International Journal of Artificial Intelligence & Robotics (IJAIR) E-ISSN: 2686-6269
Vol.6, No.2, 2024, pp.75-82

Filter Feature Selection for Detecting Mixture, Total Phenol, and pH of Civet Coffee

Shinta Widyaningtyas[1]*, Muhammad Arwani[2], Sucipto[3], Yusuf Hendrawan[4]
[1]Agriculture Engineering Department, Politeknik Negeri Jember, Indonesia
[2]Agroindustrial Technology Department, Universitas Nahdlatul Ulama Indonesia, Indonesia
[3]Agroindustrial Technology Department, Universitas Brawijaya, Indonesia
[4]Biosystem Engineering Department, Universitas Brawijaya, Indonesia
[1]shinta_widya@polije.ac.id*; [2]m.arwani@unusia.ac.id; [3]ciptotip@ub.ac.id; [4]yusuf_h@ub.ac.id
*corresponding author

ABSTRACT

Civet coffee, a highly valued specialty coffee, is susceptible to adulteration with regular coffee, resulting in economic losses and consumer fraud. This study investigates the potential of electrical spectroscopy as a non-destructive technique for detecting civet coffee adulteration. We analyzed the bioelectrical properties of civet coffee beans and their mixtures with regular coffee, focusing on impedance parameters (Z, Lp, Ls, Rp, Rs) as potential indicators of adulteration. Two machine learning models, Artificial Neural Network (ANN) and Random Forest, were trained and evaluated using Mean Squared Error (MSE) validation to identify the most informative features for predicting mixture composition, total phenol content, and pH. The findings demonstrate that impedance parameters, particularly Z, consistently exhibited high feature importance scores across different attribute evaluators and search methods. The optimal model, an ANN with a correlation attribute evaluator and ranker search method, achieved an MSE validation of 0.0479, indicating strong predictive accuracy. These results suggest that electrical spectroscopy, coupled with machine learning, offers a promising approach for developing automated, non-invasive methods for detecting civet coffee adulteration, thereby protecting consumers and ensuring the integrity of the specialty coffee market.

Keywords: Artificial Neural Network; Civet Coffee; Electrical Properties; Filter Feature Selection; Random Forest

Article History

Received : Sept, 11th 2024	Accepted : Dec, 10th 2024	Published: Dec, 28th 2024

I. INTRODUCTION

Civet coffee is arabica or robusta coffee consumed by a civet, fermented in its digestive tract, and expelled together with civet feces as whole, horn-skinned seeds. The feces were collected, washed, dried, and then processed generally [1]–[3]. The fermentation in the civet digestive tract gives the coffee an authentic aroma. This authentic aroma, uniqueness, and limited production process make civet coffee the most expensive coffee in the world [4]–[6]. As a high-priced commodity, civet coffee is prone to being counterfeited with regular coffee beans. Mixing civet coffee with green beans is common in Indonesia [7]. This fraudulent practice undermines the integrity of the specialty coffee market and erodes consumer trust, potentially diminishing the cultural prestige associated with authentic civet coffee. For producers, mixture threatens brand reputation and market share, while consumers face the risk of paying exorbitant prices for inferior products, leading to dissatisfaction and skepticism towards genuine specialty offerings. Ensuring the authenticity of civet coffee safeguards the livelihoods of legitimate producers and maintains the cultural heritage tied to its production. Moreover, protecting consumers from fraud enhances market transparency and fosters trust in premium coffee products. For this reason, an automatic tool is needed to identify the mixture in civet coffee green bean. In addition to estimating the blend in civet coffee, this study was strengthened by estimating total phenol and pH because civet coffee is low bitter, low acid, and low caffeine when compared to regular coffee.

According to [8], coffee is one of the sources of phenolic compounds that affect the bitter taste of coffee. In addition, pH estimation was also carried out to determine the acidity of the coffee. The design of an automatic device for the prediction mixture, total phenol, and pH of civet coffee can be done through an approach to the electrical properties of agricultural materials. Electrical properties are often used in quality testing of agricultural products because the moisture content in hygroscopic materials such as agricultural products is the dominant factor affecting the electrical properties [9]. Electrical properties can be measured using the dielectric method. The dielectric method is done by placing the material between capacitor plates so that the Impedance (Z), Series Resistance (Rs), Parallel Resistance (Rp), Parallel Capacitance (Cp), Series Capacitance (Cs), Series Inductance (Ls), Parallel Inductance (Lp), and Dielectric Constant [7][10]. Dielectric methods have been widely used in determining the quality of

DOI: http://dx.doi.org/10.25139/ijair.v6i2.9010

agricultural products, including sugar content estimation in sugarcane [11], ground beef [12], detection of catechin during black tea fermentation [13], and moisture content in green tea [14]. This indicates the dielectric method's ability to predict a wide range of contents in agricultural products. In this study, the dielectric method acquires Z, Rs, Rp, Cp, Cs, Ls, Lp, and dielectric constant but not all features relevant in prediction of civet coffee mixtures, total phenol, and pH. Currently, no study has reported the best combination of features (electrical properties) in the prediction of coffee blend, total phenolics, and pH. For this reason, feature selection is carried out to get the best features for prediction. There are several methods in feature selection including the filter method.

