IMAGE CLASSIFICATION OF LOCAL ROBUSTA AND ARABICA COFFEE SEEDS IN MALANG REGENCY USING GRAY LEVEL CO-OCCURRENCE MATRIX AND K-NEAREST METHODS

Devita Widiawati*[1], Muhammad Rijalun Shodaqu[2], Gilang Priambodo[3]
Dian Nuswatoro University Kediri, Jl. Penanggungan No. 41a, Bandar Lor, Kec . Mojoroto, City of Kediri, East Java 64129, (0354) 2895777).
E-mail : 611202100019@mhs.dinus.ac.id , 611202100023@mhs.dinus.ac.id ,
611202100027@mhs.dinus.ac.id
*Corresponding author

Maulana Fajar Anas[4], Titien Suhartini Sukamto S.Kom, M.Eng.[5], Aris Nurhindarto, M.Kom[6]
Dian Nuswatoro University Kediri, Jl. Penanggungan No. 41a, Bandar Lor, Kec . Mojoroto, City of Kediri, East Java 64129, (0354) 2895777).
E-mail : 611202100044@mhs.dinus.ac.id, titien.suhartini@dsn.dinus.ac.id,
arisnurhidarto@dsn.dinus.ac.id

Abstract – *Coffee* is one type results current plantation this favored by some among. Indonesia is in the order to four Becomes Robusta coffee export and producer in the world. Appearance communities coffee lovers make coffee as provider field profession for part big resident. In Indonesia, especially in the Regency of Trunk, a lot very Public around who has coffee plantations including namely Robusta coffee and Arabica coffee *(coffea arabica)* local. For some new people Do you know and love coffee yet? can differentiate type of coffee visually. In the era of increasingly digitalization, advanced like this. There are several method for differentiate something object among them that is processing digital image. Frequent problems occur that is many less consumers in determine Robusta and Arabica coffee types. From trouble that, then researcher designing a system classification on robusta and Arabica coffee beans could obtained with implementation algorithm *K-Nearest Lightweight Classification* (K-NN). [1] combined with extraction feature *Gray Level Co-Occurrence Matrix* (GLCM). Digital image dataset used that is a total of 194 pictures where inside it there is type image coffee beans. Image dataset Robusta and Arabica coffee beans each local number of 97 images. Image dataset shared into 20 test data and 174 training data. Testing conducted using *Matlab* software produce score accuracy highest at distance pixels=1 and the value of K=1 with respect to angle of 45° by 95%.

Keywords – Coffee, K-NN, GLCM, Matlab

1. INTRODUCTION

Coffee is one product from current plantation this needed by people around the world. In Indonesia, coffee plantations have very important role for economy national, then Becomes

facilitator field work for Public around [1] In Indonesia, especially in Malang Regency, people around many have coffee plantations including namely Robusta coffee and Arabica coffee local. Robusta coffee is type of coffee produced more many compared with arabica coffee. [2] Arabica coffee alone more many Chosen by coffee lovers because it feels like it's said delicious. Characteristic typical of coffee is in the size as well as form little seeds small compared to robusta coffee. Caffeine content low, delicious aroma as well as the price is high. For the new layman know coffee yet can differentiate type of coffee visually. In the era of increasingly digitalization, advanced [3] there is Miscellaneous method for differentiate something object as example that is processing digital image.

K-NN is one of the algorithm where could used in the classification process. In the classification process loaded a number of method that can implemented one that is with training data, which has similarity with object that. [4] Destination coffee classification is for share coffee image inside class that is robusta and arabica based on colors and textures contained in the image coffee beans for easy recognized in accordance with contents. [5] Classification image coffee beans with use K-NN algorithm will pass two stages namely training and testing. The test phase uses a dataset in the form of the amount of training data and the value of k that is not same.

On research this will discussed about how method designing a system purposeful testing for differentiate Robusta and Arabica coffee beans local in Malang Regency. Then will conducted testing from application K-Nearest Neighbor algorithm and extraction GLCM features in classification image that. [6] Proposed research that is involve each algorithm that is *K-Nearest Neighbor (*K-NN) and extraction feature *Gray Level Security Matrix* (GLCM). [7] Combination from both of them could classify digital image from RGB to form Grayscale images that produce good data results. Result of study on that is score only accuracy use K-NN algorithm will more low compared to with classification that uses K-NN algorithm combined with extraction GLCM features that get score more accuracy. [5]

2. RESEARCH METHOD

K-Nearest Neighbor (K-NN)

