

## Identitas Paper

<b>Paper ID</b>	17220
<b>Judul Paper</b>	Optimization of Breast Cancer Classification using Faster R-CNN
<b>Authors</b>	SUNARDI, Anton Yudhana, Anggi Rizky Windra Putri*
<b>Jurnal</b>	Revue d'Intelligence Artificielle (RIA)
<b>Penerbit</b>	Lavoisier
<b>Negara</b>	France
<b>Quartile dan SJR</b>	Scopus Q3, SJR 0.3 <a href="https://www.scimagojr.com/journalsearch.php?q=24847&amp;tip=sid&amp;clean=0">https://www.scimagojr.com/journalsearch.php?q=24847&amp;tip=sid&amp;clean=0</a>
<b>Terbit</b>	Vol 37 No 1 February 2023 pp 39-45
<b>Link Paper</b>	<a href="https://www.iieta.org/journals/ria/paper/10.18280/ria.370106">https://www.iieta.org/journals/ria/paper/10.18280/ria.370106</a>



## Proses Pengelolaan Paper

Tahap	Tanggal	Proses/Aktivitas
1	27 September 2022	Submission
2	31 October 2022	Revision required
3	03 November 2022	Submit the revision
4	08 November 2022	Accepted
5	08, 13 January 2023 09, 28 February 2023 01 March 2023	Editing process
6	25 March 2023	Published

			Sep-22				Oct-22				Nov-22				Dec-22				Jan-23				Feb-23				Mar-23			
	Quartile	SJR	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
RIA	Q3	0.30				27				31	3	8							8	13		9	28	1					25	

## Tahap 1 Submission, 27 September 2022

RIA 17220 Important suggestions on your submission "Optimization of Breast Cancer Classification using Faster R-CNN"

Dear Authors,

Some of our partners have jointly founded a new publishing house named **Acadlore**, aiming to provide young researchers and those from emerging countries with a platform for sharing and exchanging ideas. The official website of the publishing house is now accessible at <https://www.acadlore.com>. Acadlore has launched multiple peer-reviewed open-access journals, many of which are calling for high-quality submissions.

We noticed that your manuscript titled *Optimization of Breast Cancer Classification using Faster R-CNN* is perfectly in line with the scope of one of Acadlore journals, **Acadlore Transactions on AI and Machine Learning (ATAIML)**. We'd like to know if you are willing to transfer this article to that journal.

If you agree with the article transfer, you can either

- (1) have your next two submissions to Scopus-indexed IIETA journals published free of charge; or
- (2) receive a monetary reward of 200 US dollars.

Please tell us which option you'd like to choose.

Acadlore will handle your current article quickly, without charging any fee. Besides, our partners are confident that your article will be frequently cited by many scholars in various Scopus / SCIE-indexed journals, and become Scopus-indexed once the Acadlore journal gets indexed in Scopus in the coming two or three years.

## Tahap 2 Revision required, 31 October 2022

status of the article

Dear author,

We have reached a decision regarding your submission to Revue d'Intelligence Artificielle, "Optimization of Breast Cancer Classification using Faster R-CNN".

Our decision is: Revisions Required

1. Revise the current paper according to reviewers' comments. **Highlight any change** or track changes you make. **Response to reviewers is also required.**
2. Typeset your final paper according to the attached template.

Please return your revised manuscript and the response to reviewers to this e-mail before **3 November 2022**. Thanks for your cooperation.

**Editorial Board**  
Revue d'Intelligence Artificielle  
<http://www.iieta.org/Journals/RIA>  
published by  
International Information and Engineering Technology Association (IIETA)  
<http://www.iieta.org/>

### Review Comments

#### Optimization of Breast Cancer Classification using Faster R-CNN

Overall, this manuscript reads well and has the potential to provide us some useful implications for future direction. However, there are some important issues that the authors have to address in order to make this manuscript better. My major comments to this manuscript are as follows:

1. "Modern lifestyle factors that are instantaneous and a lack of knowledge about healthy lifestyles make women vulnerable to breast cancer." The importance to cure breast cancer should be emphasized in the introduction.
2. "Previous studies have detected breast cancer using multiparametric magnetic resonance methods and histogram analysis." The author lists a number of previous studies. However, he or she does not summarize the advantages or shortcomings or the space to improve.
3. "This study proposes an approach using the Faster R-CNN deep learning model." The research aim is very clear, but the research gap has not been clearly defined. Why does the author want to conduct this study? The research motivation is not sufficient.
4. The findings or research results should be introduced briefly in the introduction part, so as to readers can get a whole view of this paper when they just finish reading the introduction.
5. "This classification speeds up the detection and treatment of cancer by specialists" After this, it is suggested to add the whole structure or the paper's layout at the end of the introduction.
6. "This research uses the Faster-RCNN deep learning model. The application of the model is to detect, determine, and classify breast cancer image data." Before part 2, it is suggested to supplement "related work" or "literature review" part so as to summarize the previous studies in detail and clarify the position of this study.
7. Good research needs to explain what it means in practice, so the managerial implications should be stressed in this study.
8. "The deep learning algorithm of the Faster R-CNN model has a Regional Proposal Network structure to generate target areas and make target positioning very precise." The conclusion part should be enlarged. The contributions of this study should be strengthened in the conclusion. What are the differences between this study and previous studies? In what ways does this article enrich the study of this stream?

**Tahap 3**  
**Submit the revision, 08 November 2022**

ANGGI PUTRI <anggi2007048019@webmail.uad.ac.id>  
 kepada editor.ria ▾

Kam, 3 Nov 2022, 07:30 ☆ ↶ ⋮

Dear Admin editor RIA, IIETA

Thanks for replying to my email and reviewing my journal, comments from reviewers are very helpful in improving the writing of the journal. Here I attach the revised results based on comments from reviewers.

Thanks for attention and cooperation.

Best Regards

Anggi Rizky Windra Putri,S.Kom

...

