



Classification of Roasting Rates in Coffee Beans By Digital Image Processing Using The Naive Bayes Classifier (NBC) Method

M. Mahaputra Hidayat¹, Eko Prasetyo², and Bambang Teguh Wicaksono³

Department of Informatics Engineering, Bhayangkara University of Surabaya^{1,2,3}

ARTICLE INFORMATION

Journal of Science and
Technology – Volume 27
Number 1, May 2023

Page:
23 – 30
Date of issue :
May 30, 2023

DOI:
[10.31284/j.iptek.2023.v27i1.3697](https://doi.org/10.31284/j.iptek.2023.v27i1.3697)

E-MAIL

mahaputra@ubhara.ac.id¹
eko@ubhara.ac.id²
bambangteguh6@gmail.com³

PUBLISHER

LPPM- Adhi Tama Institute of
Technology Surabaya
Address:
Jl. Arief Rachman Hakim No.
100, Surabaya 60117, Tel/Fax:
031-5997244

*Jurnal IPTEK by LPPM-ITATS
is licensed under a Creative
Commons Attribution-
ShareAlike 4.0 International
License.*

ABSTRACT

Coffee is the bean of the coffee plant and is the source of coffee drinks. Coffee beans must pass through the coffee roasting stage or also called coffee roasting, from this stage of the process the coffee will be roasted and this process also has its own level. At this stage of the roasting process coffee shop business people often do not know this process, then in this case used Local Binary Pattern (LBP) method. LBP is a simple and very efficient texture operator by labeling pixels by doing thresholding on each pixel neighbors and considers the result as a binary number. This method is used to obtain the texture ekstrasi of an image. While for the classification method using the Naive Bayes Classifier (NBC) method. Naive Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing the frequencies and combinations of values from a given dataset. The algorithm uses Bayes' theorem and assumes all the independent or non-interdependent attributes given by the values on the class variables. From the test results by comparing training data and testing data obtained an accuracy rate of 81%. For an image-based developed system with display recognition difficulties, this performance is good.

Keywords: *Coffee beans, Coffee Bean roasting, Local Binary Pattern, Naive Bayes Classifier, Classification System.*

ABSTRACT

Kopi adalah biji dari tumbuhan kopi dan merupakan sumber dari minuman kopi. Biji kopi harus melewati tahap roasting kopi atau juga disebut penyangraian kopi, dari tahap proses ini kopi akan disangrai dan proses ini juga memiliki tingkatan tersendiri. Pada tahap proses sangrai ini para pelaku bisnis kedai kopi sering kali belum mengetahui proses ini, maka dalam kasus ini digunakan metode *Local Binary Pattern* (LBP). LBP adalah operator tekstur yang sederhana dan sangat efisien dengan cara memberikan label pada piksel dengan melakukan peng-ambangan (*thresholding*) pada setiap pixel tetangga dan mempertimbangkan hasilnya sebagai bilangan biner. Metode ini digunakan untuk mendapatkan ekstrasi tekstur dari suatu citra. Sedangkan untuk metode pengklasifikasiannya menggunakan metode Naive Bayes Classifier (NBC). Naive Bayes merupakan sebuah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma ini menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Dari hasil pengujian dengan membandingkan data training dan data testing didapatkan tingkat akurasi sebesar 81%. Untuk sebuah sistem pengenalan tampilan berbasis citra yang dikembangkan, kinerja ini bagus.

Keywords: *Biji Kopi, Roasting Biji Kopi, Local Binary Pattern, Naive Bayes Classifier, Sistem Klasifikasi.*

INTRODUCTION

Coffee is a bean from the coffee plant and is the source of coffee drinks. And along with the development of the times, many people like coffee drinks and also this moment is used by some people to start opening a business in the field of coffee drinks such as coffee shops or just selling coffee beans that are ready to drink, and people who do business in the coffee sector are competing to serve good coffee drinks. Coffee must pass the coffee roasting stage or also called coffee roasting from this stage of the process coffee will be roasted and this process also has its own level, at this stage of the roasting process the coffee shop business people often do not know this process and make the coffee served to customers have a taste that is not quite right. Customers need to know what roast of coffee will be consumed, but not everyone can recognize coffee from its appearance. The authors develop a system to recognize coffee types to solve this problem. Display features use the LBP (Local Binary Pattern) method, while for classification use the NBC method (Naive Bayes Classifier).

