Accuracy Enhancement in Early Detection of Breast Cancer on Mammogram Images with Convolutional Neural Network (CNN) Methods using Data Augmentation and Transfer Learning

Arief Broto Susilo [1*], Endang Sugiharti [1]

[1] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: ariefbros@students.unnes.ac.id

ARTICLE INFO

Article history
Received 10 February 2021
Revised 15 March 2021
Accepted 12 April 2021

Keywords
Convolutional Neural Network
Transfer Learning
Data Augmentation
Deep Learning

ABSTRACT

The advancement of computer technology has made it possible for computers to imitate the work of the human brain to make decisions that can be used in the healthcare sector. One of the uses is detecting breast cancer by using Machine Learning to increase the sensitivity and or specificity of detection and diagnosis of the disease. Convolutional Neural Network (CNN) is the most commonly used image analysis and classification method in machine learning. This study aims to improve the accuracy of early detection of breast cancer on mammogram images using the CNN method by adding the Data Augmentation and Transfer method. Learning. This study used a mammography image dataset taken from MIAS 2012. The dataset has seven classes with 322 image samples. The results of accuracy tests of early detection process of breast cancer using CNN by utilizing Data Augmentation and Transfer Learning show several findings: Model VGG-16, Model VGG-19, and ResNet-50 produced the same training accuracy rate of 86%, while for validation accuracy, Model ResNet-50 produced the highest level of accuracy (71%) compared to other models (VGG-16=64%, VGG-19=61%). The use of more image datasets may create better accuracy.

1 Introduction

The development of computer technology which is increasingly rapid in this era provides more and more benefits to human life. Nowadays, computers can work by imitating the human brain to make decisions that can be used in the healthcare sector. One of which is the detection of breast cancer. Breast cancer is the most common disease affecting women. Recent data shows 1 in 10 women develop breast cancer during their lifetime. Data from the American Cancer Society (ACS) states that breast cancer is women's primary type of cancer. A total of 57,650 women were diagnosed with breast cancer, and 39,520 died from breast cancer (Tuna *et al.*, 2014).

Global Cancer Observatory data from the World Health Organization (WHO) in 2018 shows that the most cancer cases in Indonesia are breast cancer, which is 58,256 cases or 16.7% of the total 348,809 cancer cases in the world (Widowati, 2019).

In the medical field, the diagnosis of cancer is a hot issue. With information technology, software, and hardware development, mass descriptive tumor feature data will be obtained and information on cancer research easily (Putra, Santoso, & Zahra, 2014).

A popular machine learning technique is deep learning, in which a computer model performs direct classification tasks from learning from text, images, or sound (Rawat & Wang, 2017). Models have trained on many datasets and a Convolutional Neural Network (CNN) architecture containing multiple layers (Devare, 2019). In medical imaging, deep learning automatically detects cancer cells (Charan, Khan, & Kurshid, 2018).

The growing interest in deep learning has made CNN the most commonly used image analysis and classification method. CNN has had the latest results in various classification tasks (Mikołajczyk, Agnieszka, & Grochowski, 2018). However, the inadequacy of the dataset used could affect the accuracy of CNN. This problem can be overcome by increasing the amount of test data and training data used for the training process to improve the level of accuracy (Mahmud, Adiwijaya, & Al Faraby, 2019) by applying data augmentation to improve the dataset in training ((Mikołajczyk *et al.*, 2018; Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). One other alternative to overcome the inadequacy of the dataset is to apply transfer learning (Barman *et al.*, 2019; Gopalakrishnan *et al.*, 2017; Vesal *et al.*, 2018). Transfer Learning is a technique of using previously learned features from a data set containing a large amount of data and then transferring that learning and applying it to its dataset. Previous research "Classification of Breast Cancer Histology Images Using Transfer Learning" has succeeded in increasing the accuracy by 18.3% with the transfer learning method (Ahmad, Ghuffar, & Khurshid, 2019). In addition, transfer learning can also solve computing time and the small training dataset effectively (Barman *et al.*, 2019).

