TELKOMNIKA, Vol.17, No.6, December 2019, pp.3073~3085
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v17i6.12689

Computer vision for purity, phenol, and pH detection of *Luwak* Coffee green bean

Yusuf Hendrawan*[1], Shinta Widyaningtyas[2], Sucipto Sucipto[3]
[1]Department of Agricultural Engineering, Agricultural Technology Faculty, Universitas Brawijaya, Veteran St., Malang, Indonesia, telp/fax: +62-341-580106/+62-341-568917
[2]Department of Agroindustrial Technology, Agricultural Technology Faculty, Universitas Brawijaya, Veteran St., Malang, Indonesia
[3]Halal and Qualified Industry Development (HAL-Q ID), Universitas Brawijaya, Veteran St., Malang, Indonesia
*Corresponding author, e-mail: yusufhendrawan@gmail.com[1], shintawidya1435@gmail.com[2], ciptotip@ub.ac.id[3]

Abstract

*Computer vision as a non-invasive bio-sensing method provided opportunity to detect purity, total phenol, and pH in Luwak coffee green bean. This study aimed to obtain the best Artificial Neural Network (ANN) model to detect the percentage of purity, total phenol, and pH on Luwak coffee green bean by using color features (red-green-blue, gray, hue-saturation-value, hue-saturation-lightness, L*a*b*), and Haralick textural features with color co-occurrence matrix including entropy, energy, contrast, homogeneity, sum mean, variance, correlation, maximum probability, inverse difference moment, and cluster tendency. The best ANN structure was (5 inputs; 30 nodes in hidden layer 1; 40 nodes in hidden layer 2; and 3 outputs) which had training mean square error (MSE) of 0.0085 and validation MSE of 0.0442.*

Keywords*: artificial neural network, computer vision, Luwak coffee*

1. Introduction

In the last two decades, global coffee consumption growth has continued to grow, as driven by coffee-based products and beverage formulations and the increasing number of coffee shops [1]. One type of coffee, known to be expensive and rare in the world is *Luwak* (civet) coffee [2]. As a high-priced commodity, *Luwak* coffee is prone to be mixed with regular coffee beans. At present, an internationally recognized method of distinguishing *Luwak* and regular coffee, remains unnoticed. This, therefore gives the opportunity to design a simple, fast, accurate, and non-destructive equipment, capable of detecting the percentage of mixed portion between *Luwak* coffee and regular coffee. The study results of Jumhawan [3] found out that the tastes of roasted *Luwak* coffee and regular roasted coffee are citric acid and malic acid. Research on the *Luwak* coffee green bean has never been conducted, albeit about 75% of Indonesian coffee exports are in the form of green bean. In addition to detecting *Luwak* coffee mixtures in regular coffee, this study also measures total phenol as an antioxidant and pH to measure the coffee acidity. Coffee becomes a source of antioxidants to ward off free radicals that are beneficial for health. The largest antioxidant component in coffee is phenol [4, 5]. At present, consumption of green bean extract becomes a new trend due to its low calorie content [6]. Measuring total phenol in green bean, helps measure the antioxidant activity. In addition, coffee has an acidic taste that is identical to its pH content. The trend of consuming green beans extract requires a study of pH, due to consumer sensitivity of coffee acidity, especially in arabica coffee. This research is utilized as one of the stages in designing tools for coffee inspection. Computer vision technology has been widely applied in identifying and copying coffee as an example of research as conducted by Oliveira [7], applying computer vision and computational intelligence to classify green bean coffee. The results show the performance of computer vision which achieves classification accuracy of up to 100%. Nansen [8] applied computer vision by using hyperspectral imaging to identify commercial roasted coffee brands based on their quality. Caporaso [9] detect moisture content in single green bean coffee by using computer vision. The results show optimal results for moisture

Received March 15, 2019; Revised July 2, 2019; Accepted July 18, 2019

content detection in single green bean coffee and successfully classify the types of coffee (Arabica and Robusta). Navarro [10] employed digital imaging technology to model the quality of coffee during the roasting process. The results present good performance in using a combination of digital imaging with adaptive network based fuzzy inference systems (ANFIS) to monitor coffee color during the roasting process. The use of artificial intelligent modeling such as artificial neural network (ANN) has been successfully applied in various coffee identification studies [11, 12]. However, there have been no studies that have observed light computer vision and artificial intelligent modeling performance to identify the purity of green bean coffee for *Luwak* coffee type.

Image analysis is identified as a fast, non-destructive and low-cost method for assessing the quality of food products [13, 14]. According to Patel [15], machine vision development is based on the inspection of the food quality and agricultural products, unfortunately faced several obstacles which later requires such an accurate, fast and objective technique in determining the quality of the measured material. This technology appears in the development of automated machinery in the agriculture and food industries [16]. Several studies [17-20] depict optimal results in machine vision application when using a combination of ANN modeling with color features (RGB, grey, HSL, HSV, L*a*b*) and Haralick textural feature [21].

In this study, the green bean image data as derived from a mixture of *Luwak* coffee and regular coffee are identified by using color features, such as: $Red_{(RGB)}$, $Green_{(RGB)}$, $Blue_{(RGB)}$, grey, Hue, $Saturation_{(HSL)}$, $Lightness_{(HSL)}$, $Saturation_{(HSV)}$, $Value_{(HSV)}$, L*, a*, b*, and textural features in each type of color (including entropy, energy, contrast, homogeneity, sum mean, variance, correlation, maximum probability, inverse different moment and cluster tendency). In addition, all the color and textural features in this study to select the best feature-subset combination are classified by using the feature selection method (filter method) before being used as input in ANN modeling. The selected color and textural features are then modeled by using ANN to estimate the percentage of pH, total phenol and purity of *Luwak* coffee with the lowest parameter value of Mean Square Error (MSE).

2. Research Method

This study utilizes green bean of arabica *Luwak* (civet) coffee and regular arabica coffee from Indonesian Plantation Company (PT Perkebunan Nusantara XII), Banyuwangi, Indonesia. Arabica *Luwak* coffee used in the research is Longan *Luwak* coffee. Regular arabica coffee is processed by using wet processing method. The tool for capturing pictures is digital camera (with specification of: Nikon Coolpix A10, 16 megapixels, Japan) placed in a black box, with the background of black surface, with constant fluorescent lighting and evenly distributed throughout the green bean coffee surface, and directly placed under a vertically mounted camera. The image data processing tool applies an Intel (R) Core (TM) i3 of 32 bit CPU computer 2.10 Ghz. Software used is by Windows 7 32 bit Operating System, with a self-built visual basic 6.0 based color and textural analysis software, equipped with Waikato Environment for Knowledge Analysis (WEKA) 3.8 [22], and with Matlab R2012a [23]. Green bean with a predetermined percentage, is placed on a platform with an area of 256 cm^2. The image format used is a bitmap. The image acquisition design is depicted in Figure 1. This study utilizes the green bean of arabica *Luwak* coffee and regular arabica coffee as the research object. Each data collection is gathered by using 160 coffee beans, while calculating the percentage of the mixture is performed in unit of seeds. Mixed proportions consist of: 0%, 10%, 30%, 40%, 50%, 70%, 90%, and 100% of *Luwak* coffee as shown in Figure 2. Total phenol test was measured using the Folin Ciocalteu method [24]. The pH measurement was carried out on coffee extract using a pH meter.