2 Lampiran • Dipindai dengan Gmail



NO	REVISION POINT	BEFORE REVISION	AFTER REVISION	LOCATION
1	“Modern lifestyle factors that are instantaneous and a lack of knowledge about healthy lifestyles make women vulnerable to breast cancer.” The importance to cure breast cancer should be emphasized in the introduction.	no addition The importance of curing breast cancer should be emphasized in the introduction	Handling and curing this disease is very important because it can provide opportunities and a higher quality of life, increase life expectancy; eliminate or inhibit the growth of cancer cells; and save women from breast cancer malignancy.	Page 1
2	“Previous studies have detected breast cancer using multiparametric magnetic resonance methods and histogram analysis.” The author lists a number of previous studies. However, he or she does not summarize the advantages or shortcomings or the space to improve.	Previous studies have detected breast cancer using multiparametric magnetic resonance methods and histogram analysis [5]; feature Gray Level Co-occurrence Matrix (GLCM) [7]; machine learning-based comparative analysis [8]; analysis using LBP, HOG, and GLCM. Techniques [9]; Computer-Aided Diagnosis (CAD) system [10].	The following are some of the previous studies that carried out the detection of breast cancer. Xie et al. research the multiparametric magnetic resonance method and histogram analysis. This research aims to Identify breast cancer image biomarkers using multiparametric MR imaging maps and whole-tumor histogram analysis. His research method was retrospective involving 134 patients with invasive ductal carcinoma [8]. Punithavathi and Devakumari researched the extraction of cancer features through experiments with the Gray Level Co-occurrence Matrix (GLCM) feature and the process of feature extraction testing on the Mammographic Image Analysis Society (MIAS) dataset [9].	Page 1
3	“This study proposes an approach using the Faster R-CNN deep learning model.” The research aim is very clear, but the research gap has not been clearly defined. Why does	No addition and affirmation of research motivation	Based on this, this research encourages women to be more aware of the threat of breast cancer, so early detection of breast cancer is essential for proper treatment and a healthy lifestyle	Page 1

	the author want to conduct this study? The research motivation is not sufficient.			
4	The findings or research results should be introduced briefly in the introduction part, so as to readers can get a whole view of this paper when they just finish reading the introduction.	This study proposes an approach using the Faster R-CNN deep learning model. This method studies image data in depth to recognize the unique characteristics of a particular object, thus requiring optimization steps to produce high accuracy values	This study proposes an approach using image segmentation optimization and the Faster R-CNN deep learning model. This method studies image data in depth to recognize the unique characteristics of a particular object, thus requiring optimization steps to produce high accuracy values [5]. This research uses optimization with four scenarios of the image data segmentation process [6]. The results of this study contain information on the Proposed approach with image segmentation optimization and the Faster R-CNN model deep learning method. The application of the proposed approach using a smartphone camera reached the highest accuracy of 90.43% in scenarios in recognizing image data on the MIAS breast cancer dataset in real-time.	Page 1
5	“This classification speeds up the detection and treatment of cancer by specialists” After this, it is suggested to add the whole structure or the paper’s layout at the end of the introduction.	This classification speeds up the detection and treatment of cancer by specialists	This classification speeds up the detection and treatment of cancer by specialists. This paper contains information on the proposed approach to optimizing breast cancer detection and classification, including the introduction, literature review; results and discussions; and conclusions.	Page 1
6	“This research uses the Faster-RCNN deep learning model. The application of the model is to detect, determine, and classify breast cancer image data.” Before part 2, it is suggested to supplement “related work” or “literature review” part so as to summarize the previous studies in detail and clarify the position of this study	There is no literature review section yet	there is already a literature review section	Page 1
7	Good research needs to explain what it means in practice, so the managerial implications should be stressed in this study.	So scenario2 is the best image segmentation in detecting and classifying images of breast cancer.	The results of this study can help any woman in the early detection of breast cancer independently because the Faster R-CNN system only requires breast mammogram images and a smartphone camera. Women can find out the symptoms or progression of cancer on themselves before having an examination with a specialist.	Page 6
8	The conclusion part should be enlarged.	The deep learning algorithm of the Faster R-CNN model has a	Based on the optimization results in this study, image	Page 6

	<p>The contributions of this study should be strengthened in the conclusion. What are the differences between this study and previous studies? In what ways does this article enrich the study of this stream?</p>	<p>Regional Proposal Network structure to generate target areas and make target positioning very precise. Therefore, Faster R-CNN is the leading research algorithm due to its efficacy in reliably detecting disease points. Optimization results in four scenarios, namely 63.47%, 90.43%, 85.22%, and 76.52% of the MIAS breast cancer dataset. So scenario2 is the best image segmentation in detecting and classifying images of breast cancer.</p>	<p>segmentation is very influential in increasing the accuracy of the Faster R-CNN deep learning model. Optimization results with four image segmentation scenarios, namely 63.47%, 90.43%, 85.22%, and 76.52% of the MIAS breast cancer dataset. In the second scenario, the detection and classification accuracy values are from 60.86% to 90.43%; This shows that the more color variations in the image, the more accurate it is. The faster R-CNN recognizes the target area and determines the position of the target to be very precise. The optimization and application of Faster R-CNN in this study is a perfect combination for detecting disease points reliably. The results of this study can help any woman in the early detection of breast cancer independently because the Faster R-CNN system only requires breast mammogram images and a smartphone camera. Women can find out the symptoms or progression of cancer on themselves before having an examination with a specialist. This research innovation supports the work of radiologists and specialists in detecting and classifying breast cancer.</p>	
--	--	--	---	--

## Optimization of Breast Cancer Classification using Faster R-CNN

Sunardi<sup>1</sup>, Anton Yudhana<sup>2</sup>, Anggi Rizky Windra Putri<sup>3\*</sup>

<sup>1,2</sup>Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia, 55191

<sup>3</sup>Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia, 55191

Corresponding Author Email: [anggie.windra.putri@gmail.com](mailto:anggie.windra.putri@gmail.com)

<https://doi>

### ABSTRACT

**Received:**  
**Accepted:**

**Keywords:**

*Faster R-CNN, Deep learning, MIAS, Data Augmentation, Image data, Breast cancer.*

Breast cancer results from aberrant cell division in the breast and leads to the formation of tumors. The modern lifestyle, which is instant and rarely exercises, is the main driving force for this disease. Therefore, this study aims to diagnose by recognizing the specific characteristics of cancer in a benign or malignant class in the breast area. This study approach uses the deep learning technology model Faster R-CNN and dataset Mammographic Image Analysis Society (MIAS). This model requires unique image characteristics to recognize and produce a higher accuracy value. Furthermore, this study proposes optimizing an image segmentation approach using Matlab, ImageJ, and Python software to enrich cancer-specific images. This approach plays a vital role in increasing the accuracy of cancer detection. The results of this study before optimization have an accuracy rate of 63.47% using a smartphone camera; after optimization, the highest accuracy value becomes 90.43%, therefore 9.57% requires further examination by a specialist. Based on these results, these results help assist radiologists in making decisions about the results of the initial examination of breast mammogram data.

## 1. INTRODUCTION

Modern lifestyle factors that are instantaneous and a lack of knowledge about healthy lifestyles make women vulnerable to breast cancer. In general, breast cancer susceptibility in women based on the age of 40-49 years is 92 cases (27%); vulnerable age 50-59 years there are 80 cases (23.5%); aged 30-39 years is 65 cases (19%); aged 60-69 years, there were 51 cases (15%). The lowest percentage of breast cancer occurred at 10-19 years old, with 2 cases (0.5%) [1],[2].

