Journal of Theoretical and Applied Information Technology

15th November 2023. Vol.101. No 21
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CLASSIFICATION OF FUNGAL SPECIES USING K-NN BASED ON COLOR FEATURE EXTRACTION AND GLCM

WILIS KASWIDJANTI[1] , BAMBANG YUWONO[2], INDAH WIDOWATI[3], DILA AJENG MEILIAWATI[4] , MANGARAS YANU FLORESTIYANTO[5]

[1,2,4,5]Informatics Engineering Department, UPN “Veteran” Yogyakarta, Indonesia

[3]Agribusiness Department, UPN “Veteran” Yogyakarta, Indonesia

E-mail: [1]wilisk@upnyk.ac.id, [2]bambangy@upnyk.ac.id, [3]indah.widowati@upnyk.ac.id, [4]dilaajengm@upnyk.ac.id, [5]mangaras.yanu@upnyk.ac.id

ABSTRACT

In digital image processing, feature extraction is an important task to obtain crucial information about the characteristics of the image. One of the feature extractions that can be analyzed is texture feature extraction. Grey-level Co-occurrence Matrix (GLCM) is a texture feature extraction method that uses statistical approaches and has been proven to be the strongest descriptor for data classification. Many parameters in GLCM can be used as texture feature extraction values, but some parameters are often used in research, namely ASM, Contrast, IDM, and Correlation. This study will combine several commonly used GLCM parameters with Energy and Dissimilarity parameters. The combination of these GLCM parameters will be implemented to classify the types of portobello and shiitake mushrooms K-Nearest Neighbor (K-NN) algorithm as a classification method. Based on several models that have been constructed, the best performance is achieved when the model is built using 6 parameters, which results in an accuracy of 97%. This accuracy was obtained by testing the system using a confusion matrix with several experiments based on the constructed model and the predetermined K value.

Keywords: *Grey level Co-occurrence Matrix, K-Nearest Neighbor, Fungal, Confusion Matrix*

1. INTRODUCTION

Computer vision technology is used in the field of digital image processing to teach students about digital image processing techniques [1]. The method of feature extraction in image processing is crucial for obtaining crucial data about the attributes of the image. Texture feature extraction is one of the feature extractions that can be examined [2]. One technique for statistical texture analysis is the grey-level co-occurrence matrix (GLCM), in which the texture features are taken from the co-occurrence matrix using various statistical methods. Only a few parameters are suggested by Heralick to obtain GLCM features, but there are parameter values in GLCM texture feature extraction.

Numerous studies use various parameters that are frequently applied with varied degrees of precision. These variables include ASM, Contrast, IDM, and Correlation, among others. However, certain studies that take into account additional factors, such as energy and dissimilarity, result in higher accuracy. A study by [3] that attempted to use energy, contrast, entropy, and homogeneity (IDM) as GLCM parameters to detect melasma in facial images achieved an accuracy of 98%. A study by [4] that combined the extraction of dissimilarity, correlation, and contrast features to detect potholes and asphalt roads has demonstrated that dissimilarity has good performance with an accuracy of 91.707%.

Several GLCM parameters will result in feature extraction values that can be used for classification. According to earlier research, the KNN algorithm performs rather well in classification. For instance, the study by [5] titled Identification of Orchid Types demonstrates that the KNN algorithm has a respectable success rate with an accuracy of 88.75% in the classification of meat types it was used in. With the best-case scenario with k = 1 and k = 5, the GLCM Method and the KNN Algorithm produce 80% accuracy. Additionally, [6] found that using the KNN method throughout the clustering phase to predict student graduation produced an accuracy of 85.15%.