The image is converted from RGB colour space to grey, HSL, HSV and L*a*b* colour spaces [25]. The result of feature extraction is the color co-occurrence matrix (CCM) in each color group ($Red_{(RGB)}$, $Green_{(RGB)}$, $Blue_{(RGB)}$, grey, Hue, $Saturation_{(HSL)}$, $Lightness_{(HSL)}$, $Saturation_{(HSV)}$, $Value_{(HSV)}$, L*, a*, and b*). AlQaisi [26] developed different methods used to extract texture features from a color image. Texture values extracted in each type of color based on Haralick's texture analysis. The results of image data acquisition produce the 120 color and textural features. Haralick's textural equations are as follows:

$$Energy = \sum_{i}^{M}\sum_{j}^{N} P^2[i,j] \quad (1)$$

$$Entropy = -\sum_{i}^{M}\sum_{j}^{N} P^2[i,j] \quad (2)$$

$$Contrast = \sum_{i}^{M}\sum_{j}^{N} (i-j)^2 P[i,j] \quad (3)$$

$$Homogenity = \sum_{i}^{M}\sum_{j}^{N} \frac{P[i,j]}{1+|i-j|} \quad (4)$$

$$Inverse\ Difference\ Moment = \sum_{i}^{M}\sum_{j}^{N} \frac{P[i,j]}{|i-j|^k} \ \ i \neq j \quad (5)$$

$$Correlation = \sum_{i}^{M}\sum_{j}^{N} \frac{(i-\mu)(j-\mu)P[i,j]}{\sigma^2} \quad (6)$$

$$Sum\ Mean = \frac{1}{2}\sum_{i}^{M}\sum_{j}^{N} (iP[i,j] + jP[i,j]) \quad (7)$$

$$Variance = \frac{1}{2}\sum_{i}^{M}\sum_{j}^{N} ((i-\mu)^2 P[i,j] + (j-\mu)^2 P[i,j]) \quad (8)$$

$$Cluster\ Tendency = \sum_{i}^{M}\sum_{j}^{N} (i+j-2\mu)^k P[i,j] \quad (9)$$

$$Maximum\ Probability = \max_{i,j}^{M,N} P[i,j] \quad (10)$$

where: *P(i,j)* is the *(i,j)th* element of a normalized co-occurrence matrix, and μ and σ are the mean and standard deviation of the pixel element given by the following relationships:

$$P[i,j] = \frac{N(i,j)}{M} \quad (11)$$

$$\mu = \sum_{i}^{M} i \sum_{j}^{N} P[i,j] \quad (12)$$

$$\sigma = \sum_{i}^{M} (i-\mu)^2 \sum_{j}^{N} P[i,j] \quad (13)$$

where: *N(i,j)* is the number counts in the image with pixel intensity *i* followed by pixel intensity *j* at one pixel displacement to the left, and *M* is the total number of pixels.

ANN topology optimization is conducted by using Matlab R2012a software. The results of data acquisition of digital image processing methods obtaine 528 images at a predetermined percentage. Image data is divided into 66.67% as training data and 33.33% as validation data. The distribution of training data and validation data is the initial stage in ANN [27]. Training data is applied to update weights, biases and study data patterns. The accuracy of the model uses validation data to find out the ability of the network to identify new data patterns [28]. ANN

modeling applies the backpropagation neural network (BPNN) algorithm, which is a popular algorithm, widely used in ANN [29, 30]. Before performing ANN modeling, input and output data are normalized to range of -1 and 1.

Input layers include colors and textural features. The output layer expresses the percentage of purity, total phenol and pH in *Luwak* coffee. Designing the best ANN topology is accomplished through sensitivity analysis with a variety of learning functions; activation function; learning rate and momentum (0.1, 0.5, 0.9); hidden layer (1, 2); hidden layer node (10, 20, 30, 40) with the lowest validation of MSE parameter. This study formulates the 3 activation functions i.e. purelin, tansig, and logsig [31].

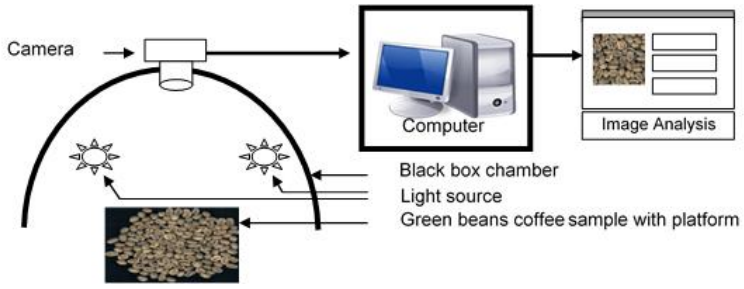


Figure 1. Design of image acquisition system

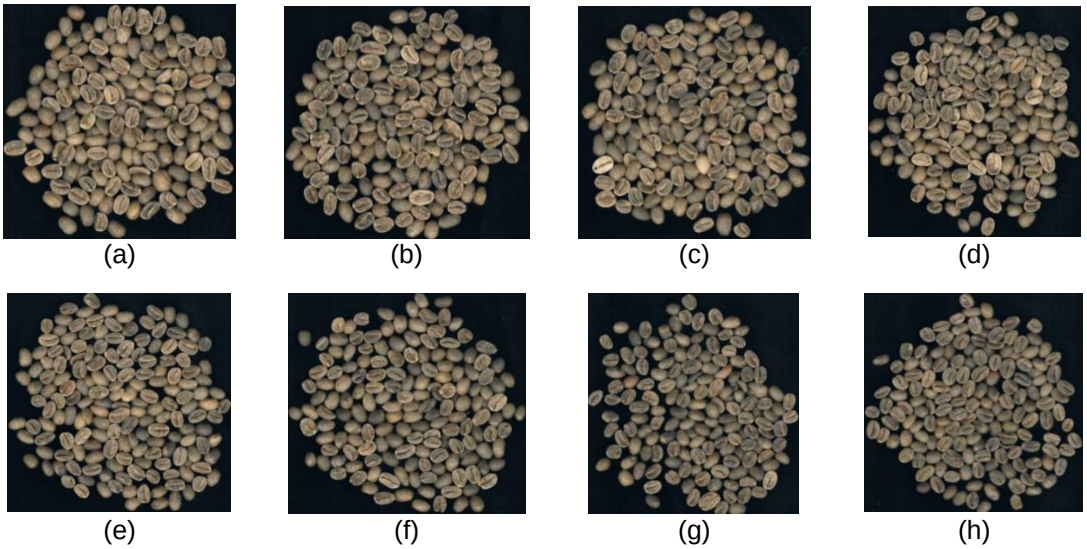


Figure 2. Mixture of *Luwak* (civet) coffee and regular coffee:
(a) 0%; (b) 10%; (c) 30%; (d) 40%; (e) 50%; (f) 70%; (g) 90%; (h) 100%

