International Journal of Artificial Intelligence in Medical Issues

Volume 2 Issue 2
ISSN 3025-4167
https://doi.org/10.56705/ijaimi.v2i2.156

Research Article

Optimizing Cardiomegaly Detection: A Random Forest Approach to Processed Chest X-ray Imagery

Nurul Alfiyyah [1*]
[1] Universitas Muslim Indonesia, nurulalfiyyah.labfik@umi.ac.id
Correspondence should be addressed to Nurul Alfiyyah; nurulalfiyyah.labfik@umi.ac.id

Received 30 September 2024; Revised 08 October 2024; Accepted 20 October 2024; Published 30 November 2024

Abstract:
This study explores the application of a Random Forest Classifier for the automated detection of Cardiomegaly from chest X-ray images, utilizing a dataset processed and derived from the NIH Chest X-ray Dataset. Given the crucial need for accurate and timely diagnosis of Cardiomegaly to inform appropriate treatment decisions, this research aims to determine the efficacy of machine learning models in augmenting diagnostic processes. Employing image pre-processing techniques such as Sobel filtering for edge detection and Hu Moments for feature extraction, the study enhances the input features for the model. The performance of the classifier was evaluated using a 5-fold cross-validation approach, yielding results with average accuracy, precision, recall, and F1-scores ranging approximately between 52% and 54%. These findings suggest a moderate level of reliability and consistency, indicating the potential utility of ensemble machine learning methods in medical imaging analysis. However, the variability in performance across different data subsets highlights the challenges and necessitates further optimization. This research contributes to the ongoing discourse on integrating machine learning into clinical settings, demonstrating the potential benefits and current limitations. Future research is recommended to expand the dataset variety, integrate advanced deep learning methodologies, and rigorously test these models in clinical environments. The findings hold significant implications for the development of automated diagnostic tools in healthcare, potentially leading to enhanced diagnostic accuracy and efficiency.

Keywords: Random Forest Classifier, Image Pre-processing, Ensemble Learning Methods, Healthcare Diagnostics.

Dataset link: https://www.kaggle.com/datasets/rahimanshu/cardiomegaly-disease-prediction-using-cnn

1. Introduction

Cardiomegaly, or an enlarged heart, is a medical condition often indicative of underlying heart disease or stress, and can lead to congestive heart failure if not diagnosed and managed effectively. The condition is typically identified in clinical settings through imaging techniques such as chest X-rays. However, traditional methods of detection rely on visual inspections and measurements made by radiologists, which are subject to variability and potential human error. The significance of achieving accurate and timely diagnosis cannot be overstated, as it directly influences treatment decisions and patient outcomes.

The primary problem addressed by this research is the need for improved diagnostic accuracy in the detection of Cardiomegaly from chest X-ray images. Manual methods not only vary due to individual practitioner skill but are also time-consuming. Consequently, there is a compelling need for automated systems that enhance both the speed and precision of diagnostic processes. Such systems can help standardize Cardiomegaly assessments and potentially reduce the rates of misdiagnosis.

The objective of this study is to develop an automated detection system for Cardiomegaly using a machine learning approach, specifically through the implementation of a Random Forest Classifier [1]. By utilizing advanced image pre-processing techniques and ensemble learning methods, this study aims to provide a robust model capable of assisting radiologists in diagnosing Cardiomegaly more accurately and efficiently.

Several research questions guide this study: How effectively can a Random Forest Classifier [2]–[4], trained on pre-processed images, detect Cardiomegaly compared to traditional diagnostic methods? What are the optimal pre-processing techniques that enhance the model's accuracy? These questions aim to explore the viability of machine learning algorithms in medical image analysis and their potential to supplement existing diagnostic procedures.

This research is bounded by certain limitations, primarily the scope of the dataset used, which is derived from a larger collection of chest X-rays provided by the NIH Clinical Center. The images have been processed and resized to a uniform dimension, potentially omitting subtle nuances crucial for diagnosis in a real-world scenario. Furthermore, the study focuses exclusively on images processed with Sobel filtering and Hu Moments for feature extraction, which may not encapsulate all features relevant to Cardiomegaly.

