


Use of Reflection Coefficients and Decision Tree Algorithm for Rapid Classification of Hazardous Chemical Liquids

¹Ebru Efeoglu, ^{*2}Gurkan Tuna

¹Kütahya Dumlupınar University, Faculty of Engineering, Department of Software Engineering, Kütahya
ebru.efeoğlu@dpu.edu.tr, 

^{*2}Trakya University, Edirne Vocational School of Technical Sciences, Department of Computer Technologies, Edirne
gurkantuna@trakya.edu.tr, 

Abstract

The purpose of occupational health and safety studies is to protect employees from work accidents and occupational diseases and to ensure that they work in a healthy environment. Most of the work accidents happen as a result of wrong storage, transportation and use of chemicals. In order to protect employees from chemical hazards and eliminate their possible risks, a risk assessment should first be carried out. Based on the results of the risk assessment, if not done before, bottles that contain hazardous chemicals must be classified and labelled according to their risk levels. The labels of bottles that contain chemical liquids must be checked and if the labels are worn or unreadable, they must be renewed. After these have been done, hazardous chemical liquids must be classified, stored and transported according to these labels. In this study, a non-contact, liquid measurement system based on microwave data is proposed to detect hazardous liquids. In order to select the most suitable algorithm for use in this measurement system, 3 different classification algorithms have been used and the performance analysis of the algorithms has made. In the study, 3 different classification processes have been applied according to the chemical properties of the liquids. It has been observed that Random Tree algorithm has achieved the best performance while Rep Tree algorithm has done the worst performance. Using this system, hazardous chemical liquids can be detected without opening the cover of the bottles that contain the liquids. Therefore, it can be used to quickly label hazardous liquids for their safe storage and transportation.

Keywords: Decision Tree Algorithm, Hazardous Chemicals, Microwave, Occupational Safety, Patch Antenna

1. INTRODUCTION

Most work accidents occur when the hazards of a working environment are not properly managed, either because they are not perceived as risks before the work starts, or because they are not fully understood. Therefore, appropriate training should be given to new employees to recognise and understand the hazards [1]. It has been reported that young workers have a higher non-fatal accident rate than older workers [2,3]. In order to create and maintain a safe working environment, hazards that may occur at workplaces should be identified and it is necessary to effectively monitor potential risks and give priority to preventive activities [4].

Although chemicals play an important role in the daily life of people around the world [5,6], some of the work accidents occur due to the chemicals. The chemical industry produces

a wide variety of substances necessary for daily use and chemical hazards generally arise from chemical synthesis or production, processing, transportation and misuse. Compared to the other substances, explosive, flammable and poisonous ones are more dangerous when released inappropriately [7]. It is well-known that the most dangerous of work accidents are explosions and fires that occur during the storage and transportation of petroleum and petroleum products, as well as other types of fuel [8,9]. On the other hand, there are ones like chemical liquids, which are less found in daily life [10]. Chemical liquids are often used in laboratories. Laboratories are one of the places where the danger and thus risks are almost the highest in terms of chemical hazards. Since flammable, explosive and poisonous chemicals are stored in chemical laboratories, they have many potential work accident risks. In addition, there is a higher risk of fire and explosion in them due to

* Corresponding Author

electrical equipment used in chemical experiments, high temperature and pressure.

Chemical liquids can be divided into 5 main groups as easily flammable liquids, corrosive liquids, toxic liquids, oxidising liquids and explosive liquids as given in Figure 1 [11,12]. Different measures have been proposed to prevent explosions related to the use of chemical liquids [12]. Figure 1 gives general information that indicates which class should be stored together or not. As it is known, the most hazardous chemicals encountered in major work accidents are sulphuric acid, hydrochloric acid and ammonia [11]. If oxidising chemicals come into contact with combustible materials, fire or explosion may occur. In the event of impact or spillage, these materials may be mixed with organic materials and fire and explosion may occur. These chemicals are also sensitive to heat and should be stored in a cool place. The flash point of a volatile substance is the lowest temperature at which the vapour of the substance will ignite when a source of ignition is given. In most cases, as the flash point decreases, the relative danger of a flammable liquid increases. Most of the solvents available in laboratories have flash points well below room temperature. Therefore, their proper use, storage and disposal is highly critical.














It has been suggested to use different techniques such as nuclear magnetic resonance and X-ray [13, 14], NQR method [15] for the detection of hazardous liquids. X-ray systems are the most widely used among these techniques [14]. In addition, liquid detection and identification can be performed using THz time domain spectroscopy [16]. However, these methods are generally effective in detecting peroxide-based liquids. And they cannot distinguish many types of liquids used in daily life. Therefore, there is a need for a system to distinguish these liquids [17].

The proposed system is based on microwave and decision trees. There are many microwave measurement methods. Of these methods the coaxial probe method is commonly used in microwave based systems. It has been used for different purposes including the examination of the microwave absorbing properties of ionic liquids at room temperature [18], electromagnetic properties of water and selected fruits and vegetables [19] and dielectric measurement of biological tissues [20]. In order to make measurements with this technique, the probe must be touched on the solid material in solids and the probe must be immersed in the liquid for liquid measurements. Metamaterial-based compact microwave liquid sensor was proposed for dielectric characterization of liquids [21].