3. Results and Analysis

The feature extraction results in 120 colors and textural features which represent information related to the image (a mixture of *Luwak* coffee and regular green bean in various percentages). The main problem emerges that not all color and textural features are capable of predicting dependent variable or objective function. This stage is intended to find out the features affecting either dependent variable or objective function. Feature selection is conducted by preprocessing data in data mining. Selection of features becomes an important stage to speed up the modeling process and to facilitate the design of tools. The main purpose of feature selection is to prevent overfitting, as characterized by high MSE validation; to reduce training time and to improve model accuracy [32-34]. The research results of the Karabulut [35] presented that feature selection might increase accuracy by 15.55% in ANN, Naive Bayes, and J48 Decision Tree modeling. This study employs 6 attribute evaluators, such as: Cfs Subset, Correlation Attribute, One R Attribute, ReliefF, Gain Ratio Attribute, and Gain Info Attribute. This

study applies filter model to find out the feature selection. The filter model which is fast and simple, assesses relevant features by knowing the intrinsic nature of data. Filter method algorithms rank features based on their proximity to the class. The filter method has the advantages of being a fast and simple computing method [36, 37].

The feature selection output for digital image data is in the top 10 rank for efficiency and simplification of ANN input. After obtaining the top 10 rank, the input data is then modeled by using ANN to select the input which produces the lowest MSE validation. ANN structures used are: the 40 nodes in 1st hidden layer and 40 nodes in 2nd hidden layer; activation function used in hidden layer and output layer was tansig; trainlm as learning function; learning rate of 0.1 and momentum of 0.9.

Table 1 shows the $Red_{(RGB)}$ sum mean, which has a strong correlation with the percentage of regular coffee mixtures in *Luwak* coffee weighing of 0.23599. After obtaining the weights and ratings, the data in Table 1 are modeled by using ANN to find out features which can predict total phenol, pH, and the percentage of regular coffee mixes in *Luwak* coffee with the lowest MSE validation parameter. ANN output for digital image data feature selection in Table 2 shows that when the 120 color and textural features are used as ANN inputs, there is no value due to network errors. This is due to incompatibility of the trainlm learning function with the amount of input data. Thus, feature selection remains necessary to be performed.

The results of feature selection, present 5 data inputs correlating with the percentage of regular coffee mixes in *Luwak* coffee. The five data results from feature selection are labeled as texture features. Extraction of image features is based on CCM. The frequently applied conventional method in texture analysis is the gray level co-occurrence matrix (GLCM), which is a popular method for representing texture features as developed by Harralick.

Figure 3 depicts the relationship of regular coffee mixture percentage in *Luwak* coffee with $Red_{(RGB)}$ sum mean. The results show that the $Red_{(RGB)}$ sum mean decreases along with the increasing percentage of *Luwak* coffee. The value of the textural feature $Red_{(RGB)}$ sum mean states the average number of red values in the image (the higher the value of $Red_{(RGB)}$ sum mean, the average number of reds in the textural feature will be greater). Figure 3 shows the textural feature $Red_{(RGB)}$ sum mean of 100% *Luwak* coffee was lower than 0% *Luwak* coffee.

Figure 4 shows $Value_{(HSV)}$ sum mean in which the value decreases with the increase in the percentage of *Luwak* coffee. Textural feature $Value_{(HSV)}$ sum mean value states the average number of value in the image (the higher the value, the textural feature will be greater). The texture value states the amount of light received by the eye regardless of the color. This is in accordance with the color of *Luwak* coffee and regular coffee which can be observed visually. *Luwak* coffee used in the study has a darker color than in regular coffee, affecting the value sum mean to decrease along with the increase in the percentage of *Luwak* coffee. Figure 5 shows $Saturation_{(HSL)}$ sum mean in which the value decreases along with the increase in the percentage of *Luwak* coffee. $Saturation_{(HSL)}$ sum mean value states the average number of saturation value in the image (the higher the value, the greater the amount of saturation). Figure 6 shows $Blue_{(RGB)}$ variance which decreases with the increase in the percentage of *Luwak* coffee. $Blue_{(RGB)}$ variance shows variations in co-occurrence matrix elements. Images with small color degree transitions will have little variance. If the variance value is high, the degree of color of the image will spread. Variance is the sum of squares of differences in intensity among the neighboring pixels. Figure 6 shows that *Luwak* (civet) coffee green bean (100%) have a less diffused blue color, while green bean in *Luwak* coffee (0%) has a diffused blue color. Figure 7 shows the Hue variance which increases along with the increase in the percentage of *Luwak* coffee. From the graph, it is obvious that *Luwak* coffee (100%) has a more diffused Hue color than the *Luwak* coffee (0%).

ANN modeling produces predictive output, weight and bias which is optimal in estimating the percentage of purity, total phenol and pH. The most important step in designing ANN structure is the selection of optimal weights and biases among neurons with high generalizations [38]. The selected ANN struecture is presented in Figure 8. The initial stage in designing ANN structure is a training error of the learning function. Learning function plays a role in changing weights and biases during training. ANN modeling results consist of weights and biases that affect MSE validation. For this reason, a training error of the learning function is carried out. The research results of Sharma and Venugopulan [39] and Aggarwal and Rajendra [40] point out that the learning function influences ANN performance.

