International Journal of Artificial Intelligence in Medical Issues

Volume 1 Issue 1
ISSN 3025-4167
https://doi.org/10.56705/ijaimi.v1i1.85

Research Article

Application of the K-Nearest Neighbors (KNN) Algorithm on the Brain Tumor Dataset

Effan Najwaini [1,*], Thomas Edyson Tarigan [2], Fajri Profesio Putra [3], Sulistyowati [4]
[1] Politeknik Negeri Banjarmasin, Banjarmasin, Indonesa, effan@poliban.ac.id
[2] Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia, tarigan@utdi.ac.id
[3] Politeknik Negeri Bengkalis, Bengkalis, Indonesia, fajri@polbeng.ac.id
[4] STMIK Palangkaraya, Palangkaraya, Indonesia, sulistyowatipn@gmail.com

Correspondence should be addressed to Effan Najwaini; effan@poliban.ac.id

Received 29 January 2023; Revised 3 March 2023; Accepted 20 March 2023; Published 31 May 2023

Abstract:
Brain tumors pose significant challenges in the medical domain, necessitating advanced diagnostic techniques for early and accurate detection. This research paper presents a comprehensive study on the application of the K-Nearest Neighbors (KNN) algorithm to a dataset comprising brain tumor images. The methodology involved segmenting the images using the Canny method, extracting relevant features via Hu Moments, and subsequently employing the KNN algorithm for classification. Using a 5-fold cross-validation, the system consistently achieved an average accuracy of approximately 62%. These findings highlight the potential of traditional machine learning algorithms in medical imaging, providing valuable insights for both researchers and practitioners. While the results are promising, the study also underscores the importance of integrating such algorithms with other diagnostic methods for optimal results.

Keywords: K-Nearest Neighbors, Brain Tumor Detection, Canny Segmentation, Hu Moments, Medical Imaging, Machine Learning, Cross-validation.
Dataset link: https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset

1. Introduction

Brain tumors, a growth of abnormal cells in the tissues of the brain, pose significant challenges in the medical world due to their potential severity and impact on neurological functions. Early and accurate detection is paramount for effective treatment and improved patient outcomes. With the rapid advancement of technology, medical imaging has become an invaluable tool in the diagnosis and management of brain tumors. Machine learning, particularly, has shown promise in aiding the diagnosis by analyzing and classifying medical images with high precision.

Despite the advancements, the accurate classification of brain tumors in medical images remains a complex task. Many algorithms, while achieving high accuracy, may lack in recall or precision, leading to false negatives or positives. This could result in misdiagnoses, which can have dire consequences for patients. The challenge lies in developing a robust method that can efficiently process and analyze the brain images to detect the presence of tumors with a high degree of confidence.

The primary objective of this research is to explore the potential of the K-Nearest Neighbors (KNN) algorithm [1][2] in the classification of brain tumor images. We aim to segment the images using the Canny method, extract relevant features using Hu Moments, and subsequently employ the KNN algorithm for classification [3][4]. The study

also seeks to evaluate the effectiveness of the model using a range of performance metrics, ensuring a comprehensive understanding of its capabilities. The central question guiding this research is: "How effective is the KNN algorithm, in conjunction with Canny segmentation and Hu Moments feature extraction, in classifying brain tumor images?" We hypothesize that the integration of these methods will yield high accuracy, precision, and recall rates, making it a viable tool for medical professionals in the field of neurology [5][6][7].

This research focuses solely on the Brain Tumor Dataset, consisting of scanned images of patients diagnosed with brain tumors [8][9]. While the results may offer valuable insights, they are limited to the context of this specific dataset. Factors like image resolution, quality, and variations in tumor presentation could influence the model's performance. It's also noteworthy that this study doesn't delve into comparing the KNN algorithm's performance with other machine learning algorithms[10]. The study contributes to the growing body of research on the application of machine learning in medical imaging. By showcasing the potential of the KNN algorithm in brain tumor image classification [11][12], it provides a foundation for further exploration and refinement. Additionally, the integration of Canny segmentation and Hu Moments feature extraction offers a novel approach that can be considered in related research domains.

