International Journal of Integrated Science and Technology (IJIST)
Vol.2, No.5, 2024: 484-493

Detect The Activity Of Benign And Malignant Breast Cancer

Ayu Fitriyani [1*], Muhamad Fatchan[2], Wahyu Hadikristanto[3]
Universitas Pelita Bangsa

Corresponding Author: Ayu Fitriyani**;** fitriyaniayu4@gmail.com

A R T I C L E I N F O

Keywords: Breast Cancer Detection, Convolutional Neural Network, Transfer Learning, Data Augmentation, VGG-16 Model.

Received : 20, March
Revised : 25, April
Accepted: 29, May

A B S T R A C T

Breast cancer detection is an important stage for early cancer diagnosis. In this study, a Convolutional Neural Network (CNN) algorithm is used to detect breast cancer. The dataset used consists of MRI scan images of benign and malignant breast cancer, which are processed through breast image cropping and data augmentation. The model was trained using CNN architecture with transfer learning method of VGG-16 model. The results of the model training showed good performance with an accuracy of 62%. These findings show the potential of using CNN and transfer learning in improving early detection of breast cancer.

DOI: https://doi.org/10.59890/ijist.v2i5.1870

ISSN : 3026-4685

https://journal.multitechpublisher.com/index.php/ijist

INTRODUCTION

One type of skin keratin commonly recognized worldwide among female groups is called breast keratin, which has been identified as a leading cause of death. Based on the experience of the cancer community in America, about 40,000 women and more than 600 men die from breast cancer (Nasser & Yusof, 2023). Breast cancer is the leading cause of death for women worldwide. According to the World Health Organization (WHO), 19.3 million cancer cases are expected by 2025. In Egypt, cancer is an ongoing problem, especially breast cancer (Ragab et al., 2019).

Currently, mammography is the most important method for accurately detecting breast cancer. The best alternative to mammography is magnetic resonance imaging (MRI). However, MRI is performed when the radiologist wants to confirm the presence of a tumor. The main thing to note about MRI is that patients may experience allergic reactions to certain materials or skin infections at the injection site (Ragab et al., 2019). This study uses photometric X-rays on breast tissue and applies them to a film surface. The results may indicate damaged breast wires that may not be physically visible or that do not seem to be causing problems (Nurbaiti et al., 2024).

One of the most significant contributing factors to the epidemic is the lack of early breast cancer education to help children understand and manage the disease non-invasively. These cancer deaths also occur because patients present to healthcare facilities at an advanced stage. If patients are already in the cancer stage, they will need more time to complete the check-in process. One of the main causes of the increase in breast cancer cases is the lack of public knowledge about cancer and how to detect it (Kusumawaty et al., 2021).

The purpose of this research is to improve the breast cancer detection method by using Convolutional Neural Networks (CNN). CNN is a type of statistical architecture that takes inspiration from the working of the human visual cortex and has been successfully used for many facial expression tasks, including raising eyebrows. The use of VGG-16 to perform transfer learning is done in (Penggunaan et al., 2023).

The CNN model known as VGG-16 was developed by Karen Simonian and Andreas Zissmann (2014). VGG-16 shows that this combination is necessary to ensure that each component of the network functions properly. The VGG-16 architecture has 16 layers that overlap the conventional and FC layers. The main feature of this model is its highly complex structure, which only generates 3x3 arcs and 2x2 layers from start to finish. The VGG16 reduction is too large to evaluate as it requires a large amount of memory, with a maximum parameter of 138M. However, instructions like this can be provided with proper guidance (Ariansyah et al., 2023).

LITERATURE REVIEW

This study examines the use of Convolutional Neural Networks (CNN), specifically the pre-trained VGG-16 model, in MRI classification for early tumor detection. Magnetic resonance imaging (MRI) is employed as a non-invasive technique to detect brain tumors, with a focus on detecting tumors early to facillitate more accurate diagnosis and treatment planning. CNN, particulary the

VGG-16 model, is used to classify MRI images into two groups: images with tumors and images without tumors (Candra et al., 2024).

