Indonesian Journal of Data and Science

Volume 4 Issue 3
ISSN 2715-9936
https://doi.org/10.56705/ijodas. v4i3.114

Research Article

Rice Leaf Disease Classification with Machine Learning: An Approach Using Nu-SVM

Rudi Setiawan [1,*], Hamada Zein [2], Rexania Agramanisti Azdy [3], Sulistyowati [4]

[1] Universitas Trilogi, Jakarta, Indonesia, rudi@trilogi.ac.id
[2] Universitas Muahammadiyah Kalimantan Timur, Samarinda, Indonesia, hz831@umkt.ac.id
[3] Universitas Bina Darma, Palembang, Indonesia, rezania.agramanisti.azdy@binadarma.ac.id
[4] STMIK Palangkaraya, Palangkaraya, Indonesia, sulistyowati@stmikplk.ac.id

Correspondence should be addressed to Rudi Setiawan; rudi@trilogi.ac.id

Received 05 November 2023; Accepted 28 November 2023; Published 31 December2023

Abstract:

This study explores the application of machine learning for classifying rice leaf diseases, employing the Nu-Support Vector Machine (Nu-SVM) algorithm, analyzed through a 5-fold cross-validation approach. The research focuses on distinguishing between healthy leaves and those affected by BrownSpot and LeafBlast diseases. The dataset, comprising segmented rice leaf images processed using Sobel edge detection and Hu Moments feature extraction, is utilized to train and test the model. Results indicate a moderate level of accuracy (52.12% to 53.81%) across the cross-validation folds, with precision and recall metrics demonstrating variability and highlighting the challenges in precise disease classification. Despite this, the model maintains a consistent ability to identify true positives. The study contributes to the field of precision agriculture by showcasing the potential and limitations of using machine learning for plant disease diagnosis. It underscores the need for advanced image processing techniques and diverse feature extraction methods to enhance model performance. The findings are pivotal for developing more effective, automated diagnostic tools in agriculture, thereby aiding in better disease management and potentially improving crop yields. This research serves as a foundational step towards integrating machine learning in agricultural disease detection, emphasizing its importance in sustainable farming practices.

Keywords: Machine Learning, Nu-SVM, Rice Leaf Diseases, Image Processing, Precision Agriculture.
Dataset link: https://www.kaggle.com/datasets/shayanriyaz/riceleafs/

1. Introduction

In the realm of agricultural science, rice stands as a crucial crop, feeding a significant portion of the global population. However, the cultivation of this vital grain faces numerous challenges, particularly from various plant diseases that can drastically reduce yield and quality. Among these, diseases affecting rice leaves, such as BrownSpot and LeafBlast, are of paramount concern. These diseases not only impair the health of the plants but also pose a significant threat to food security, especially in regions heavily reliant on rice as a staple food. In this context, the early and accurate diagnosis of rice leaf diseases is essential for effective disease management and mitigation.

Despite the critical nature of this issue, traditional methods for diagnosing plant diseases often rely heavily on manual observation by experts, a process that is time-consuming, labour-intensive, and subject to human error. The advent of machine learning in agricultural applications presents an opportunity to revolutionize this traditional approach. By leveraging advanced computational techniques, it is possible to automate the process of disease detection, thereby enhancing accuracy and efficiency.

The objective of this research is to explore the potential of machine learning in the classification of rice leaf diseases. Specifically, the study focuses on the application of a Nu-Support Vector Machine (Nu-SVM) algorithm, a variant of the traditional SVM, known for its effectiveness in handling classification problems with a nuanced balance between class separability and data distribution complexity. This research aims to develop a model that can accurately differentiate between healthy rice leaves and those afflicted with BrownSpot or LeafBlast diseases.

Central to this research are several questions: Can machine learning algorithms, particularly Nu-SVM, reliably classify rice leaf diseases? How does the performance of the Nu-SVM model compare in terms of accuracy, precision, recall, and F1-measure [1], [2] against traditional diagnostic methods? Furthermore, the study investigates the effectiveness of image segmentation using the Sobel method [3], [4] and feature extraction via Hu Moments [5], [6] in preparing the data for machine learning analysis.