The Local Binary Pattern method is one of the texture descriptors first introduced in 1994 by Ojala. LBP is a simple and highly efficient texture operator by labeling pixels by thresholding each neighboring pixel and considering the result as a binary number. Then we calculate the features of the LBP texture that is formed. Studies using LBP as feature extraction include Pulpitis Detection Through Periapical Radiographs with an accuracy of up to 80%. [1], detecting wood defects with up to 89% accuracy [2], emotion recognition based on micro-expressions with grooves up to 70% [3], Hiragana pattern recognition with up to 81% accuracy[4], classification of cow's milk with 67% accuracy[5], detect glaucoma with up to 93% accuracy[6], the highest accuracy achieved by the classification of medicinal plants 96% [7]. LBP can also be combined with other methods to achieve better accuracy such as GLCM [8], Completed Robust Local Binary Pattern (CRLBP) [9]. The combination can even reach 90%. The Local Binary Pattern method is suitable for the extraction of texture feature characteristics, the extracted characteristic data will be used as training data to determine the type and level of roasting coffee beans [10]. The Naive Bayes Classifier (NBC) method or Bayesian classification is a statistical classification that can be used to predict the probability of membership of a class. Bayesian classification is based on Bayes' theorem which has a classification ability similar to decision trees and neural networks [11]. Bayesian classification is proven to have high acubatement and speed when applied to databases with large data. NBC classifies by calculating the prior probability data for each feature and then calculating the posterior probability. The class with the highest posterior probability value is used as the predicted class [11].

Coffee classification has also been carried out by previous researchers, such as a coffee smoothing device as well as clacification [12], coffee maturity grouping [13], classification of coffee maturity using the nearest neighbor method [14], and LBP also [15]. Existing research shows that coffee classification performs well. However, research has yet to use LBP as a system feature. Whereas in previous research, classification with LBP features can achieve an accuracy of 89%. So, the authors conducted research by developing a classification of roasted coffee beans in the digital image using the naive Bayes classifier method. Based on this system, customer can recognize the level of roasting in coffee beans, so that it can be used as a medium to help baristas in serving coffee drinks.

LITERATURE REVIEW

Local Binary Pattern Method

Local Binary Pattern (LBP) is a method of texture analysis that uses statistical and structural models. LBP was first introduced by Timo Ojala. The LBP operator uses a comparison of the grayness values of neighboring pixels [8]. The base operator of the LBP is 3 x 3 using 8 pixels of the innermost of an *ic* middle pixel. The nth neighboring pixel is thresholdd using the grayness value of

the middle pixel as shown in equation (1) and the $s(x)$ thresholding function as shown in equation (2). The binary code resulting from the LBP operator of neighboring pixels will be used to represent the features of the middle pixel ic .

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n \quad \dots (1)$$

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad \dots (2)$$

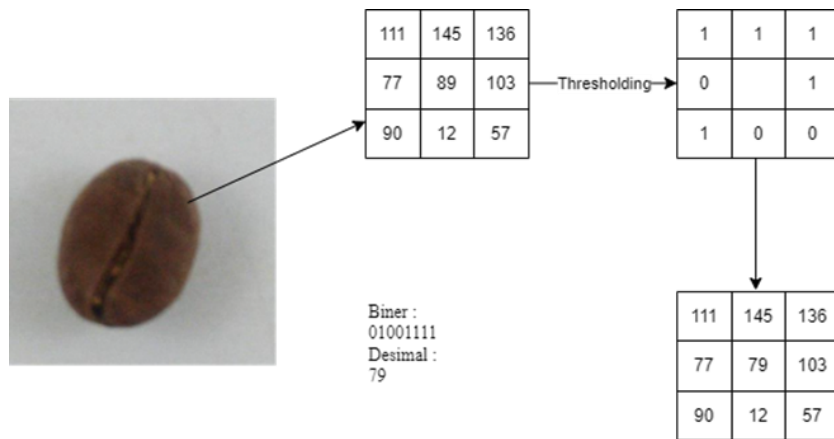


Figure 1. LBP Illustration

Figure 1 is an illustration of the LBP process. The first process is to perform a reduction of neighboring pixels with the middle pixel (1). Furthermore, the result of the reduction is thresholded using equation (2), if the result is ≥ 0 then it is given a value of 1 and if the result is < 0 then it is given a value of 0. After that, the binary values of neighboring pixels will be arranged counterclockwise and the 8 bits of the binary are converted into decimal values to replace the middle pixel value of ic .

