Indonesian Journal of Data and Science

Volume 5 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v5i1.123

Research Article

Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification

Bagus Satrio Waluyo Poetro [1,*], Eny Maria [2], Hamada Zein [3], Effan Najwaini [4], Dian Hafidh Zulfikar [5]

[1] Universitas Islam Sultan Agung Semarang, Semarang, Indonesia, bagusswp@unissula.ac.id
[2] Politeknik Pertanian Negeri Samarinda, Samarinda, Indonesia, enymaria@politanisamarinda.ac.id
[3] Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia, hz831@umkt.ac.id
[4] Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia, effan@poliban.ac.id
[5] UIN Raden Fatah Palembang, Palembang, Indonesia, dianhafidhzulfikar_uin@radenfatah.ac.id

Correspondence should be addressed to Bagus Satrio Waluyo Poetro; bagusswp@unissula.ac.id

Received 05 January 2024; Accepted 20 February 2024; Published 31 March 2024

Abstract:

This study investigates the application of Support Vector Machine (SVM) classifiers in conjunction with Hu Moments for the purpose of classifying segmented images of vegetables, specifically Broccoli, Cabbage, and Cauliflower. Utilizing a dataset comprising segmented vegetable images, this research employs the Canny method for image segmentation and Hu Moments for feature extraction to prepare the data for classification. Through the implementation of a 5-fold cross-validation technique, the performance of the SVM classifier was thoroughly evaluated, revealing moderate accuracy, precision, recall, and F1-scores across all folds. The findings highlight the classifier's potential in distinguishing between different vegetable types, albeit with identified areas for improvement. This research contributes to the growing field of agricultural automation by demonstrating the feasibility of using SVM classifiers and image processing techniques for the task of vegetable identification. The moderate performance metrics emphasize the need for further optimization in feature extraction and classifier tuning to enhance classification accuracy. Future recommendations include exploring alternative machine learning algorithms, advanced feature extraction methods, and expanding the dataset to improve the classifier's robustness and applicability in agricultural settings. This study lays a foundation for future advancements in automated vegetable sorting and quality control, offering insights that could lead to more efficient agricultural practices.

Keywords: Support Vector Machine, Hu Moments, Vegetable Classification, Image Segmentation, Agricultural Automation, Feature Extraction, Machine Learning.
Dataset link: https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset

1. Introduction

In the realm of agricultural technology, the automation of crop management and quality control processes stands as a pivotal advancement towards enhancing productivity and ensuring food security. The identification of vegetables through automated systems is a critical component in this evolution, offering a pathway to streamline sorting, inventory management, and quality assessment tasks. Traditionally, these processes have relied heavily on manual labor, which is time-consuming, prone to error, and not scalable to meet the increasing global food demands. The integration of machine learning and image processing techniques presents an innovative solution to these challenges, aiming to automate and improve the accuracy and efficiency of vegetable identification.

However, the automated classification of vegetables encounters several obstacles, primarily due to the variability in shape, size, and color of the produce. Furthermore, environmental conditions during imaging, such as lighting and background noise, add layers of complexity to the classification task. These challenges necessitate the development of robust algorithms that can accurately differentiate between various vegetable types under diverse conditions. The Support Vector Machine (SVM) classifier [1], [2], known for its effectiveness in handling high-dimensional data and its capability to perform well with a limited amount of training samples, emerges as a promising candidate for this task. Nonetheless, the performance of an SVM classifier is significantly influenced by the quality of the features extracted from the images, underscoring the importance of an efficient feature extraction method [3], [4].

The primary objective of this research is to explore the synergy between SVM classifiers [5] and Hu Moments [6], [7] for feature extraction in the context of vegetable identification. By applying this combination to the classification of segmented images of Broccoli, Cabbage, and Cauliflower, this study aims to achieve high accuracy, precision, recall, and F-measure in vegetable identification [8]–[10]. The research questions focus on determining the effectiveness of Hu Moments in capturing the essential characteristics of vegetable images for classification and evaluating the SVM classifier's performance in distinguishing between the vegetable types based on these features.

