

Detection of Breast Cancer in Mammography Using Pretrained Convolutional Neural Networks with Fine-Tuning



Cesar Muñoz-Chavez, Hermilo Sánchez-Cruz, Humberto Sossa-Azuela, and Julio Ponce-Gallegos

1 Introduction

Latin American women have higher rates of breast cancer incidence and mortality compared to women in developed countries. The incidence rate of breast cancer is 80% for women over 44 years old, and the mortality rate is 86% [1]. Young women are also affected by breast cancer, with rates as high as 15% in less developed countries such as Mexico [2]. Low- and middle-income countries have a mortality to incidence ratio that is considerably higher, between 60% and 75%, compared to high-income countries [3]. Early detection of breast cancer is crucial for better survival rates, and digital screening mammography is commonly used by radiologists for analysis, diagnosis, and categorization of breast cancer [4, 5].

Although mammography is currently one of the most reliable methods for detecting breast cancer, medical experts face challenges in interpreting mammogram images, which could lead to an incorrect diagnosis. Current methods of image classification were investigated, and those that use artificial intelligence stood out the most [6, 7]. Artificial intelligence (AI) has advanced quickly in recent years in a variety of sectors, including image processing. Since images are one of the most crucial sources of information for activities involving human intelligence, AI has been widely used in image processing [8, 9]. In this research, AI techniques were found to be beneficial, particularly in the form of computer-aided diagnosis (CAD)

C. Muñoz-Chávez · H. Sánchez-Cruz (✉) · J. Ponce-Gallegos
Centro de Ciencias Básicas, Universidad Autónoma de Aguascalientes,
Aguascalientes, Mexico
e-mail: hermilo.sanchez@edu.uaa.mx

H. Sossa-Azuela
Centro de Investigación en Computación, Instituto Politécnico Nacional,
Mexico City, Mexico

as an alternative to radiologists' time-consuming and inaccurate double-reading procedure. Some of these CADs utilize machine learning algorithms that take up the information from complex patterns [10–12]. However, traditional classifiers based on handcrafted features are considered to be complex and time-consuming, especially for tasks involving feature extraction and selection [11, 12].

Recent studies have shown that deep learning algorithms are efficient on tasks including image segmentation, detection, and classification in a range of computer vision and image processing disciplines [13–15]. Deep learning is a type of machine learning that is inspired by the structure and function of the brain. Deep learning has lately received a lot of attention for classification problems, particularly in the field of medical image analysis [16]. Therefore, for this research, current methods of image classification were analyzed, and based on the state-of-the-art, deep convolutional neural networks (DCNNs) were found to perform well on the image classification task [16–18]. These models require a large number of images, which is why we searched for well-known state-of-the-art datasets. Two datasets were used in this study: CBIS-DDSM (Curated Breast Imaging Subset of DDSM) [19] and Mini-MIAS (Mammographic Image Analysis Society) [20]. Transfer learning with fine-tuning was used to accomplish this research [21, 22], where a model built for one task is used for another and the model's output is modified to meet the new task [23, 24].

This chapter presents the contributions of a research study aimed at exploring the effectiveness of deep learning algorithms for classifying mammogram images. In particular, the study compared the performance of two deep learning models, ResNet-50 and EfficientB7. To enhance the quality of the images, the researchers applied various image processing techniques, including CLAHE, unsharp masking, and a median filter. Additionally, they used a data augmentation algorithm from a library called Albumentation to increase the number of training images and improve the robustness of the convolutional neural network (CNN). The performance of each model was evaluated using five metrics, including accuracy, precision, recall, F1-score, and confusion matrix.

The rest of this chapter is organized as follows. A review of pertinent literature is presented in Sect. 2. The methodology used and the differences between each dataset are presented in Sect. 3, along with full disclosure of all the steps used to accomplish the goal of classification tasks. Experimental results comparing every model and metric are described in Sect. 4. Finally, Sect. 5 provides some conclusions.

2 Previous Works

Deep convolutional neural network (DCNN) models have become increasingly popular in recent years due to their exceptional performance in various computer vision tasks, including image classification, segmentation, and detection. Many different models have been proposed and utilized in research studies. However,

researchers have obtained varied results, and it is important to analyze these results to identify areas where DCNN models can be improved.

