Universiti Malaysia PAHANG
Engineering · Technology · Creativity

ORIGINAL ARTICLE

Intelligent Clasification Sewage Treatment Plant (STP) Using E-Nose

Fatin Munirah Nasarudin[1], Muhammad Sharfi Najib[1], Suhaimi Mohd Daud[1], Mujahid Mohamad[1], Saiful Nizam Tajuddin[2], Che Mohd Aizal Che Mohd[2]

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

ABSTRACT – The environment is an invaluable gift. In an era of progress, economic activities and development projects are often carried out to improve living standards and keep pace with other developed countries. However, this activity has had a negative impact on the environment because some parties still fail to control the disposal of waste that can cause environmental pollution. The effects of this pollution can give to discomfort and disruption to the life of the community. It is because the effects of toxic emissions have caused air pollution to spread foul odors. Therefore, this study was conducted to classify air odor and water odor from the treatment plant in the area of Universiti Malaysia Pahang, Gambang Campus. The classification of air and water odors was done using case-based reasoning.

1. INTRODUCTION

Primarily, the main concern of sewage treatment works is the treatment of liquid waste. Solid and gaseous wastes are generated in the process of treatment, which can lead to secondary contamination if not handled effectively. Gaseous waste, resulting in odour-like pollution, may have the most significant effect on the population in the vicinity of the sewage treatment work [1].

In either solution or suspension, sewage is water-carried waste that flows away from society and is known as wastewater [2]. Wastewater can be classified into four categories: domestic, industrial, infiltration, and stormwater [3]. Domestic sewage comprises liquid wastes from urinals, lavatories, toilets, kitchen sinks, washbasins, and other fixtures in homes, businesses, and institutions. Due to the presence of human excreta, this sewage is usually highly foul. Industrial sewage is made up of liquid wastes from different industries' industrial processes, such as dyeing, paper making, brewing, etc. The quality of the industrial sewage depends mainly upon the type of industry and the chemicals used in their process waters. Sometimes, they may be very foul and require extensive treatment before being disposed of in public sewers. Storm sewage means water that is discharged from a surface due to rainfall, snowmelt or snowfall [4].

An electronic nose (E-nose) is a device that consists of a collection of partially unique electronic chemical sensors and a pattern recognition device that can detect simple or complex odors [5]. Electronic noses have a wide range of commercial applications, including agriculture, biomedicine, the environment, food, manufacturing, the military, and various scientific research fields. The invention of the device also enhances product characteristics, quality, and consistency as a result. It increases quality management capabilities provided by E-nose for monitoring all phases of the industrial manufacturing process [6].

E-nose was build-up by the chemical sensor array, odour chamber, E-nose pump and microcontroller, which have been widely used to analyse volatile organic compounds [7]. The volatile compounds react with the sensor surfaces, causing changes in the latter's chemical and physical properties [8]. The sensor array took data readings of environmental odour collected within the chamber where the pump located within the upper part of the e-nose sucks in the odour into the e-nose chamber. Considering the main function is to detect multiple aromas or odors, the sensing array should include various types of individual sensors where every sensor is responsible for detecting chemicals [9].

The data for E-nose detection will be analysed using standardisation techniques, which is normalisation technique [10]. Normalisation technique is scaling technique or mapping techniques or a pre-processing stage [11]. The features of each sample were extracted from the normalised value. Groups will be formed from the normalised value. Ten cases were obtained in each category from the mean calculation of the normalised value in which each data case will be stored into Case-based reasoning (CBR) database [14]. By measuring the similarity of the cases, it will be recovered and reused for new cases [12].

CBR is a major paradigm in automated reasoning and machine learning [13]. There are three terms in the phrase case-based reasoning and they require a brief clarification. A case is essentially a description of a problem that has been solved. A list of such cases is a case base. Since cases are the first source of reasoning, the word-based means that they are based on them. The word that is most typical of the methodology is reasoning. It implies that given a problem to be solved, the method is intended to conclude using cases [14]. The main CBR strategies typically involve four elements:

case recovery, case reuse, case revision, and case retention. The retrieval of cases entails creating a case database by defining ontological characteristics of both problem and solution attributes. Case reuse, also known as case adaptation, seeks to map previous cases' solutions to the target problem [15]. If an exact match cannot be identified, revise and modify the most similar case or group of cases as appropriate. As part of the current situation, maintaining the new solution is very useful for future problem solving by retaining the memory knowledge to solve new problems in the future [9].