This study provides an overview of the primary role of deep learning in cancer analysis, with CNN being the most widely used deep learning model. Although there is still some contreversy over the quality of the available data, The authors also highlight some challengs encountered when using deep learning for medical data analysis, such as selecting appropriate models, setting parameters, and obtaining high-speed medical data (Tandon et al., 2024).

This study discusses using the Transfer Learning (LTA) method, which utilizes Convolution Neural Networks (CNN) as it's foundation, to quickly and accurately diagnoses breast cancer using binary classification (benign and malignant). In this study, the Deep Learning technique is applied by utilizing pre-trained models such as ResNet-50 and VGG-16 to extract high-level features from mammography images. This study highlight the importance of early detection of breast cancer and the use of medical imaging technology to enable more accurate diagnoses (Murthy et al., 2024).

This study suggeststhat by employing data augmentation and transfer learning techniques, the model's performance can more accurately classify breast tumors. Two breast ultrasonography datasets are utilized in this study, where CNN models are evaluated based on accuracy, precision, and recall. It is believed that data augmentation can enhance the model's performance (BOUKAACHE et al., 2024).

METHODS

Here are the steps taken to collect, analyze, and interpret research data.

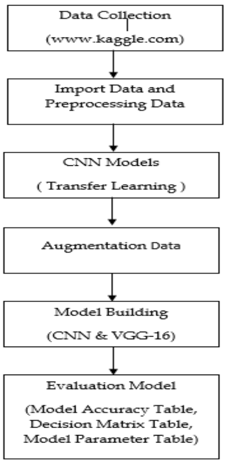

Figure 1. Research Flow

In conclusion, the steps outlined above constitute a comprehensive framework for managing and analyzing research data, particularly in the context of utilizing Convolutional Neural Network (CNN).

Data Collection

Data collection for this study is available at https://www.kaggle.com/datasets/ayufitriyani21/kanker-payudara. The Kaggle process is searching and selecting datasets from the Kaggle platform, which is a popular online community for data scientists and machine learning practitioners. Kaggle provides a variety of datasets from many domains, which are contributed by users worldwide.

Preprocessing Data

Before moving to the next stage of model development, the data needs to go through preprocessing to ensure quality and consistency. This process begins with dividing the data based on breast cancer labels: "no" for benign cases and "yes" for malignant cases. Subsequently, a cropping process is performed on the images to focus on relevant areas in the mammography images.