K-Nearest Neighbor is method grouping object based on the closest training data the distance. The K-Nearest Neighbor (K-NN) algorithm is one of them the easiest method understandable, easy tested and without estimate KNN, and a little training. [3] Method this is method simple that can be keep all available data and classify new data in accordance size similarities. [8] Most used for classify data point about how neighbor classified. K-Nearest Neighbor works best on data sets that are not big and have many features. [9]

Goal algorithm this for classification object latest in accordance with characteristics and sample data. Give it test point, will found number of K objects (point train) closest with point that. KNN has advantages, he toughness on training data and succeed when the training data no small. [10] Algorithm this work in accordance with closest distance or even the most distant with neighbors. Euclidean distance is the distance that has function as formula count distance Among test image with picture train in accordance with smallest value from score closest neighbour, can concluded as following :

$$d = \sqrt{(x1 - y1)^2 + (xn - yn)^2} \quad \text{(a)}$$

Description :
d   = Euclidean Distance

(x) = Training Data (Training Data)
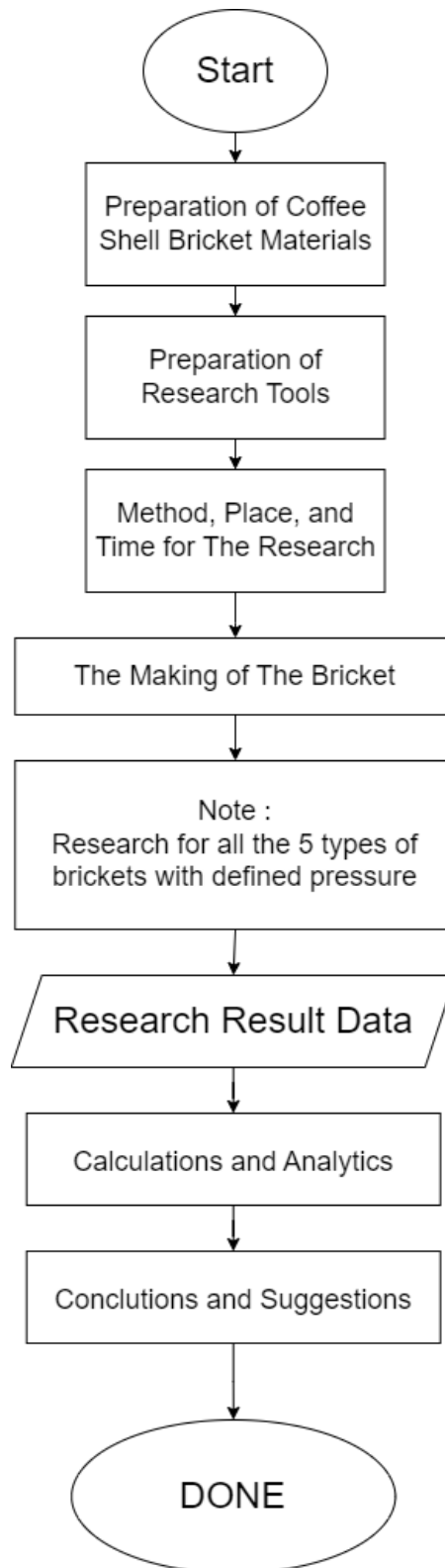(y) = Test Data
(n) = Number image
(a) = 1st Record to n

Figure 1. flow diagram process

2.1 Gray Level Co-Occurrence Matrix (GLCM)

GLCM [11] is one method used as studies texture or extraction characteristic where the calculation use order second that is count couple two pixels picture original, while on the order first the calculation use statistics based on score pixels picture original as well as without notice pixels adversity. [5] are 4 features used in extraction GLCM features, namely as following :

a. Energy

Show score high resemblance originated from high energy value.

$$\sum_{x,y} p(x-y)^2$$

Description :
x   = matrix row value
y   = value column matrix
(,) = Element value matrix row (x) and column (y) co-occurrence

b. Contrast

Contrast is calculation difference intensity Among pixels one and adjacent pixels throughout picture.

$$\sum_{x,y} (x-y)^2 p(x,y)$$

Description :
x   = matrix row value
y   = value column matrix
(,) = Element value matrix row (x) and column (y) co-occurrence

c. Correlation

Show size linear dependence of image grayscale degrees where will give instruction existence inner linear structure digital image.

