VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4077

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

EFFECTIVE BREAST CANCER DETECTION USING NOVEL DEEP LEARNING ALGORITHM

Irawadi Buyung[1]; Agus Qomaruddin Munir[2]; Putra Wanda[3]*

Dept. Electrical Engineering[1]
Dept. Information Technology[2]
Dept. Informatics[3]
Universitas Respati Yogyakarta
https://www.respati.ac.id
buyungirawadi@gmail.com, agusqmnr@respati.ac.id, putra.wanda@respati.ac.id

(*) Corresponding Author

Abstract— *Ultrasound is one of the most popular screening methods for finding breast cancer. However, the diagnosing process becomes complex due to the scarcity of experienced radiologists. We are motivated to use deep learning to address issues with medical image recognition because of its promising performance in various computer vision challenges. We present a rapid and precise breast cancer detection approach based on the Rapid-CNN. To undertake this experiment, we gather datasets related to breast cancer, pre-process them, train models, and assess the performance of the trained models. This model's bounding box detection of breast cancer has a training accuracy of up to 98.03% and a minimal loss of 0.78%. This model can detect the bounding box that is more than what it should be, so we applied NMS to eliminate the prediction of the bounding box that is less precise to increase accuracy.*

Keywords: *Breast Cancer Detection, Deep Learning, Rapid-CNN.*

Intisari— *Salah satu alat skrining yang paling umum untuk deteksi kanker payudara adalah ultrasound. Namun, kurangnya ahli radiologi yang mumpuni menyebabkan proses diagnosis menjadi tugas yang sulit. Pencapaian Deep Learning yang sangat bagus dalam berbagai masalah aplikasi komputer menginspirasi kami untuk menerapkan teknologi tersebut pada masalah pengenalan citra medis. Pada artikel ini, kami mengusulkan model deteksi Rapid-CNN untuk mendeteksi kanker payudara dengan cepat dan akurat. Eksperimen ini mengumpulkan dataset kanker payudara, melakukan pra-pemrosesan, melatih model, dan mengevaluasi kinerja model. Berdasarkan hasil percobaan diperoleh bahwa model ini dapat mendeteksi kanker payudara menggunakan bounding boxes dengan akurasi mencapai 98,03% dengan nilai loss hingga 0,78% pada proses training. Dalam model ini dimungkinkan untuk mendeteksi bounding box dengan lebih akurat, sehingga kami menerapkan NMS untuk menghilangkan prediksi bounding box yang kurang tepat untuk meningkatkan akurasi.*

Kata Kunci: *Deteksi, Kanker, Payudara,, Deep Learning, Rapid-CNN.*

INTRODUCTION

Breast cancer is one of the most common cancers in women for 15% of yearly fatalities [1]. Early identification of breast cancer can increase lifespan, reduce mortality risk, and improve quality of life [2]. Moreover, breast cancer screening can achieve early detection of ambiguous breast lesions. A systematic approach to breast screening is breast imaging diagnosis, which comprises breast MRI, mammography, and breast ultrasound [3].

Ultrasound is one of the most often utilized screening tools for finding breast cancer because of its pain-free, pleasant operation and outstanding real-time performance. Unfortunately, the ultrasonic instrument's high sensitivity makes it susceptible to the effects of various body tissues and their surroundings, resulting in many speckling noises and making diagnosis challenging. Moreover, the diagnostic effectiveness may suffer from a shortage of experienced radiologists, and 10–30% of diagnoses are missed [1] [2].

Conventional machine learning (ML) methods require large amounts of manual segmentation annotation data to train and test models for the classification or segmentation of ultrasound images. On the other hand, manual labeling is costly, time-consuming, and labor-intensive, significantly

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4077

raising the cost of system development [4]. Several papers have proposed methods for breast cancer detection. An article offered K-Nearest Neighbor (KNN) and Decision Tree to classify breast cancer. After selecting the Principal Component Analysis (PCA) technique, Wisconsin Diagnostic Breast Cancer (WDBC) dataset verified these two machine learning algorithms. Based on the findings of the experiments, the KNN classifier outperformed the decision tree classifier in the classification of breast cancer [5].