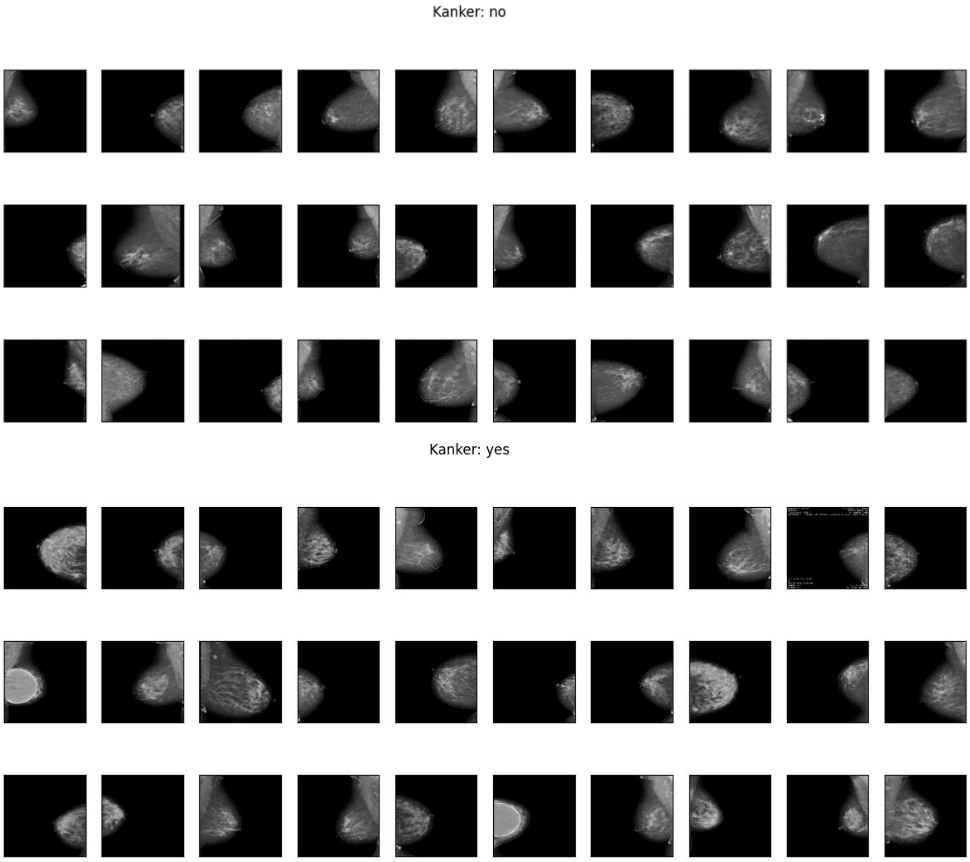

Figure 2. Sample Image

The graph above depicts thirty sample plots from the training dataset for each label. Yes, as for label number one, it represents malignant breast cancer, and Image number two serves as a sample of benign breast cancer. This dataset contains medical information on breast MRI indicating both benign and malignant conditions.

The data preprocessing stage involves several crucial steps to ensure the quality and consistency of the data to be used in the model. These steps include normalizing the image size, resizing for uniformity, and adjusting the contrast to enhance the visual quality of the images. Additionally, irrelevant or poor-quality images are removed to maintain the integrity of the dataset. This process is essential to ensure that the data used in model training and testing is of optimal quality and can provide accurate and consistent results.

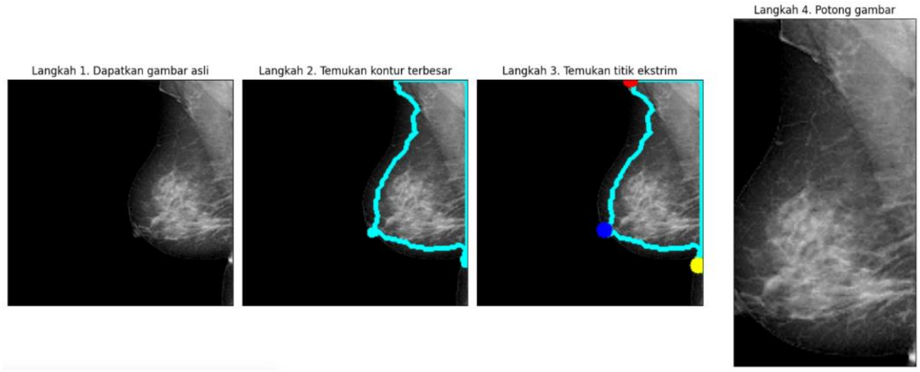

Figure 3. Brain Image Cropping

Breast Image Cropping. This diagram depicts the process of breast reorganization conducted on the images obtained from magnetic resonance imaging (MRI). This breast preparation technique allows us to automatically identify important areas of the breast for cancer detection by locating the affected region in the picture. This diagram provides a visual representation of how breast cropping affects the graph and aids in our understanding of the process.

Model CNN(Transfer Learning)

This research adopts transfer learning to utilize pre-trained Convolutional Neural Network (CNN) models. As a basis, we use the VGG-16 model which has been proven effective in various image classification tasks. We take advantage of this model's ability to understand complex image features. To adapt the model to the specific task of liveness detection of benign and malignant breast cancer, we made adjustments to the final layer of the model. Making these adjustments optimizes the model to understand the most relevant patterns and features in our image data. With this transfer learning approach, it can utilize the existing knowledge in the VGG-16 model to improve the performance and accuracy of the model in the classification task performed.

Augmentation Data

To enrich the diversity of training data and prevent overfitting tendencies, data augmentation techniques can be applied. The augmentation process involves various image transformations, including rotation, cropping, horizontal inversion, and noise addition. Through these steps, the model is introduced to a wider variety of images, which is expected to improve the generalization ability of the model. By introducing such variations, the model can learn from various situations that may occur in real data, resulting in stronger and more stable results when applied to test data that has never been seen before.