This paper will summarise the Sewage Treatment Plant (STP) classification methodology based on odor profiles using an e-nose and a case-based reasoning classifier.

2. METHODOLOGY

Figure 1 shows the flowchart for the principal analysis of STP classification using CBR and odour profiles. The approach starts with collecting raw data using an E-nose device. Next, the pre-processing process where all raw data was calculated using the normalisation and technique of mean calculation was calculated. Then, feature extraction step where all the samples' unique feature was extracted to establish the odour-profile. Later, the CBR classification technique was used to classify the odour profiles. Last but not least, the overall sensitivity, specificity, and accuracy of the STP sample's classification system will be determined by evaluating the classification result's results.
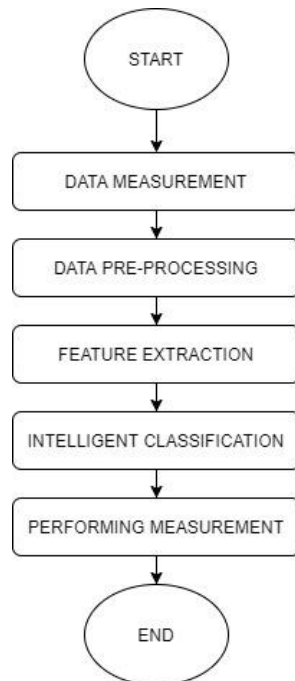
Figure 1. Flowchart for Sewage Treatment Plant Classification.

2.1 E-nose Experimental Setup and Data Measurement

An E-nose device was used to collect the STP odour data. This system embodies chemical sensor collection, odour chamber, E-nose pump and microcontroller. As figure 2 shows, the data collection was conducted and four different points.

Figure 2. Odour collection at four different points.

E-nose was positioned at each of the points to record the raw data. The odour is sucked into the E-nose chamber by a pump located in the upper part of the E-nose, and the sensor array will record the reading of the odor STP collected within the chamber. At each of the points. Five times of experiment was conducted which each of the experiment will record 600 raw data in 3 minutes and 7 seconds. In between the experiment, the sensor array will be neutralised using ethanol. The same process was run at every each of the points. Then the raw data will store on a computer via USB cable.

Table 1. Data measurement table for STP odor profile.

DATA MEASUREMENT	S1	S2	S3	S4
1	DM_{11}	DM_{12}	DM_{13}	DM_{14}
2	DM_{21}	DM_{22}	DM_{23}	DM_{24}
3	DM_{31}	DM_{32}	DM_{33}	DM_{34}
4	DM_{41}	DM_{42}	DM_{43}	DM_{44}
–	–	–	–	–
–	–	–	–	–
–	–	–	–	–
600	DM_{6001}	DM_{6002}	DM_{6003}	DM_{6004}

Table 1 above shows the 600 raw data measurement was collected from 1 experiment conducted at point 1. 5 repeated experiments at four different points were conducted and tabulated the same as Table 1. S1, S2, S3 and S4 represent the four sensors use for the data collection. DM indicate the data measurement of the STP odour sample.

2.3 Data Pre-processing

At each point, all the five repeated experiments were converted into one dataset using Equation 1 mean calculation which $\bar{X}$ stand for the mean value, $\sum_{i=1}^{n} x_i$ is the total of raw value data in a row for five experiments, and n is the total of the experiment conducted at each point. After that, all the collected raw data were normalised using equation (2) which R' stands for the normalised value, R for data raw data measurement and R_{max} is the highest value from each raw. The normalised value was gained after all value in each row was divided by the highest value in own row. Consequently, the normalised value will be 0 until 1 and was the value was tabulated into Table 2

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

$$R' = \frac{R}{R_{max}} \tag{2}$$

Table 2. Normalised value table for STP odour-profile.