Based on this background, the authors researched how to apply augmentation and transfer learning data to the Convolutional Neural Network method to increase breast cancer detection accuracy. This study is entitled "Increasing Accuracy of Early Detection of Breast Cancer on Mammogram Images with Convolutional Neural Network (CNN) Methods Using Data Augmentation and Transfer Learning".

2 Methods

In this research, the Data Augmentation method was applied to enrich the data and Transfer learning as feature extractors. The technique was employed to optimize the Convolutional Neural Network (CNN). The classification used in this study uses softmax from CNN. From the classification results, an increase in accuracy from CNN after the application of Data Augmentation and Transfer Learning methods was obtained. The flowchart of the method used in this study is shown in Figure 1.

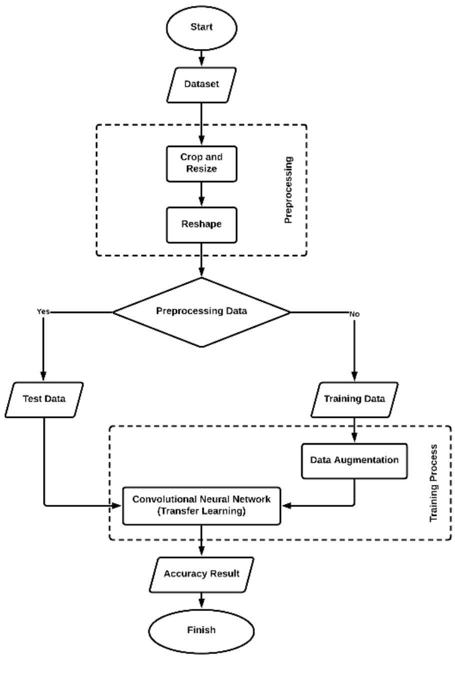

Figure 1. Research flowchart

2.1 Data Preprocessing

This study used the MIAS dataset in the form of mammographic images, consisting of 332 images divided into 7 classes, as shown in Figure 2.

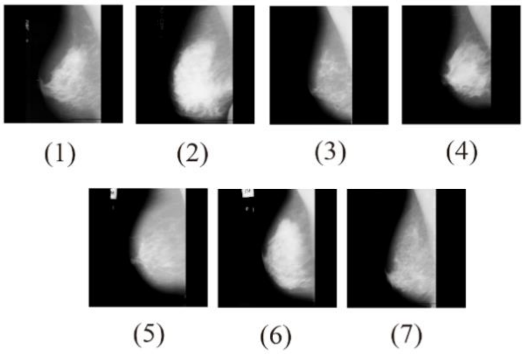

Figure 2. The example of mammography images per class

All facial images were divided into training data and testing data. This split was done automatically with the ratio of 70:30. Data pre-processing employed two stages, namely crop and resize and reshaping. Cropping was done to cut the dataset image with a scale of 250:950, 250:950 (pixel x was taken in the range 250 to 950, and pixel y was taken in the range 250 to 950). The image was 700x700 pixels in size. Cropping stages aimed to give more focus to the image area needed for the training stage later. Furthermore, the dataset was resized from 700x700 pixels to 224x224 pixels (the image size was adjusted to the input received by the model). The image results after the crop and resizes stage can be seen in Figure 3.

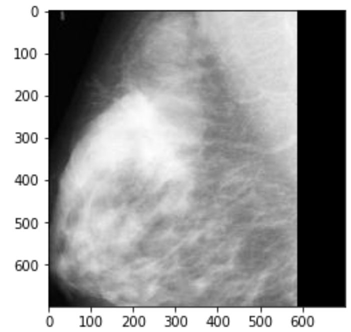

Figure 3. Image result after crop and resize

The next stage is reshaping. Initially, each mammographic image was converted into a matrix containing numbers at the reshape stage then the matrix was modified in dimensions according to what the model needs.

2.2 Data Augmentation

Data augmentation in this study used ImageDataGenerator, which the Keras library owns. The data augmentation applied were nine arguments, as shown in Table 1.

Table 1. Types of data and augmentation parameter

Data Augmentation Types	Parameters
Zoom	0.2
Rotation	5 degrees
Width shift	0.15
Heigth shift	0.15
Shear	0.01
Horizontal Flip	True
Vertical Flip	True

Fill mode	Nearest

With the augmentation type in table 1, the data augmentation results were obtained, as shown in Figure 4.