One of the primary advantages of filter methods is that they do not require the involvement of a learning algorithm to eliminate irrelevant features, thereby reducing computational costs and expediting the process [15][16]. The non-linear relationships between the input parameters (Z, Rs, Rp, Cp, Cs, Ls, Lp, and dielectric constant) and the output variables (civet coffee mixtures, total phenol, and pH) can be effectively modeled using Artificial Neural Networks (ANN) and Random Forest (RF) algorithms. These models have been extensively applied in various feature selection tasks, including nutrients in pistachio leaves [17], disease of mago leaves [18], detection purity, total phenol, and pH in civet coffee [19]. This study compares the performance of ANN and RF models based on the lowest validation Mean Square Error (MSE) criteria. The objective is to identify the optimal combination of electrical properties (features) that best predict civet coffee mixtures, total phenol, and pH. This research is part of a broader effort to design an automatic, simple, fast, and non-destructive tool for predicting these parameters. Ultimately, this study contributes to addressing the issue of adulterating civet coffee with regular coffee, a problem that affects both producers and consumers by ensuring the authenticity and quality of the product.

II. METHOD

Experimental design for filter feature selection using ANN can be seen in Fig.1.

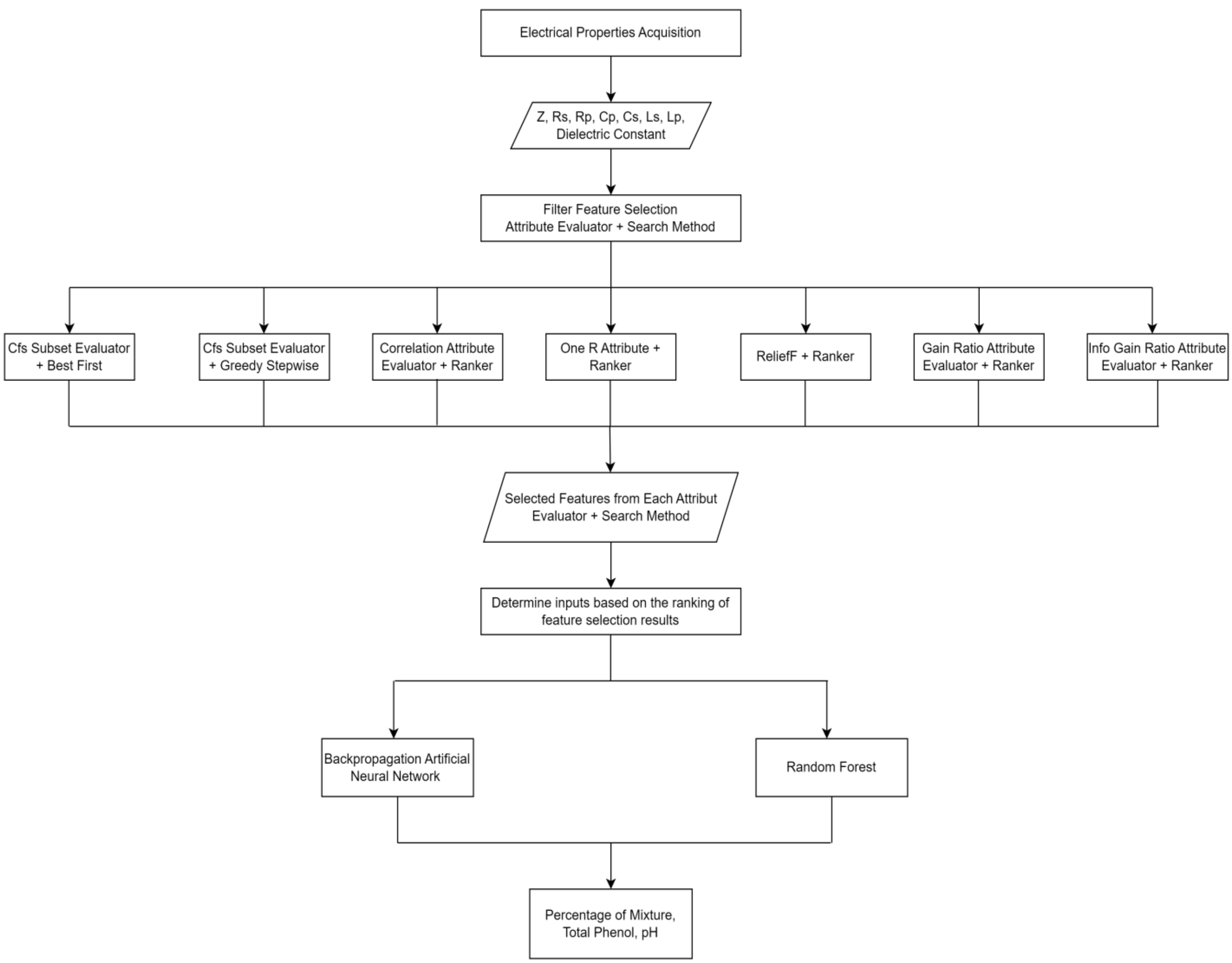

Fig. 1. Filter Feature Selection Experimental Design for Detecting Percentage of Mixtures, Total Phenol, and pH

DOI: http://dx.doi.org/10.25139/ijair.v6i2.9010

A. Materials and Equipment

This study used green beans, arabica civet coffee, and regular Arabic coffee from PT. Perkebunan Nusantara XII, Banyuwangi, Indonesia. The limitation of this study is the lack of variety of civet coffee and regular coffee used in the study. The tools for measuring electrical properties are LCR Meter HIOKI IM3523 and Stavolt. The bioelectric data was processed using Intel Core i7-1165G7 2.80 GHz 8 GB RAM. The software used is HIOKI IM3523 LCR Meter GUI, Microsoft Excel 2021, Waikato Environment for Knowledge Analysis (WEKA) 3.8.6, Python 3.11.5, and Visual Studio Code.

B. Electrical Properties Acquisition

In data collection, 160 green beans were used for each mixture, while the calculation was based on unit seed. The civet and regular coffee mixture proportions in this study were 0%, 10%, 30%, 40%, 50%, 70%, 90%, and 100% for electrical properties acquisition based on [7], where green beans at a predetermined percentage are placed between calibrated parallel plate capacitors. After that, the value of bioelectric properties, including Z, Rs, Rp, Cp, Cs, Ls, Lp, and dielectric constant at 2368-12000 Hz, was measured. Initially, measurements of the electrical properties were carried out in different frequency ranges, 50 Hz - 33000 Hz, 33,001 Hz - 66000 Hz, and 66,001 Hz - 100,000 Hz, to make it