In this study, a non-contact, a liquid classification system that can quickly detect hazardous liquids is presented. The classification system can classify liquids contained in bottles without the need of opening the caps of the bottles and without the possibility of being exposed to any breathing and skin contact with the liquids. It is a microwave-based system and based on the use of well-known Decision Tree algorithm family for classification. In the study, three different classification processes were applied to liquids according to

their chemical content and intended use. In the classification studies, 3 different decision tree algorithms were used and their performances were compared. As a result of the performance analysis study, it was understood that Random Tree algorithm was the most successful algorithm.

The rest of this paper is as follows. Experimental setup, methodology used in this paper and Decision Tree algorithms are given in Section 2. Section 3 presents experimental study and its results. Finally, the paper is concluded in Section 4.

					
Flammable	 +	-	-	-	+
Explosive	-	 +	-	-	-
Toxic	-	-	 +	-	+
Oxidising	-	-	-	 +	 ○
Irritant harmful	 +	-	 +	 ○	+

- Can not be stored
+ Storable
○ Can be stored with special precautions

Figure 1. Chemical substance storage matrix.

2. MATERIALS AND METHODS

The experimental setup used in this study for liquid classification using microwave patch antenna is shown in Figure 2. It consists of a microwave circular patch antenna design connected to a vector network analyser in order to measure of the reflection coefficient of electromagnetic wave. To build the experimental setup, an antenna with a resonant frequency of 1.5 GHz was designed. The design was constructed on a FR4 based dielectric substrate with 1.6 mm height, 4.4 relative permittivity and 10x10 cm² ground plane beneath it. The antenna is feed by 50 Ohm SMA (Sub Miniature Version A) feed probe. The design and geometry of the antenna is illustrated in Figure 3. The electromagnetic wave reflection coefficient of the liquids was measured by placing the bottle on the antenna. For each liquid measurement, the reflection coefficient was recorded at 40 measurement points between 1.52-1.42GHz frequency band. The dataset, consisting of 31 liquid measurements, formed a 31x40 matrix in total. Then, this data matrix was given as input to classification algorithms.



Figure 2. Experimental setup for liquid classification.

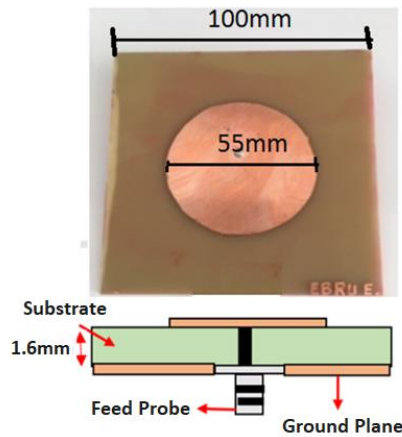


Figure 3. Design and geometry of the antenna.

Diameter of the antenna is calculated using (1) and (2).

$$F = \frac{8,791 \times 10^9}{f_r \sqrt{\epsilon_r}} \quad (1)$$

$$a = \frac{F}{\left\{ 1 + \frac{2h}{\pi \epsilon_r F \left[\ln \left(\frac{\pi F}{2h} \right) + 1,7726 \right]^{1/2}} \right\}} \quad (2)$$

where ϵ_r is relative permittivity of the substrate, f_r is the resonant frequency, h is the height of the substrate, and a is the radius of the patch.

2.1. Decision Trees

A decision tree structure consists of a root node containing data, inner nodes (branches) and end nodes (leaves). When decision tree algorithm is used, first, a decision tree is created, then the rules produced from the decision tree is used to classify the records in the database. By applying this created decision tree class on unknown data, the classes of this data are determined. In this tree structure, each node represents an attribute [22]. In Figure 4, a tree structure consisting of four dimensional attribute values belonging to three classes are shown. In this figure, x_i represents attribute values; The values a , b , c , d and e indicate the threshold values used for branching, and A , B and C indicate the class labels. There are univariate or multivariate decision tree structures according to the number of variables used in each stage of tree formation [23].

The most important step in creating decision trees is the selection of criteria that the branch of the tree will be made. Some of the approaches that can be used for this purpose are knowledge gain and knowledge gain rate [24], Gini index [25], Towing rule [24] and Chi Square probability table statistics [26]. The entropy method is used to determine which branch of the decision tree will be used in the use of information gain and information gain rate. Let C_1, C_2, \dots, C_n be a dataset consisting of several classes. If T shows class values, the probability of a class can be computed using (3).

$$P_i = (C_i / |T|) \quad (3)$$

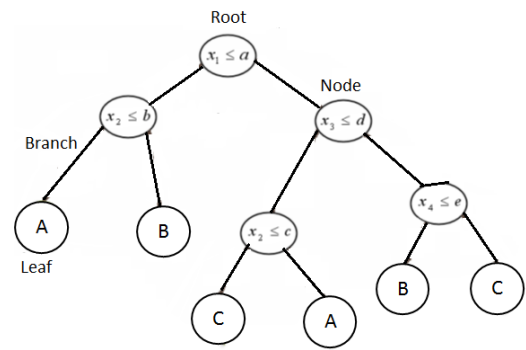


Figure 4. A decision tree structure consisting of three classes with four dimensional property spaces.