Data from the World Health Organization (WHO) Global Burden of Cancer Study (Globocan) recorded that the total breast cancer cases in Indonesia in 2020 reached 396,914, and the total number of deaths was 234,511 [3]. The disease is the most common problem and threatens women worldwide. Breast cancer occurs due to aberrant cell division in the chest area and leads to the formation of breast cancer cells [4]. Generally, the disease starts from a suspicious lump or bulge until the next stage becomes a tumour. Handling and curing this disease is very important because it can provide opportunities and a higher quality of life, increase life expectancy; eliminate or inhibit the growth of cancer cells; and save women from breast cancer malignancy.

Based on this, this research encourages women to be more aware of the threat of breast cancer, so early detection of breast cancer is essential for proper treatment and a healthy lifestyle. Therefore, this study aims to diagnose certain chest areas to detect and classify cancer into benign or malignant classes in women's breasts [4]. A detection system refers to suspicious lumps in the chest area, especially female breasts.

This study proposes an approach using image segmentation optimization and the Faster R-CNN deep learning model. This method studies image data in depth to recognize the unique characteristics of a particular object, thus

(ref 5-13 blm dirujuk) requiring optimization steps to produce high accuracy values [14]. This research uses optimization with four scenarios of the image data segmentation process [15].

The results of this study contain information on the Proposed approach with image segmentation optimization and the Faster R-CNN model deep learning method. The application of the proposed approach using a smartphone camera reached the highest accuracy of 90.43% in scenarios in recognizing image data on the MIAS breast cancer dataset in real-time. [This classification speeds up the detection and treatment of cancer by specialists [16]. This paper contains information on the proposed approach to optimizing breast cancer detection and classification, including the introduction, literature review; results and discussions; and conclusions.]

## 2. LITERATURE REVIEW

The following are some of the previous studies that carried out the detection of breast cancer. Xie et al. research the multiparametric magnetic resonance method and histogram analysis. This research aims to identify breast cancer image biomarkers using multiparametric MR imaging maps and whole-tumor histogram analysis. His research method was retrospective involving 134 patients with invasive ductal carcinoma [5].

Punithavathi and Devakumari researched the extraction of cancer features through experiments with the Gray Level Co-occurrence Matrix (GLCM) feature and the process of feature extraction testing on the Mammographic Image Analysis Society (MIAS) dataset [7]. Farhan and Kamil conducted a study using LBP, HOG, and GLCM feature extraction techniques to reduce false positives in breast diagnosis. Use of Mini-MIAS Database for this approach [9].

**Commented [AR3]:** 4. The findings or research results should be introduced briefly in the introduction part, so as to readers can get a whole view of this paper when they just finish reading the introduction.

**Commented [AR4]:** 5. "This classification speeds up the detection and treatment of cancer by specialists" After this, it is suggested to add the whole structure or the paper's layout at the end of the introduction.

**Commented [AR5]:** 6. "This research uses the Faster-RCNN deep learning model. The application of the model is to detect, determine, and classify breast cancer image data." Before part 2, it is suggested to supplement "related work" or "literature review" part so as to summarize the previous studies in detail and clarify the position of this study.

**Commented [AR1]:** 1. "Modern lifestyle factors that are instantaneous and a lack of knowledge about healthy lifestyles make women vulnerable to breast cancer." The importance to cure breast cancer should be emphasized in the introduction.

**Commented [AR2]:** 3. "This study proposes an approach using the Faster R-CNN deep learning model." The research aim is very clear, but the research gap has not been clearly defined. Why does the author want to conduct this study? The research motivation is not sufficient.

Fadlil et al. using The Mammographic Image Analysis Society (MIAS) dataset with a computer-aided detection method, and the accuracy result is 93.60% [10]. Dawngliani et al. dataset sourced from Mizoram Cancer Institute from 2009-2016. Data processing uses data mining, namely Random Forest with an accuracy rate of 97.13% [22]. Khan et al. research using The Wisconsin Breast Cancer Diagnostic (WBCD) dataset. In addition, data processing uses machine learning methods: logistic regression, decision tree, and others [8].

The results of this study obtained a logistic regression accuracy value of 98.60% and a decision tree of 94.73%. Marsilin [21] researched breast cancer classification using the query image with the database images of breast cancer. The research method uses content-based image retrieval (CBIR) and KNN Classifier, so the results have an accuracy rate of 85.00%.

The fundamental difference from previous research is the implementation of cancer image detection in this study with a smartphone camera. Therefore, this study proposes an early detection approach by optimizing image data segmentation and recognizing specific characteristics of cancer [11],[12],[13]. This study's breast cancer image data came from MIAS and has been normalized.

## 2.1 Deep Learning

Deep learning is a machine learning subfield with a network structure like the human brain [11],[17]. The difference between deep learning and machine learning is in data management performance. Deep learning can manage an increasing amount of large data and solve a problem, while machine learning optimizes small amounts of data [6],[18],[19]. Furthermore, deep learning in this research can adapt to updating data and recognizing functional and structural characteristics of disease [14].

## 2.2 Tensorflow Framework and Model Faster R-CNN

Tensorflow is an open-source deep learning framework with libraries to perform high-performance computational calculations [11]. One part is the R-CNN model. The model is based on Convolutional Neural Network (CNN) [20], which acts as a backbone and produces an algorithm that can detect objects to perform classification quickly [21].

## 2.3 Dataset MIAS Breast Cancer

This study uses a breast cancer dataset from MIAS (<https://www.kaggle.com/code/lemonweed/cancer-classification/data>) [10]. The MIAS dataset consists of 322 breast mammogram data with details of 123 abnormal and 199 normal. Unfortunately, this dataset is private, so no detailed information is available, including age, never been pregnant, overweight, giving birth over the age of 30 years, menopause, menstruating age <12 years, or exposure to radiation from radiotherapy. However, this dataset provides information on the image categories of benign and malignant cancers.

## 3. RESEARCH METHOD

This research uses the Faster-RCNN deep learning model. The application of the model is to detect, determine, and classify breast cancer image data. The following is a discussion of supporting theories about deep learning, the

Faster R-CNN model, the MIAS breast cancer dataset, and the proposed approach.

The proposed approach in this study utilizes the TensorFlow framework to run the Faster R-CNN model. The proposed approach of the Faster R-CNN model is used for the early prediction of breast cancer through detecting and classifying cancer objects in the chest and breast muscle areas.