The two GLCM parameters of energy and dissimilarity were added to the previously frequently used parameters of ASM, contrast, IDM, and correlation in this study. This study will use the K-

Nearest Neighbor (KNN) algorithm in addition to the addition of two additional factors to classify different varieties of portabello and shiitake mushrooms based on their texture and color. The goal of the parameter and algorithm combination is to determine the GLCM's best accuracy. The purpose of this research is to find the best results from several combinations of methods that have been previously planned and then tested. In this research, there is a calculation of the extraction of texture feature characteristics including ASM, Contrast, IDM, Correlation, Energy, and Dissimilarity, which have been planned for combination. The classification is built using several GLCM Models that are combined and designed with the calculation of each feature extraction that is averaged and implemented for classification. The combination models designed include; Model 1 is a combination built with the calculation of 4 feature extractions namely ASM, contrast, IDM, and correlation, Model 2 is a combination built with the calculation of 5 feature extractions namely ASM, contrast, IDM, correlation, and energy, Model 3 is a combination built with the calculation of 5 feature extractions ASM, contrast, IDM, correlation, and dissimilarity, the last is Model 4 built using 6 feature extractions namely ASM, contrast, IDM, correlation, energy, and dissimilarity. Not included in this study are image data used not with various conditions only using the same conditions, namely the same distance and lighting conditions (which have been provided by researchers), not using image data other than portabello and shitake mushroom image data, not including other classification results other than portabello and shitake mushroom types, comparison methods are 4 GLCM models that have been designed, not using other methods, not using labels other than portabello mushroom and shitake mushroom labels.

Texture feature extraction is one of the widely used methods, from some literature there are several texture feature extractions that have good results, one of which is GLCM, in GLCM texture feature extraction there are many quantities but only a few are often used. Literature screening is done from the beginning of the field selection, namely intelligent systems that are specific to digital images, and then the selection of classification topics. In the field of classification there are methods and algorithms that can be used, from many methods the author chose to use the GLCM method and the KNN algorithm with the focus of comparison by building several methods of extracting texture features from GLCM. Previous research only focused on classification performance using one method without any comparison. performance using a method without any comparison and did not include a clear clear reasons for choosing an aspect for the classification. In In this research, the author clearly writes that the 4 comparison models as aspects to be implemented into the classification of objects that will produce the best model value, where the best model is the one that is used for classification. produce the best model value, where the best model can be used to classify an image object. to classify an image object.

2. METHODS

The study approach, including the steps taken to create the system implementation of the K-Nearest Neighbor Algorithm and the Gray Level Co-Occurance Matrix Method in Mushroom Type Classification, is described in this section. Figure 1 illustrates the phases of the investigation.

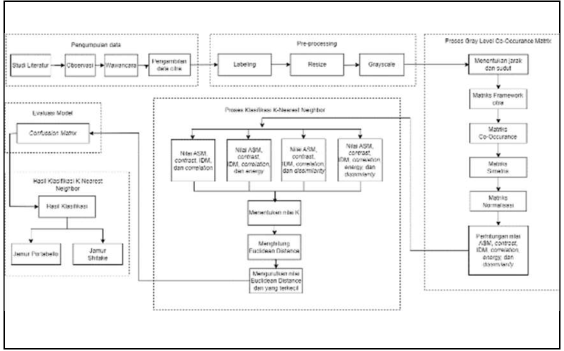
Pengumpulan data
Studi Literatur
Observasi
Wawancara
Pengambilan data citra
Pre-processing
Labeling
Resize
Grayscale
Proses Gray Level Co-Occurance Matrix
Menentukan jarak dan sudut
Matriks Framework citra
Matriks Co-Occurance
Matriks Simetrik
Matriks Normalisasi
Perhitungan nilai ASM, contrast, IDM, correlation, energy, dan dissimilarity
Proses Klasifikasi K-Nearest Neighbor
Nilai ASM, contrast, IDM, dan correlation
Nilai ASM, contrast, IDM, correlation, dan energy
Nilai ASM, contrast, IDM, correlation, dan dissimilarity
Nilai ASM, contrast, IDM, correlation, energy, dan dissimilarity
Menentukan nilai K
Menghitung Euclidean Distance
Mengurutkan nilai Euclidean Distance dari yang terkecil
Evaluasi Model
Confusion Matrix
Hasil Klasifikasi K Nearest Neighbor
Hasil Klasifikasi
Jamur Portabello
Jamur Shitake