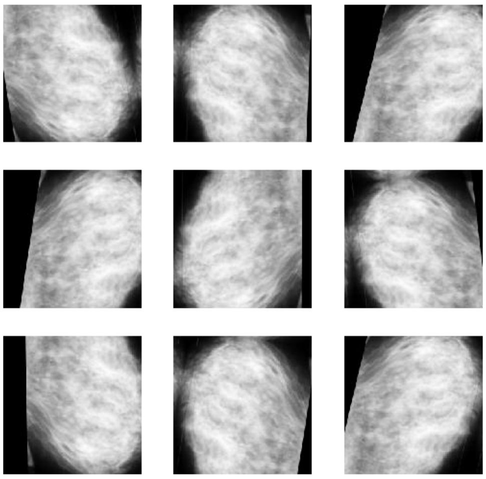

Figure 4. The results of the augmented image

2.3 Transfer Learning

Transfer Learning is the reuse of knowledge from previously trained models to perform new tasks. In this study, transfer learning was used as a feature extractor. This study used the VGG-16, VGG-19, and ResNet50 architectural models trained in ImageNet data by creating an architecture identical to VGG-16, VGG-19, and ResNet50, the fully connected layer and download weights. Keras provides several popular ImageNet models. All VGG-16, VGG-19, and ResNet50 weights have been trained using ImageNet datasets and can recognize colors, textures, etc. Therefore, the weights of VGG-16, VGG-19, and ResNet50 were used to extract features from all images in the dataset. The source code for the program is as follows.

1. Using the image feature extraction results from VGG-16, the model input value was taken from the last vector of the feature extraction process, namely 4096. The summary of the VGG-16 model is shown in Figure 5.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Model)                (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten_1 (Flatten)          (None, 25088)             0
_________________________________________________________________
dropout_1 (Dropout)          (None, 25088)             0
_________________________________________________________________
dense_1 (Dense)              (None, 7)                 175623
=================================================================
Total params: 14,890,311
Trainable params: 175,623
Non-trainable params: 14,714,688
_________________________________________________________________
```

Figure 5. Summary of the VGG-16 model

2. Using the image feature extraction results from VGG-19, the input value of the model was taken from the last vector of the feature extraction process, namely 4096. The summary of the VGG-19 model is shown in Figure 6.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg19 (Model)                (None, 7, 7, 512)         20024384
_________________________________________________________________
flatten_1 (Flatten)          (None, 25088)             0
_________________________________________________________________
dropout_1 (Dropout)          (None, 25088)             0
_________________________________________________________________
dense_1 (Dense)              (None, 7)                 175623
=================================================================
Total params: 20,200,007
Trainable params: 175,623
Non-trainable params: 20,024,384
_________________________________________________________________
```

Figure 6. Summary of the VGG-19 model

3. Using the image feature extraction results from ResNet50, the input value of the model was taken from the last vector of the feature extraction process, which is 2048. The ResNet50 model summary is shown in Figure 7.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
resnet50 (Model)             (None, 7, 7, 2048)        23587712
_________________________________________________________________
flatten_1 (Flatten)          (None, 100352)            0
_________________________________________________________________
dropout_1 (Dropout)          (None, 100352)            0
_________________________________________________________________
dense_1 (Dense)              (None, 7)                 702471
=================================================================
Total params: 24,290,183
Trainable params: 702,471
Non-trainable params: 23,587,712
_________________________________________________________________
```

Figure 7. Summary of the Resnet50 model

3 Results and Discussion

3.1 Results

The training stage was carried out using Adam's optimizer, which consisted of 30 epochs of model training (257 per epoch); validation steps = 97 were done by using Early Stopping setting, which provided a training process on a particular period where the model's performance improves on the validation dataset (Prechelt, 1998).

In this paper, the authors compared the differences that occurred using the VGG-16 Model, the VGG-19 Model, and the ResNet50 Model. The loss and accuracy values were calculated from the three models on the training dataset and the validation dataset by testing the 30 epoch model and using the Early Stopping setting. In the VGG-16 Model, the testing stopped at the 18th epoch; the VGG-19 Model testing stopped at the 21st epoch; the ResNet50 model tests stopped at the 8th epoch. The results of the loss and accuracy model values can be seen in Table 2 below.

Table 2. The result of loss and accuracy value

	Model VGG-16	Model VGG-19	Model ResNet50
Training data Loss	2,86	3,57	3,20
Validation data Loss	29,2	45,8	9,05
Training data Accuracy	0,86	0,86	0,86
Validation data Accuracy	0,64	0,61	0,71

Based on Table 2, among the three models that have gone through the training and validation stages, the most minor Training Data Loss score was generated by the VGG-16 model. However, they scored the same in Training Data Accuracy, while the RestNet50 Model performed best in Validation data Accuracy. The diagrams explaining the results can be seen in Figure 8, Figure 9, and Figure 10.

Figure 8. Diagrams of the training and validation accuracy and training and validation loss of the VGG-16 model

Figure 9. Diagrams of the training and validation accuracy and training and validation loss of the VGG-19 model

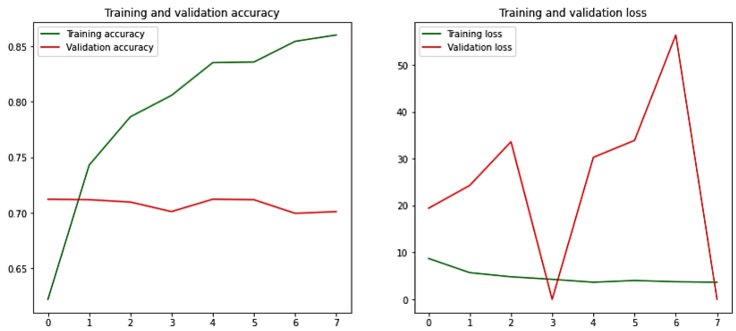

Figure 10. Diagrams of training and validation accuracy and training and validation loss model of ResNet 50

From the diagrams above, it can be seen that the training accuracy on each model tends to increase. In contrast, the validation accuracy on each model has a significant fluctuation except for the Resnet50 model. In the Training and Validation Loss diagram, the levels of training loss of all of the models tend to decrease while the levels of validation loss tend to fluctuate.

3.2 Discussion

After the results of training and validation were gained, the level of their accuracy could be calculated. The comparison of the accuracy level of the models is shown in Figure 11.

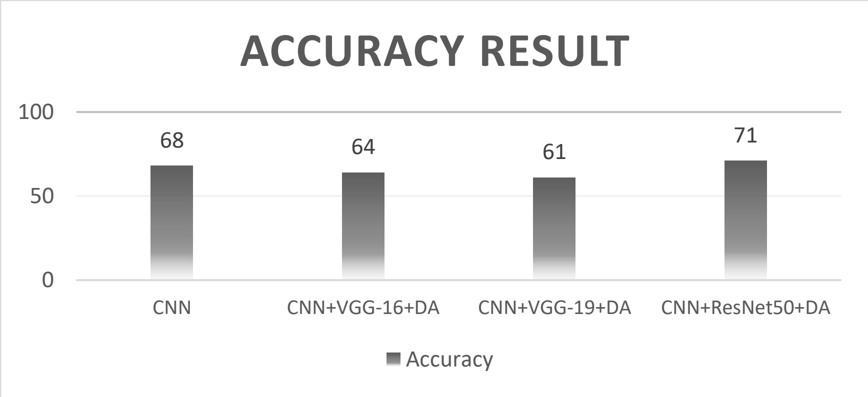

Figure 11. Accuracy results

From the results of the accuracy level above, it can be seen that the accuracy results from previous studies that only used CNN show an accuracy of 68%. The CNN test result with VGG-16 and VGG-19 architectural models and Data Augmentation decreased. The architectural model was less able to produce minimal data loss, while the most remarkable results are in system implementation using the ResNet50 Transfer Learning model plus Data Augmentation with an accuracy rate of 71%. The accuracy test results using ResNet50 increased because it was able to reduce the value of data loss as minimal as possible.

This present research yields an essential finding that by applying Transfer Learning and Data Augmentation in the Convolutional Neural Network method, optimal breast cancer detection accuracy in the ResNet50 architectural model can be increased. Furthermore, this research can be a meaningful reference for researchers seeking to conduct breast cancer detection research. Notwithstanding the advantages, this study also has several disadvantages due to the lack of apparent image differences and unbalanced data in each class.

4 Conclusion

Based on the study results, it can be concluded that the application of Transfer Learning and Data Augmentation can optimize the Convolutional Neural Network to improve the accuracy of breast cancer detection with mammography images obtained from MIAS 2012. The results of accuracy tests of early breast cancer detection process using Convolutional Neural Networks (CNN) by utilizing Data Augmentation and Transfer Learning show several findings: Model VGG-16, Model VGG-19, and ResNet-50 produced the same training accuracy rate of 86%. In contrast, for validation accuracy, Model ResNet-50 produced the highest level of accuracy (71%) compared to other models (VGG-16 = 64%, VGG-19 = 61%).

References

Ahmad, H. M., Ghuffar, S., & Khurshid, K. (2019). Classification of Breast Cancer Histology Images Using Transfer Learning. *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 328-332. doi: 10.1109/IBCAST.2019.8667221

Barman R., Deshpande S., Agarwal S., & Inamdar U. (2019, March). Transfer Learning for Small Dataset. *Proceedings of National Conference on Machine Learning*, 132-137.

Charan, S., Khan, M. J., & Khurshid, K. (2018, March). Breast cancer detection in mammograms using convolutional neural network. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1-5. doi: 10.1109/ICOMET.2018.8346384

Devare M. (2019, March). Challenges and Opportunities in High-Performance Cloud Computing. *Handbook of Research on the IoT, Cloud Computing, and Wireless Network Optimization*. Retrieved from https://www.researchgate.net/publication/329424286_Challenges_and_Opportunities_in_High_Performance_Cloud_Computing

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., & Agrawal, A. (2017). Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials*, 157, 322-330. doi: 10.1016/j.conbuildmat.2017.09.110

Mahmud, K. H., Adiwijaya, A., & Al Faraby, S. (2019). Klasifikasi Citra Multi-kelas Menggunakan Convolutional Neural Network [Multi-class Image Classification Using Convolutional Neural Network]. *e-Proceeding of Engineering*, 6(1).

Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problems. *2018 international interdisciplinary Ph.D. workshop (IIPhDW)*, 117-122. doi: 10.1109/IIPHDW.2018.8388338

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

Prechelt, L. (1998). Early stopping - But When? *Neural Networks: Tricks of the trade*, 55-69, 1524. Springer: Berlin, Heidelberg. doi: 10.1007/3-540-49430-8_3

Putra, D. K., Santoso, I., & Zahra, A. A. (2011). Identifikasi Keberadaan Kanker Pada Citra Mammografi Menggunakan Metode Wavelet Haar [Identification of Cancer Presence in Mammographic Image Using the Haar Wavelet Method]. *TRANSMISI, Jurnal Teknik Elektro*, 11(2), 100-106. doi: 10.12777/transmisi.11.2.100-106

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, *29*(9), 2352-2449. doi: 10.1162/neco_a_00990

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6, 60. doi:10.1186/s40537-019-0197-0

Tuna, A., Avdal, E. U., Yucel, S. C., Dal, N. A., Dicle, A., Ozkan, A., ... Degirmenci, M. (2014). Effectiveness of Online Education In Teaching Breast Self-Examination. *Asian Pacific Journal of Cancer Prevention (APJCP)*, 15(7), 3227-3. doi: 10.7314/APJCP.2014.15.7.3227

Vesal, S., Ravikumar, N., Davari, A., Ellmann, S., & Maier, A. (2018). Classification of Breast Cancer Histology Images Using Transfer Learning. *International Conference Image Analysis and Recognition*, 812-819. Springer: Cham.

Widowati, H. (2019). Kasus Kanker Payudara Paling Banyak Terjadi di Indonesia [Most Breast Cancer Cases Occur in Indonesia]. retrieved on 20 January 2021 in https://databoks.katadata.co.id/datapublish/2019/06/03/kasus-kankerpayudara-paling-banyak-terjadi-di-indonesia