Jurnal Mantik, 6 (1) (2022) 494-501

IOCSCIENCE

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Wood Classification For Efficiency in Preventing Illegal Logging Using K-Nearest Neighbor

Rudy Chandra

Information Technology, Faculty of Vocation, Del Institute of Technology, Sitoluama, Laguboti, 22381, North Sumatera, Indonesia.

E-mail: rudy.chandra@del.ac.id

ARTICLE INFO

Article history:
Received: Mar 17, 2022
Revised: Aprl 12, 2022
Accepted: May 30, 2022

Keywords:
classification, invariant moment K-NN, machine learning, wood

ABSTRACT

Wood is a commodity that is usually used for various purposes. Growing demands for timber have led some humans to commit fraudulent and unlawful moves. One of them is through illegal logging of wood species that are protected by the state. The Ministry of Environment and Forestry apart from forest rangers has several problems in classifying wood species. Then technology is used to overcome these problems through the use of machine learning. One suitable algorithm for classifying is K-NN. There are five types of information wood used, specifically marine resin, teak, kruing, meranti, and ironwood. The total wood photos are 1,300 with a complete dataset of 250 images for each type of wood and 10 images for classification testing. The trial was carried out by finding the most optimal K, ie K = 3, after that it was recorded with a confusion matrix. The results obtained are 76% accuracy, 78.8% recall, 76% precision, and 77.37% F1-Score. The higher the value of K, the greater the number of classifications and the lower the accuracy. The higher the value of K, the more types of classifications you want to test, and the less accurate the percentages in the classification process.

1. Introduction

Wood is a commodity that is generally used for various purposes, such as the basic material for houses, home decoration furniture, etc[1]. Some woods have different characteristics, therefore affecting the quality and cost of a type of wood, it is necessary to classify the type of wood, especially the type of wood that needs to be known, so that related parties can use the wood properly[2].

The grouping of wood types is done in two ways, namely by looking at the general characteristics and also the anatomy of the wood[3]. The general characteristic of wood is that it has bark properties that can be obtained even if you do not use an image magnification tool. Classification of wood species takes into account the characteristics of the bark or the anatomical characteristics of the wood, including the arrangement, shape, and size of intruder cells or tissues, which can only be observed with the aid of a magnifying glass such as a loupe or microscope and carried out by experienced experts[4]. Those who want to know the type of wood, really need an expert who classifies the type of wood[5].

The increasing demand for wood circulating in the community has resulted in several individuals committing fraudulent actions and violating the law. one of them by carrying out illegal logging on wood species protected by the state. Therefore, the ministry of environment and forestry and the forest police have had a little difficulty and time-consuming in classifying wood species. And the solution to this problem requires an application to help law enforcement officers such as forest rangers who are in the field be able to be responsive in classifying illegal wood quickly, in less than one to two weeks so as to save time and increase the accuracy of matching wood species. The logging program itself is illegal, including the use of the wrong means to access the forest. In practice, illegal logging is carried out in forest areas which are prohibited in principle[6]. The logging program itself is illegal, including the use of the wrong means to access the forest. There are many types of wood that are considered to have limited trade or are regulated by law, namely sea resin, teak, kruing, meranti, and ironwood. The many types of wood that have almost the same texture can make it difficult for companies to classify wood by type.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The application of technology today is believed to be able to facilitate human work. Machine learning is one alternative in assisting forestry officers in identifying wood species by conducting learning based on existing datasets[7]. A well-known algorithm in the Machine Learning field and widely applied is k-Nearest Neighbor (k-NN) by looks for the smallest value or similarity to the number of neighbors (k). To simplify the k-NN algorithm in classifying wood images, a method is needed to extract the texture characteristics of wood. The well-known feature extraction is Invariant Moment which will produce 7 characteristic values for each image by calculating an object based on position, orientation, and other parameters. Based on these problems, forestry officers hope to take advantage of developing technology to assist officers in making decisions and obtaining wood image identification results through a computer system to simplify and shorten work time.