While the study promises significant contributions to the field of agricultural disease management, it is important to acknowledge its limitations. The research relies on a dataset that, while comprehensive, may not encompass all variants of rice leaf diseases. Additionally, the computational models used are subject to the inherent limitations of machine learning, including the quality and size of the dataset and the potential for overfitting.

Nevertheless, this research aims to contribute significantly to the field of precision agriculture. By demonstrating the efficacy of machine learning in disease classification, it can pave the way for more advanced, automated, and accurate diagnostic tools in agriculture. Such advancements not only hold the potential to improve crop management and yield but also to enhance food security in rice-dependent regions of the world.

2. Method:

The study employs a quantitative research design, focusing on the application and evaluation of machine learning algorithms for image classification. The primary objective is to classify images of rice leaves into three categories: healthy, BrownSpot-infected, and LeafBlast-infected. The performance of the Nu-SVM model is assessed through various metrics such as accuracy, precision, recall, and F1-measure [7], [8]. Our research is designed in five well-structured main stages, and their aspects are illustrated in **Figure 1**.

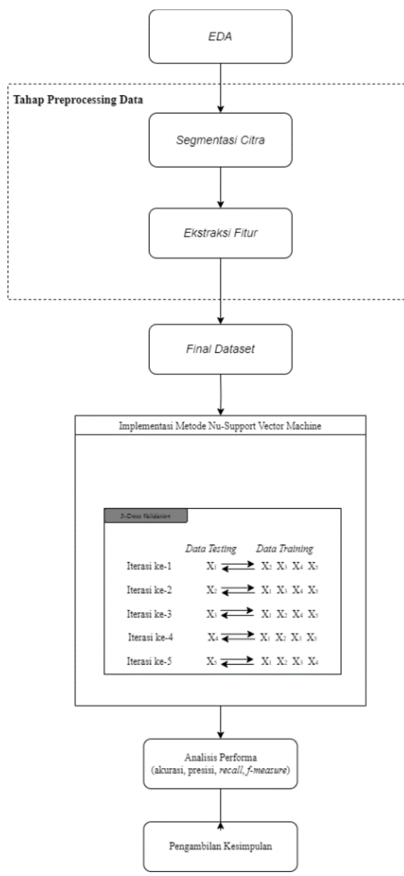

Figure 1. General Research Design Stages

Data Collection Process: **Rice leaf disease**

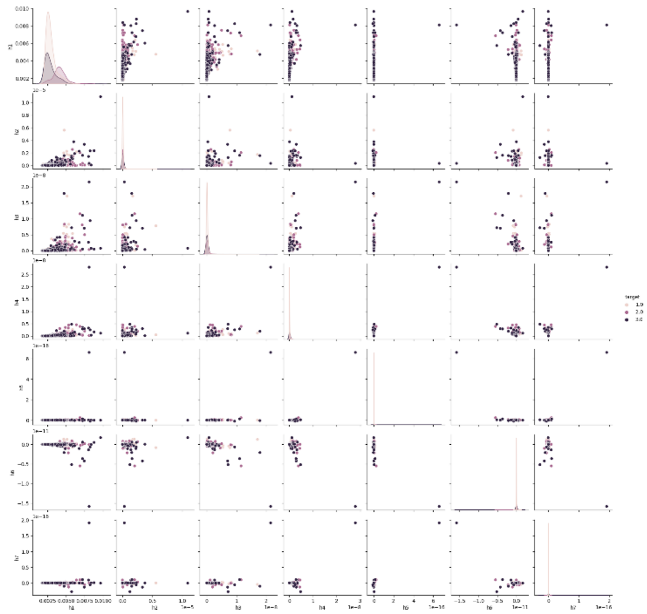

Figure 2. Scatter Plot

The dataset comprises segmented images of rice leaves, categorized into three classes: Healthy (Class 1), BrownSpot (Class 2), and LeafBlast (Class 3). These images have been pre-processed using Sobel edge detection for segmentation and Hu Moments for feature extraction. The dataset's diversity and representativeness ensure a comprehensive analysis across different disease types. The dataset was sourced from kaggle repository, ensuring a diverse range of images representing various stages of the three rice leaf conditions. Each image underwent pre-processing through Sobel segmentation and Hu Moments feature extraction before being inputted into the machine learning model.