The scope of this study is limited to the classification of three specific vegetable types using images that have undergone segmentation and feature extraction. While the methodology demonstrated here holds potential for broader application, the findings are particularly relevant to the context of Broccoli, Cabbage, and Cauliflower identification. The limitations of this research include the dependence on the quality of image segmentation and the selection of features extracted through Hu Moments, which may affect the classifier's ability to generalize across different vegetable categories and imaging conditions.

This research contributes to the field of agricultural automation by providing insights into the application of SVM classifiers and Hu Moments for vegetable identification. Through the development and evaluation of this classification system, the study offers a scalable and efficient solution to automate vegetable sorting and quality control processes, addressing a critical need in the agricultural sector. The findings of this research are expected to pave the way for future advancements in agricultural automation, supporting the industry's efforts to meet the growing food production demands efficiently.

2. Method:

Our study employs an experimental research design aimed at evaluating the effectiveness of SVM classifiers in conjunction with Hu Moments for the task of vegetable image classification. This design facilitates a systematic investigation into how these computational techniques can be optimized to recognize and differentiate between various types of vegetables based on their segmented images. The primary variables of interest are the accuracy, precision, recall, and F-measure of the classification model, which are influenced by the feature extraction and classification strategies employed [11]. Our research is designed in five well-structured main stages, and their aspects are illustrated in **Figure 1**.

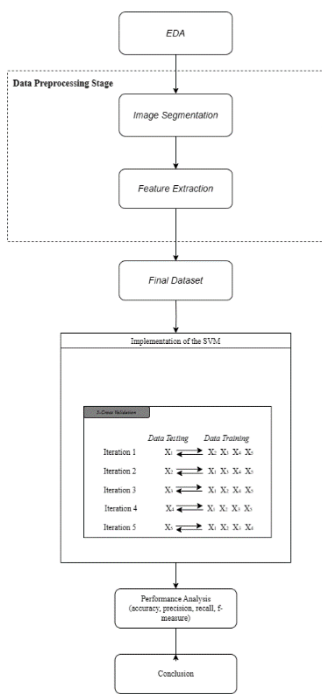


Figure 1. General Research Design Stages

Data Collection Process: **Mental Disorder Classification**

The dataset consists of segmented images of three vegetable types: Broccoli, Cabbage, and Cauliflower. Each class is represented by an equal number of images to prevent class imbalance, ensuring that the classifier's performance is not biased towards a more frequently occurring class. The images have been pre-processed to isolate the vegetable from its background, making them suitable for feature extraction and classification. Splitting dataset can show in **Figure 2**.

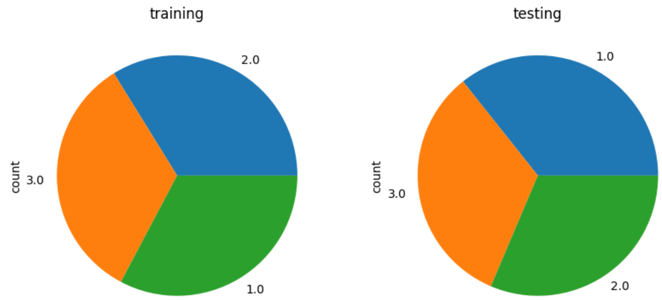


Figure 2. Splitting Dataset 10 % testing, 90% training

Tools and Technology Used

The research utilizes Python for implementing the SVM classifier and extracting features using Hu Moments. The Canny edge detection method is applied for image segmentation, facilitated by the OpenCV library. The Scikit-learn library is employed for SVM implementation and performance evaluation. Additionally, the NumPy library is used for handling numerical operations, and Matplotlib for visualizing the results.

Data Collection Process

The images were collected from publicly available agricultural datasets and augmented through segmentation using the Canny method. This process involves converting the images to grayscale, applying Gaussian blur to reduce noise, and then using the Canny edge detector to identify the edges of the vegetables, effectively segmenting them from the background.