The contributions of this research are twofold. Firstly, it advances the application of machine learning in medical imaging, particularly using ensemble methods for the detection of significant heart conditions. Secondly, it seeks to establish a benchmark for the performance of automated detection systems in identifying Cardiomegaly, offering insights into the potential integration of such technologies in clinical settings, thus paving the way for future innovations in medical diagnostics [5]–[7]. Through rigorous validation methods and comprehensive performance metrics, this study aims to substantiate the efficacy of automated systems in enhancing diagnostic accuracy for Cardiomegaly in chest X-rays.

2. Method

The study employs a quantitative research design using a supervised machine learning approach to develop a classification model capable of identifying Cardiomegaly from chest X-ray images. The effectiveness of the model is evaluated through a series of experiments involving pre-processing, feature extraction [8], model training [9], and validation using cross-validation techniques [10]–[12]. **Figure 1** presents a graphical representation of the complete research workflow.

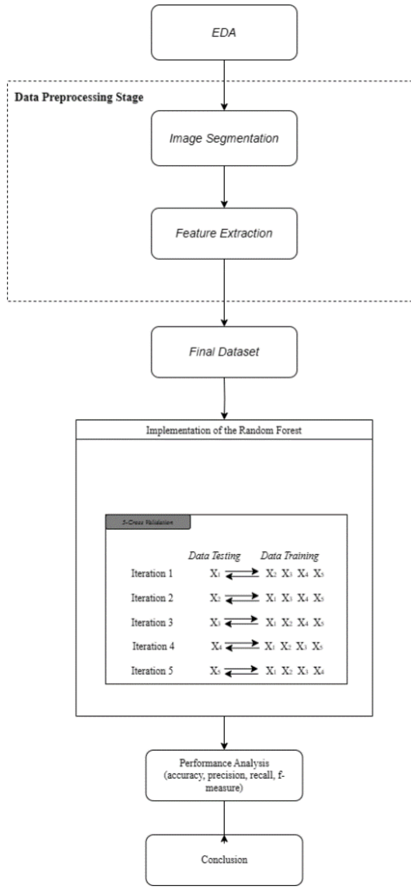

Figure 1: Procedure for Assessing a Random Forest

Sample or Data Selection:

The dataset consists of chest X-ray images labelled as either 'Positive' or 'Negative' for Cardiomegaly. These images were initially sourced from the publicly available NIH Chest X-ray Dataset and further processed to select only those pertinent to the study. The final dataset includes equal proportions of positive and negative cases, ensuring balanced class distribution which is critical for training unbiased machine learning models.

Tools and Technology Used:

The study utilizes Python programming language with several libraries:

- NumPy and Pandas for data manipulation,
- OpenCV for image pre-processing,
- Scikit-learn for implementing the Random Forest algorithm and conducting the machine learning analysis,
- Matplotlib and Seaborn for data visualization.

Data Collection Process

Images were resized to 128 × 128 pixels to normalize input size and processed using two key techniques:

a. Sobel Operator: Used for edge detection in images, enhancing the structural edges by calculating the gradient of image intensity at each pixel [13]–[15].

In **Figures 2** and **3** the results of image segmentation using sobel features on the dataset are shown.