A report proposes ANN for breast cancer classification to increase classification accuracy. The Taguchi method initially counts the quantity of matched neurons in one of the hidden layers of the ANN. In accordance with the outcomes of the Taguchi approach, the model then goes through the training process, choosing the appropriate number of hidden neurons for the hidden layer. According to the findings of the experiments, this technique can give a classification accuracy for breast cancer that is 98.8% [6]. The accuracy, sensitivity, and specificity of the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers were examined in this study. The Mammographic Image Analysis Society uses the model to conduct its training procedure (MIAS). According to the experiment's outcomes, the SVM classifier performed better than the KNN classifier [7].

Deep learning technology has allowed image recognition to discover target areas in medical images and classify detected target features. Deep learning's detection and classification technique are comparable to doctors' operating procedures to determine diagnoses based on ultrasound results. Thus, the approach becomes a new solution to the earlier issues [8]. The current paper proposed CNN and Uniform Experimental Design (UED) to classify breast cancer. UED uses regression analysis to optimize CNN parameters [9][10]. Another study explored a comparative classification of breast MRI tumors using human-engineered radionics, Transfer Learning from Deep Convolutional Neural Network (DCNN), and the Fusion Method [11]. DCNN shows excellent potential for classifying several very various fine-grained objects. Therefore, further study proposed a deep learning method based on Bilinear Convolutional Neural Networks (BCNNs) for the acceptable category of breast cancer histopathology images [12].

Therefore, we propose Rapid-CNN to detect breast cancer by utilizing breast images. We establish an effective model to solve the breast cancer detection issue. In breast cancer detection, we present several crucial contributions to this study, particularly in the categorization of breast cancer using learning methods as follows:

1. We introduce a novel technique for detecting breast cancer utilizing the Rapid-CNN algorithm to train the dataset to develop a viable model. We use a large dataset of breast ultrasound images to build our model.
2. We build a model to detect breast cancer as a solution to find the location of breast cancer more effectively than traditional machine learning techniques.
3. We test a model to achieve high accuracy in detecting breast cancer effectively using unseen image features. We modify several parameters to achieve the best accuracy value to create the best training model.

The following is a breakdown of the journal's organizational structure: Part II expands upon earlier findings. Part III discusses the issue description of the study. The experimental design, comprising a feature learning method, a dataset, and preprocessing, is described in Section IV. At the same time, the study's findings and in-depth analysis are presented in Section V. Section VI concludes by discussing the conclusion.

MATERIALS AND METHODS

Based on annotated and actual breast images, Rapid-CNN is used to detect breast cancer in this study. By inserting an RPN layer, this method addresses the sluggish R-CNN performance issue. The obtained feature mapping is entered into the RPN network to identify potential target regions, and then the mapping features and potential target regions are entered into the ROI network. Rapid CNN structural properties have a loss function comparable to multitasking [24]. This study uses a deep learning architecture to develop the detection model in multiple steps. Fig 1. depicts several steps to conduct the study.

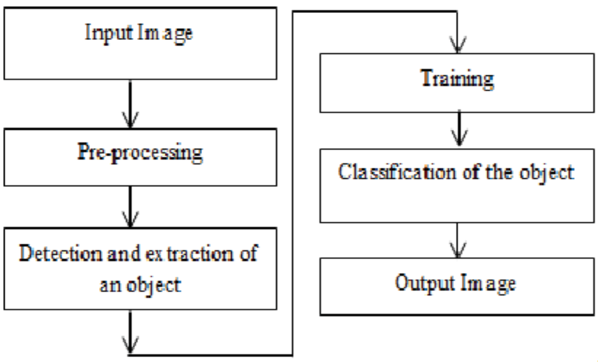

Fig.1: Experiment process Breast Cancer Detection using Rapid-CNN

In this study, classification assigns classes to images by finding similar features in images belonging to different categories and using them to identify and label images. In this part, we present a detailed explanation of the steps involved in detection using Rapid-CNN:

1. Input Image: The first step in detection using deep learning is to input an image. This can be a single image, a sequence of images, or a video feed.

2. Preprocessing: Before the image can be used for detection, it must be preprocessed to improve its quality and reduce noise. This can involve resizing the image, normalizing the pixel values, and performing other transformations such as color space conversions, image filtering, or data augmentation techniques.

3. Detection: The next step is to use a detection algorithm to identify objects within the image. This involves dividing the image into regions and analyzing each region to determine whether it contains an object. In this study, we utilize Rapid-CNN as a detection algorithm to construct a breast cancer detection model

4. Training: Before the detection algorithm can be used, it must be trained on a large dataset of images containing the objects it is meant to detect. During training, the Rapid-CNN algorithm learns to recognize the features associated with each object class, such as shape, color, and texture.