easier to see data trends. The frequency selected was 2368 Hz - 12000 Hz because the electrical properties can distinguish civet coffee based on the percentage in that range. The measurement value of electrical properties is measured using dielectric methods through parallel plate capacitors. At the selected frequency, bioelectric data was taken from 22 frequency points with three repetitions and eight mixtures so that the bioelectric data amounted to 528 data for filter model feature selection.

C. Filter Feature Selection

Filter feature selection is a machine learning technique that employs statistical analysis to identify relevant subsets of features, thereby enhancing classification and clustering outcomes [20][21]. The filter feature selection method utilizes several evaluator attributes, including:

a) Cfs Subset Evaluator: This evaluator assesses the worth of a subset of attributes by considering each feature's predictive ability and the degree of redundancy among them [22].
b) Correlation Attribute Evaluator: This evaluator measures the worth of an attribute by calculating the correlation (Pearson's correlation) between the attribute and the class label [22].
c) One R Attribute Evaluator: This evaluator assesses the worth of an attribute using the OneR classifier.
d) ReliefF: Developed by Kononenko in 1994, the ReliefF algorithm uses a weighting technique to measure the significance of features in classification contexts. ReliefF weights are continuous values that enable features to be ranked based on their relevance. The fundamental principle of ReliefF is to rank features according to their ability to differentiate between classes that are near and far from each other [23].
e) Gain Ratio Attribute Evaluator: This evaluator measures the worth of an attribute by calculating the gain ratio concerning the class.
f) Info Gain Attribute Evaluator: This evaluator assesses the worth of an attribute by measuring the information gained concerning the class.

In addition, three search methods are also used, i.e.:

a) Best First: It uses greedy hill climbing with a backtracking feature to search for the space of attribute subsets. One benefit of Best First is its backtracking capability, which allows it to return to the prior subset until it finds the features corresponding to the class.
b) Greedy Stepwise: Search the space of attribute subsets in a greedy forward or backward manner. A search strategy focused on local changes is called greedy stepwise. Adding or deleting a single feature from the subsets collection is called a local modification. Forward selection refers to algorithms based on feature addition, whereas reverse selection refers to algorithms based on feature removal. The greedy stepwise search strategy keeps adding and removing features until it finds the best feature that strongly correlates with the class.
c) Ranker: The rank attributes are based on their evaluations, providing a simple and efficient way to prioritize features.