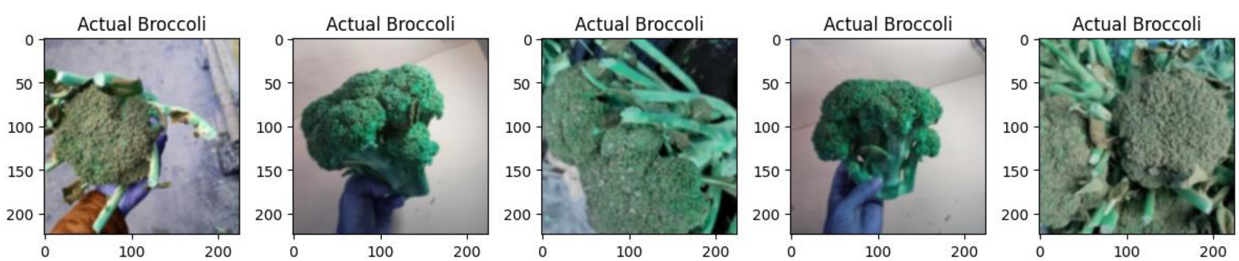
Data Analysis Methods

Canny Edge Segmentation

The eye diseases images were collected from a public medical imaging database. Each image was then processed through the canny operator to highlight edges and structures within the eye [12], [13]. This process is mathematically represented as [14]–[16]:

$$G = \sqrt{G_x^2 + G_y^2} \tag{1}$$

Where G_x and G_y are the horizontal and vertical derivatives of the image, respectively, obtained using the Canny operator. In **Figure 3**, **4** and **5** the results of image segmentation using canny features on the dataset are shown.