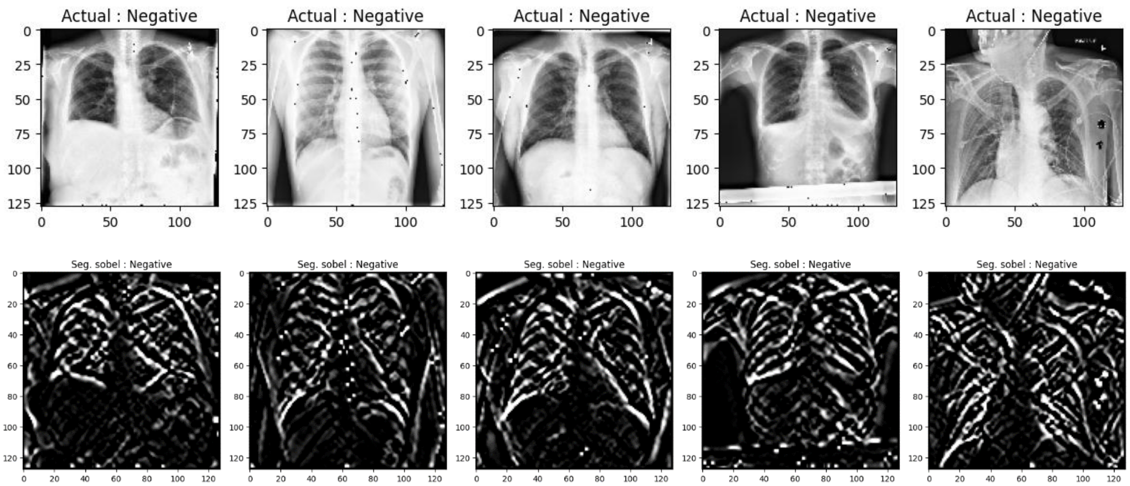

Figure 2: Sobel Results for Negative Class

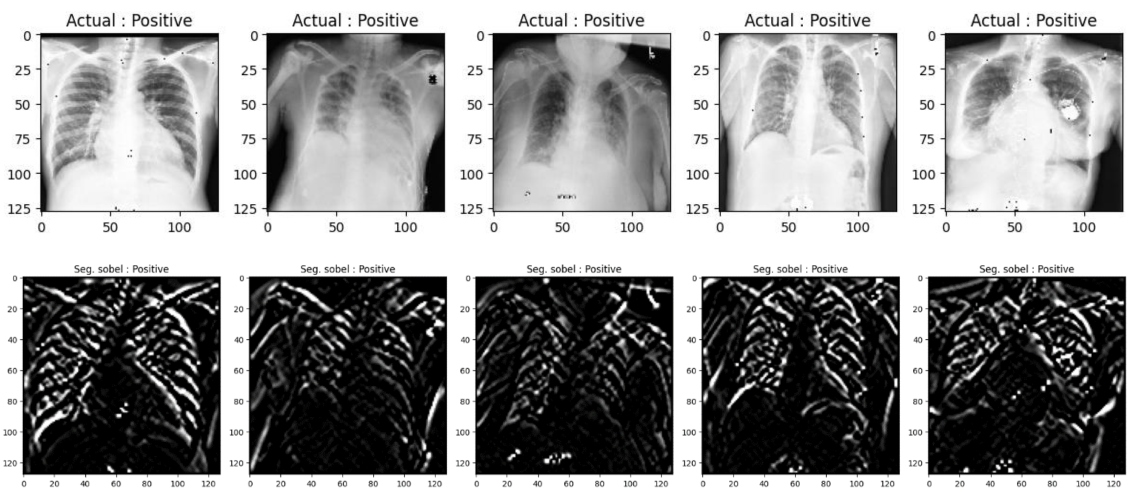

Figure 3: Sobel Results for Positive Class

b. Hu Moments: Employed to extract shape-based features from images, which are invariant to image transformations and provide a robust basis for classification.

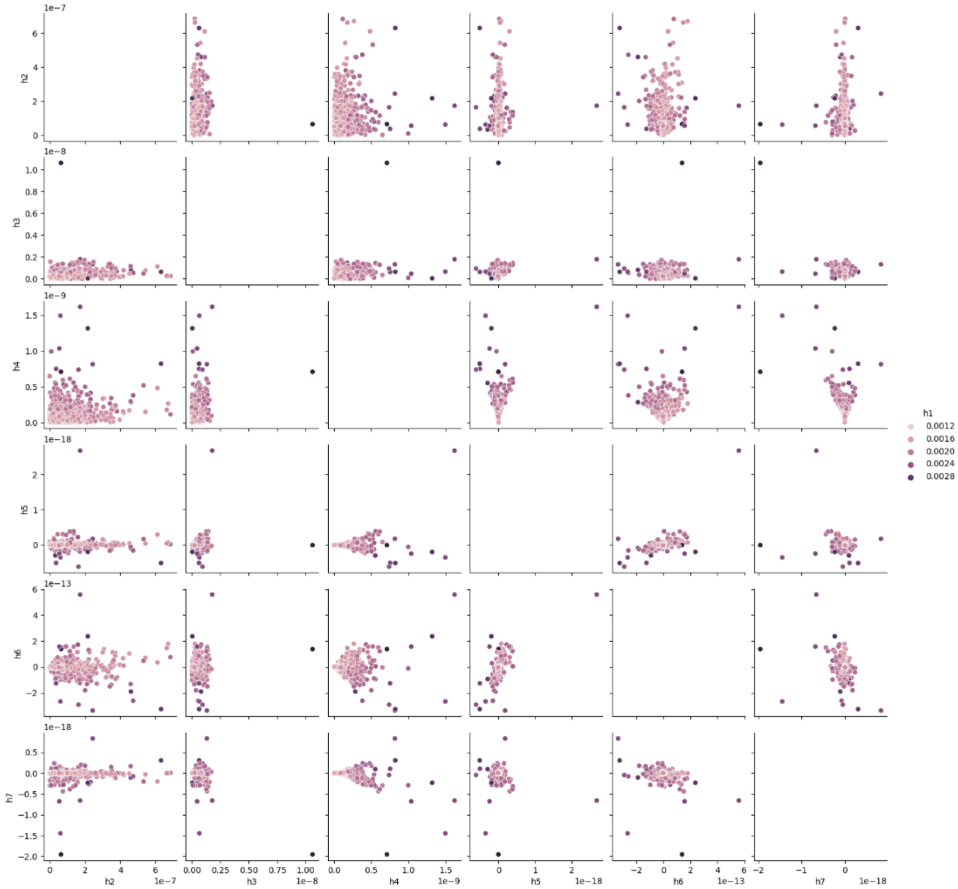

Figure 4: Scatter Plot Visualization of Extracted Hu Moments Features

Data Analysis Method

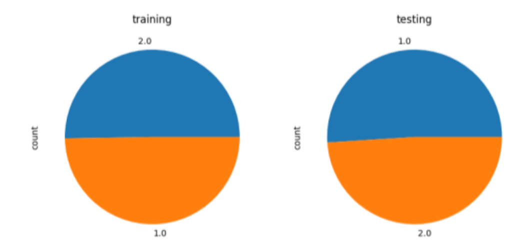

Figure 5: Spilitting Data

The dataset was split into training and testing sets with an 80/20 ratio using stratified sampling to maintain equivalent proportions of each class in both sets. The Random Forest Classifier [16] was trained on the training set. Performance evaluation was conducted using 5-fold cross-validation [17], [18] to assess the model's robustness and generalizability. The performance of the model was evaluated using accuracy, precision, recall, and F1-measure, calculated as follows [19]–[24]:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (4)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3. Result and Discussion

The Random Forest Classifier was evaluated using a 5-fold cross-validation to ascertain its effectiveness in diagnosing Cardiomegaly from chest X-ray images. The performance metrics, presented below, show a mix of variability across the different folds, which is indicative of the model's general behaviour under varied data subsets. **Table 1**.

Table 1: Performance Metrics Across 5-Fold Cross-Validation for the Random Forest Classifier

K-n	**Metrics**			
	Accuracy	***Precision***	***Recall***	***F-Measure***
K-1	53.49%	53.5%	53.49%	53.47%
K-2	50.45%	50.48%	50.45%	49.76%
K-3	51.35%	51.35%	51.35%	51.35%
K-4	53.89%	54.49%	53.89%	52.23%
K-5	52.76%	52.9%	52.76%	52.14%
$\sum \boldsymbol{Avg}$	52.39%	52.54%	52.39%	51.79%