5. Classification: Once the objects have been detected within the breast image, the next step is to classify them. This typically involves using a separate classification algorithm, such as a deep neural network, to identify the specific class of each object. For example, if the thing is a breast cancer feature, the classification algorithm would label it as such.

6. Output: The output of the detection process is typically a bounding box that surrounds each detected object, along with the label for each object. This output can be used for various purposes, such as tracking objects in a video feed or identifying objects within a scene. Additionally, the accuracy and precision of the detection algorithm can be evaluated by comparing its output with ground-truth data.

In this study, Rapid-CNN detection involves a series of steps, from inputting an image to preprocessing it, detecting objects, training the algorithm, classifying the breast cancer class, and generating output. These steps can be iterated upon and refined to improve the accuracy and precision of the detection algorithm.

A. Proposed Method

This study presents a new CNN architecture, dubbed Rapid-CNN, in which area proposal generation occurs before the convolution layer. This step is rumored to decrease performance when working with large images. This NN resolves the performance issue by implementing the RPN layer and eliminating the current production of region proposals. Rapid-CNN suggested a solution to the performance problem. After feature extraction is performed, the model estimates RPN [25].

Rapid-first CNN's component is the region suggestion method, which gives bounding boxes or locations for likely picture objects. Commonly, a CNN is used to extract features from these objects during the second stage, which is feature creation. The third layer is a classification layer that predicts the object's class membership. The fourth layer is a regression layer that establishes the object's bounding box coordinates. This issue is addressed by the Rapid-CNN study, which generates regional suggestions using the RPN, reduces region proposal time, and permits the region proposal stage to share layers with future detection stages [26].

In the proposed network, an image's objective function is defined as:

$$L(\{P_i\}),\{t_i\} = \frac{1}{N_{cls}} \sum_i L_{cls}\ (P_i, P_i^*) + \lambda \sum_i P_i^*\ \ L_{reg}\,(t_{i,} t_i^*)$$

P_i is the probability that the anchor will be predicted as a target.

$$P_i^* = \begin{Bmatrix} 0 & negative\_lable \\ 1 & positive\_lable \end{Bmatrix}$$

$t_i = \{t_x, t_y, t_w, t_h\}$ is a vector representing the predicted bounding box's four-parameter coordinates. t_i^* is the coordinate vector of the positive anchor in the ground truth bounding box.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4077

Table 1. Mathematic notation of the Rapid-CNN

Notation	Description
P_i^*	The probability that the anchor is predicted to be the target.
t_i^*	The coordinate vector of the ground truth bounding box corresponds to the positive anchor.
t_i	The vector, represents the four-parameter coordinates of the predicted bounding box.
L_{cls}	The cross-entropy loss of binary classification (target & non-target)
L_{reg}	the regression loss
R	Smooth L1 function.
$L1$	Function.

In the Rapid-CNN process, each suggested region for each image requires feed-forward from CNN, even though the regions may overlap. Second, Rapid-CNN runs three separate models: a feature extraction model, a classification model, and a regression model. Rapid-CNN is a new solution to deal with both problems. The overall picture, as well as the extraction stage, have both been improved. Unlike R-CNN, it integrates all models into a single network, including feature extraction, classification, and detection. Second, the number of times a regional CNN has to be run has been reduced to one per image [24].

B. Main Idea

The main goal of this paper is to use the Rapid-CNN algorithm to develop a model based on breast ultrasound images to detect breast cancer. Rapid-CNN comprises two key processes for detecting and categorizing breast lesions [27]. Rapid-CNN starts with an ultrasonic image as input and outputs a rectangle box around the desired item. Second, the RPN, trained with ground truth data to give Regional Proposals, passes the convolution feature map through it. Therefore, the feature map is sent into the RPN, which generates a set of predictive-score areas [25] [28].