2. Method

2.1 Wood

Wood is a commodity that is generally used for various purposes, such as the basic material for houses, home decoration furniture, etc[2]. Some woods have different characteristics, therefore affecting the quality and cost of a type of wood, it is necessary to classify the type of wood, especially the type of wood that needs to be known, so that related parties can use the wood properly. There are many types of wood that are considered to have limited trade or are regulated by law, namely sea resin, teak, kruing, meranti, and ironwood. The many types of wood that have almost the same texture can make it difficult for companies to classify wood by type.

2.2 Invariant Moment

Feature extraction invariant moment provide the properties of invariance to scale, position, and rotation[8][9]. The moment that transforms the image function f(x, y) is defined as[10]:

$$\boldsymbol{m_{pq}} = \sum_{X=0}^{H-1} \sum_{y=0}^{W-1} \boldsymbol{x^p\, y^q\, f(x,y)} \tag{1}$$

Where $\boldsymbol{m}$ is the moment you are looking for then p and q are integers i.e. 0.1,2,…, H is the image height, W is the image width, x is the row, y is the column, and f(x,y) is the image intensity value. Furthermore, the central moment for an image is expressed in the equation 2.

$$\mu_{pq} = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} (\boldsymbol{x}-\overline{\boldsymbol{x}})^{p}(\boldsymbol{y}-\overline{\boldsymbol{y}})^{q}\, \boldsymbol{f(x,y)}, \text{ where : } \overline{\boldsymbol{x}} = \frac{m_{10}}{m_{00}} \text{ dan } \overline{\boldsymbol{y}} = \frac{m_{01}}{m_{00}} \tag{2}$$

After getting the values of μ_{20}, μ_{02}, μ_{30}, μ_{03}, μ_{12} and μ_{21} for each object, then to equation 2.5 normalize the value of the center moment.

$$\boldsymbol{\eta_{pq}} = \frac{\mu_{pq}}{\mu_{00}{}^{\gamma}}, \text{ where: } \boldsymbol{\gamma} = \frac{\boldsymbol{p+q}}{\boldsymbol{2}} + \boldsymbol{1} \tag{3}$$

Then the value of the normalization of the center moment of each object will be obtained μ_{20}, μ_{02}, μ_{30}, μ_{03}, μ_{12} and μ_{21}. After that, calculate with equation 3 to get 7 moment invariant values can be used for scale, position, and rotation invariant pattern identification as below[11].

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - 3\eta_{03})^2 \\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\phi_5 &= (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - 3\eta_{03})^2] \\
&\quad +(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} - \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\phi_6 &= (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})[(\eta_{30} - \eta_{12})^2 - (\eta_{21} - \eta_{03})^2] \\
&\quad +4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - 3\eta_{03})^2] \\
&\quad +(3\eta_{21} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} - \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{4}$$

2.3 k-NN

k-Nearest Neighbor (k-NN) is a classification algorithm based on a data set that learn previous data[12].

The learning data is searched and then produces the *k* value of the dataset that is closest to the test data using the Euclidean distance equation[13]. At the dataset stage, the algorithm only stores the vector features and the data for classification. In the classification stage, the same features are calculated for the test data, where the test data whose classification is unknown, or the dataset is retested to determine the accuracy of the classification[14]. Calculate the distance from the test data vector to all classify data vectors, and get the closest k segments. The k-NN algorithm uses the neighborhood classification as a predictive value from new data or test data. Usually calculated based on Euclidean distance[15]:

$$\boldsymbol{d(p,q)} = \sqrt{\sum_{i=1}^{n}(\boldsymbol{q_i} - \boldsymbol{p_i})^2} \quad (5)$$

where :
$\boldsymbol{d(p,q)}$ = Distance from points p and q
$\boldsymbol{q_i}$, $\boldsymbol{p_i}$ = vector Euclidean
$\boldsymbol{n}$ = lots of dots

The steps for calculating the K-Nearest Neighbor method include:

1. Determine K (Distance to nearest neighbor);
2. Calculate the distance between the tested data and the dataset. At this stage, calculating the distance data that you want to predict is calculated with the dataset;
3. Sorting distance values or ranking based on *k* into groups that have the smallest Euclid (sorting distance results from smallest to largest);
4. Collect the Y category (Nearest Neighbor Classification) based on the *k* value or take the nearest neighbor data after being sorted previously, and determine a good *k* value;
5. The next step is to determine the classification. From the results of the dataset, data that dominates will be obtained

The value of *k* is used in all classes. Since the distribution of local samples in the classes is very different, the value of *k* for selecting the most similar locale for each class is usually very different. The K-NN illustration can be depicted on the gray circle as test data while the red and blue circles are dataset, where all red circles are class 1 and all blue circles are class 2 can be seen in the figure 4.

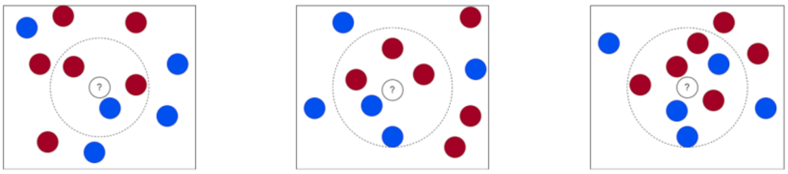


Figure 1. Illustration of k-NN with k = 3, k = 5 and k = 7

2.4 Confusion Matrix

Confusion Matrix is a concept for calculating actual data and predictive results from a classification/identification method used. The Confusion Matrix table has 2 dimensions, including the dimensions of the actual data and the predicted results.

TABEL 1
CONFUSION MATRIX

		True Values	
		True	*False*
Prediction	*True*	TP *Correct Result*	FP *Unexpected result*
	False	FN *Missing result*	TN *Correct absence of result*

The prediction performance calculation is described as follows:

1. Accuracy

Accuracy is the percentage of the total correct number of each identification process which explains the accuracy of the model in identifying.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

$$Accuracy = \frac{TP+TN}{n} \tag{6}$$

2. Recall

Recall the percentage of positive data compared to the overall positive data

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

3. Precision

Precision percentage correct positive with overall positive predicted result

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

4. F1-Score

Comparison of the average precision and recall obtained

$$F1\text{-}score = \frac{2 \times (All\ Precision \times All\ Recall)}{All\ Precision + All\ Recall} \text{ x}100\% \tag{9}$$

2.5 Research Methodology

a. Research Data

This study was conducted to classify the anatomical pattern of wood which consists of 5 types, namely:

1) Sea Resin (Hopea dryobalanoides),
2) Teak (Tectona grandis),
3) Kruing (Dipterocarpus borneensis),
4) Meranti (Shorea leprosula),
5) Ironwood (Eusideroxylon zwageri)

Anatomical images of each wood will be taken using a cellphone camera with a universal clip 60x zoom microscope camera lens with led. The image used is taken from the top and bottom sides of the wood latitude that was taken using a camera. Size of the image resolution obtained is 300 x 300.

Each wood data is divided into 2 parts, namely dataset who contain invariant moment value and classification data. The dataset has a total of 250 images and the testing data to classify has a total of 10 image. so that the total of all data is 1,300 wood images. Detailed pictures of each type of wood can be seen as follows:

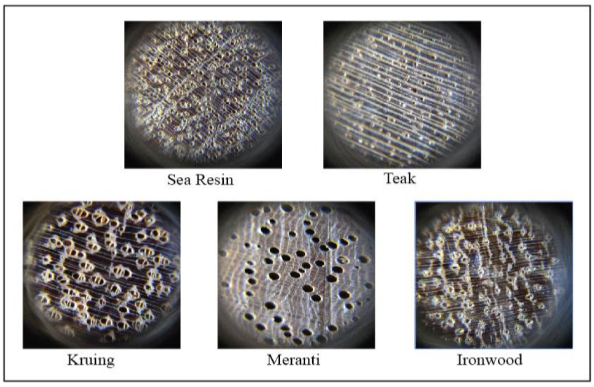


Figure 2. Wood Images

b. Research Analysis Model

The identification process must require several preparations or steps to obtain more optimal results. The following is the procedure for the analysis process of the research model:

1. Wood Dataset Process:

a. Perform image preparation to get the appropriate image results for the Invariant Moment feature extraction process.
b. Choose a dataset according to the type of wood to be trained
c. After getting 7 values of the moment invariant feature, do the training process and save the values in the available database

2. Classification Process:
   a. Select the wood image you want to test
   b. Generate values and will sort the values from the smallest to the largest in the classification process.
   c. Observing the highest number of decisions for the specified K[16]
3. Record the results of the test with the Confusion Matrix table

The scheme in this study is shown in Figure 6 as follows:

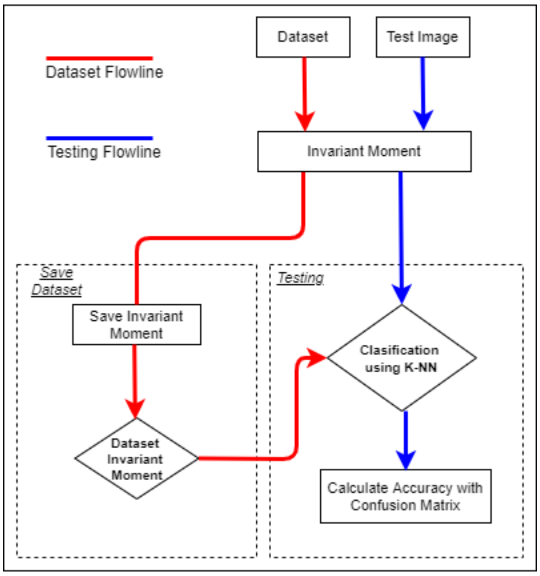


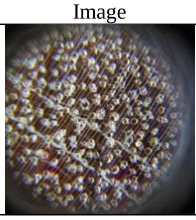
Figure 3. Scheme of Research Analysis Model

3. Result and Discussion

3.1 Feature Extraction

Feature extraction helps you find characteristic values in an image and process them for a classification process using K-NN. Feature extraction using the invariant moment algorithm produces seven values for each image. Each value is used according to the following process. An example is shown in Table 2. When calculated with an invariant moment, seven values are obtained as follows.:

TABEL 2
EXAMPLE OF INVARIANT MOMENT IMAGE

Image	ϕ	Value
	ϕ_1	-1,6198
	ϕ_2	-0,7906
	ϕ_3	-0,1399
	ϕ_4	-0,3831
	ϕ_5	1,3262
	ϕ_6	0,2867
	ϕ_7	1,3205

3.2 Wood Dataset

Each feature extraction process of this process is calculated and stored in the database. If you use the K-NN algorithm to find the minimum value in the classification process, the value stored in the database is the default. Table 3 shows an example of the saved database values.

TABLE 3
EXAMPLE OF DATASET INVARIANT MOMENT

File name	Invariant Moment Value	Class
Sea resin 1.JPG	-1,6198 -0,7906 -0,1399 -0,3831 1,3262 0,2867 1,3205	Sea resin
Teak 1.JPG	-1,5975 -0,644 -0,4517 -0,36 1,2896 0,391 1,3724	Teak
Kruing 1.JPG	-1,5793 -0,7053 -0,4417 -0,3233 1,3233 0,3795 1,3467	Kruing
Meranti 1.JPG	-1,543 -0,7929 -0,3024 -0,3388 1,2884 0,2534 1,4353	Meranti
Ironwood 1.JPG	-1,5204 -0,8757 -0,3317 -0,3038 1,287 0,3523 1,3923	Ironwood

3.3 Classification

In the classification process, the properties of the tested image are calculated using invariant moments. After the seven sample values are obtained, the system calculates these values using the seven sample values in the database for each saved image. The next step is sorting from minimum to maximum. Then calculate the number of neighbors you need, the K value, and note the number of closest neighbors. The number or minimum of the closest neighborhoods is the class of the image.

TABLE 4
TEST RESULT OF WOOD CLASSIFICATION

Wood Classification	Result				
	k=1	k=3	k=5	k=7	k=9
Sea resin	6	7	5	5	6
Teak	6	7	5	6	5
Kruing	8	9	5	5	7
Meranti	8	8	8	8	8
Ironwood	7	7	8	7	7
Total Classification Target	35	38	31	31	33

After adding the correct result of the K value classification, the K value with the highest percentage is selected. From Table 5, we can see that the value with K = 3 has the highest percentage of the other K values. Therefore, all classification results using K = 3 are entered into the confusion matrix table. See Table 5 for a score table that includes the confusion matrix.

3.4 Confusion Matrix

The classification process is recorded and calculated using the confusion matrix to calculate the accuracy and performance of the constructed system. Table 5 shows the results of the confusion matrix table.

TABEL 5
CONFUSION MATRIX

Actual / Prediction		*Confusion matrix Table*				
		Sea Resin	Teak	Kruing	Meranti	Ironwood
Classification Result	Sea Resin	7	0	0	2	1
	Teak	0	7	2	0	1
	Kruing	0	0	9	1	0
	Meranti	0	0	1	8	1
	Ironwood	1	0	0	2	7

The confusion matrix consists of several scoring parameters: accuracy, precision, recall, and F1 score. Accuracy is a performance metric that represents a percentage of the accuracy of all tests. To calculate the accuracy, all the classification values declared correct are summed and divided by all the data.

$$Accuracy = \frac{7+7+9+8+7}{50} x100\% = 76\%$$

The recall is a performance measure that provides information from a positive predicted negative class. Then the calculated recall value of each type of wood can be written into the recall results table like table 6

TABLE 6
RECALL CALCULATION

Classification	Sea Resin	Teak	Kruing	Meranti	Ironwood
True Positif	7	7	9	8	7
False Negatif	1	0	3	5	3
Recall (TP/(TP+FN))	0,875	1	0,75	0,615	0,7

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

To find the overall recall value of the test on the type of wood, it can be calculated using equation 7:

$$All\ Recall = \frac{0{,}875+1+0{,}75+0{,}615+0{,}7}{5} \text{x}100\% = 78{,}8\%$$

Precision is a measure of performance that reports the results of a prediction as a positive class that is positive. You can then enter the calculated precision values for each wood type into the accuracy result table as shown in Table 7.

TABLE 7
PRECISION CALCULATION

Classification	Sea Resin	Teak	Kruing	Meranti	Ironwood
True Positif	7	7	9	8	7
False Positif	3	3	1	2	3
Precission (TP/(TP+FP))	0,7	0,7	0,9	0,8	0,7

To find the overall precision value of the test on the type of wood, it can be calculated using equation 8:

$$All\ Precision = \frac{0{,}7+0{,}7+0{,}9+0{,}8+0{,}7}{5} \text{x}100\% = 76\%$$

Then there is the F1 score, which is the last parameter used to evaluate the classification result. F1Score is a comparison of the average fit and recall obtained. In other words, the F1 Score shows the balance between precision and recall. The calculation is based on equation 9.

$$F1\text{-}score = \frac{2 \times (0{,}76 \times 0{,}788)}{0{,}76 + 0.788} \text{ x } 100\% = 77{,}37\%$$

With the confusion matrix table, the accuracy value is 76%, recall is 78.8%, Precision is 76% and F-Score is 77.37%.

4. Conclusions

The higher the value of K, the greater the number of classifications and the lower the accuracy. The higher the value of K, the more types of classifications you want to test, and the less accurate the percentages in the classification process. Using the wood data, we can see that K is the best value, that is, K = 3. When applied using the confusion matrix, the accuracy value is 76%, recall value is 78.8%, recall value is 76%, and the F1-Score is 77.37%.

