Nonlinear Dynamics and Systems Theory, **23** (3) (2023) 323–337

Electronic Nose for Classifying Civet Coffee and Non-Civet Coffee

D. B. Magfira [1], R. Sarno [2], F. A. Susanto [1], T. Herlambang [1], K. Oktafianto [3], W. Hartawan [4] and I. W. Farid [5*]

[1] *Information System Department, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[2] *Department of Informatics, Institute of Teknologi Sepuluh Nopember, Indonesia.*
[3] *Department of Mathematics, University of PGRI Ronggolawe, Indonesia.*
[4] *Department of Computer Science, University of Pat Petulai, Indonesia.*
[5] *Department of Electrical Automation Engineering, Vocational Faculty, Institute of Teknologi Sepuluh Nopember, Surabaya, Indonesia.*

Received: December 23, 2022; Revised: July 26, 2023

Abstract: Several Electronic Nose (E-nose) studies on coffee classification have been conducted. The E-nose uses gas sensors to detect the aroma of coffee and generate signals. Then the signals are classified using machine learning algorithms. In this study, the E-nose used five gas sensors to classify civet coffee and non-civet coffee, and the machine learning algorithms used were SVM, KNN and Decision Tree. The coffee variant used was Arabica coffee with the types of civet coffee (kopi luwak) and non-civet coffee (kopi non-luwak) originating from Aceh, Arjuno Malang, Bengkulu. In this study, the mixture of civet coffee and non-luwak coffee was made with a percentage of 100: 0, 90:10, 10:90, 80:20, 20:80, 75:25, 25:75, 50:50. The accuracy of the classification of Aceh civet coffee (LA) and Aceh non-civet coffee (NLA) was 90% (SVM), 100% (KNN), 100% (Decision Tree). The accuracy of the classification of Arjuno civet coffee (LAR) and Arjuno non-civet coffee (NLAR) was 100% (SVM, KNN, Decision Tree). The accuracy of the classification of Bengkulu civet coffee (LB) and Bengkulu non-civet (NLB) was 45% (SVM), 100% (KNN, Decision Tree). And the accuracy of coffee mixture classification (Aceh civet and Aceh non-civet) was 90% (SVM), 93.75% (KNN), and 95% (Decision Tree). The accuracy level obtained was affected by the age of coffee storage, the data collection process when detecting the coffee aroma, and the number of class attributes used.

Keywords: *E-nose; SVM; KNN; decision tree; civet coffee; non-civet coffee.*

Mathematics Subject Classification (2010): 62H30, 68W40, 68Q17.

* Corresponding author: mailto:wahyu@ee.its.ac.id

1 Introduction

Coffee is a plant that produces fruit that can be extracted into several types of products, one of the famous and easy-to-find coffee products are beverages. Coffee beans produced from coffee cherries go through several stages of processing [1]. In Indonesia, there are two variants of coffee commonly found, that is, arabica coffee and robusta coffee, and there are also two types of coffee called civet coffee resulted from the civet digestive process. Coffee beans that have gone through the civet digestion process can reduce their acid level so that the coffee beans produced become the best quality coffee [2]. In the process of odor stimulation, the gas-stimulating molecules are small or few in number. In this process, more decisive is not the amount of all gas entering through the nose but the number of gas molecules per unit time that touch odor sensitive cells in the nasal cavity [3]. Smells are usually produced from very low concentrations. The arising of the aroma of food and drink is caused by the formation of volatile compounds. The aroma that is released by every food and drink varies [4]. In addition, different cooking methods will cause different aromas. Likewise, coffee has a distinctive smell after passing through the roasting process [5].

Based on the raw material of each coffee cultivar, civet coffee has different chemical characteristics from the others. The chemical compounds contained in coffee beans greatly influence the taste and aroma of steeping coffee [6]. Therefore, the differences in chemical characteristics make the taste of Arabica civet coffee better than the taste of robusta civet coffee [7]. The coffee drink generally smells less when its temperature decreases because volatile substances will evaporate at high coffee temperatures. The sharp aroma of coffee can be smelled because coffee has polyphenol compounds [8]. The air containing volatile substances from a food will flow turbulently through the crevices of the nasal cavity. The gas molecules in the inhaled air stimulate and touch odor sensitive cells in the nasal cavity. The aroma of coffee can be detected by measuring the gas contained in it. The electronic nose technology can be used to detect the aroma present in coffee with the help of digital data from signaling that appears through the Arduino screen [9].

Classification research for the introduction of the coffee aroma is needed, especially in the introduction of civet coffee and non-civet coffee. This is followed by the increasing demand for the best arabica civet coffee from Indonesia, in meeting domestic and foreign market needs. With the high demand for arabica civet coffee, it is very susceptible to fraud committed by businessmen who expect greater profits, whereas the price of civet coffee is higher than the price of non-civet coffee. In this paper, the authors conducted a study of the types of civet arabica coffee and non-civet arabica coffee from three coffee-producing regions in Indonesia, namely, Aceh coffee, Malang coffee (Arjuno), Bengkulu coffee, the three types of coffee from these regions have not previously been studied simultaneously with arabica civet coffee and arabica non-civet coffee samples from Aceh, Malang (Arjuno coffee) and Bengkulu. This research detects coffee aroma with a series of gas sensors called the electronic nose, the E-nose circuit consists of five MQ-type gas sensors, they are MQ2, MQ3, MQ4, MQ7 and MQ135, from the aroma detection data shown in the displayed signal by the sensor via a monitor screen in the form of a digital signal. The data obtained from the detection of coffee aroma will be processed in main data to find information from the results of coffee aroma detection using the machine learning algorithm for data to be classified, and analyzed based on the accuracy produced so that to find the authenticity of arabica civet coffee and arabica non-civet coffee from the coffee samples used [10], [11], [12].

2 Research Method

The research aims to design an electronic nose used to detect the aroma of coffee. The variants used are arabica coffee with civet coffee and non-civet coffee originating from the three best coffee-producing regions in Indonesia: Aceh, Arjuno and Bengkulu. The aroma of coffee is detected with an electronic nose which generates digital data. The detected digital data are processed and tested in the classification process by applying machine learning classification algorithms by using Weka software.

2.1 Electronic nose design

The electronic nose circuit is designed with five MQ-type gas sensors, Arduino, a USB cable, a small fan and one tightly closed container. Table 1 shows the sensitivity of each sensor used in the study [13].