Table 1. Feature Selection of Digital Image Data

No.	Attribute Evaluator	Search Method	Image Features	Weight	Rank
1.	Cfs Subset Evaluator	Best First	$Red_{(RGB)}$ Entropy	-	1
			Hue Contrast	-	2
			Hue Inverse	-	3
			Hue Correlation	-	4
			$Saturation_{(HSL)}$ Correlation	-	5
			$Saturation_{(HSV)}$ Correlation	-	6
			$Red_{(RGB)}$ Sum Mean	-	7
			Hue Sum Mean	-	8
			$Saturation_{(HSL)}$ Sum Mean	-	9
			$Saturation_{(HSV)}$ Sum Mean	-	10
		Greedy Stepwise	$Red_{(RGB)}$ Entropy	-	1
			Hue Contrast	-	2
			Hue Inverse	-	3
			$Saturation_{(HSL)}$ Correlation	-	4
			$Saturation_{(HSV)}$ Correlation	-	5
			$Red_{(RGB)}$ Sum Mean	-	6
			Hue Sum Mean	-	7
			$Saturation_{(HSL)}$ Sum Mean	-	8
			$Saturation_{(HSV)}$ Sum Mean	-	9
			$Blue_{(RGB)}$ Variance	-	10
2.	Correlation Attribute Evaluator	Ranker	$Green_{(RGB)}$ Sum Mean	0.324	1
			Grey Sum Mean	0.324	2
			$Red_{(RGB)}$ Sum Mean	0.323	3
			$Value_{(HSV)}$ Sum Mean	0.323	4
			$Lightness_{(HSL)}$ Sum Mean	0.323	5
			$L_{(Lab)}$ Mean	0.318	6
			$Blue_{(RGB)}$ Variance	0.317	7
			$Lightness_{(HSL)}$ Variance	0.317	8
			$Blue_{(RGB)}$ Cluster	0.316	9
			Grey Variance	0.316	10
3.	One R Attribute	Ranker	$Saturation_{(HSL)}$ Sum Mean	65.530	1
			Hue Homogeneity	61.932	2
			Hue Inverse	61.932	3
			Hue Energy	61.932	4
			Hue Contrast	60.795	5
			Hue Sum Mean	60.606	6
			Hue Entropy	59.848	7
			Hue Variance	57.576	8
			$Blue_{(RGB)}$ Variance	56.061	9
			Hue Cluster	55.303	10
4.	ReliefF	Ranker	$Red_{(RGB)}$ Sum Mean	0.236	1
			$Value_{(HSV)}$ Sum Mean	0.235	2
			$Saturation_{(HSL)}$ Sum Mean	0.234	3
			$Blue_{(RGB)}$ Variance	0.231	4
			Hue Variance	0.231	5
			Grey Sum Mean	0.227	6
			$Lightness_{(HSL)}$ Variance	0.227	7
			Grey Variance	0.225	8
			$Green_{(RGB)}$ Variance	0.225	9
			$Green_{(RGB)}$ Sum Mean	0.223	10
5.	Gain Ratio Attribute Evaluator	Ranker	Grey Sum Mean	0.716	1
			$Red_{(RGB)}$ Sum Mean	0.698	2
			$Lightness_{(HSL)}$ Sum Mean	0.688	3
			$Value_{(HSV)}$ Sum Mean	0.683	4
			$Green_{(RGB)}$ Sum Mean	0.682	5
			$Value_{(HSV)}$ Variance	0.665	6
			$Saturation_{(HSL)}$ Sum Mean	0.663	7
			$Lightness_{(HSL)}$ Variance	0.662	8
			Grey Variance	0.643	9
			Grey Cluster	0.640	10
6.	Info Gain Attribute Evaluator	Ranker	$Saturation_{(HSL)}$ Sum Mean	2.082	1
			Hue Inverse	2.010	2
			Hue Contrast	1.919	3
			Hue Energy	1.919	4
			Hue Entropy	1.908	5
			Hue Homogeneity	1.883	6
			Hue Correlation	1.872	7
			Hue Sum Mean	1.852	8
			Hue Variance	1.838	9
			Hue Cluster	1.808	10

Table 2. ANN Performance using Colour and Textural Feature Selection

No.	Attribute Evaluator	Search Method	Input	MSE Training	MSE Validation
1.	-	-	All Feature (120 Inputs)	-	-
2.	Cfs Subset Evaluator	Best First	Feature rank 1~2	0.0099	0.1094
			Feature rank 1~3	0.0098	0.2485
			Feature rank 1~4	0.0090	0.7921
			Feature rank 1~5	0.0098	0.1740
			Feature rank 1~6	0.0099	0.0973
			Feature rank 1~7	0.0083	0.0802
			Feature rank 1~8	0.0100	0.1239
			Feature rank 1~9	0.0079	0.3220
			Feature rank 1~10	0.3204	0.6605
		Greedy Stepwise	Feature rank 1~2	0.0099	0.1094
			Feature rank 1~3	0.0098	0.2485
			Feature rank 1~4	0.0098	0.1167
			Feature rank 1~5	0.0084	0.2642
			Feature rank 1~6	0.0088	0.3582
			Feature rank 1~7	0.0098	0.1245
			Feature rank 1~8	0.0060	0.3213
			Feature rank 1~9	0.0093	0.1699
			Feature rank 1~10	0.0062	0.1853
3.	Correlation Attribute Evaluator	Ranker	Feature rank 1~2	0.0099	0.2878
			Feature rank 1~3	0.0100	0.4090
			Feature rank 1~4	0.0097	0.1789
			Feature rank 1~5	0.0100	0.2674
			Feature rank 1~6	0.0097	0.0621
			Feature rank 1~7	0.0097	0.0918
			Feature rank 1~8	0.0100	0.0684
			Feature rank 1~9	0.0099	0.0672
			Feature rank 1~10	0.0803	0.3246
4.	One R Attribute	Ranker	Feature rank 1~2	0.0088	0.3919
			Feature rank 1~3	0.0097	0.5325
			Feature rank 1~4	0.0082	0.7760
			Feature rank 1~5	0.0091	0.3968
			Feature rank 1~6	0.0048	0.9021
			Feature rank 1~7	0.0098	0.2642
			Feature rank 1~8	0.0086	0.1098
			Feature rank 1~9	0.0096	0.1249
			Feature rank 1~10	0.0097	0.1622
5.	RelieF	Ranker	Feature rank 1~2	0.0135	0.0689
			Feature rank 1~3	0.0100	0.0872
			Feature rank 1~4	0.0098	0.0826
			Feature rank 1~5	0.0097	0.0491
			Feature rank 1~6	0.0094	0.0552
			Feature rank 1~7	0.0095	0.0569
			Feature rank 1~8	0.0098	0.0542
			Feature rank 1~9	0.0097	0.2612
			Feature rank 1~10	0.3099	0.3937
6.	Gain Ratio Attribute Evaluator	Ranker	Feature rank 1~2	0.0099	0.1528
			Feature rank 1~3	0.0099	0.1014
			Feature rank 1~4	0.0100	0.1741
			Feature rank 1~5	0.3970	0.9287
			Feature rank 1~6	0.0100	0.0900
			Feature rank 1~7	0.0098	0.0588
			Feature rank 1~8	0.0099	0.0531
			Feature rank 1~9	0.0099	0.0568
			Feature rank 1~10	0.0095	0.1432
7.	Info Gain Ratio Attribute Evaluator	Ranker	Feature rank 1~2	0.0095	0.5348
			Feature rank 1~3	0.0095	0.1452
			Feature rank 1~4	0.0099	0.4116
			Feature rank 1~5	0.0097	0.6437
			Feature rank 1~6	0.0099	0.1161
			Feature rank 1~7	0.0098	0.1787
			Feature rank 1~8	0.0100	0.1561
			Feature rank 1~9	0.6231	0.2797
			Feature rank 1~10	0.0097	0.2618

