Indonesian Journal of Data and Science

Volume 4 Issue 3
ISSN 2715-9936
https://doi.org/10.56705/ijodas. v4i3.115

Research Article

Automated Classification of Empon Plants: A Comparative Study Using Hu Moments and K-NN Algorithm

Hayatou Oumarou [1], Nurul Rismayanti [2,*]

[1] Department of Computer Science, HTTC, University of Maroua, Cameroon, Oumarou.hayatou@univ-maroua.cm
[2] Universitas Negeri Malang, Malang, Indonesia, nurulrismayanti.labfik@umi.ac.id

Correspondence should be addressed to Nurul Rismayanti: nurulrismayanti.labfik@umi.ac.id

Received 15 November 2023; Accepted 30 November 2023; Published 31 December 2023

Abstract:

The study "Automated Classification of Empon Plants: A Comparative Study Using Hu Moments and K-NN Algorithm" investigates the potential of image processing and machine learning techniques in the classification of empon plants, specifically ginger and turmeric. Utilizing a dataset of leaf images, the research employed the Canny method for image segmentation and Hu Moments for feature extraction, followed by classification using the K-Nearest Neighbors (K-NN) algorithm. The performance of the model was evaluated through a 5-fold cross-validation method, focusing on metrics such as accuracy, precision, recall, and F1-score. The results showcased the model's variable performance, with the highest accuracy reaching 65.33%. The study contributes to the field by demonstrating the application of Hu Moments in plant classification and by assessing the K-NN algorithm's effectiveness in this context. These findings offer insights into the potential of combining image processing techniques with machine learning for accurate plant classification, paving the way for further research in the area.

Keywords: Empon Plants, Image Processing, Hu Moments, K-Nearest Neighbors (K-NN), Plant Classification.
Dataset link: https://www.kaggle.com/datasets/owenlie/empon-dataset

1. Introduction

In the rapidly evolving field of agricultural technology, the classification of plant species has emerged as a crucial task, particularly in the diverse and rich flora of tropical regions. Empon plants [1], primarily ginger and turmeric, hold significant medicinal and economic value but often pose a challenge for accurate classification due to their similar morphological characteristics. Traditional methods for plant classification, while effective, are time-consuming and often require expert knowledge, prompting a need for automated, efficient, and accurate systems.

This research aims to address the problem of distinguishing between ginger and turmeric plants, leveraging advancements in image processing and machine learning. The main objective of this study is to develop an automated classification system that can accurately identify these plants using their leaf images. This system proposes the use of the Canny [2], [3] method for image segmentation and Hu Moments [4] for feature extraction, coupled with the K-Nearest Neighbors (K-NN) [5] algorithm for classification. By integrating these techniques, the study seeks to enhance the precision and efficiency of plant classification.

The research revolves around several key questions: How effectively can the Canny method segment empon plant images for feature extraction? What is the reliability of Hu Moments in capturing distinguishing features of these plants? And most importantly, how accurately can the K-NN algorithm classify the empon plants based on these extracted features?

While the study presents a novel approach in the classification of empon plants, it is not without limitations. The primary focus on ginger and turmeric may limit the generalizability of the findings to other plant species. Additionally, the effectiveness of the K-NN algorithm may vary based on the chosen value of K and the dataset's diversity [6]–[8].

Despite these limitations, this research contributes significantly to the field of botanical classification and agricultural technology. By automating the classification process, this study not only aids in the efficient identification of empon plants but also sets a precedent for future research in the automated classification of other plant species. The findings of this study could pave the way for broader applications in agricultural research, horticulture, and the medicinal plant industry, where accurate and quick classification of plant species is paramount.

2. Method:

In the methodological framework of this study, we adopted a systematic approach to classify empon plants, focusing on ginger and turmeric, using digital image processing and machine learning techniques. This study employed an experimental design involving image processing and classification. The primary goal was to evaluate the effectiveness of Hu Moments [9] combined with the K-Nearest Neighbors (K-NN) [8], [10] algorithm in accurately classifying images of empon plants. This research is designed in five well-structured main stages, and their aspects are illustrated in **Figure 1**.