Figure 1: Research Flow

2.1 Data Collection

Interviews, observations, and a sampling of picture data were the three stages of the data-collecting approach in this study. Literature reviews, research-related papers, journals, books, and articles that explore the GLCM method and object categorization using the KNN algorithm were the subject of literature reviews. Observation, the author conducted observation by visiting PT Volva Indonesia, a site for mushroom cultivation, on July 30, 2022, to observe the different types of mushrooms in person and learn about the variations in mushroom cultivation methods that contribute to the significance of identifying each type of mushroom. Interview, on July 30, 2022, an interview was held at PT Volva Indonesia to gather information regarding the study's intended use of the data. A direct interview with Mr. Sardjito, a member of the PT Volva Indonesia staff who is in charge of overseeing and managing the mushroom growth in that location, was used to gather information. Sampling image data, data was collected by photographing mushrooms with an iPhone 11 Pro Max smartphone camera at a resolution of 3024 ×

4032 pixels and ring light illumination. 248 samples of mushrooms, including 124 pictures of portabello mushrooms and 124 pictures of shitake mushrooms, were photographed.

2.2 Pre-Processing

The pre-processing stage involves generalizing the data to make the subsequent processing of the data simpler. Labeling, scaling, and grayscale are some of the preprocessing phases performed in this study.

The labeling procedure involves assigning names to each predetermined class to all image data. The Shitake mushroom type has label 1, whereas portabello mushroom type has label 2.

When resizing an image, the original image size is replaced with a predefined image size. Resize is utilized in this study to resize the image to a 161x181 size to balance the image size for processing. The conversion of RGB values to grayscale or greyscale values is the final step in pre-processing. To make three-dimensional images easier to analyze and use as input images for classification, grayscale images must be used to reduce them to one dimension with the same intensity value [7]. It is necessary to complete the RGB calculation process before beginning the greyscale computation.

The process of determining the Red, Green, and Blue values of each pixel, which will subsequently be employed in the conversion to greyscale, is used to extract the RGB color properties. Then, determine the average (mean) of all picture data, including training and test data that help identify objects in digital images. One of the parameters for establishing the classification of fungus uses this average calculation. The following formula can be used to calculate the mean:

$$\mu = \frac{1}{N}\sum_{t1}^{N} Ai \tag{1}$$

There are various methods for converting data. The calculation formula employed in this study to convert RGB to grayscale is as follows. Greyscale is equal to 0.299R, 0.587G, and 0.114B.

2.3 Gray-Level Co-occurrence Matrix-Processing

Calculate the likelihood of an adjacent association between two pixels at a specific distance and angular orientation to provide second-order statistical characteristics. The relationship between two pixels with distance d and direction defines each component of the GLCM. The purpose of the orientation angle function is to ascertain each pixel's direction of the adjacent relationship in the image [8].

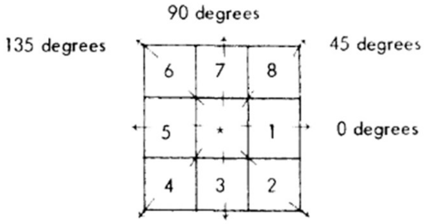

Figure 2: Direction for GLCM with angles of 0°, 45°, 90°, and 135° [9]

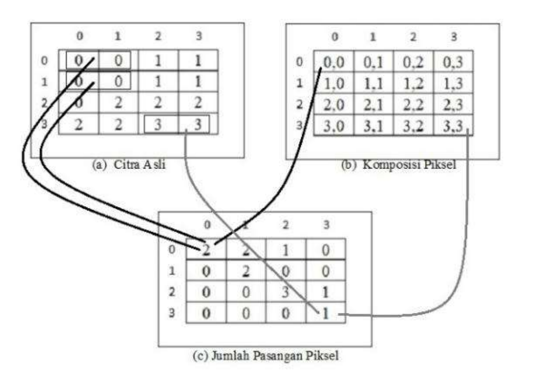

Figure 3: Example of initial determination of GLCM matrix based on two-pixel pairs [10]

Only a few of the criteria suggested by Haralick are utilized to acquire GLCM features. In this study, the angular second moment (ASM), contrast, inverse different moment (IDM), and correlation are the only GLCM parameters used.