The research by Rampun et al. [25] focuses on developing an automated method for breast pectoral muscle segmentation in mediolateral oblique mammograms. To achieve this, they employed a convolutional neural network (CNN) inspired by the holistically nested edge detection (HED) network, which is capable of learning complex hierarchical features and resolving spatial ambiguity in estimating the pectoral muscle boundary. The CNN is also designed to detect “contour-like” objects in mammograms. The study utilized several datasets, including MIAS, INBreast, BCDR, and CBIS-DDSM, to evaluate the performance of the proposed method. An ensemble approach was employed by Altameem et al. [26], in which the Gompertz function was used to build fuzzy rankings of the base classification techniques and the decision scores of the base models were adaptively combined to construct final predictions. Using Inception V4, ResNet 16, VGG 11, and DenseNet 121, as well as other deep CNN models, a deep learning approach using convolutional neural networks (CNNs) was used to classify breast cancer histopathological images from the BreaKHis dataset. The approach introduced by Wei et al. [27] enables the use of high-resolution histopathological images as input to existing convolutional neural networks (CNNs) without requiring complex and computationally expensive modifications to the network architecture. This is achieved through the extraction of image patches from the high-resolution image, which are then used to train the CNN. The final classification is obtained by combining the predictions from these patches. This method allows existing CNNs to be used for histopathological image analysis without the need for extensive modifications or the development of a new architecture. Additionally, a network was trained and validated by Auccahuasi et al. [28] using a database of images containing microcalcifications classified as benign and malignant from mammographic images of MIAS. One recent study introduced “double-shot transfer learning,” which is a revolutionary method built on the idea of transfer learning. This strategy, presented by Alkhaleefah et al. [29], significantly improved categorization accuracy. The following research, proposed by Charan et al. [30], uses deep learning and neural networks for the classification of normal and abnormal breast detection in mammogram images using the Mammograms-MIAS dataset, which contains 322 mammograms with 189 images of normal breasts and 133 images of abnormal breasts. The study used a convolutional neural network (CNN) and obtained promising experimental results that suggest the efficacy of deep learning for breast cancer detection in mammogram images. Saber et al. [31] used pretrained convolutional neural networks (CNNs) to detect and classify breast tumors in the INbreast dataset using mammography images. The proposed model preprocesses the images to improve image quality and reduce computation time and then transfers the learned parameters from the CNNs to improve the classification. Another recent study proposes, by Qasim et al. [32], the use of a convolutional neural network (CNN) to detect breast cancer in mammography images by classifying them as noncancerous or cancerous abnormalities using the DDSM dataset. A set of mammogram images is preprocessed using histogram equalization, and the resulting images are used as a training source for the CNN. The proposed system, called

BCDCNN, is compared to the MCCANN system, and the results show that BCDCNN has higher classification accuracy and a higher resolution compared to other existing systems.

The use of pretrained models is good for discovering new models for specific tasks; a new method proposed by Montaha et al. [33] utilizing the fine-tuned VGG-16 model called BreasNet-18 is introduced. This methodology used the CBIS-DDSM dataset, and preprocessing was applied to these images. Artifact removal was the name of the initial stage of the preprocessing. It made use of techniques like binary masking, morphological opening, and detecting the largest contour. “Remove line” is the name of the second stage. The following techniques were used on certain images that included a bright, straight line attached to the breast contour: in-range operation, Gabor filter, morphological operations, and invert mask. Some algorithms, like gamma correction, CLAHE, and Green Fire Blue, enhance images in the third stage. After the preprocessing, the data augmentation technique is used to acquire additional images and solve issues with over- and under-fitting. Finally, the BreastNet-18 model was used to evaluate the classification problem of four classes. Additionally, Allugunti et al. [34] show a computer-aided diagnostic (CAD) method is recommended to classify three classes for the methodology using certain traditional methods (cancer, no cancer, and noncancerous). Convolutional neural networks (CNNs), support vector machines (SVM), and random forests (RF) were the three classifiers they employed. They used a dataset with a total of 1000 images from the Kaggle website. Following the test of each classifier, it was discovered that CNN performed better, achieving an accuracy of 99.6%.