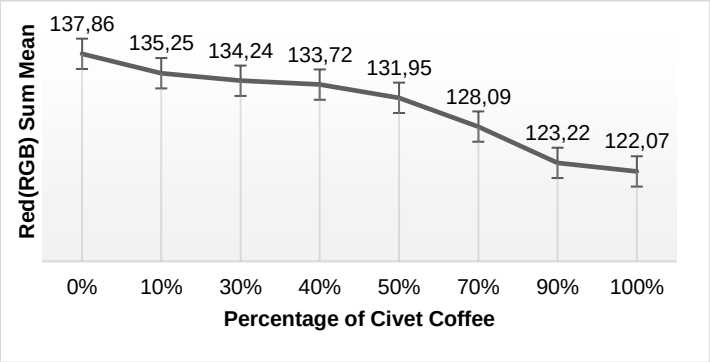


Figure 3. The relationship of *Luwak* coffee purity to red$_{(RGB)}$ sum mean

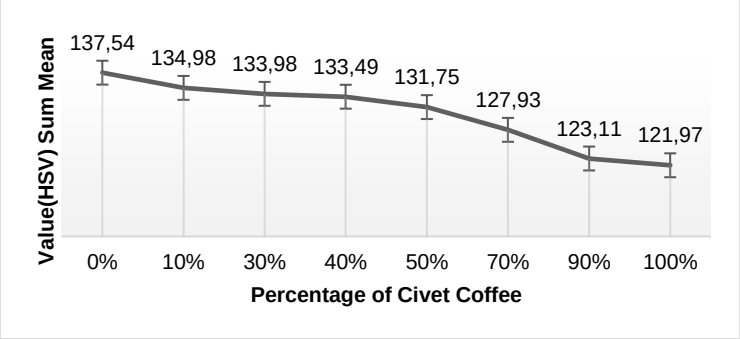


Figure 4. Relationship of *Luwak* coffee purity to value$_{(HSV)}$ sum mean

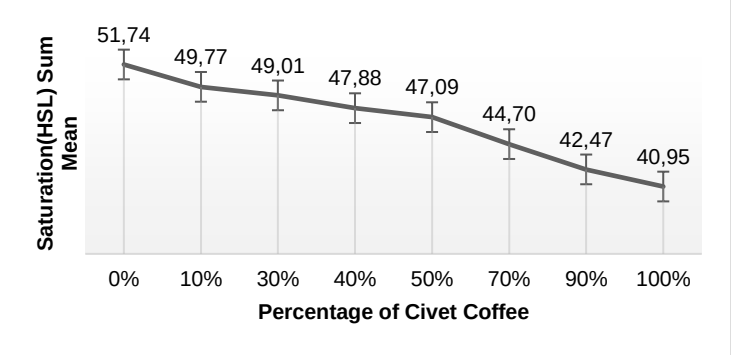


Figure 5. The relationship of *Luwak* coffee purity to saturation$_{(HSL)}$ sum mean

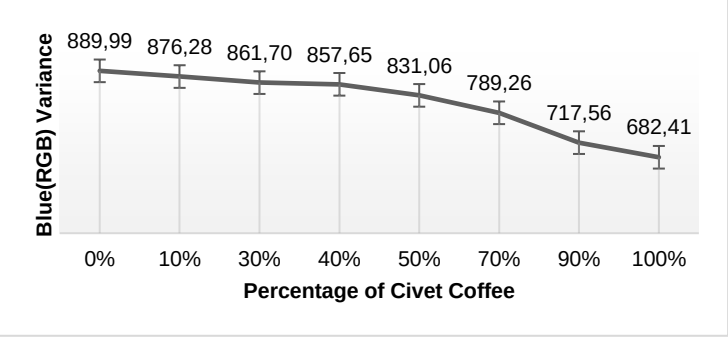


Figure 6. Relationship of *Luwak* coffee purity to blue$_{(RGB)}$ variance

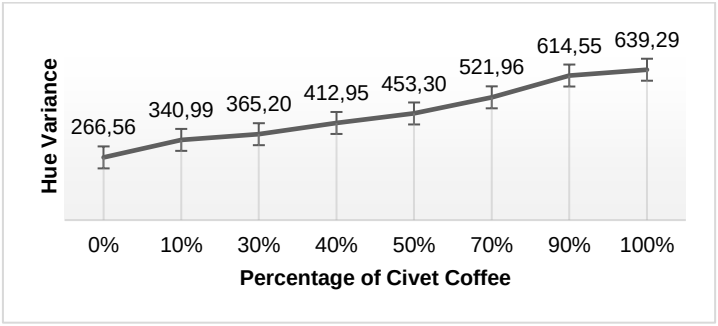


Figure 7. Relationship of luwak (civet) coffee purity to hue variance

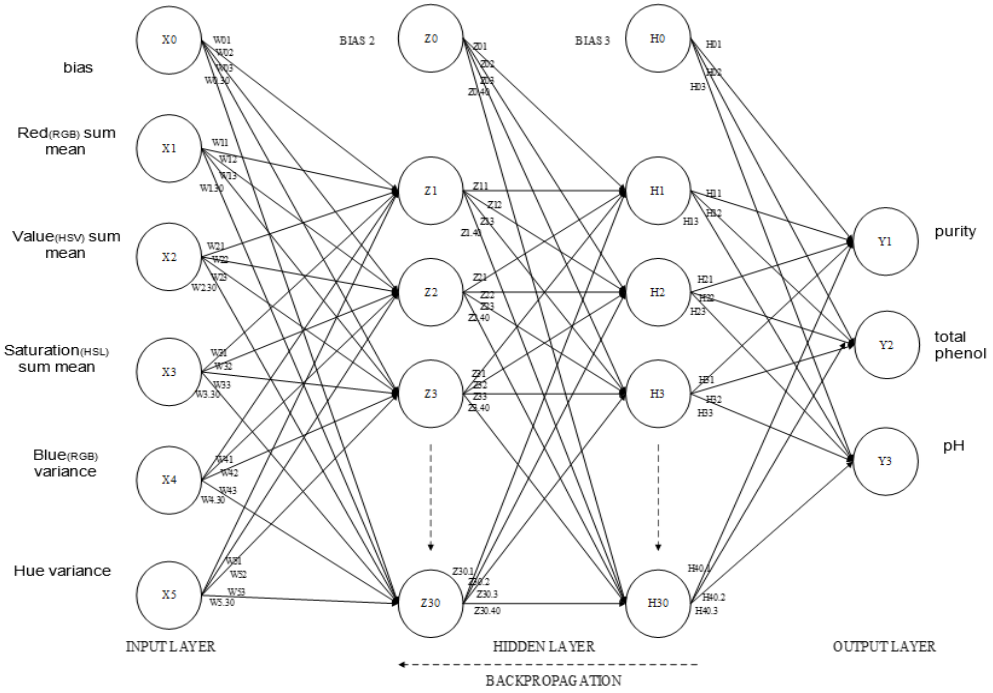


Figure 8. ANN structure with 5 selected input parameter images for estimating the percentage of *Luwak* (civet) coffee mixture, total phenol and pH

Table 3 presents trainlm as the selected learning function which produces the lowest MSE validation. Trainlm is a learning function that updates the weights and biases based on Lavenberg Marquadt optimization. Trainlm is categorized as the fastest algorithm and is recommended as the first supervision algorithm despite entailing more memory than other algorithms. Trainlm is proceeded by using Jacobian Matrix calculations, while network performance is measured from MSE. Trainlm is designed to have a two-level training speed, which is faster without calculating the Hessian matrix. After obtaining the best learning function that gives the lowest MSE validation, then training error is managed in the activation function as illustrated in Table 4.