Sensor	Sensitivity to
MQ 2	Hydrogen, Methane, Alcohol, Propane, Butane
MQ 3	Alcohol, Methane, Benzine, Hexane, LPG, Carbon Monoxide
MQ 4	Methane, Hydrogen, Carbon Monoxide, Alcohol, Smoke
MQ 7	CO (Carbon Monoxyde)
MQ 135	Air Quality
	(SnO_2, Ammonia, Gasoline vapor, Sulfide, and other harmful gases)

Table 1: Types and sensitivity of sensors.

These five sensors can detect some of the same gases and what distinguishes them is the sensitivity level of each sensor to the gas, which can be detected based on sensor technical data [14]. In order to detect the aroma in coffee, all sensors are arranged in a circuit called the Electronic nose [15].

2.1.1 Electronic nose (E-nose)

The electronic nose (E-nose) is an instrument used to detect odors or aromas [16]. This system is built of an array of gas sensors known as the electronic olfactory system because the Electronic nose has the ability to imitate the work of the human sense of smell [17]. The output of the Electronic nose system is a signal forming patterns that represent each scent so that it is applicable for identification, comparison, quantification and classification based on the aroma [18], [19].

2.2 Coffee

Coffee is one of the largest plantation commodities in Indonesia, spread in the highlands. Coffee plants growing in the highlands produce the best quality coffee [20]. The regions producing the best quality coffee being in great demand on the international market and by domestic consumers include Gayo Arabica coffee from Aceh, Arjuno Arabica coffee from Malang and Kepahiang Arabica coffee from Bengkulu. The Gayo Highlands, situated at an altitude of 1200 meters above sea level, make coffee plants grow well, with the cool natural conditions of the Gayo Highlands giving an impact on the quality of the aroma and taste of the coffee produced [21]. Kepahiang is one of the districts in Bengkulu province with the majority of the population being coffee farmers. The coffee plantations in Kepahiang, situated at an altitude of 900 to 1300 meters above sea level, were originally planted with Robusta coffees only. With the increasing demand for Arabica coffee in the international market, farmers in Bengkulu were getting convinced to plant Arabica coffee, among them were those farmers in Kepahiang district. Currently, Arabica coffee in Kepahiang is very popular in the international market, especially on the European continent [22]. The foot of Mount Arjuna in Malang Regency, East Java, situated at an altitude of 900 to 1500 meters above sea level, is an ideal site for coffee plantations. The coffee produced on the slopes of Mount Arjuno in Karangploso sub-district-Malang Regency often wins the national and even international coffee competitions. Arabica coffee from the slopes of Mount Arjuno has a soft and fragrant taste due to the effect of volcanic soil and forest vegetation which is a heterogeneous forest. In addition, another very influential factor is the smell of sulfur from Mount Wilerang adjacent to Mount Arjuno so that different aromas and tastes give characteristics to Arjuno Arabica coffee [23].

2.2.1 Civet coffee and non-civet

Kopi luwak (civet coffee) is one of the typical Indonesian coffee products produced from the feces of a civet animal (Paradoxurus hermaphroditus) after the animal consumes ripe coffee cherries [24]. The luwak (civet) selects coffee cherries with an optimum maturity level based on taste and aroma, eats them by peeling the outer skin, then swallows the seeds and mucus. In the civet's digestive system, the coffee beans undergo a natural fermentation process at an optimal temperature level with the help of microbes and enzymes present in the civet's digestion. The fermentation process provides changes in the chemical composition of the coffee beans, which can improve the quality of the taste of Luwak coffee to be different from ordinary coffee so that Luwak coffee has a specific and special taste and aroma. The improvement in the taste quality of Luwak coffee is caused by the low protein content and high fat content compared to ordinary coffee. Low protein content can reduce bitter taste, while high fat content can increase body weight [25].

2.2.2 Data mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various databases [26]. The data mining process is done by applying a classification machine learning algorithm in recognizing information from a data [27]. The machine learning algorithms applied in this research are Support Vector Machine (SVM), k-nearest neighbor (KNN) and Decision Tree.

2.2.3 Support vector machine (SVM)

This method uses a two-step classification process. First, a kernel function transforms low-dimensional features into high-dimensional features [28]. Such transformation transforms non-linearly separable data into linearly separable data of higher dimensions. There are various kinds of kernels to use such as the Polynomial and Radial Basis Function

(RBF). The second step is to construct the maximum margin of the hyperplane to determine the decision limit for each class. The concept of maximum separation prevents misclassification of outliers, thus making the SVM a high-accuracy classification method [29]. In classifying the data, the data are divided into two types, that is, the training data and the test data, the training data set is labeled as $T = \{(x_b, l_i), i = 1, 2, \dots, L\}$ with $x_i \in R^P$ and $l_i \in \{-1, 1\}$, and the test data as $f(x) = sign \sum_{i=1}^{L} a_i \cdot l_i \cdot K(x_i, x) + b$, where a_i is the Lagrange multiplier, b is the limit value, K is the kernel function, so the SVM is a subset of the training data with $a_i > 0$ [30].

2.2.4 K-nearest neighbor (KNN)

The K-Nearest Neighbor (KNN) is a classification algorithm that will determine the label (class) of a test data based on many classes from the closest distance to k in the training data group [31]. The value of k used is 3 and 5 to be used in applying the KNN method, while the distance calculation uses the Euclidean Distance method [32]. The KNN will classify the test image into the class with the highest number of members [33]. The working principle of KNN is to find the shortest distance between the data to be evaluated and its k-closest neighbors in the training data [34]. The following is the equation for calculating the distance to the nearest neighbor: $D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ with x being the sample data, y being the test data, and D being the distance.

2.2.5 Decision tree (C4.5)

This algorithm has the input in the form of training samples and test samples [35], [36]. The training samples are in the form of sample data to be used to build a tree that has been tested for its credibility, while the test samples are data fields to be used later as parameters in classifying data [37]. In general, the C4.5 algorithm builds a decision tree following the steps as below: 1. Select an attribute as a root. 2. Create a branch for each value. 3. Divide cases in branches. 4. Repeat the process for each branch until all cases in the branch have the same class. Select the root attribute based on the highest gain value of the existing attributes. To calculate the gain, use the formula Gain (S.A) = Entropy (S), $Z_i = i$ Entropy (S_i), where S is the set of cases, A is the attribute, N is the number of partitions of the attribute $A, (S_i)$ is the number of cases in the i-th partition, and $|S_r|$ is the number of cases in S.