C. Dataset

In this experiment, breast ultrasound photos are collected from Kaggle.com to build our detection model. The data is then separated into training and testing datasets. Learning models are constructed using training datasets, whereas testing datasets are used to evaluate the performance or accuracy of the models. Here, the researcher divides the dataset into 80 percent for training and 20 percent for testing. Table 2 shows the details in terms of the distribution of the dataset used in the study, as follows:

Table 2. Distribution of the dataset

Dataset	Sample
Data Training (80%)	840
Data Testing (20%)	210
Total	**1.050**

D. Data Pre-Processing

Pre-processing is a stage to process high-resolution photographs; because high-resolution photo processing takes a long time, it must reduce the image size. Then, we convert the image to grayscale. After that, we perform noise removal to find and remove unwanted noise from the digital image. We sort each sample price by magnitude. The sample median in the window is the middlemost value, which can be a filter output. Grouping photos into segments is necessary for recording changes to image attributes. Image analysis is performed pixel by pixel after segmentation, and each pixel is labeled based on whether the gray level pixel is greater or less than the threshold value. As a result, segmented image analysis becomes easier [29].

E. Detection Method

To conduct our study, we used Rapid-CNN to detect breast cancer. We collected a dataset containing original and annotated breast images. This study builds a detection model of breast cancer with a training and testing process. Before the pre-processing stage, we split the dataset into two classes: Annotate and image, with 80% as the training dataset and 20% as the testing dataset.

During pre-processing, a detection model is simplified by resizing the image, removing undesired noise from the digital image, creating a bounding box for the position of the breast cancer target, and then labeling it based on the image. We extract vector values from features before putting them into the training and testing procedure.

After preprocessing, the training dataset is used to train the model. At the testing phase of the procedure, we utilize the testing dataset to evaluate model performance via data validation. In addition, a viable model was constructed and then validated using vector test data to assess the model's ability to detect breast cancer spots. Fig 1. depicts several steps to conduct the study.

RESULTS AND DISCUSSION

This work incorporates original and annotated breast photos in the training process. A low error rate proves that the model's performance is satisfactory. In this procedure, we set epochs = 50 to train the model by modifying various hyperparameters for optimal performance. Using the hyperparameter setting, the loss result is 0.2743, while at epoch 49, the loss result is 0.0780. These data suggest that more epochs can result in a small score drop. Our proposed model is adequate for detecting breast cancer with robust detection outcomes based on the training data.

In the testing process, we analyze 88 annotated breast cancer data to test the model and get a red bounding box to mark the location of breast cancer. The testing process produces expected output and model output detection. Fig. 2 shows desired output and model output detection.

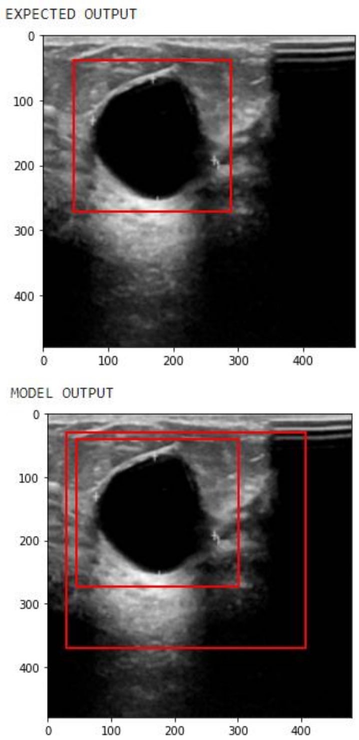

Fig. 2 Detection results

In Fig. 2, the expected outcome demonstrates accurate breast image recognition. In contrast, the model output implies that our model can detect the target. The testing results show that the intended output contains one bounding box, whereas the model output may contain many bounding boxes. In the subsequent testing phase, we determine actual and model detection based on the number of breast cancer spots denoted by a bounding box. The actual detection and model detection outcomes are graphically represented. Fig. 3 shows the actual detection results and the detection results of our model.

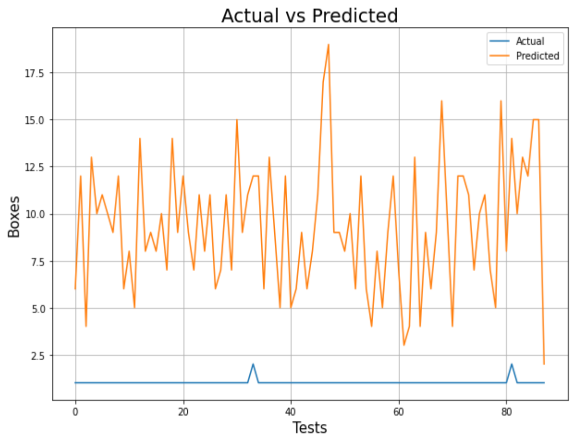

Fig. 3 Actual vs Predicted testing process

Fig. 3 shows the blue line that indicates the actual test result of accurate breast image detection, while the orange line indicates the effect of model detection. Based on the testing result, our model can detect cancer more accurately than the real detection image approach.

