	0.2500	0.249991
S3	0.365706	0.999931	0.999931	1.000	0.2500	0.249982
S4	1	1	1	1.000	0.2500	0.25
	Total Similarity Between Two Cases				0.999957	

The weight vector is displayed in Table 6. The characteristic of this electronic nose is represented by its sensors. Each sensor was assigned a local weight value of 1 by the expert. To execute the odour-profile classification, the classification must construct the local weight for each attribution, and the weight's value were modified heuristically to achieve a more accurate classification. For all characteristics, the value of local weight was set to 1.

3.3 CBR Performance Measure

Table 7 demonstrates the confusion matrix. This table provides the extracted value from the voting table, which displays the actual and predicted output of the CBR similarity function, as seen in Table 6. From each of the samples that was utilized to validate the classifier's performance, the three highest values from the similarity prediction result were selected.

Table 7. Confusion Matrix

		Predicted				
		sample 1	sample 2	sample 3	sample 4	sample 5
actual	sample 1	30	0	0	0	0
	sample 2	0	30	0	0	0
	sample 3	0	0	30	0	0
	sample 4	0	0	0	30	0
	sample 5	0	0	0	0	30
Total		30	30	30	30	30

Table 8 and 9 displays the performance evaluation for classifying tap water odour profiles using the CBR classifier. Using equations (3)-(5), the classification's sensitivity, specificity, and accuracy were calculated.

Table 8. CBR Performance Evaluation

	ACTUAL				
	sample 1	sample 2	sample 3	sample 4	sample 5
Total Cases	30	30	30	30	30
Condition Positive(P)	30	30	30	30	30
Condition Negative(N)	120	120	120	120	120
True Positive(TP)	30	30	30	30	30
False Positive (FP)	0	0	0	0	0
True Negative (TN)	120	120	120	120	120
False Negative(FN)	0	0	0	0	0
Sensitivity=TP/(TP+FN)	1	1	1	1	1
Specify=TN/(TN+FP)	1	1	1	1	1
Accuracy= (TP+TN)/(P+N)	1	1	1	1	1
Overall Sensitivity(%)	100				
Overall Specificity(%)	100				
Overall Accuracy(%)	100				

Table 9. Overall Performances

OVERALL (%)	SENSITIVITY	100
	SPECIFICITY	100
	ACCURACY	100

The examples for samples 1 through 5 with a result of 1 exhibit an evaluation value of 1 for each water sample in terms of sensitivity, specificity, and precision. Consequently, the overall sensitivity, specificity, and accuracy of the CBR classification approach for classifying tap water samples are all 100%. It demonstrates that the CBR classifier was used to classify the odour of tap water with 100 percent performance.

4.0 CONCLUSION

This investigation reveals that the sources of five samples of tap water at temperatures ranging from 10 to 50 degrees Celsius have distinct odour. Even if the patterns and odour of the tap water samples are relatively similar, E-nose sensor array can distinguish between them. Using Case-Based Reasoning, the classifier is able to classify extracted data even when the source cases stored in memory have several differences in a small data dimension (CBR). It is demonstrated when this classification method results in a 100% classification rate.

5.0 ACKNOWLEDGEMENT

The authors would like to acknowledge to Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang for funding this study under the Research Grant (PDU223205).

6.0 REFERENCES

[1] D. Grey and C. W. Sadoff, "Sink or Swim? Water security for growth and development," *Water Policy*, vol. 9, no. 6, pp. 545–571, 2007, doi: 10.2166/WP.2007.021.

[2] A. Azlan, H. Khoo, M. Idris, … A. I.-… K. M., and undefined 2011, "Evaluation of selected metal elements in commercial drinking water and tap water in peninsular Malaysia," *ejournal.ukm.my*, Accessed: Oct. 22, 2021. [Online]. Available: http://ejournal.ukm.my/jskm/article/view/1319.

[3] A. K. Biswas and K. E. Seetharam, "Asian water development outlook, 2007: achieving water security for Asia.," *Int. J. Water Resour. Dev.*, vol. 24, no. 1, pp. 145–176, 2008.

[4] WHO and Unicef, "Global Water Supply and Sanitation Assessment 2000 Report," *Water Supply*, p. 87, 2000, Accessed: Dec. 01, 2021. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Global+Water+Supply+and+Sanitation+Assessment+2000+Report#0.

[5] WHO, "Global leprosy update, 2017: reducing the disease burden due to leprosy," *Wkly. Epidemiol. Rec.*, vol.

93, no. 35, pp. 445–456, 2018.
[6] "Laporan Kualiti Alam Sekeliling 2017 – Enviro Knowledge Center." https://enviro2.doe.gov.my/ekmc/digital-content/malaysia-environmental-quality-report-2017-2/ (accessed Dec. 01, 2021).
[7] "(PDF) Major inorganic elements in tap water samples in Peninsular Malaysia." https://www.researchgate.net/publication/221804247_Major_inorganic_elements_in_tap_water_samples_in_Peninsular_Malaysia (accessed Dec. 01, 2021).
[8] M. S. Aini, A. Fakhrul-Razi, O. Mumtazah, and J. C. M. Chen, "Malaysian households' drinking water practices: A case study," *https://doi.org/10.1080/13504500709469749*, vol. 14, no. 5, pp. 503–510, Oct. 2009, doi: 10.1080/13504500709469749.
[9] V. Diz, M. Cassanello, and R. M. Negri, "Detection and discrimination of phenol and primary alcohols in water using electronic noses," *Environ. Sci. Technol.*, vol. 40, no. 19, pp. 6058–6063, Oct. 2006, doi: 10.1021/ES052322E.
[10] A. D. Wilson and M. Baietto, "Applications and advances in electronic-nose technologies," *Sensors*, vol. 9, no. 7, pp. 5099–5148, Jul. 2009, doi: 10.3390/S90705099.
[11] "JEECIE-Journal of Electrical, Electronics, Control and Instrumentations Engineering." http://apps-cfm.ump.edu.my/research/jeecie/index.cfm (accessed Dec. 01, 2021).
[12] J. A. Pino, R. Marbot, A. Delgado, C. Zumárraga, and E. Sauri, "Volatile constituents of propolis from honey bees and stingless bees from yucatán," *J. Essent. Oil Res.*, vol. 18, no. 1, pp. 53–56, 2006, doi: 10.1080/10412905.2006.9699384.
[13] P. Perner, "Mining sparse and big data by case-based reasoning," *Procedia Comput. Sci.*, vol. 35, no. C, pp. 19–33, 2014, doi: 10.1016/j.procs.2014.08.081.