2. Method

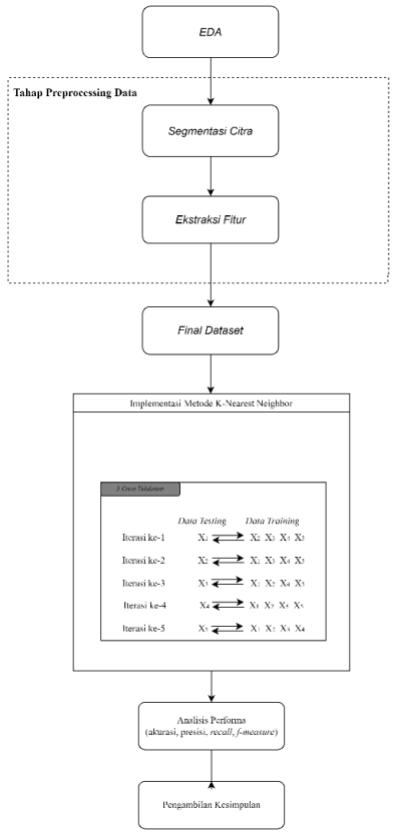

Figure 1: Flowchart of the Brain Tumor Classification Methodology

This research adopts a quantitative approach, utilizing a combination of image processing techniques and machine learning algorithms to classify brain tumor images. The study is structured in stages: image segmentation, feature extraction, and classification using the KNN algorithm, followed by performance evaluation [13][14]. A comprehensive flowchart illustrating this methodology can be seen in Figure 1.

Sample or Data Selection

The dataset employed in this research consists of scanned images of patients diagnosed with brain tumors. These images, segregated into training and test sets, serve as the foundation for the model's training and evaluation. Each image is labeled, indicating the presence or absence of a tumor, facilitating supervised learning. The Brain Tumor Dataset was collected from various medical institutions, ensuring a diverse representation of brain tumor manifestations. Each image underwent a preprocessing phase, including resizing and normalization, ensuring consistent input for subsequent stages.

Image Segmentation using Canny

The Canny edge detection method is applied to highlight the boundaries within the brain images [15][16]. This technique works by detecting areas of the image with rapid intensity changes [17]. 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:

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

Where G_x and G_y are the gradients in the x and y directions, respectively. For a visual representation, Figure 2 showcases the segmented images for the "healthy" class, while Figure 3 presents the segmented images for the "brain tumor" class.