In the following process, we calculate the ratio value of actual and model detection. Fig. 4 shows the actual vs. Predicted ratio

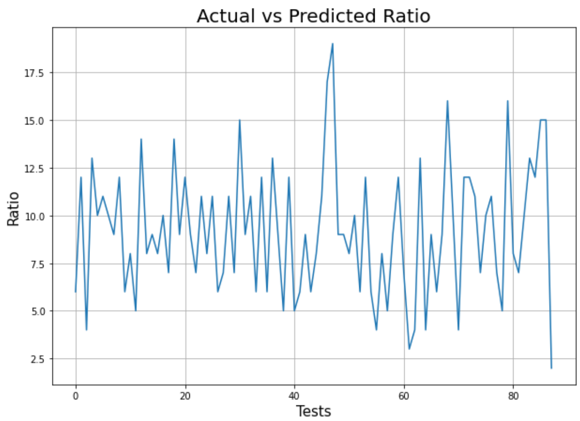

Fig. 4 Actual vs predicted ratio of the testing process

Fig. 4 displays a graph of the actual detection ratio with our model's detection. The ratio graph shows the number of detection results from our model divided by actual detection results. Our model can harvest better detection than the actual detection in breast image detection issues. In the previous dataset, there were redundant and overlapping bounding boxes. Therefore, we can conclude that our proposed model can detect more objects. Finally, after several stages, the Rapid-CNN can detect areas of breast cancer with several settings to obtain the best performance detection and accuracy value.

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4077

Table 3. Training and testing results with different optimizer function with Rapid-CNN.

Hyperparameter	Training Accuracy	Testing Accuracy
Epoch = 50 Batch size = 64 *Learning rate = 0.09*	98.03%	96.23%

Hyperparameter	Training Loss	Testing Loss
Epoch = 50 Batch size = 64 *Learning rate = 0.09*	0.78%.	2,13%

CONCLUSION

Detecting breast ultrasound pictures is a key obstacle in detecting breast cancer. The existing literature suggests employing conventional machine-learning techniques to address this difficulty. Yet, manual feature engineering is costly and time-consuming. To improve the performance of the detection model through autonomous feature engineering, Rapid-CNN is recommended for the development of a breast detection model.

Based on the experimental result, the Rapid-CNN can produce higher accuracy and tiny loss during the training process. In the training process, the study set some hyperparameters Epoch = 50. The training process has a loss of 0.0780, region box loss of 0.0471, abjectness loss of 0.0005, and RPN box loss of 0.0014. In the testing process, the study produces more bounding boxes with a testing accuracy of up to 96.23% and a minimum loss of 2,13%. Therefore, the model can be a promising solution to deal with breast cancer detection challenges accurately and in real-time.

Future studies can adopt another method, such as GAN architecture, to enhance this model. GAN can produce high-quality images and provide an accurate medical image analysis solution. With the development of additional features, the usage of a neural network that is dynamic should produce greater precision.