The results of the training error show that the tansig function in the hidden layer and purelin in the output layer gives the lowest MSE validation. Purelin activation function is only used in the output layer. Purelin produces y=x, which cannot solve non-linear problems on hidden layer nodes. ANN model is identical to finding the relationship of non-linear data between input and output. For this reason, activation functions such as tansig and logsig are recommended in the hidden layer. Training error activation function is performed because it affects MSE validation. Research by Chang and Chung [41] shows that the results of the sensitivity analysis in the activation function, affect the performance of ANN in producing

the lowest MSE validation. The unsuitable activation function causes an increase in MSE validation. After obtaining the best activation function, ANN structure is further designed with a variety of learning rates, hidden layer nodes, and the number of hidden layers.

Table 5 shows the best structure which is the 5-30-40-3 with learning rate of 0.1 and momentum of 0.5, which produces MSE validation of 0.0442. Determination of the number of hidden layer nodes and the number of hidden layers plays as the most important stage in designing ANN structure. The results show that two hidden layers can predict the output variable which is better than 1 hidden layer, due to the 2-hidden-layer ability to solve non-linear problems which is better than 1 hidden layer. However, more hidden layers hinder the computer running. Therefore, the hidden layer sensitivity analysis is needed. In this study, the maximum number of hidden layers is determined, as Karsoliya [42] stated that the 2-hidden-layers can solve non-linear problems. Sensitivity analysis of learning rate and momentum is required because both of these indicators play a role in changes of weight and bias during training.

Table 3. Training Error based on Learning Function

No.	Learning Function	R training	R validation	MSE training	MSE validation
1.	Traincgb (Conjugate Gradient BP with Powell – Beale Restart)	0.98891	0.98231	0.0100	0.0482
2.	Traincgf (Conjugate Gradient BP with Fletcher Reeves Update)	0.98882	0.98849	0.0100	0.0517
3.	Traincgp (Conjugate Gradient BP with Polak Ribiere Update)	0.98885	0.98713	0.0100	0.0488
4.	Traingd (Gradient Descent BP)	0.98306	0.98884	0.0151	0.0509
5.	Traingda (Gradient Descent with Adaptive Learning Rate BP)	0.98302	0.98792	0.0152	0.0518
6.	Traingdm (Gradient Descent with Momentum BP)	0.98883	0.98836	0.0096	0.0534
7.	Traingdx (Gradient Descent with Momentum Adaptive Learning Rate BP)	0.98680	0.98876	0.0118	0.0508
8.	Trainlm (Lavenberg Marquadt BP)	0.99085	0.98932	0.0092	0.0464
9.	Trainoss (One Step Secant BP)	0.98887	0.98772	0.0100	0.0542
10.	Trainrp (Resilient BP)	0.98885	0.98729	0.0100	0.0474
11.	Trainscg (Scaled Conjugate Gradient BP)	0.98887	0.98650	0.0100	0.0539

Table 4. Training Error based on Activation Function

Learning function	Activation function			R training	R validation	MSE training	MSE validation
	Hidden Layer 1	Hidden Layer 2	Output Layer				
TRAINLM	Tansig	Tansig	Purelin	0.99085	0.98932	0.0092	0.0464
	Tansig	Tansig	Tansig	0.99085	0.98932	0.0092	0.0579
	Tansig	Tansig	Logsig	0.88223	0.86926	0.2300	0.2671
	Logsig	Logsig	Purelin	0.98915	0.97422	0.0097	0.0596
	Logsig	Logsig	Tansig	0.98925	0.85381	0.0100	0.2136
	Logsig	Logsig	Logsig	0.42882	0.86914	0.3747	0.3791

Table 5. Accuracy and MSE of ANN Performance based on Learning Rate and Momentum

Learning Rate	Momentum	ANN Structure	R training	R validation	MSE training	MSE validation
0.1	0.5	5-30-3	0.98895	0.97281	0.0099	0.0684
		5-40-3	0.98900	0.98386	0.0098	0.0445
		5-30-30-3	0.99178	0.97802	0.0075	0.0540
		5-30-40-3	0.99052	0.97933	0.0085	0.0442
		5-40-40-3	0.98894	0.97398	0.0099	0.0516
0.1	0.9	5-30-3	0.98910	0.96839	0.0098	0.0684
		5-40-3	0.98896	0.96914	0.0099	0.0546
		5-30-30-3	0.98887	0.97890	0.0100	0.0509
		5-30-40-3	0.98929	0.98191	0.0096	0.0476
		5-40-40-3	0.98894	0.97398	0.0099	0.0516

Figure 9 shows three descriptions, which are: the blue line representing the training, the blue dashed line representing the best, and the black dashed line representing the goal. The training shows an iterative relationship to MSE during training. Figure 9 shows decreasing error along with increasing iteration due to the stable network ability to recognize data patterns. ANN is a "black box" modeling that is often used in dealing with non-linear problems. ANN has the ability to learn from iterations, which is widely used in various fields of science. The advantages of ANN are being able to adopt, study and generalize [43, 44]. The advantages of ANN are also to quickly and accurately study data patterns compared to conventional

mathematical modeling [45]. The blue dash line represents the best which shows the convergent network at iteration 9 and MSE of 0.008536. Figure 9 shows the network reaching the goal in iteration 9. The maximum number of iterations is 10,000 with a goal of 0.01, meaning that the training will stop at 10,000 iterations or when the goal has reached 0.01. MSE validation reaches 0.01 goal because the network is considered accurate in predicting the objective function or variable "y". Setting a MSE goal which is too small, results in overfitting the model. Determination of the number of iterations and goals are obligatory to avoid overfitting the model. Overfitting occurs when the model is very exclusive in recognizing training data patterns affecting the generalization to decrease, as indicated by the high MSE validation.