The entropy value of the classes is computed using (4).

$$Entropi(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (4)$$

T class values according to B attribute in the data set T_1, T_2, \dots, T_n . Considering that it is divided into sub-sets, the gain to be obtained by using B attribute values is computed using (5).

$$Gain(B, T) = Entropi(T) - \sum_T \frac{|T_i|}{|T|} Entropi(T_i) \quad (5)$$

Segmentation information for determining the value of B attribute for the T set is computed using (6).

$$B = - \sum_{i=1}^k \frac{|T_i|}{|T|} \log_2 \left(\frac{|T_i|}{|T|} \right) \quad (6)$$

Then, the gain rate can be computed using (7).

$$Gain\ rate = \frac{Gain(B, T)}{B} \quad (7)$$

Using this criterion, the T training set is separated repeatedly, with the maximum rate of gains at each node of the tree. This process is repeated until each leaf node contains observation values of only one class. The decision tree classification algorithm divides the training data into subsets containing only one class, resulting in a very large and complex tree structure. For this reason, it is possible to replace a subtree with a leaf. This process is called pruning and removes the parts of the decision tree that do not affect or contribute to the classification accuracy in order to obtain a more understandable tree structure [27]. Pre-pruning is done while creating the tree structure. On the other hand, post-pruning is done after the tree structure is created [24].

In this study, Random Tree, Rep Tree and Extra Trees algorithms are used for the classification process. Random Tree is a classification algorithm that creates a tree by taking a certain number of randomly selected properties in each node. There is no pruning. It also has an option that allows estimation of class possibilities based on the data set held.

Rep Tree algorithm was first proposed by Quinlan [27]. It is fast machine learning that creates a decision tree by pruning to reduce the effects of noise in the training examples [28]. With the regression tree logic, the best one is selected from the decision trees created after creating more than one tree in different iterations. The tree is pruned by reduced error pruning from the bottom up [29]. Because, the pruned tree reduces complexity in the classification process. The leaves of the decision tree always contain the exit values. The outputs at each branch point in the decision tree are selected based on the reduction of the "special threshold value" variance. To sum up, Rep Tree algorithm is based on the principle of minimizing the error caused by variance and the principle of gaining knowledge with entropy [29] and only works with numerical data. Extra Trees algorithm creates a community of decision trees that have not been pruned to the classic top-down procedure. The two main differences with

other tree-based community methods are that they select the breakpoints completely randomly, separating nodes and using the entire learning example to grow trees [30].

Decision tree algorithms are supervised learning algorithms. In this study, the purpose of using them is to predict which class the new liquid belongs to, using the model created from the data in the database when data from a new liquid arrives. Schematic representation of the classification processes held in this study is shown in Figure 5. As shown in this figure, in this study, 3 different classification experiments were carried out. In the first experiment, a total of 30 liquids listed in Table 1 were classified into two categories: hazardous or non-hazardous. Extra Trees, Random Tree and Rep Tree algorithms were used to classify the liquids given in Table 1.

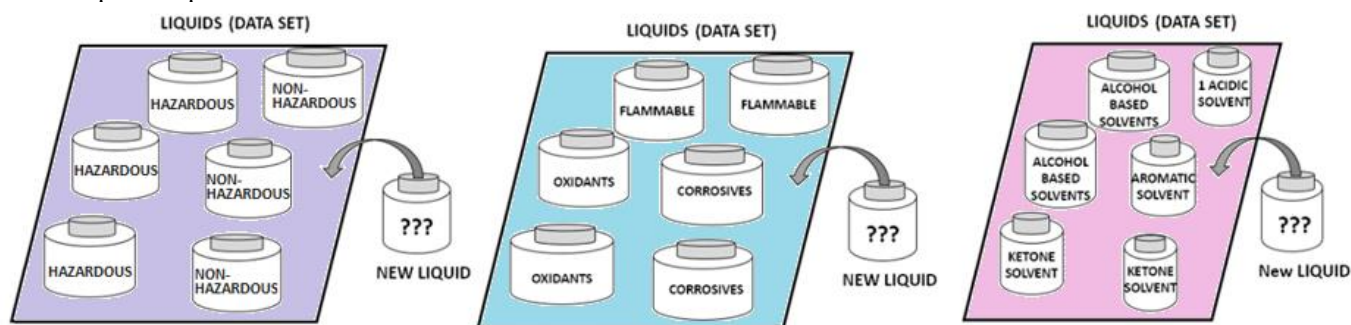


Figure 5. Classification of hazardous liquids.