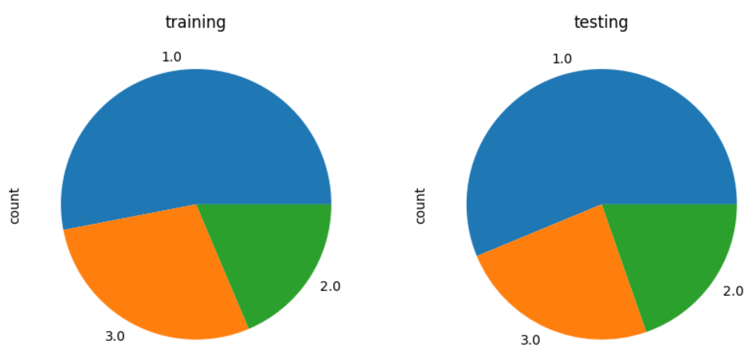

Figure 3. Splitting Dataset 10 % testing, 90% training

Image Segmentation: Sobel

The Sobel operator is used for edge detection, highlighting the contours of rice leaves [9]–[11]. It works by convolving the image with a pair of 3×3 kernels, one estimating the gradient in the x-direction (horizontal) and the other in the y-direction (vertical). The gradients G_x and G_y for each pixel are combined to give the overall gradient G at that point [12]–[14].

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

Where A is the image matrix, G_x and G_y are the horizontal and vertical gradients, respectively. The overall gradient magnitude for each pixel is then computed as **Figure 4** to **6**:

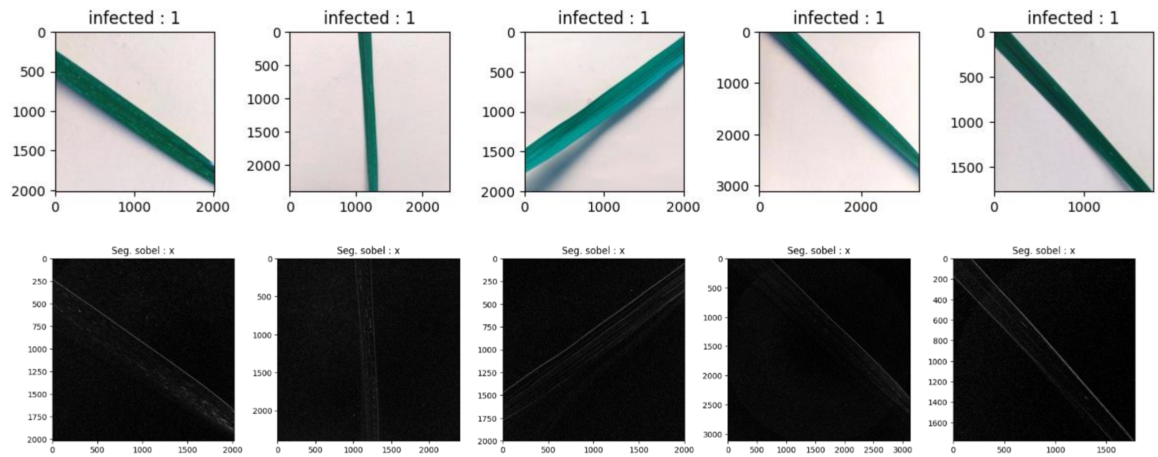

Figure 4. Sobel Detection Results for Healthy Class

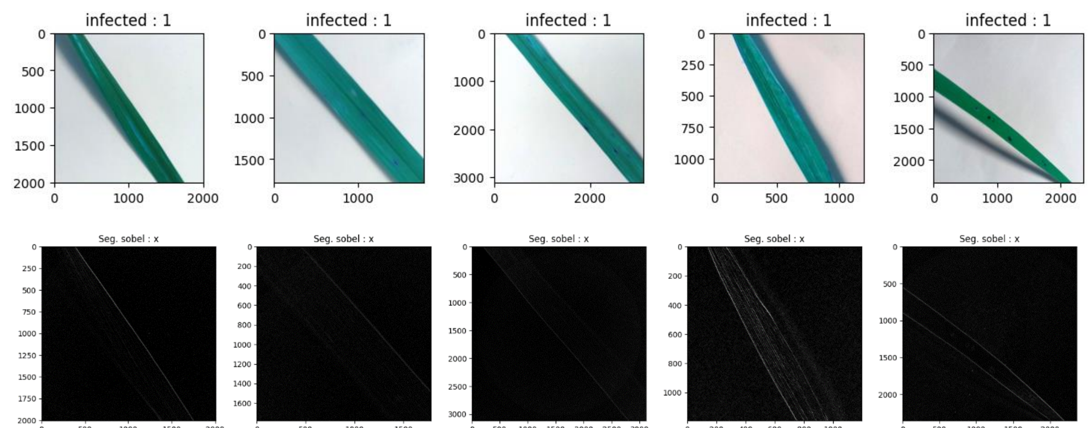

Figure 5. Sobel Detection Results for BrownSpot Class

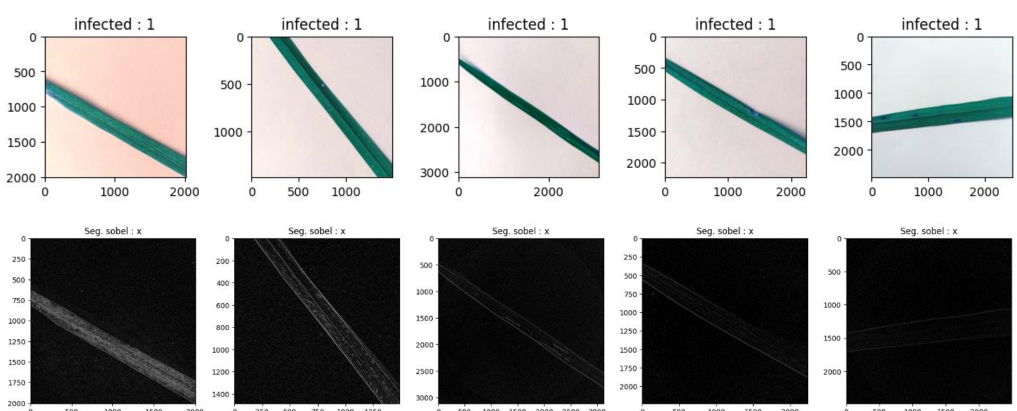

Figure 6. Sobel Detection Results for LeafBlast Class

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 [5], [6], [15]. The n^{th} order central moment is defined as:

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

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

The Hu moments are derived from these central moments as follows [16]–[18]:

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

Classification Algorithm: Nu-Super Vector machine (Nu-SVM)

Nu-SVM [19]–[22], a variant of SVM, is used for classification. It includes a parameter 'nu' which controls the number of support vectors and the margin errors. The optimization problem for Nu-SVM is formulated as [23]–[26].

$$min_{w,b,\varepsilon} \frac{1}{2} W^T W + C \sum\nolimits_{i-1}^{n} \varepsilon_i \tag{4}$$

Subject to $y_i\ (W^T \emptyset(x_i) + b \geq 1 - \varepsilon_i, \varepsilon_i \geq 0$. Where w is the weight vector, b is the bias, ε are slack variables, C is the penalty parameter, $\emptyset$ is the kernel function, and y_i are the target labels.

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 [27]–[29]. 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.