2.2.6 Confusion matrix

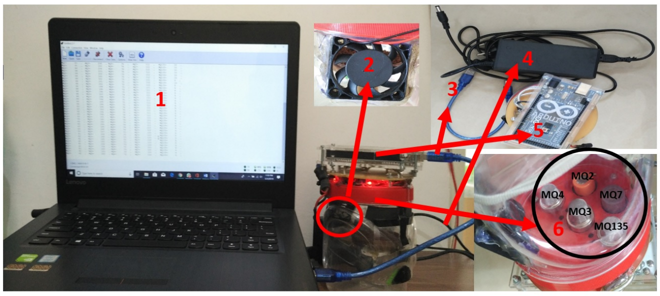
The confusion matrix is the most common way to show classification results, especially in multiclass data, to present it in the form of a confusion matrix also known as a contingency table [38]. For instance, $x_{r,c}$ of the confusion matrix $C \in N^{lxl}$ with r indicating the prediction class and c indicating the correct prediction results, and those outside the diagonal being the wrong prediction results. Table 2 shows the form of the matrix to be built in displaying the classification results [39].

3 Results and Analysis

This research aimed to design an Electronic nose and to perform the process of analyzing data from coffee aroma detection results by applying machine learning algorithms for recognizing and finding information from the data resulted from the detection of the

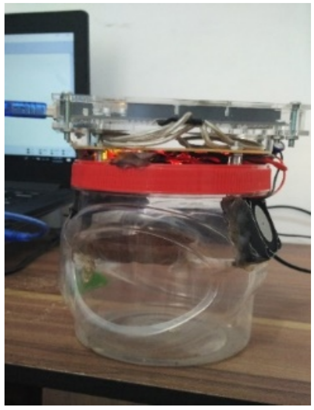
Prediction (r)	Original (c)			
	Class 1	Class 2	...	Class n
Class 1	X_{11}	X_{12}	...	X_{1n}
Class 2	X_{21}	X_{22}	...	X_{2n}
Class n	X_{n1}	X_{n2}	...	X_{nn}

Table 2: Confusion Matrix.



1. Computer 3. Cable USB 5. Arduino
2. Fam 4. Adapter 6. 5 Sensor (MQ2, MQ3, MQ4, MQ7 and MQ135

(a) Detailed Hardware Design.



(b) Finished Design

Figure 1: Design of the Electonic Nose.

aromas of the civet coffee and non-civet coffee originating from three coffee-producing regions in Indonesia: Aceh, Arjuno Malang and Bengkulu.

3.1 Hardware design of electonic nose

Five gas sensors (MQ2, MQ3, MQ4, MQ7 and MQ 135) were assembled into one above the Printed Circuit Board (PCB) and connected to the Arduino device so as to be able to read the aroma detection of coffee.

The hardware designed to detect the aroma of civet coffee and non-civet coffee is called the Electronic nose. The Electronic nose was connected to a computer and run on the Arduino IDE software, and it displayed the results of the coffee aroma detection on the cooltrem screen. The results of the coffee aroma detection displayed on the cooltrem screen can be seen in Figure 2.

Each line in Figure 2 displays the value of the detection results from the MQ2, MQ3, MQ4, MQ7, and MQ135 sensors every 2 seconds. The data collection in each experiment was done for 15 minutes.

The experiment conducted in this study was to detect the aroma of Arabica coffee variants with civet and non-civet coffee types from the best three coffee-producing areas: Aceh, Arjuno Malang, Bengkulu. This study was also conducted for an experimental mixture of Aceh civet coffee with Aceh non-civet coffee. The Aceh coffee mixture is divided into eight mixtures with the percentage of each mixture shown in Table 3. The experiments on each type of mixture, Aceh civet coffee with Aceh non-civet coffee, Arjuno civet coffee with Arjuno non-civet coffee, Bengkulu civet coffee with Bengkulu non-civet

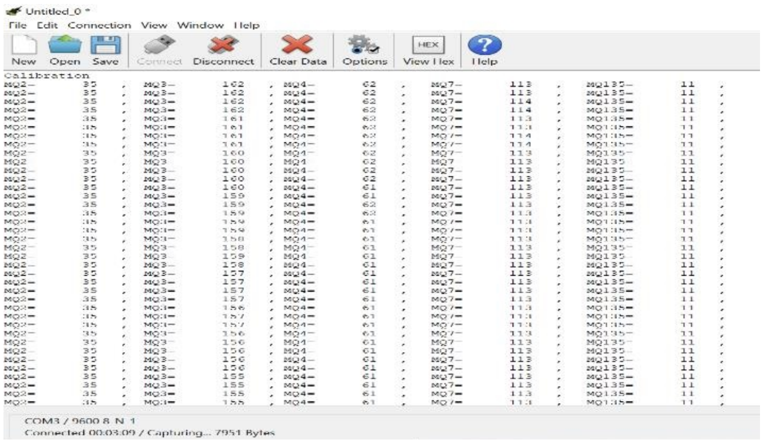


Figure 2: Digital data display of coffee aroma detection results.

No.	Aceh Civet	Aceh Non-civet	Class
1	100%	0%	L100NL0
2	90%	10%	L90NL10
3	10%	90%	L10NL90
4	80%	20%	L80NL20
5	20%	80%	L20NL80
6	75%	25%	L75NL25
7	25%	75%	L25NL75
8	50%	50%	L50NL50

Table 3: Percentage of Aceh Coffee Mixture.

coffee, and coffee mixture (Aceh civet with Aceh non-civet), are conducted as many as 50 times of data collection using the Electronic nose. One-time data collection is carried out for 15 minutes at room temperature. The coffee used as an experimental material is ground coffee with an ideal grinding level, from coarse to medium, with a coffee weight of 15 grams per data collection. The output of the coffee aroma detection produces a digital value from each sensor and is displayed on the cooltrem screen (Figure 2). To classify the coffee aroma detection data, the digital data will go through the process of calculating the average value and standard deviation. The calculations are made in each column based on the sensor name contained in the Electronic nose circuit. The calculation of the average value and standard deviation is made for all aroma detection results in each data sampling. The data from the calculation of the average value and standard deviation of each sensor can be seen in Figure 3, Figure 4, Figure 5, and Figure 6.