3. RESULTS AND DISCUSSION

For the first classification experiment, a database was created from S_{11} parameter (reflection coefficient) measurements and two different methods were used to test the success of Extra Trees, Random Tree and Rep Tree algorithms. In the first method, all the data in the database was used as the training. In the second method cross validation was used. The purpose of doing this was to test the success of the algorithms in classifying liquids that were not in the database. Frequency-dependent reflection coefficient measurements of the liquids used in the classification experiment with the whole data set are given in Figure 6.

When Figures 7, 8 and 9 are considered, it can be seen that Extra Trees and Random Tree algorithms correctly classified all the liquids in the classification without cross validation but Rep Tree algorithm misclassified 3 liquids. When cross-validated, Extra Trees algorithm misclassified 3 of the liquids, Random Tree algorithm misclassified 2 of the liquids and Rep Tree algorithm did 6 of the liquids, respectively.

Other classification metrics of the algorithms are presented in Figure 10. As it can be seen, high accuracy, precision and recall values are obtained when Random Tree algorithm was used.

Hazardous chemical liquids can be generally divided into 3 groups as shown in Table 2. These are flammable (ethanol, methanol, acetone, 1-propanol, 2 propanol, butanol), corrosives (sulphuric acid, nitric acid, acetic acid), oxidants (hydrogen peroxide). In the second classification experiment, hazardous chemical liquids were divided based on the abovementioned groups. Cross validation was not used in the remaining experimental studies since the number of liquids was limited. In other words, the entire dataset was used in the classification process.

Frequency-dependent reflection coefficient measurements of the liquids used in the second classification experiment are given in Figure 11. Hazardous liquids were divided into 3 classes using Extra Trees, Random Tree and Rep Tree algorithms. The confusion matrices of the algorithms are presented in Figure 12. Other classification metrics of the algorithms are presented in Figure 13.

Table 1. Liquids used in the classification experiments.

Hazardous liquids			Non-Hazardous liquids		
Ethanol	1-propanol	Water	Turnip juice	Raki	Peach juice
Toluene	Methanol	Soda	Milk	Champagne	Ice-tea(Peach)
Butanol	Acetone	Beer	Liquid hand soap	Tequila	Apricot juice
Nitric acid	Acetic acid	Mineral water	Buttermilk	Whiskey	Vinegar
Sulphuric acid	Isopropanol	Rose juice	Gin	Vodka	Liqueur

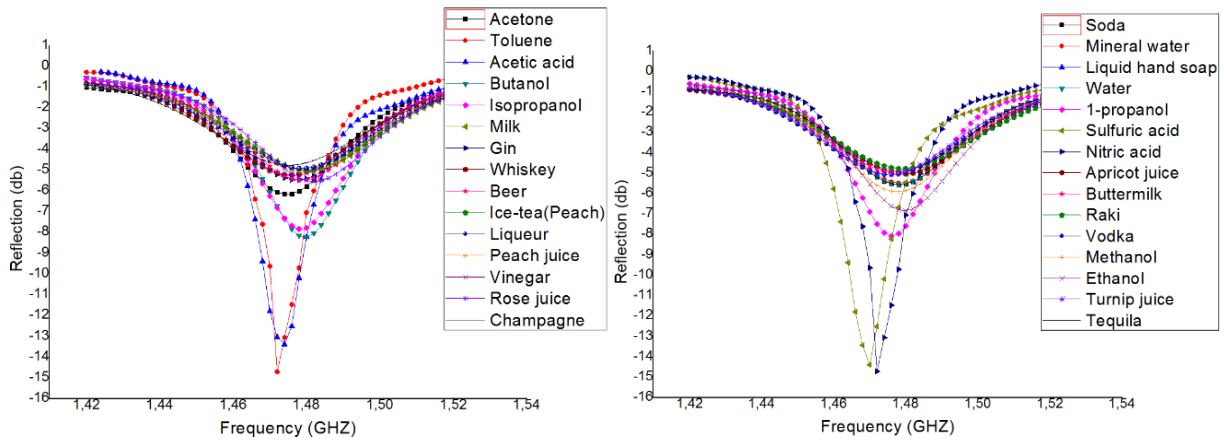


Figure 6. Reflection coefficient measurements of the liquids used in the first classification experiment.

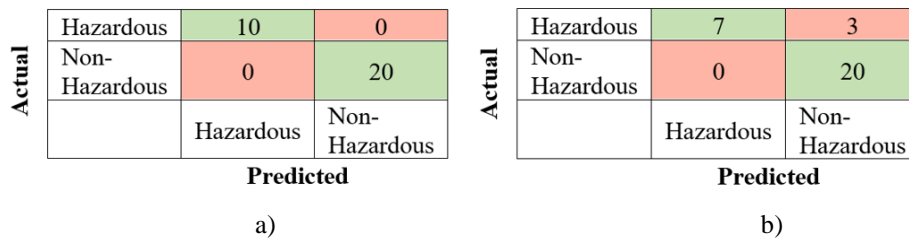


Figure 7. Confusion matrix of Extra Trees a) without cross validation b) after cross validation.