Normalized Value	S1	S2	S3	S4
1	NV_{11}	NV_{12}	NV13	NV_{14}
2	NV_{21}	NV_{22}	DM23	NV_{24}
3	NV_{31}	NV_{32}	DM33	NV_{34}
4	NV_{41}	NV_{42}	DM43	NV_{44}
–	–	–	–	–
–	–	–	–	–
–	–	–	–	–
600	NV_{6001}	NV_{6002}	NV_{6003}	NV_{6004}

Table 2 show the normalisation table for the STP sample where the tabulated normalised value consists 600x4 point of S1, S2, S3 and S4 represent the four sensors used respectively. NV indicate the normalised value of the STP odour sample.

2.4 Feature Extraction

Based on the normalised value, all the unique feature was extracted from each of the point. In each of the point, the normalised value will be grouped into ten cases based on the normalised value's mean calculation. Each of the cases consists of 60 normalised value. The cases for each point were tabulated and stored into CBR memory as "stored cases" for the classification process.

Table 3. Clustered table for STP odour-profile.

Case ID	S1	S2	S3	S4
Case1	NV_{11}	NV_{12}	NV_{13}	NV_{14}
Case2	NV_{21}	NV_{22}	DM_{23}	NV_{24}
Case3	NV_{31}	NV_{32}	DM_{33}	NV_{34}
Case4	NV_{41}	NV_{42}	DM_{43}	NV_{44}
–	–	–	–	–
–	–	–	–	–
–	–	–	–	–
Case40	NV_{40_1}	NV_{40_2}	NV_{40_3}	NV_{40_4}

Table 3 shows the clustered table for the STP sample in which the tabulated clustered cases consist of 10 x 4 points where S1, S2, S3 and S4 represent the four sensors used, respectively. NV indicates the normalised value of the STP odour sample with the Case ID represented for each point.

2.5 Intelligent Classification

Cased-Based Reasoning (CBR) was the technique used in this research. CBR was one of the most techniques use in the classification problem. To perform classification, the CBR method uses four cycles: retrieve, reuse, rewrite, and retain. CBR cycle for STP classification shows as in Figure 3.
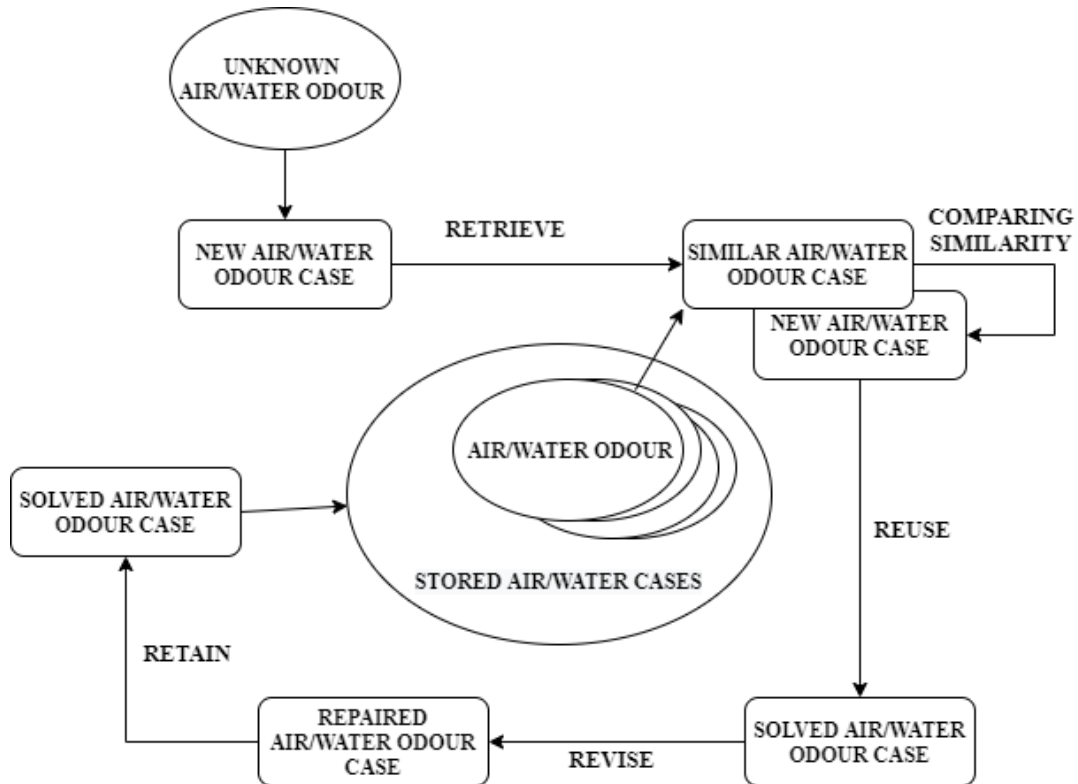
Figure 3. CBR cycle for STP odour-profile classification.