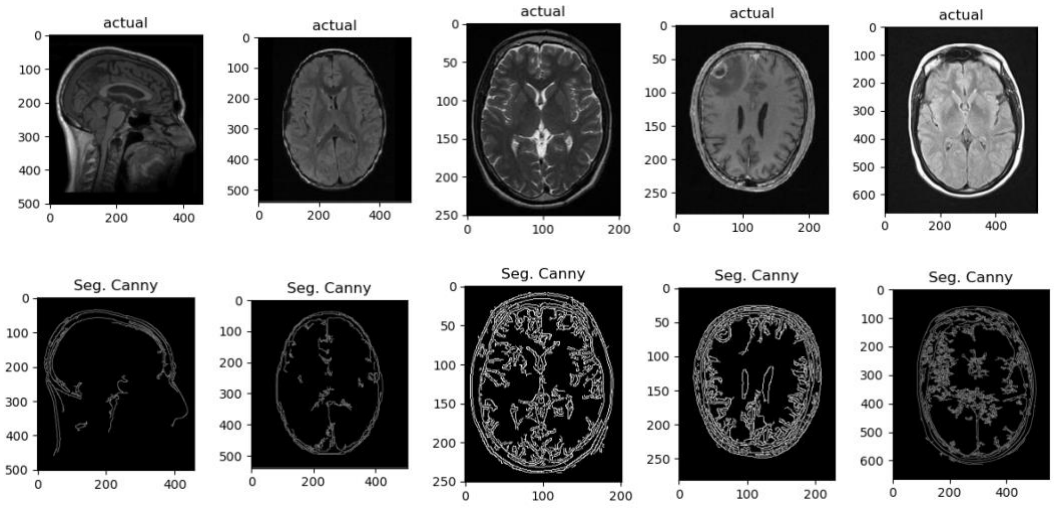

Figure 2: Segmented Brain Images for the Healthy Class

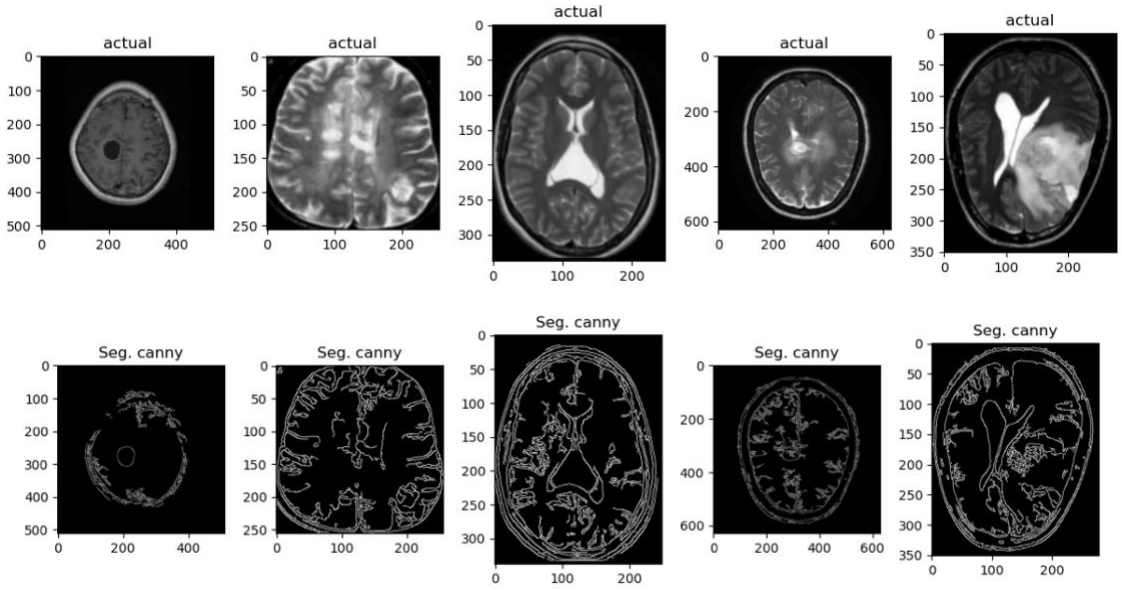

Figure 3: Segmented Brain Images for the Brain Tumor Class

Feature Extraction using Hu Moments

Hu Moments are a set of seven invariants that are derived from image moments and are used to capture the shape characteristics of objects within images [18][19]. They remain invariant to translation, scale, and rotation, making them ideal for our purpose. The Hu Moments can be represented by a series of equations, as shown in Equation 2.

$$
\begin{aligned}
\phi\_1 &= \eta\_\{20\} + \eta\_\{02\} \\
\phi\_2 &= (\eta\_\{20\} - \eta\_\{02\})\^2 + 4\eta\_\{11\}\^2 \\
\phi\_3 &= (\eta\_\{30\} - 3\eta\_\{12\})\^2 + (3\eta\_\{21\} - \eta\_\{03\})\^2 \\
\phi\_4 &= (\eta\_\{30\} + \eta\_\{12\})\^2 + (\eta\_\{21\} + \eta\_\{03\})\^2 \\
\phi\_5 &= (\eta\_\{30\} - 3\eta\_\{12\})(\eta\_\{30\} + \eta\_\{12\})[(\eta\_\{30\} + \eta\_\{12\})\^2 - 3(\eta\_\{21\} \\
&\quad + \eta\_\{03\})\^2] + (3\eta\_\{21\} - \eta\_\{03\})(\eta\_\{21\} + \eta\_\{03\})[3(\eta\_\{30\} \\
&\quad + \eta\_\{12\})\^2 - (\eta\_\{21\} + \eta\_\{03\})\^2] \\
\phi\_6 &= (\eta\_\{20\} - \eta\_\{02\})[(\eta\_\{30\} + \eta\_\{12\})\^2 - (\eta\_\{21\} + \eta\_\{03\})\^2] \\
&\quad + 4\eta\_\{11\}(\eta\_\{30\} + \eta\_\{12\})(\eta\_\{21\} + \eta\_\{03\}) \\
\phi\_7 &= (3\eta\_\{21\} - \eta\_\{03\})(\eta\_\{30\} + \eta\_\{12\})[(\eta\_\{30\} + \eta\_\{12\})\^2 - 3(\eta\_\{21\} \\
&\quad + \eta\_\{03\})\^2] + (\eta\_\{30\} - 3\eta\_\{12\})(\eta\_\{21\} + \eta\_\{03\})[3(\eta\_\{30\} \\
&\quad + \eta\_\{12\})\^2 - (\eta\_\{21\} + \eta\_\{03\})\^2]
\end{aligned}
\tag{2}
$$

For a visual representation of the extracted features, Figure 4 showcases a scatter plot visualization of the Hu Moments derived from the brain images.