Figure 3 displays the results of calculating the average value and standard deviation of each sensor from the detection of Aceh Arabica coffee aroma. The Aceh Arabica coffee data are divided into two classes: Aceh civet (LA) and Aceh non-civet (NLA). Each class has 50 data, with the data range of 1-50 for the Aceh civet class (LA) and of 51-100 data for the Aceh non-civet class (NLA). Aceh Arabica coffee has 10 attributes: avrMQ2, avrMQ3, avrMQ4, avrMQ7, avrMQ135, stdMQ2, stdMQ3, stdMQ4, stdMQ7,

No	Class	avrMQ2	avrMQ3	avrMQ4	avrMQ7	avrMQ135	stdMQ2	stdMQ3	stdMQ4	stdMQ7	stdMQ135
1	LA	50.87755	99.14286	84.40816	175.5102	15.28571	0.917687	1.355262	0.924019	1.51351	0.494872
2	LA	54.72	108.12	91.86	186	15.94	1.990377	0.886341	5.321691	8.270429	0.732393
..	..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..	..
50	LA	51.2	107.2	87.42	177.2	15.86	1.296148	1.232883	3.826696	1.865476	0.824864
51	NLA	32.41	55.04333	42.49333	87.03	8.516667	1.114406	0.813299	0.499956	0.801519	0.499722
..	..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..	..
100	NLA	32.49	56.49	49.46333	99.53	10	0.4999	0.4999	0.498654	0.499099	0

Figure 3: Data of Aceh Arabica Coffee Aroma : Civet and Non-Civet.

stdMQ135. The attributes in the aroma data of Aceh Arabica coffee are the values from the calculation results of the average and standard deviation of five sensors in the Electronic nose circuit.

No	Class	avrMQ2	avrMQ3	avrMQ4	avrMQ7	avrMQ135	stdMQ2	stdMQ3	stdMQ4	stdMQ7	stdMQ135
1	LAR	29.87667	77.75	43.13333	90.39	9.006667	0.338805	3.778117	0.418994	0.760197	0.081377
2	LAR	31.00667	66.62667	45.54333	93.34667	9.5	0.553735	1.230429	1.004053	1.59786	0.5
..	..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..	..
50	LAR	29.49333	76.01667	45.06	92.66667	10.00333	0.499956	1.438653	1.369818	1.958457	0.806219
51	NLAR	30.47667	50.54667	43.49667	90.53667	10.95	0.499455	0.497817	0.499989	0.498654	0.232737
..	..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..	..
100	NLAR	30.44667	50.50333	43.48667	90.55333	159.5	0.497147	0.499989	0.499822	0.497147	86.60206

Figure 4: Data of Arjuno Arabica Coffee Aroma: Civet and Non-Civet.

The data of Arjuno civet Arabica coffee in Figure 4 has 2 classes and 10 class attributes. Each class has 50 data, 1-50 data for the Arjuno civet class (LAR) and 51-100 data for the Arjuno non-civet class (LAR). There are 10 attributes in Arjuno Arabica coffee aroma data, that is, avrMQ2, avrMQ3, avrMQ4, avrMQ7, avrMQ135, stdMQ2, stdMQ3, stdMQ4, stdMQ7, stdMQ135. These attributes are the results of the calculation of the average and standard deviation of five sensors in the Electronic nose circuit.

No	Class	avrMQ2	avrMQ3	avrMQ4	avrMQ7	avrMQ135	stdMQ2	stdMQ3	stdMQ4	stdMQ7	stdMQ135
1	LB	32.20667	90.53333	50.35667	112.88	9.776667	0.733	5.148031	1.917322	3.351457	0.41648
2	LB	47.11667	127.7767	108.2867	248.7133	16.98	0.754799	2.051046	1.031741	1.269313	0.14
..	..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..	..
50	LB	50.17667	106.3833	113.9667	204.2233	20.32	1.11001	2.820707	2.283029	3.415668	0.507543
51	NLB	41.53667	121.49	94.44	180.9833	13.52	0.498654	0.4999	0.496387	0.84245	0.4996
..	..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..	..
100	NLB	41.51333	121.52	94.44667	180.9933	13.49667	0.499822	0.4996	0.497147	0.808263	0.499989

Figure 5: Data of Bengkulu Arabica Coffee Aroma : Civet and Non-Civet.

Figure 5 displays the data of Bengkulu Arabica coffee aroma, divided into two classes, that is, Bengkulu civet (LB) and Bengkulu non-civet (NLB). Each class has 50 data with the order range of the data as in the table: 1-50 for the Bengkulu civet (LB) class and 51-100 for the Bengkulu non-civet class (NLB). There are 10 data attributes for each class: avrMQ2, avrMQ3, avrMQ4, avrMQ7, avrMQ135, stdMQ2, stdMQ3, stdMQ4, stdMQ7, stdMQ135. The class attributes are the result of the calculation of the average and standard deviation of five sensors in the Electronic nose circuit.

The coffee mixture of Aceh civet coffee and non-civet Aceh coffee, divided into eight mixture classes, has a total of 400 data. There are 50 data for each coffee mixture class.