The study by Shen et al. [35] introduces a novel deep learning algorithm for detecting breast cancer on mammograms. The algorithm utilizes an “end-to-end” training approach, which decreases the dependence on lesion annotations and enables the use of image-level labels. The model is trained on two separate datasets, the Digital Database for Screening Mammography (CBIS-DDSM) and the INbreast database, and the results indicate excellent performance with high accuracy on both heterogeneous mammography platforms. These findings have the potential to improve clinical tools and decrease the incidence of false-positive and false-negative screening results, which can lead to more accurate diagnoses and improved patient outcomes. The research conducted by Khamparia et al. [36] presented an approach for classifying mammograms using deep learning models. The authors experimented with various models, such as a pretrained VGG model, a residual network, and a mobile network, to determine their effectiveness. They found that their fine-tuned VGG16 model, with data augmentation and pretrained ImageNet weights, outperformed the other models in terms of accuracy. They utilized the DDSM dataset for their experiments, and their approach yielded an accuracy of 88.30% and an AUC value of 93.30%. Hameed et al. [37] proposed a deep learning methodology for the accurate detection of cancerous and noncancerous tissue using pretrained convolutional neural networks (CNNs). They used both the VGG16 and VGG19 models as is and with modifications to enhance performance. The study collected 544 whole slide images (WSIs) from 80 patients with breast cancer from the

pathology division of Colsanitas Colombia University in Bogota, Colombia. The images were normalized, and the data was divided into 80% for training and 20% for testing. They used a data augmentation strategy during training and achieved an accuracy and F1-score of 95.29%. This study has significant implications for the development of tools to aid in the accurate diagnosis of breast cancer, potentially improving patient outcomes. Additionally, the three stages of the Multi-View Feature Fusion (MVFF) methodology proposed by Nasir Khan et al. [38] are as follows: In the first step, mammography is binary-classified as abnormal or normal, and in the second, mass and calcification are classified. The final step is to classify the condition as malignant or benign. They utilized the CBIS-DDSM dataset for this investigation. The AUC values for mass and categorization were 93.20% and 0.84% for malignant and benign tumors, respectively. A new computer-aided detection (CAD) system is proposed for classifying benign and malignant mass tumors in breast mammography images using deep learning and segmentation techniques. The CAD system uses two segmentation approaches, one involving manual determination of the region of interest (ROI) and the other using threshold- and region-based techniques. AlexNet, a deep convolutional neural network (DCNN) used for feature extraction and fine-tuning to classify two classes, Ragab et al. present the performance in their paper [39]. VGGNet models that have been adjusted may perform better if classifications of masses and calcifications from mammography are performed using transfer learning. Xi et al. presented research about it [40]. The research by Hepsa et al. [41] showed that breast biopsies based on mammography and ultrasound results have a high rate of being diagnosed as benign (40–60%), which can lead to negative impacts such as unnecessary operations, fear, pain, and cost. To address this, they apply deep learning using convolutional neural networks (CNNs) to classify abnormalities in mammogram images as benign or malignant using two databases: Mini-MIAS and BCDR. While Mini-MIAS has valuable information such as the location and radius of the abnormality, BCDR does not. Initially, accuracy, precision, recall, and F1-score values range from 60 to 72%. To improve results, the authors implement preprocessing methods including cropping, augmentation, and balancing image data. They create a mask to find regions of interest in BCDR images and observe an increase in classification accuracy from 65% to around 85%.

3 Material and Methods

3.1 CNN Architectures

In this research, four popular CNN architectures were evaluated for their suitability for the classification of mammograms at different stages. These architectures have been widely cited in recent studies and were customized for this specific task using the suggested methodology. The evaluation of different CNN architectures is

important in identifying the best model for a particular task, as different architectures have different strengths and weaknesses. By comparing the performance of these architectures, we can determine which one is most suitable for the task at hand. The customization of the architectures for mammogram classification involved adjusting various hyperparameters, such as the learning rate and batch size, to optimize performance. This process is crucial for achieving the best possible results and improving the accuracy of mammogram classification, which can have important implications for breast cancer diagnosis and treatment.

VGG19

Simonyan et al. [42] contribute a thorough evaluation of networks of increasing depth using an architecture with 3×3 convolution filters. They indicate the model's ability to significantly improve 16–19 weight layers deep in both models.

Very deep convolutional neural network layers, totaling 19 layers, are used in the VGG19 architecture. It consists of multiple fully connected layers that are followed by a string of convolutional and max pooling layers. The network has been trained on a sizable dataset of photos in order to learn to recognize a wide range of objects and scenarios. The network is intended to be used for image classification tasks. The use of tiny convolutional filters with 3×3 pixels is one of the distinguishing characteristics of the VGG19 design. Fine-grained characteristics can then be learned by the network from the input photos, which is helpful for tasks like object recognition.

VGG19 has achieved good results on a variety of image classification tasks, which suggests that it may be a reliable and robust model for this type of problem.

ResNet-50 and ResNet152

He et al. [43] introduced ResNet, a CNN architecture that uses skip connections to enable residual function learning. The ResNet models include ResNet-34 A, ResNet-34 B, ResNet-34 C, ResNet-50, ResNet-101, and ResNet-152. An ensemble of these models achieved a 3.57% error rate on the ImageNet test set and won first place in the 2015 ILSVRC classification challenge. In this study, ResNet-50 and ResNet-152 were used for testing.

The ResNet architecture allows for the learning of residual functions through skip connections, which skip over layers that are not essential for the current task. This improves learning efficiency and speed. ResNet is known for its ability to train very deep networks without the vanishing gradient problem, thanks to the use of skip connections. ResNet-50 has an efficient and simple architecture, making it faster to train and easier to implement compared to other models. This feature is especially useful for tasks that involve processing a large amount of data, such as mammogram classification.

EfficientNetB7

Tan et al. [44] introduced a family of models called EfficientNets. EfficientNet is an efficient model that can achieve state-of-the-art accuracy on ImageNet and is commonly used for image classification transfer learning tasks. This architecture was formed by leveraging a multi-objective neural architecture search that optimizes accuracy as well as floating point operations per second (FLOPS).

The main idea behind EfficientNet is to scale up CNNs in a more efficient manner. Conventional CNNs increase the depth and width of the network to improve accuracy, but this also increases the number of parameters and computation required. EfficientNet, instead, proposes to scale the network up in a more balanced way by also increasing the resolution of the input image. This allows the network to improve the accuracy while keeping the computational cost constant.

In order to scale up CNNs in a more organized way, this model suggests a novel model scaling technique that makes use of a straightforward but incredibly powerful compound coefficient. Our method uniformly scales each dimension with a fixed set of scaling coefficients, in contrast to existing approaches that arbitrarily scale network dimensions like width, depth, and resolution.

EfficientNetB7 has been designed to be highly efficient in terms of both accuracy and resource usage. It has achieved state-of-the-art performance on a number of image classification and object detection benchmarks and has been widely used in a variety of applications.

3.2 Datasets

The Curated Breast Imaging Subset of DDSM, also known as CBIS-DDSM [19], is a subset of images that have been selected and curated by radiologists with specialized training from the original DDSM dataset. These images are stored in the standard DICOM format, which is commonly used for storing medical images such as CT and MRI scans.

Another curated mammographic dataset that is widely available is the Mammographic Imaging Analysis Society (MIAS) dataset [20]. Both of these datasets are often used for training and testing machine learning algorithms for tasks such as image classification, object recognition, and segmentation. In addition, the Mini-MIAS dataset is often used as a benchmark for image compression techniques.

Both CBIS-DDSM and Mini-MIAS datasets are available for free and provide numerous images that can be used for medical image analysis research. The CBIS-DDSM dataset contains over 2620 scanned film mammography studies, while the Mini-MIAS dataset contains over 322 images. The location of the tumor has already been indicated for both datasets, making them particularly useful for breast cancer detection and diagnosis research.

3.3 *Experiment Environment*

The Python programming language was utilized in this study and was run on a workstation equipped with an NVIDIA Geforce RTX 3070ti 8GB graphics card, Ryzen R9 5900x processor, 32 of DDR4 3200Mhz RAM, and an XPG Spectrix 512GB SSD. TensorFlow and Keras were employed in this research, both of which are open-source libraries designed for deep learning applications [45].

TensorFlow was launched by Google to aid in the development of deep learning models, while Keras is a neural network library that was written in Python. To achieve faster and more accurate results, we utilized the GPU, which necessitated the installation of the Deep Neural Network library (cuDNN) [46] and Compute Unified Device Architecture (CUDA).

