PAPER • OPEN ACCESS

Classification of *Civet* and *Canephora* coffee using Support-Vector Machines (SVM) algorithm based on order-1 feature extraction

To cite this article: R Z H Suyoto *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1173** 012006

View the article online for updates and enhancements.

Classification of *Civet* and *Canephora* coffee using Support-Vector Machines (SVM) algorithm based on order-1 feature extraction

R Z H Suyoto, M Komarudin[*], G F Nama, and T Yulianti

Informatics Engineering Study Program, Department of Electrical Engineering, Faculty of Engineering, Universitas Lampung, Jl. Prof. Sumantri Brojonegoro No.1 Bandar Lampung 35145, Indonesia

*Email: m.komarudin@eng.unila.ac.id

Abstract. Identification of coffee beans types is a process carried out in determining the category of civet and robusta coffee. The selection of civet coffee manually by the community in determining the variety of coffee beans. Manual selection makes the process slow. There is no certainty whether the selected coffee stated original civet coffee or a mixture with Robusta coffee. It needs to be built a system in identifying the types of coffee beans. In this study, the classification technology used for the selection process of Civet and Robusta coffee beans. This research uses first-order feature extraction with the Support-Vector Machines (SVM) algorithm, which aims to recognize civet and robusta coffee beans' patterns using texture analysis on grayscale images and feature extraction. In the study, the authors used the Rapid Application Development approach to build the selection system. The data used in this study were 120 images, consisting of 110 training data and 10 test data. The accuracy in identifying the types of coffee beans by using this method is 87.27%.

Keywords: Civet coffee, robusta coffee, classification, SVM

1. Introduction

Lampung Province is the second-largest producer of Robusta coffee exports, with production in 2013-2017 of 110.12 thousand tons of coffee per year. In 2015 coffee production was concentrated in five districts. West Lampung Regency with the highest production reached 52.65 thousand tons or 47.81% of the total Robusta coffee production in Lampung Province. Other districts namely, Tanggamus Regency with a production yield of 26.92% or 29.64 tons, North Lampung Regency with production results 9.41% or 10.37 tons, Way Kanan Regency with a production yield of 8.38% or 9.23 tons, Pesisir Barat Regency with a production yield of 4.06% or production reaching 4.47 tons. Center for Agricultural Data and Information Systems - Ministry of Agriculture 2017 [1].

Civet Coffee is a delicious taste of coffee beans that are loved by coffee lovers or coffee connoisseurs in the world because coffee produced from imperfect digestion by civet civets makes the most expensive coffee on the world market [2]. Unique coffee whose coffee beans pass through the fermentation process in wet processing can improve the body (thick taste) and Milky (fat taste) [3]. Civet itself is a wild animal that consumes good quality fruits in the forest or farmer's garden, including coffee fruit. Civet

has a very sensitive sense of smell, so clever to choose which coffee fruit is the best and has been cooked perfectly so that the quality Civet coffee produced is of high quality.

Because coffee that is eaten by civet coffee is quality coffee making civet coffee known as the most expensive and distinctive coffee (unique) to date Civet coffee production is still in very limited quantities. Civet coffee itself has a physical color of dark skin and brown (enzymatic browning), has epidermis and tends to be darker [4].

Civet coffee prices are expensive and limited production, leading to fraud by unscrupulous by mixing civet coffee with other types of coffee such as Robusta. The fraud that occurred in the person gave rise to the idea to create an application that could distinguish between Civet coffee and Robusta coffee. In this study distinguish between civet and robusta coffee by classifying the image of civet and robusta coffee beans, general the stages in the image classification process are, inserting images, converting RGB images to grayscale, extracting features, training, testing and measuring the accuracy of images. classification aims to classify the object into certain classes based on the value of attributes associated with the object being observed.

This study aims to identify the image of coffee beans using texture analysis on grayscale images and color feature extraction in color images. Grayscale image elements have elements of mean, entropy, variance, skewness, and kurtosis. The result of feature contraction will be the input value for SVM to classify the type of coffee beans used is the image of coffee beans consisting of 120 coffee beans consisting of 60 types of robusta coffee beans and 60 types of civet beans.