Naive Bayes Classifier Method

Naive bayes is a method for binary and multiclass classification. This method, also known as naive bayes classifier, applies the technique of supervised classification of objects in the future by assigning class labels using conditional probabilities [11]. A gaussian distribution is an assumption of continuous value distribution associated with each numerical value feature. The gauss method is:

$$P(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \quad \dots (3)$$

Where μ is the average value and δ is the value of the standard deviation, to calculate the average and standard deviation using the following formula:

$$\mu = \frac{\sum_{i=1}^n xi}{n} \quad \dots (4)$$

$$\delta^2 = \frac{\sum_{i=1}^n (xi - \mu)^2}{n - 1} \quad \dots (5)$$

METHOD

Coffee is a coffee bean producing plant which will later be managed into a coffee drink. Along with the development of the era, people began to like coffee drinks a lot and along with the demand and interest in coffee drinks, many people also opened a coffee shop or café business by serving coffee drinks as one of the attractions of coffee shops. Of the many specializations and

demands for coffee drinks, many people do not know about how to process coffee itself, one of which is in the way of roasting coffee or frying coffee in this case roasting coffee or frying coffee has several levels in the process. Therefore, a system is needed that can help the community in recognizing the level of roasting coffee beans using the Local Binary Pattern and Naive Bayes methods. This system serves to help the community in recognizing the level of roasting in coffee beans.

The data used in the classification process is coffee bean image data from dark robusta, medium robusta, light robusta, dark arabica, medium arabica, and light arabica by identifying the texture of the coffee beans. For the data used, there are 1500 data to be used as training and test data with a composition that will be explained further in the results and discussion section. The design of the system aims to provide an overview of the planning of the system to be built. in this stage will also be given an overview of the flow of information and processes in the system. Here are the steps that will be done in designing the system.

System Flowchart

A system flowchart is a graphical depiction of the steps and sequences of procedures of a program. Flowcharts help analysts and programmers to solve problems into smaller segments and assist in analyzing other alternatives in program operation. The following is an overview of the flowchart system to be created as shown in Figure 2.

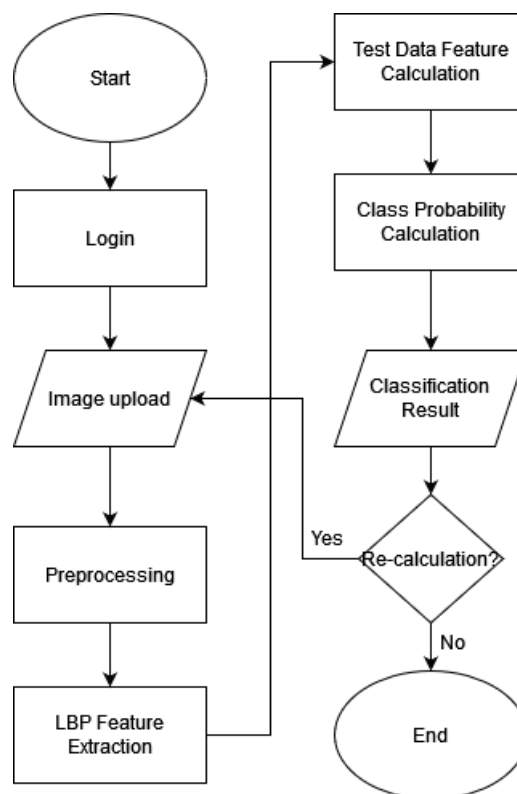


Figure 2. System Flowchart

Preprocessing

Data preprocessing is an initial technique to transform raw data into a more efficient format and to remove unnecessary parts from the input image for further processing. Some of the processes that can be done at the preprocessing stage include resizing, cropping, greyscaling, and segmentation.

LBP Feature Extraction

LBP uses grayscale images with a 3x3 neighborhood, where the pixel value in the center is used as the threshold. The neighboring pixels are then binarized, with the rule that if the neighboring pixel value is smaller than the threshold, it will be assigned a value of 0, while if it is greater than or equal to the threshold, it will be assigned a value of 1. After that, the resulting binary values will be converted to decimal numbers to represent the characteristic pattern structure. The following is the result of the image that has been processed with Local Binary Pattern as shown in Figure 3.