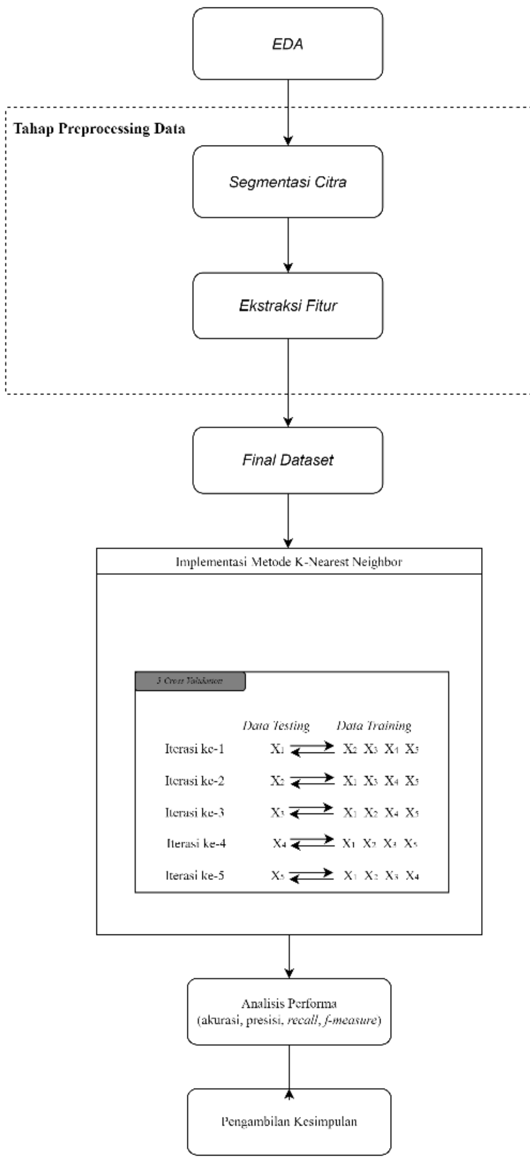

Figure 1. General Research Design Stages

Data Collection Process: **Empon plants**

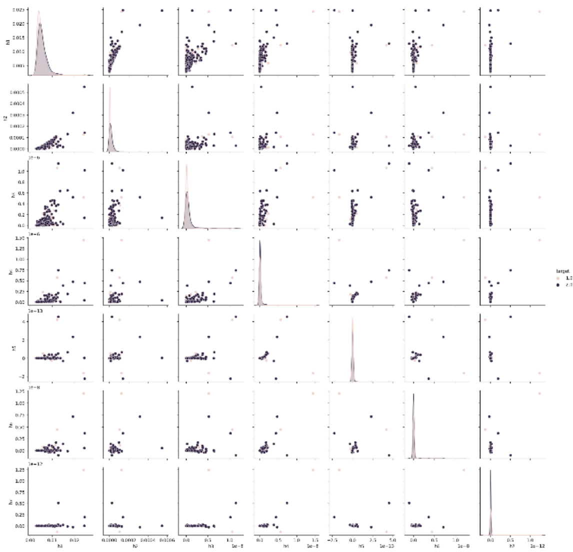

Figure 2. Scatter Plot

The dataset comprised leaf images of empon plants, specifically ginger and turmeric. Each image was labeled with its corresponding plant type, forming the basis for supervised learning. The dataset was divided into a training set and a test set, following the standard practice in machine learning for model validation. These images have been pre-processed using Canny edge detection [11] for segmentation and Hu Moments for feature extraction.