1. Angular Second Moment (ASM)

ASM is used to determine the value of image uniformity, which is a measure of homogeneity. The following is the ASM calculation:

$$ASM = \sum_{i=1}^{L} \sum_{j=1}^{L}(GLCM(i,j))^{2} \tag{1}$$

2. Contrast

An attribute to gauge the intensity differences in an image is contrast. If the intensity fluctuation in the image is large, the contrast value will be higher; conversely, if the intensity variation is low, the contrast value will be lower. The calculation of contrast is as follows:

$$Contrast = \sum_{i} k^{2} = [\sum_{i} \sum_{j} GLCM\ (i,j)] \tag{2}$$

3. IDM, or Inverse Different Moment

When the GLCM levels are uniform, IDM is a local homogeneity that is high, and it is low otherwise. homogeneous Images will have a high IDM. The IDM computation is as follows:

$$IDM = \sum_{i1}^{L}\sum_{j=1}^{L}\frac{GLCM\ (i,j)}{1+(i,j)^{2}} \tag{3}$$

4. Correlation

Correlation is a measure of linear dependence between gray-level values that shows the linear structure in the image. The following is the correlation calculation:

$$Correlation = \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{(i-\mu i')(j-\mu j')(GLCM(i,j))}{(\sigma i' \sigma j')} \quad (4)$$

5. Energy

Energy is a feature used to gauge how many intensity pairs are present in the GLCM matrix. If the pixel pairings that have satisfied the co-occurrence matrix requirements are concentrated on a few coordinates, the energy value will be higher; conversely, if they are dispersed, the energy value will be lower. The calculation of energy is as follows:

$$Energy \sqrt{\sum_{i=1}^{L} \sum_{j=1}^{L} (GLCM(i,j))^{2}} \quad (5)$$

6. Dissimilarity

Distance between pairs of objects (pixels) in the region of interest is measured by dissimilarity. The dissimilarity calculation is as follows:

$$Dissimilarity = \sum_{i} \sum_{j} |i-j| \, p(i,j) \quad (6)$$

K- Nearest Neighbor is a supervised learning technique in which the majority of the nearest neighbor category is used to classify the new instance results [5]. K-NN is a classification technique that works well with huge amounts of data and is robust to noisy training data [11]. Since the KNN algorithm is non-parametric and the simplest of the numerous machine learning algorithms investigated [12], it can be used for any data distribution.

Using previously categorized and stored training data, the K-NN approach is designed to categorize a fresh batch of objects. This algorithm is used to identify k groups of unclassified items in the testing data and then compare them to the training data that share the most similarities. K-NN analysis is used to determine how similarity measurements and the value of k utilized affect the system's ability to accurately categorize digital images. If the likelihood of similarity is equal, using an odd value for k is meant to reduce algorithm errors. Let's take the scenario where a meat image needs to have its kind determined based on information about other meat images that are already known.

The closest case of the new meat image to all the previous meat image data is calculated to determine which meat image to employ. The method for categorizing the new meat image will be based on the old meat image that is closest to the new meat image.

Figure 4 shows a KNN classification example with a value of k = 3. There are two different kinds of classes: class O and class X, as well as q1 and q2 as undetermined class symbol nodes.