2. Literature review

2.1. Previous research
Sri Citra, conducted research with use image processing techniques for coffee bean processing. Classifying coffee beans into four classes based on the image area, height, width, and parameters. However, due to the time of manual verification, it was not very thorough, so from the results of the program around 81% success [5]. Wahyu aji, researched the identification of coffee fruit maturity using artificial neural networks to study patterns and identify coffee maturity using color features in the fruit. Based on the results of the test produces the best image with 100% accuracy without error in every test performed [6]. Other research on applying the Rapid Application Development (RAD) methodology in developing training application systems, using RAD enables rapid implementation in real environmental systems. RAD's research, it proved to be the right method for this work project because it served the purpose of rapid implementations for small systems [7]. Zhibin Liau, in his research on evaluating the ability of cooperative income villages by using SVM algorithm to evaluate Supplier Cooperative Design Ability (SCDA), introduces 2 specific gravity factors and sample weight factors. As a result, the accuracy and ability of SVM have a strong classification function and good learning so it has broad application prospects in the SCDA section [8]. Beynon-Davies, in his research on integrating Rapid Application Development (RAD), discusses the development of systems using RAD on business and uncertainty and Participatory Design (PD). RAD and PD systems are reviewed in the area of possible cross-fertilization, which relates to aspects of RAD approach research, specifically scenario concepts, and detailed design for work on RAD [9]. Yun Lin, in his research on the classification of texts based on the SVM and K-Nearest Neighbor (KNN) algorithm, combines the SVM and KNN algorithm. As a result, the classification of SVM and KNN can obtain better accuracy from the grouping of SVM and KNN and can improve computing that is very small [10].

2.2. Feature extraction
Feature extraction is a method of characterization based on image histogram characteristics. The histogram shows the probability value of the appearance of the degree of gray pixel values in an image. Values that exist in the resulting histogram, can be calculated with several parameters of first-order characteristics, including Mean, Skewness, Variant, Kurtosis, and Entropy [11].

Based on the values in the resulting histogram, several characteristic parameters can be calculated, including the mean, variance, skewness, kurtosis, and entropy, namely:

1. Mean Feature (μ) shows the size of the dispersion of an image

$$\mu = \sum_{n=0}^{N} f_n \, P(f_n) \tag{1}$$

Explanation:
f_n = gray intensity value
$P(f_n)$ = histogram value

2. Variance Feature (σ2) shows the variation of elements in the histogram of an image

$$\sigma^2 = \sum_{n=0}^{N}(f_n - \mu)^2 \, P(f_n) \tag{2}$$

Formula Description:
f_n = gray intensity value
μ   = mean value
$P(f_n)$ = histogram value

3. Skewness Feature (a3) indicates the relative skewness level of the histogram curve of an image.

$$\alpha_3 = \frac{1}{\sigma^3} \sum_{n=0}^{N}(f_n - \mu)^3 P(f_n) \tag{3}$$

Formula Description:
a3   = standard deviation of gray intensity values
f_n = gray intensity value
μ   = mean value
$P(f_n)$ = histogram value

4. Characteristics of Kurtosis (a4) shows the relative degree of shock of the histogram curve of an image

$$\alpha_4 = \frac{1}{\sigma^4} \sum_{n=0}^{N}(f_n - \mu)^4 P(f_n) - 3 \tag{4}$$

Formula Description:
$\alpha\,4$ = standard deviation of gray intensity values
f_n = gray intensity value
M   = mean value
$P(f_n)$ = histogram value

5. Entropy (H) Characteristics Indicates the size of the irregular shape of an image

$$H = \sum_{n=0}^{N} P(f_n).^2 \log P(f_n) \tag{5}$$

Formula Description:
f_n = gray intensity value
$P(f_n)$ = histogram value

2.3. Support Vector Machine (SVM)
Support Vector Machine one of the methods in supervised learning that is usually used for classification (such as Support Vector Classification) and regression (Support Vector Regression). Related to the number of decision boundaries cannot be modeled with a form or linear equation, one must model a decision boundary as a non-linear function [12]. SVM is used to find the best hyperplane by maximizing the distance between classes. The hyperplane is a function that can be used for separating classes. In 2-D functions used for class, classifications are referred to as lines whereas, functions used for class classifications in 3-D are called planes similarly, whereas functions used for classifications in classrooms of higher dimensions are called hyperplanes. In SVM the outermost data object that is closest to the hyperplane is called a support vector. Objects called support vectors are the most difficult to classify due to positions that almost overlap with other classes Related to the number of decision

boundaries, it cannot be modeled with a form or linear equation, it must model a decision boundary as a non-linear function.