No	Class	avrMQ2	avrMQ3	avrMQ4	avrMQ7	avrMQ135	StdMQ2	stdMQ3	stdMQ4	stdMQ7	stdMQ135
1	LA100NLA0	50.9	99.14	84.44	175.56	15.3	0.921954	1.34179	0.941488	1.512085	0.5
..	–	..	..	..	..	..	..	..	..	..	..
50	LA100NLA0	51.2	107.2	87.42	177.2	15.86	1.296148	1.232883	3.826696	1.865476	0.824864
51	LA90NLA10	67.34	87.91	125.64	197.73	10.37333	6.381567	3.662681	12.32114	14.81431	0.50394
..	–	..	..	..	..	..	..	..	..	..	..
100	LA90NLA10	67.34	87.91	125.64	197.73	10.37333	6.381567	3.662681	12.32114	14.81431	0.50394
101	LA10NLA90	39.7833	79.49	72.1466	163.51	8.673333	0.695022	2.787454	1.251062	2.385351	0.468994
..	–	..	..	..	..	..	..	..	..	..	..
150	LA10NLA90	54.1166	75.4866	105.916	187.056	9.956667	1.393935	0.499822	1.403468	0.836733	0.203606
151	LA80NLA20	48.13	83.1533	84.9333	148.696	7.46	1.089235	3.618907	2.552559	5.359539	0.498397
..	–	..	..	..	..	..	..	..	..	..	..
200	LA80NLA20	60.8233	79.7	114.113	186.103	9.95	0.803402	0.493288	0.860517	0.687499	0.217945
201	LA20NLA80	51.39	92.69	93.056	188.093	10.96667	0.48775	2.694544	0.812479	0.494593	0.179505
..	–	..	..	..	..	..	..	..	..	..	..
250	LA20NLA80	55.5266	95.5566	98.4566	198.093	10.45667	0.567998	0.535527	0.498119	2.037471	0.498119
251	LA75NLA25	67.8495	90.1170	128.337	211.645	10.99666	1.218897	2.032267	2.660804	2.623805	0.057735
..	–	..	..	..	..	..	..	..	..	..	..
300	LA75NLA25	59.5866	92.2666	116.39	200.07	9.993333	3.416015	2.833529	7.666675	11.41279	0.804128
301	LA25NLA75	55.5233	91.58	105.916	194.303	9.993333	0.550263	3.812733	1.209568	0.937011	0.081377
..	–	..	..	..	..	..	..	..	..	..	..
350	LA25NLA75	59.01	82.5866	122.026	210.933	10.98333	0.806164	1.087423	0.820135	0.825967	0.151841
351	LA50NLA50	47.9933	84.5566	88.5966	205.82	11	0.081377	2.499358	0.808696	1.381159	0
..	–	..	..	..	..	..	..	..	..	..	..
400	LA50NLA50	51.42	88.9833	97.8	219.2	11.49333	1.607358	2.050135	3.1241	4.693968	0.499956

Figure 6: Data of Aceh Arabica Coffee Aroma for 8 Mixtures of Civet and Non-Civet Coffee.

The data are: for the LA100NLA class, the order range is 1-50; for the LA90NLA10 class, the order range is 51-100; for the LA10NLA90 class, the order range is 101-150; for the LA80NLA20 class, the order range is 151-200, for the LA20NLA80 class, the order range is 201-250; for the LA75NLA25 class, the order range is 251-300; for the LA25NLA75 class, the order range is 301-350; and for the LA50NLA50 class, the order range is 350-400. Each class has 10 class attributes, that is, avrMQ2, avrMQ3, avrMQ4, avrMQ7, avrMQ135, StdMQ2, stdMQ3, stdMQ4, stdMQ7, stdMQ135. The class attribute is the result of the calculation of the average and standard deviation of five sensors in the Electronic nose circuit. The aroma data from each type of the regional coffee and coffee mixture were tested in the classification process by the SVM, KNN, and Decision Tree algorithms. The tests were done based on the attribute values existing in each class. The test data in the classification process were divided into two types, that is, the training data and the test data. The percentage distribution of the training data and the test data is 80% of the training data and 20% of the test data from the total data entered. The results of the tests carried out on the data for each regional coffee aroma (Aceh, Arjuno, Bengkulu) and coffee mixture (Aceh civet coffee with Aceh non-civet coffee) are shown in the tables below.

<table>
<tr><td rowspan="4">Prediction</td><td colspan="10">Target</td></tr>
<tr><td colspan="3">SVM</td><td></td><td colspan="2">KNN</td><td></td><td colspan="2">Decision Tree</td></tr>
<tr><td></td><td>LA</td><td>NLA</td><td></td><td>LA</td><td>NLA</td><td></td><td>LA</td><td>NLA</td></tr>
<tr><td>LA</td><td>10</td><td>1</td><td></td><td>11</td><td>0</td><td></td><td>11</td><td>0</td></tr>
<tr><td></td><td>NLA</td><td>1</td><td>8</td><td></td><td>0</td><td>9</td><td></td><td>0</td><td>9</td></tr>
</table>

Table 4: Confusion Matrix of Aceh Arabika Coffee.

The classification results on testing the aroma data of Aceh civet coffee (LA) with Aceh non-civet coffee (NLA) by the SVM classification algorithm are: 10 data classified according to the prediction class and target class and 1 data classified in Aceh non-civet target class (NLA) based on the prediction of the Aceh civet class (LA). The Aceh non-civet prediction class (NLA) classified 1 data in the target class of Aceh civet (LA) and 8 data in the target class of Aceh non-civet (NLA). The results of the classification by the SVM algorithm based on the prediction of the same class and target class have an accuracy of 90%. The classification results on testing the aroma data of Aceh civet coffee (LA) with Aceh non-civet coffee (NLA) by the KNN and Decision Tree algorithms are: 11 data classified in the same prediction class and target class in the Aceh civet class (LA). The classification according to the prediction of the Aceh non-civet class (NLA) indicates 9 data in the Aceh non-civet target class (NLA). The classification results displayed in the confusion matrix of the KNN and Decision Tree algorithms represent the data classified in each prediction class and target class having the same amount of data. The results of the classification by the KNN and Decision Tree algorithms show the same accuracy in each algorithm, which is 100%.

Prediction	Target								
	SVM				KNN			Decision Tree	
		LAR	NLAR		LAR	NLAR		LAR	NLAR
	LAR	11	0		11	0		11	0
	NLAR	0	9		0	9		0	9

Table 5: Confusion Matrix of Arjuno Arabica Coffee.

The results of the classification of Arjuno civet coffee aroma data (LAR) with Arjuno non-civet coffee (NLAR) by the SVM, KNN, and Decision Tree algorithms represent the amount of data classified by each algorithm, having the same amount, 11 data classified in the Arjuno civet target class (LAR) from the Arjuno civet prediction class (LAR), in the Arjuno non-civet prediction class (NLAR), 9 data classified in the Arjuno non-civet target class, so the accuracy of these three algorithms is the same, that is, 100%.

Prediction	Target								
	SVM				KNN			Decision Tree	
		LB	NLB		LB	NLB		LB	NLB
	LB	0	11		11	0		11	0
	NLB	0	9		0	9		0	9

Table 6: Confusion Matrix of Bengkulu Arabica Coffee.