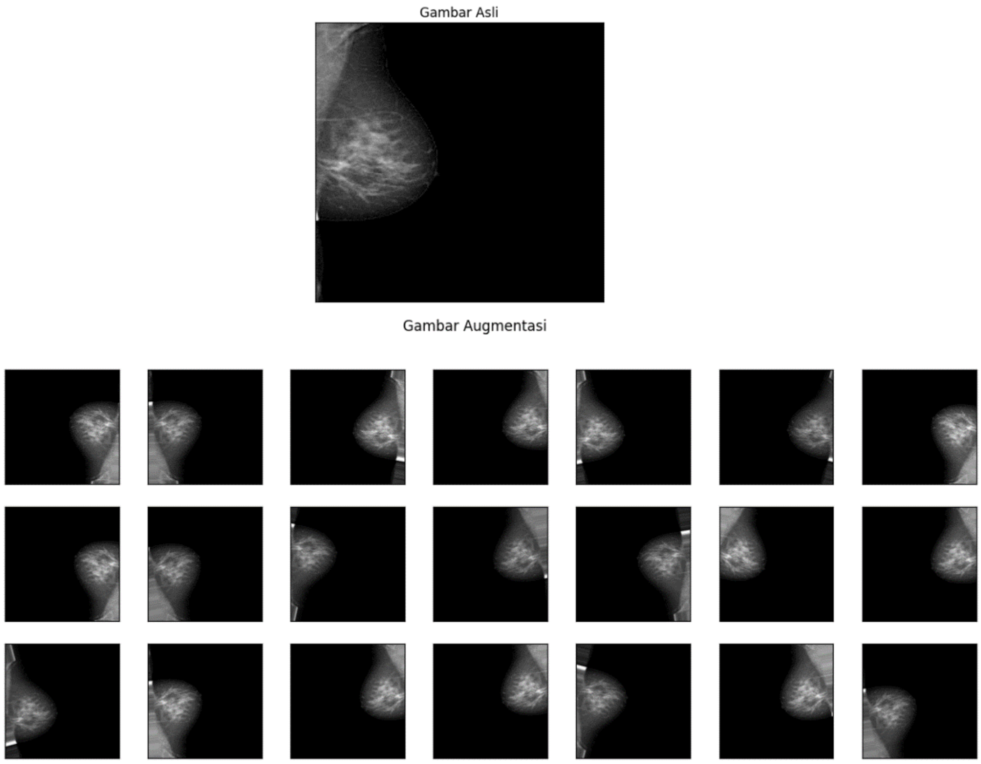

Figure 4. Augmented Images

Model Building(CNN & VGG-16)

The model building process involves two main approaches: first, using a pre-built Convolutional Neural Network (CNN), and second, utilizing the VGG-16 model with transfer learning techniques. The pre-built CNN was designed with various convolution and pooling layers to extract important features from the images. This model was developed specifically for the purpose of liveness detection of benign and malignant breast cancer. Meanwhile, the pre-trained VGG-16 model is customized by adding some dense layers at the end for the classification of two classes, namely benign and malignant. By combining these two approaches, we hope to improve the model's ability to identify and distinguish images relevant to breast cancer liveness detection.

Evaluation Model

Once the model has been trained, an evaluation step is performed to assess the performance of the model using test data that has been previously separated from the training data. The evaluation metric used is accuracy, which measures how well the model classifies breast cancer liveness. In addition, we also look at the confusion matrix, which gives an idea of how well the model predicts the positive and negative classes. The confusion matrix helps us understand how good the model is at avoiding classification errors, such as identifying breast cancer as inactive (false negative) or classifying inactive breast cancer as active (false positive). With this evaluation, we can gain the necessary insight into the quality and performance of the model that has been developed.

RESULTS

In this section, the steps taken to complete the study as well as the results obtained from the implemented model are presented. The results of the model are shown through the confusion matrix generated during the validation evaluation. This confusion matrix gives an idea of the performance of the model in classifying the data into correct labels.