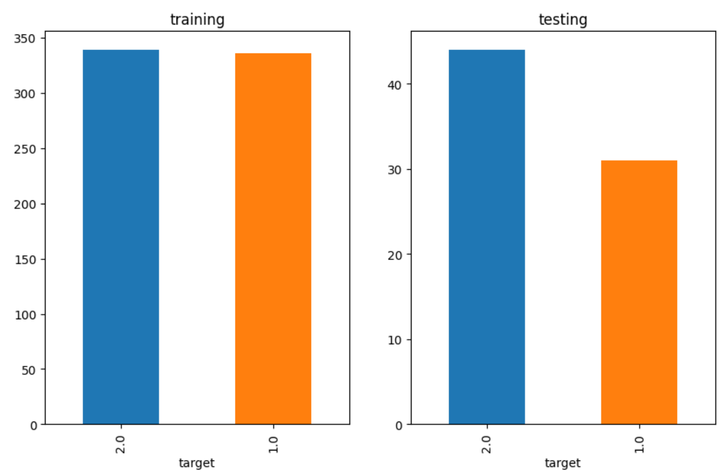

Figure 3. Splitting Dataset 10 % testing, 90% training

Image Segmentation: Canny

The Canny edge detection method is applied to highlight the boundaries within the brain images. This technique works by detecting areas of the image with rapid intensity changes. The Canny method involves several steps, including:

- Gaussian filtering to remove noise.
- Finding intensity gradients of the image.
- Non-maximum suppression to get rid of spurious response to edge detection.
- Double thresholding to determine potential edges.
- Edge tracking by hysteresis.

The mathematical formulation for the gradient magnitude is given by [12]:

$$M(x,y) = \sqrt{G_x^2 + G_y^2} \tag{1}$$

The overall gradient magnitude for each pixel is then computed as **Figure 4** and **5**:

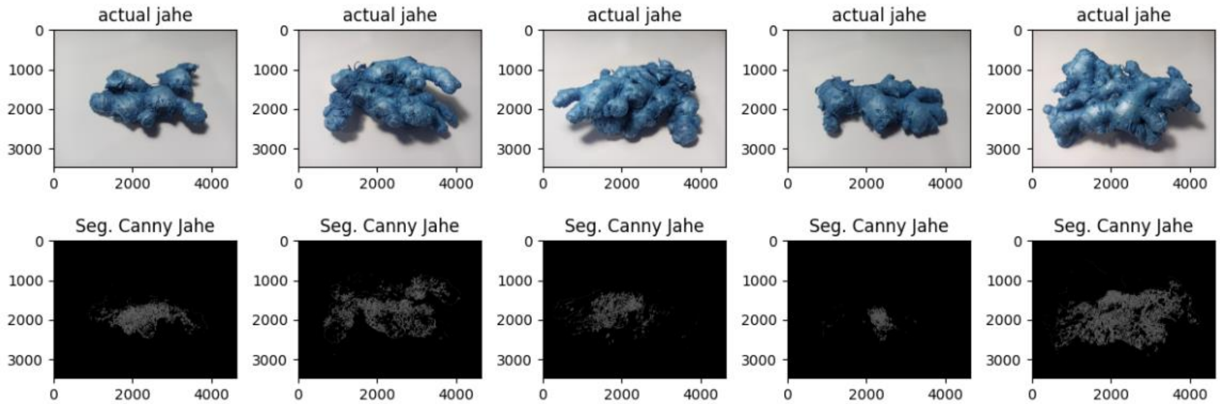

Figure 4. Canny Detection Results for Jahe Class

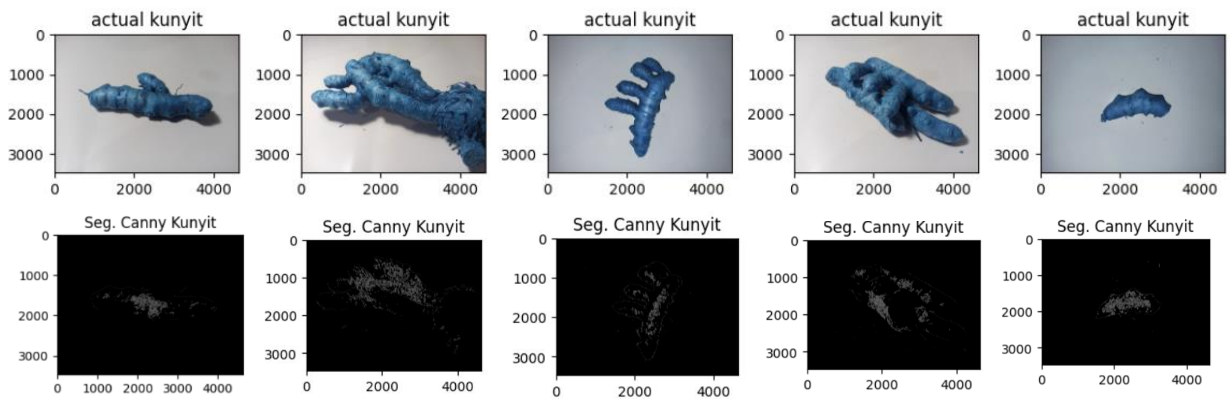

Figure 5. Canny Detection Results for Kunyit Class

Feature Extraction: Hu Moments

Hu Moments are invariant to image transformations and provide a robust feature set for pattern recognition. The seven Hu Moment invariants are calculated from the normalized central moments of the image. The n^{th} order central moment is defined as [13], [14]:

$$\mu_{pq} = \sum_{x,y} (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{2}$$

Where $f(x,y)$ is the pixel intensity at (x,y), and $(\bar{x},\bar{y})$ is the centroid of the image [15], [16].

The Hu moments are derived from these central moments as follows:

$$\begin{aligned} H_1 &= \mu_{20} + \mu_{02} \\ H_2 &= (\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2 \\ &\vdots \\ H_7 &= \mu_{30}\mu_{12} - \mu_{21}\mu_{03} - 3\mu_{12}^2\mu_{03} + 3\mu_{21}^2\mu_{12} \end{aligned} \tag{3}$$

Classification Algorithm: K-NN

The KNN algorithm classifies an image based on how its features compare to the features of images in the training dataset [17], [18]. Given an input image, the algorithm calculates the Euclidean distance between the features of this image and every image in the training set. It then selects the 'K' training images that are closest to the input image and classifies based on the majority label among these 'K' images.

The formula for Euclidean distance in a two-dimensional space is [19]:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \tag{4}$$

K-fold Cross-validation:

The model's performance was evaluated using 5-fold cross-validation and metrics such as accuracy, precision, recall, and F1-score. This comprehensive methodological approach ensures a thorough evaluation of the proposed classification system. This process ensures that each sample is used for validation exactly once. The method's formulaic representation is [19].

$$\mathrm{CV}_{(K)} = \frac{1}{K}\sum_{i=1}^{K} \mathrm{Error}_i \tag{5}$$

Performance Comparison Analysis

Post-validation, the model's performance was assessed using metrics such as accuracy, precision, recall, and F-measure [5], [20]. Their respective formulae are [21].

$$\begin{gathered} Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \\ Pericision = \frac{TP}{(TP + FP)} \\ Recall = \frac{TP}{(TP + FN)} \\ F - measure = \frac{2(presisi \times recall)}{(presisi + recall)} \end{gathered} \tag{6}$$

The above formulas explain:
True Positive (TP): The number of cases correctly predicted as positive by the model.
True Negative (TN): The number of cases correctly predicted as negative by the model.
False Positive (FP): The number of cases incorrectly predicted as positive by the model.
False Negative (FN): The number of cases incorrectly predicted as negative by the model.

These metrics provided a comprehensive understanding of the model's performance, highlighting its strengths and areas of improvement.

3. Results and Discussion

Results

The results of our research, "Automated Classification of Empon Plants: A Comparative Study Using Hu Moments and K-NN Algorithm," demonstrate the performance of the K-NN classifier across different metrics. Utilizing a 5-fold cross-validation approach, we observed variations in accuracy, precision, recall, and F1-score across each fold. The accuracy ranged from a low of 51.33% to a high of 65.33%, precision varied similarly, and the recall and F1-scores followed a comparable pattern.

A detailed table was constructed to encapsulate these metrics, presenting a clear picture of the model's performance across each fold. The data suggested a certain level of variability in the model's ability to classify empon plants accurately. The detailed results are presented in **Table 1** and visualized in **Figure 7** for a clearer understanding and comparison of the metrics across different iterations.

Table 1. Performance Metrics Across 5-Fold Cross-Validation for the Nu-SVM

K-n	Performa			
	Accuracy	***Precision***	***Recall***	***F-Measure***
K-1	57%	57%	57%	56%
K-2	63%	63%	63%	63%
K-3	51%	52%	51%	51%
K-4	59%	59%	59%	59%
K-5	65%	66%	65%	65%
$\sum \boldsymbol{Avg}$	59%	59%	59%	59%

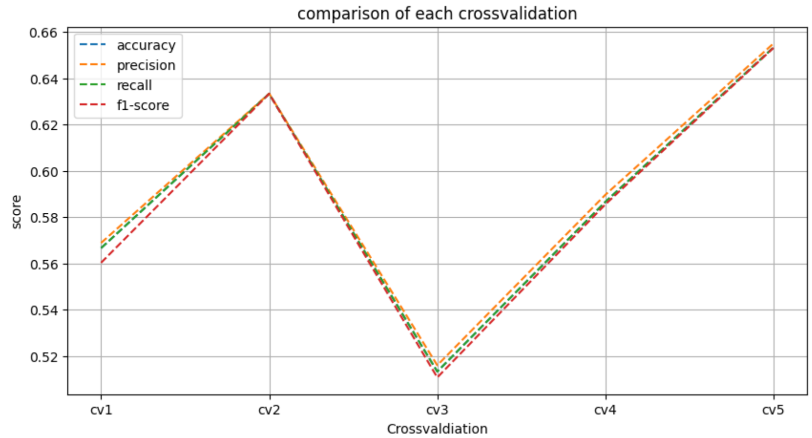

Figure 6. Visualisation Performance Metrics Across 5-Fold Cross-Validation for the Decision Tree

A tabular representation of these results offers a clear visualization, allowing for an immediate grasp of the model's performance nuances across different folds. Such a tabular format is crucial for understanding the specific areas where the model excels and where improvement is needed.

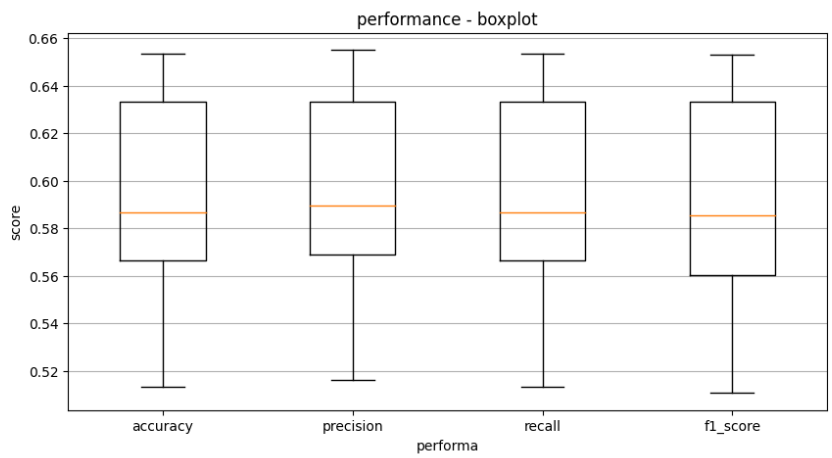

Figure 7. Boxplot Performance Metrics Across 5-Fold Cross-Validation for the Decision Tree

The boxplot provided in Figure 7 offers a visual representation of the performance metrics derived from our research on the automated classification of empon plants using the K-Nearest Neighbors algorithm and Hu Moments for feature extraction. This graphical summary showcases the distribution of accuracy, precision, recall, and F1-score across the 5-fold cross-validation process. The orange line within each box represents the median of the data, while the box itself delineates the interquartile range (IQR), reflecting the middle 50% of scores for each performance metric. The 'whiskers' extend to the minimum and maximum values within 1.5 times the IQR from the lower and upper quartile, respectively, providing insight into the variability and consistency of our model's performance.

Discussion

The interpretation of these results indicates that while the K-NN algorithm, combined with Hu Moments for feature extraction, holds promise for the classification of empon plants, there are noticeable fluctuations in performance metrics across different folds. This variability might be attributed to the dataset's characteristics or the choice of features extracted for classification. A significant finding of our study is the relatively high performance in certain folds (e.g., Fold 5), which suggests that under specific conditions, the model can achieve a high level of accuracy. However, the variability across folds points to a need for further refinement. When compared to previous research, our approach of using Hu Moments for feature extraction in plant classification is relatively novel. Most existing studies focus on more conventional methods, suggesting that our research could contribute a new perspective to the field.

The practical implications of our findings are significant for the field of agricultural technology, particularly in the automated classification of plants. This technology could be beneficial for large-scale agricultural operations and botanical studies where quick and accurate plant classification is essential. However, the research is not without limitations. The variability in the model's performance across different folds indicates that the model may not be entirely robust. Additionally, the study focused only on two types of empon plants, which limits the generalizability of the findings.

For future research, we recommend exploring additional feature extraction methods and alternative machine learning algorithms to enhance the model's accuracy and consistency. Investigating a more diverse set of empon plants could also provide insights into the model's applicability to a broader range of species. Additionally, integrating other forms of data, such as climatic and soil conditions, might improve the classification accuracy and robustness.

4. Conclusion

The study "Automated Classification of Empon Plants: A Comparative Study Using Hu Moments and K-NN Algorithm" aimed to address the challenge of classifying empon plants, specifically ginger and turmeric, using image processing and machine learning techniques. The results revealed varying degrees of success in the model's performance, as indicated by the metrics of accuracy, precision, recall, and F1-score across different folds of cross-validation. The highest accuracy achieved was 65.33%, highlighting potential in the chosen methodology, but also variability that suggests the need for further refinement. This research has contributed to the field by exploring the application of Hu Moments for feature extraction in plant classification, an approach less common in existing literature, and by examining the efficacy of the K-NN algorithm in this context.

Our findings lead to several recommendations for future research and practice. Considering the variability in the model's performance, future studies could explore a combination of different feature extraction techniques and machine learning algorithms to enhance accuracy and reliability. Expanding the research to include a broader range of plant species could also test the model's applicability and robustness in different scenarios. Furthermore, integrating additional data types, like environmental factors, might provide a more comprehensive understanding of plant classification. In practical terms, these advancements could significantly aid in large-scale agricultural operations and botanical research, where efficient and accurate plant classification is crucial.

References:

[1] M. Hanafi and B. Agustian, "Deep Neural Network Method to Classify Empon-Empon Herb Based on E-Nose," *2020 Fifth Int. Conf. …*, 2020, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9288553/.

[2] B. Iqbal, "Canny edge detection and Hough transform for high resolution video streams using Hadoop and Spark," *Cluster Comput.*, vol. 23, no. 1, pp. 397–408, 2020, doi: 10.1007/s10586-019-02929-x.

[3] J. Trivedi, "Canny edge detection based real-time intelligent parking management system," *Sci. J. Silesian*

Univ. Technol. Ser. Transp., vol. 106, pp. 197–208, 2020, doi: 10.20858/sjsutst.2020.106.17.

[4] Y. Jusman, "Machine Learnings of Dental Caries Images based on Hu Moment Invariants Features," *Proc. - 2021 Int. Semin. Appl. Technol. Inf. Commun. IT Oppor. Creat. Digit. Innov. Commun. within Glob. Pandemic, iSemantic 2021*, pp. 296–299, 2021, doi: 10.1109/iSemantic52711.2021.9573208.

[5] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.