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

Performance Comparison Analysis

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

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

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 research aimed at classifying rice leaf diseases using the Nu-SVM algorithm, and the results have been insightful. The model's performance was evaluated through a 5-fold cross-validation technique, ensuring a robust assessment. The accuracy across the five folds ranged from 52.12% to 53.81%, indicating a moderate level of predictive capability. However, a closer look at precision, recall, and F1-scores, which provide a more nuanced view of the model's performance, revealed some variability. Precision values fluctuated significantly across the folds, with the highest being 51.95% in K-1 and the lowest at 37.66% in K-3. Recall, mirroring the accuracy, remained relatively consistent, suggesting that the model's ability to identify true positives was stable across different data subsets. The F1-Score, a balance between precision and recall, hovered around 39% to 43%, underscoring the challenges in achieving high precision without sacrificing recall. 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	52%	52%	52%	43%
K-2	53%	46%	53%	43%
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%

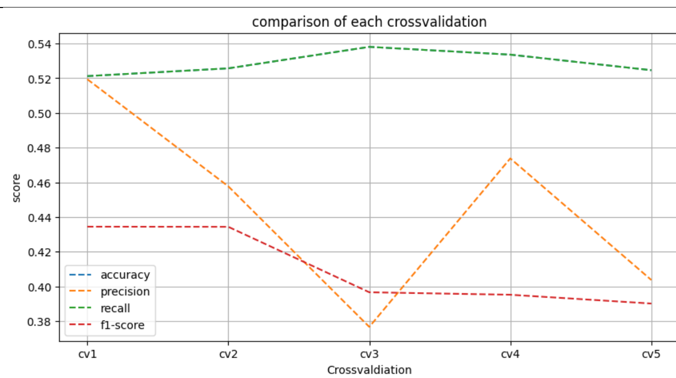

Figure 7. Visualisation Performance Metrics Across 5-Fold Cross-Validation for the Decision Tree

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.

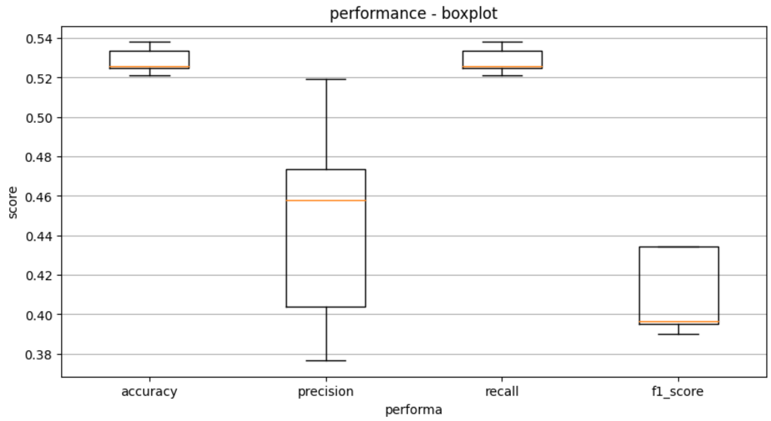

Figure 8. Boxplot Performance Metrics Across 5-Fold Cross-Validation for the Decision Tree

The boxplot presented in **Figure 8** illustrates the variability and distribution of performance metrics—accuracy, precision, recall, and F1-score—derived from the 5-fold cross-validation of the Nu-SVM algorithm applied in rice leaf disease classification. Each boxplot encapsulates the interquartile range, median, and outliers for the respective performance scores across the cross-validation folds, providing a visual summary of the model's classification consistency and areas of potential improvement.

Discussion

Interpreting these results, it's evident that while the Nu-SVM model demonstrates a reasonable ability to classify rice leaf diseases, there is room for improvement, especially in precision. The moderate accuracy indicates that the model, with its current feature set and parameters, captures significant patterns in the data but also struggles with some intricacies of the classification task. This variation in precision across folds might be attributed to the inherent complexities in the dataset, such as similarities between disease symptoms or variations in image quality.

Comparing these findings with existing literature, it's clear that machine learning offers a promising avenue for agricultural disease classification, yet the challenge lies in fine-tuning models for higher precision without compromising on recall. The results align with previous studies that emphasize the difficulty in achieving high accuracy in image-based disease classification due to the subtle and varied nature of disease symptoms.