References

[1] V. Ristiawanto, B. Irawan, and C. Setianingsih, "Wood classification with transfer learning method and bottleneck features," *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 111–116, 2019, doi: 10.1109/ICOIACT46704.2019.8938428.

[2] J. Chen and C. C. Yang, "The impact of the covid-19 pandemic on consumers' preferences for wood furniture: An accounting perspective," *Forests*, vol. 12, no. 12, 2021, doi: 10.3390/f12121637.

[3] G. J. M. C. Van Vliet, J. Koek-Noorman, and B. J. H. ter Welle, "Wood anatomy, classification and phylogeny of Melastomataceae," *Biodiversity, Evol. Biogeogr. Plants*, vol. 27, pp. 463–473, 1981.

[4] C. K. Wang and P. Zhao, "Classification of wood species using spectral and texture features of transverse section," *Eur. J. Wood Wood Prod.*, vol. 79, no. 5, pp. 1283–1296, 2021, doi: 10.1007/s00107-021-01728-9.

[5] M. I. Taqyudin, B. Irawan, and C. Setianingsih, "Wood Classification Based on Fiber Texture Using Backpropagation Method," *ICSECC 2019 - Int. Conf. Sustain. Eng. Creat. Comput. New Idea, New Innov. Proc.*, pp. 245–250, 2019, doi: 10.1109/ICSECC.2019.8907197.

[6] Z. Guan, Y. Xu, P. Gong, and J. Cao, "The impact of international efforts to reduce illegal logging on the global trade in wood products," *Int. Wood Prod. J.*, vol. 9, no. 1, pp. 28–38, 2018, doi: 10.1080/20426445.2017.1419541.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

[7] G. Carleo *et al.*, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, no. 4, p. 45002, 2019, doi: 10.1103/RevModPhys.91.045002.

[8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, vol. second ed. 2002.

[9] R. A. Asmara, M. Mentari, N. S. Herawati Putri, and A. Nur Handayani, "Identification of Toga Plants Based on Leaf Image Using the Invariant Moment and Edge Detection Features," *4th Int. Conf. Vocat. Educ. Training, ICOVET 2020*, pp. 75–80, 2020, doi: 10.1109/ICOVET50258.2020.9230343.

[10] E. P. Wibowo, S. A. Harseno, and R. K. Harahap, "Feature extraction using histogram of oriented gradient and hu invariant moment for face recognition," *Proc. 3rd Int. Conf. Informatics Comput. ICIC 2018*, no. 1, pp. 1–5, 2018, doi: 10.1109/IAC.2018.8780493.

[11] H. Ming-Kuei, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, pp. 179–188, 1962.

[12] W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," *IEEE Access*, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.

[13] A. I. Saleh, S. A. Shehata, and L. M. Labeeb, "A fuzzy-based classification strategy (FBCS) based on brain–computer interface," *Soft Comput.*, vol. 23, no. 7, pp. 2343–2367, 2019, doi: 10.1007/s00500-017-2930-y.

[14] S. Zhang, "Cost-sensitive KNN classification," *Neurocomputing*, vol. 391, pp. 234–242, 2020, doi: 10.1016/j.neucom.2018.11.101.

[15] S. Li, K. Zhang, Q. Chen, S. Wang, and S. Zhang, "Feature Selection for High Dimensional Data Using Weighted K-Nearest Neighbors and Genetic Algorithm," *IEEE Access*, vol. 8, pp. 139512–139528, 2020, doi: 10.1109/ACCESS.2020.3012768.

[16] S. Zhang *et al.*, "IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS 1 Efficient kNN Classification With Different Numbers of Nearest Neighbors," *Ieee Trans. Neural Networks Learn. Syst.*, pp. 1–12, 2017, [Online]. Available: http://ieeexplore.ieee.org.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.