3. Research methods
The research method used in the implementation and execution of this thesis is the Rapid Application Development (RAD) method, which is a model of an international software development process, especially for short processing times [13]. The research flow is in Figure 1.
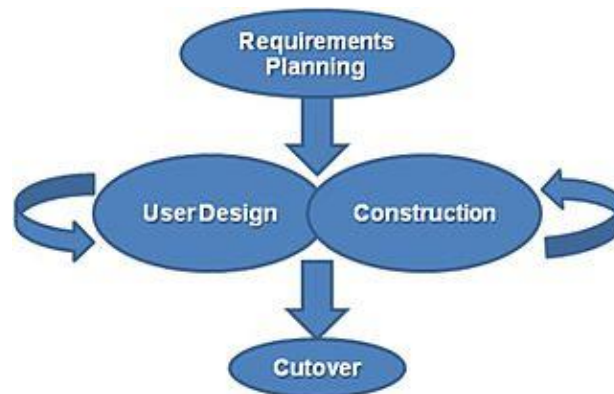

Figure 1. RAD Research stage.

The stages of the RAD method:
1. Requirements planning or Need Planning.
   At this stage, the authors design the needs and planning carried out to determine the goals and expectations for research as well as current and potential problems that need to be addressed. In this stage the writer collected the materials and made observations on the needs of the civet and robusta coffee bean samples, the methods used and the MATLAB application compared to the information analyzed to obtain system specifications.
2. User Design or the Design Process
   At this stage, the authors carry out the design process in the form of modeling the required data based on application modeling and defining its attributes and their relationships with other data. This stage the author uses ERD and LRS for database modeling so that it can know what attributes are needed and how the relation of the data, as well as making the application mockup process to be used. At this stage the authors also made changes during the process, starting from working on the system, in the first experiment the authors used GLCM feature extraction, but during the research work, the authors experienced changes in GLCM feature extraction into first-order feature extraction due to the value of features that could implement research.

   Furthermore, in the classification method used, the previous author used the KNN classification, however, there was a change to the SVM classification because SVM classification uses the separating elements between classes using hyperplane lines, and the mockup display on the Matlab used. Furthermore, the authors use the use case as a process identification, use case definition, and activity diagrams as modeling the application process. This stage is done to get the best design of an application system.
3. Construction (construction) and Implementation.
   At this stage, it breaks down into preparations for a fast workmanship framework, the development of programs and applications, coding in systems, units, integration, and system testing. At the time of the research, the author made the application using the MATLAB 2017b GUI, then coding the system until the application was successfully made so that it could be implemented on the system. In making and implementing changes occur when the author does the work process such as adding features to the Matlab GUI display that was not previously displayed then it is displayed in the GUI,

and the display successfully identifies from coffee beans that were not displayed before, the author displays the results of the identification of the application system.

4. Cutover

At this stage, the application was made successfully. Through this stage, the writer also implements the application functions that have been defined about defining the data[14]. After a successful application, the program will be tested whether there is an error or not before it is published. The most important thing is user involvement is needed so that the system can be developed if something goes wrong and gives satisfaction to the user, other than that the old system does not need to be run in parallel with the new system or the system that has been successfully published.

4. Results and discussion

4.1. Implementation

This study uses a comparison of training (training data) and testing (test data) of 70:30 because the typical image research uses the ratio of 70:30. The training data used amounted to 110 images and the test data amounted to 10 images. The image used as training data is taken from datasets with sequence numbers 1 to 55 in each class, and images used as test data are taken from datasets with sequence numbers 1 to 5 in each class. The following is the display to enter the image.
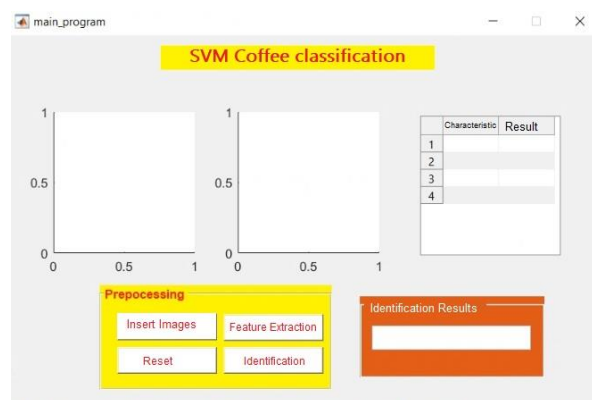

Figure 2. Display of the Matlab GUI application.