$$\sum_{x,y} \frac{(x-\mu x)(y-\mu y)p(x,y)}{\sigma x \sigma y}$$

Description :
x   = matrix row value
y   = value column matrix
$p\,(\ ,y\,)$ = Element value matrix row (x) and column (y) co-occurrence
$\mu\,x, \mu\,y$ = the average value of the elements in the row and column
(,)   = value standard deviation in rows and columns

d. Homogeneity

Show number of gray levels in image if pixels same then homogeneity becomes high.

$$\sum_{x,y} \frac{p(x,y)}{1 + |x - y|}$$

Description :
x   = matrix row value
y   = value column matrix
(,) = Element value matrix row (x) and column (y) co-occurrence

3. RESULTS AND DISCUSSION

Processing image containing about algorithm that takes image as input and return image as the output. Processing image Becomes tool help done in many branch knowledge knowledge like knowledge computer, engineering electricity and electronics, robotics, physics, chemistry, science environment, biology and psychology. [6] Generally processing digital images have destination repair quality from digital image.

Image that is similarity from something object. Image is split Becomes two that is analog images and digital images. Analog image is continuous image like picture on television. Whereas digital image is image that can computer processed. Digital images are formed by several element limited, each element the have score as well as coordinate certain.

Matrix Laboratory is a system in which there is information in the form of an array so that user easy complete problem perspective. [12] With Thing the will allows the user to break problem technical based on calculation, as for example that is related to matrix and formula vector, where problem that Becomes dangerous if the user is required complete the problem with use language low level programming.

Grayscale image that is picture two dementia whose components showing quality gray image (pixels). Grayscale has a number of type which gray is different with picture black white. [13] Image modified Becomes picture gray will different with picture color. Inside it RGB image loaded three matrix layers, called the R-layer, G-layer, and B-layer. For get image gray, change obtained with take the average value of R, G and B, so that produce S value as image gray. [14]

GLCM is one of the method used as studies texture or extraction characteristic where the calculation use order second that is count couple two pixels picture original, while on the order first the calculation use statistics based on score pixels picture original as well as without notice pixels adversity. [5]

Testing in research data taken direct, from results testing with use GLCM as extraction feature with angles 0°, 45°, 90°, 135° and use KNN method for classify image based on class with the value of K used i.e. K=1, K=3, K=5, K=9. [15] Research results will taken in accordance yielding predictions score accuracy tall with training data experiment. [2] From research this our will could determine how many percent amount resulting accuracy of the created programs.

a. Confusion Matrix against 0° angle

Table 1. Confusion Matrix Against 0° angle

Actual Results	Prediction Results 0° angle	
	Arabica	Robusta
Arabica	9	1
Robusta	1	9

$$Confusion\ Matrix = \frac{jumlah\ citra\ uji\ benar}{jumlah\ citra\ latih}\ x\ 100\%$$

$$= \frac{9+9}{20}\ x\ 100\%$$

$$= 90\ \%$$

b. Confusion Matrix against 45° angle

Table 2. Confusion Matrix Against 45° angle

Actual Results	Prediction Results corner 45	
	Arabica	Robusta
Arabica	10	0
Robusta	1	9

$$Confusion\ Matrix = \frac{jumlah\ citra\ uji\ benar}{jumlah\ citra\ latih}\ x\ 100\%$$

$$= \frac{10+9}{20}\ x\ 100\%$$

$$= 95\ \%$$

c. Confusion Matrix against 90° angle

Table 3. Confusion Matrix Against 90° angle

Actual Results	Prediction Results angle 90°	
	Arabica	Robusta
Arabica	9	1
Robusta	1	9

$$Confusion\ Matrix = \frac{jumlah\ citra\ uji\ benar}{jumlah\ citra\ latih}\ x\ 100\%$$

$$= \frac{9+9}{20}\ x\ 100\%$$

$$= 90\ \%$$

d. Confusion Matrix against angle 135°

Table 4. Confusion Matrix Against 135° angle

Actual Results	Prediction Results	
	Arabica	Robusta
Arabica	9	1
Robusta	1	9

$$Confusion\ Matrix = \frac{jumlah\ citra\ uji\ benar}{jumlah\ citra\ latih}\ x\ 100\%$$

$$= \frac{9+9}{20}\ x\ 100\%$$

$$= 90\ \%$$

4. CONCLUSION

After conducted study with use extraction features and methods that have been proposed, withdrawn conclusion in the form of :

