Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Breast Cancer Classification Through CT Scan Using Convolutional Neural Network (CNN)

Anita Loi[1)*], Ruth N Panjaitan[2)], Saut Dohot Siregar[3)], Allwin M Simarmata[4)]
[1,2,3,4)]Universitas Prima Indonesia, Medan, Indonesia
[1)]anitaloi7373@gmail.com, [2)]panjaitan.ruth14@gmail.com, [3)]sautdohotsiregar@gmail.com, [4)] allwinsimarmata@unprimdn.ac.id,

Submitted : May 27, 2024 | **Accepted** : Jun 27, 2024 | **Published** : Jul 1, 2024

Abstract: A common disease suffered by Indonesian women is breast cancer. Early awareness of breast cancer is very important to minimize the negative impact and increase the chances of recovery for breast cancer patients. Breast cancer detection efforts using CT scan image technology. CT scan images provide a detailed picture of the internal structure of the breast, allowing the identification of pathological changes that may be early signs of breast cancer. The purpose of the study is to utilize CNN algorithm for breast cancer classification using CT scan images. The dataset used consists of three labels namely benign cancer, malignant cancer, normal. The three data sets consist of 1096 data. CNN is a type of algorithm in the field of artificial intelligence that has proven successful in pattern recognition on image data. The collected breast CT scan image dataset includes breast cancer and non-breast cancer cases. The data is used to train and test the CNN model. Furthermore, breast cancer classification through CT scans is carried out by applying the CNN method. The results of the research conducted obtained an accuracy of 97.26%. In Benign classification with precision 0.99 (99%), recall 0.96 (96%), f1-score 0.98 (98%), support 186, then Malignant classification with precision 93% or with points 0.93, recall 98% with points 0.98, and f1-score 96% with points 0.96, and support 202. The last is the normal classification with 99% precision with 0.99 points, 97% recall with 0.97 points, 98% f1-score with 0.93 points, and 269 support.

Keywords: Breast Cancer, CT Scan, Classification, Convolutional Neural Network

INTRODUCTION

A disease characterized by uncontrolled and abnormal proliferation of body cells is known as cancer. Cells with the potential to reproduce rapidly and spread to other tissues and organs of the body are known as tumors. Cancer arises when normal cells undergo genetic mutations that turn them into malignant cells. For example, breast cancer.

Breast cancer is a condition where abnormal cells grow in the breast area. Breast cancer can be detected through the observation of three key proteins that make up breast cancer. According to information from the North Sumatra Provincial Government article in 2022, there were 3,206 individuals with breast cancer recording the highest number of 393 cases, followed by lung cancer with 313 cases. Early awareness is a key factor in breast cancer treatment. Patients can get the right treatment to treat breast cancer by correctly classifying the images at an early stage. In the research conducted shows that the use of CT scans by combining deep learning techniques can improve accuracy and efficiency in classifying breast cancer. The use of CT scans is preferred due to its high accuracy and sensitivity compared to chest X-rays. A deep analysis approach to breast CT images supports radiologist diagnosis and second opinion. CNNs are often used to strengthen such image analysis.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13706 p-ISSN : 2541-044X

LITERATURE REVIEW

One type of artificial neural network developed from the Multilayer Perceptron (MLP) technique is the Convolutional Neural Network (CNN). CNNs can process images or two-dimensional data with high resolution. CNN can automatically extract hierarchical features from medical images, understanding patterns from simple to complex. CNN remains effective in recognizing breast cancer features despite changes in tumor location and size. The use of convolution filters allows CNN capture relevant visual information for breast cancer classification. CNNs can handle large medical images and support transfer learning, enabling adaptation of prepared models to large data for identifying breast cancer.

In previous research, the CNN algorithm achieved 85% accuracy in the classification of breast cancer in histopathology images. The goal is to classify breast cancer using histopathology images and focuses on analyzing biopsy tissue. For the research conducted using CT Scan images because previous research only used CNN, but was less accurate. Using CT Scan image technology can help identify structural changes in 2 breast tissue before clinical symptoms appear, allowing earlier identification of potential cancer problems. The advantage of this research is that using CT Scan image technology is faster in diagnosing, because CT Scan images can be generated quickly, reducing the waiting time for patients to get early information about their condition. This research focuses on breast classification for early patient awareness. By considering the problems described, it is concluded to make a research entitled "Breast Cancer Classification Through CT Scan Using Convolutional Neural Network (CNN)". Through the research carried out, the hope is to increase the speed and accuracy in the diagnosis of breast cancer and facilitate the medical work by using CT Scan image technology..

METHOD

The framework of the research, the approach used is experimental, where researchers apply the CNN algorithm to classify breast cancer in individuals. The research utilizes deep learning methods to process images, which will then be used as research subjects. Through deep learning, the image will be processed and classified using the CNN algorithm to determine whether the subject's image depicts malignant cancer, benign cancer, and normal.

The duration of the research lasted for 6 months, with testing of the proposed method and was carried out at the Prima Indonesia University Laboratory. The data used in the research comes from https://www.kaggle.com/datasets/anissaapril/kanker-payudara/ which has been collected by the research team.

Fig 1. Flowchart of Research Stages

In the context of the research conducted, the working procedure of image processing classification of breast cancer images with CT Scan images using the Convolutional Neural Network algorithm can be carried out in the following stages:

1. The dataset was downloaded from Kaggle and then loaded into a zip after which the zip was extracted and then grouped into each folder with the labels malignant, benign, and normal. Then the three folders were put together and split into 876 image sets for training, 220 image sets for testing.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13706 p-ISSN : 2541-044X

2. Pre-processing is performed on the image training data by applying rotation, contrast enhancement, and noise reduction on the image.
3. The next step involves inputting images for processing so as to classify individuals suffering from breast cancer using the CNN model.
4. The layers in the model function as classifiers by extracting features and evaluating probabilities on objects in the image test data to get the final result of the model performance. 5. The last step shows how many models successfully classify breast cancer in individuals and normal individual breasts by visualizing in the form of a confusion matrix

RESULT

In this section, the researcher will explain the results of the research obtained. Researchers can also use images, tables, and curves to explain the results of the study. These results should present the raw data or the results after applying the techniques outlined in the methods section. The results are simply results; they do not conclude.

Data Preparation

This part of the stage aims to provide a visual understanding of the image dataset to be applied for the experiments conducted. In this research, one example image from each class (label) in the dataset will be taken to provide a visual representation of the various categories or classes in the dataset. The author provided the data by taking a dataset available from open source at https://www.kaggle.com/datasets/anissaapril/kankerpayudara/. The dataset used is about 1096 data consisting of benign, malignant, and normal cancer labels. See the following image:

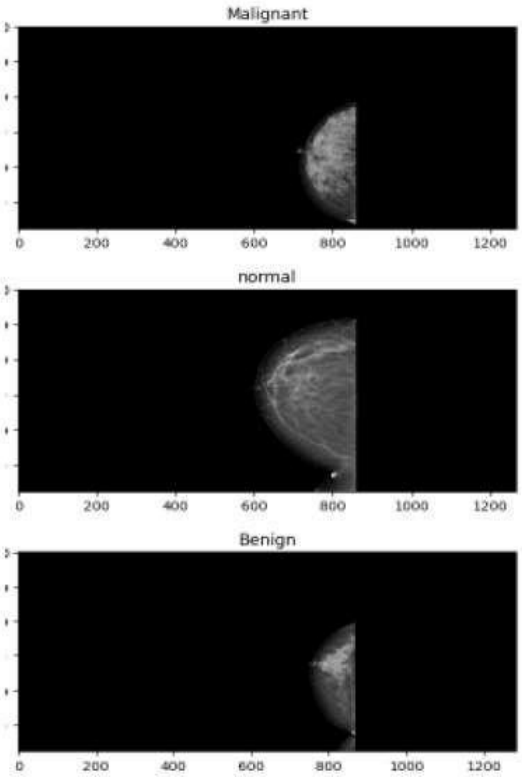

Fig. 2 database

Testing

In this work, a classification model will be created using CNN, which is recommended for image analysis. However, it is necessary to be aware of the possibility of overfitting in CNN models, so the use of activation functions is necessary to overcome it. CNN, 9 which stands for Convolutional Neural network, belongs to the category of deep neural networks and is commonly used in image analysis. In this CNN model, there are several layers that are used to classify the model. First create a first

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

convolution layer, with 32 3x3 filters. Each filter is used to extract important features from the input image. The ReLU activation function is used to activate the neurons in this layer. Input shape indicates the size of the input image, which is a 150x150 pixel image with 3 RGB color channels. Next there is a max-pooling layer that reduces the image resolution to half. Max-pooling helps to reduce the image dimension and improve computational efficiency.

This process is repeated three times with different convolution and max-pooling layers, with 65, 128 and 512 filters respectively. This aims to extract increasingly complex features as the process progresses. After a series of convolution and max-pooling layers, we have the Flatten layer which converts the 2D tensor into a 1D vector. This layer is required before going to the fully connected layer. Then, we have a dense (fully connected) layer with 512 units and ReLU activation as shown below.

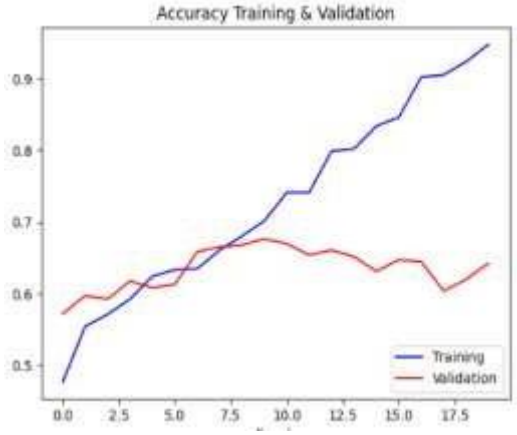

Fig. 3 Training and Testing Accuracy Graph

Fig. 4 Training and Testing Loss Graph

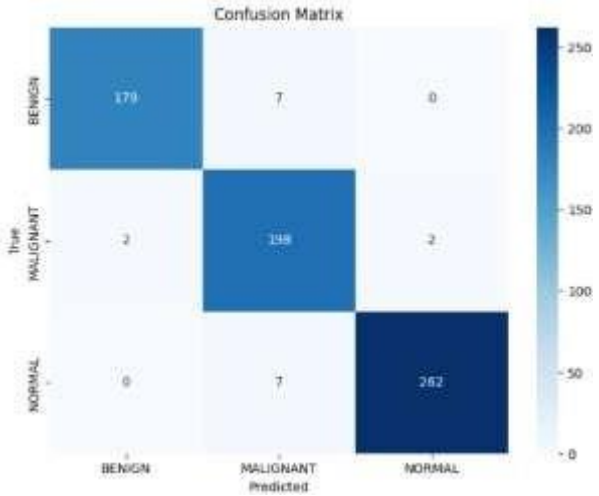

Fig. 5 Confusion Matrix of Breast Cancer Classification

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The displayed images reflect the results of the Confusion Matrix of breast cancer classification. Confusion Matrix is a standard evaluation table in deep learning to measure model performance. This matrix provides information about the extent of the model's ability to classify data into the right category. The following is a table of accuracy and how many are classified by the Convolutional Neural Network (CNN) model.

Table. 1 CNN Training and Validation Accuracy

Model	Training Accuracy	Validation Accuracy
CNN	97,26%	70%

The classification results show that there are 179 benign images that have been successfully classified, while for malignant cancer there are 198 images that have been successfully classified, then finally there are 262 normal images that have been classified correctly. The following is a table of precision, recall, f-1 score, support for each label

Table. 2 Image Classification Results Based on Labels

Label	**precision**	**recall**	**F1-score**	**Number of Image Classifications**
Benign	0.99	0.96	0.98	179
Malignant	0.93	0.98	0.96	198
Normal	0.99	0.97	0.98	262

DISCUSSIONS

In the research conducted by R. Rokhana et al. achieved 95.3% accuracy in using CNN as a method for fracture detection in B-Mode Ultrasonic images. Then in the research of F. A. A. Harahap et al classified brain tumors with MobileNetV2 CNN through transfer learning, with an average accuracy of 86%. However, this result is still below the previous accuracy which reached 94%. Then in the research of M. Lestandy et al 11 detected breast cancer tissue with CNN, achieving an average accuracy of 80% on a dataset of 277,524. In research conducted by M. A. Hanin et al. classify skin diseases with CNN, achieving 96.53% accuracy. The results of this study achieved 97.26% accuracy, possibly due to the use of customized CNN methods, larger and more representative datasets, and sophisticated data preprocessing techniques, including data normalization and data augmentation, as well as careful hyperparameter tuning. After conducting research using CT scan images with the CNN method, several significant changes are expected. First, this research is expected to improve the accuracy in breast cancer diagnosis due to the more detailed information provided by CT Scan images. In addition, the use of CT Scan images is also expected to speed up the diagnosis process as images can be generated quickly, reducing the waiting time for patients to get early information about their condition. With this research, it is hoped that there will be an understanding of the importance of classifying breast cancer, and medical professionals are expected to be more assisted in the process of diagnosing and treating breast cancer by using CT Scan image technology and the CNN method. Overall, it is hoped to have a good impact on efforts to prevent, diagnose, and treat breast cancer more effectively. A visualization image of the model results classifying breast cancer based on the label as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

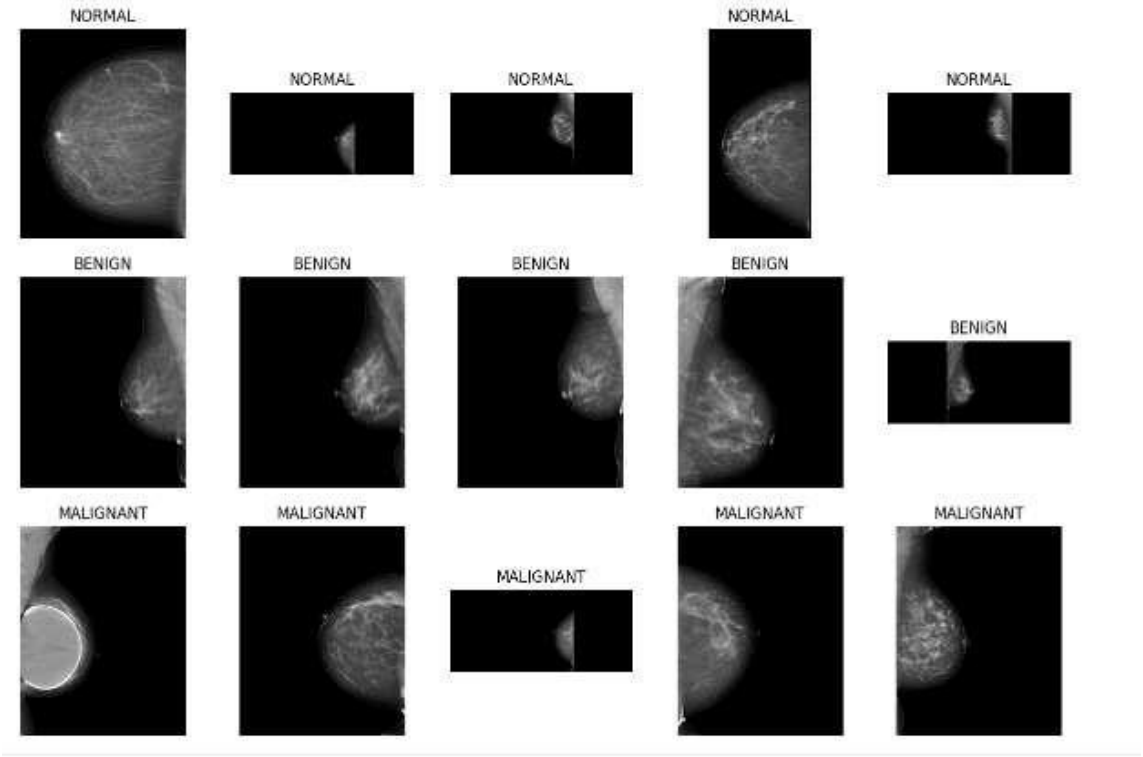

Fig. 6 Breast Cancer and Normal Classification Results

From the visualization of breast cancer classification results visualized in Figure 3.5, it is found that the distribution of images in each class is relatively balanced, but there are variations in the visual representation of each class. The model successfully classifies the images based on the labels correctly so that the images appear according to the labels.

CONCLUSION

In a study conducted with CNN algorithm to classify breast cancer through CT scan images. The aim is to provide support to medical personnel by providing tools that facilitate accurate diagnosis of breast conditions, as well as facilitating appropriate treatment options for breast care. Based on the findings from the results so far, it is concluded that:

1. The result of 97% accuracy research
2. Providing medical personnel with tools to improve accuracy in diagnosing and treating breast cancer
3. CNN algorithm in classifying breast cancer through CT scan images can be an innovative step in medical service delivery.

REFERENCES

M. D. Hapsari and E. I. Setyawan, Potensi Antikanker Ekstrak Buah Jamblang (Syzygium cumini L.) sebagai Bahan Pangan Fungsional, Vol. 2, PP. 356–368, 2023.

Z. Momenimovahed and H. Salehiniya, Epidemiological characteristics of and risk factors for breast cancer in the world, *Breast Cancer Targets Ther.*, Vol. 11, PP. 151–164, 2019, doi: 10.2147/BCTT.S176070.

M. Pegram, C. Jackisch, and S. R. D. Johnston, Estrogen/HER2 receptor crosstalk in breast cancer: combination therapies to improve outcomes for patients with hormone receptor-positive/HER2-positive breast cancer, *npj Breast Cancer*, Vol. 9, no. 1, 2023, doi: 10.1038/s41523-023-00533-2.

Pemprov, Data Kanker Payudara 2022 Sumut, 2023. https://sumutprov.go.id/artikel/artikel/peringatan-hari-ibu-ke-99-pemprov-sumut-beri-edukasi-penanganan-kanker

*name of corresponding author

S. L. Anwar *et al.*, Risk factors of distant metastasis after surgery among different breast cancer subtypes: A hospital-based study in Indonesia, *World J. Surg. Oncol.*, Vol. 18, no. 1, PP. 1–16, 2020, doi: 10.1186/s12957-020-01893-w.

C. B. S. Maior, J. M. M. Santana, I. D. Lins, and M. J. C. Moura, Convolutional neural network model based on radiological images to support COVID-19 diagnosis: Evaluating database biases, *PLoS One*, Vol. 16, No. 3 March, 2021, doi: 10.1371/journal.pone.0247839.

D. Singh, V. Kumar, Vaishali, and M. Kaur, Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks, *Eur. J. Clin. Microbiol. Infect. Dis.*, Vol. 39, no. 7, PP. 1379–1389, 2020, doi: 10.1007/s10096-020-03901-z..

S. Serte and H. Demirel, "Deep learning for diagnosis of COVID-19 using 3D CT scans," *Comput. Biol. Med.*, Vol. 132, No. October 2020, PP. 104306, 2021, doi: 10.1016/j.compbiomed.2021.104306..

M. Sahu and R. Dash, *A survey on deep learning: Convolution neural network* (CNN), Vol. 153, Springer Singapore, 2021. doi: 10.1007/978-981-15-6202-0_32.

M. W. Muhammad Afrizal Amrustian, "mplementasi Metode Convolutional Neural Network untuk Klasifikasi Breast Cancer pada Citra Histopatologi, *Media Inform.*, Vol. 7, No. 2, PP. 10, 2023, doi: 10.30865/mib.v7i1.5194..

M. Harahap, S. K. Anjelli, W. A. M. Sinaga, R. Alward, J. F. W. Manawan, and A. M. Husein, Classification of diabetic foot ulcer using convolutional neural network (CNN) in diabetic patients, *J. Infotel*, Vol. 14, No. 3, PP. 196–202, 2022, doi: 10.20895/infotel.v14i3.796.

R. Rokhana *et al.*, Convolutional Neural Network untuk Pendeteksian Patah Tulang Femur pada Citra Ultrasonik B–Mode, *J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 8, No. 1, PP. 59, 2019, doi: 10.22146/jnteti.v8i1.491.

M. Lestandy, Deteksi Dini Kanker Payudara Menggunakan Metode Convolution Neural Network (CNN), *Inspir. J. Teknol. Inf. dan Komun.*, Vol. 12, No. 1, PP. 65, 2022, doi: 10.35585/inspir.v12i1.2667.

F. A. A. Harahap, A. N. Nafisa, E. Purba, and ..., Implementasi Algoritma Convolutional Neural Network Arsitektur Model MobileNetV2 dalam Klasifikasi Penyakit Tumor Otak Glioma, Pituitary dan Meningioma, *J. Teknol. …*, Vol. 5, No. 1, PP. 53–61, 2023, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/article/view/234

M. A. Hanin, R. Patmasari, and R. Y. Nur, Sistem Klasifikasi Penyakit Kulit Menggunakan Convolutional Neural Network (CNN) Skin Disease Classification System Using Convolutional Neural Network (CNN), *e-Proceeding Eng.*, Vol. 8, No. 1, PP. 273–281, 2021.

R. Andre, B. Wahyu, and R. Purbaningtyas, Klasifikasi Tumor Otak Menggunakan Convolutional Neural Network Dengan Arsitektur Efficientnet-B3, *J. IT*, Vol. 11, No. 3, PP. 55–59, 2021, [Online]. Available: https://jurnal.umj.ac.id/index.php/just-it/index

J. Kecerdasan, T. Informasi, D. H. Firdaus, B. Imran, L. D. Bakti, and E. Suryadi, Klasifikasi Penyakit Katarak Pada Mata Menggunakan Metode Convolutional Neural Network (CNN) Berbasis Web Web-Based Classification of Cataract in the Eyes Using Convolutional Neural Network (CNN) Method, *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 1, no. 3, pp. 18–26, 2022.

*name of corresponding author