Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 155-167
DOI: 10.38016/jista.918795

Analysis of the Parameters that Affect the Measurements of Reflection Coefficients and Evaluation of the Effects of Parameters for K Nearest Neighbors-Based Liquid Classification

Ebru Efeoglu [1], Gurkan Tuna[2*]

[1] İstanbul Gedik University, Faculty of Economics, Administrative and Social Sciences, Department of Management Information Systems, İstanbul/Turkey

[2] Trakya University, Edirne Vocational College of Technical Sciences, Edirne/Turkey

ebru.efeoglu@gedik.edu.tr, gurkantuna@trakya.edu.tr

Abstract
In this study, microwave spectroscopy method has been used in liquid measurements and K nearest neighbors algorithm has been used for classifying liquids. For this aim, firstly an experimental setup consisting of a vector network analyzer, a patch antenna and a bottle have been built to measure the reflection parameter of each liquid used in classification experiments. The aim of this study is to examine both the parameters that may affect the measurements taken with the proposed system and the algorithm parameters that may affect the performance in the classification of liquids and the effects of these parameters. Measurements have been taken by leaving different distances between the antenna and the liquid in order to examine whether the distance of the liquids to the antenna affects the measurement result, and if so, what effect. For examining the parameters of K nearest neighbors algorithm that may affect the classification, the scattering parameters of different liquids measured using the patch antenna have been used as microwave dataset. In addition, the effect of container type has been analyzed. Performance tests have been conducted by weighting and without weighting the algorithm, by measuring the accuracy rate when different numbers of nearest neighbors and different distance metrics have been used. The results reveal that the classification made by applying weighting is more successful than the classification made without weighting regardless of the number of nearest neighbors and used distance metrics.

Keywords: Microwave measurement, circular patch antenna, K nearest neighbors algorithm, liquid classification, distance metric, weighting, number of nearest neighbors.

K En Yakın Komşular Tabanlı Sıvı Sınıflandırması İçin Yansıma Katsayıları ve Parametrelerin Etkilerinin Değerlendirilmesi

Öz
Bu çalışmada sıvı ölçümlerinde mikrodalga spektroskopi yöntemi kullanılmış ve sıvıların sınıflandırılmasında K en yakın komşular algoritması kullanılmıştır. Bu amaçla, öncelikle sınıflandırma deneylerinde kullanılan her bir sıvının yansıma parametresini ölçmek için bir vektör ağ analizörü, bir yama anteni ve bir şişeden oluşan deney düzeneği oluşturulmuştur. Bu çalışmanın amacı, hem önerilen sistemle alınan ölçümleri etkileyebilecek parametreleri hem de sıvıların sınıflandırılmasında performansı etkileyebilecek algoritma parametrelerini ve bu parametrelerin etkilerini incelemektir. Sıvıların antene olan mesafesinin ölçüm sonucunu etkileyip etkilemediğini, etkiliyorsa etkisini incelemek için anten ile sıvı arasında farklı mesafeler bırakılarak ölçümler yapılmıştır. Sınıflandırmayı etkileyebilecek en yakın komşu algoritmasının parametrelerini incelemek için, yama anten kullanılarak ölçülen farklı sıvıların saçılma parametreleri mikrodalga veri seti olarak kullanılmıştır. Ayrıca kap tipinin etkisi analiz edilmiştir. Farklı sayıda en yakın komşu ve farklı mesafe ölçütleri kullanıldığında doğruluk oranı ölçülerek ağırlıklandırılarak ve algoritma ağırlıklandırılmadan performans testleri yapılmıştır. Sonuçlar, ağırlıklandırma uygulanarak yapılan sınıflandırmanın, en yakın komşu sayısına ve kullanılan uzaklık ölçütlerine bakılmaksızın ağırlıklandırma yapılmadan yapılan sınıflandırmaya göre daha başarılı olduğunu ortaya koymaktadır.

Anahtar Kelimeler: Mikrodalga ölçümü, dairesel yama anten, K en yakın komşular algoritması, sıvı sınıflandırması, mesafe ölçüsü, ağırlıklandırma, en yakın komşu sayısı.

* Corresponding Author.
E-mail: gurkantuna@trakya.edu.tr

Recieved : 17 April 2021
Revision : 07 June 2021
Accepted : 12 July 2021

1. Introduction

The measure of a liquid being a flammable liquid is the flash point. Flash point is the lowest temperature at which a liquid will emit sufficient vapor to form an air-flammable mixture. Alcohols cannot be distinguished visually because they are 100% liquid and colorless liquids, and they are in a small group of chemicals that can spontaneously ignite (Q. Chen, Kang, Zhou, & Wang, 2017). The flash point of alcohol-water solutions diluted with water will increase and as the flash point value of flammable liquids increases, the risk of fire hazards decreases (Cheremisinoff, 1999), However, these liquids, besides their flammable properties, also contain toxic substances and endanger human health. For instance, drinking methanol accidentally causes serious health problems (Slaughter, Mason, Beasley, Vale, & Schep, 2014). From this perspective, the classification of liquids is important to manage the hazards of chemicals and take necessary measures.

Classification of flammable and explosive liquids using THz time history spectroscopy in classification of liquids (Tan et al., 2017), characterization of aqueous alcohol solutions in bottles and determination of the alcohol content of an aqueous solution were performed (Jepsen, Jensen, & Møller, 2008). An electronic nose using machine learning was proposed to detect mixtures of water, methanol and ethanol (Hayasaka et al., 2020). Flammable liquids were detected using the X-ray spectroscopy method (Orachorn, Chankow, & Srisatit, 2019), (H. Chen, Hu, Wang, Xu, & Hou, 2020). Raman spectroscopy method was used for screening and determination of methanol content in ethanol-based products (Wirasuta et al., 2019).

Liquids differ in complex permeability and reflection and transmission coefficients. Microwave frequency bands can be used to determine complex permeability, reflection and transmission coefficients of liquids and to characterize liquids. Liquid characterization is also important for food safety and quality. For fruit quality control, non-destructive control experiments with microwave method were performed (Jawad et al., 2017). It was also used to calculate the permeability, reflection coefficient, S_{11}, and transmission coefficient, S_{21}, (Li, Haigh, Soutis, Gibson, & Sloan, 2018), (Jiang, Ju, & Yang, 2016) of the liquids. Microwave measurement method is fast, non-hazardous and not affected by environmental conditions (Li, Haigh, Soutis, Gibson, & Sloan, 2017b). It was used to measure the permeability of thin layer materials (Borisov & Karpenko, 2001) and to measure the parameters of silicon (Yurchenko, Novikov, & Kitaeva, 2012).

There are many microwave measurement methods used in fluid measurements in the literature, such as open-ended coaxial probe techniques (Li et al., 2017b) and Free space method (Jose, Varadan, & Varadan, 2001). When the coaxial probe method is used, the probability of inaccuracy in solid material measurements is high. The cost of the measurement method to be used is also important. For example, Time-Domain Reflectometers (TDRs) are expensive (Venkatesh & Raghavan, 2005). For Free space technique, measurements vary according to the choice of the horn antenna, the design of the specimen holder and the geometry and location of the specimen. An improperly determined measurement location and an unsuitable sample geometry increase the likelihood of erroneous measurements (Li, Haigh, Soutis, Gibson, & Sloan, 2017a).

K Nearest Neighbors (KNN) algorithm does not require a training step and is resistant to noisy training data (Bhatia, 2010). Therefore, it is commonly used as a basic classifier in many field problems (Jain, Duin, & Mao, 2000). KNN is known as instance-based learning. In it, training samples are stored exactly and the classification of an unknown, i.e., a new test sample, takes into account the similarity between the samples in the training set. The similarity is, for example, its proximity to the data in the training set. The distance metric is used to decide which member of the training set is the nearest. Once the nearest training sample is found, the class is estimated for the test sample (Chakrabarti et al., 2008). For this, the test sample is compared with the records that are most similar in the current training set at hand (Larose & Larose, 2014). In the literature, there are several distance metrics. However, the most commonly used distance function for KNN is the Euclidean distance metric. The microwave measurement methods and KNN algorithm were used together for different purposes including classification of kidney stones (Saçlı et al., 2019), detection of deep tissue injuries (Moghadas & Mushahwar, 2018) and detection of breast cancer (Aydın & Kaya Keleş, 2017).

Although the performance of KNN has been heavily studied, it has not been evaluated for classifying alcoholic liquids with different distance metrics, the different number of nearest neighbors, and the weighting process. Different from the literature, in this study it is evaluated whether weighting application using different distance metrics and changing the number of nearest neighbors can affect the performance of KNN algorithm when it is used for alcoholic liquid classification made with microwave datasets. Another issue examined in the study is the parameters that may affect the measurements. Determining these parameters and paying attention to them while making measurements ensure more reliable measurements. The remainder of this paper is as follows. First, the factors affecting microwave measurements are examined and the effects of different parameters on the measurement results are presented. Parameters used for KNN algorithm and their implementation are described in Section 2. Experimental setup of this study which was used to collect and use microwave measurement data is explained in Section 3. Finally, Section 4 concludes this paper.

2. Parameters of K Nearest Neighbors Algorithm for Liquid Classification and Their Implementation

In KNN, the similarities of the data to be classified with the data in the training set are computed. As a result of the computation, the data to be classified is assigned to the nearest classes in the training set. The nearest number of neighbors and similarity function criteria affect the performance of KNN (Kresse & Danko, 2012). In KNN, training samples are defined with n-dimensional numerical properties. Each sample shows a point in n-dimensional space. Thus, all training samples are stored in an n-dimensional sample space. The objective is to find the nearest *k* training samples to the unknown sample. The distance between two points, such as X= (x_1, x_2,…., X_n) and Y= (y_1, y_2, ……., Y_n) is expressed by different distance metrics (Chakrabarti et al., 2008).

Basic parameters of KNN algorithm are distance metric, number of nearest neighbors, *k*, and weighting application. *k* expresses the number of neighbors, and classification is made based on this value. For instance, if *k* value is set to 1, the nearest 1 neighbor is taken into consideration and the tested sample is assigned to the class where this neighbor is located. In this study, four different distance metrics will be used.

As given in (1) Minkowski distance metric is calculated by summing the absolute difference between the two points by taking the *p* prime in the distance criterion. Then 1/*p* of this sum is taken. This equation gives Euclid distance if *p* value is set to 2, Manhattan distance if $p \to \infty$, and Chebyshev distance if *p* value is set to 1 (Kresse & Danko, 2012).

$$\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \quad (1)$$

Euclidean distance metric, defined as a straight line distance between two points in any number of dimension spaces, is calculated by taking the square root of the sum of the squares of the differences between the respective coordinates of each point, as given in (2) (Kresse & Danko, 2012).

$$\left(\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\right) \quad (2)$$

As given in (3) Manhattan distance metric calculates the linear distance between actual vectors using the sum of absolute differences (Kresse & Danko, 2012).

$$\left(\sum_{i=1}^{n}|x_i - y_i|\right) \quad (3)$$

Finally, Chebyschev distance metric, also known as the maximum value distance, is calculated using (4) (Rey, Kordon, & Wells, 2012).

$$lim_{p\to\infty}\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} = max_{i=1}^{n}|x_i - y_i| \quad (4)$$

In KNN, weight values are assigned to all neighbors. The weight values of neighboring samples that are closer to the sample to be classified in the weighting application are higher than the other neighbors. Generally, the most preferred method of assigning weight is the method in which the weight of each neighbor is taken in 1/*d*. Here *d* represents the distance between neighbors (Doad & Bartere, 2013).

2.1. Performance of the Classifier

There are several performance metrics used to evaluate how well a classifier is performing at the end of the classification process (Chakrabarti et al., 2008). The metrics used in this study were confusion matrix, accuracy, precision, recall, Kappa, Area Under the Receiver Operating Characteristic (ROC) Curve (AUC), Matthews Correlation Coefficient (MCC) and Root Mean Square (RMS). Confusion matrix is often used to determine the performance of the classification model with a series of test data with actual values known (Chakrabarti et al., 2008), (Larose & Larose, 2014). True Positive (TP) values, i.e. actual alcoholic liquids, are positive values that have been predicted correctly. True Negative (TN) values, i.e. actual non-alcoholic liquids, are negative values that have been predicted correctly. These values indicate that, for the selected sample, the actual class is the same as the predicted class. They are the diagonal elements of the matrix and are shown in green in Figure 1. False Positive (FP) values, i.e. non-alcoholic liquids misclassified as alcoholic liquids, and False Negative (FN) values, i.e. alcoholic liquids misclassified as non-alcoholic liquids, occur when the actual class is different from the predicted class. That is, they indicate the number of incorrectly classified samples. They are shown in red in Figure 1. The increase in TP and TN values and the decrease in FN and FP values indicate that the classification performance is good.

Using the confusion matrix, the accuracy value can be calculated as in (5). Accuracy is the ratio of accurately estimated samples to the total number of samples. High accuracy rate is an indicator of high classification performance.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (5)$$

		Predicted	
		TP (Alcoholic Liquids)	TN (Non-Alcoholic Liquids)
Actual	TP (Alcoholic Liquids)	TP	FN
	TN (Non-Alcoholic Liquids)	FP	TN

Figure 1. Confusion matrix

Other performance metrics are also calculated using a confusion matrix. For example, precision is calculated using the left side of the matrix (Equation (6)). It is a measure of the precision of the classification algorithm. Recall, which is a measure of the integrity of the classification algorithm, is calculated using (7).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN + TP}} \quad (7)$$

The harmonic mean of precision and recall values gives F-measure value (8). It is difficult to compare the two models with low recall and high precision and vice versa. In this case, the value of F-measure is checked.

$$\text{F\_Measure} = \frac{2*\text{Recall}*\text{ Precision}}{\text{Recall+ Precision}} \quad (8)$$

Kappa value is used to measure how much agreement there is between the classification made as a result of the classification and the actual classifications in a dataset.

$$\text{Kappa} = \frac{\text{P(x)}-\text{P(y)}}{1-\text{P(y)}} \quad (9)$$

where P(x) is a value that shows probabilistic accuracy of the classification algorithm and P(y) is the weighted average of the probability of classifications made in the same dataset.

In a ROC curve, the horizontal axis shows the false positive rate (FPR), the vertical axis the correct positive rate (TPR). The area under this curve (AUC) is used as the classification metric. FPR and TPR values are calculated using (10) and (11), respectively.

$$\text{FPR} = \frac{FP}{FP+TN} \quad (10)$$

$$\text{TPR} = \frac{TP}{TP+FN} \quad (11)$$

MCC is used as a measure of the quality of binary classifications in machine learning and is calculated using (12).

$$MCC = \frac{TPxTN-FPxFN}{\sqrt{(TP+FP)x(TP+FN)x(TN+FP)x(TN+FN)}} \quad (12)$$

RMS is used to scale the differences between the actual values and the values predicted by the model. It is determined by taking the square root of the mean square error and calculated using (13).

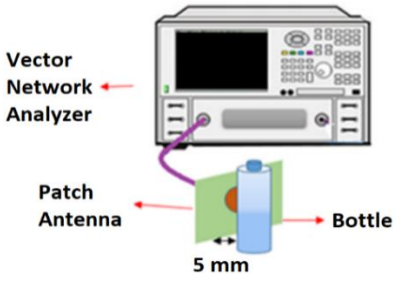
$$\text{Rms} = \sqrt{\frac{1}{\text{n}}\sum_{\text{k=1}}^{\text{n}}(\text{T}_{\text{ik}} - \text{A}_{\text{k}})^2} \quad (13)$$

where T_{ik} is the predicted value and A_k is the objective value. If the error value approaches zero, it means that the correct prediction of the classification algorithm increases.

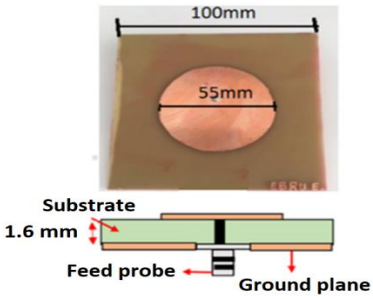
When the values of accuracy, precision, recall, F-measure, AUC, MCC and Kappa are 1, it indicates perfect classification. Therefore, these values are desired to be as close to 1 as possible.

3. Experimental Setup for Collecting and Using Microwave Measurement Data

Scattering parameters (S parameters) describe the electrical behavior of linear electrical networks when they are exposed to various steady-state stimuli by electrical signals. The measurement system shown in Figure 2 consists of a vector network analyzer (VNA) and a circular patch antenna that can send signals between a specific frequency band and record the reflection coefficient of the reflected signals. The reflection coefficient (S_{11}) of the reflected signals from the source is expressed as the ratio of the amplitude of the reflected signal to the amplitude of the transmitted signal. The resonance frequency of the antenna fed with the 50 Ohm SMA (SubMiniature version A) feed probe is 1.5 GHz. The measurement setup used in this study is given in Figure 2. The reflection parameter (S_{11}) of each liquid was measured so that the distance between the antenna and the bottle remains 5 mm without touching the patch antenna to the liquids in 0.5 liter pet bottles with the microwave measurement device. Measurements were made between 1.42-1.54 Ghz.



(a)



(b)

Figure 2. a) Experimental setup, b) Schematic view of the patch antenna

There were 56 S_{11} values for each liquid. Then, these values were divided into two classes as alcoholic and non-alcoholic liquids by using KNN algorithm. The liquids used in the study were non-alcoholic liquids such as cola, soap, shampoo, water, milk, bath cream, shower gel, ice-tea (peach), cherry juice, ayran and alcoholic liquids such as cologne, whiskey, white wine, raki.

Apart from these liquids, Ethanol, Methanol, 1-Propanol, Isopropanol and their aqueous solutions with different volume concentrations were used. The total number of liquids used was 54, including 44 alcoholic and 10 non-alcoholic. In analyzing the effect of the type of container in which the liquid is on the measurement result in liquid measurements, measurements were taken by placing the liquids in glass and plastic bottles and the measurement results were compared. The aim of the experiments was to determine the factors affecting S parameter measurements and to examine the effects of KNN parameters on classification performance in order to characterize the liquid with high accuracy in measurements. In determining the factors affecting the measurements in the experiments, the effect of the distance of the bottle to the antenna and the effect of the container that the liquid is in were analyzed. The parameters used to examine the effects of KNN parameters on classification performance are *k* value, distance metric and weighting application. In order to examine the classification parameters, a separate classification was made for each parameter in the classification of alcoholic and non-alcoholic liquids using microwave data. The best values were tried to be determined by comparing the classification results.

3.1. Results and Discussion

In order to find out whether the distance between the antenna and the liquid affects the microwave measurement data, measurements were taken by leaving different distances between the antenna and the liquid. The measurement of liquids in pet bottles was made by leaving 5 mm, 1 mm and 25 mm between antenna and liquid. The results of alcohol measurements taken for three different distances are given in Figure 3. This step was repeated for non-alcoholic liquids to examine the effect of measurement distance, and the results are given in Figure 4. As it can be seen in the figures, it is seen that the resonance peak increases with the increase of the distance between the antenna and the liquid in the measurements of all liquids. Since the resonance peak gives its highest value in the air environment, an increase in the value of the peak as it moves away from the antenna in liquid measurements indicates that the sensitivity of the antenna decreases. In other words, as the liquid moves away from the antenna, the sensitivity of the antenna to detect the liquid decreases. In order to measure the liquid accurately, the distance between the liquid and the antenna should be as small as possible. Another parameter whose effect on measurement data is examined is the container effect. For this, measurements were taken using different containers and the results are given in Figure 5 and Figure 6. Glass bottles and plastic bottles were used as container types in the measurements.

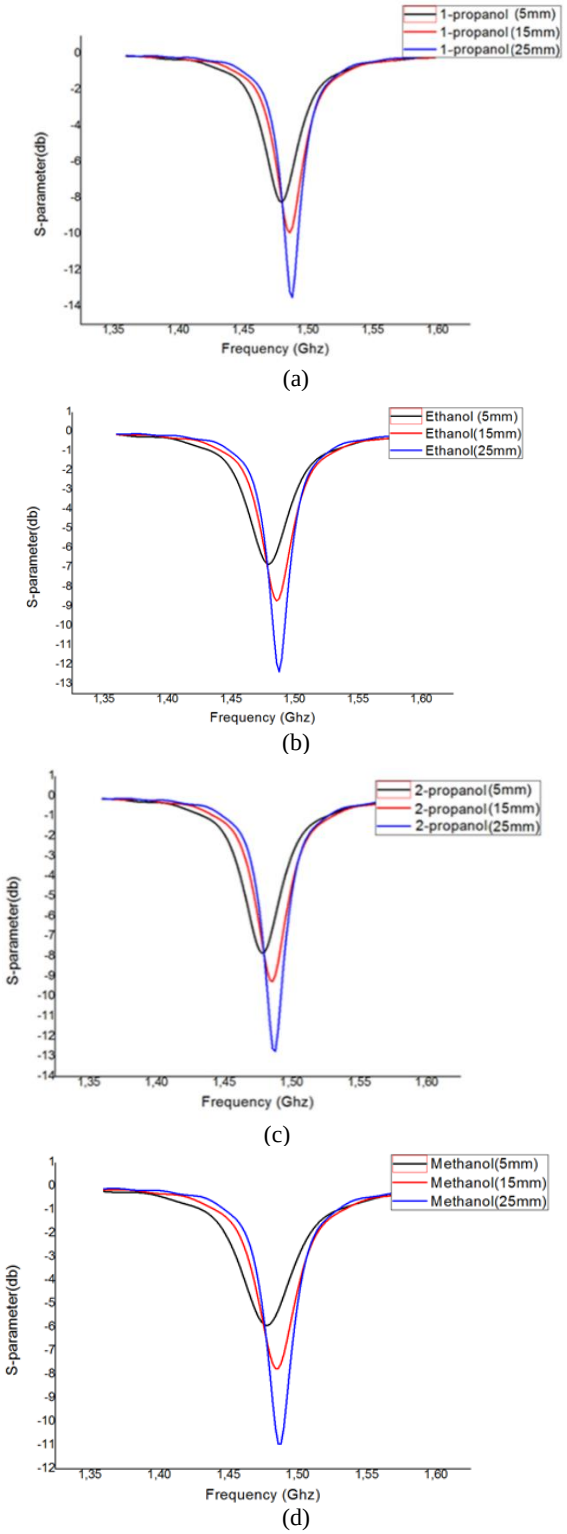


Figure 3. S parameter measurements of alcohols at different antenna-bottle distances a) Isopropyl (1-propanol) b)Ethanol c)Isopropanol (2-propanol) d)Methanol

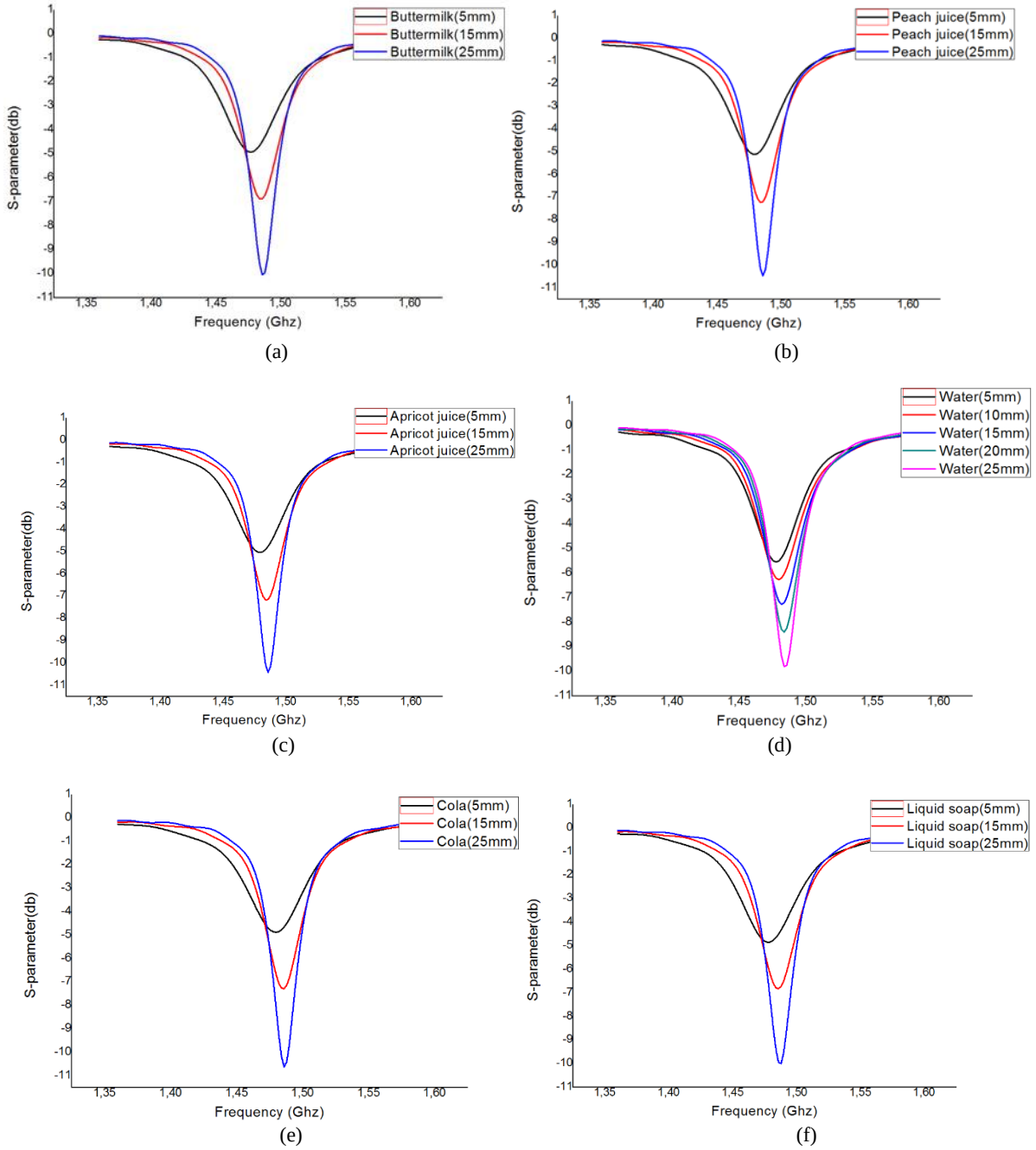
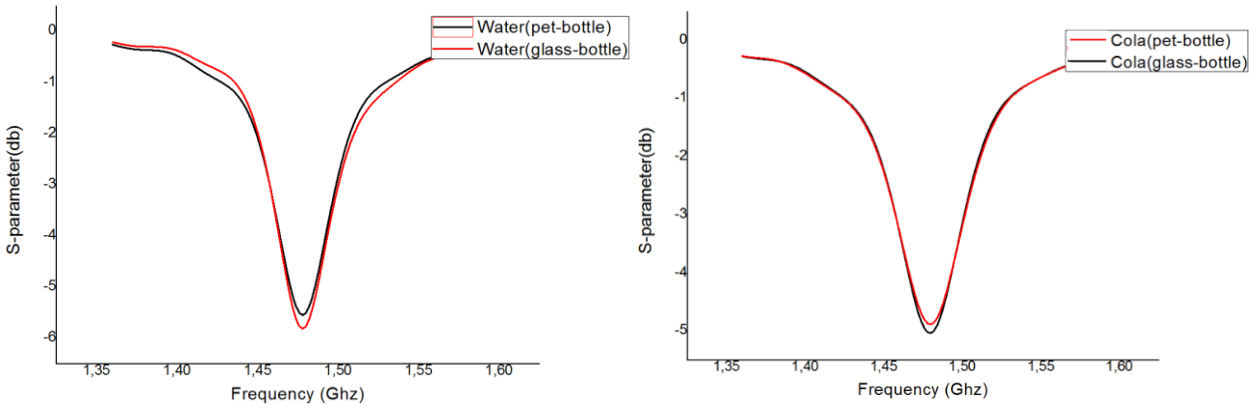


Figure 4. S parameter measurements of everyday liquids at different antenna-bottle distances a) Buttermilk b) Peach juice c) Apricot juice d) Water e) Cola f) Liquid soap



(a) (b)

Figure 5. S parameter measurements of everyday liquids in different bottles a) Water b) Cola

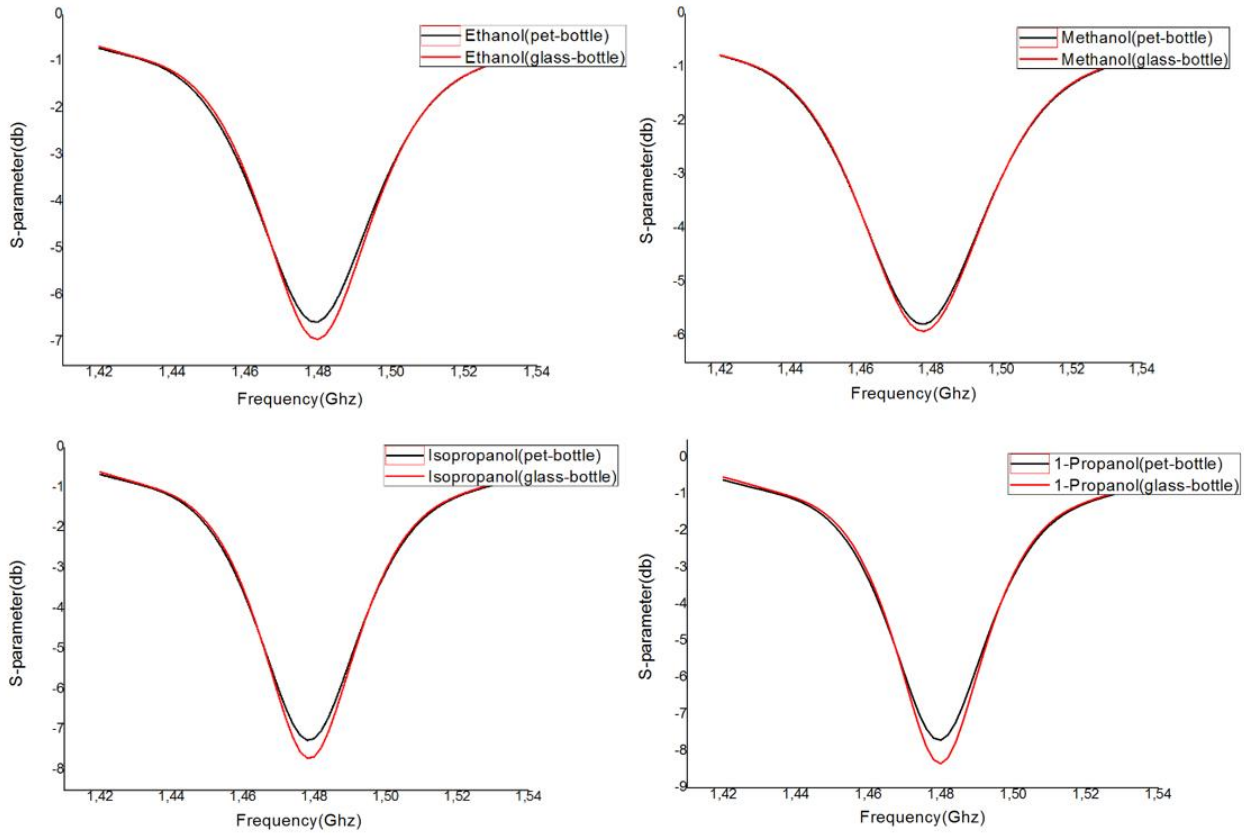


Figure 6. S parameter measurements of alcohols in different bottles

The measurement results taken using different containers show us that the resonance peak gives a higher value in the measurements taken using the glass container (bottle). S_{11} parameter measurements of the liquids were used in the study and S_{11} parameter is the reflection coefficient. Signals reflected from the object to be measured are detected by the antenna. Therefore, taking measurements in (using) a highly reflective container like glass will suppress the reflected signals from the liquid. Therefore, since the signals reflected from the liquid are important for us, it is more appropriate to use pet bottles with low reflectivity in liquid measurements. It was understood that the measurement results, the distance and the type of container used in microwave measurements affect the results. Classifications were made using different parameters to investigate the KNN parameters affecting the classification. The dataset used in the classification studies is given in Figure 7.

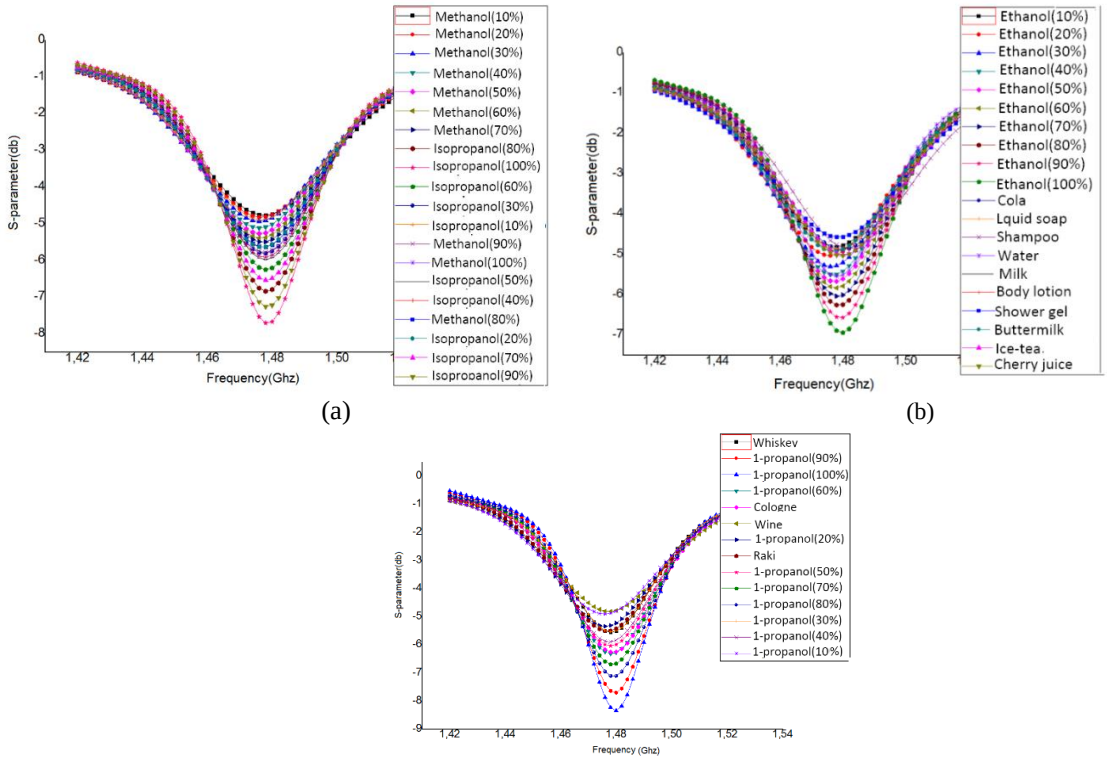


Figure 7. S parameter measurement results used in classification

For each liquid, the measurements were repeated twice and a dataset was created. The dataset was created from 108 measurement data belonging to 54 different liquids. Then, using WEKA, 10-fold cross-validation was applied. As shown in Figure 8, the dataset was divided into 10 parts, 9 parts were used for the training and 1 part was used as the test data.

Cross Validation	1st group	2nd group	3rd group	4th group	5th group	6th group	7th group	8th group	9th group	10th group
1st	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
2nd	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
3rd	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
4th	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
5th	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
6th	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
7th	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
8th	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
9tn	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
10th	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

Figure 8. 10-fold cross-validation

After the cross-validation, tradeoff between *k* value and the accuracy rate of the classifier for different distance metrics is shown in Figure 9. As can be seen, the accuracy rate of the classifier mostly decreased as the number of neighbors increased for all the distance metrics. The highest accuracy rate was 100% and the lowest accuracy rate was 90%. As can be seen in Figure 10 and Figure 11, the number of correctly classified liquids was 108 when *k* was set to 1 for different distance metrics, while the number of correctly classified liquids decreased to 97 when Chebyshev distance metric was used and *k* was set to 7. Likewise, although there was no misclassified liquid when *k* was set to 1, it was seen that a total of 11 liquids were misclassified when *k* was set to 7. After the classification experiments with these neighbor numbers and distance metrics, weighting was applied. As a result of this, for all the k values and distance metrics 100% accuracy was obtained and all the liquids were classified correctly. The confusion matrices obtained for different *k* values and the distance metrics are given in Figure 12. When the confusion matrix is analyzed, it can be seen that all the liquids were correctly classified when *k* was set to 1 and Euclidean distance metric was preferred. However, when *k* was set to 2 and Euclidean distance metric was preferred, 2 alcoholic liquids were incorrectly classified as non-alcoholic liquids and 6 non-alcoholic liquids were incorrectly classified as alcoholic liquids.

When the confusion matrix is analyzed, it can be seen that the average number of alcoholic liquids that was predicted incorrectly for different *k* values was 6.8 when Euclidean distance metric was preferred, was 6.8 when Chebyshev distance metric was preferred, was 7.2 when Manhattan distance metric was preferred, and finally it was 6.8 when Minkowski distance metric was preferred. This reveals that the distance measure which is the least affected by the change of *k* value in the detection of alcoholic liquid is Manhattan distance.

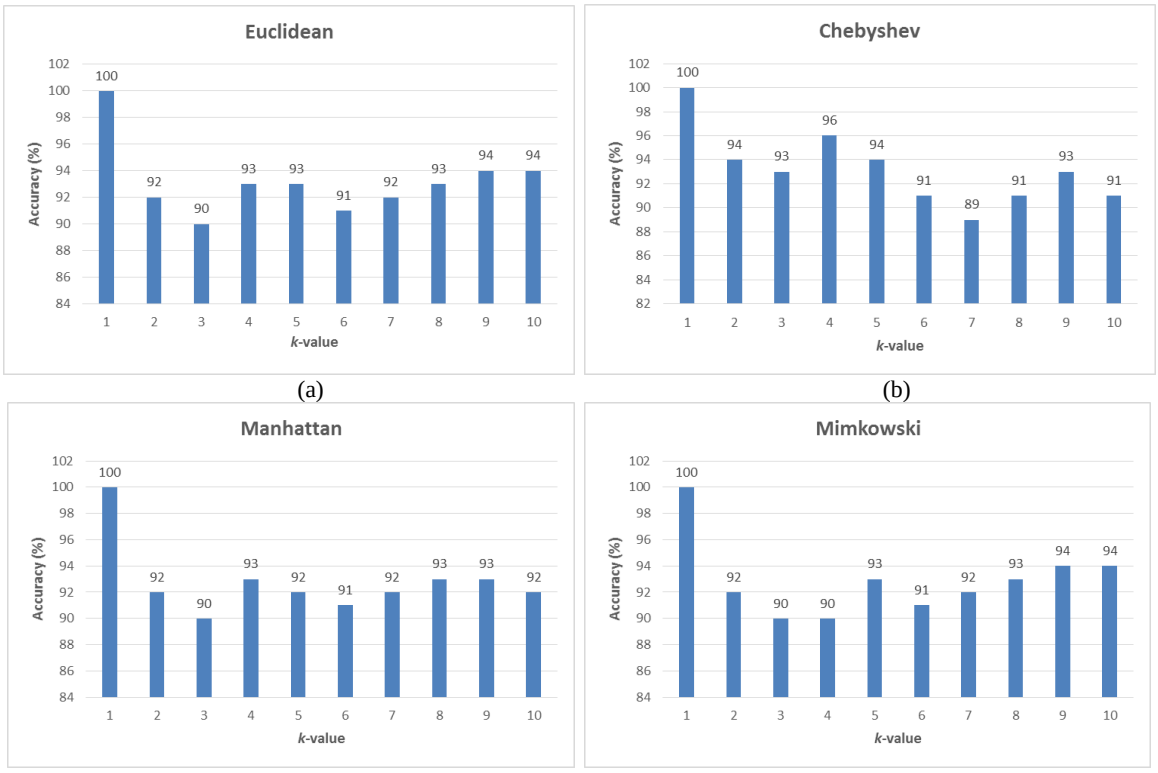


Figure 9. Tradeoff between *k* value and accuracy rate when the distance metric was a) Euclidean, b) Chebyshev, c) Manhattan, d) Minkowski

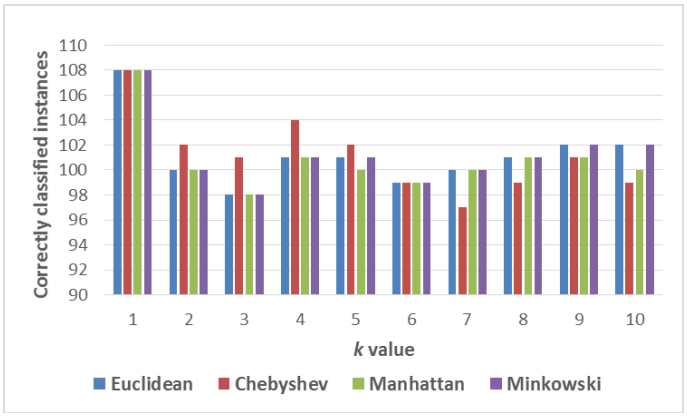


Figure 10. Tradeoff between *k* value and the number of correctly classified instances

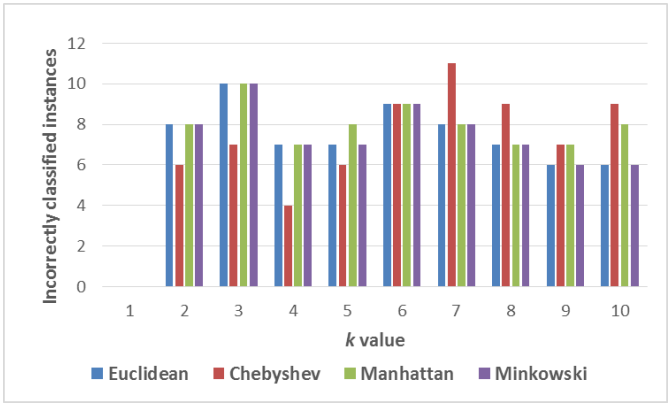


Figure 11. Tradeoff between *k* value and the number of incorrectly classified instances

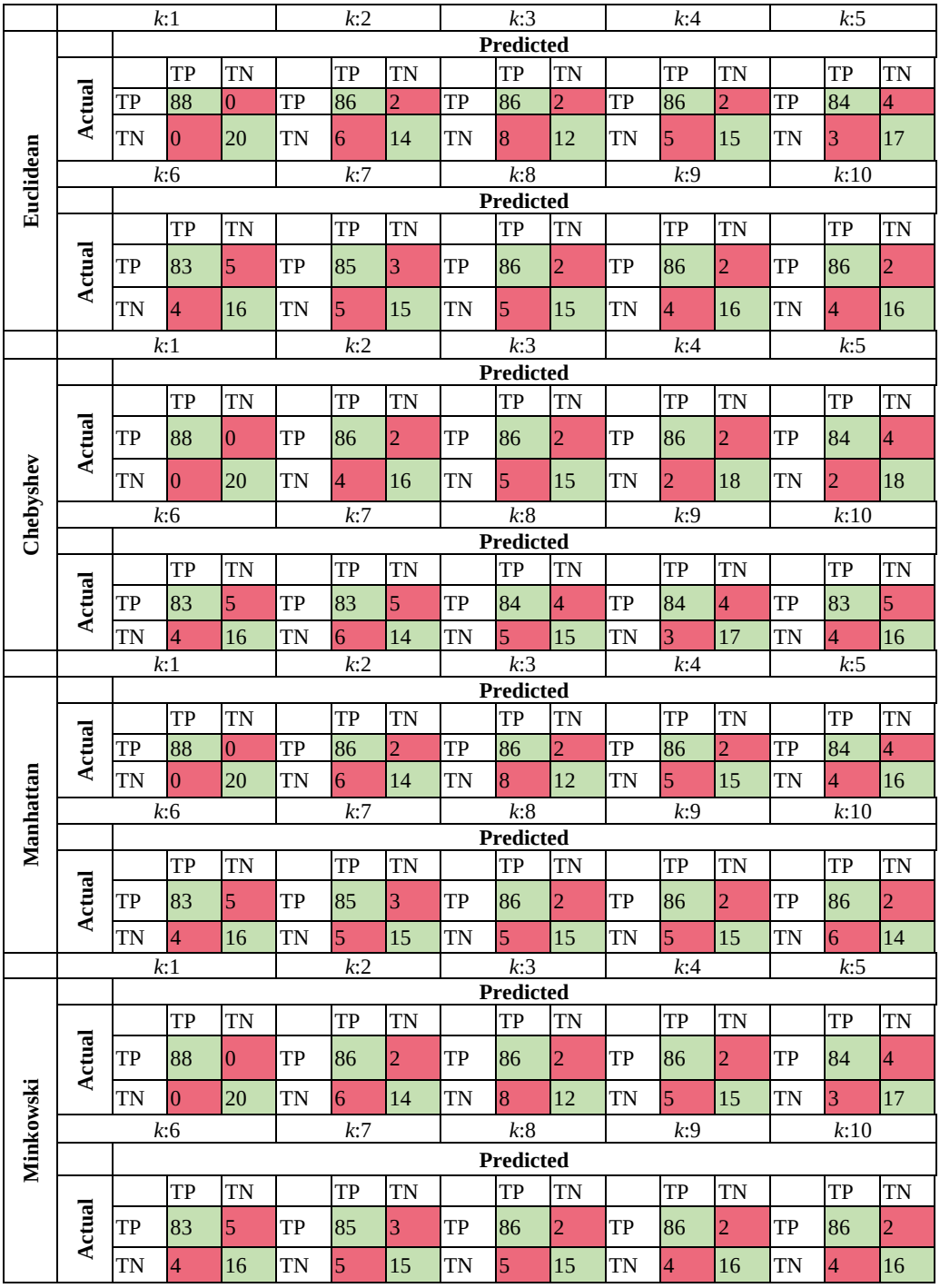


Figure 12. Confusion matrix for different *k* values and different distance metrics

When the confusion matrix is considered to examine the average number of non-alcoholic liquids predicted correctly, it can be seen that the results were different. The average number of non-alcoholic liquids that was predicted correctly for different *k* values was 15.6 when Euclidean distance metric was preferred, was 16.5 when Chebyshev distance metric was preferred, was 15.2 when Manhattan distance metric was preferred, and finally it was 15.6 when Minkowski distance metric was preferred.

Performance metrics obtained from classifications using different *k* values and different distance metrics are given in Table 1. When *k* was set to 1, all the performance metrics reached the maximum value and RMS was at the minimum value. With the increase in *k* value, there was a decrease in the performance metrics. When *k* was set to 7, the highest decrease was seen in Precision and Recall values when using Chebyshev distance metric. In the classifications made, an increase in RMS was observed with the increase of the *k* value.

For Chebyshev distance metric, RMS became 0.09 when *k* was set to 1 and it became 0.22 when *k* was set to 10. The lowest Kappa value was 0.65 and it was obtained when *k* was set to 3 for Euclidean, Manthattan and Minkowski distance metrics and when *k* was set to 7 for Chebyshev distance metric.

Since the number of alcoholic and non-alcoholic data used in this classification study was not equal, the dataset was an unbalanced dataset. For unbalanced datasets, the use of MCC values is recommended (Chicco & Jurman, 2020). When MCC values obtained for different *k* values were examined, the average MCC values were computed as 0.78 (for Euclidean), 0.78 (for Chebyshev), 0.76 (for Manhattan) and 0.78 (for Minkowski). This result shows that the distance metric most sensitive to the value of *k* was Manhattan.

Table 1. Performance metrics for different distance metrics

Distance metric	Metric	*k*=1	*k*=2	*k*=3	*k*=4	*k*=5	*k*=6	*k*=7	*k*=8	*k*=9	*k*=10
Euclidean	Precision	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	Recall	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	F-measure	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	MCC	1	0.74	0.66	0.77	0.79	0.72	0.74	0.77	0.81	0.81
	AUC	1	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
	KAPPA	1	0.73	0.65	0.77	0.78	0.72	0.74	0.77	0.80	0.80
	RMS	0.09	0.17	0.20	0.21	0.22	0.22	0.22	0.21	0.18	0.18
Chebyshev	Precision	1	0.94	0.93	0.96	0.94	0.91	0.89	0.91	0.93	0.91
	Recall	1	0.94	0.93	0.96	0.94	0.91	0.89	0.91	0.93	0.91
	F-measure	1	0.94	0.93	0.96	0.94	0.91	0.89	0.91	0.93	0.91
	MCC	1	0.81	0.77	0.87	0.82	0.72	0.65	0.71	0.79	0.72
	AUC	1	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.98	0.97
	KAPPA	1	0.80	0.77	0.87	0.82	0.72	0.65	0.71	0.78	0.72
	RMS	0.09	0.15	0.17	0.17	0.19	0.20	0.22	0.21	0.19	0.21
Manhattan	Precision	1	0.92	0.90	0.93	0.92	0.91	0.92	0.93	0.93	0.92
	Recall	1	0.92	0.90	0.93	0.92	0.91	0.92	0.93	0.93	0.92
	F-measure	1	0.92	0.90	0.93	0.92	0.91	0.92	0.93	0.93	0.92
	MCC	1	0.74	0.66	0.77	0.75	0.72	0.74	0.77	0.77	0.74
	AUC	1	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
	KAPPA	1	0.73	0.65	0.77	0.75	0.72	0.74	0.77	0.77	0.73
	RMS	0.09	0.17	0.20	0.20	0.22	0.22	0.22	0.21	0.19	0.19
Minkowski	Precision	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	Recall	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	F-measure	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	MCC	1	0.74	0.66	0.77	0.79	0.72	0.74	0.77	0.81	0.81
	AUC	1	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
	KAPPA	1	0.73	0.65	0.77	0.78	0.72	0.74	0.77	0.80	0.80
	RMS	0.09	0.17	0.20	0.21	0.22	0.22	0.22	0.21	0.18	0.18

4. Conclusion

The classification of liquids, which is important to manage the hazards of chemicals, is an interesting research topic in recent years. Therefore, various methods were proposed for liquid classification. As it is easier to implement than other algorithms, KNN algorithm was often preferred for classification problems. However, the use of KNN algorithm for classifying liquids that contain alcohol is still limited. In this study, the parameters affecting S_{11} parameter measurements were analyzed in order to characterize the liquid in the most appropriate way. In addition, the application of weighting in the performance of KNN algorithm for classification, the use of different number of nearest neighbors and different distance metrics was examined.

It was observed that the increase in the number of the nearest neighbors reduced the classification performance. Although generally reducing the value of the nearest neighbor number increases the algorithm's sensitivity to noisy data, the low number of nearest neighbors led to the better results due to the very low noise in the S_{11} parameters. Considering the effect of distance metrics, when the nearest neighbor number value was 1, all the distance metrics led to the same result.

After weighting was applied, all the liquids were classified correctly with 100% accuracy. By applying weighting, the performance of the classification can be made independent of distance metrics and *k* values. Therefore, it is recommended to apply weighting to KNN algorithm in the classification made with S_{11} data. Moreover, the results obtained in this study made it clear that in order to increase the sensitivity of the measurements, it is recommended to make liquid measurements in pet bottles with low reflective

properties and to take measurements by keeping the liquid as close to the antenna as possible.

References

Aydın, E. A., & Kaya Keleş, M. (2017). Breast cancer detection using K-nearest neighbors data mining method obtained from the bow-tie antenna dataset. *International Journal of RF and Microwave Computer-Aided Engineering, 27*(6), e21098.

Bhatia, N. (2010). Survey of nearest neighbor techniques. *arXiv preprint arXiv:1007.0085*.

Borisov, V., & Karpenko, A. (2001). Using of the Michelson microwave interferometer for the measurement of permittivity of thin-layer materials. *Russian journal of nondestructive testing, 37*(8), 597-599.

Chakrabarti, S., Cox, E., Frank, E., Güting, R. H., Han, J., Jiang, X., . . . Neapolitan, R. E. (2008). *Data mining: know it all*: Morgan Kaufmann.

Chen, H., Hu, Z., Wang, P., Xu, W., & Hou, Y. (2020). *Application of spectral droplet analysis method in flammable liquids identification.* Paper presented at the 2019 International Conference on Optical Instruments and Technology: Optical Sensors and Applications.

Chen, Q., Kang, G., Zhou, T., & Wang, J. (2017). *Fire hazard analysis of alcohol aqueous solution and Chinese liquor based on flash point.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Cheremisinoff, N. P. (1999). *Handbook of industrial toxicology and hazardous materials*: CRC Press.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*, 6.

Doad, P., & Bartere, M. (2013). A Review: Study of Various Clustering Techniques. *International Journal of Engineering Research & Technology, 2*(11), 3141-3145.

Hayasaka, T., Lin, A., Copa, V. C., Lopez, L. P., Loberternos, R. A., Ballesteros, L. I. M., . . . Lin, L. (2020). An electronic nose using a single graphene FET and machine learning for water, methanol, and ethanol. *Microsystems & Nanoengineering, 6*(1), 1-13.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence, 22*(1), 4-37.

Jawad, H., Lanteri, J., Migliaccio, C., Pichot, C., Platt, I. G., Tan, A. E.-C., . . . Woodhead, I. M. (2017). *Microwave modeling and experiments for non destructive control improved quality of fruit.* Paper presented at the 2017 IEEE Conference on Antenna Measurements & Applications (CAMA).

Jepsen, P. U., Jensen, J. K., & Møller, U. (2008). Characterization of aqueous alcohol solutions in bottles with THz reflection spectroscopy. *Optics express, 16*(13), 9318-9331.

Jiang, Y., Ju, Y., & Yang, L. (2016). Nondestructive in-situ permittivity measurement of liquid within a bottle using an open-ended microwave waveguide. *Journal of Nondestructive Evaluation, 35*(1), 7.

Jose, K., Varadan, V., & Varadan, V. (2001). Wideband and noncontact characterization of the complex permittivity of liquids. *Microwave and Optical Technology Letters, 30*(2), 75-79.

Kresse, W., & Danko, D. M. (2012). *Springer handbook of geographic information*: Springer Science & Business Media.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4): John Wiley & Sons.

Li, Z., Haigh, A., Soutis, C., Gibson, A., & Sloan, R. (2017a). Evaluation of water content in honey using microwave transmission line technique. *Journal of Food Engineering, 215*, 113-125.

Li, Z., Haigh, A., Soutis, C., Gibson, A., & Sloan, R. (2017b). Microwaves sensor for wind turbine blade inspection. *Applied Composite Materials, 24*(2), 495-512.

Li, Z., Haigh, A., Soutis, C., Gibson, A., & Sloan, R. (2018). A simulation-assisted non-destructive approach for permittivity measurement using an open-ended microwave Waveguide. *Journal of Nondestructive Evaluation, 37*(3), 1-10.

Moghadas, H., & Mushahwar, V. K. (2018). Passive microwave resonant sensor for detection of deep tissue injuries. *Sensors and Actuators B: Chemical, 277*, 69-77.

Orachorn, P., Chankow, N., & Srisatit, S. (2019). An Alternative Method for Screening Liquid in Bottles at Airports Using Low Energy X-ray Transmission Technique. *Radiation environment and medicine: covering a broad scope of topics relevant to environmental and medical radiation research, 8*(2), 77-84.

Rey, T., Kordon, A., & Wells, C. (2012). *Applied data mining for forecasting using SAS*: SAS Institute.

Saçlı, B., Aydınalp, C., Cansız, G., Joof, S., Yilmaz, T., Çayören, M., . . . Akduman, I. (2019). Microwave dielectric property based classification of renal calculi: Application of a kNN algorithm. *Computers in biology and medicine, 112*, 103366.

Slaughter, R., Mason, R., Beasley, D., Vale, J., & Schep, L. (2014). Isopropanol poisoning. *Clinical toxicology, 52*(5), 470-478.

Tan, X., Huang, S., Zhong, Y., Yuan, H., Zhou, Y., Xiao, Q., . . . Qi, C. (2017). *Detection and identification of flammable and explosive liquids using THz time-domain spectroscopy with principal component analysis algorithm.* Paper presented at the 2017 10th UK-Europe-China Workshop on Millimetre Waves and Terahertz Technologies (UCMMT).

Venkatesh, M., & Raghavan, G. (2005). An overview of dielectric properties measuring techniques. *Canadian biosystems engineering, 47*(7), 15-30.

Wirasuta, I. M. A. G., Dewi, N. K. S. M., Purwaningsih, N. K. P. A., Heltyani, W. E., Aryani, N. L. P. I., Sari, N. M. K., . . . Ramona, Y. (2019). A rapid method for screening and determination test of methanol content in ethanol-based products using portable

Raman spectroscopy. *Forensic Chemistry, 16*, 100190.

Yurchenko, A. V., Novikov, A., & Kitaeva, M. V. (2012). A resonator microwave sensor for measuring the parameters of Solar-quality silicon. *Russian journal of nondestructive testing*, *48*(2), 109-114.