Figure 3 shows the sequence of the CBR cycle used for STP odour-profile classification, which starts with the unknown air and water odour profile. To retrieve the stored cases in the store memory, the unknown odour profile from

each stage went through the retrieval process. CBR was the method learned from the previous case where the unknown air water odour cases was compared with the previous cases stored in the system. Suppose the similarity of the new cases has the highest similarity with the previous case. In that case, the system will use the information in the stored case to provide the decision or answer for the problem statement. Calculation of the similarity percentage was done between new and previous cases using equation (3). One out of the 40 cases was used, while the other 39 cases will be stored cases.

$$Similarity(N,S) = \frac{\sum_{i=1}^{n} f(N_i, S_i) \times w_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

The calculation was conducted with f, similarity function between N and S cases was time with w_i, and index weight for matching feature and divide with new w_i, index weight for matching feature. N represent the new case, and S represent the store case. One out of the 40 cases was used, while the other 39 cases will be stored cases. If the similarity percentage was high, it explained that the new and previous cases could be from the same group.

2.6 Performance measure

The next step was performance measure from the CBR classification result where the result was evaluated confusion matric. In this step, the evaluation was conducted based on the case sensitivity, specificity and accuracy of overall STP classification using Equation 4 until 6.

$$Accuracy = \frac{TP+TN}{TN+FP+FN+TP} \tag{4}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{5}$$

$$Specificity = \frac{TN}{TN+FP} \tag{6}$$

The accuracy was about the clones of the measurement to a specific value where the sum of TN, false positive (FP), false negative (FN), and true positive (TP) was separated by the total of TN, false positive (FP), false negative (FN), and true positive (TP). The sensitivity was to measure the proportional of the actual negative that are correctly where the TP was divided with total of TP and FN. Specificity measured the actual positives that are correctly identified where TN was divided with the total TN and FP.

The TP, TN, FP, and TP concept was gained through the CBR voting process based on the CBR performance table. Example situation for the TP, the predicted says that the case was 'P', which the actual result was 'P1'. The concept also applied to the TN which the predicted case was 'P2', and the actual result was 'P2'. For FP, the concept was the predicted result 'P1', but the actual result 'P2' and the concept also applied to FN.