This approach focuses on the detection algorithm so that it becomes a key indicator of recognizing cancer objects. In deep learning, Faster R-CNN is the main algorithm for detecting targets (breast cancer objects) because it can determine unique characteristics and produce high accuracy values. The following flowchart of the proposed approach model system is shown in Fig. 1.

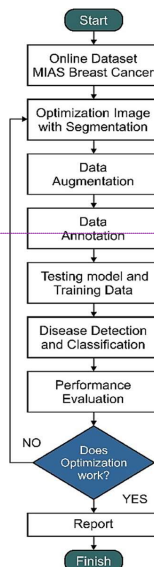


Figure 1. System model proposed approach using Faster R-CNN

Fig. 1 provides information on the working principle of the approach proposed in this study. The working principle consists of eight stages. The first stage of the online MIAS breast cancer dataset is determining and collecting breast cancer image data from online datasets. The second stage is preprocessing by optimizing the image data. The optimization uses a segmentation technique or algorithm according to the needs and functions. Optimization plays a role in increasing the value of object detection accuracy by manipulating the image. These manipulations include size, resolution, colour enhancement, and more.

The third stage is data augmentation, which expands the data set to improve model performance by producing various forms of image variations. This stage can enrich the sample data and add references to the image data training process. The fourth stage is data annotation; this stage is the process of labelling each breast cancer image. This labelling is the key to determining the image's position and a class of disease object areas. In addition, it is possible to detect multiclass objects.

The fifth stage is the system learns to recognize image data in depth to determine its unique characteristics. This introduction stage includes two essential processes: model

Commented [AR6]: "Previous studies have detected breast cancer using multiparametric magnetic resonance methods and histogram analysis." The author lists a number of previous studies. However, he or she does not summarize the advantages or shortcomings or the space to improve.

testing (fixed variable) and training data (independent variable). Model testing aims to identify breast cancer image data that have been optimized. Training data is the introduction of image objects with Faster R-CNN, which produces loss values. The loss value becomes an error parameter in data recognition—calculating loss value during training using Eq. (1).

$$L(\{pi_i\}, \{ti\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(pi_i, pi_i^*) + \lambda \frac{1}{N_{reg}} \sum_i pi_i * L_{reg}(ti, ti^*) \quad (1)$$

The following is detailed information on Eq. (1), where  $i$  is the anchor index,  $pi_i$  is the anchor prediction probability, and  $pi_i^*$  is the fundamental truth label.  $ti$  acts as a vector representing the four-parameter coordinates of the prediction bounding box, and  $ti^*$  is the ground-truth box associated with the positive anchor. Representative classification and regression losses with symbols  $L_{cls}$  and  $L_{reg}$ . In addition, it serves as a balancing parameter. Each parameter is normalized with  $N_{cls}$  and  $N_{reg}$  weighted  $\lambda$ . The second process, namely model testing, is the stage of calling the label map function and introducing image data based on labels. Finally, the label predicts disease so that the system will measure the recognition accuracy value based on the predicted classification results.

The sixth stage is system testing in detecting, recognizing, and classifying breast cancer from image data with a smartphone camera. The seventh stage is evaluating system performance in detecting and classifying breast cancer. The evaluation involves the performance of the system algorithm, the value of Accuracy, Precision, Recall, Specificity, and F1 Score in recognition of breast cancer. The following is the equation for each value using Eq. (2) to (6)[22].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

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

$$Specificity = \frac{TN}{TP + FP} \quad (5)$$

$$F1\ Score = \frac{2 * (Recall * Precision)}{Recall + Precision} \quad (6)$$

The measurement of each detection and classification value is based on Eqs (2) to (6) with a TP (True Positive) value which is the truth value of a prediction. The value of TN (True Negative) means the predicted value that is rejected in a positive state. The FP (False Positive) value is the identification of the wrong class value, and the FN (False Negative) value acts as the class value that is rejected in a negative state [11]. The final stage is delivering the optimization results of Faster R-CNN in detecting breast cancer. This study's proposed approach uses a Ryzen5 4000 series hardware processor, 4 GB NVIDIA GeForce GTX1650Ti GPU, 16 GB RAM, and Windows 10 64-bit.

## 4 RESULT AND DISCUSSION

This section describes optimizing breast cancer detection and classification system with Faster R-CNN. The cancer classification system will determine whether it is benign or malignant. This study utilizes datasets as research subjects to evaluate the performance of the proposed approach, while breast cancer is the object of research.

### 4.1 Online Dataset MIAS Breast Cancer

This study uses the MIAS online dataset amount of 322 image data. The data consists of two categories, namely Benign and Malignant breast X-rays. This study took 63 Benign and 52 Malignant data, which served as the primary model—263 data for training and 115 for testing deep learning capabilities.

### 4.2 Optimization Image

This study's optimization section seeks to improve breast cancer image data [23]. The data comes from the normalized MIAS dataset. The following is a list of optimization scenarios in this study, as shown in Table 1.

Table 1. Image data optimization list

Prop. Approach	Optimization	Software
Scenario1	Threshold	Mathlab
Scenario2	16 Colors	ImageJ
Scenario3	5 Ramps	ImageJ
Scenario4	Matplotlib	Python

Table 1 provides information on this study's image data optimization list. The optimization serves to generate special features in an image. The following is an example of an image from the MIAS dataset before optimization, as shown in Fig. 2.

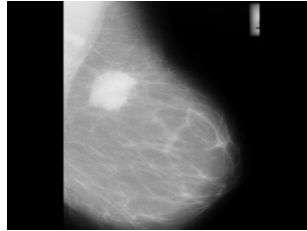
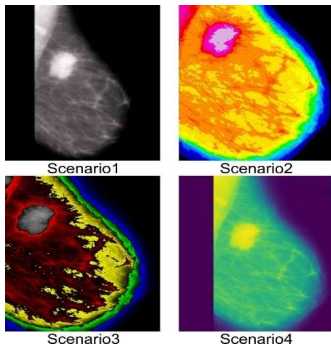


Figure 2. MIAS image data

Fig. 2 shows an image of breast cancer originating from MIAS and has been normalized. Next is the optimization process with four segmentation scenarios. The results of each scenario produce different images, as shown in Fig. 3.



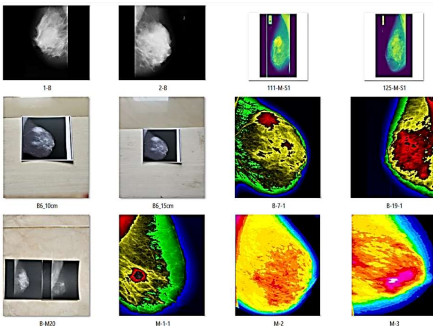


**Figure 3.** Four image segmentation scenarios

Fig. 3 provides an overview of the results of the image optimization process on the MIAS dataset. The optimization consists of four scenarios: the first threshold, the second 16 Colors, the third 5 Ramps, and the fourth matplotlib. Next, perform data augmentation of image data from each scenario.