1. From the research that has been conducted testing with 174 training data and 20 test data which have two classes in it that is image seed robusta and arabica.
2. Processed extraction feature with GLCM against angles 0°, 45°, 90°, 135° as well method classification namely KNN against values K=1, K=3, K=5, K=7, K=9. Testing to 0° angle yields 18 correct data where 9 is correct for class arabica and 9 true for class robusta from the number of test data is 20 image data, 19 data is correct where 10 is classified Correct for class arabica and 9 true class robusta from the amount of test data is 20 images to 45° angle, 18 classified data Correct where 9 is correct for class arabica and 9 true for class robusta from the number of test data is 20 images at an angle of 90°, 18 classified data Correct where 9 is correct for class arabica and 9 true for class robusta from the amount of test data is 20 images to 90° angle.
3. In the test image test data shows score high accuracy with using 20 test image data and 174 image data train so that yield 95% value available accuracy in angle 45° with distance pixels used is 2, data is correct a total of 19 images coffee beans with a total of 20 test data.

Suggestion :

Based on the conclusions that have been spelled out, the advice you want be delivered that is in the form of :

1. Expected next study this could developed in mobile form (android).
2. For study next expected could use other objects like classify fresh and rotten bananas.
3. Application extraction GLCM features expected could combined with another method for produce score more accuracy high.

References

[1] J. Zeniarja, A. Ukhifahdhina, and A. Salam, "Diagnosis Of Heart Disease Using K-Nearest Neighbor Method Based On Forward Selection," 2019.

[2] V. Sari, F. Firdausi, and Y. Azhar, "Comparison Prediction Arabica Coffee Quality with Use SGD Algorithm, Random Forest and Naive Bayes," *Edumatic : Journal of Informatics Education*, vol. 4, no. 2, pp. 1–9, Dec. 2020, doi : 10.29408/edumatic.v4i2.2202.

[3] A. Rachmad Syulistyo and V. Meliana Agustin, "Predicting News Article Popularity with Multi Layer Perceptron Algorithm," 2022. [Online]. Available: https://www.kaggle.com/waseemakramkhan/the-tribune-news-articles.

[4] E. Hari Rachmawanto, C. Atika Sari, H. Pramono, and W. Shinta Sari, "Visitor Prediction Decision Support System at Dieng Tourism Objects Using the K-Nearest Neighbor Method," 2022.

[5] YN Yaspin, DW Widodo, and J. Sulaksono, " Classification Clove Flower Quality for Increase Quality With Utilization Gray Level Co-Occurence Matrix (GLCM) characteristics."

[6] "4".

[7] D. Care Khrisne and D. Putra, "Automatic Image Annotation Using Block Truncation Method and K-Nearest Neighbor," vol. 4, no. 1, 2013.

[8] "Implementation K-Nearest Neighbor Algorithm As Classification Decision Support Receiver PPA and BBM Scholarships Sumarlin STIKOM Uyelindo Kupang.

[9] C. Atika Sari and E. Hari Rachmawanto, "Sentiment Analyst on Twitter Using the K-Nearest Neighbors (KNN) Algorithm Against Covid-19 Vaccination," 2022.

[10] D. Ikhsan, E. Utami, and FW Wibowo, "METHOD OF CLASSIFICATION OF QUALITY GREENBEAN ARABIKA LANANG COFFEE AND USUALLY USING K-NEAREST NEIGHBOR BASED ON SHAPE," *Journal SINUS SCIENCE*, vol. 18, no. 2, p. 1, Jul. 2020, doi : 10.30646/sinus.v18i2.456.

[11] D. Aditya Nugraha and A. Sartika Wiguna, "COLOR FEATURE SELECTION OF COFFEE BEAN DIGITAL IMAGE USING PRINCIPAL COMPONENT ANALYSIS METHOD Digital Image Selection of Coffee Seed Using Component Analysis Method," 2020.

[12] HP Hadi and H. Rachmawanto, "Extraction Feature Analysis Features and Colors In Classification Process Maturity K-Nearest Neighbor Based Rambutan Fruit," *SKANIKA: System Computer and Informatics Engineering*, vol. 5, no. 2, pp. 177–189, 2022.

[13] HP Hadi and EH Rachmawanto, "JIP (Journal Informatics Polynema) COLOR FEATURE EXTRACTION AND GLCM IN KNN ALGORITHM FOR CLASSIFICATION OF HAIR Maturity".

[14] "garuda1429115".

[15] Y. Prastyaningsih, W. Kusrini, P. Tanah Laut, JA Yani KM, D. Stage KecPelaihari Land District Laut, and South K.," System Image Retrieval at the Coffee Bean Roasting Level Using Color Feature Extraction," vol. 6, no. 2, p. 2021.