3. RESULT AND DISCUSSION

3.1 Raw Data Measurement

The data were collected at four different points where at each point, five repeated experiments were performed that resulted in 12,000 data measurements. Thus, all the data measurement that represents for all of the points was tabulated. Figure 4 shows the raw data measure against sensor array for four samples represent fours points at STP. The raw data calculation is shown on the y-axis as a resistance value, while the sensor array is shown on the x-axis. S1, S2, S3 and S4 represent the sensor used in the E-nose. Sensor 3 has the highest sensor reading for all samples, while sensor 2 has the lowest sensor reading for all samples at each stage. The patterns of sensor 3 and sensor 4 are almost identical, as shown in the figure below, even though they contain major differences within each sample, which can be very useful for the classification process. For the data to be more consequential, data pre-processing needs to be performed. Data pre-processing is needed for the data to be more significant.
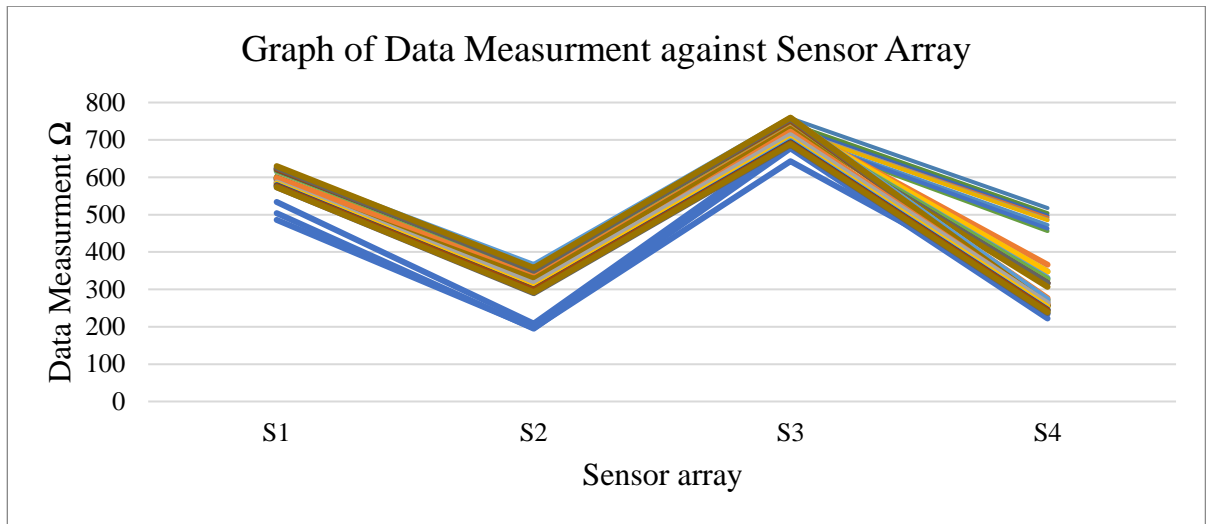
Figure 4. Graph of the data measurment against sensor array.

3.2 Data Pre-processing

All of the data measurements obtained, about 12,000, were standardised by dividing the absolute value in each row by the highest value from its row. The 12000 normalised values will then be divided into four groups, one for each sample. Following that, 3000 normalised values from each category were clustered into ten cases from which the odour feature would be extracted. Figure 5 shows the normalised value against the sensor array where the y-axis represents the normalised resistance value, while the x-axis represents the sensor array.
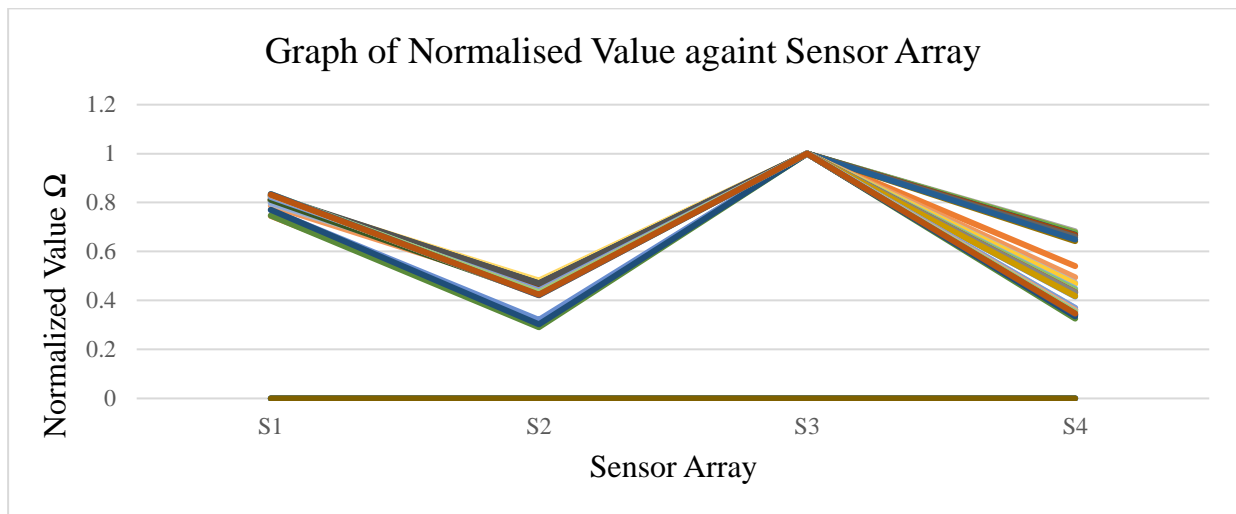
Figure 5. Graph of the normalised value against sensor array.