4 Methodology

This section describes the data preprocessing steps, data selection process, data augmentation strategy, and CNN architectures used for the classification tasks in this research.

Deep convolutional neural network (DCNN) models require a significant amount of data for training to achieve good performance [47]. To ensure accurate diagnosis and better outcomes, it is crucial to use trustworthy datasets. Deep learning has shown that more data can improve results. We searched through many sources to find a dataset that could provide the information we needed, recognizing that the lack of data could create issues. Ultimately, we identified two of the best datasets for mammography images used in state-of-the-art research: Mini-MIAS and CBIS-DDSM. We utilized images from both datasets in this investigation. Mini-MIAS contains 322 images, 133 of which show abnormalities (63 benign and 51 malignant), and the remaining 208 do not. In contrast, CBIS-DDSM comprises 2620 scanned film mammography studies. Once we obtained these datasets, we divided them into four different classification problems: normal and abnormal for Mini-MIAS; normal and abnormal for CBIS-DDSM; masses and calcifications for CBIS-DDSM; and finally, masses, calcifications, and normal for CBIS-DDSM. We tackled each of these classification problems in four stages (Fig. 1).

To achieve good performance, the deep convolutional neural network (DCNN) models required a significant amount of data, which is often not readily available. Therefore, transfer learning and fine-tuning techniques were applied to improve the performance of the models, even with limited data availability.

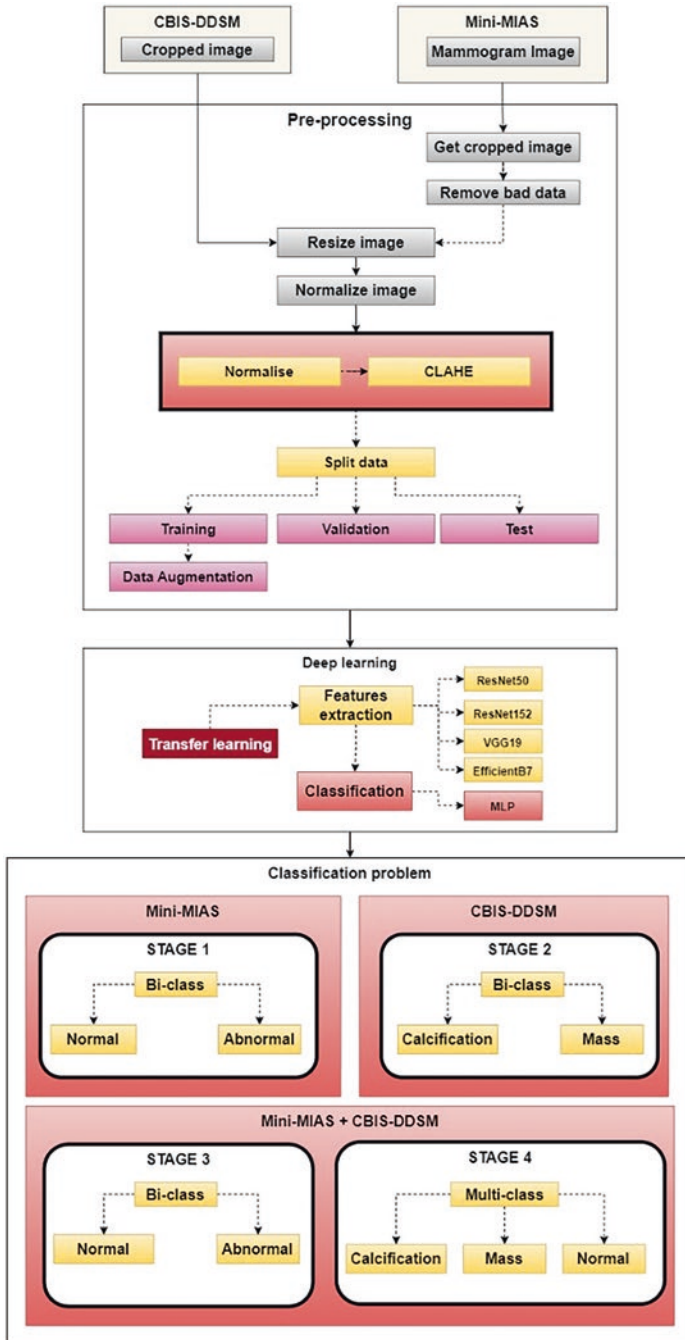


Fig. 1 Proposed methodology for this research

4.1 Stage 1

In the first stage, it involved a comparison of images without any abnormalities to those with tumors, both benign and malignant. Preprocessing of the final images with anomalies received special consideration at this stage because an algorithm was necessary to crop the images using the coordinates provided on the website page. Since it was only necessary to take random crops from images without any anomalies, there were no issues with the normal images.

4.2 Stage 2

In stage 2, a binary classification is applied to distinguish between the two types of mammography anomalies: masses and calcifications. The dataset is divided into masses (benign, benign without callback, and malignant) and calcifications (benign, benign without callback, and malignant), making CBIS-DDSM suitable for this scenario. To solve this classification problem, benign masses are combined with malignant masses and benign calcifications with malignant calcifications. Images labeled as “benign without callback” were not used at this stage.

4.3 Stage 3

In stage 3, something different was attempted. It was not possible to compare healthy mammograms without anomalies to abnormal mammograms in CBIS-DDSM since there is no category for healthy images. To address this, images were obtained from the Mini-MIAS dataset, which includes various abnormalities such as architectural distortion, calcification, well-defined or confined masses, spiculated masses, ill-defined masses, and asymmetry. The Mini-MIAS dataset also enables us to determine the severity of the abnormality (benign or malignant). However, to perform the binary classification, normal images were required. A random cut was made on each mammogram labeled as “normal,” and these images were used alongside abnormal images from CBIS-DDSM for the classification.

4.4 Stage 4

In the final stage, the potential of this methodology for a multiclass problem was tested using labeled images of masses, calcifications, and normal images. As CBIS-DDSM does not include healthy or normal images, images from the Mini-MIAS

dataset were used instead. As a result, the output changed from a binary problem to a trinary problem due to the difference in data.

This methodology has been utilized, as depicted in Fig. 1. The primary difference is that CBIS-DDSM does not require an algorithm to crop images, unlike Mini-MIAS, which needed it. Cropped images can be obtained directly from the website's archives.

4.5 Preprocessing

As previously stated, we examined four stages that utilized Mini-MIAS and CBIS-DDSM image databases. The first database, Mini-MIAS, contained images in PGM format, which we obtained from the website. However, we stored the images in PNG format because it preserves the quality of edited images. While downloading the images, we encountered an issue: The digital mammograms had a resolution of 1024×1024 , which could hinder processing. Luckily, the necessary information to identify anomalies was available on the Mini-MIAS website, and we developed an algorithm to crop the images based on their coordinates.

To identify regions in images with anomalies, we used the coordinates to obtain each cropped image. However, several images that contained anomalies lost sharpness and quality when expanded, resulting in their exclusion from the evaluation. Conversely, healthy images were easier to obtain and were randomly selected for patches with a resolution of 112×112 . Using an inter-cubic interpolation algorithm, we resized each image (normal and abnormal) to the standard 224×244 size after obtaining the ROI.

The CBIS-DDSM dataset, which contains cropped images of mammograms with masses and calcifications, was downloaded. The images had different sizes, so we created an algorithm to automatically resize and extract them from the folders to a standard size of 244×244 using an inter-cubic interpolation algorithm. For each lesion (masses and calcifications), we combined benign and malignant images to test whether the architectures used in this study could distinguish between them. The dataset contained 1555 images of masses and 1331 images of calcifications.

After obtaining the cropped images, we normalized each one to ensure that the pixel values ranged from 0 to 255, as the original mammography image had a resolution of 16 bits and pixel values ranging from 0 to 65,535. We used the CLAHE technique to enhance image contrast, limiting contrast amplification to reduce noise amplification [47–49]. A clip limit of 0.01 was used for CLAHE. Finally, we compared the results of two different approaches, namely, normalize (NO) and CLAHE.

Data augmentation was necessary after preprocessing to increase the amount of data available [50–54]. We used the Albumentation library [55] to develop an algorithm that employed horizontal and vertical flips, with the option of adding rotations. Researchers can increase the number of rotational samples required by adjusting the sample variable (Fig. 2).

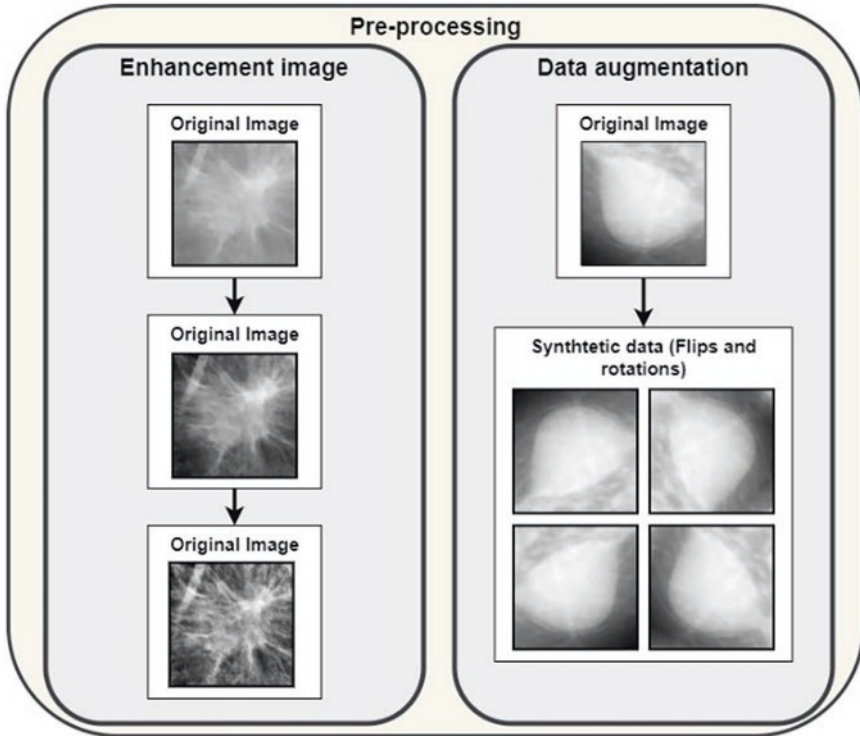


Fig. 2 To preprocess the images, several steps were taken. Firstly, the images were cropped and then normalized by adjusting the pixel values to fit within a certain range. After that, the contrast of the images was enhanced using the CLAHE technique. To increase the number of images available for the training phase, data augmentation was used to create synthetic data

In the conducted research, transfer learning models were utilized, and ImageNet weights were used to achieve better results. By employing pretrained models with learned features, the models were fine-tuned on the mammography image classification task. The ImageNet dataset, which is a large-scale dataset used for pretraining deep neural networks for image classification tasks, was chosen due to its millions of images and thousands of object categories, making it a valuable resource for transfer learning [56, 57]. The models were initialized with weights pretrained on ImageNet, which allowed them to leverage the learned features and adapt to the mammography image classification task more efficiently [21–24].

5 Results and Evaluations

The preprocessed dataset was divided into three parts: 80% for training data, 10% for validation data, and 10% for test data. The ResNet-50 and EfficientNetB7 neural networks were used for classification with the following parameters settings: Adam optimizer, with a learning rate of 0.0001 [58], batch size set to 32, and binary and categorical cross-entropy used as loss functions [59]. The training process was set to run for 20 epochs.

For the classification neural network, we employ these layers in order to improve the outcome (Fig. 3).

5.1 Metrics

It is important to understand the metrics, including the confusion matrix which shows the number of true positive, true negative, false-positive, and false negative predictions made by the model. True positive (TP) refers to the number of instances that were correctly classified as positive, while true negative (TN) refers to the number of instances that were correctly classified as negative. False positive (FP) refers