The classification results on testing the Bengkulu civet coffee aroma data with Bengkulu non-civet coffee by the SVM algorithm show 11 data classified in the Bengkulu non-civet target class (NLB) from the Bengkulu civet prediction class (LB). The classified data from the prediction of the Bengkulu non-civet class with the target of the Bengkulu non-civet class indicate there are 9 data. In the Bengkulu civet prediction class (LB) with the same target class, there is no data classified in the Bengkulu civet class (LB), so the accuracy of the SVM algorithm is only 45%. Based on the confusion matrix of the classification results by the SVM algorithm, it shows that the test data entered in the classification process do not meet the right target class, this greatly affects the accuracy obtained from the classification results by the SVM algorithm. The results of testing data on the aroma of the Bengkulu civet coffee (LB) with the Bengkulu non-civet coffee (NLB) by the KNN and Decision Tree algorithms show the same amount of classified data, 11 data classified in the Bengkulu civet target class (LB) from the Bengkulu civet prediction class (LB), in the Bengkulu non-civet prediction class (NLB), 9 data classified, 9 in the Bengkulu non-civet target class (NLB). Thus, the KNN and Decision Tree algorithms show the same accuracy of 100%.

The classification results for the Aceh coffee mixture with 8 mixture classes and 10 class attributes, tested by the machine learning algorithms are as shown in the following table. The results of the Aceh coffee mixture classification (Aceh civet coffee and Aceh

		Target							
		LA100NLA0	LA90NLA10	LA10NLA90	LA80NLA20	LA20NLA80	LA75NLA25	LA25NLA75	LA50NLA50
Prediction	LA100NLA0	13	0	0	0	0	0	0	0
	LA90NLA0	0	7	0	2	0	1	0	0
	LA10NLA90	0	0	5	0	0	1	0	0
	LA80NLA20	0	0	3	8	0	0	0	0
	LA20NLA80	0	0	0	0	12	0	0	0
	LA75NLA25	0	0	0	0	0	11	0	0
	LA25NLA75	0	0	0	0	0	1	9	0
	LA50NLA50	0	0	0	0	0	0	0	7

Figure 7: Confusion Matrix of the Aceh Coffee Mixture by the SVM Algorithm.

non-civet coffee) show an accuracy of 90% by the SVM algorithm. The data tested were 80 data, and the classification results are presented in the form of a matrix (Table 4). The matrix formed based on the classification results shows the predicted distribution of data in each class. The data for the coffee mixture of the LA100NLA0 prediction class classified 13 data in the target class. In the prediction class of LA90NL10 there are 7 data classified according to the target class, 2 data classified in the target class of LA80NLA20 and 1 data classified in the target class of LA75NLA25. In the prediction class of LA10NLA90 there are 5 data classified according to the target class and 1 data classified in the target class of LA75NLA25. In the prediction class of LA80NLA20, there are 8 data classified according to the target class and 3 data classified in the target class of LA10NLA90. In the LA20NLA80 class there are 12 data classified with the right target class. In the LA75NLA25 class there are 11 data classified with the same target class. In the LA25NLA75 class there are 9 data classified according to the target class and 1 incoming data classified in the LA75NLA25 class. The LA50NLA50 class is classified according to the target class of 7 data. In the LA90NLA10 class classification, there are 2 data that fall into the LA80NLA20 target class and 1 data into the LA75NLA25 target class, meaning that the detection data among the coffee aromas from three types of mixtures have similarities in some of the data when detecting the coffee aroma. The similarity in the aroma detection data makes the data predictable in classes that are not in accordance with the target class. The mixture of 90% of Aceh civet coffee and 10% of Aceh non-civet coffee and that of 90% of Aceh civet coffee and 20% of Aceh non-luwak coffee have a similar aroma detected by the Electronic nose. The number of data samples used for each class can help minimize the occurrence of target class errors during the classification process.

The results of the coffee mixture classification by using the KNN algorithm are observed. For the prediction class of LA100NLAO, there are 13 data classified, predicted in accordance with the target class, and for the prediction class of LA90NLA10, there are 7 data classified matching the prediction with the target of the same class, 2 data in the target class of LA80NLA20 and 1 data in the target class of LA75NLA25. As predicted in accordance with the target class, 2 data are classified correctly, 2 data are classified into LA80NLA20 and 1 data is classified into LA75NLA25. LA10NLA90 is classified

	Target								
		LA100NLA0	LA90NLA0	LA10NLA90	LA80NLA20	LA20NLA80	LA75NLA25	LA25NLA75	LA50NLA50
Prediction	LA100NLA0	13	0	0	0	0	0	0	0
	LA90NLA0	0	7	0	2	0	1	0	0
	LA10NLA90	0	0	6	0	0	0	0	0
	LA80NLA20	0	0	0	11	0	0	0	0
	LA20NLA80	0	0	0	0	12	0	0	0
	LA75NLA25	0	0	0	0	1	10	0	0
	LA25NLA75	0	0	0	0	1	0	9	0
	LA50NLA50	0	0	0	0	0	0	0	7

Figure 8: Confusion Matrix of Aceh Coffee Mixture by the KNN Algorithm.

as 6 data correctly. 11 data are classified in LA80NLA20. 12 data are classified in the LA20NLA80 class. The LA75NLA25 class is classified as 10 data predicted correctly and 1 data included in the class of LA20NLA80. For LA25NLA75, there are 11 data classified correctly in accordance with the class, and 1 data is classified in LA20NLA80. LA50NLA50 has correctly predicted 7 data in accordance with its prediction class, then the accuracy of the classification by the KNN algorithm is 93.75% of the test input data amounting to 80 data.

	Target								
		LA100NLA0	LA90NLA0	LA10NLA90	LA80NLA20	LA20NLA80	LA75NLA25	LA25NLA75	LA50NLA50
Prediction	LA100NLA0	13	0	0	0	0	0	0	0
	LA90NLA0	0	8	0	2	0	0	0	0
	LA10NLA90	0	0	6	0	0	0	0	0
	LA80NLA20	0	0	0	11	0	0	0	0
	LA20NLA80	0	0	0	0	12	0	0	0
	LA75NLA25	0	0	0	0	0	11	0	0
	LA25NLA75	0	0	0	0	0	2	8	0
	LA50NLA50	0	0	0	0	0	0	0	7

Figure 9: Confusion Matrix of Aceh Coffee Mixture by the Decision Tree Algorithm.