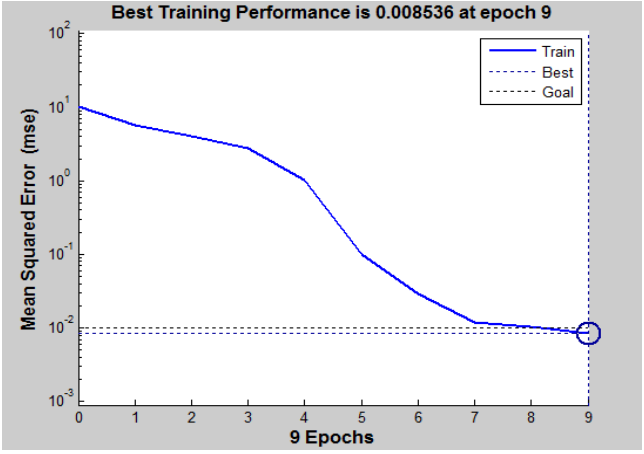
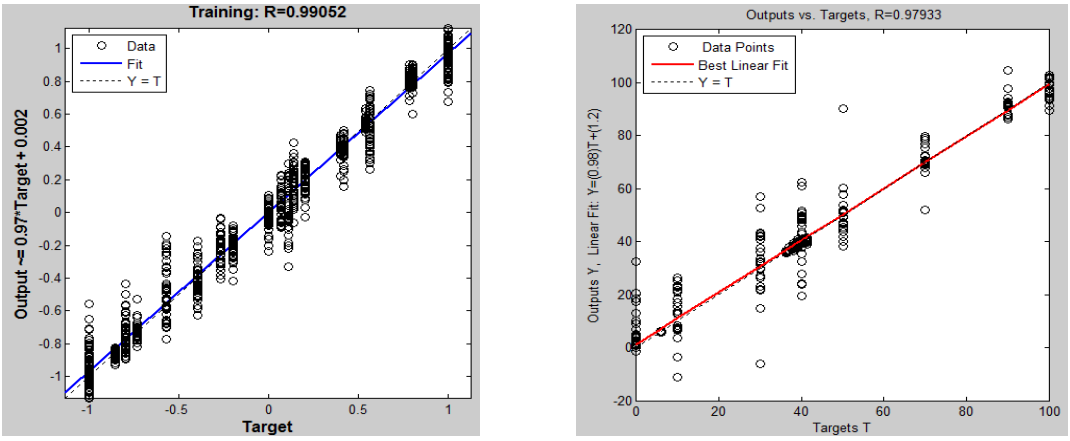


Figure 9. Relationship of the number of iterations with MSE in the ANN training process for identification of *Luwak* (civet) coffee purity

Figure 10 shows a regression plot which has the blue line marking a training simulation while the red line shows a validation simulation. In the regression plot, the data distribution training approaches the linear fit line which indicates the prediction is getting closer to the actual value, as notified from the training R value of 0.99052. The validation regression plot can be seen from the deviant data distribution points in linear fit, causing a low R validation. The value of R indicates the correlation between input and output variables. The closer it is to 1, the correlation becomes stronger. Based on the results of the research, the value of R approaches 1 indicating that the relationship of textural feature data, which are ($Red_{(RGB)}$ sum mean, $Value_{(HSV)}$ sum mean, $Saturation_{(HSL)}$ sum mean, $Blue_{(RGB)}$ variance, Hue variance) to the percentage of total phenol, pH, and *Luwak* (civet) coffee purity is very strong.



(a) (b)

Figure 10. The regression plot of the simulation results: (a) training data (b) validation data

4. Conclusion

The detection results of the regular coffee mixes in *Luwak* (civet) coffee, total phenol and pH applying the digital images present that the 5 features of the image are selected as ANN inputs, which are: $Red_{(RGB)}$ sum mean, $Value_{(HSV)}$ sum mean, $Saturation_{(HSL)}$ sum mean, $Blue_{(RGB)}$ variance, and Hue variance by using the feature selection filter technique with the relief method. Image data using ANN produces selected structure of 5-30-40-3 (5 inputs, 30 nodes in 1st hidden layer, 40 nodes in 2nd hidden layer, 3 outputs) with a learning rate of 0.1 and momentum of 0.5 trainlm learning function, tansig activation function in hidden layer and purelin at the output layer. Selected ANN structure produces R training of 0.99502; R validation of 0.97933 and MSE training of 0.0085; and MSE validation of 0.0442. The results indicate that digital image processing and ANN model are potential to be a sensor in detecting the percentage of total phenol, pH, and *Luwak* coffee purity.

References

[1] Bohl MT, Gross C, Souza W. The role of emerging economies in the global price formation process of commodities: Evidence from Brazilian and U.S. coffee markets. *International Review of Economics & Finance*. 2019; 60: 203-215. doi.org/10.1016/j.iref.2018.11.002.

[2] Marcone MF. Composition and Properties of Indonesian Palm Civet Coffee (Kopi Luwak) and Ethiopian Civet Coffee. *Food Research International*. 2004; 37: 901-912.

[3] Jumhawan U, Putri SP, Yusianto, Marwani E, Bamba T, Fukusaki E. Selection of Discriminant Markers for Aunthetication of Asian Palm Civet Coffee (Kopi Luwak): A Metabolomics Approach. *Journal of Agricultural and Food Chemistry*. 2013; 61(33): 7994-8001.

[4] Cammerer B, Kroh LW. Antioxidant Activity of Coffee Brews. *European Food Research and Technology*. 2006; 223(4): 469-474.

[5] Cheong MW, Tong KH, Ong JJM, Liu SQ, Curran P, Yu B. Volatile composition and antioxidant capacity of Arabica coffee. *Food Research International*. 2013; 51(1): 388-396.

[6] Naidu M, Madhava G, Sulochanamma SR, Sampathu P, Srinivas. Studies of Extraction and Antioxidant Potential of Green Coffee. *Food Chemistry*. 2008; 107(1): 377-384.

[7] Oliveira EM, Leme DS, Barbosa BHG, Rodarte MP, Pereira RGFA. A computer vision system for coffee benas classification based on computational intelligence techniques. *Journal of Food Engineering*. 2016; 171: 22-27.

[8] Nansen C, Singh K, Mian A, Allison BJ, Simmons CW. Using hyperspectral imaging to characterize consistency of coffee brands and their respective roasting classes. *Journal of Food Engineering*. 2016; 190: 34-39.

[9] Caporaso N, Whitworth MB, Grebby S, Fisk ID. Rapid prediction of single green coffee bean moisture and lipid content by hyperspectral imaging. *Journal of Food Engineering*. 2018; 227: 18-29.

[10] Navarro LV, Lopez EJH, Gonzalez RIC, Guevara EA, Morales GMG. Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed. *Expert System with Applications*. 2016; 54: 162-169.