[6] A. Aisyah and S. Anraeni, "Analisis penerapan metode K-Nearest Neighbor (K-NN) pada dataset citra penyakit malaria," *Indones. J. Data Sci.*, 2022, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijodas/article/view/22.

[7] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: https://doi.org/10.33096/ijodas.v1i3.20.

[8] I. P. Putri, "Analisis Performa Metode K- Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 21–28, 2021, doi: 10.33096/ijodas.v2i1.25.

[9] S. S. Gornale, "Automatic Detection and Classification of Knee Osteoarthritis Using Hu's Invariant Moments," *Front. Robot. AI*, vol. 7, 2020, doi: 10.3389/frobt.2020.591827.

[10] A. M. Argina, "Application of the K-Nearest Neighbor Classification Method on a Dataset of Diabetes Patients," *Indones. J. Data Sci.*, 2020.

[11] D. N. Lohare, "Comparative Study of Prewitt and Canny Edge Detector Using Image Processing Techniques," *Adv. Intell. Syst. Comput.*, vol. 1187, pp. 705–713, 2021, doi: 10.1007/978-981-15-6014-9_86.

[12] A. Sinra, B. S. W. Poetro, H. Angriani, H. Zein, and ..., "Optimizing Neurodegenerative Disease Classification with Canny Segmentation and Voting Classifier: An Imbalanced Dataset Study," *... Artif. Intell. ...*, 2023, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijaimi/article/view/97.

[13] R. Setiawan, A. Parewe, A. J. Latipah, and ..., "Assessing Bagging-meta Estimator in Imbalanced CT Kidney Disease Classification: A Focus on Sobel and Hu Moment Techniques," *... Artif. Intell. ...*, 2023, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijaimi/article/view/100.

[14] N. Rismayanti, A. Naswin, U. Zaky, M. Zakariyah, and D. A. Purnamasari, "Evaluating Thresholding-Based Segmentation and Humoment Feature Extraction in Acute Lymphoblastic Leukemia Classification using Gaussian Naive Bayes," *Int. J. Artif. Intell. Med. Issues*, vol. 1, no. 2, 2023.

[15] S. AbuRass, "Enhancing Convolutional Neural Network using Hu's Moments," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 12, pp. 130–137, 2020, doi: 10.14569/IJACSA.2020.0111216.

[16] Y. Jusman, "Classification System of Malaria Disease with Hu Moment Invariant and Support Vector Machines," *Proc. - 2022 2nd Int. Conf. Electron. Electr. Eng. Intell. Syst. ICE3IS 2022*, pp. 365–368, 2022, doi: 10.1109/ICE3IS56585.2022.10010304.

[17] H. Basri, P. Purnawansyah, H. Darwis, and F. Umar, "Klasifikasi Daun Herbal Menggunakan K-Nearest Neighbor dan Convolutional Neural Network dengan Ekstraksi Fourier Descriptor," *J. Teknol. dan Manaj. ...*, 2023.

[18] A. Nurjulianty, P. Purnawansyah, and ..., "Perbandingan Metode Naïve Bayes dan K-NN dengan Ekstraksi Fitur GLCM pada Klasifikasi Daun Herbal," *J. MEDIA ...*, 2023, [Online]. Available: http://www.ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/6262.

[19] E. Najwaini, T. E. Tarigan, and F. P. Putra, "Application of the K-Nearest Neighbors (KNN) Algorithm on the Brain Tumor Dataset," *... Artif. Intell. ...*, 2023, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijaimi/article/view/85.

[20] H. Azis, P. Purnawansyah, N. Nirwana, and ..., "The Support Vector Regression Method Performance Analysis in Predicting National Staple Commodity Prices," *Ilk. J. ...*, 2023, [Online]. Available: https://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/1686.

[21] N. Rismayanti and A. P. Utami, "Improving Multi-Class Classification on 5-Celebrity-Faces Dataset using Ensemble Classification Methods," *Indones. J. Data* ..., 2023, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijodas/article/view/78.