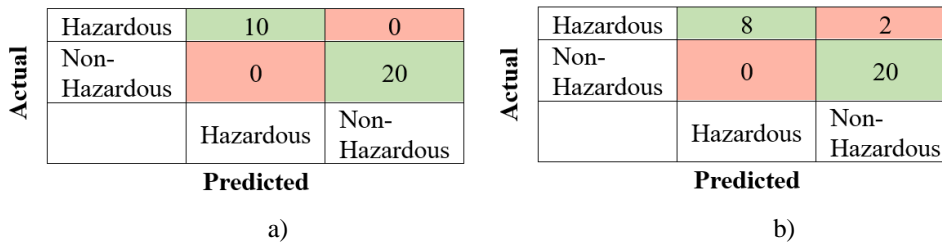


Figure 8. Confusion matrix of Random Trees a) without cross validation b) after cross validation.

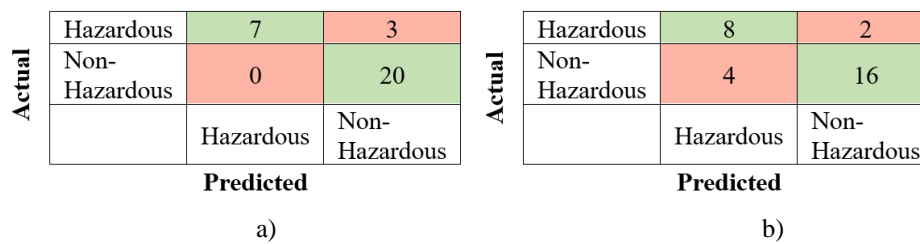


Figure 9. Confusion matrix of Rep Trees a) without cross validation b) after cross validation.

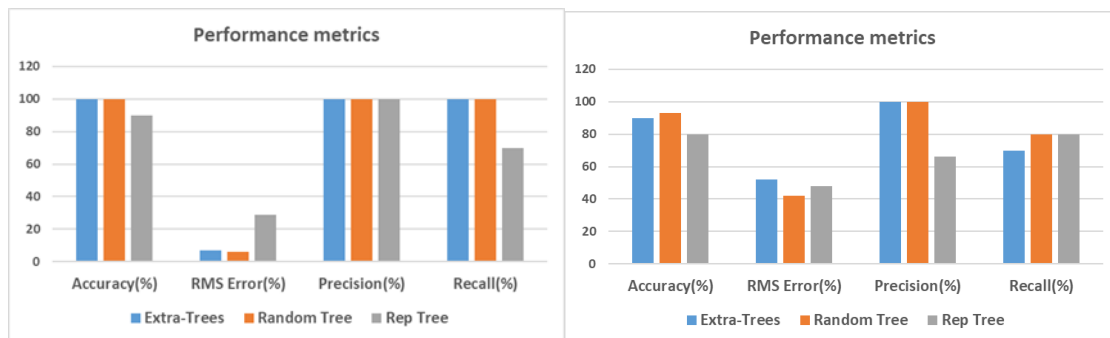


Figure 10. Performance metrics a) without cross validation b) after cross validation.

Table 2. Hazardous chemical liquids divided into 3 groups.

Flammable	Corrosives	Oxidants
Ethanol	Sulphuric acid	Hydrogen peroxide
Butanol	Acetic acid	
1-propanol	Nitric acid	
Methanol		
Acetone		
Isopropanol		

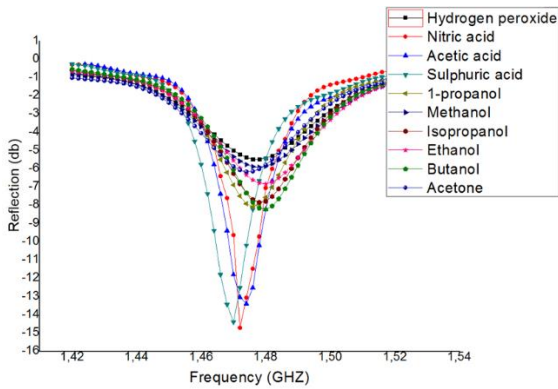


Figure 11. Reflection coefficient measurements of the liquids used in separating hazardous chemical liquids into 3 groups.

Table 3. Solvents used in the classification process.

Alcohol-Based Solvents	Ketone Solvents	Aromatic Solvents	Acidic Solvents
Ethanol	Acetone	Toluene	Acetic
Isopropanol			
Butanol			
1-propanol			
Methanol			

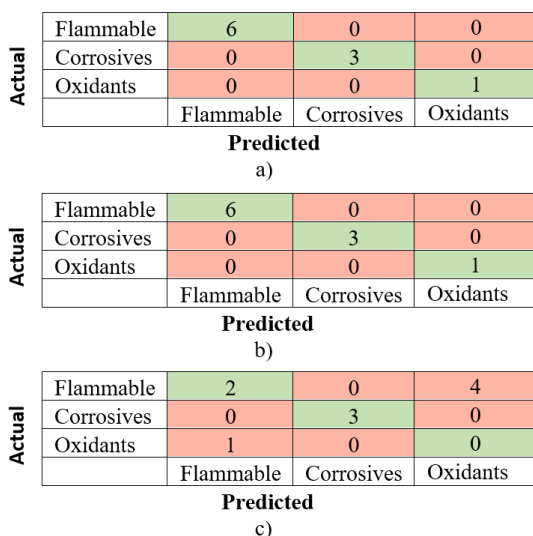


Figure 12. Confusion matrix a) Extra tree b) Random tree

c) Rep tree.

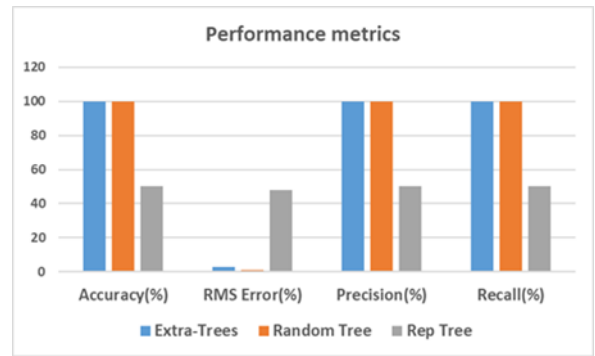


Figure 13. Performance metrics.

In this classification experiment, Extra Trees algorithm provided 100% accuracy, Random Tree did 100% accuracy, and Rep Tree algorithm did 50% accuracy. Random Tree had a better RMS error value compared to Extra Trees algorithm. Frequently used organic solvents both create risks for human and environmental health during their use and cause an environmental problem that is difficult to dispose when it becomes waste. However, there are situations where the use of solvents is mandatory. In these cases, it is necessary to choose the safest alternatives when choosing the solvent. As listed in Table 3, in the third classification experiment a total of 8, 5 of which are alcohol based solvents (Methanol, Ethanol, 1 propanol, 2-propanol, Butanol), 1 ketone solvent (Acetone), 1 aromatic solvent (Toluene) and 1 acidic solvent (acetic acid), were used and the solvents were divided into 4 main groups among themselves. Frequency-dependent reflection coefficient measurements of the liquids used in the third classification experiment are given in Figure 14. Confusion matrices and other performance metrics are presented in Figure 15 and Figure 16. In this classification experiment, Extra Trees algorithm achieved 100% accuracy, Random Tree achieved 100% accuracy, and Rep Tree algorithm did 62% accuracy. Random Tree had a better RMS value compared to Extra Trees algorithm. In this classification, Rep Tree algorithm had the highest error value and the worst performance metrics.

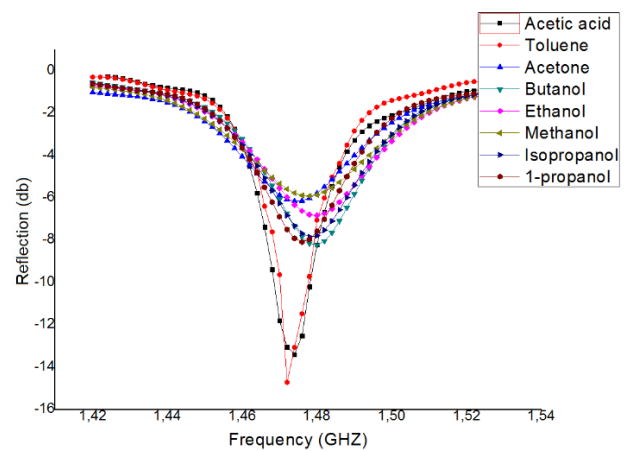


Figure 14. Reflection coefficient measurements of liquids used in the classification of solvent liquids

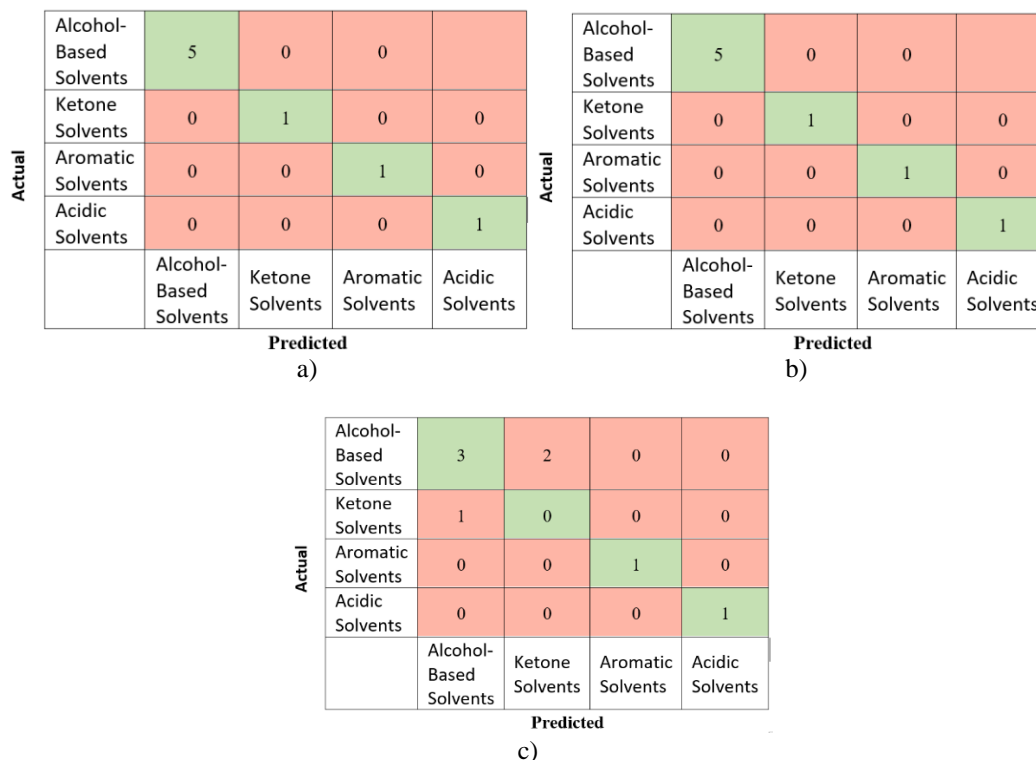


Figure 15. Confusion matrices a) Extra Trees b) Random Tree c) Rep Tree.

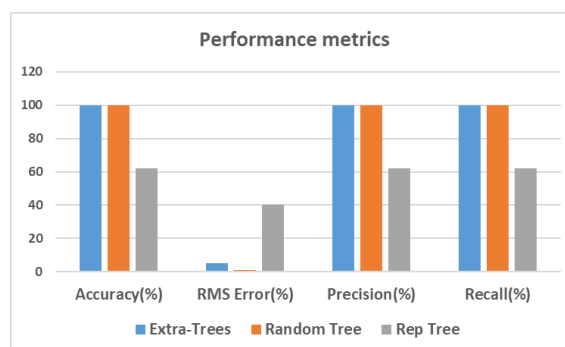


Figure 16. Performance metrics.

4. CONCLUSION

Thousands of people are at the risk of exposure to hazardous chemicals in their workplaces. If the potential risks associated with hazardous chemicals are not identified beforehand and not controlled by appropriate methods, serious work accidents, occupational diseases and even deaths can occur. On the other hand, if appropriate precautions are taken, both the health and safety of the workers can be secured and the loss of working days due to temporary and permanent incapacity or death can be prevented. In this study, a non-contact, liquid classification system that can detect hazardous liquids was presented. The classification system can quickly classify liquids contained in bottles without the need of opening the caps of the bottles. The classification system is based on a group of algorithms that are members of the well-known, Decision Tree algorithm family. The success of the algorithms was tested with a group of liquids and satisfactory results were obtained. Three classification experiments were carried out

and in each of them three different classification algorithms were used. In the first classification experiment, hazardous liquids and non-hazardous liquids were distinguished from

each other. In this classification, Extra Trees algorithm and Random Tree algorithm achieved 100% accuracy in the classification process when the whole data set was used. For liquids not found in the database, Random Tree algorithm performed better than Extra Trees algorithm with an accuracy rate of 80%. The worst result in this classification was obtained by Rep Tree algorithm. This algorithm obtained 90% accuracy in the classification when the whole dataset was used and 60% in the classification of liquids that are not in the database. In the second classification experiment, hazardous chemical liquids were divided into three groups: flammable, corrosive and oxidants. In this classification, Extra Trees and Random Tree algorithms classified the liquids with 100%.

The proposed system in the study is still under development. In the future study, the classification process will be applied by increasing the number of liquids in the database. After increasing the number of measurement data for liquids and applying many tests, the system can be used to ensure occupational safety.

Author contributions: Concept – E.E., G.T.; Data Collection and Processing - E.E.; Literature Search – E.E.; Discussion and Writing – E.E., G.T.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