Based on the image is the display after pressing the "Enter image" then the image conversion will exit from the RGB Image to Grayscale Image. In pre-processing stages are carried out as is done. Next to the Grayscale image is a table that contains the conversion of images or the value of an image and the features that are in it.

Table 1. The result of the data training feature extraction value

Sample	Mean	Ent	Var	Skew	Kurto
Civet 1	119	6.62	4318	0.29	-1.72
Civet 2	115	6.64	4441	0.34	-1.70
Civet 3	108	6.68	5142	0.28	-1.69
Civet 4	124	6.50	4942	0.06	-1.81
Civet 5	113	6.87	4457	0.21	-1.68
Robusta 1	124	6.95	4395	0.08	-1.67
Robusta 2	114	6.69	3761	0.24	-1.50
Robusta 3	112	6.61	3746	0.25	-1.70
Robusta 4	118	6.84	3311	0.32	-1.42
Robusta 5	111	6.89	3045	0.39	-1.34

Table 3 is the result of feature extraction testing using the first-order method which included 5 sample data from each type of coffee from a total of 110 sample data types. The results of feature extraction can be seen from the output values such as mean, entropy, variance, skewness and kurtosis. The above value is used as a reference to classify 110 test sample data in accordance with the standard values.

4.2. Testing and identification

In the testing stage, new samples are used that have been distinguished from training data samples or (training) consisting of 5 images or image objects of each type of coffee bean. In the application, the test is carried out to see whether the application has been able to read or identify an image in the training data that has produced output in accordance with the expected pattern. The following results identify the types of coffee beans as Table 2.

Table 2. Application identification results

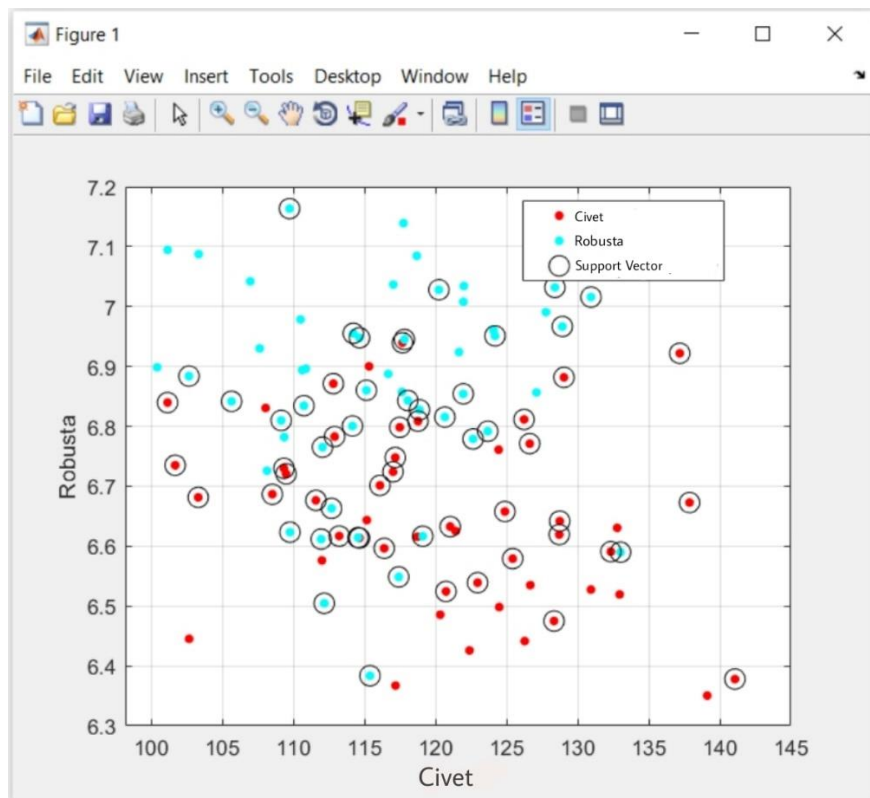| Sample	Input	Output	Information
1	Civet	Civet	True
2	Civet	Civet	True
3	Civet	Civet	True
4	Civet	Civet	True
5	Civet	Civet	True
1	Robusta	Robusta	True
2	Robusta	Robusta	True
3	Robusta	Robusta	True
4	Robusta	Robusta	True
5	Robusta	Robusta	True

Figure 3. Support Vector Machine (SVM) classification.