The classification by the Decision Tree algorithm indicates LA100NLA0 classified as 13 data, LA90NLA10 predicted 8 data in accordance with the prediction class, 2 data predicted in the LA80NLA20 class. For LA10NLA90, 6 data are correctly predicted in the right class. 11 data are predicted in the LA80NLA20 class. 12 data are classified in the LA20NLA80 class. For the LA75NLA25 class, 11 data predicted are classified in accordance with the class. For the LA25NLA75 class, 8 data are classified in the right class and 2 data fall into the incorrect prediction in the LA75NLA25 class. And for the LA50NLA50 class, 7 data are classified as correct so that from all the correct class predictions, an accuracy of 95% is obtained.

After testing the aroma data of the civet and non-civet arabica coffee from each region (Aceh, Arjuno, Bengkulu) and of the coffee mixture (Aceh civet and Aceh non-civet), it is suggested that the data can be effectively classified by using the machine learning algorithms: SVM, KNN, and Decision Tree. During the research, the researcher observed that in detecting the coffee aroma, it is more effective if the coffee used is ground coffee with a medium-sized grinding level, thereby, to maximize the aroma detection stage because at the time of data collection, the coffee aroma lifted by a fan to the lip of the sensor does not leave coffee powder making it easier for further data collection, and with this medium size the aroma released by the coffee can be smelled well so that it is detected optimally. The storage age of coffee, especially ground coffee, greatly affects the level of aroma contained in the coffee. The longer the ground coffee is stored, the less the aroma contained in the coffee, thus reducing the taste of the coffee. During the collection of the data on the aroma of the civet coffee and non-civet coffee from three coffee producing regions (Aceh, Arjuno Malang, Bengkulu) and Aceh coffee mixture (Aceh civet and Aceh non-civet), the room temperature greatly affected the data produced. The time used to collect coffee aroma data also affects the final results of the data to be used, the longer the data collection process, the better data provided for classification. In data classifying, the data used greatly affects the resulting accuracy value, the more attributes used in the classification process, the higher accuracy value gained. The accuracy of the classification results greatly affected the number of class attributes used, it is suggested to also use the calculated values of min, max, mean, range, kurtosis and skewness.

4 Conclusion

In this research, an Electronic nose (E-nose) was developed with five MQ gas sensors (MQ2, MQ3, MQ4, MQ7 and MQ135) which can detect the aroma of each coffee variant. The results of the aroma detection were shown by sensor signals displayed in digital data. For the digital data resulted from each sensor, the average and standard deviations were calculated. The result of calculating the average and standard deviation of the detection results of each sensor was named the class attribute. The classification by using the SVM, KNN, and Decision Tree algorithms was based on the average value and standard deviation of each coffee variant. The classification results showed an accuracy of above 90% for all the variants of coffee mixtures, and there was still 1 variant having an accuracy of below 50%, that is, the classification of Bengkulu civet coffee with Bengkulu non-civet coffee. As seen from the average of all accuracy produced, the classification of the civet coffee with the non-civet was effectively done with the data resulted from the aroma detection carried out by the Electronic nose. In classifying data, the data used greatly affected the resulting accuracy, the more attributes used in the classification process, the higher accuracy value gained.

Acknowledgment

The authors would like to thank: Institute of Technology Sepuluh Nopember , Sekar Gayo coffee Bener Meriah - Aceh, Apresio Kopi Malang, Permata Sari Kepahiang Farmers Group - Bengkulu, who have supported the author in conducting this research. This research was supported by LPPM - Nahdlatul Ulama Surabaya of University (UNUSA). Constructive criticism and suggestions from readers are highly expected, to improve research and research achievements.

References

[1] M. I. Sofyan and T. E. Sebayang. Investigation of Coffee Export Dynamics in Indonesia. *Jurnal Bisma Unesa* **11** (1) (2018) 67–76.

[2] S. A. W. Kristanti and R. Sutamih. Tingkat Kematangan Biji Kopi Arabika (Coffee arabica L) dalam Menghasilkan Kadar Kafein. *Sain Natural* **9** (1) (2019) 22–28.

[3] D. R. Wijaya, R. Samo and E. Zulaika. Sensor Array Optimization for Mobile Electronic Nose: Wavelet Transform and Filter Based Feature Selection Approach. *International Review on Computers and Software (I.RE.CO.S.)* **11** (8) (2016) 659–671.

[4] D. R. Wijaya, R. Samo and E. Zulaika. Information Quality Ratio as a novel metric for mother wavelet selection. *Elsevier, Chemometrics and Intelligent Laboratory Systems* **160** (2017) 59–71.

[5] D. Giacalonea, T. D. Kreuzfeldt, N. Yang, C. Liu and I. Fisk. Common roasting defects in coffee: Aroma composition, sensory characterization and consumer perception. *Elsevier* (2018) 1–12.

[6] J. Towaha and B. T. Eka. Kopi Luwak Budidaya Sebagai Diversifikasi Produk Yang Mempunyai Citarasa Khas. *SIRINOV* **3** (1) (2015) 19–30.

[7] E. Wahyuni, A. Karim and A. Anhar. Analysis of Taste Quality of Organic Arabica Coffee in Several Altitudes and Processing Techniques in Gayo Highlands. *Jurnal Manajemen Sumberdaya Lahan* **2** (3) (2013) 261–269.

[8] A. T. Toci and M. V. Z. Boldrin. Coffee Beverages and Their Aroma Compounds. *Natural and Artificial Flavoring Agents and Food Dyes, Elsevier* (2018) 379–423.

[9] K. C. Hsu, T. H. Fang, Y. J. Hsiao and P. C. Wu. Response and Characteristics of TiO2/Perovskite Heterojunctions for CO Gas Sensors. *Journal of Alloys and Compounds* **1** (1) (2019) 2–39.

[10] R. Radi, S. Ciptohadijoyo, W. S. Litananda, M. Rivai and M. H. Purnomo. *Electronic Nose Based on Partition Column Integrated With Gas Sensor for fruit Identification and Classification.* Yogyakarta, Indonesia, 2016.

[11] U. Jumhawan, S. Prama Putri, Y. T. Bamba and E. F. Application of gas chromatography/flame ionization detector-based metabolite fingerprinting for authentication of Asian palm civet coffee (Kopi Luwak). *Journal of Bioscience and Bioengineering* **20** (20) (2015) 1–7.

[12] V. Vapnik. *Support Vector Networks.* Kluwer Academic Publishers, Boston. Manufactured in The Netherlands (1995) 273–297.