- [1] R. H. Hill Jr, "Recognizing and understanding hazards—The key first step to safety," *Journal of Chemical Health and Safety*, vol. 26, no. 3, pp. 5-10, 2019.
- [2] S. Salminen, "Have young workers more injuries than older ones? An international literature review," *Journal of safety research*, vol. 35, no. 5, pp. 513-521, 2004.
- [3] A. Parent-Thirion et al., *Sixth European Working Conditions Survey: Overview Report*. Eurofound (European Foundation for the Improvement of Living and Working ...), 2016.
- [4] A. Byzov, A. Telegina, I. Korotkiy, and J. Veber, "Consequence assessment of explosions for fuel-air mixtures at hazardous production facilities," in *E3S Web of Conferences*, vol. 140: EDP Sciences, p. 08014, (2019).
- [5] J. Casal, *Evaluation of the effects and consequences of major accidents in industrial plants*. Elsevier, 2017.
- [6] A. Z. Mendiburu, J. A. de Carvalho Jr, and C. R. Coronado, "Method for determination of flammability limits of gaseous compounds diluted with N₂ and CO₂ in air," *Fuel*, vol. 226, pp. 65-80, 2018.
- [7] D. K. Horton, Z. Berkowitz, G. S. Haugh, M. F. Orr, and W. E. Kaye, "Acute public health consequences associated with hazardous substances released during transit, 1993–2000," *Journal of hazardous materials*, vol. 98, no. 1-3, pp. 161-175, 2003.
- [8] K. Kempna et al., "Fire Safety Protection Assessment of Industrial Technologies," in *Journal of Physics: Conference Series*, 2018, vol. 1107, no. 4: IOP Publishing, p. 042036.
- [9] A. Nikulin and A. Y. Nikulina, "Assessment of occupational health and safety effectiveness at a mining company," *Ecology, Environment and Conservation*, no. 23, p. 1.
- [10] C. Wei, W. J. Rogers, and M. S. Mannan, "Application of screening tools in the prevention of reactive chemical incidents," *Journal of Loss Prevention in the Process Industries*, vol. 17, no. 4, pp. 261-269, 2004.
- [11] T. Hoppe, N. Jaeger, and J. Terry, "Safe handling of combustible powders during transportation, charging, discharging and storage," *Journal of Loss Prevention in the Process Industries*, vol. 13, no. 3-5, pp. 253-263, 2000.
- [12] W. L. Welles, R. E. Wilburn, J. K. Ehrlich, and C. M. Florida, "New York hazardous substances emergency events surveillance: learning from hazardous substances releases to improve safety," *Journal of hazardous materials*, vol. 115, no. 1-3, pp. 39-49, 2004.
- [13] S. Kumar, "Liquid-contents verification for explosives, other hazards, and contraband by magnetic resonance," *Appl. Magn. Reson.*, 2004, vol. 25, nos. 3–4, pp. 585–597.
- [14] S. Singh and M. Singh, "Explosives detection systems (EDS) for aviation security," *Signal Process.*, vol. 83, no. 1, pp. 31–55, 2003.
- [15] L. Cardona, J. Jiménez and N. Vanegas, "Nuclear quadrupole resonance for explosive detection," *Ingeniare Revista chilena de ingeniería*, vol. 23, no. 3, pp. 458–472, 2015.
- [16] K. Choi, T. Hong, K. I. Sim, T. Ha, B.C. Park, J.H. Chung, et al. "Reflection terahertz time-domain spectroscopy of RDX and HMX explosives," *J. Appl. Phys.*, vol. 115, no. 2, p. 023105, 2014.
- [17] Z.Z. Abidin, F.N. Omar, P. Yogarajah, D.R.A. Biak, and Y.B.C. Man, "Dielectric characterization of liquid containing low alcoholic content for potential halal authentication in the 0.5-50 GHz range," *Am. J. Appl. Sci.*, vol. 11, no. 7, pp. 1104–1112, 2014.
- [18] F. Yang , J. Gong , E. Yang , Y. Guan , X. He , S. Liu, X. Zhang , Y. Deng, " Microwave-absorbing properties of room-temperature ionic liquids, " *Journal of Physics D: Applied Physics*, 52(15):155302, 2019.
- [19] A. La Gioia et al., "Open-ended coaxial probe technique for dielectric measurement of biological tissues: Challenges and common practices," *Diagnostics*, vol. 8, no. 2, p. 40, 2018.
- [20] T. Karpisz, B. Salski, P. Kopyt, J. Krupka, "Measurement of Electromagnetic Properties of Food Products and Liquids," In: 2018 12th International Conference on Electromagnetic Wave Interaction with Water and Moist Substances (ISEMA): 2018: IEEE: 1-9, (2018).
- [21] S. Kayal, T. Shaw, D. Mitra, "Design of metamaterial based compact and highly sensitive microwave liquid sensor," *Applied Physics A* ,126:13, 2020.
- [22] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote sensing of environment*, vol. 86, no. 4, pp. 554-565, 2003.
- [23] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote sensing of environment*, vol. 61, no. 3, pp. 399-409, 1997.
- [24] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.
- [25] W. Y. Loh, "Classification and regression trees," *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 1, no. 1, pp. 14-23, 2011.
- [26] J. Mingers, "An empirical comparison of pruning methods for decision tree induction," *Machine learning*, vol. 4, no. 2, pp. 227-243, 1989.
- [27] J. R. Quinlan, "Simplifying decision trees," 1986.
- [28] J. Li, S. Zhang, Y. Lu, and J. Yan, "Real-time P2P traffic identification," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, 2008: IEEE, pp. 1-5.
- [29] M. F. Amasyali and O. Ersoy, "Evaluation of regression ensembles on drug design datasets," 2009.
- [30] Y. Freund and L. Mason, "The alternating decision tree learning algorithm," in *icml*, 1999, vol. 99, pp. 124-133.