#### 4.3 Data Augmentation

At this stage, multiply the training data from 115 to 263 image data. Multiplying the data is to provide a more comprehensive reference for training data. The following is the application of augmentation data in this study, as shown in Fig. 4.

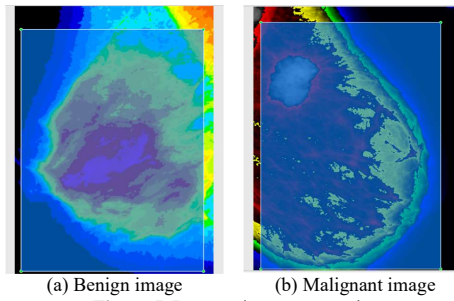


**Figure 4.** Data augmentation

Fig. 4 shows the application of the augmentation process, resulting in a large variety of image data. The augmentation process in this study was done by segmenting, printing and taking breast X-rays through a smartphone camera. The image capture process tests the system's success in detecting and classifying into Benign and Malignant classes.

#### 4.4 Data Annotation

At this stage, perform annotation, namely labelling the image data with python tools, namely labelling.py; labelling process based on information from the MIAS breast cancer dataset. The label serves as a learning target for the system in recognizing special features in image data. The following is the labelling process in this study, as shown in Fig. 5.



**Figure 5.** Process data augmentation

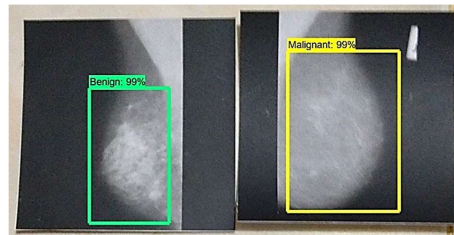
Fig. 5 presents assigning a label to each breast cancer image—the label functions for cancer attractions in the breast area. For example, point a is the image of Benign class cancer, and point b is the image of Malignant class cancer.

#### 4.5 Testing Model and Training Data

The system performs model testing and training data to identify breast cancer images at this stage. In addition, this stage reduces the value of loss or error when the system detects breast cancer images. The process of testing the model and training data for the Faster R-CNN model is shown in Fig. 6. Fig. 6 provides an overview of image processing by R-CNN in recognizing special features in the image. The total time of R-CNN in training and testing is 12 hours, with each scenario running for four hours.

#### 4.6 Disease Detection and Classification

This stage describes system testing results from CNN and Faster R-CNN models for detecting image data classification. The process of testing the proposed approach system in this study through a smartphone camera in real-time. System testing at this stage uses the results of training data scenario1. The following are the results of the detection and system classification of breast cancer images, as shown in Fig. 7.

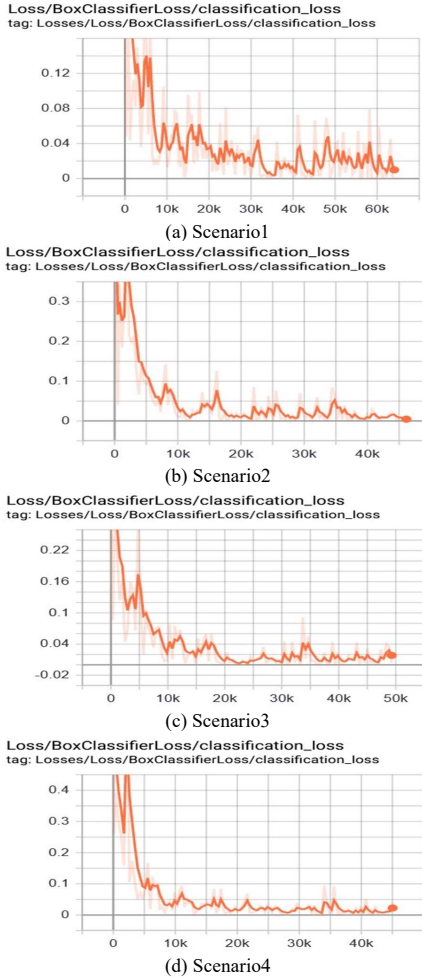


**Figure 7.** The process of detecting and classifying breast cancer images

Fig. 7 provides information on the accuracy of the proposed approach system. This system's accuracy level is based on model testing and training data results, resulting in an accuracy of 99% in Fig. 7. In addition, the system has successfully detected and classified breast cancer into benign and malignant classes.

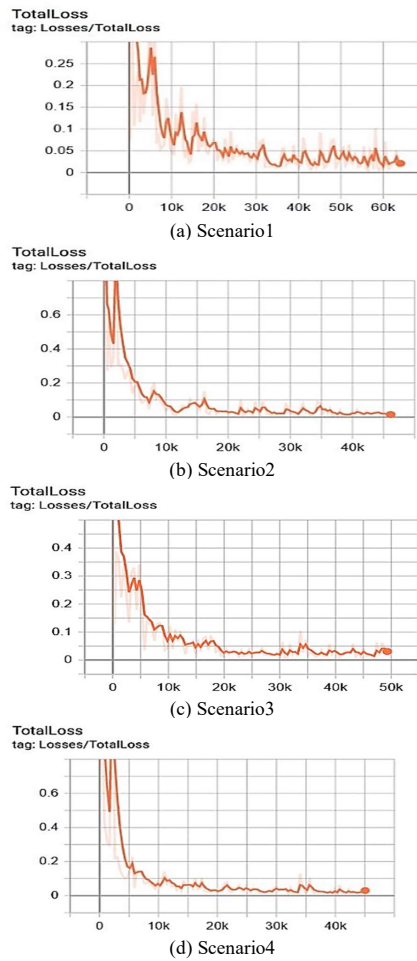
#### 4.7 Performance Evaluation

This stage describes the results of the evaluation and analysis of the performance of the proposed approach system. The following is a graph of the classification performance of each scenario during the research, as shown in Fig. 8.



**Figure 8.** Classification loss performance results

Fig. 8 provides information on each scenario's performance in this study's proposed approach system. The x-axis in Fig. 8 shows the loss value and the y-axis is the number of steps during the training process. Scenario1 obtains an average loss classification value in the range of 0.02-0.01; scenario2 in loss value  $>0.05$ ; scenario3 at value  $>0.02$ ; and scenario4 with a loss value of  $>0.05$  during the training process image data. The following is a graph of the total loss for each scenario, as shown in Fig. 9.