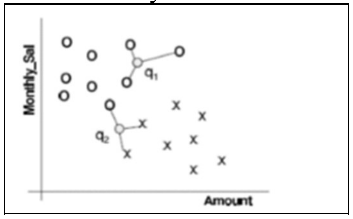

Figure 4: Illustration of Solutions on K-NN [13]

Node q1 is categorized as class O since it receives its three nearest neighbors, all of whom are in class O. Node q2 can be categorized into class X since it has two nearest neighbors from class X and one neighbor from class O.

There are several methods, including Euclidean distance and Manhattan distance (city block distance), which are frequently used, to calculate the closeness distance between new data and old data (training data). The closeness distance between x and y is represented by the symbol d (x, y), where x is the training data and y is the test data. By adding together as many as n attribute values from the x and y data, the x and y data are calculated.

At this point, the research findings on the method of obtaining architecture and the incorporation of portabello and shitake mushroom classification in the earlier-created design will be explained. Using a confusion matrix, this study's testing phase. Some of the outcomes of the study's average red, green, and blue values are shown below.

Table 1: RGB feature extraction results.

No	Label	Nilai *Red*	Nilai *Green*	Nilai *Blue*
1.	1	17.775897470791424	13.457078702054508	7.319166990205772
2.	1	14.38779463738938	9.998999363595248	5.551322641199803
3.	2	9.346040081491829	6.757757934046054	3.362462526166642
4.	2	12.272145084273598	9.348297517220953	4.9992275086955305

By adding together the total red, green, and blue values and dividing them by the number of pixels, this average value will be utilized as a criterion to establish the classification of different species of mushrooms. An illustration of the typical value of red, green, and blue in mushroom imagery is shown in Table 1.

The value of the outcomes of the GLCM calculation, as seen in Tables 2 and 3, is then shown in the following table.

Table 2: Example of Feature Extraction Value GLCM Label 1.

	ASM	Contrast	IDM	Correlation	Energy	Dissimilarity
0	0.59962091215545563	117.68680939226525	0.79251484785797787	0.9660323186385855	0.7743519304266351	3.4018646408839768
45	0.59499833803530051	181.417500000000042	0.78655685099699901	0.9478596074914373	0.7713613537346197	4.391041666666674
90	0.59884648738021161	137.03126293995814	0.79122936699508803	0.9604280271988345	0.7738517218306206	3.685990338164239
135	0.59521292438272213	172.40388888888927	0.78816675317373335	0.9504501691586017	0.7715004370593197	4.210069444444454

Table 3: Example of Feature Extraction Value GLCM Label 2.

	ASM	**Contrast**	**IDM**	**Correlation**	**Energy**	**Dissimilarity**
0	0.78405162042424448	38.95055248618785	0.89900962736127533	0.94550521091476 48	0.88546689403062265	1.32196132596685 2
45	0.78120472668306684	61.968055555555665	0.89441833883448155	0.91372845056566678	0.88385786565661158	1.76916666666666708
90	0.78364129184544452	49.9028295376120 2	0.89721323317127 97	0.93013920723582258	0.88523516188945525	1.5175293305728055
135	0.78089808063272204	58.993680555555569	0.89458531804888864	0.91786935732074417	0.88368437840256609	1.6986805555555586

The input matrix for the GLCM calculation procedure must be created from the image once it has successfully undergone pre-processing. The estimate in this study is based on four angles—0°, 45°, 90°, and 135°—and an adjacent distance of 1 pixel. The normalizing matrix is the matrix that will be employed in the feature extraction utilizing GLCM calculations. Finding the co-occurrence matrix yields this normalization matrix, which is then added to its transpose matrix to produce what is known as a symmetrical matrix. Each pixel is divided by the total number of pixels after the symmetrical matrix has been obtained, creating a normalizing matrix. The next step is to choose the parameter values for the GLCM feature extraction. Six GLCM parameters—ASM, contrast, IDM, correlation, energy, and dissimilarity—are combined in this study.

The next step is to convert each parameter's average RGB value and GLCM texture feature extraction value into parameters that may be used to determine classification using K-Nearest Neighbor. In this study, we divided the data into training and test groups with a 70:30 ratio. From several additional split data % experiments, the percentage comparison of training and test data produced the greatest results.

The accuracy of the KNN approach employing the best model and 6 parameters with various specified K values is demonstrated in the example below. The confusion matrix method was used in this work as a system test. The 150 pieces of information utilized for testing are divided into 75 pieces for the portobello class and 75 pieces for the shitake class. The table below shows the test implementation for the best model, which uses 6 parameters and K = 5.

Table 4: Confussion Matrix Testing

Aktual	Prediksi	
	Portabello	**Shitake**
Portabello	36 (TP)	1 (FN)
Shitake	1 (FP)	37 (TN)

$$Accuracy = \frac{\text{TP+TN}}{\text{TP + TN + FP + FN}} \text{ x } 100 = \frac{36+37}{37 + 37 + 1 + 0} \text{ x } 100 = 97{,}3\%$$

$$Precision \text{ Portabello} = \frac{\text{TP}}{\text{TP+FP}} \text{ x } 100 = \frac{36}{36+1} \text{ x } 100 = 97{,}29\%$$

$$Precision \text{ Shitake} = \frac{\text{TP}}{\text{TP+FP}} \text{ x } 100 = \frac{37}{37+1} \text{ x } 100 = 97{,}36\%$$

$$Recall \text{ Portabello} = \frac{\text{TP}}{\text{TP+F}} \text{ x } 100 = \frac{36}{36+1} \text{ x } 100 = 97{,}29\%$$

$$Recall \text{ Shitake} = \frac{\text{TP}}{\text{TP+F}} \text{ x } 100 = \frac{37}{37+1} \text{ x } 100 = 97{,}36\%$$

From the calculation of the confusion matrix above, the model performance results of the portabello and shitake classes are obtained with an accuracy of 97.3%, precision of 97.3%, and recall of 97.3%.

In this research, it has succeeded in finding one combination model of the GLCM magnitude, namely in the combination model of the ASM, contrast, IDM, correlation, energy, and dissimilarity magnitudes. where in the research objectives it is said that the author wants to find the best model from several combinations of GLCM magnitudes built. However, with the limited image data used, this research has achieved its goal with the following evaluation: the image data used does not vary so that when testing, the system built produces poor performance. For example, when the image data to be used for testing has different brightness, distance, and noise from the training data.

4. CONCLUSION

The implementation of Gray Level Co-Occurence Matrix and K-Nearest Neighbor feature extraction has been demonstrated to produce good performance in classifying mushroom types with accuracy of 97%, precision of 97%, and recall of 97%, according to the research results on the application of these techniques in mushroom type classification. The best model performance is obtained when the model is built using 6 Gray Level Co-Occurrence Matrix parameters, namely ASM, Contrast, IDM, Correlation, Energy, and Dissimilarity in the K-Nearest Neighbor method for mushroom type classification with an accuracy increase of 4% from models that are only built using 4 models, namely ASM, Contrast, IDM, and Correlation, 1% from the energy combination model, and 5% from the dissimilarity combination model. Additionally, image modifications like rotating and adjusting brightness have a significant impact on how accurately the classification process works.

The addition of new classes to detect classes other than those predetermined when testing and the addition of more and more varied data sets are suggestions for this research. It is hoped that future research will apply parameters to various feature extractions as a comparison to this research. In this study, testing was carried out using image data with different brightness and rotation conditions. Testing that has been done produces performance with an accuracy of 41.67% for testing using image data with different brightness conditions and an accuracy of 58.3% for testing using image data with different rotation conditions from previously created training data. With the limitations and lack of variety of the training data used caused a significant change in the accuracy of the classification process This is because the system is not used to and does not recognise image data with different conditions. not recognise image data with different conditions.

There are several shortcomings in this research that are expected to be an improvement for future research, including data that is less varied so that the resulting performance is not good for testing data with different conditions.