D. Model Design

1) Artificial Neural Network (ANN): ANN is a non-linear statistical model that is widely used to solve pattern recognition, modeling, and control problems in various fields such as engineering and science [24]–[26]. ANN is inspired by the operating system of the human brain [27]. ANN works based on the training stage and determines the error rate through the validation and testing stages. ANN can solve complex (non-linear) problems and predict more than one output. This study used backpropagation (BP) ANN prediction of civet coffee mixtures, total phenol, and pH. The BP is a supervised algorithm with multiple layers, including an input layer representing the input variables of the problem, an output layer representing the dependent variables, and one or more hidden layers consisting of nodes to learn non-linear data. ANN was designed with two hidden layers comprising 40 nodes each. This configuration was selected based on previous studies [19], demonstrating that two hidden layers balance model accuracy and computational efficiency. Models with fewer hidden layers were found inadequate for capturing complex data patterns. In comparison, those with more hidden layers resulted in excessively long training times, hindering the practicality of the proposed automatic detection tool. The ANN employed a tansig activation function in both hidden and output layers, chosen for its ability to handle non-linear transformations effectively. The trainlm learning algorithm was selected due to its combination of the steepest descent and Gauss-Newton methods, offering rapid convergence and enhanced stability during training. A learning rate of 0.1 and momentum of 0.9 was optimized through sensitivity analyses to ensure efficient weight updates and to prevent the model from becoming trapped in local minima.

2) Random Forest (RF): RF is a supervised machine-learning technique used for classification and regression [28][29]. The main benefit of RF is its ability to use the bagging approach to manage large data dimensionality and multicollinearity [30]. Many different training sample sets will be generated using the random forest technique using randomly chosen qualities, and each sample set will construct a decision tree [31]. With the ntree parameter set to 100, this investigation used RF as a comparative model. The number of trees in the forest is known as the ntree. Ntree 100 was selected because it maintains training times while balancing model correctness and computational load, guaranteeing adequate diversity across the decision trees.

3) Evaluation Criteria: Evaluation criteria are used to determine the error rate of the model built. In this study, the model evaluation criteria are based on the lowest Validation MSE value using Equation (1).

$$MSE\ Validation = \left[\frac{1}{N}\sum_{i=1}^{N}(Y_{model,i} - Y_{measurement,i}\right]^2 \quad (1)$$

III. RESULT AND DISCUSSION

This study used bioelectric data to estimate total phenolics, acidity (pH), and percentage of regular coffee blends. The dielectric method produces eight parameters of electrical properties in the form of impedance (Z), series inductance (Ls), parallel inductance (Lp), series capacitance (Cs), parallel capacitance (Cp), series resistance (Rs), parallel resistance (Rp), and dielectric constant. Feature selection was employed to identify the most influential properties for predicting the target variables [32]–[34]. Table 1 reveals that impedance (Z) consistently exhibited the highest weight across various attribute evaluators, indicating its strong correlation with the percentage of civet coffee mixture, total phenol content, and pH. This finding aligns with previous research highlighting the significant influence of moisture content on impedance measurements in hygroscopic materials like coffee beans [7]. The relationship between impedance and civet coffee blend is shown by the results of research [7], where 100% civet coffee has the lowest impedance value and increases as the percentage of mixture increases. The results of [7] also show that 100% civet coffee has a higher moisture content than 0% civet coffee. The lower the impedance, the higher the conductivity of the material so that the moisture content of the material is high, and vice versa. The dielectric method has two frequency ranges: Radio Frequency (RF) from 3 kHz - 300 MHz and Microwave Heating from 300 - 300 GHz [35]. This research includes a frequency range of 2368 Hz - 12000 Hz (2.368 kHz - 12 kHz) in the RF. In both frequency ranges, water content significantly affects electrical properties due to its polar nature and ability to absorb electric current. Due to its fermentation process, the high water content in coffee beans, particularly in civet coffee, likely contributes to the strong correlation between impedance and the target variables. The physical structure of the coffee beans, including variations in density and composition, also influences bioelectric properties [36].

Table 1 shows the order of the selected features based on the weights of each attribute evaluator and search method. The determination of input in ANN and RF according to WEKA output ranking. The ANN topology used in this electrical properties data feature selection is based on [19]: 40 nodes in 1st hidden layer and 40 nodes in 2nd hidden layer; tansig activation function that is used in the hidden layer and output layer; trainlm learning algorithm, learning rate at 0.1 and momentum at 0.9. This sensitivity analysis is carried out on several parameters, i.e., nodes in the hidden layer, hidden layer, activation function, learning rate, and momentum is conducted because they play a vital role in the performance of ANN [7][19][37]–[39]. The hidden layer is between the input and output layers in the ANN topology. This research uses two hidden layers because, based on [40], two hidden layers provide excellent accuracy and time complexity results.