The practical implications of these results are significant for precision agriculture. The ability to accurately identify disease types can lead to more targeted and effective treatment strategies, potentially saving significant resources and improving yields. However, the limitations of the current research must be acknowledged. The variability in precision across folds suggests that the model may be sensitive to the specific data it's trained on, and the overall moderate performance highlights the need for more sophisticated or diverse feature extraction techniques.

Future research should focus on exploring different image processing and feature extraction methods to enhance model accuracy and precision. Additionally, incorporating larger and more diverse datasets could improve the model's robustness and generalizability. Further studies might also explore hybrid models or deep learning techniques, which have shown promise in other image classification tasks. The ultimate goal is to develop a highly reliable and efficient tool for farmers and agricultural specialists to diagnose and manage rice leaf diseases effectively.

4. Conclusion

In summary, this research on the classification of rice leaf diseases using the Nu-SVM algorithm has yielded important insights, although it has also highlighted certain challenges. The model demonstrated moderate accuracy, with values ranging between 52.12% and 53.81% across different folds in the 5-fold cross-validation process. However, precision varied considerably, suggesting a need for further refinement in the model's ability to precisely classify the diseases. The consistency in recall across folds indicated a stable detection of true positives, but the F1-scores suggested a balance between precision and recall needs improvement. These findings answer our primary research question, affirming that while machine learning, particularly the Nu-SVM algorithm, can be effective in classifying rice leaf diseases, achieving high precision and accuracy remains a complex task. The study contributes to the growing body of knowledge in agricultural technology, specifically in the application of machine learning for disease detection in crops, highlighting both the potential and the limitations of current methodologies.

For future research, it is recommended to explore advanced image processing techniques and more diverse feature extraction methods to enhance the accuracy and precision of the classification model. The incorporation of larger and more varied datasets could also improve the model's robustness and generalizability. Furthermore, experimenting with deep learning approaches, which have shown promising results in other areas of image classification, might offer new pathways for advancements in this field. In practice, these findings underscore the need for continuous development and integration of machine learning tools in agriculture, aiming to provide more efficient, accurate, and cost-effective solutions for disease management in crops, thereby supporting sustainable agricultural practices and food security.

References:

[1] E. Taşkesen, "Predicting heat transfer performance of FeCu water hybrid nanofluid under constant magnetic field using ANN," *J. Therm. Eng.*, vol. 9, no. 3, pp. 811–822, 2023, doi: 10.18186/thermal.000000.

[2] N. Mamat, "Enhancement of water quality index prediction using support vector machine with sensitivity analysis," *Front. Environ. Sci.*, vol. 10, 2023, doi: 10.3389/fenvs.2022.1061835.

[3] J. N. Archana, "Enhancement of digital chest images using a modified Sobel edge detection algorithm," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, no. 3, pp. 1718–1726, 2021, doi: 10.11591/ijeecs.v24.i3.pp1718-1726.

[4] C. Xiu, "Image Segmentation of CV Model Combined with Sobel Operator," *Proc. 32nd Chinese Control Decis. Conf. CCDC 2020*, pp. 4356–4360, 2020, doi: 10.1109/CCDC49329.2020.9164450.

[5] D. V Kondusov, "Comparison of 3D Models Using Hu Moment Invariants," *Russ. Eng. Res.*, vol. 40, no. 7, pp. 570–574, 2020, doi: 10.3103/S1068798X20070199.

[6] 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.

[7] 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.

[8] A. A. Ewees, "Performance analysis of Chaotic Multi-Verse Harris Hawks Optimization: A case study on solving engineering problems," *Eng. Appl. Artif. Intell.*, vol. 88, 2020, doi: 10.1016/j.engappai.2019.103370.

[9] S. K. Chen, "An Enhanced Adaptive Sobel Edge Detector Based on Improved Genetic Algorithm and Non-Maximum Suppression," *Proceeding - 2021 China Autom. Congr. CAC 2021*, pp. 8029–8034, 2021, doi: 10.1109/CAC53003.2021.9727626.