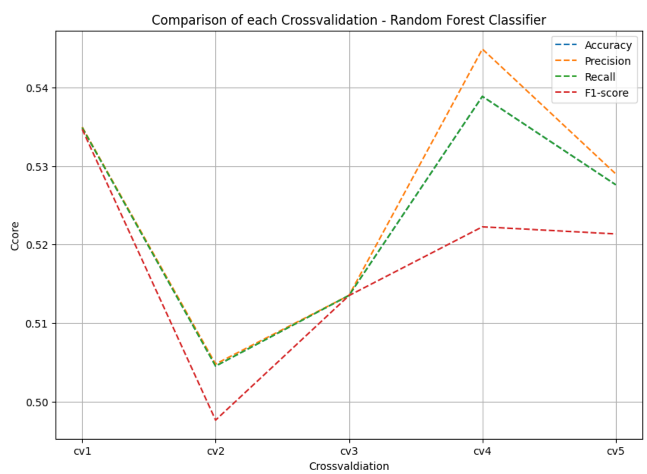

Figure 6: Performance Metrics Across 5-Fold Cross-Validation for the Random Forest Classifier

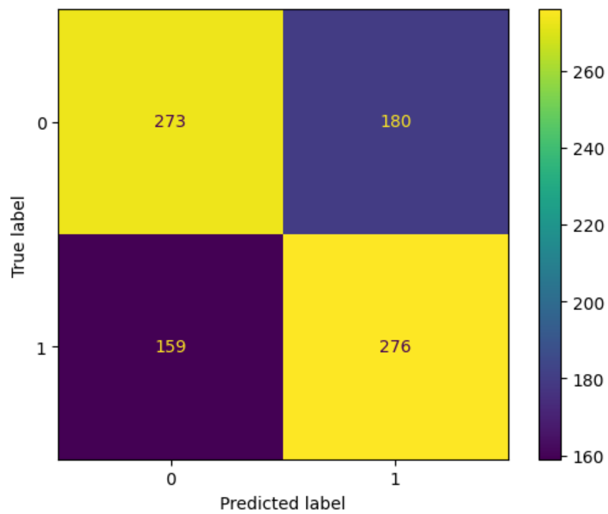

Figure 7: Confusion Matrix

Figure 7 The accompanying confusion matrix visualizes the true and predicted classifications of the model, showing a total of 273 true negatives, 180 false positives, 159 false negatives, and 276 true positives. This indicates a balanced sensitivity and specificity across the dataset. A series of graphs further elucidate the performance metrics across the folds, with the line graph depicting the fluctuations in accuracy, precision, recall, and F1-score. A boxplot illustrates the spread and central tendency of these metrics, providing a visual summary of the model's consistency.

Discussion

The interpretation of the results reveals that the Random Forest Classifier, while generally consistent, displays moderate variability in performance across different subsets of the data, which could be attributed to the intrinsic heterogeneity of medical imaging data. The classifier's precision generally aligns with its accuracy, suggesting that when it predicts a case as positive, it is fairly reliable. However, the recall and F1-score variability highlight potential areas for improvement, particularly in minimizing false negatives, which are critical in medical diagnostics.

These findings align with previous research suggesting that ensemble methods like Random Forest can effectively handle complex patterns in data but may require careful tuning of parameters and feature selection to optimize performance for specific applications such as medical image analysis. The practical implications of this research are significant, offering a potential tool for assisting radiologists in diagnosing Cardiomegaly more efficiently and with greater accuracy, thereby facilitating earlier and more effective treatment interventions.

However, the study is not without limitations. The reliance on processed and resized images may result in the loss of critical diagnostic information, potentially affecting the classifier's ability to generalize to new, unprocessed images. Furthermore, the balanced dataset used for training and testing might not perfectly represent the true distribution of Cardiomegaly in the general population, possibly skewing the model's predictive power in real-world scenarios.

Future research should focus on expanding the dataset to include a wider variety of X-ray images, exploring more sophisticated image processing methods that preserve critical diagnostic features, and testing the model across unbalanced datasets to better simulate real clinical environments. Further, integrating newer deep learning approaches could potentially enhance the model's ability to learn more complex features without manual feature extraction, driving improvements in both accuracy and reliability.