Based on Table 3 is a truth table 10 of 110 sample data types of coffee beans. In the test, there were 96 coffee beans declared correct and stated 14 coffee beans wrong because there was an error in this classification into an SVM algorithm having an accuracy rate of 87.27%.

Figure 3 shows the classification formed by a support vector machine or SVM. Distinguishing between types of civet coffee and robusta coffee with red color as civet coffee and robusta coffee in blue and circle indicates a vector machine that divides it using 110 training data samples.

4.3. Cutover

In addition to conducting their testing or functional, the author conducts testing of the community to conduct testing. The author tests the community or user. In testing conducted on the system using training data that is tested by first the user runs the program, then the user enters an image, then the user performs the feature extraction and the user identifies. In testing the user program successfully identified by the program.

5. Conclusions

Based on the results of trials that have been carried out in this study, it is proven that the SVM algorithm can be implemented to classify the types of coffee beans. At an accuracy of 87.27% by successfully recognizing 96 coffee beans out of 110 and not recognizing 14 beans, then for subsequent research, it can be developed with other algorithms and methods. Coffee set image data types need to be added so that the level of accuracy can be increased. The author approves the process of introducing types of coffee beans using an android application that is far more easily accessed by the public.

Acknowledgements

References

[1]   Suwandi, Outlook 2017 Agricultural Commodities, Coffee Plantation Sub Sector. Center for Agricultural Data and Information Systems General Secretariat - Ministry of Agriculture 2017, 2017.

[2]   B. Ikhwan, "Enchantment of Kopi Luwak," Export News, Ministry of Trade, pp. 1–20, 2013.

[3]   J. Towaha, A. Aunillah, E. H. Purwanto, and H. Supriadi, "Effects of Elevation and Processing on Chemical Content and Taste of Lampung Robusta Coffee," J. Tanam. Ind. and Refresher, vol. 1, no. 1, p. 57, 2019, doi: 10.21082 / jtidp.v1n1.2014.p57-62.

[4]   Yusianto, "Kopi Luwak and testing its authenticity," Pus. Researcher. Coffee and Cocoa Indones., pp. 1–44, 2014.

[5]   S. C. Madi yuliana, "Deciding Coffee Beans by Using Image Processing [Thesis]," Inst. pertan. Bogor, 2010.

[6]   W. Aji Pulungan, "Identification of Coffee Fruit Maturity Using Artificial Vector Neutral Neural Networks," University of Lampung, 2019.

[7]   M. Daud, A. Bakar, and H. Rusli, "Implementing Rapid Application Development (RAD) methodology in developing practical training application systems," Proc. 2010 Int. Symp. Inf. Technol. - Syst. Dev Appl. Knowl. Soc. ITSim '10, vol. 3, pp. 1664–1667, 2010, doi: 10.1109 / ITSIM.2010.5561634.

[8]   Z. Liu and L. Bai, "Evaluating the supplier cooperative design ability using a novel support vector machine algorithm," Proc. 2008 12th Int. Conf. Comput Support. Coop Work Des. CSCWD, vol. 2, pp. 986–89, 2008, doi: 10.1109 / CSCWD.2008.4537113.

[9]   P. Beynon-Davies and S. Holmes, "Integrating rapid application development and participatory design," IEE Proc. Softw., Vol. 145, no. 4, pp. 105–112, 1998, doi: 10.1049 / ip-sen: 19982196.

[10]   Y. Lin and J. Wang, "Research on text classification based on SVM-KNN," Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci ICSESS, pp. 842–844, 2014, doi: 10.1109 / ICSESS.2014.6933697.

[11] A. Fadlil, Pattern Recognition Lecture Module, no. 002. Yogyakarta: Ahmad Dahlan University, 2012.

[12] J. Wira, Introduction to the Concept of Machine Learning and Deep Learning Edition 1.3, vol. 1.3. Tokyo, Japan: Wiragotama, 2019.

[13] O. Irnawati and et al, "Rapid Application Development (RAD) Method in the Design of Inventory Website of PT. Sarana Abadi Makmur Bersama (S.A.M.B), "vol. 6, pp. 12–18, 2018.

[14] A. Ismail and A. Aziz, "ITourism travel buddy mobile application," Int. Conf. Next Gener. Mob Appl. Serv. Technol., Pp. 82–87, 2016, doi: 10.1109 / NGMAST.2016.22.