# Teakwood Grade Identification with GLCM and K-NN with Adaboost Optimization

## (Case Study at KPH Cepu)

Nirma Ceisa Santi[1], Hastie Audytra[2]

[1,2] University of Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia

[1] nirmaceisa@unugiri.ac.id, [2] hastie@unugiri.ac.id

*Abstract*— **Teak is one type of tree that has many functions and uses. Teak wood have a very high quality to choose as resource for the manufacture of home furniture such as tables, chairs, cabinets, and others. But middle testers (Perhutani staff) who test the quality of wood grade have limitations if the classification uses the five senses of sight and there are still many furniture entrepreneurs who are often mistaken about teak wood quality assessment. This resulted in a lack of quality grade teak wood used as raw material for making home appliances or for furniture and commerce in Perhutani Corporation, especially KPH Cepu. The teak wood image data is then acquired through preprocessing data ready to be processed. By using GLCM as an image feature extraction both training data and testing data. After the image characteristics are obtained, the image is classified by the K-Nearest Neighbor method with adaboost optimization. The final result is obtained in the form of wood grade quality classification namely grade A, B, C and D according to the class.**

*Index Terms*— **Classification; GLCM; K-nn; Teakwood.**.

## I. INTRODUCTION

Teak is a type of woody tree that has high quality. A large tree, straight trunk and can grow to a height of 30-40 m straight towering. Teak has the biological name *Tectona grandis L.f.* In today's industrial bussines, teak wood is mostly made into veneers to coat the outer skin of expensive plywood, besides that, parquet is also made for floor coverings, even twigs that cannot be used as raw material for processed furniture can still be used as fuel because teak can produce high degrees of heat high enough so that in the past it could be used as fuel in steam locomotives. Not only domestic commodities, Indonesia also exports processed teak wood in the form of outdoor furniture [1]. The largest teak production in Central Java is in Perum Perhutani Kph Cepu which has an area of 33,017.29 Ha. The teak forest area is located in Blora Regency and Tuban Regency and Bojonegoro Regency. Production management in the forest area in the Cepu KPH area is divided into two Forest Management Sub Units (SKPH), namely SKPH north Cepu Utara and SKPH South Cepu. The demand for teakwood production at Kph Cepu is the highest to date. Kph Cepu carries out production activities in accordance with the legal basis of central regulations. The production function is more focused on logging and testing the quality of teak wood. Teak wood grouping serves to determine with grades A, B, C, and D. This test was carried out by the staff of Perum Perhutani Kph Cepu Middle examiner. So far, the intermediate examiners have been doing it manually based on the five senses of sight with teak logs that have been classified by the length and diameter by the loggers. The grouping or classification of the quality of teak wood types aims to facilitate the sale of teak on the Perum Perhutani trading portal at Tokoperhutani.com and facilitate the fulfillment of buyer needs.

Identification of teak wood quality based on the image of the vertical cross-section is still done manually and takes a long time so that the production of teak wood for sale is not optimal. Many studies have been carried out by completing the classification of data record numbers taken through measurement actions during felling [2]. In the processed image data, texture feature extraction is the most recommended considering the homogeneity of the data features that are still in image form. The steps in feature extraction can use Gray level co-occurrence matrix which is a statistical method used as texture feature extraction. After the feature extraction results are complete, the classification is not only processed. The classification process is carried out to group images that have been identified using GLCM into classes that have been determined by processing data information. One of the specific problems with texture images is that the image edges are considered as textures in the search for image features.

R. Qayyum, K. Kamal ,T. Zafar and S. Mathava uses the Particle Swarm Optimization algorithm to train a feed forward neural network used to classify defects using GLCM-based features. The proposed technique shows promising results for the classification of wood defects. Means found that the Square Error of the network for the training dataset was 0.3483, whereas, for the test dataset, the accuracy was found to be 78.26% [3].