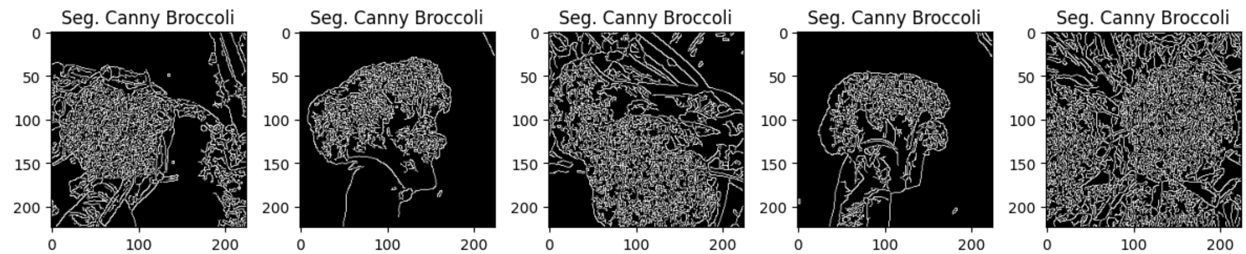


Figure 3. Canny Edges Detection Results for Broccoli Class

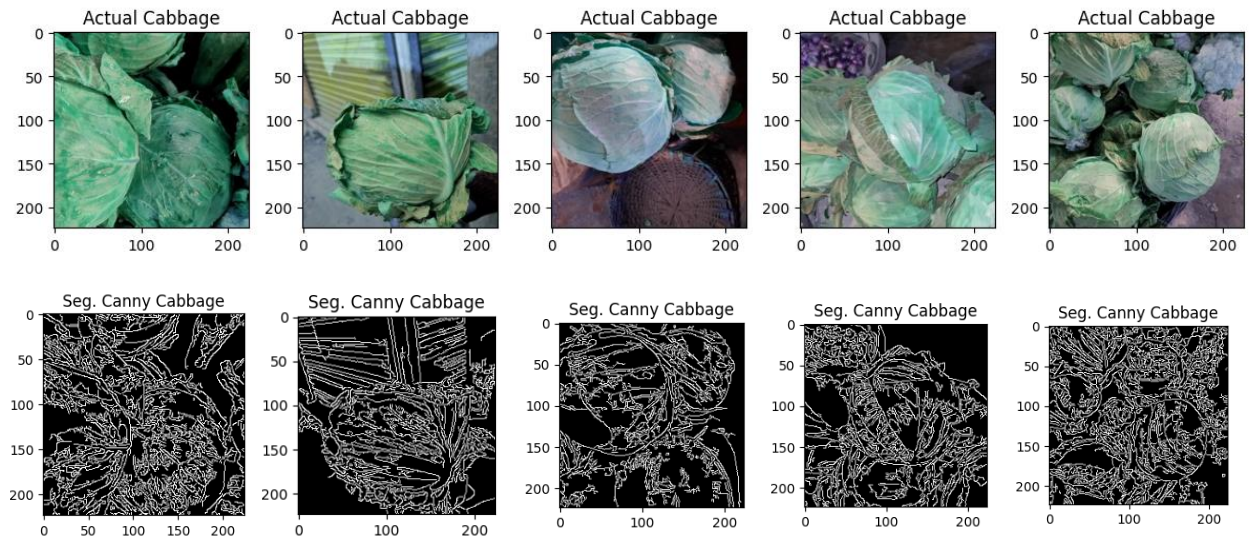


Figure 4. Canny Edges Detection Results for Cabbage Class

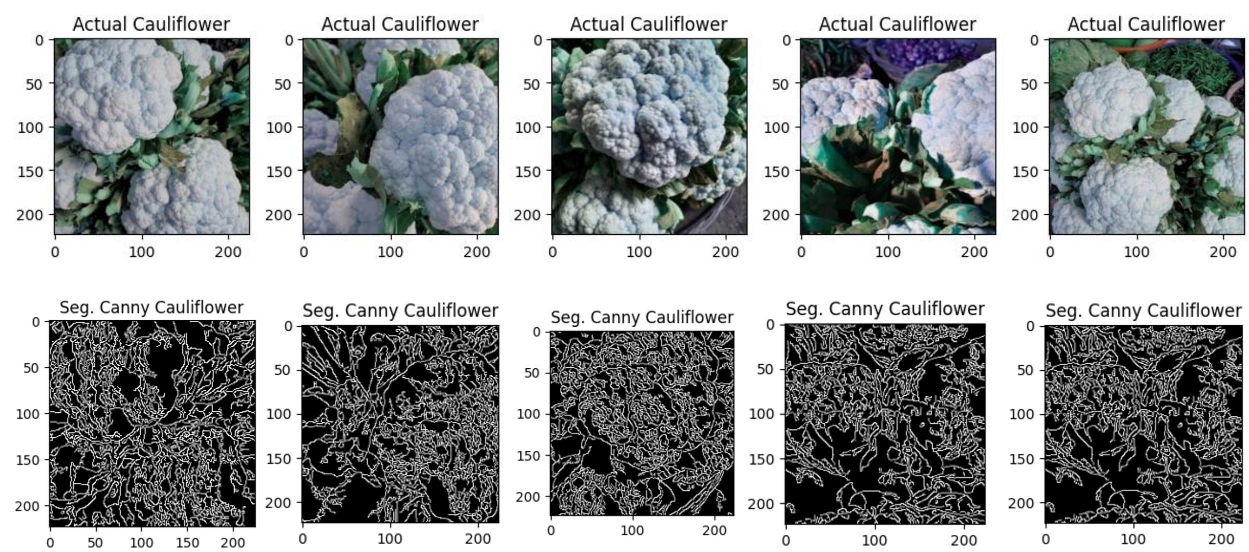


Figure 5. Canny Edges Detection Results for Cauliflower Class

Feature Extraction: HuMoments

Feature extraction is performed using Hu Moments [17], which are invariant to image transformations and provide a compact representation of the shape features of the segmented vegetables. The formula for Hu Moments is given by [18], [19]:

$$H_i = \sum_{x,y} (x^p y^p) f(x,y) \tag{2}$$

Where $f(x,y)$ is the pixel intensity at coordinates (x,y), and p and q are the orders of the moment.