On the other hand, one hidden layer is not suitable for complex data. The results of [41] using five hidden layers require a long training time, so it is not suitable for tool design as a major goal of this research. Therefore, this research uses two hidden layers. A neuron's output is determined by the activation function, which is based on the operation performed on the input. In this study, we used the tansig activation function based on the sensitivity analysis conducted before this study. The trainlm algorithm combines the steepest descent method and the Gauss-newton algorithm. The trainlm algorithm combines the Gauss-newton algorithm with the steepest descent method. Combining the two algorithms in trainlm aims to improve the model accuracy and reach the target faster when the error reaches the minimum point [42][43]. The trainlm learning algorithm was also selected based on the sensitivity analysis in the previous study. A sensitivity analysis of learning rate and momentum is necessary as both indicators play a role in weight and bias changes during training. During training, momentum plays a role in updating weights and preventing local minima, while the learning rate controls weight changes.

TABLE I
FEATURE SELECTION OF ELECTRICAL PROPERTIES

DOI: http://dx.doi.org/10.25139/ijair.v6i2.9010

No	Attribute Evaluator	Search Method	Electrical Properties	Weight	Ranking
1	Cfs Subset Evaluator	Best First	Z, Lp	-	-
		Greedy Stepwise	Z, Lp	-	-
2	Correlation Attribute Evaluator	Ranker	Z	0.0311	1
			Lp	0.0272	2
			Ls	0.026	3
			Rp	0.0225	4
			Rs	0.0198	5
			Cs	0	6
			Cp	0	7
			Dielectric Constant	0	8
3	One R Attribute	Ranker	Z	16.429	1
			Rs	16.25	2
			Rp	14.464	3
			Lp	13.125	4
			Cs	12.5	5
			Dielectric Constant	12.5	6
			Cp	11.789	7
			Ls	11.696	8
4	ReliefF	Ranker	Rs	0.0000734	1
			Rp	0.0000155	2
			Dielectric Constant	0.0000132	3
			Cp	0	4
			Cs	0	5
			Lp	-0.0000292	6
			Ls	-0.000195	7
			Z	-0.0002398	8
5	Gain Ratio Attribute Evaluator	Ranker	Z	0.0311	1
			Lp	0.0272	2
			Ls	0.026	3
			Rs	0.0225	4
			Cp	0.0198	5
			Cs	0	6
			Rp	0	7
			Dielectric Constant	-0.0000188	8
6	Info of Gain Ratio Attribute Evaluator	Ranker	Z	0.0602	1
			Lp	0.0514	2
			Ls	0.0421	3
			Rs	0	4
			Cp	0	5
			Cs	0	6
			Rp	0	7
			Dielectric Constant	0	8

Table II shows the lowest MSE Validation for ANN modeling, the Correlation Attribute Evaluator with the Ranker search method. The selected features, i.e., Z, Lp, Ls, Rp, and Rs, have an MSE Validation of 0.0479. For random forest modeling, the parameter used is ntree.

TABLE II
MSE VALIDATION OF ARTIFICIAL NEURAL NETWORK AND RANDOM FOREST FOR ELECTRICAL PROPERTIES FEATURE SELECTION

No	Attribute Evaluator	Search Method	Input	ANN	RF
1	Cfs Subset Evaluator	Best First	Z, Lp	0.3358	0.0962
		Greedy Stepwise	Z, Lp	0.3358	0.0962
2	Correlation Attribute Evaluator	Ranker	Z, Lp	0.3358	0.0962
			Z, Lp, Ls	0.3358	0.0759
			Z, Lp, Ls, Rp	0.0567	0.1038
			Z, Lp, Ls, Rp, Rs	**0.0479**	0.1775
			Z, Lp, Ls, Rp, Rs, Cs	0.3687	0.1745
			Z, Lp, Ls, Rp, Rs, Cs, Cp	0.0904	0.1755
			Z, Lp, Ls, Rp, Rs, Cs, Cp, Dielectric Constant	0.0686	0.1760
3	One R Attribute	Ranker	**Z, Rs**	0.9767	**0.0623**
			Z, Rs, Rp	0.0542	0.7975
			Z, Rs, Rp, Lp	0.1246	0.2014
			Z, Rs, Rp, Lp, Cs	0.8352	0.1995
			Z, Rs, Rp, Lp, Cs, Dielectric Constant	0.0808	0.2002
			Z, Rs, Rp, Lp, Cs, Dielectric Constant, Cp	0.0485	0.1960
			Z, Rs, Rp, Lp, Cs, Dielectric Constant, Cp, Ls	0.0686	0.1760
4	ReliefF	Ranker	Rs, Rp	0.9049	0.9027
			Rs, Rp, Dielectric Constant	0.2026	0.9916
			Rs, Rp, Dielectric Constant, Cp	0.0480	0.9985