[13] F. R. Rozak. *Flowing System Design of Indirect Elektronic Nose to Tea Aroma Dedection.* Universitas Gadjah Mada, Yogyakarta, Indonesia, 2015.

[14] S. S. Izza, R. Sarno and J. Siswantoro. Estimating Gas Concentration using Artificial Neural Network for Electronic Nose. In: *Elsevier, 4th Information Systems International Conference (ISICO 2017)* Bali, Indonesia, 2017.

[15] H. Hanwei. Technical Data MQ135, MQ4, MQ7, MQ2 and MQ3 Gas Sensor. *http://www.hwsensor.com.* Hanwei Electronics Group Corporation, 1998.

[16] D. W. Rahman , R. Sarno and E. Zulaika. Gas Concentration Analysis of Resistive Gas Sensor Array. *International Symposium on Electronics and Smart Devices (ISESD)* (2017) 337–342.

[17] D. B. Magfira and R. Sarno. Classification Of Arabika and Robusta Coffee Using Elektronic Nose. In: *ICOIACT* Yogyakarta, Indonesia, 2018.

[18] S. Baskara, D. Lelono and T. W. Widodo. Pengembangan Hidung Elektronik untuk Klasifikasi Mutu Minyak Goreng dengan Metode Principal Component Analysis. *IJEIS* **6** (2) (2016) 221–230.

[19] R. Sarno and D. W. Rahman. Recent Development in Electronic Nose Data Processing for Beef Quality Assessment. *Telkomnika* **17** (1) (2019) 337–348.

[20] F. Chamim. *https://mlgcoffee.com/2014/09/19/definisi-kopi-dan-sejarah-penyebaran-kopi-di-dunia/.* 27 February 2014. [Online]. [Accessed 2019].

[21] Y. W. Teniro, Z. Zulfan and H. Husaini. Perkembangan Pengolahan Kopi Arabika Gayo Mulai Dari Panen Hingga Pasca Panen Di Kampung Simpang Teritit Tahun 2010-2017. *Jurnal Ilmiah Mahasiswa (JIM) Pendidikan Sejarah FKIP Unsyiah* **3** (3) (2018) 52–63.

[22] D. Listyati, B. Sudjarmoko and A. M. Hasibua. Farming Analysis and Marketing Chain of Robusta Coffee in Bengkulu. *Jurnal of Industrial and Beverage Crops* **4** (3) (2017) 145–154.

[23] T. Ada. *https://www.timesindonesia.co.id/read/187708/20181028/000141/kopi-arabika-arjuno-kopi-unggulan-dari-malang/*. 29 April 2019. [Online]. [Accessed 2019].

[24] M. F. Marcone. Composition and properties of Indonesian palm civet coffee (Kopi Luwak) and Ethiopian civet coffee. *Elsevier* Canada, 2004.

[25] R. Rubiyo and J. Towaha. The Effect of Fermentation on Flavor Quality of Probiotic Civet Coffee. *RISTI* **4** (2) (2013) 175–182.

[26] T. Herlambang, H. Nurhadi, A. Muhith, A. Suryowinoto, and K. Oktafianto. Estimation of Forefinger Motion with Multi-DOF Using Advanced Kalman Filter. *Nonlinear Dynamics and Systems Theory.* **23** (1) (2023) 24–33.

[27] A. Lukman and M. Marwana. Machice Learning Multi Klasifikasi Citra Digital. *Konfrensi Nasional Ilmu Komputer (KONIK)* Makassar, 2014.

[28] M. Jupri and R. Sarno. Taxpayer Compliance Classification Using C4.5, SVM, KNN, Naive Bayes and MLP. *ICOIATC* Yogyakarta, 2017.

[29] R. Sarno, J. A. Ridoean and D. Sunaryono. Classification of Music Mood Using MPEG-7 Audio Features and SVM with Confidence Interval. *International Journal on Artificial Intelligence Tools, World Scientific Publishing Company* **27** (5) (2018) 1–18.

[30] K. Oktafianto, A. Z. Arifin, E. F. Kurniawati, T. Herlambang and F. Yudianto. Tsunami Wave Simulation in the Presense of a Barrier. *Nonlinear Dynamics and Systems Theory.* **23** (1) (2023) 69–78.

[31] J. Gou, H. Ma, W. Ou, S. Zeng and Y. Rao. A Generalized Mean Distance-Based K-Nearest Neighbor Classifier. *ELSEVIER, Expert Systems with Applications* **115** (1) (2018) 356–372.

[32] K. Zhao and C. Huang. Air Combat Situation Assessment for UAV Based on Improved Decision Tree. *Chinese Control And Decision Conference (CCDC)* Shenyang, China, 2018.

[33] J. Gou, W. Qiu, Z. Yi and X. Shen. Locality Constrained Representation-Based K-Nearest Neighbor Classification. *Knowledge-Based Systems* **167** (1) (2019) 38–52.

[34] A. Utku, I. D. Alper and M. A. Akcayol. Decision Tree Based Android Malware Detection System. *Signal Processing and Communications Applications Conference (SIU)* Izmir, Turkey, 2018.

[35] H. Elaidi, Y. Elhaddar, Z. Benabbou and H. Abbar. An Idea of a Clustering Algorithm using Support Vector Machines Based on Binary Decision Tree. *International Conference on Intelligent Systems and Computer Vision (ISCV)* Fez, Maroko, 2018.

[36] M. Y. Anshori, I. H. Santoso, T. Herlambang, K. Oktafianto and P. Katias. Forecasting of Occupied Rooms in the Hotel Using Linear Support Vector Machine. *Nonlinear Dynamics and Systems Theory.* **23** (2) (2023) 129–140.

[37] E. S. Sankari and D. Manimegalai. Predicting Membrane Protein Types Using Various Decision Tree Classifiers Based on Various Modes of General PseAAC for Imbalanced Datasets. *Journal of Theoretical Biology* **435** (12) (2018) 208–217.

[38] A. Luque, A. Carrasco, A. Martin and A. H. Las. The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix. *Pattern Recognition* **91** (6) (2019) 216–231.

[39] S. Ruuska, W. Hamalainen, S. Kajava and M. Mughal. Evaluation of the Confusion Matrix Method in the Validation of an Automated System for Measuring Feeding Behaviour of Cattle. *Behavioural Processes* **148** (4) (2018) 56–62.