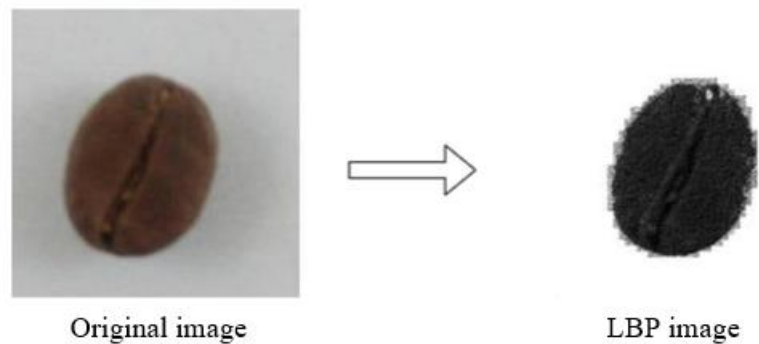


Figure 3. Result of LBP Feature Extraction

Data Feature Calculation

The result of the LBP process will be taken as a histogram value criterion to be used in the classification process. Then the data normalization process will be conducted, which is the process of making variables have the same range of values, with no values that are too large or too small. This will make statistical analysis easier. In this case, the author uses Simple Feature Scaling as the method for data normalization, which will result in a new value after normalization within the range of 0 to 1.

Classification

After data normalization, the class probability will be calculated using Naive Bayes, and then the data will be converted into a table form to speed up the search for solutions. There are several features that will be classified to determine which class the tested image will belong to. There are 6 output classes whose results will be known, namely Robusta Dark, Robusta Medium, Robusta Light, Arabica Dark, Arabica Medium, and Arabica Light. The data used by the author is numerical data. In Naive Bayes Classifier, there are several stages that will be conducted before testing, namely calculating the number of each class, calculating the mean of each class feature, and calculating the standard deviation. To measure how well the naive Bayes method can classify data correctly, accuracy is calculated by comparing the number of correct classifications to the total number of classifications. In the context of image classification, for example, accuracy can be calculated by comparing the number of images classified correctly to the total number of images tested. The higher the accuracy value, the better the classification model is at classifying data.

RESULT AND DISCUSSION

Application testing in this study was carried out 8 times with a different composition of test data and training data. The purpose of the test is to obtain a level of accuracy, where each test with a different composition will produce a different accuracy value. With the hope that the author can find out on which composition produces a high accuracy value and the lowest accuracy value.

Table 1. The composition of test data and training data.

Test	Test Data	Train Data
1	100	60
2	100	100
3	100	200
4	100	300
5	100	500
6	100	700
7	100	900
8	100	1500

Table 1 shown the composition of test data and training data used. There is total 1500 data to be used as training and test data. The testing was carried out 8 times, with each test using 100 test image data. Meanwhile, for each training data, a variable number of image data was used, ranging from 60 to 1500 image data. After the testing process is carried out, the accuracy value will be calculated. Accuracy is a measurement of how well a classification model can correctly classify data. Accuracy is calculated by comparing the number of correct classifications with the total number of classifications. In this research, accuracy can be calculated by comparing the number of test data that are classified correctly with the total number of data tested. The higher the accuracy value, the better the classification model is in classifying data.

Table 2. Test result.

Test	Accuracy (%)	Error (%)
1	56	44
2	70	30
3	70	30
4	75	25
5	77	23
6	81	19
7	80	20
8	72	18
Average	73,87	26,12

From Table 2, after 8 tests using the composition of training data and different test data, the highest accuracy rate was obtained in the 6th test with an accuracy level of 81% using 100 test data and 700 training data. While the lowest accuracy was obtained in the 1st test with an accuracy rate of 56% using 100 test data and 60 training data.

CONSLUSIONS

Conclusion

Based on the results of research that has been carried out regarding the application of the Local Binary Pattern method to the imagery process and the Naive Bayes method as a classification in determining classes, it can be concluded as follows: The results of predicting the roasting rate of coffee beans were obtained, from the tests carried out by comparing training data and test data obtained the highest level of accuracy of 81%; The classification process using the Naive Bayes method utilizes data training to produce probabilities for each different class, so that the probability values of these criteria can be optimized to determine the results of coffee bean classification.