[11] Valencia RM, Jurado JM, Magana SGC, Alcazar A, Diaz JH. Characterization of Mexican coffee according to mineral contents by means of multilayer perceptrons artificial neural networks. *Journal of Food Composition and Analysis*. 2014; 34(1): 7-11.

[12] Kouadio L, Deo RC, Byrareddy V, Adamowski JF, Mushtaq S, Nguyen VP. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. *Computer and Electronics in Agriculture*. 2018; 155: 324-338.

[13] Vithu P, Moses JA. Machine vision system for food grain quality evaluation: A review. *Trends in Food Science & Technology*. 2016; 56: 13-20.

[14] Patricio DI, Rieder R. Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. *Computer and Electronics in Agriculture*. 2018; 153: 69-81.

[15] Patel KK, Kar A, Jha SN, Khan MA. Machine Vision System : A Tool for Quality Inspection of Food and Agricultural Products. *Journal of Food Science and Technology*. 2012; 49(2): 123-141.

[16] Sadegaonkar VD, Kiran HW. Automatic Sorting using Computer Vision & Image Processing for Improving Apple Quality. *International Journal of Innovative Research & Development*. 2015; 4(1): 11-14.

[17] Hendrawan Y, Al Riza D. Machine Vision Optimization using Nature-Inspired Algorithms to Model Sunagoke Moss Water Status. *International Journal on Advanced Science, Engineering and Information Technology*. 2016; 6(1): 45-57.

[18] Hendrawan Y, Murase H. Neural-Intelligent Water Drops Algorithm to Select Relevant Textural Features for Developing Micro Precision Irrigation System using Machine Vision. *Computers and Electronics in Agriculture*. 2011; 77(2): 214-228.

[19] Hendrawan Y, Murase H. Neural-Discrete Hungry Roach Infestation Optimization to Select Informative Textural Features for Determining Water Content of Cultured Sunagoke Moss. *Environmental Control in Biology*. 2011; 49(1): 1-21.

[20] Hendrawan Y, Murase H. Bio-inspired Feature Selection to Select Informative Image Features for Determining Water Content of Cultured Sunagoke Moss. *Expert Systems with Applications*. 2011; 38(11): 14321-14335.
[21] Haralick RM, Shanmugam K, Its'hak Dinstein. Textural features for image classification. *IEEE Transaction on Systems, Man, and Cybernetics*. 1973; 3(6): 610-621.
[22] Mark H, Eibe F, Geoffrey H, Bernhard P, Peter R, Ian H W. The WEKA Data Mining Software: An Update. *SIGKDD Explorations.* 2009; 11(1).
[23] Mathworks. MATLAB Release 2010b. http://www.mathworks.com. 2010.
[24] Singleton VL, Rossi J. Calorimetry of Total Phenolic with Phosphomolybdic-Phosphotungstic Acid Agents. *American Journal of Enology and Viticulture*. 1965; 16: 144-158.
[25] Philipp I, Rath T. Improving plant discrimination in image processing by use of different colour space transformations. *Computer and Electronics in Agriculture*. 2002; 35: 1-15.
[26] AlQaisi A, AlTarawneh M, Alqadi ZA, Sharadqah AA. Analysis of color image features extraction using texture methods. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2019; 17(3): 1220-1225.
[27] Ahmadi MA, Ebadi M, Hosseini SM. Prediction Breakthrough Time of Water Coning in the Fractured Reservoirs by Implementing Low Parameter Support Vector Machine Approach. *Fuel*. 2014; 117: 579-589.
[28] Ahmadi MA, Reza S, Moonyong L, Tomoaki K, Alireza B. Determination of Oil Well Production Performance using Artificial Neural Network (ANN) linked to the Particle Swarm Optimization (PSO) Tool. *Petroleum*. 2015; 1: 1180-132.
[29] Rumelhart DE, Hinton GE, Williams RJ. Learning Internal Representation by Error Propagation. Cambridge: MIT Press. 1986.
[30] Basheer IA, Hajmeer M. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *Journal of Microbiological Methods*. 2000; 43: 3-31.
[31] Artrith N, Alexander U. An Implementation of Artificial Neural Network Potentials for Atomistic Materials Simulations: Performance for TiO_2. *Computational Material Science*. 2016; 114: 135-150.
[32] Canedo VB, Betanzos AA. Ensembles for feature selection: A review and future trends. *Information Fusion*. 2019; 52: 1-12. doi.org/10.1016/j.inffus.2018.11.008
[33] Urbanowicz RJ, Meeker M, Cava WL, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*. 2018; 85: 189-203.
[34] Gao W, Hu L, Zhang P, He J. Feature selection considering the composition of feature relevancy. *Pattern Recognition Letters*. 2018; 112: 70-74.
[35] Karabulut EM, Ozel SA, Turgay I. A Comparative Study on the Effect of Feature Selection on Classification Accuracy. *Procedia Technology*. 2012; 1: 323-327.
[36] Zeng Z, Hongjun Z, Rui Z, Cheng XY. A Novel Feature Selection Method Considering Feature Interaction. *Pattern Recognition*. 2015; 48(8): 2656-2666.
[37] Ghaemi M, Mohammad RFD. Feature Selection using Forest Optimization Algorithm. *Pattern Recognition*. 2016; 60: 121-129.
[38] Hejmanowski R, Wojciech TW. Suitability Assessment of Artificial Neural Network to Approximate Surface Subsidence due to Rock Mass Drainage. *Journal of Suitable Mining*. 2015; 14: 101-107.
[39] Sharma B, Venugopulan. Comparison of Neural Netwrok Training Functions for Hematoma Classification in Brain CT Images. *IOSR Journal of Computer Engineering*. 2014; 16(1): 31-35.
[40] Aggarwal R, Rajendra K. Effect of Training Function of Artificial Neural Networks (ANN) on Time Series Forecasting. *International Journal of Computer Applications*. 2015; 109(3): 14-17.
[41] Chang CL, Chung SL. Parameter Sensitivity Analysis of Artificial Neural Network for Predicting Water Turbidity. *International Journal of Environmental, Chemical, Ecological, Geological, and Geophysical Engineering*. 2012; 6(10): 657-660.
[42] Karsoliya S. Approximating Number Of Hidden Layer Neuron In Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends And Technology*. 2012; 3(6): 714-717.
[43] Haykin S. Neural Networks and Learning MachSines. 3rd Edition. Pearson. 2009.
[44] Kheirkhah A, Azadeh A, Saberi M, Azaron A, Shakouri H. Improved Estimation of Electricity Demand Function by using of Artificial Neural Network, Principal Component Analysis and Data Envelopment Analysis. *Computers & Industrial Engineering*. 2012; 64: 425-441.
[45] Jaiswal S, Benson ER, Bernard JC, Van Wicklen GL. Neural Network Modelling and Sensitivity Analysis of a Mechanical Poultry Catching System. *Biosystem Engineering*. 2005; 92(1): 59-68.