**Figure 9.** Total loss performance results

Fig. 9 shows a graph of the total loss of the Faster R-CNN model for each scenario. The x-axis in Fig. 9 shows the loss value and the y-axis is the number of steps during the training process. The average value of loss for each scenario is about  $>0.05$ ,  $>0.1$ ,  $>0.05$ , and  $>0.1$  during the image data training process. Performance evaluation measurements in this study include the Accuracy value for the ratio of Correct predictions (positive and negative) to the overall data; Precision value plays a role in knowing the ratio of correct positive predictions compared to the overall positive predicted results; the value of Recall (Sensitivity) is the ratio of true positive predictions compared to the overall data that is true positive;

Specificity value is the truth of predicting negative compared to the overall harmful data, and the F1 Score is a weighted comparison of the average precision and recall. This research focuses on the value of accuracy. The following is the mathematical calculation of the Accuracy, Precision, Recall, Specificity, and F1 Score levels for each scenario in this study using Eq. (2) to (6), as shown in Table 2.

Table 2 provides information on the performance of the Faster R-CNN model algorithm for each image scenario. In addition, the table provides information on the Accuracy values before and after optimization. Before the optimization process, the accuracy value is 60.86%, and after going through the optimization process with four scenarios, it becomes 63.47%, 90.43%, 85.22%, and 76.52%. Based on the calculation of the Accuracy value results, scenarios have a higher level of accuracy in detecting and classifying breast cancer.

#### 4.8 Report

This section presents each breast cancer image scenario before and after optimization efforts. The visualization of image data on the Faster R-CNN system is as shown in Fig. 10 which provide an image data distribution overview, including benign and malignant classes. In addition, the image shows how the system determines special features in breast cancer images based on benign or malignant cancer image classes.

```
Administrator: Anaconda Prompt (machine_Learning) - python train.py --logtostderr --train_dir=training/ --pipeline_config_path=tra...
I0513 17:35:59.740720 19516 learning.py:507] global step 165: loss = 0.5349 (0.233 sec/step)
INFO:tensorflow:global step 166: loss = 1.5722 (0.240 sec/step)
I0513 17:35:59.986310 19516 learning.py:507] global step 166: loss = 1.5722 (0.240 sec/step)
INFO:tensorflow:global step 167: loss = 1.0396 (0.268 sec/step)
I0513 17:36:00.259372 19516 learning.py:507] global step 167: loss = 1.0396 (0.268 sec/step)
INFO:tensorflow:global step 168: loss = 1.4010 (0.248 sec/step)
I0513 17:36:00.512096 19516 learning.py:507] global step 168: loss = 1.4010 (0.248 sec/step)
INFO:tensorflow:global step 169: loss = 0.7894 (0.286 sec/step)
I0513 17:36:00.805143 19516 learning.py:507] global step 169: loss = 0.7894 (0.286 sec/step)
```

Figure 6. Data training process

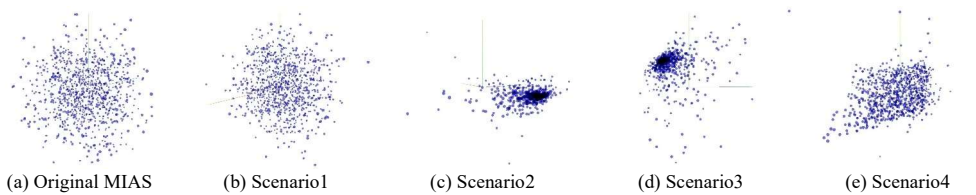


Figure 10. Visualization of image data on the Faster R-CNN model system

Table 2. Performance of the Faster R-CNN model algorithm for each scenario

Models	TP	FP	TN	FN	Accuracy	Precision	Recall	Specificity	F1 Scr.	Information
Org. MIAS	36	13	34	29	60.86%	73.46%	55.38%	72.34%	63.15%	Before Opt.
Scenario1	37	15	36	27	63.47%	71.15%	57.81%	70.59%	63.79%	Prop. Approach
Scenario2	55	8	49	3	90.43%	87.30%	94.83%	85.96%	91.01%	Prop. Approach
Scenario3	53	10	45	7	85.22%	84.13%	88.33%	81.82%	86.18%	Prop. Approach
Scenario4	45	18	43	9	76.52%	71.43%	83.33%	70.50%	76.92%	Prop. Approach

## 5 CONCLUSION

Based on the optimization results in this study, image segmentation is very influential in increasing the accuracy of the Faster R-CNN deep learning model. Optimization results with four image segmentation scenarios, namely 63.47%, 90.43%, 85.22%, and 76.52% of the MIAS breast cancer dataset. In the second scenario, the detection and classification accuracy values are from 60.86% to 90.43%; This shows that the more color variations in the image, the more accurate it is. The faster R-CNN recognizes the target area and determines the position of the target to be very precise. The optimization and application of Faster R-CNN in this study is a perfect combination for detecting disease points reliably. The results of this study can help any woman in the early detection of breast cancer independently because the Faster R-CNN system only requires breast mammogram images and a smartphone camera. Women can find out the symptoms or progression of cancer on themselves before having an examination with a specialist. This research innovation supports the work of

radiologists and specialists in detecting and classifying breast cancer.

#### CONFLICTS OF INTEREST

The authors declare no conflict of interest. In addition, the authors have thoroughly examined ethical issues, including plagiarism, double publication or submission, and redundancy.

#### REFERENCES

- [1] C. E. DeSantis *et al.*, "Breast cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.
- [2] A. T. Jalil, S. H. Dildi, and A. Karevskiy, "SURVEY OF BREAST CANCER IN WASIT PROVINCE, IRAQ," *Global Journal of Public Health Medicine*, vol. 1, no. 2, pp. 33–38, Nov. 2019, doi: 10.37557/gjphm.v1i2.7.
- [3] S. K. Nur Handayani, "Kanker dan Serba-Serbinya (Hari Kanker Sedunia 2022) – RSP Respira," Feb. 01,

Commented [AR7]: 8. The conclusion part should be enlarged. The contributions of this study should be strengthened in the conclusion. What are the differences between this study and previous studies? In what ways does this article enrich the study of this stream?

Commented [AR8]: 7. Good research needs to explain what it means in practice, so the managerial implications should be stressed in this study.