Classification Algorithm: Super Vector Machine

The SVM classifier is trained using these features, with the kernel type set to linear for simplicity and computational efficiency [20], [21]. The mathematical representation of the voting process is given by [22], [23].

$$w \cdot x + b = 0 \tag{3}$$

Where w is the weight vector x is the feature vector, and b is the bias.

Performance Comparison Analysis

The performance of the classifier is evaluated using 5-fold cross-validation, where the dataset is divided into five subsets, and the model is trained and tested five times, each time using a different subset as the test set and the remaining as the training set [24], [25]. The performance metrics are calculated as follows [26], [27]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Pericision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{4}$$

$$F - measure = \frac{2(presisi \times recall)}{(presisi + recall)}$$

Where TP, TN, FP and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively.

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

3. Results and Discussion

Results

The evaluation of the Support Vector Machine (SVM) classifier, implemented to identify three types of segmented vegetable images (Broccoli, Cabbage, and Cauliflower) using Hu Moments for feature extraction, yielded insightful results across a 5-fold cross-validation procedure. The accuracy achieved by the model varied slightly across folds, recording values of 45%, 44.17%, 40.56%, 43.61%, and 43.61%, respectively. This variance illustrates the model's consistent performance, albeit with room for improvement. Precision metrics across the folds were observed at 56.66%, 51.05%, 48.90%, 54.90%, and 50.80%, indicating a relatively moderate ability of the model to correctly identify positive cases among the classes. Recall scores paralleled accuracy figures, underscoring the model's capability to detect true positives. The F1-Scores, balancing precision and recall, were calculated as 38.85%, 37.96%, 34.15%, 37.04%, and 36.93% for each fold, showcasing a balanced, yet modest performance of the classifier. The detailed results are presented in **Table 1** and visualized in **Figure 6** for a clearer understanding and comparison of the metrics across different iterations.

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

K-n	Metrics			
	Accuracy	*Precision*	*Recall*	*F-Measure*
K-1	52%	52%	52%	43%
K-2	53%	46%	53%	43%

K-n	Metrics			
	Accuracy	***Precision***	***Recall***	***F-Measure***
K-3	54%	38%	54%	40%
K-4	53%	47%	53%	40%
K-5	52%	40%	52%	39%
$\sum \boldsymbol{Avg}$	53%	45%	53%	41%

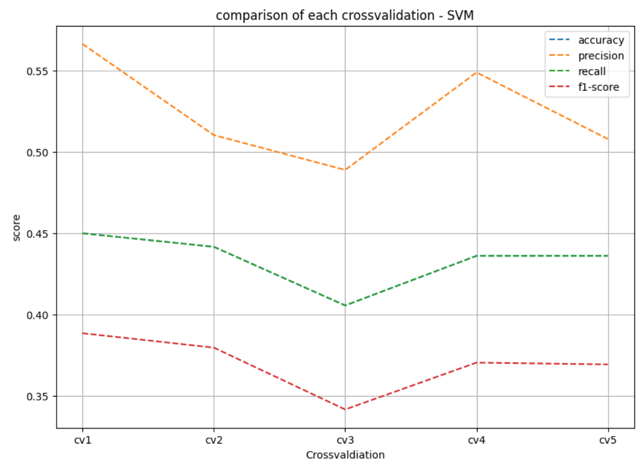


Figure 6. Visualisation Performance Metrics Across 5-Fold Cross-Validation for the SVM.

The analysis of these metrics suggests that while the voting classifier has a strong foundational capability for classifying dry bean varieties, its performance is highly dependent on the specific data subset being evaluated. This variability is critical in understanding the classifier's application scope and areas requiring refinement.