REFERENCES:

[1] E. J. Kusuma, C. A. Sari, E. H. Rachmawanto, and D. R. I. M. Setiadi, “A Combination of Inverted LSB, RSA, and Arnold Transformation to get Secure and Imperceptible Image Steganography,” *J. ICT Res. Appl.*, vol. 12, no. 2, p. 103, 2018, doi: 10.5614/itbj.ict.res.appl.2018.12.2.1.

[2] N. Purwaningsih, I. Soesanti, H. A. Nugroho, S. Pengajar, J. T. Elektro, and D. Teknologi, “EKSTRAKSI CIRI TEKSTUR CITRA KULIT SAPI BERBASIS CO-OCCURRENCE MATRIX,” pp. 6–8, 2015.

[3] W. I. Praseptiyana, A. W. Widodo, and M. A. Rahman, “Pemanfaatan Ciri Gray Level Co-occurrence Matrix (GLCM) Untuk Deteksi Melasma Pada Citra Wajah,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 11, pp. 10402–10409, 2019, Accessed: Aug. 18, 2023. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6685

[4] “Three combination value of extraction features on GLCM for detecting pothole and asphalt road | Arbawa | Jurnal Teknologi dan Sistem Komputer.” https://jtsiskom.undip.ac.id/article/view/13828/12667 (accessed Aug. 16, 2023).

[5] D. P. Pamungkas and A. B. Setiawan, “IMPLEMENTASI EKSTRASI FITUR DAN K-NEAREST NEIGHTBOR UNTUK IDENTIFIKASI WAJAH PERSONAL,” *Joutica J. Inform. Unisla*, vol. 3, no. 2, pp. 187–193, Sep. 2018, doi: 10.30736/JTI.V3I2.233.

[6] A. Rohman, “MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) UNTUK PREDIKSI KELULUSAN MAHASISWA,” *Neo Tek.*, vol. 1, no. 1, Mar. 2015, doi: 10.37760/NEOTEKNIKA.V1I1.350.

[7] N. IBRAHIM, N. IBRAHIM, T. F. BACHERAMSYAH, B. HIDAYAT, and S. DARANA, “Pengklasifikasian Grade Telur Ayam Negeri menggunakan Klasifikasi K-Nearest Neighbor berbasis Android,” *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 6, no. 2, p. 288, Jul. 2018, doi: 10.26760/elkomika.v6i2.288.

[8] M. Ramadhani, S. Suprayogi, and H. B. Dyah, “Klasifikasi Jenis Jerawat Berdasarkan Tekstur Dengan Menggunakan Metode Glcm,” *eProceedings Eng.*, vol. 5, no. 1, Apr. 2018, Accessed: Aug. 18, 2023. [Online]. Available: https://openlibrarypublications.telkomuniversi

ty.ac.id/index.php/engineering/article/view/6049

[9] R. M. Haralick, I. Dinstein, and K. Shanmugam, “Textural Features for Image Classification,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.

[10] A. Kadir, *Dasar pengolahan citra dengan Delphi*. 2013.

[11] D. Rohpandi, A. Sugiharto, M. Yoga, and S. Jati, “Klasifikasi Citra Digital Berbasis Ekstraksi Ciri Berdasarkan Tekstur Menggunakan GLCM Dengan Algoritma K-Nearest Neighbor”.

[12] Khamis and H. S, “APPLICATION OF k-NEAREST NEIGHBOUR CLASSIFICATION IN MEDICAL DATA MINING IN THE CONTEXT OF KENYA,” *Sci. Conf. Proc.*, vol. 0, no. 0, 2014, Accessed: Aug. 18, 2023. [Online]. Available: http://41.204.187.99/index.php/jscp/article/view/1144

[13] P. Cunningham, “k-Nearest neighbour classifiers,” *Mult. Classif. Syst.*, Mar. 2007, Accessed: Sep. 29, 2023. [Online]. Available: https://www.academia.edu/2617752/k_Nearest_neighbour_classifiers