The CBR case library for the sewage treatment plant odour sample is shown in Table 4. The table contains 40 cases, each of which represents ten cases. The first ten cases (case 1 through case10) reflect the sample at point one. (Case 11 through Case20), (Case 21 through Case30) and (Case 31 through Case40) represent Point2, Point3, and Point4, respectively, for the subsequent ten cases. The four sensors used in the E-nose are designated by the letters S1, S2, S3, and S4. These cases will be added to the CBR memory to conduct the classification process as "stored cases".

Table 4. CBR case library for the Sewage Treatment Plant (STP).

CASE ID	S1	S2	S3	S4
Case1	0.7458	0.3154	1.0000	0.5405
Case2	0.8044	0.4335	1.0000	0.6826
Case3	0.8124	0.4371	1.0000	0.6776
Case4	0.8096	0.4363	1.0000	0.6765

Case5	0.8060	0.4549	1.0000	0.6774
Case6	0.8109	0.4486	1.0000	0.6684
Case7	0.8100	0.4471	1.0000	0.6676
Case8	0.8090	0.4398	1.0000	0.6599
Case9	0.8082	0.4311	1.0000	0.6429
Case10	0.8086	0.4371	1.0000	0.6494
Case11	0.7694	0.3209	1.0000	0.4160
Case12	0.7879	0.4392	1.0000	0.4943
Case13	0.7959	0.4589	1.0000	0.4719
Case14	0.8075	0.4709	1.0000	0.4677
Case15	0.8121	0.4543	1.0000	0.4489
Case16	0.8108	0.4440	1.0000	0.4438
Case17	0.8125	0.4523	1.0000	0.4328
Case18	0.8130	0.4586	1.0000	0.4336
Case19	0.8155	0.4537	1.0000	0.4324
Case20	0.8170	0.4487	1.0000	0.4181
Case21	0.7472	0.2907	1.0000	0.3258
Case22	0.7993	0.4614	1.0000	0.3696
Case23	0.8153	0.4772	1.0000	0.3610
Case24	0.8186	0.4770	1.0000	0.3580
Case25	0.8231	0.4815	1.0000	0.3596
Case26	0.8233	0.4676	1.0000	0.3411
Case27	0.8241	0.4702	1.0000	0.3403
Case28	0.8264	0.4704	1.0000	0.3420
Case29	0.8275	0.4626	1.0000	0.3378
Case30	0.8293	0.4702	1.0000	0.3412
Case31	0.7707	0.3028	1.0000	0.3356
Case32	0.8106	0.4258	1.0000	0.3623
Case33	0.8228	0.4412	1.0000	0.3688
Case34	0.8305	0.4338	1.0000	0.3605
Case35	0.8315	0.4231	1.0000	0.3539
Case36	0.8333	0.4215	1.0000	0.3500
Case37	0.8338	0.4313	1.0000	0.3536
Case38	0.8348	0.4315	1.0000	0.3532
Case39	0.8350	0.4219	1.0000	0.3446
Case40	0.8322	0.4238	1.0000	0.3468

The highest normalised value in this table is in column S3, made up entirely of the letter' 1'. Column 3 has the same value for each case because the raw data were divided with the highest value in each row. The different values from each sensor are different because each set-up sensor has a different sensitivity. Sensor 3 shows the high sensitivity on the odour of STP.