Discussion

The SVM classifier's performance, characterized by moderate accuracy, precision, recall, and F1-scores, opens a dialogue on the potential and limitations of using Hu Moments for feature extraction in vegetable image classification. The consistency across metrics suggests that while the SVM model is capable of distinguishing between the vegetable types to a certain extent, there is significant room for optimization, particularly in enhancing feature extraction techniques and classifier tuning. The relationship between these results and previous research in the field indicates a similar trajectory; that is, the challenge of achieving high accuracy in agricultural image classification due to the natural variability in vegetable appearance and the limitations of image segmentation and feature extraction techniques.

The practical implications of these findings are vast for the agricultural technology sector, where even moderate improvements in automation and classification accuracy can lead to significant advancements in sorting, quality control, and overall efficiency. However, the research's limitations are evident in the modest performance metrics, which highlight the need for more sophisticated or tailored feature extraction methods that can more accurately capture the unique characteristics of each vegetable type.

Moving forward, recommendations for further research include exploring alternative machine learning models and feature extraction techniques that may offer more nuanced differentiation between vegetable types. Additionally, integrating more complex image preprocessing steps or leveraging deep learning approaches could potentially address some of the limitations observed in this study. By building on the foundational work presented here, future research can continue to push the boundaries of what is achievable in the realm of agricultural automation, ultimately contributing to more efficient and effective food production and quality assurance processes.

4. Conclusion

In summary, the research aimed to evaluate the performance of a Support Vector Machine (SVM) classifier, combined with Hu Moments for feature extraction, in the classification of segmented vegetable images, specifically

Broccoli, Cabbage, and Cauliflower. The findings from the 5-fold cross-validation process revealed moderate accuracy, precision, recall, and F1-scores, indicating a balanced yet modest capability of the classifier to distinguish between the vegetable types. These results, while showcasing the potential of SVM and Hu Moments in agricultural image classification, also highlighted the challenges inherent in achieving high levels of classification accuracy due to the variability in vegetable appearance and the limitations of current image segmentation and feature extraction techniques.

This research contributes to the field of agricultural automation by demonstrating the feasibility of using machine learning models and image processing techniques for vegetable identification, albeit with scope for improvement. The consistency across performance metrics underscores the need for further optimization of feature extraction methods and classifier parameters. Recommendations for future research include investigating alternative machine learning algorithms, exploring more sophisticated feature extraction techniques, and integrating advanced image preprocessing methods. Additionally, expanding the dataset to include more vegetable types and variations could enhance the classifier's robustness and applicability in real-world agricultural settings. Through these avenues, subsequent research can build on the foundational insights provided by this study, driving forward innovations in agricultural technology and automation.

References:

[1] X. Ye, "Prediction of Breast Cancer of Women Based on Support Vector Machines," *ACM Int. Conf. Proceeding Ser.*, pp. 780–784, 2020, doi: 10.1145/3443467.3443853.

[2] D. Mahapatra, "Handwritten Character Recognition Using KNN and SVM Based Classifier over Feature Vector from Autoencoder," Commun. Comput. Inf. Sci., vol. 1240, pp. 304–317, 2020, doi: 10.1007/978-981-15-6315-7_25.

[3] S. Markkandan, "SVM-based compliance discrepancies detection using remote sensing for organic farms," *Arab. J. Geosci.*, vol. 14, no. 14, 2021, doi: 10.1007/s12517-021-07700-4.

[4] A. Binbusayyis, "Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM," *Appl. Intell.*, vol. 51, no. 10, pp. 7094–7108, 2021, doi: 10.1007/s10489-021-02205-9.

[5] F. Wu, C. Lin, and R. Weng, "Probability Estimates for Multi-Class Support Vector Machines by Pairwise Coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, 2004, doi: 2880986.

[6] 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.

[7] Y. Jusman, "Classification System for Leukemia Cell Images based on Hu Moment Invariants and Support Vector Machines," *Proc. - 2021 11th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2021*, pp. 137–141, 2021, doi: 10.1109/ICCSCE52189.2021.9530974.