E-ISSN: 2686-6269

No	Attribute Evaluator	Search Method	Input	ANN	RF
			Rs, Rp, Dielectric Constant, Cp, Cs	0.3157	0.9991
			Rs, Rp, Dielectric Constant, Cp, Cs, Lp	0.5421	0.7892
			Rs, Rp, Dielectric Constant, Cp, Cs, Lp, Ls	0.4167	0.7109
			Rs, Rp, Dielectric Constant, Cp, Cs, Lp, Ls, Z	0.0686	0.1760
5	Gain Ratio Attribute Evaluator	Ranker	Z, Lp	0.3358	0.0962
			Z, Lp, Ls	0.3358	0.0759
			Z, Lp, Ls, Rs	0.0587	0.7189
			Z, Lp, Ls, Rs, Cp	0.0869	0.8109
			Z, Lp, Ls, Rs, Cp, Cs	0.5421	0.7229
			Z, Lp, Ls, Rs, Cp, Cs, Rp	0.0904	0.8229
			Z, Lp, Ls, Rs, Cp, Cs, Rp, Dielectric Constant	0.0686	0.1760
6	Info of Gain Ratio Attribute Evaluator	Ranker	Z, Lp	0.3358	0.0962
			Z, Lp, Ls	0.3358	0.0759
			Z, Lp, Ls, Rs	0.0587	0.7189
			Z, Lp, Ls, Rs, Cp	0.0869	0.8109
			Z, Lp, Ls, Rs, Cp, Cs	0.5421	0.7229
			Z, Lp, Ls, Rs, Cp, Cs, Rp	0.0904	0.8229
			Z, Lp, Ls, Rs, Cp, Cs, Rp, Dielectric Constant	0.0686	0.1760

Based on [44], ntree affects the training degree and accuracy of the RF model. The RF modeling's lowest MSE Validation at 0.0623 uses the One R Attribute with the ranker search method and the selected features, i.e., Z and Rs. The ntree value used is 100. From the research results, the best model for predicting the percentage mixture of civet coffee, total phenol, and pH uses ANN with five selected features and MSE Validation of 0.0479. ANN demonstrated superior predictive accuracy, indicating a strong fit to the data and effective modeling of the non-linear relationships between the electrical properties and the target variables. While RF also effectively predicted the target variable, ANN provided a more accurate model for this application. Feature selection significantly enhanced model performance. For the ANN model using the Correlation Attribute Evaluator, MSE Validation was reduced by 30.174% from an initial higher value when all features were considered. Similarly, there was a 64.602% reduction in MSE Validation for the RF model utilizing the One R Attribute. It can be seen that when eight electrical properties are used as input for both ANN and RF, the validation MSE is higher than when only a few features are selected. This substantial decrease underscores the importance of selecting relevant features to improve model accuracy and reduce error rates. Especially if it is to be implemented in the tool, when this feature is implemented, it is necessary to explore civet coffee and regular coffee samples from various regions in Indonesia to generalize the model.

IV. CONCLUSION

This study investigated the potential of electrical properties for predicting key quality parameters of civet coffee: the percentage mixture, total phenol content, and pH. Employing feature selection techniques, we identified impedance (Z) as a highly informative feature, consistently ranking among the top across multiple attribute evaluators. This finding suggests a strong correlation between impedance and the target variables, likely attributed to the influence of water content in the coffee blends, which significantly impacts impedance measurements. The comparative analysis indicates that ANN outperforms RF in predicting percentages of a mixture of civet coffee, pH, and total phenol based on electrical properties. The best-performing model, an Artificial Neural Network (ANN), achieved a Mean Squared Error (MSE) validation of 0.0479. MSE Validation was reduced by 30.174% from an initial higher value when all features were considered using the Correlation Attribute Evaluator and Ranker search method. This model utilized five selected features: Z, Lp, Ls, Rp, and Rs. While these results demonstrate the feasibility of using electrical properties to predict civet coffee quality, further research is warranted to validate these findings with a larger and more diverse dataset. Additionally, exploring the influence of other factors, such as bean origin, on the electrical properties could enhance the model's predictive capabilities.