3.3 CBR Voting

Table 5 displays the results of CBR voting for odour sample classification at a Sewage Treatment Plant (STP). Case ID, real, K=1, K=2, and K=3 are all mentioned in the table. The actual column was determined based on the point of the data collection. Case1-case10, case11-case20, case21-case30 and case31-case40 represent Point1, Point2, Point3 and Point4 respectively.

The voting procedure was carried out by crossing 40X40 similarity matrices in the same category and other groups. The highest value (K=1), second highest value (K=2), and third highest value (K=3) signify the highest value, second

highest value, and third highest value, respectively. In each row, all K=1, K=2, and K=3 were voted in the same category.

Table 5. CBR voting result.

CASE ID	ACTUAL	VOTING, K=1	VOTING, K=2	VOTING, K=3
CASE1	POINT1	P2	P2	P3
CASE2	POINT1	P1	P1	P1
CASE3	POINT1	P1	P1	P2
CASE4	POINT1	P1	P2	P2
CASE5	POINT1	P2	P2	P2
CASE6	POINT1	P1	P1	P1
CASE7	POINT1	P1	P1	P1
CASE8	POINT1	P1	P1	P1
CASE9	POINT1	P1	P1	P1
CASE10	POINT1	P1	P1	P1
CASE11	POINT2	P4	P3	P1
CASE12	POINT2	P2	P2	P2
CASE13	POINT2	P2	P2	P2
CASE14	POINT2	P2	P2	P2
CASE15	POINT2	P2	P2	P2
CASE16	POINT2	P2	P2	P2
CASE17	POINT2	P2	P2	P2
CASE18	POINT2	P2	P2	P2
CASE19	POINT2	P2	P2	P2
CASE20	POINT2	P2	P2	P2
CASE21	POINT3	P4	P2	P4
CASE22	POINT3	P3	P4	P3
CASE23	POINT3	P3	P3	P3
CASE24	POINT3	P3	P3	P3
CASE25	POINT3	P3	P3	P3
CASE26	POINT3	P3	P3	P3
CASE27	POINT3	P3	P3	P3
CASE28	POINT3	P3	P3	P3
CASE29	POINT3	P3	P3	P3
CASE30	POINT3	P3	P3	P3
CASE31	POINT 4	P3	P2	P4
CASE32	POINT 4	P4	P4	P4
CASE33	POINT 4	P4	P4	P4
CASE34	POINT 4	P4	P4	P4
CASE35	POINT 4	P4	P4	P4
CASE36	POINT 4	P4	P4	P4
CASE37	POINT 4	P4	P4	P4
CASE38	POINT 4	P4	P4	P4
CASE39	POINT 4	P4	P4	P4
CASE40	POINT 4	P4	P4	P4

3.4 CBR Performance Measure

The confusion matrix for CBR voting results is shown in Table 6. For every group, K=1, K=2 and K=3 was voted to be in their group. There are a total of 40 cases in this method. Each of the points consists of 10 cases. In the confusion matrix table, there is the actual case and predicted case. The actual case is the sample's real case, while the expected case is based on the voting results in table 2. In table 3, 21 cases, 27cases, 26 cases and 28 cases were predicted to be in groups P1, P2, P3 and P4, respectively. As a result, the total true positive for each category is represented by the value of the expected events, and the total true positive for the entire sample is 102.

Table 6. Confusion Matrix for CBR voting result.

		Predicted Result			
		P1	P2	P3	P4
Actual Result	P1	21	8	1	0
	P2	1	27	1	1
	P3	0	1	26	3
	P4	0	1	1	28

Table 7. CBR voting result.