2022. <http://rsprespira.jogjaprovo.go.id/kanker-dan-serba-serbinya-hari-kanker-sedunia-2022/> (accessed May 01, 2022).
- [4] Z. Momenimovahed and H. Salehiniya, "Epidemiological characteristics of and risk factors for breast cancer in the world," *Breast Cancer: Targets and Therapy*, vol. 11, pp. 151–164, 2019, doi: 10.2147/BCTT.S176070.
- [5] T. Xie *et al.*, "Differentiation of triple-negative breast cancer from other subtypes through whole-tumor histogram analysis on multiparametric MR imaging," *European Radiology*, vol. 29, no. 5, pp. 2535–2544, 2019, doi: 10.1007/s00330-018-5804-5.
- [6] D. A. Zebari, D. Q. Zeebaree, A. M. Abdulazez, H. Haron, and H. N. A. Hamed, "Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images," *IEEE Access*, vol. 8, pp. 1–20, 2020, doi: 10.1109/ACCESS.2020.3036072.
- [7] P. V and D. D. D., "Detection of Breast Lesion Using Improved GLCM Feature Based Extraction in Mammogram Images," *SSRN Electronic Journal*, no. May, pp. 1040–1047, 2020, doi: 10.2139/ssrn.3663875.
- [8] M. Monirujjaman Khan *et al.*, "Machine Learning Based Comparative Analysis for Breast Cancer Prediction," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–15, Apr. 2022, doi: 10.1155/2022/4365855.
- [9] A. H. Farhan and M. Y. Kamil, "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 7, 2020, doi: 10.1088/1757-899X/928/7/072098.
- [10] A. F. Fadhil and H. K. Ormek, "A Computer-Aided Detection System for Breast Cancer Detection and Classification," *Selcuk University Journal of Engineering Sciences*, vol. 20, no. 1, pp. 23–31, 2021, [Online]. Available: <https://sujes.selcuk.edu.tr/sujes/article/view/534>
- [11] C. Orozco and C. Rebong, "Vehicular Detection and Classification for Intelligent Transportation System: A Deep Learning Approach Using Faster R-CNN Model," *International journal of simulation: systems, science & technology*, pp. 1–7, Jul. 2019, doi: 10.5013/IJSSST.a.20.S2.11.
- [12] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, May 2017, pp. 650–657. doi: 10.1109/FG.2017.82.
- [13] B. S. Bari *et al.*, "A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework," *PeerJ Computer Science*, vol. 7, p. e432, Apr. 2021, doi: 10.7717/peerj-cs.432.
- [14] H.-P. Chan, R. K. Samala, and L. M. Hadjiiski, "CAD and AI for breast cancer—recent development and challenges," *The British Journal of Radiology*, vol. 93, no. 1108, p. 20190580, Apr. 2020, doi: 10.1259/bjr.20190580.
- [15] E. Michael, H. Ma, H. Li, and S. Qi, "An Optimized Framework for Breast Cancer Classification Using Machine Learning," *BioMed Research International*, vol. 2022, pp. 1–18, Feb. 2022, doi: 10.1155/2022/8482022.
- [16] K. Govindaswamy and S. Ragunathan, "Genre Classification of Telugu and English Movie Based on the Hierarchical Attention Neural Network," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 54–62, 2021, doi: 10.22266/ijies2021.0228.06.
- [17] M. S. Dawngliani, N. Chandrasekaran, R. Lalmawipuii, and H. Thangkhannau, "Breast Cancer Recurrence Prediction Model Using Voting Technique," in *EAI/Springer Innovations in Communication and Computing*, 2021, pp. 17–28. doi: 10.1007/978-3-030-49795-8\_2.
- [18] A. Peryanto, A. Yudhana, and R. Umar, "Rancang Bangun Klasifikasi Citra Dengan Teknologi Deep Learning Berbasis Metode Convolutional Neural Network," *Format : Jurnal Ilmiah Teknik Informatika*, vol. 8, no. 2, p. 138, Feb. 2020, doi: 10.22441/format.2019.v8.i2.007.
- [19] V. Sstla, V. K. K. Kollu, L. K. Voggu, and R. Bhavanam, "Revue d ' Intelligence Artificielle," *International Information and Engineering Technology Association*, vol. 34, no. 3, pp. 323–330, 2020.
- [20] A. Peryanto, A. Yudhana, and R. Umar, "Convolutional Neural Network and Support Vector Machine in Classification of Flower Images," *Khazanah Informatika: Jurnal ...*, vol. 8, no. 1, pp. 1–7, 2022, doi: 10.23917/khif.v8i1.15531.
- [21] M. F. Rahman and B. Bambang, "Deteksi Sampah pada Real-time Video Menggunakan Metode Faster R-CNN," *Applied Technology and Computing Science Journal*, vol. 3, no. 2, pp. 117–125, Mar. 2021, doi: 10.33086/atesj.v3i2.1846.
- [22] S. Uma and M. Rao, "Medical Big Data Analysis using LSTM based Co-Learning Model with Whale Optimization Approach," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 4, Aug. 2022, doi: 10.22266/ijies2022.0831.56.
- [23] J. R. Marsilin, "An Efficient CBIR Approach for Diagnosing the Stages of Breast Cancer Using KNN Classifier," *Bonfring International Journal of Advances in Image Processing*, vol. 2, no. 1, pp. 01–05, Mar. 2012, doi: 10.9756/BIJAIP.1127.

# Tahap 4 Accepted, 08 November 2022

The screenshot shows a Gmail interface with a sidebar on the left containing navigation options like Compose, Mail, Chat, Spaces, and Meet. The main content area displays an email titled "Regarding Acknowledgement RIA 17220" from editor.ria.iieta.org. The email body contains the following text:

Dear author,

We have reached a decision regarding your submission to *Revue d'Intelligence Artificielle*.

Manuscript Title: Optimization of Breast Cancer Classification using Faster R-CNN  
Manuscript ID: 17220

Our decision is to: Accept Submission.

Before we proceed with the publication of your article, please complete the arrange payment of your article processing charge (USD 400) in 15 days by the following ways:

USD Remittance Path  
BENEFICIARY NAME: IIEETA-RSMC Beijing LTD  
ACCOUNT NUMBER: 0200296409116032240  
BENEFICIARY ADDRESS: Room 302, Building 3, Jinmao Plaza, Fengtai District, Beijing, PRC  
BENEFICIARY BANKER'S NAME: Industrial and Commercial Bank of China Beijing Municipal Branch, Beijing, PRC  
SWIFT CODE: ICBKCNBJBJM

Ps: 1. Remark the manuscript ID in the Remittance.  
2. Send the remittance receipt or payment screenshot, and name of the remitter to this email.

International Information & Engineering  
Technology Association  
#2020, Scotia Place Tower One, 10060 Jasper Avenue, Edmonton,  
AB T5J 3R8, Canada  
Tel: +1 780 218 9926; Fax: +1 780 341 0600  
www.iieta.org



## Acceptance Letter

### Revue d'Intelligence Artificielle

November 8, 2022

**Anggi Rizky Windra Putri**

Master Program of Informatics,  
Universitas Ahmad Dahlan, Yogyakarta,  
Indonesia, 55191

Dear Sunardi, Anton Yudhana, Anggi Rizky Windra Putri

MS: Optimization of Breast Cancer Classification using Faster R-CNN

I am pleased to inform you that as per the recommendation of the editorial board, your above-mentioned manuscript has been accepted for publication in *Revue d'Intelligence Artificielle* (ISSN 0992-499X).

Please note the following points, and ensure compliance:

- 1) Page Proofs of your paper will also be sent to you for minor corrections and approval.
- 2) Provide us with your E-mail ID for future correspondence.
- 3) Provide complete postal address with correct postal-code.
- 4) The corresponding author would receive a PDF of the published paper.
- 5) If your paper is found with plagiarism content after the release of Acceptance Letter, it will not be published in the Journal.

It is recommended that you cite this and other published works from International Information and Engineering Technology Association ([www.iieta.org](http://www.iieta.org)) in your papers to be published in other journals.

Yours sincerely,



Date: November 8, 2022  
Place: Edmonton, Canada

## Tahap 5a Editing Process, 08 January 2023

UNIVERSITAS  
AHMAD DAHLAN

Sunardi sunardi <sunardi@mti.uad.ac.id>

### RIA 17220

editor.ria.iieta.org <editor.ria@iieta.org>  
To: "anggi2007048019@webmail.uad.ac.id" <anggi2007048019@webmail.uad.ac.id>  
Cc: "sunardi@mti.uad.ac.id" <sunardi@mti.uad.ac.id>, "eyudhana@mti.uad.ac.id" <eyudhana@mti.uad.ac.id>

Sun, Jan 8, 2023 at 8:43 AM

Dear author,

Please check the attached files and follow the requirements:

1. Please download the attachment we sent you, and modify your article according to our mark based this version. Please **highlight** any change you make.
2. Please fill in "Copyright Transfer Agreement". Please note that "corresponding author's signature" in the agreement shall be manually signed.
3. Please add 2 references, citing 2 articles in RIA between 2019-2022, which need to be appropriate for this article.
4. If possible, please add 2 references, citing 2 appropriate articles in ATAIML ([https://www.acadlore.com/journals/ATAIML/2022\\_1\\_1](https://www.acadlore.com/journals/ATAIML/2022_1_1)), which is optional.

To ensure fast publication of your paper, please return your revised manuscript and the answers to all queries to this e-mail **before 8-9 January 2023**. Thus, we have enough time to process your manuscript in the next step. For further assistance, please do not hesitate to contact us via this e-mail.

#### Editorial Board

Revue d'Intelligence Artificielle

<http://www.iieta.org/Journals/RIA>

published by

International Information and Engineering Technology Association (IIETA)

<http://www.iieta.org/>

## Tahap 5b Editing Process, 13 January 2023

editor.ria.iieta.org <editor.ria@iieta.org>  
to anggi2007048019@webmail.uad.ac.id, me, eyudhana@mti.uad.ac.id

Fri, Jan 13, 11:19 PM ☆ ↶ ⋮

Dear author,

Please double check, Figures need to be in order, so Figure 6 needs to be between Figure 5 and Figure 7. And please reply to me as soon as possible, we plan to publish this article in these few days.

Best regards

RIA

One attachment • Scanned by Gmail



## Tahap 5c Editing Process, 09 February 2023

UNIVERSITAS  
AHMAD DAHLAN

Sunardi sunardi <sunardi@mti.uad.ac.id>

RIA 17220

editor.ria.iieta.org <editor.ria@iieta.org>  
To: "anggi2007048019@webmail.uad.ac.id" <anggi2007048019@webmail.uad.ac.id>  
Cc: "sunardi@mti.uad.ac.id" <sunardi@mti.uad.ac.id>, "eyudhana@mti.uad.ac.id" <eyudhana@mti.uad.ac.id>

Thu, Feb 9, 2023 at 3:38 PM

Dear author

There are two issues regarding reference, I have note them in the attached file, please check. Additionally, please supply **Orcid** of all authors. Thank you for your support and consideration.

Best regards

Editorial Board

Revue d'Intelligence Artificielle

<http://www.iieta.org/Journals/RIA>

published by

International Information and Engineering Technology Association (IIETA)

<http://www.iieta.org/>

## Tahap 5d Editing Process, 28 February 2023

editor.ria.iieta.org <editor.ria@iieta.org>  
to anggie.windra.putri@gmail.com, anggi2007048019@webmail.uad.ac.id, me, eyudhana@mti.uad.ac.id

Tue, Feb 28, 4:16 PM ☆ ↶ ⋮

Dear author

We are going to publish this article as soon as possible, please reply to me soon. Thank you for your support and consideration.

...

One attachment • Scanned by Gmail



## Tahap 5e Editing Process, 01 March 2023

RIA 17220 External Inbox x



editor.ria.iieta.org <editor.ria@iieta.org>  
to anggi2007048019@webmail.uad.ac.id, me, eyudhana@mti.uad.ac.id, anggie.windra.putri@gmail.com

Wed, Mar 1, 3:30 PM ☆ ↶ ⋮

Dear author,

Please supply all authors' **ORCID**. We will publish this article soon.

Before publication, it is decided that this final proof should be sent to the authors once again for careful reading and re-check, to rule out the mistakes / errors of all kinds.

Download "final proof". Read it carefully with particular reference to the following points:

1. If some corrections are left out, **highlight them** and attach notes on how to correct directly in the PDF file.
2. Check with care all the symbols in the text.

Please return the corrected final proof before **March 1-2 2023**.

It is our mutual responsibility that the academic works published in the journal should be "mistake or error free" and of quality as well. Our attention and efforts to this would make the journal still better, besides enhancing the utility of your published research.

Editorial Board

Revue d'Intelligence Artificielle

<http://www.iieta.org/Journals/RIA>



**Tahap 6**  
**Published, 25 March 2023**

RIA 17220 External Inbox x



editor.ria iieta.org <editor.ria@iieta.org>  
to angg2007048019@webmail.uad.ac.id, me, eyudhana@mti.uad.ac.id, anggiewindra.putri@gmail.com

Sat, Mar 25, 7:33 AM ☆ ↶ ⋮

Dear author  
We plan to publish this article on March 27, but we find a small issue which needs to be revised, please make modifications. Look forward your reply.

**Editorial Board**  
Revue d'Intelligence Artificielle

<http://www.iieta.org/Journals/RIA>

published by  
International Information and Engineering Technology Association (IIETA)  
<http://www.iieta.org/>



One attachment • Scanned by Gmail