REFERENCES

- [1] L. Listyani Ayuningtyas, D. Bambang Hidayat, D. Suhardjo, and M. SpRKG, "Simulasi Dan Analisis Deteksi Pulpitis Melalui Periapikal Radiograf Menggunakan Metode Local Binary Pattern Dengan Klasifikasi Fuzzy Logic," *eProceedings Eng.*, vol. 2, no. 2, Aug. 2015, Accessed: May 12, 2023. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/647>
- [2] F. N. Achسانی, R. D. Atmaja, and R. Purnamasari, "Deteksi Adanya Cacat Pada Kayu Menggunakan Metode Local Binary Pattern," *eProceedings Eng.*, vol. 2, no. 1, Apr. 2015, Accessed: May 12, 2023. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/1685>
- [3] N. Amynarto, Y. A. Sari, and R. C. Wihandika, "Pengenalan Emosi Berdasarkan Ekspresi Mikro Menggunakan Metode Local Binary Pattern," *JPTIHK*, vol. 2, no. 10, 2018.
- [4] R. Amat, J. Y. Sari, and I. P. Ningrum, "IMPLEMENTASI METODE LOCAL BINARY PATTERNS UNTUK PENGENALAN POLA HURUF HIRAGANA DAN KATAKANA PADA SMARTPHONE," *JUTI J. Ilm. Teknol. Inf.*, vol. 15, no. 2, p. 152, Jul. 2017, doi: 10.12962/J24068535.V15I2.A612.
- [5] D. S. Auladi, B. Hidayat, and S. Darana, "Identifikasi Dan Klasifikasi Kemurnian Susu Sapi Berdasarkan Pemrosesan Sinyal Video Menggunakan Metode Local Binary Pattern (lbp) Dan Learning Vector Quantization (lvq)," *eProceedings Eng.*, vol. 4, no. 2, Aug. 2017, Accessed: May 12, 2023. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/483>
- [6] F. NURFAJAR, R. MAGDALENA, and S. SA'IDAH, "Deteksi Glaukoma pada Citra Fundus Retina menggunakan Metode Local Binary Pattern dan Support Vector Machine," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 10, no. 4, p. 769, Oct. 2022, doi: 10.26760/elkomika.v10i4.769.
- [7] Z. Y. Lamasigi, M. Hasan, and Y. Lasena, "Local Binary Pattern untuk Pengenalan Jenis Daun Tanaman Obat menggunakan K-Nearest Neighbor," *Ilk. J. Ilm.*, vol. 12, no. 3, pp. 208–218, Dec. 2020, doi: 10.33096/ilkom.v12i3.667.208-218.
- [8] J. Arifin, "KLASIFIKASI CITRA TEKSTUR KAYU MENGGUNAKAN GRAY LEVEL CO-OCCURANCE MATRIX DAN LOCAL BINARY PATTERN," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 1, pp. 34–40, Feb. 2022, doi: 10.26798/JIKO.V6I1.557.
- [9] A. Kurniawardhani, N. Suciati, and I. Arieshanti, "KLASIFIKASI CITRA BATIK MENGGUNAKAN METODE EKSTRAKSI CIRI YANG INVARIANT TERHADAP ROTASI," *JUTI J. Ilm. Teknol. Inf.*, vol. 12, no. 2, pp. 48–60, 2014, Accessed: May 12, 2023. [Online]. Available: www.ee.oulu.fi/research/imag/texture/image_data/image_data/Brodatz32.zip.
- [10] Pulung Nurtantio Andono, T Sutojo, and Muljono, *Pengolahan Citra Digital*. Andi Offset, 2017.
- [11] E. Prasetyo, *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab*. Andi Offset, 2012.
- [12] Ansar, Sukmawaty, Murad, S. A. Muttalib, R. H. Putra, and Abdurrahim, "Design and performance test of the coffee bean classifier," *Processes*, vol. 9, no. 8, Aug. 2021, doi: 10.3390/PR9081462.
- [13] M. Rioarda Irfā'i, B. Fatkhurrozi, and I. Setyowati, "Klasifikasi Tingkat Kematangan Buah

Kopi Menggunakan Algoritma Fuzzy C-Means,” *THETA OMEGA J. Electr. Eng.*, 2021.

- [14] S. Raysyah, V. Arinal, and D. I. Mulyana, “KLASIFIKASI TINGKAT KEMATANGAN BUAH KOPI BERDASARKAN DETEKSI WARNA MENGGUNAKAN METODE KNN DAN PCA,” *JSiI (Jurnal Sist. Informasi)*, vol. 8, no. 2, pp. 88–95, Sep. 2021, doi: 10.30656/JSiI.V8I2.3638.
- [15] B. Cahyono, B. E. Cahyono, A. T. Nugroho, and I. W. Maulinida, “Klasifikasi Jenis Biji Kopi dengan Menggunakan Metode Gray Level Co-occurrence Matrix (GLCM),” *TEKNOTAN*, vol. 16, no. 3, pp. 191–196, Mar. 2023, doi: 10.24198/jt.vol16n3.9.