	P1	P2	P3	P4
TOTAL CASE	30	30	30	30
TPR = TP/(TP+FN)	0.7	0.9	0.87	0.93
TNR = TN/(TN+FP)	0.99	0.89	0.97	0.96
FPR = FP/(FP+TN)	0.01	0.11	0.03	0.04
FNR = FN/(FN+TP)	0.30	0.10	0.13	0.07
Average TPR	85%			
Average TNR	95%			
Average FPR	5%			
Average FNR	15%			
Overall accuracy = (TP+TN)/P+N)	85%			
Overall sensitivity = TP/(TP+FN)	85%			
Overall specificity = TN/(TN+FP)	95%			

Table 7 show the overall accuracy, sensitivity and specificity of the STP classification. For the accuracy, the evaluation value was 85%. Meanwhile, the evaluation value was 85% and 95% for the sensitivity and specificity, respectively.

4. CONCLUSION

In conclusion, this study shows that at four different points, the STP releases a different kind of aroma which may be influenced by the air surrounding and the presence of wind where E-Nose was able to collect the odours data of the STP. The research has shown that the odor-profile for the surrounding air at the STP was successfully established as the data collected at the four different points was successfully normalised. The feature of the sample was extracted. The extracted feature become the attribute input and is successfully stored in the CBR database to undergoes the retrieval phase. Four different samples were successfully classified to 85% accuracy and sensitivity and 95% specificity.

5. ACKNOWLEDGEMENT

REFERENCES

[1]   G. Dotro, P. Molle, J. Nivala, J. Puigagut, and O. Stein, *Biological Waste Water Treatment Series*, vol. 7. 2017.

[2]   S. S. Pattanshetti and S. M. Gawande, "Characteristics Study of Wastewater in Gadhinglaj," no. June, pp. 10–13, 2017.

[3]   Y. Y. Choi *et al.*, "Characteristics and biodegradability of wastewater organic matter in municipal wastewater treatment plants collecting domestic wastewater and industrial discharge," *Water (Switzerland)*, vol. 9, no. 6, 2017.

[4]   Asaduzzaman, "Dhaka University of Engineering & Technology, Gazipur," no. April, p. 2017, 2017.

[5]   J. Yan *et al.*, "Electronic nose feature extraction methods: A review," *Sensors (Switzerland)*. 2015.

[6]   J. He, L. Xu, P. Wang, and Q. Wang, "A high precise E-nose for daily indoor air quality monitoring in living environment," *Integr. VLSI J.*, 2017.

[7]   S. Deshmukh, R. Bandyopadhyay, N. Bhattacharyya, R. A. Pandey, and A. Jana, "Application of electronic nose for industrial odors and gaseous emissions measurement and monitoring - An overview," *Talanta*, vol. 144, pp. 329–340, 2015.

[8]   L. Eusebio, L. Capelli, and S. Sironi, "Electronic nose testing procedure for the definition of minimum performance requirements for environmental odor monitoring," *Sensors (Switzerland)*, vol. 16, no. 9, 2016.

[9]   S. M. Daud, M. S. Najib, N. Zahed, M. H. M. Jusof, M. F. C. Jusoh, and M. I. I. N. Hassim, "Classification of lubricant oil adulteration level using case-based reasoning," *J. Fundam. Appl. Sci.*, vol. 9, no. 4S, p. 256, 2018.

[10]   "Classification of Ammonia in water for Oil and Gas Industry using Case Based Reasoning," vol. 1, no. 7, p. 2462, 2016.

[11]   S. G. K. Patro and K. K. Sahu, "Normalisation: A Pre-processing Stage," *Iarjset*, no. March, pp. 20–22, 2015.

[12]   X. Hu, B. Xia, M. Skitmore, and Q. Chen, "The application of case-based reasoning in construction management research: An overview," *Autom. Constr.*, 2016.

[13]   S. National, "Case-Based Reasoning Case-Based Reasoning," no. September 2001, 2016.

[14]   M. S. Najib *et al.*, "Fish Quality Study Using Odor-Profile Case-Based Reasoning (CBR) Classification Technique," vol. 11, no. 10, pp. 6691–6696, 2016.

[15]   Y. Su, S. Yang, K. Liu, K. Hua, and Q. Yao, "Developing a case-based reasoning model for safety accident pre-control and decision making in the construction industry," *Int. J. Environ. Res. Public Health*, vol. 16, no. 9, 2019.