4. Conclusion

The investigation into the use of a Random Forest Classifier for diagnosing Cardiomegaly from chest X-ray images demonstrates moderate success, with the model achieving average accuracy, precision, recall, and F1-scores around 52% to 54% across a 5-fold cross-validation. The results reveal that while the model is somewhat reliable in its predictions, there is noticeable variability across different data subsets, highlighting the challenge of using machine learning in medical image analysis where data heterogeneity is prevalent. The analysis confirms that a well-tuned ensemble method like Random Forest can manage complex patterns in medical imaging data but requires careful consideration of feature selection and model parameters to optimize performance.

This research substantiates the hypothesis that machine learning models can aid in the automated diagnosis of Cardiomegaly, providing a foundation for future technological integration into clinical diagnostics. The study contributes to the growing body of knowledge demonstrating the viability of machine learning in enhancing diagnostic accuracy and efficiency in healthcare settings. For future endeavors, it is recommended to expand the dataset to include a broader range of image types and conditions, explore advanced deep learning techniques that might obviate the need for manual feature extraction, and assess the model's performance in real-world clinical settings to better gauge its practical effectiveness and reliability. Further research should also consider the impact of dataset imbalances and strive to develop models that are robust to such challenges, potentially increasing their applicability and utility in diverse medical environments.

References:

[1] S. Khomsah and E. Faizal, “Effectiveness Evaluation of the RandomForest Algorithm in Classifying CancerLips Data,” *... Artif. Intell. Med. ...*, 2023.

[2] H. Nhat-Duc, “Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification,” *Autom. Constr.*, vol. 148, 2023, doi: 10.1016/j.autcon.2023.104767.

[3] L. Zhang, “Prediction of prognosis in elderly patients with sepsis based on machine learning (random survival forest),” *BMC Emerg. Med.*, vol. 22, no. 1, 2022, doi: 10.1186/s12873-022-00582-z.

[4] Y. Zhao, “Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period,” *Ecol. Indic.*, vol. 135, 2022, doi: 10.1016/j.ecolind.2021.108529.

[5] R. A. Azdy, R. F. Syam, E. Faizal, and ..., “Performance Evaluation of Bagging Meta-Estimator in Lung Disease Detection: A Case Study on Imbalanced Dataset,” *Int. J. ...*, 2023.

[6] 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.

[7] 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.

[8] 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.

[9] R. A. Disha, "Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique," *Cybersecurity*, vol. 5, no. 1, 2022, doi: 10.1186/s42400-021-00103-8.

[10] 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.

[11] J. Zhang, "Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting," *Comput. Electron. Agric.*, vol. 173, 2020, doi: 10.1016/j.compag.2020.105384.

[12] T. A. Reist, "Cross validation of aerodynamic shape optimization methodologies for aircraft wing-body optimization," *AIAA J.*, vol. 58, no. 6, pp. 2581–2595, 2020, doi: 10.2514/1.J059091.

[13] D. R. D. Varma, "Performance Monitoring of Novel Iris Detection System using Sobel Algorithm in Comparison with Canny Algorithm by Minimizing the Mean Square Error," *Proc. 3rd Int. Conf. Intell. Eng. Manag. ICIEM 2022*, pp. 509–512, 2022, doi: 10.1109/ICIEM54221.2022.9853127.

[14] 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.

[15] 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.

[16] X. Yu, "Random forest algorithm-based classification model of pesticide aquatic toxicity to fishes," *Aquat. Toxicol.*, vol. 251, 2022, doi: 10.1016/j.aquatox.2022.106265.

[17] S. Ortiz-Toquero, "Classification of Keratoconus Based on Anterior Corneal High-order Aberrations: A Cross-validation Study," *Optom. Vis. Sci.*, vol. 97, no. 3, pp. 169–177, 2020, doi: 10.1097/OPX.0000000000001489.

[18] K. M. Bain, "Cross-validation of three Advanced Clinical Solutions performance validity tests: Examining combinations of measures to maximize classification of invalid performance," *Appl. Neuropsychol.*, vol. 28, no. 1, pp. 24–34, 2021, doi: 10.1080/23279095.2019.1585352.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] N. A’ayunnisa, Y. Salim, and H. Azis, “Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab,” *… J. Data Sci.*, 2022.

[24] 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.