From the figure above, the confusion matrix shows that the model has a validation accuracy of 0.62. This means that the model managed to correctly predict 62% of the total cases tested. Specifically, the model correctly identified 17 negative cases (label 0) and 18 positive cases (label 1). However, there were also misclassifications, where 12 negative cases were classified as positive and 9 positive cases were classified as negative.

These results show that while the model has a fairly good ability to classify the data, there is still room for improvement, especially in reducing the number of misclassifications. This confusion matrix provides valuable information to identify areas where the model may need to be further adjusted or improved.

For ease of understanding, these results have been summarized in the following table:

Validation Accuracy = 0.62

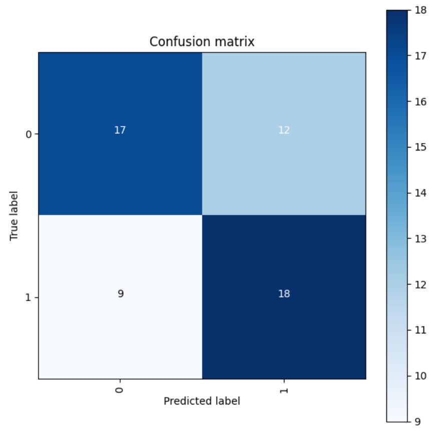

Figure 5. Validation Accuracy

Test Accuracy = 0.62

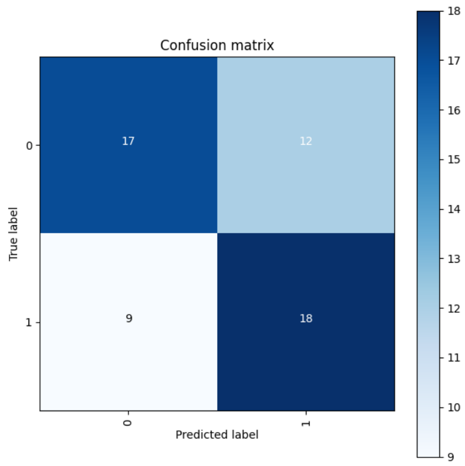

Figure 6. Test Accuracy

As such, the tables and graphs presented above provide a clearer and more structured view of the model's performance in this study. The test accuracy of 0.62 indicates that the model has a solid foundation, but still needs further improvement to increase precision and reduce prediction error.

DISCUSSION

The model evaluation results in this study show that although the model is able to achieve an accuracy of 0.62, there are still several aspects that need attention to improve the quality and reliability of the model. One of the main aspects that need attention is the handling of classification errors that occur. From the confusion matrix, it can be seen that there are 12 negative cases that are misclassified as positive and 9 positive cases that are misclassified as negative. This shows the potential to improve the precision of the model, especially in identifying true positive cases.

CONCLUSIONS AND RECOMMENDATIONS

This research has developed and evaluated a benign and malignant breast cancer liveness detection model using Convolutional Neural Network (CNN) and transfer learning techniques with the VGG-16 model. Although the resulting model achieved an accuracy of 0.62, this result shows that there is still room for improvement, especially in terms of reducing classification errors. The built model shows a basic ability in classifying mammography images into benign and malignant categories, with a fairly good initial accuracy as a first step. Errors in the classification of negative to positive and vice versa indicate that the model requires further customization to improve its accuracy, especially in detecting true positive cases. The implementation of transfer learning with the VGG-16 model shows its effectiveness in handling medical image classification tasks, although it still requires further optimization.

Recommendations for future research include several important aspects. Improving the quality and quantity of training data, including conducting additional data collection and filtering to ensure only high-quality images are used. Further tuning the hyperparameters of the model to improve classification performance. Experiments with different configurations of convolutional and dense layers should also be conducted. Using more diverse data augmentation techniques to improve the generalization ability of the model, such as the addition of a variety of image transformations. Conduct further validation with more varied clinical data to ensure that the model is reliable in real clinical environments. Finally, explore the use of other deep learning models that may be more effective in this task, such as ResNet or Inception, to compare performance and determine the best model to use. With these steps, it is hoped that the breast cancer liveness detection model can be further improved, thus making a significant contribution to the early diagnosis and treatment of breast cancer.