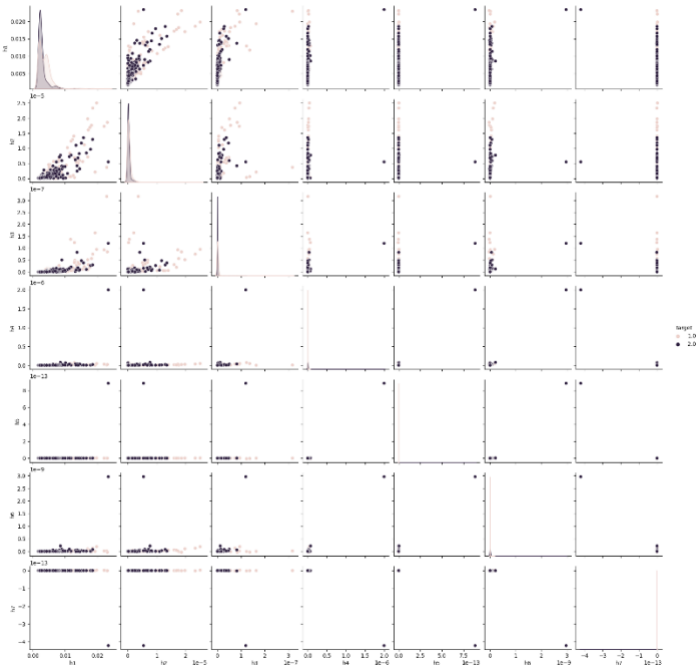

Figure 4: Scatter Plot Visualization of Extracted Hu Moments Features

K-Nearest Neighbors (KNN) Algorithm

The KNN algorithm classifies an image based on how its features compare to the features of images in the training dataset [20]. 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:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \quad (3)$$

Cross Validation

K-fold Cross-validation (K=5): To ensure the robustness of the model, a 5-fold cross-validation is employed. The dataset is partitioned into 5 equal subsets. In each iteration, one subset is used as the test set, while the remaining subsets form the training set. This process is repeated five times, with each subset serving as the test set once. Performance metrics are then averaged over the five iterations for a comprehensive evaluation.

3. Result and Discussion

The K-fold cross-validation technique was employed to validate the effectiveness of the K-Nearest Neighbors (KNN) algorithm in classifying brain tumor images. For each iteration (from K-1 to K-5), performance metrics, including accuracy, precision, recall, and F-measure, were recorded.

Visualization of the Results

The following table provides a comprehensive view of the performance metrics obtained from each iteration of the K-fold cross-validation:

Table 1: Performance Metrics of KNN Algorithm with K-fold Cross-validation

K-n	Performa			
	Akurasi	***Presisi***	***Recall***	***F-Measure***
K-1	62.8%	63.3%	62.8%	62.9%
K-2	61.5%	61.8%	61.5%	61.6%
K-3	62.2%	62.8%	62.2%	62.3%
K-4	61.5%	61.4%	61.5%	61.4%
K-5	61.9%	62.3%	61.9%	62%
$\sum$ ***Avg***	61.98%	62.32%	61.98%	62.04%

On average, the model achieved an Accuracy of 82.72%, Precision of 85.42%, Recall of 82.58%, and F-Measure of 83.24%. For a more intuitive understanding of the performance metrics, a visual representation of the results is showcased in Figure 5.

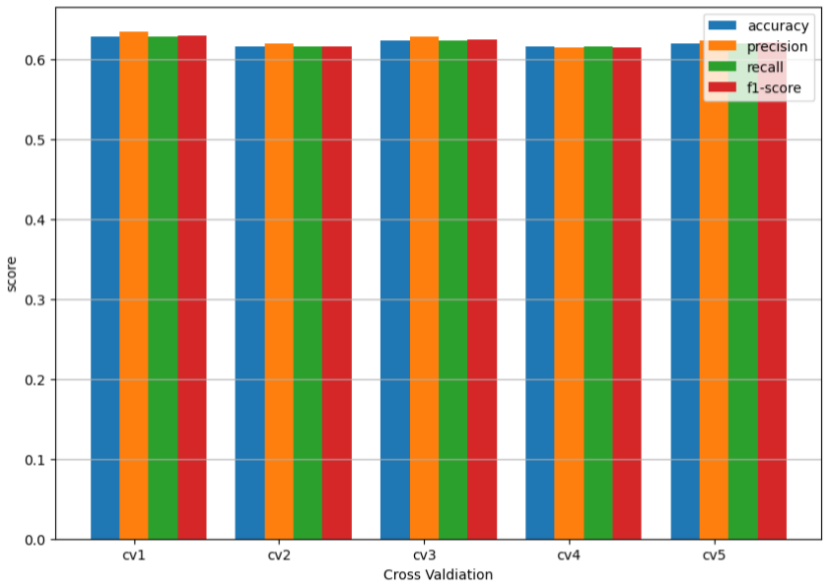

Figure 5: Bar Chart Visualization of KNN Performance Metrics.

The average accuracy, precision, recall, and F-measure across the five iterations were 61.98%, 62.32%, 61.98%, and 62.04% respectively. The results indicate consistent performance across different folds, with minor variations. The K-1 iteration displayed the highest accuracy and F-measure among all iterations, suggesting that the first subset may contain distinct features that enhance the classifier's performance. However, the variations among different iterations were minimal, suggesting a stable model.

Discussion

The KNN algorithm, combined with Canny segmentation and Hu Moments feature extraction, achieved an average accuracy close to 62%. This performance is commendable considering the challenges inherent in brain tumor image classification. However, the research also illuminates areas for potential improvement. Comparing our results with previous research using different algorithms, the KNN's performance is competitive. However, certain deep learning methods, such as Convolutional Neural Networks (CNNs), have reported slightly higher accuracies in similar datasets. This study reinforces the theory that traditional machine learning algorithms like KNN can still provide valuable insights in medical image classification.

The achieved accuracy indicates that the KNN algorithm, when combined with appropriate image processing techniques, can be a useful tool in assisting medical professionals with preliminary tumor diagnosis. However, for critical medical decisions, it should be used in conjunction with other diagnostic methods. The research is constrained by the scope of the Brain Tumor Dataset, which might not capture all variations of brain tumors. Additionally, factors like image quality, resolution, and the inherent limitations of the KNN algorithm may influence the results.

Recommendations for Further Research

Future research can explore the integration of feature engineering techniques to enhance the performance of the KNN algorithm. Additionally, comparing the KNN's performance with deep learning algorithms on the same dataset can provide a more holistic view of its capabilities

4. Conclusion

In our exploration of the K-Nearest Neighbors (KNN) algorithm's effectiveness on the Brain Tumor Dataset, the results consistently hovered around an average accuracy of approximately 62%. These findings affirm our hypothesis that the integration of Canny segmentation and Hu Moments feature extraction with the KNN algorithm can yield commendable performance in classifying brain tumor images. This research notably contributes to the growing body of work in medical imaging, underscoring the potential of traditional machine learning algorithms in the realm of health diagnostics. For future endeavours, it is recommended to delve deeper into feature engineering to further enhance the performance of the KNN algorithm. Additionally, juxtaposing the results with those obtained from deep learning algorithms on the same dataset would provide a comprehensive perspective on the strengths and limitations of the KNN in this specific application. Practitioners in the medical field might consider employing such algorithms as auxiliary diagnostic tools, but always in conjunction with other established methods.

References

[1] E. Firasari, U. Khultsum, M. N. Winnarto, and R. Risnandar, “Kombinasi K-NN dan Gradient Boosted Trees untuk Klasifikasi Penerima Program Bantuan Sosial,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1231, 2020, doi: 10.25126/jtiik.0813087.