REFERENCE

[1] Pang, Ting, et al. "Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification". Elsevier Inc. 2021.
[2] Zhuang, Zhemin, et al. "Breast Ultrasound Tumor Image Classification Using Image Decomposition and Fusion Based on Adaptive Multi-Model Spatial Feature Fusion". Elsevier Inc. 2021.
[3] Chiao, Jui Ying., et al. "Detection and Classification the Breast Tumors using Mask R-CNN on Sonograms". Medicine. 2019.
[4] Shu, Xin, et al. "Deep Neural Networks with Region-based Pooling Structures for Mammographic Image Classification". IEEE Transaction on Medical Imaging. 2020
[5] Rajaguru, Harikumar, et al. "Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer". Asian Pasific Journal of Cancer Prevention, Vol 20.2019.
[6] Rahman, Md Akizur., et al. "Artificial Neural Networks with Taguchi Method for Robust Classification Model to Improve Classification Accuracy of Breast Cancer". PeerJ Computer Science.2021.
[7] Osman, Yahia and Umar Alqasemi. "Breast Cancer Computer-Aided Detection System based on Simple Statistical Features and SVM Classification". International Journal of Advanced Computer Science and Applications (IJACSA).2020
[8] Lin, Bor-Shing, et al. "Using Deep Learning in Ultrasound Imaging of Bicipital Peritendinous Effusion to Grade Inflamation Severity". Journal of Biomedical and Health Informatics.2020.
[9] Mewada, Hiren K., et al. "Spectral-Spatial Features Integrated Convolutional Neural Network for Breast Cancer Classification". Sensors.2020.
[10] Lin, Cheng-Jian and Shiou-Yun Jeng. "Optimization of Deep Learning Network Parameters Using Uniform Experimental Design for Breast Cancer Histopathological Image Classification". Diagnostics.2020.
[11] Whitney, Heather M., et al. "Comparison of Breast MRI Tumor Classification Using Human-Engineered Radiomics, Transfer Learning From Deep Convolutional Neural Networks, and Fusion Methods". Proceedings of The IEEE.2019.
[12] Liu, Weihuang., et al. "Fine-Grained Breast Cancer Classification with Bilinear Convolutional Neural Networks". Frontiers in Genetics.2020.
[13] GK, Rajini, et al. "Statistical Detection of Breast Cancer by Mammogram Image". Asian Journal of Pharmaceutical and Clinical Research. 2017
[14] Nithya, et al. "Robust Minimal Spanning Tree Using Intuitionistic Fuzzy C-means Clustering Algorithm for Breast Cancer Detection". American Journal of Neural Networks and Applications.2019
[15] Nascimento, Carmina Dessana Lima, et al. "Breast Tumor Classification in Ultrasound Images using Support Vector Machine and

Neural Networks". Research on Biomedical Engineering.2016
[16] Chtihrakkannan,R , et al. "Breast Cancer Detection using Machine Learning". International Journal of Innovative Technology and Exploring Engineering (IJITEE).2019
[17] Li, Mochen, et al. "Machine Learning-Based Decision Support System for Early Detection of Breast Cancer". Indian Journal of Pharmaceutical Education and Research.2020
[18] Ibeni, Wan Nor LWH, et al. "Comparative Analysis on Bayesian Classification for Breast Cancer Problem". Bulletin of Electrical Engineering and Informatics.2019
[19] Hamzah, Mohammad Diqi, and Antomy David Ronaldo. "Effective Soil Type Classification Using Convolutional Neural Network". International Journal of Informatics and Computation (IJICOM) Vol. 3, No.1. 2021
[20] Paryadi, Catur, M.Diqi, and Sri Hasta Mulyani. "Implementation of CNN for Plant Leaf Classification". International Journal of Informatics and Computation (IJICOM) Vol. 2, No.2. 2020
[21] Abidin, Adnan, Hamzah, and Marselina Endah H. "Efficient Fruits Classification Using Convolutional Neural Network". International Journal of Informatics and Computation (IJICOM) Vol. 3, No.1. 2021
[22] Shamy, S., and J.Dheeba. "A Research on Detection and Classification of Breast Cancer using K-Means GMM and CNN Algorithms". International Journal of Engineering and Advanced Technology (IJEAT).2019.
[23] Hameed, Zabit, et al. "Breast Cancer Hispathology Image Classification Using an Ensemble of Deep Learning Models". Sensors.2020.
[24] Zhao, Yuntau, et al. "A Malware Detection Method of Code Texture Visualization Based on an Improved Faster RCNN Combining Transfer Learning". IEEE.2020.
[25] Rahmat, Taufik, et al. "Chest X-Ray Image Classification using Faster R-CNN". Malaysian Journal of Computing.2019.
[26] Singh, Sunil, et al. "Face Mask Detection using YOLOv3 and Faster R-CNN Models: Covid -19 Environment". Multimedia Tools and Applications. 2021
[27] Wei, Kaizhen, et al. "Faster Region Convolutional Neural Networks Applied to Ultrasonic Images for Breast Lesion Detection and Classification". IEEE. 2020
[28] Abd-Ellah, Mahmoud Khaled, et al. "Automatic Diagnosis of Common Carotid Artery Disease using Different Learning Techniques". Journal of Ambient Intelligence and Humanized Computing. 2021
[29] AG, Neela, et al. "A Breast Cancer Detection using Image Processing and Machine Learning Techniques". International Journal of Recent Technology and Engineering (IJRTE).2019

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri