

## Article

# Development of a Mammography Calcification Detection Algorithm Using Deep Learning with Resolution-Preserved Image Patch Division

Miu Sakaida <sup>1</sup>, Takaaki Yoshimura <sup>2,3,4,5</sup> , Minghui Tang <sup>5,6</sup>, Shota Ichikawa <sup>7,8</sup>  and Hiroyuki Sugimori <sup>4,5,9,\*</sup> 

- <sup>1</sup> Graduate School of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan
- <sup>2</sup> Department of Health Sciences and Technology, Faculty of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan
- <sup>3</sup> Department of Medical Physics, Hokkaido University Hospital, Sapporo 060-8648, Japan
- <sup>4</sup> Global Center for Biomedical Science and Engineering, Faculty of Medicine, Hokkaido University, Sapporo 060-8648, Japan
- <sup>5</sup> Clinical AI Human Resources Development Program, Faculty of Medicine, Hokkaido University, Sapporo 060-8648, Japan
- <sup>6</sup> Department of Diagnostic Imaging, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo 060-8638, Japan
- <sup>7</sup> Department of Radiological Technology, School of Health Sciences, Faculty of Medicine, Niigata University, Niigata 951-8518, Japan
- <sup>8</sup> Institute for Research Administration, Niigata University, Niigata 950-2181, Japan
- <sup>9</sup> Department of Biomedical Science and Engineering, Faculty of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan
- \* Correspondence: sugimori@hs.hokudai.ac.jp; Tel.: +81-11-706-3410

**Abstract:** Convolutional neural networks (CNNs) in deep learning have input pixel limitations, which leads to lost information regarding microcalcification when mammography images are compressed. Segmenting images into patches retains the original resolution when inputting them into the CNN and allows for identifying the location of calcification. This study aimed to develop a mammographic calcification detection method using deep learning by classifying the presence of calcification in the breast. Using publicly available data, 212 mammograms from 81 women were segmented into  $224 \times 224$ -pixel patches, producing 15,049 patches. These were visually classified for calcification and divided into five subsets for training and evaluation using fivefold cross-validation, ensuring image consistency. ResNet18, ResNet50, and ResNet101 were used for training, each of which created a two-class calcification classifier. The ResNet18 classifier achieved an overall accuracy of 96.0%, mammogram accuracy of 95.8%, an area under the curve (AUC) of 0.96, and a processing time of 0.07 s. The results of ResNet50 indicated 96.4% overall accuracy, 96.3% mammogram accuracy, an AUC of 0.96, and a processing time of 0.14 s. The results of ResNet101 indicated 96.3% overall accuracy, 96.1% mammogram accuracy, an AUC of 0.96, and a processing time of 0.20 s. This developed method offers quick, accurate calcification classification and efficient visualization of calcification locations.

**Keywords:** deep learning; mammography; image classification



**Citation:** Sakaida, M.; Yoshimura, T.; Tang, M.; Ichikawa, S.; Sugimori, H. Development of a Mammography Calcification Detection Algorithm Using Deep Learning with Resolution-Preserved Image Patch Division. *Algorithms* **2023**, *16*, 483. <https://doi.org/10.3390/a16100483>

Academic Editors: Christian Mata, Raul Benitez and Gilberto Ochoa-Ruiz

Received: 15 September 2023

Revised: 17 October 2023

Accepted: 18 October 2023

Published: 18 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Breast cancer is a cancer of the mammary gland tissues and most often arises from the ducts, with some cases arising from the lobules and other tissues. Cancer cells may metastasize to lymph nodes or other organs, resulting in damage. Breast cancer is the leading cause of cancer in women and the second leading cause of cancer-related mortality worldwide [1]. Mammography, magnetic resonance imaging (MRI), and ultrasonography are common methods used to diagnose breast cancer [2].

In this study, we focus on mammography, which is an essential tool for breast cancer diagnosis. This study aims to develop a method for detecting calcification in mammography

using deep learning techniques. We base our classification on the presence or absence of calcification in mammography. Deep learning uses a neural network called CNN, which has a limitation on the number of pixels in its input. However, there are a large number of pixels included in mammography images, and compressing the images results in the loss of information of small lesions, such as microcalcifications [3]. If an entire image is input and classified based on the presence or absence of calcification, it becomes challenging to determine where calcification is located. There are no previous studies that analogize the location of calcification solely on the basis of image classification. In this study, the accuracy of the classifier as well as the accuracy and processing time for each mammography image were calculated and investigated. There are no studies that have been found that include these factors. Therefore, our objective is to divide the image into patches, enabling input into a CNN with the original resolution. This allows us to classify the presence or absence of calcification in each patch and display the integrated image to show the location of calcification in the original image.

## 2. Related Work

Mammography is one of the diagnostic imaging modalities for breast cancer. It is an effective screening tool for breast cancer diagnosis and is the only method proven to reduce breast cancer mortality [4]. According to a recent report [5], it is one of the most effective ways to detect breast cancer in its early stages. Mammography detects approximately 80% to 90% of breast cancer cases in asymptomatic women and has been reported to reduce breast cancer mortality by 38–48% among those screened [6,7]. This also has been reported [8–10] in recent years to reduce mortality by 30% or mortality by 20%. Mammography is mainly used for qualitative diagnosis and to observe masses, focal asymmetric shadows, calcifications, and structural abnormalities. Calcification is a common finding in mammography. It is caused by the deposition of calcium oxalate and calcium phosphate in the breast tissue and appears as bright spots on mammography [11,12]. The distribution of calcification is useful for differentiating benign from malignant disease. Diffuse or scattered calcifications are generally benign, while zonal or linear calcifications are suspected to be malignant [13]. Because clustered microcalcifications are found in 30–50% of cases of cancer diagnosed by mammography, the detection, evaluation, and classification of calcifications are important [6,7]. However, microcalcifications can be difficult to accurately detect and diagnose due to their size, shape, and heterogeneity in the surrounding tissue [6].

Computer-aided diagnosis (CAD) systems have been used in mammography to assist in the reading of mammograms and were developed to detect abnormal breast tissue and reduce the number of false-negative results for breast cancer detected by radiologists [14]. The sensitivity and specificity of CAD for all breast lesions have been reported to be 54% and 16%, respectively. CAD can easily misclassify parenchyma, connective tissue, and blood vessels as breast lesions, vascular calcifications, and axillae as microcalcifications. Because of its low specificity and high false-positive rate, CAD should not be used alone in mammography for breast examinations [15].

Deep learning technologies, specifically convolutional neural networks (CNNs), have recently been used in a variety of fields, including medical imaging [16,17]. Deep learning techniques have a wide range of applications, including classification [18,19], object detection [20,21], semantic segmentation [21–23], and regression [24–26]. Deep learning techniques have also been applied to breast diagnostics. For MRI, studies have used U-Net for breast mass segmentation [27] and ResNet18 pretrained on ImageNet for classification of benign and malignant breast tumors [28]. These studies are limited due to the small number of available breast MRI data and often rely on transfer learning. Similarly, deep learning methods for breast ultrasound also typically use transfer learning and pretraining with ImageNet due to the small number of available breast ultrasound image training sets [18]. Deep learning models in mammography have been used not only for the detection of potential malignancies but also for tasks like risk stratification, lesion classification, and

prognostic evaluation based on mammographic patterns [29,30]. Researchers have also explored the potential of employing transfer learning to overcome challenges associated with the limited availability of supervised image training mammography data [31].

### 3. Materials and Methods

#### 3.1. Subjects

In this study, we used image data from the Categorized Digital Database for low-energy and subtracted contrast-enhanced spectral mammography (CDD-CESM) images available at the Cancer Imaging Archive (<https://www.cancerimagingarchive.net/> accessed on 15 September 2023). These public data were converted from DICOM format to lossless JPEG with an average pixel size of  $2355 \times 1315$  pixels. Of the 2006 mammography image datasets of 326 patients included in these publicly available data, we used 212 low-energy images from 81 patients with more than 50 images in both the left and right imaging orientations for this study. Table 1 provides a detailed description of the image data used.

**Table 1.** Number of mammography images in this study.

	Right		Left	
	CC <sup>a</sup> Image	MLO <sup>b</sup> Image	CC Image	MLO Image
No. of images	53	51	52	56

<sup>a</sup> Craniocaudal; <sup>b</sup> mediolateral oblique.

#### 3.2. Data Preprocessing

The data preprocessing in this study was performed using in-house MATLAB software (MATLAB 2023a; The MathWorks, Inc., (Natick, MA, USA)) and a desktop computer with an NVIDIA RTX A6000 graphics card (Nvidia Corporation, Santa Clara, CA, USA). Table 2 shows the specifications of the computers used for image processing in this study.

**Table 2.** Software and equipment used in this study.

Environment	Contents
Software	MATLAB 2023a (Mathworks)
OS	Windows 11
CPU	Intel core i9-10980XE 3.0 GHz
GPU	NVIDIA RTX A6000 48 GB x 2
Memory	DDR4 2933 Quad-Channel 64 GB

In this study, as shown in Figure 1, high-resolution mammography was divided into small patches, and images that preserved the resolution of the original images were used for image classification to identify whether calcification was present in the images within the patches using a deep learning image classifier. In this study, we used a pretrained ResNet for transfer learning in ImageNet. To achieve the pixel size of  $224 \times 224$  pixels, which is the input image size for ResNet18, ResNet50, and ResNet101, 212 mammography images were divided into  $224 \times 224$ -pixel size. The right breast image was divided into  $224$ -pixel patches by starting from the bottom right side, and the left breast image was divided into  $224$ -pixel patches by starting from the bottom left side; the missing pixels were filled with zeros. The 15,049 patches were reviewed visually by four radiologists with 24, 14, 6, and 1 years of experience. These radiologists were certified by the Japan Central Organization on Quality Assurance of Breast Cancer Screening. As shown in Table 3, the presence or absence of calcification was classified by consensus.

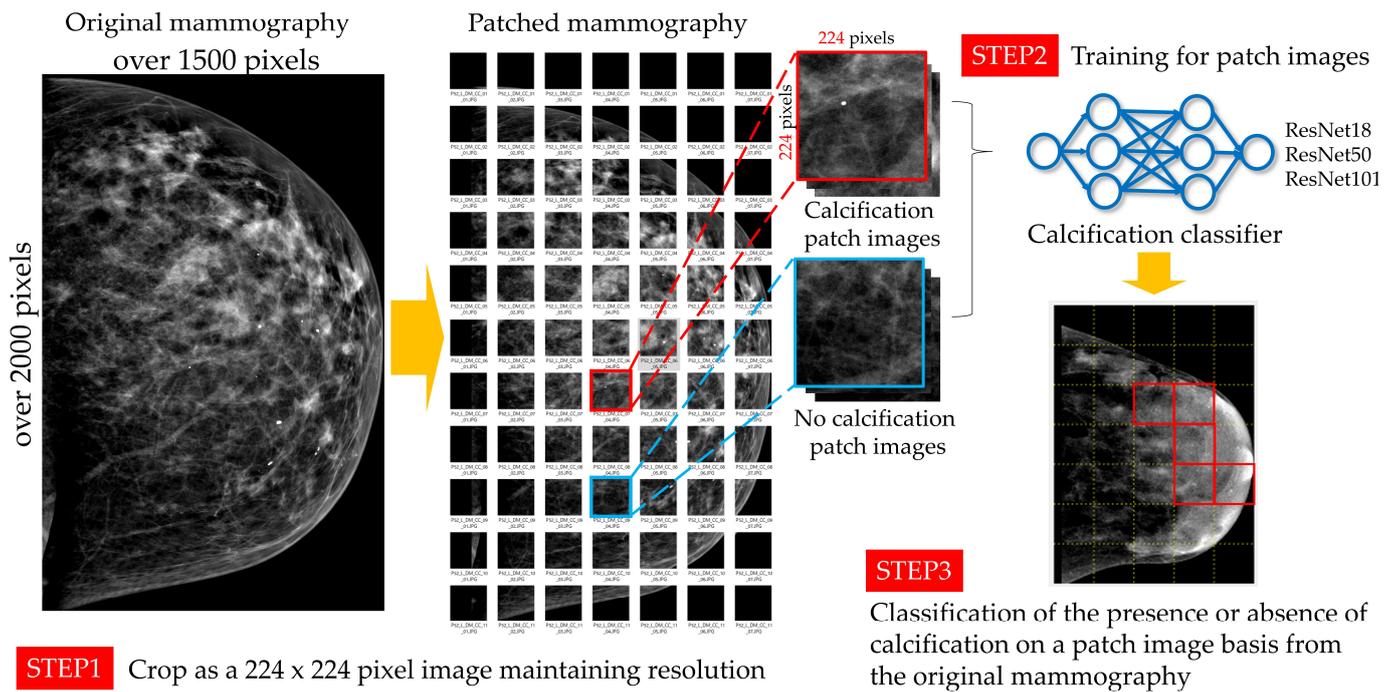


Figure 1. Schematic diagram of the training and evaluation from patch imaging of mammograms.

Table 3. Number of classifications of the presence or absence as visual calcification.

Calcification	No. of Images
Yes	1029
No	14,020

We performed fivefold cross-validation by dividing the generated patches into five subsets, A through E. The number of mammographic images in each subset was kept similar to avoid including mammographic images of the same patient in different subsets. Table 4 shows a summary of the subset structure.

Table 4. Detail of images and calcification in patches within subset.

Subset	No. of Patients	No. of Mammography Images	Calcification in the Patch	
			Yes	No
A	16	43	194	3352
B	16	42	131	3001
C	17	41	238	2329
D	15	43	196	2989
E	17	43	270	2349

Five datasets, folds 1 through 5, were created based on the subsets, with the training data for fold 1 being subsets B through E and the test data being subset A. The other datasets were created in the same way. Table 5 shows each of the datasets.

Since a classifier created by training on unbalanced training data may be biased toward classes with a large number of images in the training data, a model created on training data with many images without calcification may be more easily classified as having no calcification, leading to missing calcification [32]. Therefore, we performed data augmentation on the training data in each dataset to make the number of images with and without calcification equivalent. First, the training data patches with calcification were rotated by 90°, 180°, and 270°. Then, the original and rotated patch was flipped. Next,

these patches were expanded by increasing the pixel values by a factor of 0.5 and 0.75. Training data without calcification were expanded by applying only the flipping process. We did not perform shearing or rotations of less than 90° in this study. Table 6 shows the changes in imaging data that occurred after data augmentation.

**Table 5.** Number of images in each fold.

	No. of Images Used for Training		No. of Images Used for Testing	
	Without Calcification	With Calcification	Without Calcification	With Calcification
Fold 1	10,668	835	3352	194
Fold 2	11,019	898	3001	131
Fold 3	11,691	791	2329	238
Fold 4	11,031	833	2989	196
Fold 5	11,671	759	2349	270

**Table 6.** Number of images after data augmentation in training data.

	Without Calcification		With Calcification	
	Original	Augmented	Original	Augmented
Fold 1	10,668	21,336	835	20,040
Fold 2	11,019	22,038	898	21,552
Fold 3	11,691	23,382	791	18,984
Fold 4	11,031	22,062	833	19,992
Fold 5	11,671	23,342	759	18,216

### 3.3. Image Training

Mammography images were input in patch form, and a classifier was created to perform two-class classification with respect to the presence or absence of calcification. In order to compare the accuracy of different CNN structures in training the patched images, we used ResNet18, ResNet50, and ResNet101 for the training models, and each CNN was used to create a classifier that performs two-class classification of the presence or absence of calcification. The parameters used to create the classifiers are as follows: The optimizer was the stochastic gradient descent and momentum optimization, the batch size for the number of training samples was 128, the number of epochs was 10, and the initial learning rate was 0.001. The learning rate drop factor was 0.3, the learning rate drop period was 1, the L2 regularization was 0.005, and the momentum was 0.9. In order to evaluate the difference in accuracy due to CNN structure, we did not optimize the training parameters in this study, and the training conditions were identical.

### 3.4. Evaluation of Created Models

#### 3.4.1. Classification Accuracy

The confusion matrix is used to calculate the accuracy of the classifier as a whole. This matrix classifies the images into two classes: one with calcification and one without calcification. In addition, it determines the accuracy of each mammographic image. The sensitivity, specificity, F1 score, positive predictive value (PPV), and negative predictive value (NPV) of the model can be calculated using true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Their definitions and calculation formulas are shown below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{F1 score} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

In this study, we performed fivefold cross-validation and then averaged the results of the five models to calculate the accuracy, sensitivity, specificity, PPV, and NPV of the classifiers created by each CNN. In addition, receiver–operating characteristic (ROC) analysis was performed, and the area under the curve (AUC), which is the area under the ROC curve, was calculated by averaging the results of the five models and the AUC of the classifier created by each CNN.

### 3.4.2. Time for Classification

We calculated the time taken to classify patches generated from a single mammography image by the presence or absence of calcification. For the trained model, the processing time was defined as the time required to infer all patched images per mammogram in the test data.

### 3.4.3. Accuracy of Each Whole Mammogram

Considering that a mammogram is divided into  $N$  patches, the accuracy of each whole mammogram image can be expressed by the following equation using the number of TP patches and the number of TN patches.

$$\text{Accuracy of each whole mammogram} = \frac{\text{TP} + \text{TN}}{N} \times 100[\%]$$

In this study, we calculated accuracy for each of the 212 mammograms. We subsequently calculated the average accuracy of each CNN for each of the mammograms.

## 4. Results

### 4.1. Accuracy of the Created Classifier

Table 7 shows the accuracy of the classifier for the two-class classification of mammography images in each CNN, with and without calcifications. The classifier trained with ResNet50 had the highest values for all CNNs.

**Table 7.** Accuracy of classifiers in each CNN.

CNN <sup>a</sup>	Sensitivity	Specificity	F1 Score	PPV <sup>b</sup>	NPV <sup>c</sup>	Accuracy
ResNet18	73.4	97.8	71.4	70.5	97.8	96
ResNet50	76.2	98.1	74.7	74.4	98.1	96.4
ResNet101	75.6	98	73.8	72.9	98	96.3

<sup>a</sup> Convolutional neural network; <sup>b</sup> positive predictive value; <sup>c</sup> negative predictive value.

### 4.2. ROC Analysis

Table 8 shows the AUC of the two-class classifier for the presence or absence of calcification using mammography images input as patches in each CNN. The classifier trained by ResNet50 had the highest value.

**Table 8.** AUC of classifiers in each CNN.

CNN <sup>a</sup>	AUC <sup>b</sup>
ResNet18	0.956
ResNet50	0.963
ResNet101	0.961

<sup>a</sup> Convolutional neural network; <sup>b</sup> area under the curve.

#### 4.3. Time for Classification

Table 9 shows the inference time taken by each CNN to classify a patch generated from a single mammogram image as calcification or no calcification.

**Table 9.** Inference time by classifier for each CNN.

CNN <sup>a</sup>	Time for Classification [s]
ResNet18	0.072
ResNet50	0.136
ResNet101	0.203

<sup>a</sup> Convolutional neural network.

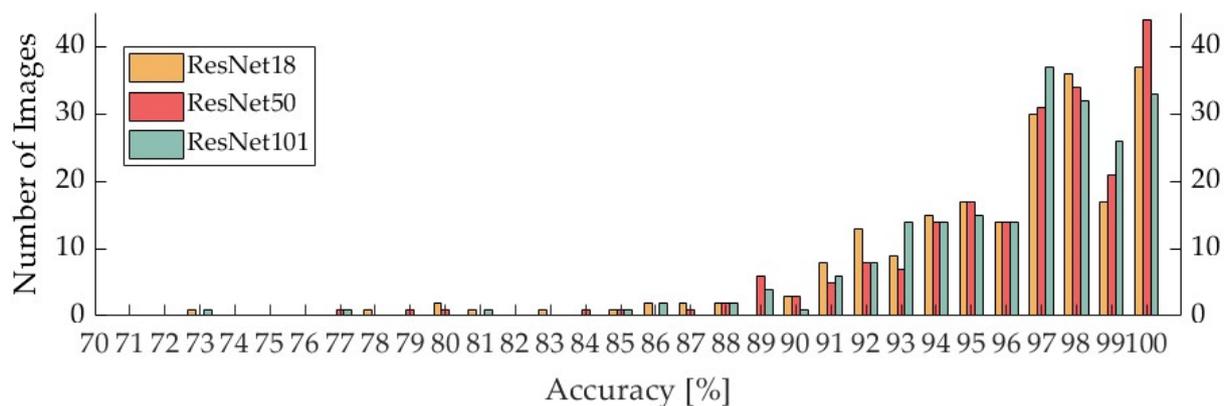
#### 4.4. Accuracy for Each Whole Mammogram

Table 10 shows the average accuracy per mammogram for each CNN, and Figure 2 summarizes the accuracy per mammogram for the classifiers trained on ResNet18, ResNet50, and ResNet101.

**Table 10.** Average accuracy for each mammography image in each CNN.

CNN <sup>a</sup>	Average Accuracy [%]
ResNet18	95.8
ResNet50	96.3
ResNet101	96.1

<sup>a</sup> Convolutional neural network.

**Figure 2.** Accuracy of one mammography image at each CNN.

The average accuracy of each mammogram image for the classifiers trained on ResNet18, ResNet50, and ResNet101 was 95.8%, 96.3%, and 96.1%, respectively, with the ResNet50 trained model having the highest value.

## 5. Discussion

The results in Section 4.1 demonstrate that the overall accuracy of the classifiers was 96.0% for the ResNet18 classifier, 96.4% for the ResNet50 classifier, and 96.3% for the ResNet101 classifier. The accuracy of the ResNet50 classifier was 96.4%, and that

of the ResNet101 classifier was 96.3%. Alternatively, in a previous study of two-class classification of the presence or absence of calcification, the authors reported that the most accurate classifier was 96% [33]. The findings obtained in the current study are comparable with those of the previous research. However, the ResNet18, ResNet50, and ResNet101 classifiers had sensitivities of 73.4%, 76.2%, and 75.6%, respectively, whereas the highest sensitivity in the previous study [33] was 98%. Compared with the results of the prior report, the sensitivity obtained in our study was relatively lower. One of the reasons for this is that in previous studies, patches were extracted from mammographic images in such a way that there was slight overlap in the top, bottom, left, and right sides of the image, which was reported to reduce the FN rate (1—sensitivity) [33,34]. A recent study [32] using the ResNet network showed that deep learning tools aid radiologists in mammogram-based breast cancer diagnosis. The study addressed class imbalance issues in training data by testing techniques like class weighting, over-sampling, under-sampling, and synthetic lesion generation. The results indicated a bias toward the majority class due to class imbalance, which was partially mitigated by standard techniques but did not significantly improve AUC-ROC. Synthetic lesion generation emerged as a superior method, particularly for out-of-distribution test sets. The morphology and distribution of the calcifications in the breast vary. Obvious benign calcifications include skin and vascular calcification, fibroadenoma calcification, calcification associated with ductal dilatation, round calcification, central translucent calcification, calcareous calcification, calcareous milk calcification, suture calcification, and heterotopic calcification. The morphology of calcification, which must be distinguished between good and bad, includes microcircular, pale and indistinct, pleomorphic, fine-linear and fine-branched, diffuse, scattered, regional, clustered, linear, and zonal distribution. Because of the variations in the morphology and distribution of calcifications, we assume that the diversity of features was not captured in this study, resulting in a low sensitivity. In addition, FN patches had a smaller area of calcification as compared with TP patches. Most coarse calcifications were correctly classified as calcified. Other calcifications were often judged as calcified when a patch contained many calcifications or slightly larger calcifications, but they were incorrectly classified as no calcified when a patch contained only a few small calcifications, such as one or two small calcifications (i.e., when the area of calcification in the patch was small). In addition, FN patches were often whiter than TP patches, especially when the entire patch was white and each patch contained only a few small calcifications (i.e., one or two). Conversely, FP patches included some patches with white patches. In addition, as with the FN patches, patches that were entirely white were often incorrectly classified as having calcification. These results suggest that calcification is often misjudged because it is difficult to distinguish between calcification and noncalcification when patches are whitish because calcification is depicted with high brightness. Therefore, we believe that the use of thresholding can improve the accuracy of the classifier by reducing the number of FN and FP patches via the extraction of areas of particularly high brightness. With regard to the point that a small area of calcification in a patch is often incorrectly classified as no calcification, in this study, we divided the mammography image into patches of  $224 \times 224$  pixels to create the input image size for ResNet; however, it is possible to divide the mammography image into patches of a smaller matrix size and to use the generated patches as the input image size for ResNet. However, if the patches are divided into a smaller matrix size and the generated patches are enlarged to  $224 \times 224$  pixels by resizing, the area of calcification in the patches will be larger and the accuracy may be improved.

We will next discuss the AUC of the created classifiers. The AUC of the classifiers created by all CNNs was greater than 0.95, a value that is considered to be very accurate and excellent. Regarding the processing time for classifying the presence or absence of calcification for the created classifiers, ResNet18, ResNet50, and ResNet101 took the shortest processing times, in that order. This is consistent with the fact that deeper CNN layers require more processing time to extract more advanced and complex features. The classifier trained on ResNet50, which had the highest accuracy and AUC, took 0.136 s to classify the

patches generated from a single mammogram image according to the presence or absence of calcification, which means that seven mammogram images can be processed per second.

Next, we discuss the accuracy of the system for one mammogram at a time. The average accuracy for each mammogram was 95.3%. The interior of the breast is composed mainly of mammary glands and fat and is classified into four categories in accordance with the proportion of these components: fatty, scattered mammary, heterogeneous hyperintense, and extremely hyperintense. The percentage of mammary tissue is larger in the latter category. The two types of tissue, heterogeneous and extremely dense, are referred to as “dense breasts”. Densely concentrated breast has less fat and more mammary tissue, resulting in a whiter image and a lower rate of lesion detection [35]. This makes it difficult to distinguish between calcification and noncalcification and is thought to reduce the accuracy of classifying the presence or absence of calcification. Because microcalcifications appear as localized areas of high brightness, the use of thresholding is expected to increase the accuracy of the classifier as a whole as well as the accuracy of each mammographic image.

The specificity of microcalcification detection in CAD, which has been conventionally used as a reading aid, is 45%, and the FP rate is reported to be high [15]. The specificity of the classifiers used in this study was 97.8% for ResNet18, 98.1% for ResNet50, and 98.0% for ResNet101. A low false-positive rate and high NPV indicate its usefulness as a diagnostic aid that may reduce the burden on physicians.

One limitation of this study is that the dataset we used consisted of mammographic images from Europeans, Americans, and other foreign nationals. Japanese individuals generally have more dense breasts and a smaller breast size than Europeans and Americans. Therefore, it is necessary to use a dataset of Japanese mammographic images or to mix the data used in this study with Japanese mammographic images in the future. In addition, the training data used in this study contained more patches without calcification than patches with calcification, and the data were biased toward patches without calcification. Therefore, in the future, it will be necessary to take care to avoid bias in the training data.

In this study, the parameters were not changed in the training process, so changing the batch size or the number of epochs may improve the accuracy. In the previous study [33], the number of epochs was set to 200, and thus, it will be necessary to study the optimal parameters in the future. We used ResNet for classification in this study; however, further improvement in accuracy can be achieved by changing the network model used. This study aimed to perform a comparison with the intention of serving as an indicator when utilizing the time taken for classification in a clinical setting. While ResNet is a relatively common network model in medical image classification [18,19], it is essential to compare different network models. However, because we believed that altering individual layers or comparison between network models with different characteristics might yield varying sensitivity due to the background information inherent in images, we sought to elucidate the extent to which accuracy and processing time differ with varying depths of the same CNN. To achieve this, we conducted training and comparison using three CNNs: ResNet18, ResNet50, and ResNet101. In the context of evaluating the validity of the three CNNs employed in this study, it is essential to discuss them in an objective manner, particularly when compared to the state of the art (SOTA) in image classification. Utilizing the effective optimization algorithm implemented in BASIC-L (Lion, fine-tuned) [36], an accuracy of 76.45% was observed, which compares favorably to the 76.22% achieved using the original ResNet-50. Furthermore, in the case of using Vision Transformer (ViT-H/14) [37] for transfer accuracy, the metric was higher at 88.08% compared to ResNet-50's 77.54%. Additionally, in the assessment of test error rates using Sharpness-Aware Minimization (SAM) [38], which seeks parameters that lie in neighborhoods with uniformly low loss, the SAM-augmented ResNet-50 yielded a rate of 22.5% against the original ResNet-50's 22.9%. In a similar comparison using ResNet-101, SAM achieved a rate of 21.2% as opposed to the original ResNet-50's 20.2%. Moreover, employing multitask learning [39] showed classification accuracies of 88.22% when based on ResNet-18, which compares favorably with the reported SOTA figure of 87.82%. It is worth noting that ResNet serves as a

foundational convolutional neural network (CNN) for classification and is consistently used for comparative evaluation against new algorithms and methods. However, in this study, optimization of the learning algorithm was not undertaken, highlighting an area for future research. Nonetheless, it is crucial to consider that the application of SOTA algorithms could potentially alter inference times. Therefore, future studies should integrate considerations for optimizing CNNs, taking into account references to existing SOTA systems.

There is another limitation to this study. There are several limitations in approaching the task of detecting calcification while maintaining the resolution of mammography images. In the application of deep layer training models, it can be difficult to directly evaluate the individual contributions of the different intermediate layers in capturing image features. This recognition requires a comprehensive understanding of how each component affects the overall performance of the model. Considering the limited use of annotated mammography datasets, there are limitations in generalizing our findings in this study, and the performance of the models we present may show variability when testing them on a broader and more diverse dataset. They should therefore be applied and evaluated on a broader set of external datasets. In addition, regarding the optimization of parameters during training, ResNet, a structurally similar network model differing only in its number of layers, was used for comparison. However, for the optimization of parameters in image training, it is necessary to evaluate true training accuracy; this is a limitation of the research in this study and will be an issue to consider in addition to its implementation in other network models in the future.

For future consideration, with the aim of increasing the sensitivity of calcification classification, there are several ways to improve the accuracy, such as training with different CNNs and parameters, as described above; using a large number of diverse training data with calcification; creating patches in which the top, bottom, left, and right sides slightly overlap; using threshold processing; and devising patch partitioning. In addition, by using a classifier trained on the shape and distribution of calcification in patches classified as having calcification, it is possible to develop a tool for determining whether calcification is benign or malignant or for determining the category of calcification. In addition, the development of such a tool could contribute to the early detection and diagnosis of breast cancer; thus, it is necessary to evaluate its usefulness in clinical practice.

## 6. Conclusions

In this study, we developed a mammographic calcification detection method that only uses image classification to determine the presence or absence of calcification in the breast using deep learning technology. The developed method is fast and has relatively high accuracy in classifying the presence or absence of calcification and facilitating the visualization of the location of calcification; thus, its usefulness has been confirmed. In the future, calcification classification sensitivity can be improved in several ways, such as using diverse training data, varying CNNs and parameters, overlapping patches, applying threshold processing, and optimizing patch partitioning. Additionally, a classifier trained on calcification features within patches could help develop a tool for distinguishing calcification types and aiding in early breast cancer detection, requiring clinical evaluation.

**Author Contributions:** M.S. contributed to data analysis, algorithm construction, and writing and editing of the manuscript. T.Y., M.T. and S.I. reviewed and edited the manuscript. H.S. proposed the idea and contributed to the data acquisition, performed supervision and project administration, and reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The models created in this study are available on request from the corresponding author. The source code used in this study is available at <https://github.com/MIA-laboratory/MMGpatchCL> (accessed on 31 August 2023).

**Acknowledgments:** The authors would like to thank the laboratory members of the Medical Image Analysis Laboratory for their help.

**Conflicts of Interest:** The authors declare that no conflict of interest exist.

## References

1. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer Statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [[CrossRef](#)] [[PubMed](#)]
2. McDonald, E.S.; Clark, A.S.; Tchou, J.; Zhang, P.; Freedman, G.M. Clinical Diagnosis and Management of Breast Cancer. *J. Nucl. Med.* **2016**, *57*, 9S–16S. [[CrossRef](#)] [[PubMed](#)]
3. Jung, H.; Kim, B.; Lee, I.; Yoo, M.; Lee, J.; Ham, S.; Woo, O.; Kang, J. Detection of Masses in Mammograms Using a One-Stage Object Detector Based on a Deep Convolutional Neural Network. *PLoS ONE* **2018**, *13*, e0203355. [[CrossRef](#)]
4. Berry, D.A.; Cronin, K.A.; Plevritis, S.K.; Fryback, D.G.; Clarke, L.; Zelen, M.; Mandelblatt, J.S.; Yakovlev, A.Y.; Habbema, J.D.F.; Feuer, E.J. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *N. Engl. J. Med.* **2005**, *353*, 1784–1792. [[CrossRef](#)]
5. Altameem, A.; Mahanty, C.; Poonia, R.C.; Saudagar, A.K.J.; Kumar, R. Breast Cancer Detection in Mammography Images Using Deep Convolutional Neural Networks and Fuzzy Ensemble Modeling Techniques. *Diagnostics* **2022**, *12*, 1812. [[CrossRef](#)] [[PubMed](#)]
6. Wang, J.; Yang, Y. A Context-Sensitive Deep Learning Approach for Microcalcification Detection in Mammograms. *Pattern Recognit.* **2018**, *78*, 12–22. [[CrossRef](#)]
7. Broeders, M.; Moss, S.; Nystrom, L.; Njor, S.; Jonsson, H.; Paap, E.; Massat, N.; Duffy, S.; Lynge, E.; Paci, E. The Impact of Mammographic Screening on Breast Cancer Mortality in Europe: A Review of Observational Studies. *J. Med. Screen.* **2012**, *19* (Suppl. 1), 14–25. [[CrossRef](#)]
8. Bjurstam, N.G.; Björneld, L.M.; Duffy, S.W. Updated Results of the Gothenburg Trial of Mammographic Screening. *Cancer* **2016**, *122*, 1832–1835. [[CrossRef](#)]
9. Beau, A.B.; Andersen, P.K.; Vejborg, I.; Lynge, E. Limitations in the Effect of Screening on Breast Cancer Mortality. *J. Clin. Oncol.* **2018**, *36*, 2988–2994. [[CrossRef](#)]
10. Duffy, S.W.; Yen, A.M.F.; Tabar, L.; Lin, A.T.Y.; Chen, S.L.S.; Hsu, C.Y.; Dean, P.B.; Smith, R.A.; Chen, T.H.H. Beneficial Effect of Repeated Participation in Breast Cancer Screening upon Survival. *J. Med. Screen.* **2023**, *ahead-of-print*. [[CrossRef](#)]
11. Azam, S.; Eriksson, M.; Sjölander, A.; Gabrielson, M.; Hellgren, R.; Czene, K.; Hall, P. Mammographic Microcalcifications and Risk of Breast Cancer. *Br. J. Cancer* **2021**, *125*, 759–765. [[CrossRef](#)]
12. Wilkinson, L.; Thomas, V.; Sharma, N. Microcalcification on Mammography: Approaches to Interpretation and Biopsy. *Br. J. Radiol.* **2017**, *90*, 20160594. [[CrossRef](#)] [[PubMed](#)]
13. Demetri-Lewis, A.; Slanetz, P.J.; Eisenberg, R.L. Breast Calcifications: The Focal Group. *AJR. Am. J. Roentgenol.* **2012**, *198*, W325–W343. [[CrossRef](#)] [[PubMed](#)]
14. Dromain, C.; Boyer, B.; Ferré, R.; Canale, S.; Delalogue, S.; Balleyguier, C. Computed-Aided Diagnosis (CAD) in the Detection of Breast Cancer. *Eur. J. Radiol.* **2013**, *82*, 417–423. [[CrossRef](#)] [[PubMed](#)]
15. Dominković, M.D.; Ivanac, G.; Radović, N.; Čavka, M. What Can We Actually See Using Computer Aided Detection in Mammography? *Acta Clin. Croat.* **2020**, *59*, 576. [[CrossRef](#)]
16. Ribli, D.; Horváth, A.; Unger, Z.; Pollner, P.; Csabai, I. Detecting and Classifying Lesions in Mammograms with Deep Learning. *Sci. Rep.* **2018**, *8*, 4165. [[CrossRef](#)]
17. Al-antari, M.A.; Al-masni, M.A.; Choi, M.T.; Han, S.M.; Kim, T.S. A Fully Integrated Computer-Aided Diagnosis System for Digital X-Ray Mammograms via Deep Learning Detection, Segmentation, and Classification. *Int. J. Med. Inform.* **2018**, *117*, 44–54. [[CrossRef](#)]
18. Chan, H.P.; Samala, R.K.; Hadjiiski, L.M. CAD and AI for Breast Cancer—Recent Development and Challenges. *Br. J. Radiol.* **2020**, *93*, 20190580. [[CrossRef](#)]
19. Sugimori, H.; Hamaguchi, H.; Fujiwara, T.; Ishizaka, K. Classification of Type of Brain Magnetic Resonance Images with Deep Learning Technique. *Magn. Reson. Imaging* **2021**, *77*, 180–185. [[CrossRef](#)]
20. Ichikawa, S.; Itadani, H.; Sugimori, H. Toward Automatic Reformation at the Orbitomeatal Line in Head Computed Tomography Using Object Detection Algorithm. *Phys. Eng. Sci. Med.* **2022**, *45*, 835–845. [[CrossRef](#)]
21. Yang, R.; Yu, Y. Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. *Front. Oncol.* **2021**, *11*, 638182. [[CrossRef](#)]
22. Pham, D.L.; Xu, C.; Prince, J.L. Current Methods in Medical Image Segmentation. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 315–337. [[CrossRef](#)]
23. Asami, Y.; Yoshimura, T.; Manabe, K.; Yamada, T.; Sugimori, H. Development of Detection and Volumetric Methods for the Triceps of the Lower Leg Using Magnetic Resonance Images with Deep Learning. *Appl. Sci.* **2021**, *11*, 12006. [[CrossRef](#)]