FURTHER STUDY

Every study has limitations, including this one. One of the main limitations in this study is the number of epochs used for model training, which is only 50 epochs. This limited number of epochs most likely contributed to the suboptimal performance of the model. For future research, it is recommended that the model training be conducted with more epochs. Increasing the number of epochs is expected to help the model to better understand the patterns in the training data, resulting in better and more accurate performance.

In addition, another limitation that needs to be considered is the variety of data used. Future research should consider expanding the dataset by collecting more images from different sources. This will help improve the model's ability to generalize and provide more accurate results when applied to data that has not been seen before.

Data augmentation techniques can also be further explored to increase the diversity of the training data. By introducing a greater variety of image transformations, the model will be trained to recognize different image shapes and conditions, thereby reducing overfitting and improving generalization.

Future research could also consider using more advanced deep learning models or a combination of multiple models to obtain better results. Experiments with more robust regularization techniques or more rigorous hyperparameter adjustments can also help improve model performance.

ACKNOWLEDGMENT

The authors would like to thank colleagues at Pelita Bangsa University for their valuable advice and support in the preparation of this paper. Support from supervisors and friends was also very helpful in completing this research. Thanks also to the family who always provide motivation.

REFERENCES

Ariansyah, M. H., Winarno, S., & Sani, R. R. (2023). Monkeypox and measles detection using CNN with VGG-16 Transfer Learning. *Journal of Computing Research and Innovation*, *8*(1), 32–44. https://doi.org/10.24191/jcrinn.v8i1.340

BOUKAACHE, A., NASSER EDİNNE, B., & BOUDJEHEM, D. (2024). Breast Cancer Image Classification using Convolutional Neural Networks (CNN) Models. *International Journal of Informatics and Applied Mathematics*, *6*(2), 20–34. https://doi.org/10.53508/ijiam.1407152

Candra, D., Wibisono, G., Ayu, M., & Afrad, M. (2024). LEDGER: Journal Informatic and Information Technology Transfer Learning model Convolutional Neural Network menggunakan VGG-16 untuk Klasifikasi Tumor Otak pada Citra Hasil MRI. *Open Access Ledger*, *3*(1).

Kusumawaty, J., Noviati, E., Sukmawati, I., Srinayanti, Y., & Rahayu, Y. (2021). Efektivitas Edukasi SADARI (Pemeriksaan Payudara Sendiri) Untuk Deteksi Dini Kanker Payudara. *ABDIMAS: Jurnal Pengabdian Masyarakat*, *4*(1), 496–501. https://doi.org/10.35568/abdimas.v4i1.1177

Murthy, D. G. N. K., H V, D. C., E V, D. V., B M, D. M., & Srinivas, D. C. (2024). A Novel Transfer Learning Approach to Improve Breast Cancer Diagnosing on Screening Mammography. *International Journal of Electrical and Electronics Research*, *12*(1), 300–307. https://doi.org/10.37391/ijeer.120141

Nasser, M., & Yusof, U. K. (2023). Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction. *Diagnostics*, *13*(1). https://doi.org/10.3390/diagnostics13010161

Nurbaiti, N., Hidayat, E. P. S., Anwar, K., Hermawan, D., & Izzuddin, S. (2024). Development of AI Models from Mammography Images with CNN for Early Detection of Breast Cancer. *Generation Journal*, *8*(1), 42–53. https://doi.org/10.29407/gj.v8i1.21601

Penggunaan, A., Terhadap, S., Emosional, K., Kelas, S., Sekolah, I. I. I., Desa, D. I., & Mayong, P. (2023). *1 , 2 , 3 123*. *09*(2), 94–98.

Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, *2019*(1), 1–23. https://doi.org/10.7717/peerj.6201

Tandon, R., Agrawal, S., Rathore, N. P. S., Mishra, A. K., & Jain, S. K. (2024). A systematic review on deep learning-based automated cancer diagnosis models. *Journal of Cellular and Molecular Medicine*, *28*(6), 1–20. https://doi.org/10.1111/jcmm.18144