[10] R. Tian, "Sobel edge detection based on weighted nuclear norm minimization image denoising," *Electron.*, vol. 10, no. 6, pp. 1–15, 2021, doi: 10.3390/electronics10060655.

[11] A. K. Singh, "Optimization of Multi-Class Non-Linear SVM Image Classifier Using A Sobel Operator Based Feature Map and PCA," *3rd Int. Conf. Range Technol. ICORT 2023*, 2023, doi: 10.1109/ICORT56052.2023.10249196.

[12] R. A. A. S, "Comparative Analysis of Eight Direction Sobel Edge Detection Algorithm for Brain Tumor MRI Images," *Procedia Comput. Sci.*, vol. 201, pp. 487–494, 2022, doi: 10.1016/j.procs.2022.03.063.

[13] Y. Harshavardhan, "Comparative analysis of accuracy in identification of bone fracture detection using Prewitt edge detection with Sobel edge detection approach," *AIP Conf. Proc.*, vol. 2822, no. 1, 2023, doi: 10.1063/5.0173412.

[14] K. Hu, "Real-time CNN-based Keypoint Detector with Sobel Filter and Descriptor Trained with Keypoint Candidates," *Proc. SPIE - Int. Soc. Opt. Eng.*, vol. 12701, 2023, doi: 10.1117/12.2679944.

[15] 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.

[16] 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.

[17] 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.

[18] B. S. W. Poetro, E. Maria, H. Zein, and ..., “Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification,” *Indones. J. …*, 2024, doi: 10.56705/ijodas.v5i1.123.

[19] 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.

[20] R. A. Dharmmesta, “Classification of Foot Kicks in Taekwondo Using SVM (Support Vector Machine) and KNN (K-Nearest Neighbors) Algorithms,” *Proceedings of the 2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2022*. pp. 36–41, 2022, doi: 10.1109/IAICT55358.2022.9887475.

[21] Y. Boer, “Classification of Heart Disease: Comparative Analysis using KNN, Random Forest, Gaussian Naive Bayes, XGBoost, SVM, Decision Tree, and Logistic Regression,” *2023 5th International Conference on Cybernetics and Intelligent Systems, ICORIS 2023*. 2023, doi: 10.1109/ICORIS60118.2023.10352195.

[22] S. Hafeez and N. Kathirisetty, “Effects and Comparison of different Data pre-processing techniques and ML and deep learning models for sentiment analysis: SVM, KNN, PCA with SVM and CNN,” *2022 First Int. Conf. …*, 2022, doi: 10.1109/ICAITPR51569.2022.9844192.

[23] J. Sun, “Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting,” *Inf. Fusion*, vol. 54, pp. 128–144, 2020, doi: 10.1016/j.inffus.2019.07.006.

[24] O. M. A. Ali, “Evaluation of Electrocardiogram Signals Classification Using CNN, SVM, and LSTM Algorithm: A review,” *8th IEC 2022 - International Engineering Conference: Towards Engineering Innovations and Sustainability*. pp. 185–191, 2022, doi: 10.1109/IEC54822.2022.9807511.

[25] 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.

[26] A. Dhar, “Comparative Analysis of Deep Learning, SVM, Random Forest, and XGBoost for Email Spam Detection: A Socio- Network Analysis Approach,” *Proceedings - 4th IEEE 2023 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2023*. pp. 701–707, 2023, doi: 10.1109/ICCCIS60361.2023.10425771.

[27] M. Rafało, “Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis,” *ICT Express*, vol. 8, no. 2, pp. 183–188, 2022, doi: 10.1016/j.icte.2021.05.001.

[28] A. T. Huynh, “A machine learning-assisted numerical predictor for compressive strength of geopolymer concrete based on experimental data and sensitivity analysis,” *Appl. Sci.*, vol. 10, no. 21, pp. 1–16, 2020, doi: 10.3390/app10217726.

[29] M. H. D. M. Ribeiro, “Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series,” *Appl. Soft Comput. J.*, vol. 86, 2020, doi: 10.1016/j.asoc.2019.105837.