24. Shibahara, T.; Wada, C.; Yamashita, Y.; Fujita, K.; Sato, M.; Kuwata, J.; Okamoto, A.; Ono, Y. Deep Learning Generates Custom-Made Logistic Regression Models for Explaining How Breast Cancer Subtypes Are Classified. *PLoS ONE* **2023**, *18*, e0286072. [[CrossRef](#)] [[PubMed](#)]
25. Usui, K.; Yoshimura, T.; Tang, M.; Sugimori, H. Age Estimation from Brain Magnetic Resonance Images Using Deep Learning Techniques in Extensive Age Range. *Appl. Sci.* **2023**, *13*, 1753. [[CrossRef](#)]
26. Inomata, S.; Yoshimura, T.; Tang, M.; Ichikawa, S.; Sugimori, H. Estimation of Left and Right Ventricular Ejection Fractions from Cine-MRI Using 3D-CNN. *Sensors* **2023**, *23*, 6580. [[CrossRef](#)] [[PubMed](#)]
27. Zhang, J.; Saha, A.; Zhu, Z.; Mazurowski, M.A. Hierarchical Convolutional Neural Networks for Segmentation of Breast Tumors in MRI With Application to Radiogenomics. *IEEE Trans. Med. Imaging* **2019**, *38*, 435–447. [[CrossRef](#)]
28. Truhn, D.; Schrading, S.; Haarbuerger, C.; Schneider, H.; Merhof, D.; Kuhl, C. Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-Enhancing Lesions at Multiparametric Breast MRI. *Radiology* **2019**, *290*, 290–297. [[CrossRef](#)]
29. Yala, A.; Lehman, C.; Schuster, T.; Portnoi, T.; Barzilay, R. A Deep Learning Mammography-Based Model for Improved Breast Cancer Risk Prediction. *Radiology* **2019**, *292*, 60–66. [[CrossRef](#)]
30. Kim, H.; Lim, J.; Kim, H.G.; Lim, Y.; Seo, B.K.; Bae, M.S. Deep Learning Analysis of Mammography for Breast Cancer Risk Prediction in Asian Women. *Diagnostics* **2023**, *13*, 2247. [[CrossRef](#)]
31. Jones, M.A.; Faiz, R.; Qiu, Y.; Zheng, B. Improving Mammography Lesion Classification by Optimal Fusion of Handcrafted and Deep Transfer Learning Features. *Phys. Med. Biol.* **2022**, *67*, 054001. [[CrossRef](#)]
32. Walsh, R.; Tardy, M. A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer. *Diagnostics* **2023**, *13*, 67. [[CrossRef](#)] [[PubMed](#)]
33. Kobayashi, T.; Haraguchi, T.; Nagao, T. Classifying Presence or Absence of Calcifications on Mammography Using Generative Contribution Mapping. *Radiol. Phys. Technol.* **2022**, *15*, 340–348. [[CrossRef](#)] [[PubMed](#)]
34. Valvano, G.; Santini, G.; Martini, N.; Ripoli, A.; Iaconi, C.; Chiappino, D.; Della Latta, D. Convolutional Neural Networks for the Segmentation of Microcalcification in Mammography Imaging. *J. Healthc. Eng.* **2019**, *2019*, 9360941. [[CrossRef](#)] [[PubMed](#)]
35. Mann, R.M.; Athanasiou, A.; Baltzer, P.A.T.; Camps-Herrero, J.; Clauser, P.; Fallenberg, E.M.; Forrai, G.; Fuchsjäger, M.H.; Helbich, T.H.; Killburn-Toppin, F.; et al. Breast Cancer Screening in Women with Extremely Dense Breasts Recommendations of the European Society of Breast Imaging (EUSOBI). *Eur. Radiol.* **2022**, *32*, 4036–4045. [[CrossRef](#)] [[PubMed](#)]
36. Chen, X.; Liang, C.; Huang, D.; Real, E.; Wang, K.; Liu, Y.; Pham, H.; Dong, X.; Luong, T.; Hsieh, C.-J.; et al. Symbolic Discovery of Optimization Algorithms. *arXiv* **2023**, arXiv:2302.06675v4.
37. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16X16 Words: Transformers for Image Recognition at Scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual, 3–7 May 2021.
38. Foret, P.; Kleiner, A.; Mobahi, H.; Neyshabur, B. Sharpness-Aware Minimization for Efficiently Improving Generalization. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual, 3–7 May 2021.
39. Kabir, H.M.D. Reduction of Class Activation Uncertainty with Background Information. *arXiv* **2023**, arXiv:2305.03238.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.