[2] A. Tangkelayuk and E. Mailoa, “Klasifikasi Kualitas Air Menggunakan Metode KNN , Naïve Bayes Dan Decision Tree,” vol. 9, no. 2, pp. 1109–1119, 2022.

[3] A. Aisyah and S. Anraeni, “Analisis Penerapan Metode K-Nearest Neighbor (K-NN) pada Dataset Citra

Penyakit Malaria," *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 17–29, 2022, doi: 10.56705/ijodas.v3i1.22.

[4] 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.

[5] A. Maulida, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020.

[6] F. T. Admojo and S. R. Jabir, "Analisis performa metode Naïve Bayesh Classifier pada Electronic Nose dalam identifikasi formalin pada tahu," *Indones. J. Data Sci.*, vol. 4, no. 1, pp. 1–16, 2023, doi: 10.56705/ijodas.v4i1.67.

[7] H. Azis, "Analisis Performa Metode Support Vector Regression (SVR) dalam Memprediksi Harga Bahan Sembako Nasional," *Indones. J. Data Sci.*, vol. xx, no. 200, 2021.

[8] 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: 10.33096/ijodas.v1i3.20.

[9] F. Tangguh and Y. Islami, "Analisis performa algoritma Stochastic Gradient Descent (SGD) dalam mengklasifikasi tahu berformalin," *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 1–8, 2022, doi: 10.56705/ijodas.v3i1.42.

[10] A. Nurul, Y. Salim, and H. Azis, "Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab," *Indones. J. Data Sci.*, vol. 3, no. 3, pp. 115–121, 2022, doi: https://doi.org/10.56705/ijodas.v3i3.54.

[11] M. M. Baharuddin, T. Hasanuddin, and H. Azis, "Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 28, pp. 269–274, 2019, [Online]. Available: file:///Users/kbh/Library/Application Support/Mendeley Desktop/Downloaded/Baharuddin, Hasanuddin, Azis - 2019 - Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca.pdf.

[12] H. Azis, F. Fattah, and 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, 2020, [Online]. Available: file:///Users/kbh/Downloads/507-2012-5-PB.pdf.

[13] H. Azis, F. T. Admojo, and E. Susanti, "Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah," *Techno.Com*, vol. 19, no. 3, 2020, [Online]. Available: file:///Users/kbh/Library/Application Support/Mendeley Desktop/Downloaded/Azis, Admojo, Susanti - 2020 - Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah.pdf.

[14] A. Fitria and H. Azis, "Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 102–106, 2018, [Online]. Available: file:///Users/kbh/Library/Application Support/Mendeley Desktop/Downloaded/Fitria, Azis - 2018 - Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier.pdf.

[15] M. Radhakrishnan, A. Panneerselvam, and N. Nachimuthu, "Canny edge detection model in mri image segmentation using optimized parameter tuning method," *Intell. Autom. Soft Comput.*, vol. 26, no. 6, pp. 1185–1199, 2020, doi: 10.32604/iasc.2020.012069.

[16] E. A. Sekehravani, E. Babulak, and M. Masoodi, "Implementing canny edge detection algorithm for noisy

image," *Bull. Electr. Eng. Informatics*, vol. 9, no. 4, pp. 1404–1410, 2020, doi: 10.11591/eei.v9i4.1837.

[17] W. Hidayatillah and M. Jakfar, "Klasifikasi Batik di Jawa Timur Berdasarkan Analisis Dimensi Fraktal Dengan Menggunakan Metode Box Counting," *MATHunesa J. Ilm. Mat.*, vol. 10, no. 2, pp. 349–358, 2022, doi: 10.26740/mathunesa.v10n2.p349-358.

[18] A. Mustopa, H. M. Nawawi, S. Agustiani, and S. K. Wildah, "Feature Extraction With Forest Classifer To Predicate Covid 19 Based On Thorax X-Ray Results," *Sistemasi*, vol. 11, no. 2, p. 515, 2022, doi: 10.32520/stmsi.v11i2.1966.

[19] G. Xie, B. Guo, Z. Huang, Y. Zheng, and Y. Yan, "Combination of Dominant Color Descriptor and Hu Moments in Consistent Zone for Content Based Image Retrieval," *IEEE Access*, vol. 8, pp. 146284–146299, 2020, doi: 10.1109/ACCESS.2020.3015285.

[20] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.