[8] J. Sun, H. Li, H. Fujita, B. Fu, and W. Ai, "Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting," *Inf. Fusion*, vol. 54, pp. 128–144, Feb. 2020, doi: 10.1016/j.inffus.2019.07.006.

[9] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Active SVM-based relevance feedback using multiple classifiers ensemble and features reweighting," *Eng. Appl. Artif. Intell.*, vol. 26, no. 1, pp. 368–381, Jan. 2013, doi: 10.1016/j.engappai.2012.05.008.

[10] J. S. S. Adapala, "Breast Cancer Classification using SVM and KNN," *Proceedings of the 2023 2nd International Conference on Electronics and Renewable Systems, ICEARS 2023*. pp. 1617–1621, 2023, doi: 10.1109/ICEARS56392.2023.10085546.

[11] M. Khushi, "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021, doi: 10.1109/ACCESS.2021.3102399.

[12] K. A. N. Pelliza, "Analysis of the efficiency of the adaptive canny method for the detection of icebergs at open sea," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 3, pp. 459–464, 2020,

doi: 10.5194/isprs-archives-XLII-3-W12-2020-459-2020.

[13] 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.

[14] K. Benhamza, “Canny edge detector improvement using an intelligent ants routing,” *Evol. Syst.*, vol. 12, no. 2, pp. 397–406, 2021, doi: 10.1007/s12530-019-09299-0.

[15] 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.

[16] J. Kuzmic, “Comparison between Filtered Canny Edge Detector and Convolutional Neural Network for Real Time Lane Detection in a Unity 3D Simulator,” *Int. Conf. Internet Things, Big Data Secur. IoTBDS - Proc.*, vol. 2021, pp. 148–155, 2021, doi: 10.5220/0010383701480155.

[17] 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.

[18] 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.

[19] B. P. Sari, “Classification System for Cervical Cell Images based on Hu Moment Invariants Methods and Support Vector Machine,” *2021 Int. Conf. Intell. Technol. CONIT 2021*, 2021, doi: 10.1109/CONIT51480.2021.9498353.

[20] A. I. Maghsoodi, “A machine learning driven multiple criteria decision analysis using LS-SVM feature elimination: Sustainability performance assessment with incomplete data,” *Eng. Appl. Artif. Intell.*, vol. 119, 2023, doi: 10.1016/j.engappai.2022.105785.

[21] T. Cuong-Le, “An efficient approach for damage identification based on improved machine learning using PSO-SVM,” *Eng. Comput.*, vol. 38, no. 4, pp. 3069–3084, 2022, doi: 10.1007/s00366-021-01299-6.

[22] X. Tian, “Predicting non-uniform indoor air quality distribution by using pulsating air supply and SVM model,” *Build. Environ.*, vol. 219, 2022, doi: 10.1016/j.buildenv.2022.109171.

[23] P. Manoharan, “SVM-based generative adverserial networks for federated learning and edge computing attack model and outpoising,” *Expert Syst.*, 2022, doi: 10.1111/exsy.13072.

[24] K. Nidhul, “Enhanced thermo-hydraulic performance in a V-ribbed triangular duct solar air heater: CFD and exergy analysis,” *Energy*, vol. 200, 2020, doi: 10.1016/j.energy.2020.117448.

[25] A. Das, “Assessment of peri-urban wetland ecological degradation through importance-performance analysis (IPA): A study on Chatra Wetland, India,” *Ecol. Indic.*, vol. 114, 2020, doi: 10.1016/j.ecolind.2020.106274.

[26] 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, doi: 10.56705/ijaimi.v1i2.99.

[27] A. Naswin and A. P. Wibowo, “Performance Analysis of the Decision Tree Classification Algorithm on the Pneumonia Dataset,” *… Artif. Intell. Med. …*, 2023, doi: 10.56705/ijaimi.v1i1.83.