While another study by Stefanus Santosa, R. A. Pramunendar, D. P. Prabowo, and Yonathan P. Santosa proved that GLCM displays the adjacency relationship between pixels in the image. The angles used are 0°, 45°, 90° and 135°. In the experiments carried out, it was produced 99.14% with the combination of BPNN GA [4].

Previous research can be concluded if the use of Gray Level Co-occurrence Matrix (GLCM) and Euclidean Distance as a calculation of the distance between images that can produce accuracy accurately. In this study, texture feature extraction for teak wood that has the possibility to be used is Gray Level Co-occurrence Matrix (GLCM) with 5 features, namely Contrast, Energy, Correlation, Homogeneity, and Entropy. As for the classification used is the K-NN classifier because the use of training data to be tested is quite a lot so it is expected to have a sufficient and good level of accuracy. To cover the weakness of the k-nn algorithm, adaboost optimization is used to reduce the presence of weak data or data outliers so that all data can be processed optimally. The data classified is of the type of teak wood texture image data which does not have information that can be processed for classification and ith the K-nn algorithm method, there are still some missing values in the k-nn classification model so that the classification algorithm's performance is not optimal. Processing data classified as teak texture image data in order to have information that can be processed for classification as data information using Gray level co-occurance matrix for feature extraction on teak board texture images. The creation of a grade classification model for teak wood boards using k-nn with adaboost optimization for classification and improving the performance of the classification algorithm.

## II. METHOD

### A. Extraction Texture Feature

Sentiment analysis is one of the methods used to Rizky Andhika Surya, Abdul Fadlil, Anton Yudhana explained that the research he did using the image extraction method which was converted through the gray level coocurance matrix method could be used as raw material for batik image classification. This is done because it is felt that the results of GLCM can show differences between images with different texture [5].

### B. Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrix (GLCM) is a matrix whose elements are the number of pairs of pixels that have a certain brightness level, where the pixel pairs are separated by distance and angle. Angle orientation is formed based on four corner directions, namely, 0°, 45°, 90° and 135°, and the distance between pixels is 1 pixel [5]. The input value from GLCM is a matrix which is a representation of the grayscale image,

the output from GLCM is a co-occurence matrix which we can then extract characteristics based on second-order statistical feature parameters such as contrast, correlation, homogeneity and energy.

### C. Data Mining

Data mining has an important function in this research. data mining processes teak wood chip data to be able to make the information needed and increase knowledge by users of this research. Basically, data mining has four main functions in processing the teak wood image that has been converted through GLCM. The main thing is to predict, including where the testing data has been taken, then to describe what information is in the processed data, then to classify the data based on predetermined groups and finally to see if the data is associated with data that is close to its characteristics.

Classification is the process of finding a model or function that describes and differentiates data into classes. Classification involves the process of examining the characteristics of an object and assigning the object to one of the predefined classes.

### D. Adaboost in KNN

The K-Nearst Neighbor classification works based on an analogy, where test and training data are compared and conclusions are drawn based on the similarity of the data generated by the comparison [6]. The calculation is carried out based on the distance (closeness of the data) which is then known as Euclidean Distance

$$d_{ik} = \sqrt{\sum_j^m (Cij - Ckj)^2} \qquad (2.1)$$

Where :

Cij  = training data

Ckij = testing data

one of the supervised algorithms for data mining with a classification model function is the AdaBoost Method. Initially this algorithm was applied to the regression model, along with the rapid development of computer technology, this method can also be applied to other statistical models. The adaBoost method is an ensemble technique using the loss function of the exponential function to improve the accuracy of the predictions made. Basically the boosting method (AdaBoost) can increase the accuracy in the classification and prediction process by generating a combination of a model, but the results of the classification or prediction chosen are the model that has the greatest weight value. The adaptive boosting method has been reported as a meta-technique to overcome the class imbalance problem.

Adaboost and its variants are very successful in application to domains because of their strong

theoretical basis, accurate predictions and great simplicity. In theory adaboost functions as a performance optimizer for the k-nn classification algorithm so that its performance can be maximized. In the case of an unbalanced data set, the application of boosting will not change the structure of the dataset, which means that the condition of the dataset remains in an unbalanced form.

