Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

Exploring Black Tea Classification Through The Use of QCM-Based Electronic Nose

[1]Sumana Banerjee*, [2]Prolay Sharma

[1]Assistant professor in Department of Biomedical Engineering, JIS College of Engineering (Orcid ID: https://orcid.org/0009-0007-4641-6078)
[2]Department of Instrumentation and Electronics, Jadavpur university (Orcid ID: https://orcid.org/0000-0002-2617-0126)

Email address: sumana28banerjee@gmail.com (author name1), prolay28sharma@gmail.com (author name2)
**Corresponding Author* sumana28banerjee@gmail.com,

Abstract: The traditional method for assessing tea quality relies on human sensory evaluation by "Tea Tasters." However, this approach is highly subjective and its reliability is questionable. Consequently, various instrumental setups have been explored in recent times. One such device is the Electronic Nose, which has been employed for tea quality assessment. This electronic nose features an array of five AT-cut 10 MHz quartz crystal microbalance (QCM) sensors. These sensors have been exposed to the aroma of different types of black tea, such as cut-tear-curl (CTC) and orthodox, with responses monitored online through a data acquisition system. The collected data has been analyzed using principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA). A comparative study of different clustering algorithms has been conducted based on cluster validity indices, and the BPMLP classifier has been utilized with a 10-fold cross-validation technique for data classification.

Keywords: QCM; PCA; LDA; ICA; BPMLP; Dunn's Index.

I. Introduction

The aroma of orthodox tea plays a crucial role in determining its quality. Key chemical compounds responsible for tea aroma are outlined in [1] and listed in Table I. Assessing the aroma quality of tea and its classification holds significant commercial importance. Traditionally, this assessment is conducted by a human sensory panel called "Tea Tasters." However, this method is susceptible to various errors due to factors such as fatigue, individual differences, infections, and adverse mental states. Additionally, scoring can vary from one taster to another, even for the same tea sample.

TABLE I. Biochemical Compounds In Black Tea Responsible For Flavour

Compounds	Flavour
Linalool, Linalool Oxide	Sweet
Geraniol, Phenylacetaldehyde	Floral
Nerolidol, Benzaldehyde, Phenyl ethanol	Fruity
Trans-2-Hexenal, n-Hexanal, Cis-3-Hexenol, Grassy, b-Ionone	Fresh flavour

An electronic nose offers a solution to these challenges. It serves as a highly useful device for fast, reliable, non-invasive, continuous, and real-time monitoring of tea aroma, aiding in quality evaluation. In this study, a QCM-based electronic nose has been developed to primarily differentiate between two broad varieties of tea, namely orthodox and CTC. The clustering information has been visualized using various algorithms, and the classification task has been performed by a feed-forward backpropagation neural network. The classifier's performance has been validated using a 10-fold cross-validation method. The results are compiled to demonstrate the utility of the e-nose system, along with data processing units, as detailed in this paper.

Copyright @ JCSEIR

Published by JCSEIR

Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

II. EXPERIMENTAL

A. The QCM Sensor

A Quartz Crystal Microbalance (QCM) determines the mass per unit area through changes in the frequency of a quartz crystal resonator. The resonance is affected by any alteration in the mass at the surface of the resonator. By applying certain simplifications and assumptions, the frequency change can be measured and correlated to the mass change using Sauerbrey's Equation [2]:

$$\Delta f = \frac{-f_0^2}{A\sqrt{\rho_q \mu_q}} \Delta m$$

f_0 – Resonant Frequency, Δf – Change in Frequency, Δm – Change in Mass, A – Piezoelectrically Active Crystal Area, ρ_q – Density of Quartz, μ_q – Shear Modulus of Quartz.

It's clear that any mass deposited on the resonator surface results in a decrease in the oscillation frequency from its initial value. This principle has been effectively utilized in both gas sensing and liquid phase applications in recent times. This concept has also inspired the development of a QCM-based Electronic Nose system for evaluating the quality of black tea.

B. The Electronic Nose Setup

An Electronic Nose setup comprising QCM sensors has been designed for the evaluation and classification of black tea quality. The schematic diagram of the entire setup is illustrated in Figure 1. Adsorbent materials suitable for three key volatile compounds of black tea aroma (Linalool, Geraniol, and Trans-2-Hexenal) have been selected. Solutions of these compounds with appropriate solvents (Table II) have been prepared and deposited on the quartz crystal blanks. The sensors have been placed in a sensor chamber. A suction pump connected to the sensor chamber draws in the tea aroma for sampling and fresh air (as selected by the selector switch) for purging. The oscillation frequency of the crystals has been generated by IC 8284, and the sensor response has been monitored online using a PCI-6602; NI DAQ card.

C. Experimental Conditions

The experimental conditions are as follows:

- Volume of the sensor chamber: 490 ml
- Air flow rate through the suction pump: 4 L/min
- Amount of dry tea sample used: 15 gm
- Sampling duration: 30 seconds
- Purging time: 3 minutes

D. Samples

Two major types of black tea are available in the market: CTC and Orthodox. Samples of these types have been collected as follows:

1. CTC with a market price of INR 300 (denoted as CTC 300)
2. CTC with a market price of INR 500 (denoted as CTC 500)
3. Orthodox with a market price of INR 500 (denoted as Orthodox 500)
4. Orthodox with a market price of INR 1000 (denoted as Orthodox 1000)

Each type has been subjected to experimentation, and 20 experimental values for each type have been considered for analysis purposes.

TABLE II. List Of Coating Materials, Their Solvents And Concentrations Used For Preparation Of The Sensors

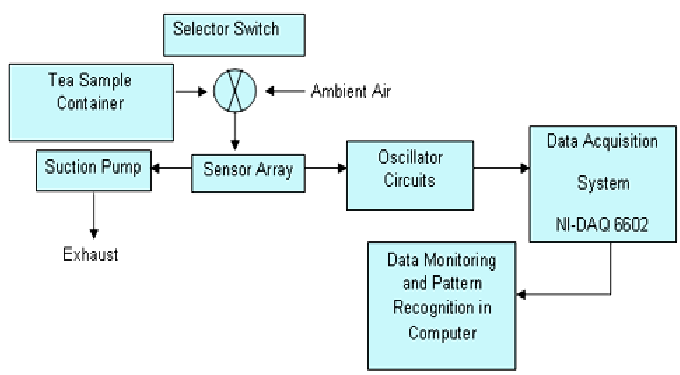

Figure 1. Schematic Diagram of the QCM based Electronic Nose Setup

III. DATA ANALYSIS

The data obtained from experiments with black tea samples using the E-Nose setup was five-dimensional. Since

Copyright @ JCSEIR

JCSEIR

Journal of Computer Science and Engineering in Innovations and Research

RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

high-dimensional data cannot be graphically represented, identifying patterns in raw data becomes challenging. Therefore, various algorithms have been employed to transform the data, enhancing clustering and reducing dimensions to facilitate easier analysis. To assess the quality of class separability, Dunn's Index—a Cluster Validity Index—has been computed for each case. Additionally, a BPMLP network has been utilized for data classification.

IV. Results And Discussion

Data has been collected for four types of black tea samples. Each sample was presented five times and exposed to all five sensors for four repeated sniffs. This resulted in a dataset of 80 entries (4 tea samples x 5 presentations x 4 repeated samplings) x 5 sensors.

Initially, the data was transformed using PCA, which revealed well-defined clusters in the obtained data. It was found that nearly 98% of the information was captured by the first two principal components. Given the data classification capability of LDA, it was applied to the raw data, resulting in immediate improvements in clustering, as shown in Figure 3 and Table 5. The fact that each relevant constituent compound of black tea produces its own response in each sensor led to the application of ICA on the raw data to extract the original source data. This approach resulted in better separation within the dataset. For dimensionality reduction, LDA was applied to the ICA-transformed data, further improving clustering as evidenced by the corresponding cluster validity indices (Table 5). The combination of ICA and LDA for clustering yielded the best results based on Dunn's Indices.

A. *Principal Component Analysis (PCA)*

PCA is a mathematical technique that transforms a set of observations into a set of linearly uncorrelated values known as Principal Components, using orthogonal transformation. This transformation is defined so that the first principal component has the highest possible variance, while each succeeding component has the highest variance possible under the constraint of being orthogonal to the previous ones. PCA helps identify patterns in data and express similarities and differences effectively [3].

In this study, 80 sample values (20 from each of the four types of tea) were transformed using PCA. The result, shown in Figure 2, indicates that over 97% of the significant information is captured by the first two principal components.

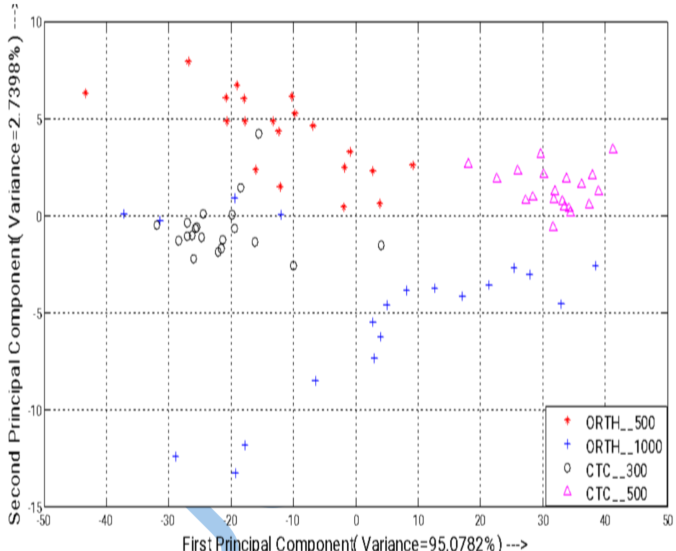

Figure 2. Result as obtained after transformation of the raw data by PCA

B. *Linear Discriminant Analysis (LDA)*

While PCA focuses on dimensionality reduction by preserving as much variance in the high-dimensional space as possible, LDA aims to reduce dimensionality while maintaining class discriminatory information. PCA performs feature classification, whereas LDA performs data classification. Unlike PCA, LDA does not alter the location and shape of the dataset but strives to enhance class separability and define decision regions between the given classes [4].

The same sample set used previously has been transformed using LDA, with a class-independent transformation applied. The result is depicted in Figure 3.

C. *Independent Component Analysis (ICA)*

In a simplified experimental assumption, it was believed that each specific compound in the tea vapor would affect only a single sensor within the five-sensor array. However, this is not the case in reality. This necessitated the extraction of individual independent sources from the raw data we obtained.

D. *ICA in Conjunction with LDA*

Although applying ICA has improved separability (see Table III), the extracted data remains high-dimensional (five dimensions). To reduce the dimensionality, the extracted data has been further transformed using LDA. This additional

Published by

Copyright @ JCSEIR

JCSEIR

Journal of Computer Science and Engineering in Innovations and Research

RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

transformation has led to even better clustering, as shown in Figure 4.

E. Dunn's Index

Validity of clusters is worth being evaluated after the algorithms are applied, to find whether the result is appreciable or not. Different cluster validity algorithms are available for this purpose.

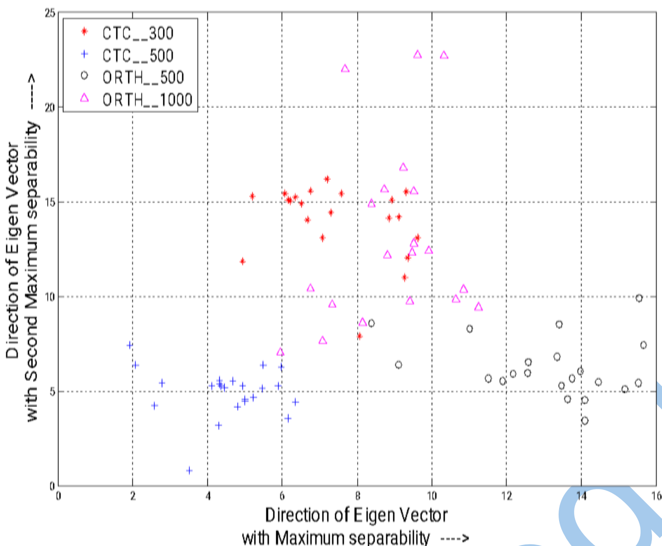

Figure 3. Result as obtained after transformation of the raw data by LDA

$$D = \min_{1 \le i \le n} \{ \min_{1 \le j \le n, i \ne j} \{ \frac{d(i, j)}{\max_{1 \le k \le n} d'(k)} \} \}$$

Independent Component Analysis (ICA) is an effective computational method used to separate a multivariate signal into additive subcomponents, assuming the source signals are mutually independent. It is essentially a specialized form of blind source separation. ICA identifies independent components by maximizing the statistical independence of the estimated components. In this paper the method of Maximisation of Non-Gaussianity has been adopted for the necessary data analysis. The Central Limit Theorem of Probability Theory says that the distribution of a sum of independent random variables tends toward a Gaussian distribution under certain conditions. If a data vector x is assumed to be a mixture of independent components as in the ICA model, it can be represented as x = As where s is the source vector and A is the mixing matrix which contains the mixing coefficients. Evidently the source s can be extracted as $s = A^{-1}x$. The coefficient matrix and the sources are estimated by numerous iterations keeping in mind that the sum of independent random variables is more Gaussian than the original variables and it becomes least Gaussian when the estimated signal equals the original source [5]. In this study, the targeted chemical constituents of black tea are considered the source, while the sensors are regarded as the mixers.

In this paper Dunn's Index has been adopted. It aims to identify dense and well separated clusters. The index is defined as the ratio between the minimal inter-cluster distance to maximal intra-cluster distance. It can be depicted by the following expression:

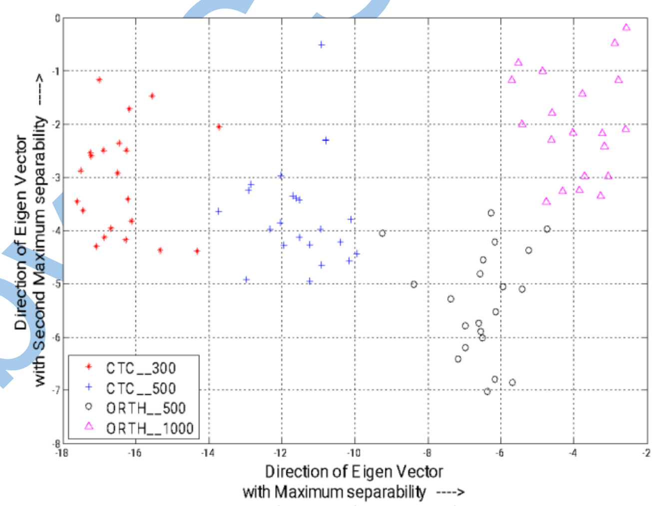

Figure 4. Result as obtained after transformation of the raw data by ICA in conjunction with LDA

where d(i,j) represents the distance between clusters i and j and d'(k) measures the intra-cluster distance of cluster k. The inter-cluster distances have been measured as the distance of centroids of the corresponding clusters. As internal criterion seeks clusters with high intra-cluster similarity and low inter-cluster similarity, algorithms that produce clusters with high Dunn's Indices are considered to be better and desirable. neurons in each and one output layer with a single neuron. The neurons of the hidden layer are assigned with the transfer function 'Tan-Sigmoid' and the network has been trained with the Levenberg-Marquardt back propagation model.

Published by JCSEIR

JCSEIR

Journal of Computer Science and Engineering in Innovations and Research

RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

A 10-fold Cross Validation method has been used for the purpose of training and testing of the network. The results obtained are shown in Table IV.

F. BPMLP Model

A BPMLP NETWORK HAS BEEN USED FOR CLASSIFICATION OF THE DATA AVAILABLE. THE NETWORK CONSISTS OF THREE SENSORY UNITS - ONE INPUT LAYER WITH FIVE NEURONS, TWO HIDDEN LAYERS WITH FIVE neurons in each and one output layer with a single neuron. The neurons of the hidden layer are assigned with the transfer function 'Tan-Sigmoid' and the network has been trained with the Levenberg-Marquardt back propagation model. A 10-fold Cross Validation method has been used for the purpose of training and testing of the network. The results obtained are shown in Table IV.

I. CONCLUSION

As shown in the visualization plots, the developed sensors can produce distinct clusters separating orthodox tea varieties from CTC tea samples. The clustering algorithms used include PCA, LDA, and ICA. The effectiveness of cluster formation has been assessed using Dunn's indices, which demonstrated improvement when transitioning from PCA to LDA, ICA, and a combined plot of ICA and LDA. Data classification has been performed using back-propagation neural networks, achieving an average classification rate of 92.5% with 10-fold cross-validation. This indicates that the developed QCM sensors can effectively differentiate among various types of tea for further quality analysis.

TABLE III. DUNN'S INDICES FOR CLUSTERS OF RAW DATA AND THOSE OBTAINED FROM TRANSFORMATIONS OF DIFFERENT ALGORITHMS

Dataset	Dunn's Index (approximate values)
Raw data obtained from QCM sensor response	0.0873
Data transformed by PCA	0.2037
Data transformed by LDA	0.2126
Data transformed by ICA	0.5030
Data transformed by ICA in conjunction with LDA	0.5338

TABLE IV. RESULTS FOR 10 FOLD VALIDATION TEST FOR DATA PATTERNED FROM E-NOSE SYSTEM

Fold	Number of Data: Train/Test	Data Misclassified	Classification rate (%)
1	72/8	2	75
2	72/8	2	75
3	72/8	2	75
4	72/8	0	100
5	72/8	0	100
6	72/8	0	100
7	72/8	0	100
8	72/8	0	100
9	72/8	0	100
10	72/8	0	100
Total number of data misclassified		6	
Average Classification Rate		92.50%	
Standard deviation		12.07%	

REFERENCES

[1] P.K. Mahanta and R. Singh, "Flavour components of assam and darjeeling teas in relation to agropractices and processing," Proc. Of the international conference on R & D in tea, pp. 148-153, 1990.

[2] G. Sauerbrey, "The use of quartz crystal oscillators for weighing thin layer and for microweighing", Z. Phys, vol. 155, pp.206-222, 1959.

[3] Lindsay I. Smith, "A tutorial on Principal Component Analysis", Lindsay, February, 2002.

[4] S. Balakrishnama, , A. Ganapathiraju, "Linear Discriminant Analysis – A Brief Tutorial".

A. Hyvarinen, J. Karhunen and E. Oja, "Independent Component Analysis". John Wiley & Sons, New York: (Chapter 8). 2001.

Authors Profile

Prolay Sharma received his Ph.D. degree in the year 2016 from Jadavpur University, Kolkata, India. He is an Associate Professor in the Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata, India. His research interests include the electronic nose, quartz crystal microbalance sensor, pattern recognition, and signal processing.

Sumana Banerjee received his M.E. degree in Biomedical Engineering from the School of Bio-Science and Engineering, Jadavpur University, Kolkata, India, in 2010. He is currently an Assistant Professor with the Department of Biomedical Engineering, JIS College of Engineering, Kalyani, Nadia. His research interests include sensor development, pattern recognition and image processing.

Published by JCSEIR

Copyright @ JCSEIR