REFERENCES

[1] M. Muzaifa, D. Hasni, and F. Rahmi, "What is kopi luwak? A literature review on production, quality and problems," in *IOP Conference Series: Earth and Environmental Science*, 2019, vol. 365, no. 1, p. 12041.
[2] A. Patria, A. Abubakar, and M. Muzaifa, "Physicochemical and sensory characteristics of luwak coffee from Bener Meriah, Aceh-Indonesia," in *IOP Conference Series: Earth and Environmental Science*, 2018, vol. 196, no. 1, p. 12010.
[3] I. E. Panggabean, *Buku pintar kopi*. AgroMedia, 2011.
[4] A. H. Ishak, S. Moidin, S. N. M. Rejab, and M. H. Mahaiyadin, "The Commercialisation of Halal and Tayyib Palm Civet Coffee," *al-Qanatir Int. J. Islam. Stud.*, vol. 29, no. 1, pp. 46–56, 2023.
[5] M. Yulia and D. Suhandy, "Indonesian palm civet coffee discrimination using UV-visible spectroscopy and several chemometrics methods," in *journal of physics: conference series*, 2017, vol. 835, no. 1, p. 12010.
[6] U. Jumhawan, S. P. Putri, T. Bamba, and E. Fukusaki, "Quantification of coffee blends for authentication of Asian palm civet coffee (Kopi Luwak) via metabolomics: A proof of concept," *J. Biosci. Bioeng.*, vol. 122, no. 1, pp. 79–84, 2016.

DOI: http://dx.doi.org/10.25139/ijair.v6i2.9010

[7] S. Widyaningtyas, S. Sucipto, and Y. Hendrawan, “DETEKSI PEMALSUAN KOPI LUWAK MENGGUNAKAN SIFAT BIOLISTRIK DAN JARINGAN SARAF TIRUAN,” *J. Teknol. Pertan.*, vol. 19, no. 3, pp. 161–172, 2018.
[8] M. Jeszka-Skowron, A. Zgoła-Grześkowiak, and T. Grześkowiak, "Phenolic compounds in coffee and tea beverages," *Anal. methods Determ. Bioact. Compd. Elem. food*, pp. 31–81, 2021.
[9] S. O. Nelson and S. Trabelsi, "Factors Influencing the Dielectric Properties of Agricultural and Food Products," *J. Microw. Power Electromagn. Energy*, vol. 46, no. 2, pp. 93–107, 2012, doi: 10.1080/08327823.2012.11689828.
[10] S. Sucipto, M. Anna, M. Arwani, and Y. Hendrawan, "A rapid classification of wheat flour protein content using artificial neural network model based on bioelectrical properties," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 2, pp. 920–927, 2019.
[11] R. Sumranbumrung, A. Siritaratiwat, and A. Kruesubthaworn, "The simulation of sugar content estimation of sugarcane based on dielectric properties for microwave imaging technique," *J. Eng. Innov.*, vol. 17, no. 3, pp. 35–42, 2024.
[12] A. Purohit, A. Kumari, A. Roy, and A. Mohan, "Determination of dielectric properties and predictive modeling for designing radio-frequency heating of ground beef," *Front. Food Sci. Technol.*, vol. 2, p. 960471, 2022.
[13] C. Dong *et al.*, "Rapid detection of catechins during black tea fermentation based on electrical properties and chemometrics," *Food Biosci.*, vol. 40, p. 100855, 2021.
[14] J. Sun, Y. Tian, X. Wu, C. Dai, and B. Lu, "Non-destructive detection for moisture content in green tea based on dielectric properties and VISSA-GWO-SVR algorithm," *J. Food Process. Preserv.*, vol. 44, no. 5, p. e14421, 2020.
[15] G. Ansari, T. Ahmad, and M. N. Doja, "Hybrid filter–wrapper feature selection method for sentiment classification," *Arab. J. Sci. Eng.*, vol. 44, pp. 9191–9208, 2019.
[16] X. Zhang, Q. Zhang, M. Chen, Y. Sun, X. Qin, and H. Li, "A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method," *Neurocomputing*, vol. 275, pp. 2426–2439, 2018.
[17] B. Pourmohammadali, M. H. Salehi, S. J. Hosseinifard, I. E. Boroujeni, and H. Shirani, "Studying the relationships between nutrients in pistachio leaves and its yield using hybrid GA-ANN model-based feature selection," *Comput. Electron. Agric.*, vol. 172, p. 105352, 2020.
[18] T. N. Pham, L. Van Tran, and S. V. T. Dao, "Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection," *IEEE access*, vol. 8, pp. 189960–189973, 2020.
[19] Y. Hendrawan, S. Widyaningtyas, and S. Sucipto, "Computer vision for purity, phenol, and pH detection of Luwak Coffee green bean," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 6, pp. 3073–3085, 2019.
[20] Z. Zhao, L. Wang, H. Liu, and J. Ye, "On similarity preserving feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 3, pp. 619–632, 2011.
[21] T. Dokeroglu, A. Deniz, and H. E. Kiziloz, "A comprehensive survey on recent metaheuristics for feature selection," *Neurocomputing*, vol. 494, pp. 269–296, 2022.
[22] M. A. Hall, "Correlation-based feature subset selection for machine learning," *Thesis Submitt. Partial fulfilment Requir. degree Dr. Philos. Univ. Waikato*, 1988.
[23] E. R. M. Saleh, E. Noor, T. Djatna, and I. TIP, “Seleksi Parameter Dielektrik Penentuan Masa Kadaluwarsa Biskuit (Wafer) dengan Pendekatan Regresi Linier, Feature Selection (Relieff) dan Artificial Neural Network,” *J. Teknol. Ind. Pertan.*, vol. 23, no. 2, 2013.
[24] H. Liang, G. Liu, H. Zhang, and T. Huang, "Neural-network-based event-triggered adaptive control of nonaffine non-linear multiagent systems with dynamic uncertainties," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 5, pp. 2239–2250, 2020.
[25] O. I. Abiodun *et al.*, "Comprehensive review of artificial neural network applications to pattern recognition," *IEEE access*, vol. 7, pp. 158820–158846, 2019.
[26] O. M. Adisa *et al.*, "Application of artificial neural network for predicting maize production in South Africa," *Sustainability*, vol. 11, no. 4, p. 1145, 2019.
[27] M. Islam, G. Chen, and S. Jin, "An overview of neural network," *Am. J. Neural Networks Appl.*, vol. 5, no. 1, pp. 7–11, 2019.
[28] U. A. Umoh, I. J. Eyoh, V. S. Murugesan, and E. E. Nyoho, "Fuzzy-machine learning models for the prediction of fire outbreaks: A comparative analysis," in *Artificial Intelligence and Machine Learning for EDGE Computing*, Elsevier, 2022, pp. 207–233.
[29] P. Dutta, S. Paul, and A. Kumar, "Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19," in *Electronic devices, circuits, and systems for biomedical applications*, Elsevier, 2021, pp. 521–540.
[30] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.
[31] Y. Liu, "Random forest algorithm in big data environment," *Comput. Model. new Technol.*, vol. 18, no. 12A, pp. 147–151, 2014.
[32] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *J. Appl. Sci. Technol. Trends*, vol. 1, no. 1, pp. 56–70, 2020.
[33] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Feature selection for high-dimensional data," *Prog. Artif. Intell.*, vol. 5, pp. 65–75, 2016.