## III. RESULT AND DISCUSSION

### A. Data Collection

The main data in this study is the image data of teak wood boards acquired by intermediate examiners (experts) in the amount of 20 wooden boards for Grade A, 20 wooden boards for grade B, 20 wooden boards for Grade C and 20 wooden boards for grade D, respectively. each taken from 4 angles, namely 0°, 45°,90°,135°. In total there are 316 data for training data and 20 wooden boards as testing data with 4 corners also producing 80 test data.



Fig. 1.   Picture of Main Data

There are 4 classes in the research that will be tested, namely grade A which indicates the grade of wood that has the best durability, density, and humidity which must be valued the most expensively by the seller, in this case Perum Perhutani Kph Cepu, then Grade B shows that durability, density, and humidity good enough with no wood defects or cambium sources. Grade C is a wood plank grade that has sufficient durability, density, and moisture and is in the category of being able to be used as raw material for furniture or equipment. While the last grade is grade D which has durability, density, and moisture. It can still be used for raw materials but with the condition of the wood. which must be adjusted to the needs (the load to

be borne by the strength of the wood). Data collection techniques are taken directly or commonly referred to as Image Acquisition. This process captures or scans an analog image using a tool to obtain a digital image that has been determined by PUSBANGHUT. When taking photos, the teak wood is placed parallel and 15 cm from the recording device (camera). After that the data is collected in the form of images [7].

### B. Preprocessing

Eliminating the Background, the original image data that has been obtained with different backgrounds and less same angles will be cut off the background with the aim of simplifying calculations in the feature extraction process. Then equalize pixel size piece per piece. The image data resulting from the uniformity of the background will be equalized with the pixel size with the aim of simplifying calculations during the feature extraction process. The pixel size of the image data will be standardized to 400 x 406 pixels. After second step is converting image to grayscale. Converting an image to grayscale can be done by taking all the pixels in the image then the color of each pixel will be retrieved information about the 3 basic colors, namely red, blue and green (via the colortoRGB function), these three basic colors will be added up and then divided by three to get the value average. This average value will be used to color the image pixels so that the color becomes grayscale.
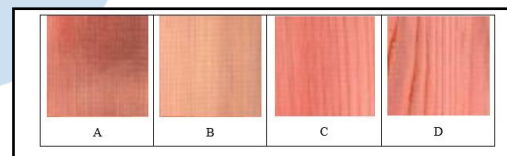


Fig. 2.   Image Data After Preprocessing

### C. Classify with Knn

K-nn calculation is an algorithm that functions to classify some learning data or training data. The training data in this study is in the form of images that have been extracted to produce image characteristics in the form of numbers. The numerical data from the feature extraction will be projected with some test data. The training data sample used in this manual calculation contains 19 training data and 1 data whose class/label is unknown and is called testing data. The training data used are 4 class A data, 5 class B data, 6 class C data and 4. class D data. Because the training data is even, it is recommended to use odd k. The criteria record is c90 means correlation on 90°angle etc.

TABLE I.  SAMPLE TRAINING DATA

| C90 | C135 | L0 | L45 | L90 | L135 | Label |
|---|---|---|---|---|---|---|
| 0,040414 | 0,076267 | 0,91083 | 0,88181 | 0,93734 | 0,88094 | A |
| 0,03165 | 0,051598 | 0,87263 | 0,82111 | 0,89269 | 0,82069 | A |

| 0,02443 | 0,062726 | 0,88888 | 0,87363 | 0,94998 | 0,87169 | A |
|---------|----------|---------|---------|---------|---------|---|
| 0,038466 | 0,088221 | 0,80603 | 0,77591 | 0,90258 | 0,77488 | A |
| 0,043501 | 0,084148 | 0,83501 | 0,80578 | 0,89541 | 0,79785 | B |
| 0,029763 | 0,046631 | 0,65786 | 0,51001 | 0,71078 | 0,52107 | B |
| 0,043065 | 0,15724 | 0,62025 | 0,59862 | 0,89714 | 0,59345 | B |
| 0,024143 | 0,043709 | 0,74565 | 0,68661 | 0,83128 | 0,68507 | B |
| 0,023148 | 0,047316 | 0,91926 | 0,90381 | 0,95439 | 0,90679 | B |
| 0,041837 | 0,10722 | 0,76995 | 0,74034 | 0,89948 | 0,74183 | B |
| 0,042542 | 0,10765 | 0,77484 | 0,7497 | 0,908 | 0,76786 | C |
| 0,007576 | 0,040417 | 0,63174 | 0,63349 | 0,92763 | 0,61665 | C |
| 0,005649 | 0,039971 | 0,26669 | 0,26773 | 0,8938 | 0,25417 | C |
| 0,046941 | 0,13563 | 0,53493 | 0,49927 | 0,82617 | 0,50017 | C |
| 0,023442 | 0,075833 | 0,85024 | 0,83771 | 0,95101 | 0,84152 | C |
| 0,016924 | 0,051342 | 0,57847 | 0,54986 | 0,85269 | 0,55322 | D |
| 0,034852 | 0,080501 | 0,79101 | 0,73985 | 0,9085 | 0,78819 | D |
| 0,024959 | 0,070892 | 0,83014 | 0,81902 | 0,93249 | 0,80603 | D |
| 0,03657 | 0,09994 | 0,83019 | 0,81694 | 0,93241 | 0,81452 | D |

TABLE II.    SAMPLE TESTING DATA

| 0,14777 | 0,15522 | 0,043065 | 0,15724 | 0,62025 | … | 0,9322 |
|---------|---------|----------|---------|---------|---|--------|
| 0,088611 | 0,11898 | 0,048778 | 0,083889 | 0,86263 | … | 0,95806 |

1. Finding the Euclidean Distance

Calculating the Euclidean distance is calculating the root in the test data minus the training data and then squared on each training data. After calculating it produces the table below.

2. Sorting data by 5 nearest neighbors

The determination of the value of K is determined as k = 5, meaning that 5 data records will be taken with the closest distance value.

3. After ranking the 5 smallest data and then checking the class on the nearest neighbor.

TABLE III.    KNN WITH K=5

| ID | Label | D | RANK |
|----|-------|---|------|
| 2A | A | 0,110198 | 3 |
| 4A | A | 0,110253 | 4 |
| 162C | C | 0,156476 | 5 |
| 166C | C | 0,078499 | 1 |
| 320D | D | 0,100962 | 2 |

From this data processing, the closest neighbor sequence is data with data id 166C which has a label C, in sequence 2 there is data with id 320D with label D. In the order of closest distances, sequences 3 and 4 with test data are data labeled A with data id 2A and 4A. In the last order, there is data 162C with label C. From the data processing, the test data has the majority of proximity to data labeled A and C.

### D. Optimation with Adaboost

In the adaboost optimization process, the same sample data will be used with the data above in the knn process [8].

- The first stage is to determine the sample weight of the data. Sample weights are selected from 1 divided by the total number of datasets. There are 20 datasets used, so the weight on the sample records is 1/20.
- The next stage is to determine the total number of possible errors that can occur. And errors that may occur 1/20.
- The third step is to calculate the performance value.
- The last stage is normalization of weight.

### E. Modeling Process

1. Experiment without Adaboost

The next stage is to apply the KNN algorithm in the workspace. First click and put the training data into the blankpage, test data and look for the KNN algorithm and apply the model and performance for the confusion matrix. Search can be done by typing in the toolbar. After everything is there and complete, connect the training data to the KNN algorithm as shown below.
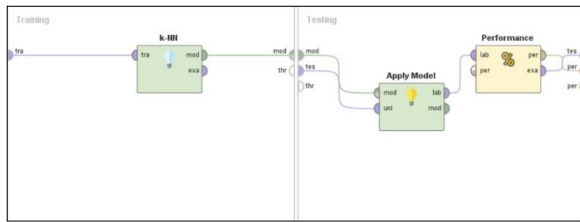
Fig. 3.   Rapid Miner Knn

## 2.   Experiment with Adaboost

In implementing this adaboost, special attention needs to be paid to the rapid miner step. The important thing is the implementation of adaboost for optimization of the training data rule attributes and the k-nn algorithm in the workspace.
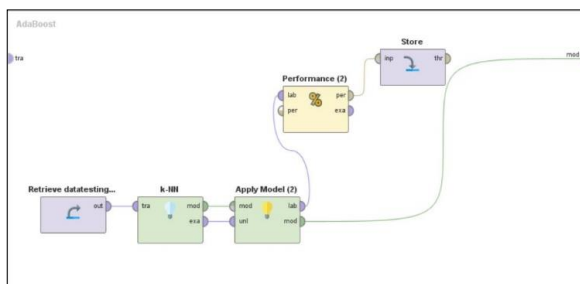


Fig. 4.   Rapid Miner with Adaboost

Details of the experimental value of k on knn adaboost from numbers 5 – 30, the greatest accuracy value is obtained at the value of k = 5.

TABLE IV.   ACCURACY COMPARISON KNN WITH KNN OPTIMIZATION ADABOOST

| K | KNN | KNN & adaboost |
|---|---|---|
| 5 | 80,31 | 91,56 |
| 7 | 67,50 | 91,56 |
| 9 | 69,06 | 83,44 |
| 11 | 60,94 | 76,56 |
| 13 | 62,19 | 66,88 |
| 15 | 59,69 | 64,06 |
| 17 | 59,69 | 66,25 |
| 19 | 59,19 | 64,06 |
| 21 | 61,25 | 65 |
| 23 | 60,62 | 64,06 |
| 25 | 58,75 | 63,44 |
| 27 | 58,13 | 65,31 |
| 29 | 55,93 | 64,69 |

## IV.   CONCLUSIONS

To improve the accuracy of the Knn algorithm in the process of classifying the quality of wood board grades at Kph Perhutani Cepu which will be forwarded to the District Research and Development Center. Semarang, data processing is carried out with the knn process optimized with adaboost. This is because Adaboost can improve accuracy in data processing with the Knn algorithm. The k-nearest neighbor algorithm produces 80.31% accuracy results, while adaboost

optimization can increase the accuracy to almost 100% with an increase of about 19.9%.

TABLE V.   ACCURACY WITH KNN ALGORITHM

| KNN | 80,31% | | | | |
|---|---|---|---|---|---|
| | True. A | True. B | True. C | True. D | Class Precission |
| Pred. A | 76 | 21 | 12 | 10 | 63,87% |
| Pred. B | 0 | 57 | 5 | 6 | 83,82% |
| Pred. C | 4 | 0 | 63 | 3 | 90,00% |
| Pred. D | 0 | 2 | 0 | 61 | 96,83% |
| | | | | | |
| Class Recall | 95,00% | 71,25% | 78,75% | 76,25% | |

TABLE VI.   ACCURACY KNN WITH ADABOOST ALGORITHM

| KNN + adaboost | 100,00% | | | | |
|---|---|---|---|---|---|
| | True. A | True. B | True. C | True. D | Class Precission |
| Pred. A | 80 | 0 | 0 | 0 | 100,00% |
| Pred. B | 0 | 80 | 0 | 0 | 100,00% |
| Pred. C | 0 | 0 | 80 | 0 | 100,00% |
| Pred. D | 0 | 0 | 0 | 80 | 100,00% |
| | | | | | |
| Class Recall | 100,00% | 100,00% | 100,00% | 100,00% | |

From the data above, the boosting method, namely adaboost, has been proven to increase the accuracy of the KNN algorithm on wood sample data.

In the cross validation experiment, trials have been carried out with different k values, namely from numbers 10 – 30 with a value of k at knn k = 5 in the experiment. So that the difference in accuracy obtained is conveyed in the following chart.

So we can conclude that the addition of adaboost optimization in k-NN processing can increase accuracy so that the testing data input process can provide more precise results than using only the knn classification algorithm.
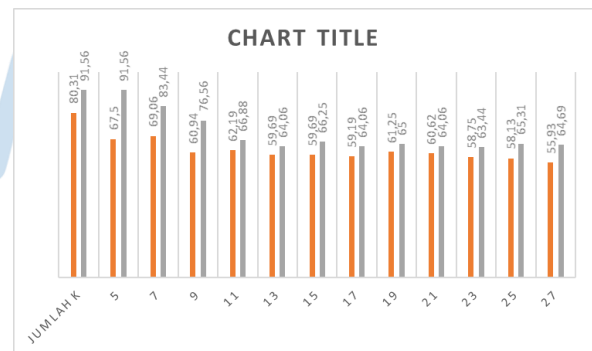


Fig. 5.   Comparasion of Accuracy by Knn and Knn Adaboost

This research contributes ideas for the development of business processes in Perhutani, especially Kph Cepu, especially in the process of buying and selling wooden boards. A good and correct dataset plays an important role in the accuracy produced, so that in future research it is necessary to pay attention to taking datasets in both numbers and are expected to update the data set.

## REFERENCES

[1]   K. CEPU, Pengelolaan Lingkungan Perum Perhuani, Blora, 2014.

[2] S. R. Widodo, "Pemanfaatan Ciri Grey Level CoOccurance Matrix (GLCM) Citra Batik," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. Nov, pp. 5769 - 5776, 2018.

[3] K. T. Z. a. S. M. R.Qoyyum, "Wood defects Classification Using GLCM Based Feature and PSO Trained Neural Network," IEEE, 2016.

[4] R. P. D. P. S. Santosa, "Wood type classification using back Propagation," IAENG International Journal of Computer Science, vol. 2, p. 46, 2019.

[5] A. a. A. Y. R.A Surya, "Ekstraksi Ciri Metode GLCM dan Filter Gabor untuk Klasifikasi Citra Batik Peklaongan," Jurnal Informatika : Jurnal Pengembangan IT, vol. 02 no 02, no. ISSN : 2477-5124, 2017.

[6] Informatikalogi, "Algoritma K-Nearest Neighbor," 2017. [Online]. Available: https://informatikalogi.com/algoritma-k-nn.

[7] PERHUTANI, "TATA KELOLA PERUM PERHUTANI," KPH CEPU.

[8] K. W. a. Y. S.Y Mohamed, "Cost- Sensitive Boosting," pp. 3358 - 3378, 2007.

[9] A. K. a. s. ajib, "Sistem Pendukung keputusan produk olahan kayu jati menggunakan naive bayes".

[10] K. S. a. A.Remesh, "Efficient CBIR Technique Using Color, Texture and Histogram," Internationa; jurnal of engineerin technologi and engineering, Vols. 41 - 44, p. 2, 2015.

[11] E.Acar, "Extraction of texture features from local iris areas by GLCM and iris recognition system based on KNN," European Journal of technic, vol. 6, 2016.

[12] R. w. a. C. Fitriyah, "Perbandingan Akurasi Klasifikasi Citra kayu jati menggunakan Naive Bayes dan KNN," JEPIN, 2019.

[13] J. A. A. R. M.I Fatturahman, "Sistem Klasifikasi Kualitas Kayu Jati berdasarkan tekstur dengan JST menggunakan GLCM," e-Proceding engineering , vol. 6, p. 9537, 2019.

[14] K. S. a. M. Putra, "Klasifikasi Citra Mamogram menggunakan metode K-Means, GLCM, dan Support Vector Machine(SVM)," MERPATI, p. 2252, 2018.

[15] S.Sudrajat, "Deteksi Penyakit Gigi Dengan Menggunakan Metode Filter Gabor dan K-Nearest Neighbour Berdasarkan tekstur citra," 2018.

[16] s. l. a. K. w.song, "Hyperspectral image classification based on knn sparse," IGARSS, no. 978-1-5090-3332-4/16, 2016