İSTANBUL UNIVERSITY
CERRAHPAŞA

ORIGINAL ARTICLE
Electrica 2022; 22(2): 237-243

Use of Hybrid Clustering and Scattering Parameters for Liquid Classification

Ebru Efeoğlu[1], Gürkan Tuna[2]

[1]Department of Software Engineering, Kütahya Dumlupınar University, Kütahya, Turkey
[2]Department of Computer Programming, Trakya University, Edirne, Turkey

Cite this article as: E. Efeoğlu and G. Tuna, "Use of hybrid clustering and scattering parameters for liquid classification," *Electrica.*, 22(2), 237-243, 2022.

ABSTRACT

With the advancement of technology, the use of machine learning techniques has increased. The need for the prevention of terrorist attacks has brought upon the use of machine learning techniques to explosive detection. Flammable liquids such as alcohol are easily available and widely used in various terrorist attacks. In this study, a new microwave measurement system is developed and a hybrid clustering approach is proposed to classify liquids. With the proposed measurement system, the reflection coefficient (S_{11} parameter) of liquids in bottles is measured at room temperature and these measurements are used as inputs by the proposed clustering algorithm. The results obtained using the proposed clustering algorithm are compared with the results obtained using a set of well-known clustering algorithms, that is, K-means, hierarchical clustering, farthest first, and fuzzy C-means, in order to make a fair comparison. The results show that the proposed clustering algorithm provides 100% accuracy and is superior to the well-known algorithms used in this study. The results will enable us to manufacture a low-cost liquid scanner for railway stations and shopping malls as well as small airports. The proposed liquid scanner's design was completed, and the manufacturing phase has been started.
***Index Terms*—**Accuracy, explosive liquids, flammable liquids, hybrid clustering algorithm, liquid classification.

I. INTRODUCTION

Machine learning makes inferences from a given data set using mathematical and statistical methods and makes predictions about the unknown with these inferences. For instance, a machine learning application can recognize images and classify them according to the items they contain, or when you listen to music, it may suggest different tracks similar to the music you are listening to. There are two different learning techniques used in machine learning: supervised learning, which is the process of learning from labeled observations, and unsupervised learning, which is the process of learning from unlabeled observations. While in supervised learning example inputs and their desired outputs are presented so that the system learns a general rule that maps inputs to outputs can be learnt, in unsupervised learning, no labels are given to the system leaving it on its own to find structure in its input. Unsupervised learning has been heavily investigated in recent years. Diagnosis of thyroid cancer [1], K-means clustering approach for short-term wind energy prediction [2], evaluation of unsupervised clustering methods in hyperspectral image data sets [3], round randomized learning vector quantization method for brain tumor imaging using hybridized K-average algorithm [4], estimating the distance between centers in data clustering problems [4], classification of Landsat 8 images, and learning of vector quantization using Kosaten's self-regulating maps [5] are examples of the literature.

Liquids can be classified for different purposes. By using machine learning techniques, the heat discharge rate of liquid fuels such as diesel and gasoline was analyzed [6]. The determination of the content of methanol, ethanol, and ethylene glycol in coconut wine in the differentiation of gasoline brands [7] was made by means of machine learning techniques [8]. According to the correlation coefficient and cluster analysis, machine learning techniques were also used in the optimization of an electronic nose sensor array [9] for tea flavor detection.

In the classification of liquids, there are expensive measurement methods that classify the liquids according to scattering spectroscopy, such as Raman spectroscopy, Fourier transform infrared (FTIR) spectroscopy, x-ray, and Electronic nose and Electronic tongue. Raman spectroscopy has many advantages such as wide detection range, sharp spectral peaks, and high resolution, and it was used to

Corresponding author: Ebru Efeoğlu
E-mail: ebru.efeoglu@dpu.edu.tr

Received: November 23, 2021

Revised: January 31, 2022

Accepted: February 7, 2022

Available Online Date: March 20, 2022

DOI: 10.54614/electrica.2022.21160

classify gasoline by distributor and to identify common additives such as methanol and toluene [10]. Using portable Raman spectroscopy, the screening and determination of methanol content in ethanol-based products was carried out [11]. The FTIR spectroscopy method was used to estimate the properties and amount of additives in gasoline [12]. Using the near-infrared spectroscopy method, the detection and prediction of the Premium 91 gasoline and the Super premium 95 gasoline mixture was made [13]. An electronic nose that utilizes machine learning techniques was proposed for the detection of water, methanol, and ethanol mixtures [14]. It was shown that flammable liquids could be detected using the x-ray method and spectral droplet analysis methods [15,16]. Clustering algorithms were also used in intelligent bioelectronic tongues and electronic noses for food and beverage control [17,18].

Microwave measurement method that relies on K-means algorithm and support vector machines was successfully used for different purposes including stroke localization and classification [19]. K-means algorithm was also used with microwave scatterometer and radiometer data to classify ice in the Arctic sea [20,21]. The microwave measurement method along with machine learning techniques was also used to classify kidney stones according to their dielectric properties [22] and to detect lesions in breast images [23]. It is known that prolonged consumption of alcoholic drinks causes various diseases. For instance, methanol poisoning causes blindness, organ failure, and even death when detected too late [24]. Methanol consumption is extremely dangerous, if left untreated, there may be significant morbidity and mortality [25]. Various types of alcohol are easily available and can cause other hazards such as being used in making liquid explosives.

In recent years, the use of microwave methods to ensure human health and safety has brought upon new approaches to the detection of explosive liquids. In this study, with the goal of classifying hazardous liquids, a measurement system capable of remote and non-contact measurement based on the reflection coefficient measurements of liquids was developed and a hybrid clustering algorithm for the classification of liquids was proposed. The paper is as follows. In the following section, the measurement system is presented and the methodology used in this paper is explained. In the third section, the algorithms used in this study are reviewed. Finally, the fourth section concludes the paper.

II. EXPERIMENTAL DESIGN FOR THE COLLECTION AND USE OF MICROWAVE MEASUREMENT DATA

Vector network analyzer shown in Fig. 1(a) is a device that can send and record signals in a certain frequency range. It can be used in either two ports or one-port mode. Transmission coefficient S_{21} measurement is used for two-port uses, and reflection coefficient S_{11} measurement is used for single-port uses. In this study, one-port mode was preferred and S_{11} parameter was measured. Liquid measurements were made using a square antenna connected to a single port with a 50 Ω SubMiniature version A feed probe shown in Fig. 1. The antenna's resonance frequency is 1.5 GHz. The measurement can be done quickly by bringing the antenna 5–6 mm closer to the liquid container, without the need for the antenna to touch the liquids in the container. It should be noted that the distance between the antenna and the liquid was the same in all measurements. In the measurements made in this study, the liquids were in pet bottles and at room temperature. Measurements were taken by holding the antenna at a distance of 5 mm from the bottle without opening the bottle cap.

A. Data Set

In this study, measurements of 15 liquids, a total of 4 hazardous and a total of 11 non-hazardous, were used. The liquids used are given in Table I, and their *S* parameter measurements are given in Fig. 2. The *S* parameter measurements were collected at 41 points in the frequency range of 1.42–1.54 GHz.

III. REVIEW OF THE ALGORITHMS USED IN THIS STUDY

A. K-Means

K-means algorithm is a simple and effective clustering method and is widely used. It divides a data set into separate sets. The aim is to ensure that clusters obtained at the end of the partitioning process have a maximum in-cluster similarity and a minimum between–cluster similarity. The value of *k*, which indicates the number of clusters sought, should be known at the beginning. *K* also reports the number of cluster centers. *K* cluster centers are randomly selected initially. The remaining objects are assigned to the cluster where the nearest cluster center is located. Distance criteria are used when assigning. The most commonly used distance criterion is Euclidean distance and it is calculated using (1). An object belongs to the cluster which is close to that cluster. The center points are updated according to the placed objects, and again the objects are assigned to the cluster with the cluster center close to them. This process continues until the new center points of the clusters and the previous center points are the same:

$$distance(x,y)=\left\{\sum_{i=1}^{n}(x_i-y_i)\right\}^{1/2} \tag{1}$$

B. Farthest First

It is a type of K-means algorithm and uses a greedy approach. This algorithm selects center points by maximizing the distances

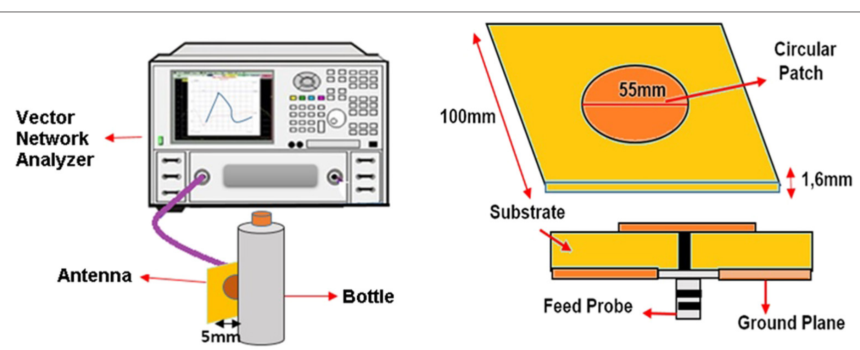


Fig. 1. Measurement system and antenna.

TABLE I. LIQUIDS USED IN THIS STUDY

	Number	Liquids
Hazardous	1	Ethanol
	2	Methanol
	3	1-Propanol
	4	2-Propanol
Non-hazardous (safe)	5	Cola
	6	Liquid soap
	7	Shampoo
	8	Water
	9	Milk
	10	Cream
	11	Shower gel
	12	Buttermilk
	13	İce-tea
	14	Apricot juice
	15	Screen cleaning liquid

and then sets the objects according to these selected center points. The selection of the first center points is made randomly. The second center point is selected with the greedy approach, which is the farthest distance from the first center point selected. Each remaining center is determined by selecting the farthest point from the previously selected set of centers with the greedy approach, and the remaining points are added to the nearest center set.

C. Fuzzy C-means

Fuzzy C-means algorithm, one of the division clustering techniques, is a goal function-based method. Clustering is ended when the goal function converges and reaches a predetermined value. The algorithm tries to minimize the objective function given in (2):

$$J_n = \sum_{i=1}^{N}\sum_{j=1}^{c} M_i^n x_i - c_j^{2}, \quad 1 \le n < \infty \tag{2}$$

where M represents the membership matrix and c represents the cluster centers. Each data has a membership value ranging from 0 to 1 for each of the sets. Clustering is done according to this membership value. If the member of the cluster belonging to the cluster is large, the object is considered closer to the cluster center and assigned to that cluster. Membership values of the data for all classes must be "1." Initially, the membership matrix is randomly assigned. Then, the center vectors are calculated. The membership matrix is updated according to the calculated cluster centers. Clustering is completed when the difference between the membership matrices reaches the value of the determined convergence criterion.

D. Hierarchical Clustering

Hierarchical algorithms can be examined under two headings: AGNES (Agglomerative Nesting), namely Agglomerative Clustering, and DIANA (Divisive Analysis), namely divisive hierarchical clustering. In AGNES, each of the data is considered to be a cluster at the initial stage. Until there are no more data to be clustered, the two most clustered clusters are combined. In DIANA, divisions into clusters from top to bottom until a single data remains are performed. There are different approaches used to calculate the distance between the two sets. Single link is based on the shortest distance basis, complete link takes into account the maximum distance between pairs of clusters, centroid considers the distance between the center points of two clusters, average link takes the average of all distances between two clusters, aiming to minimize the total intra-cluster variance, ward link method calculates the sum of error squares using intra-cluster squared deviations, and adjusted complete link method

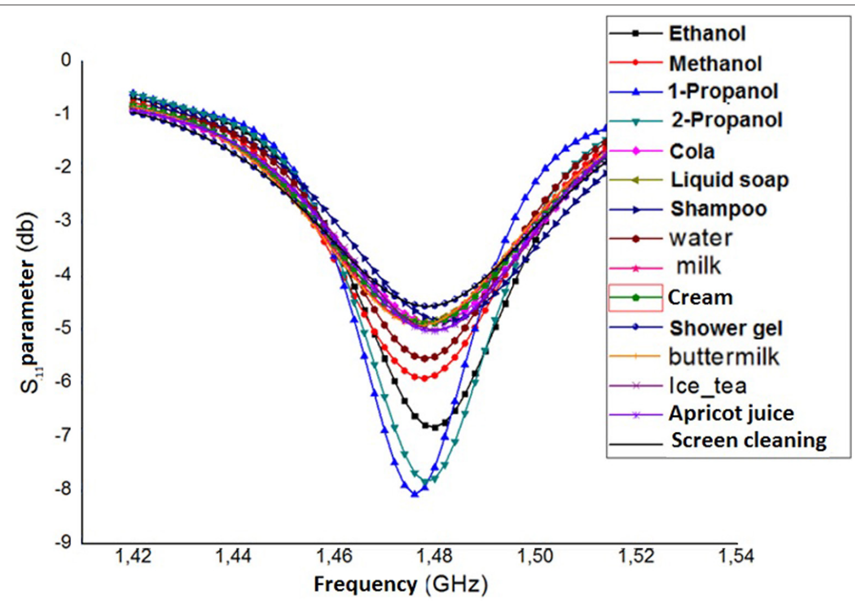


Fig. 2. Measurements of *S* parameters.

TABLE II. DISTANCES BETWEEN CLUSTERS

Single link	$D(m_1,m_2)=min_{x_1\in m_1,x_2\in m_2},D(x_1,x_2)$
Complete link	$D(m_1,m_{12})=max_{x_1\in m_1,x_2\in m_2},D(x_1,x_2)$
Average link	$D(m_1,m_2)=\frac{1}{\lvert m_1\rvert\lvert m_2\rvert}\sum_{x_1\in m_1}\sum_{x_2\in m_2}D(x_1,x_2)$
Centroid	$D(m_1,m_2)=\left(\left(\frac{1}{\lvert m_1\rvert}\sum_{x\in m_1}\vec{x}\right),\left(\frac{1}{\lvert m_2\rvert}\sum_{x\in m_2}\vec{x}\right)\right)$
Ward	$TD_{m_1\cup m_2}=\sum_{x\in m_1\cup m_2}D(x,\mu_{m_1\cup m_2})^2$
Adjusted complete link	$D(m_1,m_{12})=max_{x_1\in m_1,x_2\in m_2}D(x_1,x_2)-max_{i\in\{1,2\}}CID(m_i)$

m_1 and m_2 represent sets, D represents the distance between these sets, x_1 and x_2 represent the distance between the closest elements of m_1 and m_2 sets, and CID represents the intra-cluster distances.

subtracts the greater of the cluster intra-cluster distance values from the distance between the two farthest elements of two clusters. The computations of these methods are given in Table II.

E. Proposed Algorithm (Farthest First Hierachical Clustering Algorithm)

It is a hybrid clustering algorithm consisting of the farthest first algorithm and the processing steps of hierarchical clustering algorithms. The steps of the proposed algorithm are given below. This algorithm has been developed to find a solution to the problem of misclassification of methanol experienced in the classification of liquids using clustering algorithms.

Pseudocode of the algorithm:

1: Enter the data set and the number of clusters (*k*).

2: The first cluster center is randomly selected.

3: The distance of each remaining point to the cluster center point is calculated using the Euclidean distance metric.

$$distance(x,y)=\left\{\sum_{i=1}^{n}(x_i-y_i)\right\}^{1/2}$$

4: Select the point with the maximum distance to the center point and update the center point.

5: Calculate the distance to the cluster center for the remaining points and assign the remaining points to the cluster close to the cluster center.

6: The number of objects of the clusters is calculated.

7: Find the cluster with the least number of objects.

8: The proximity matrix between objects in the cluster with the least number of objects is calculated.

$$D(m_1,m_{12})=max_{x_1\in m_1,x_2\in m_2}D(x_1,x_2)-max_{i\in\{1,2\}}CID(m_i)$$

9: Select a random point and combine it with the other point closest to this point.

10: *Repeat*

- The proximity matrix is updated by calculating the distance of the connecting points to each other.
- Merge nearby clusters.
- *Until* there are no clusters.

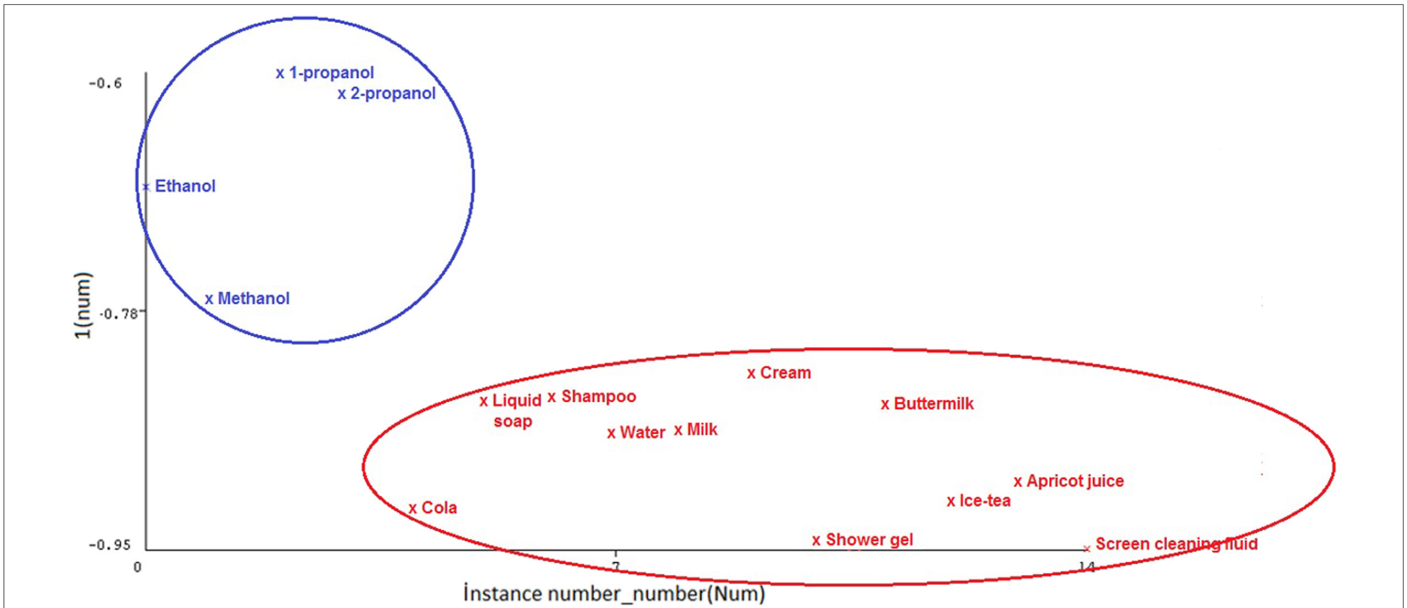


Fig. 3. Desired classification results.

F. Results

Fig. 3 shows the classification results desired to be achieved when the proposed hybrid clustering algorithm and the others were applied. As can be seen in Fig. 4, when the proposed hybrid clustering algorithm was applied, 4 hazardous and 11 non-hazardous liquids were all classified into the correct classes. On the other hand, as can be seen clearly, when Table III is considered, some of the algorithms failed to classify these liquids correctly. For instance, shampoo, which is among the non-hazardous liquids, was classified as hazardous by the farthest first algorithm. Similarly, in the cluster analysis made with K-means, hierarchical clustering, and fuzzy C-means algorithm, methanol was included in the non-hazardous class. Except for the proposed hybrid clustering algorithm, all of the algorithms used in this study classified 14 of 15 liquids correctly. To sum up, while the proposed hybrid clustering algorithm achieved 100% accuracy, the remaining ones achieved 93.3% accuracy.

IV. CONCLUSION

Illegal transportation of flammable and explosive liquids threatens human and environmental safety. For this reason, it is important to carry out safety inspection of flammable and explosive liquids in crowded places such as public transport, concert areas, and shopping malls. In this study, a measurement system capable of remote and non-contact measurement based on the reflection coefficient measurements of liquids was developed and a hybrid clustering algorithm for the classification of liquids was proposed. Measurements taken with the measurement system were used to classify suspicious liquids using the proposed hybrid clustering algorithm and its well-known alternatives, and a fair performance evaluation was made. The results of a set of experiments performed in this study showed that the proposed clustering algorithm provided 100% accuracy and showed a better performance than its alternatives when classifying hazardous liquids. A prototype system that implements the proposed approach is under development. The main advantage of the proposed system will be its estimated low cost compared to highly expensive, commercial liquid scanners so the system will have the chance of having widespread usage in railway stations and shopping malls in addition to airports.

Ethics Committee Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

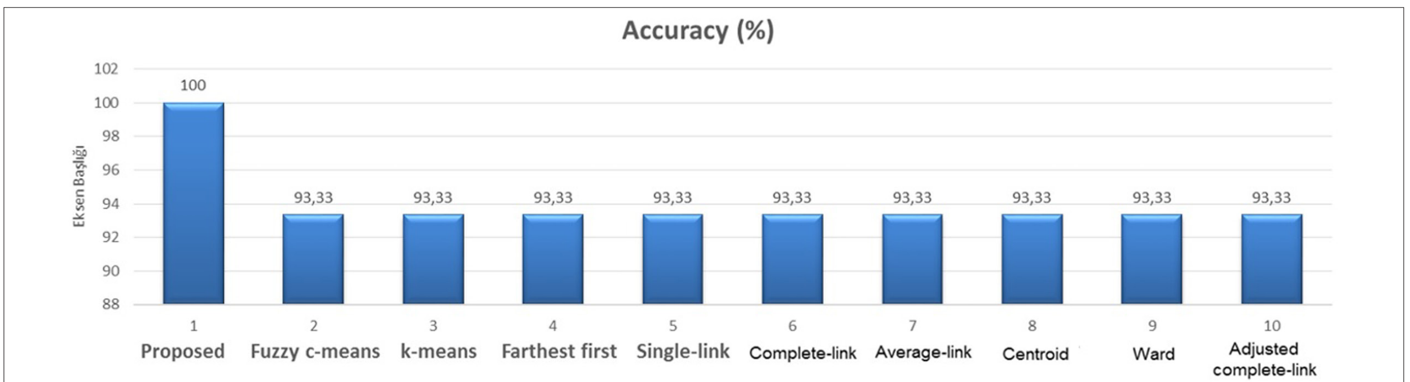


Fig. 4. Accuracy rate of the algorithms.

TABLE III. OVERALL COMPARISON OF THE PERFORMANCE OF ALL THE ALGORITHMS USED IN THIS STUDY

	Total Number of Liquids	Correctly Classified Liquids	Incorrectly Classified Liquids	Name of the Incorrectly Classified Liquid
K-means	15	14	1	Methanol
Farthest first	15	14	1	Shampoo
Fuzzy C-means	15	14	1	Methanol
Proposed hybrid clustering algorithm	15	15	0	–
Hierarchical clustering				
Single link	15	14	1	Methanol
Complete link	15	14	1	Methanol
Average link	15	14	1	Methanol
Centroid	15	14	1	Methanol
Ward	15	14	1	Methanol
Adjusted complete link	15	14	1	Methanol

Informed Consent: Informed consent is not required. We didnt carry out any experiments with human participants or animals. Thanks.

Peer-review: Externally peer-reviewed.

Author Contributions: Author Contributions: Concept – E.E.; Design – E.E., G.T.; Supervision – G.T.; Materials – E.E.; Data Collection and/or Processing – E.E.; Analysis and/or Interpretation – E.E., G.T.; Literature Search – E.E.; Writing Manuscript – E.E., G.T.; Critical Review – G.T.

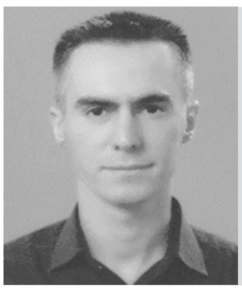
Conflict of Interest: The authors declare that there are no conflicts of interest.

Financial Disclosure: No funding was received for this study.

REFERENCES

1. E. Davicioni, and Z. Haddad, "Thyroid cancer diagnostics," Google Patents, 2018.
2. K. Wang, X. Qi, H. Liu, and J. Song, "Deep belief network based k-means cluster approach for short-term wind power forecasting," *Energy*, vol. 165, pp. 840–852, 2018.
3. W. Zhang, Z. Lian, and C. Huang, "Evaluation of unsupervised clustering methods on hyperspectral image data sets," in IEEE International Conference on Progress in Informatics and Computing (PIC), 2018, pp. 117–121.
4. S. N. H. Sheikh Abdullah et al., "Round randomized learning vector quantization for brain tumor imaging," *Comp. Math. Methods Med.*, vol. 2016, 8603609, 2016.
5. B. R. Shivakumar, and S. V. Rajashekararadhya, "Classification of Landsat 8 imagery using Kohonen's self organizing maps and learning vector quantization," in *Adv. Commun. Signal Process. VLSI Embedded Syst.* Berlin: Springer, 2020, pp. 445–462.
6. J. Lee et al., "Classification of diesel and gasoline dual-fuel combustion modes by the analysis of heat release rate shapes in a compression ignition engine," *Fuel*, vol. 209, pp. 587–597, 2017.
7. I. Barnett, and M. Zhang, "Discrimination of brands of gasoline by using DART-MS and chemometrics," *Forensic Chem.*, vol. 10, pp. 58–66, 2018.
8. V. D. Baquiran, J. A. M. Bate, M. J. B. Sembrano, J. F. Villaverde, and G. V. Magwili, "Determination of methanol, ethanol, and ethylene glycol content in coconut wine using PCA and SIMCA," in Proc. 2020 12th Int. Conf. Comput. Automat. Eng., 2020, pp. 150–154.
9. J. Wang, M. Chang, W. He, X. Lu, S. Fei, and G. Lu, "Optimization of electronic nose sensor array for tea aroma detecting based on correlation coefficient and cluster analysis," *Sens. Actuators B*, p. 128068, 2020.
10. J. A. Ardila, F. L. F. Soares, M. A. dos Santos Farias, and R. L. Carneiro, "Characterization of gasoline by Raman spectroscopy with chemometric analysis," *Anal. Lett.*, vol. 50, no. 7, pp. 1126–1138, 2017.
11. J. Dadson, S. Pandam, and N. Asiedu, "Modeling the characteristics and quantification of adulterants in gasoline using FTIR spectroscopy and chemometric calibrations," *Cogent Chem.*, vol. 4, no. 1, p. 1482637, 2018.
12. I. M. A. G. Wirasuta et al., "A rapid method for screening and determination test of methanol content in ethanol-based products using portable Raman spectroscopy," *Forensic Chem.*, vol. 16, p. 100190, 2019.
13. F. Mabood et al., "Detection and estimation of Super premium 95 gasoline adulteration with Premium 91 gasoline using new NIR spectroscopy combined with multivariate methods," *Fuel*, vol. 197, pp. 388–396, 2017.
14. T. Hayasaka et al., "An electronic nose using a single graphene FET and machine learning for water, methanol, and ethanol," *Microsyst. Nanoeng.*, vol. 6, no. 1, p. 50, 2020.
15. P. Orachorn, N. Chankow, and S. Srisatit, "An alternative method for screening liquid in bottles at airports using low energy X-ray transmission technique," in *Radiation environment and medicine: Covering a broad scope of topics relevant to environmental and medical radiation research*, vol. 8, no. 2, pp. 77–84, 2019.
16. H. Chen, Z. Hu, P. Wang, W. Xu, and Y. Hou, "Application of spectral droplet analysis method in flammable liquids identification," in 2019 Int. Conf. Optical Instruments Tech.: Optical Sensors Appl., International Society for Optics and Photonics, 2020, vol. 11436, p. 1143609.
17. P. Skládal, "Smart bioelectronic tongues for food and drinks control," *TrAC Trends Anal. Chem.*, vol. 127, p. 115887, 2020.
18. A. Gorji-Chakespari, A. M. Nikbakht, F. Sefidkon, M. Ghasemi-Varnamkhasti, J. Brezmes, and E. Llobet, "Performance comparison of fuzzy ARTMAP and LDA in qualitative classification of Iranian Rosa damascena essential oils by an electronic nose," *Sensors (Basel)*, vol. 16, no. 5, p. 636, 2016.
19. L. Guo, and A. Abbosh, "Stroke localization and classification using microwave tomography with k-means clustering and support vector machine," *Bioelectromagnetics*, vol. 39, no. 4, pp. 312–324, 2018.
20. Z. Zhang, Y. Yu, X. Li, F. Hui, X. Cheng, and Z. Chen, "Arctic sea ice classification using microwave Scatterometer and radiometer data during 2002–2017," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5319–5328, 2019.
21. Z. Zhang, F. Hui, X. Cheng, and Z. Chen, "Arctic Sea ice classification using microwave scatterometer and radiometer data," *AGUFM*, vol. 2018, p. C31E-1573, 2018.
22. B. Saçlı et al., "Microwave dielectric property based classification of renal calculi: Application of a kNN algorithm," *Comput. Biol. Med.*, vol. 112, p. 103366, 2019.
23. S. P. Rana et al., "Machine learning approaches for automated lesion detection in microwave breast imaging clinical data," *Sci. Rep.*, vol. 9, no. 1, 10510, 2019.
24. J. Van den Broek, S. Abegg, S. E. Pratsinis, and A. T. Güntner, "Highly selective detection of methanol over ethanol by a handheld gas sensor," *Nat. Commun.*, vol. 10, no. 1, p. 4220, 2019.
25. J. V. Ashurst, and T. M. Nappe, "Methanol toxicity," *StatPearls Publishing*, Treasure Island (FL), PMID: 29489213, 2019.

Ebru Efeoğlu is currently an assistant professor at Kutahya Dumlupinar University, Department of Software engineering. She received her BS degree in Geophysics Engineering from Kocaeli University and Management Information Systems from Anadolu University, Turkey. She received her PhD degree in Computer Engineering from Trakya University, Turkey in 2021. She has authored several papers in international conference proceedings and SCI-Expanded journals. Her research interests include machine learning and data mining and their applications in various research domains.



Gurkan Tuna is currently a professor in the Department of Computer Programming at Trakya University, Turkey. He is also the head of the graduate program of Mechatronics Engineering at the same university. His current research interests include wireless networks, wireless sensor networks, mobile robots, multi-sensor fusion, and smart cities.