[34] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 science and information conference*, 2014, pp. 372–378.
[35] H. Ramaswamy and J. Tang, "Microwave and radio frequency heating," *Food Sci. Technol. Int.*, vol. 14, no. 5, pp. 423–427, 2008.
[36] S. Ryynänen, "The electromagnetic properties of food materials: a review of the basic principles," *J. Food Eng.*, vol. 26, no. 4, pp. 409–429, 1995.
[37] H. Suprajitno, "Experimental Analysis of Training Parameters Combination of ANN Backpropagation for Climate Classification.," *Math. Model. Eng. Probl.*, vol. 9, no. 4, 2022.
[38] M. Uzair and N. Jamil, "Effects of hidden layers on the efficiency of neural networks," in *2020 IEEE 23rd international multitopic conference (INMIC)*, 2020, pp. 1–6.
[39] S. Sucipto *et al.*, "Prediction of water content, sucrose and invert sugar of sugarcane using bioelectrical properties and artificial neural network," *Int. Food Res. J.*, vol. 25, no. 6, pp. 2674–2680, 2018.
[40] G. T. Kumar and R. Khalid, "Optimizing Deep Neural Network Architecture: A Tabu Search Based Approach," *Comput. Sci. springer*, 2018.
[41] S. Asthana, A. Pandit, and A. Bahrdwaj, "Analysis of multiple hidden layers vs. accuracy in performance using back propagation neural network," *Indian J. Sci. Technol.*, vol. 10, no. 4, pp. 1–4, 2017.
[42] M. R. S. Mohd, J. Johari, F. A. Ruslan, N. A. Razak, S. Ahmad, and A. S. M. Shah, "Analysis of Training Function for NNARX in Solar Radiation Prediction Modeling," in *Proceedings of the 8th International Conference on Computational Science and Technology: ICCST 2021, Labuan, Malaysia, 28–29 August*, 2022, pp. 619–632.
[43] K. Mehranzamir, Z. Abdul-Malek, H. N. Afrouzi, S. V. Mashak, C. Wooi, and R. Zarei, "Artificial neural network application in an implemented lightning locating system," *J. Atmos. Solar-Terrestrial Phys.*, vol. 210, p. 105437, 2020.
[44] Y. Liu, H. Chen, L. Zhang, and Z. Feng, "Enhancing building energy efficiency using a random forest model: A hybrid prediction approach," *Energy Reports*, vol. 7, pp. 5003–5012, 2021.