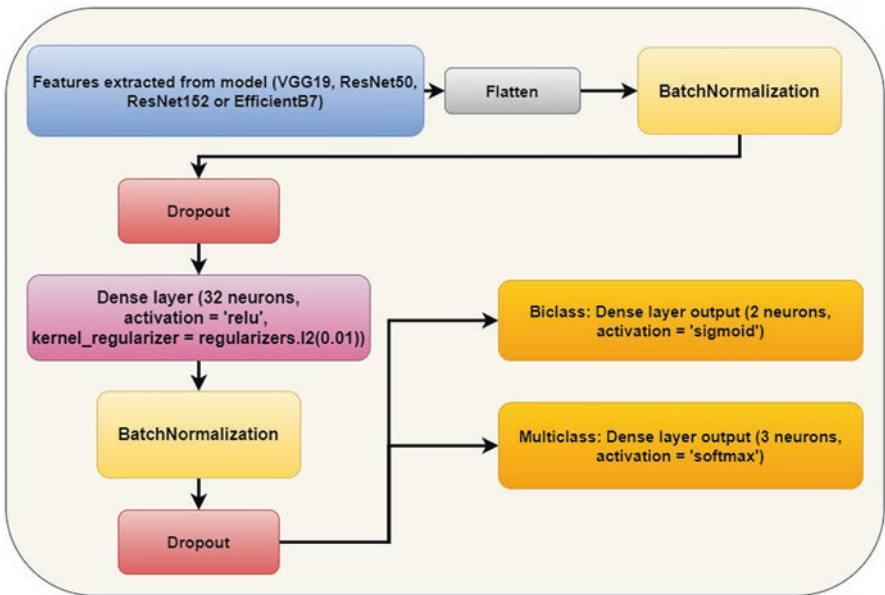


Fig. 3 Using a neural network to classify the features extracted from the models (VGG19, ResNet-50, ResNet152, or EfficientB7). Multiple layers were used, such as the dropout layer [60], batch normalization layer [61], and dense layers. The dense layer that classified the features extracted was assigned a 0.01 for the regularizer l2 value [62]

to the number of instances that were incorrectly classified as positive, and false negative (FN) refers to the number of instances that were incorrectly classified as negative.

Evaluating the performance of a machine learning model requires the use of various metrics, with accuracy being one of the most commonly used. This metric measures the percentage of correct predictions made by the model, making it simple and intuitive. However, accuracy can be misleading in cases where the classes in the dataset are imbalanced, meaning that one class is significantly more prevalent than others (Eq. 1).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (1)$$

Precision is another crucial metric that measures the proportion of true positives, or correct predictions of the positive class, among all positive predictions made by the model. This metric is particularly relevant in tasks where minimizing false positives is essential, such as medical diagnosis or spam detection (Eq. 2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (2)$$

Conversely, recall measures the proportion of true positives among all actual positive examples in the dataset. This metric is especially valuable in tasks where detecting as many positive examples as possible is critical, such as fraud detection or cancer screening (Eq. 3).

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{FP}}, \quad (3)$$

To achieve a balance between precision and recall in a model, a useful metric is the F1-score. This metric is the harmonic mean of precision and recall, providing a more comprehensive evaluation of a model's performance by taking both precision and recall into account (Eq. 4).

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (4)$$

The AUC (area under the curve) metric is used to evaluate the performance of a binary classifier. It is a measure of the classifier's ability to distinguish between positive and negative classes.

AUC is calculated by plotting the true positive rate (TPR) against the false-positive rate (FPR) at various classification thresholds. The true positive rate is the proportion of positive cases that are correctly identified as positive, while the false-positive rate is the proportion of negative cases that are incorrectly classified as positive.

It is essential to select the right metric for the task at hand because various metrics may highlight various elements of model performance. Model selection, hyperparameter adjustment, and model evaluation are a few of the uses for deep learning metrics. Deep learning metrics in this research is used to evaluate the performance of a deep learning model on a particular task or problem. They provide a way to quantitatively measure how well the model is able to solve the problem and can be used to compare the performance of different models or to track the progress of a model as it is being trained.

Overall, the purpose of using deep learning metrics is to help you understand how well your model is performing and to identify areas where it may be underperforming. This can help you to fine-tune your model and improve its performance.

5.2 Tables

For the training set, data augmentation was used on the training set to introduce variations to the model by applying changes to the datasets, which can increase the robustness of machine learning and reduce training costs.

To increase the amount of data available for training, various data augmentation techniques were applied to the Mini-MIAS and CBIS-DDSM datasets. The specific techniques used for each stage are summarized in Table 1 for stage 1 images in the Mini-MIAS dataset, Table 2 for stage 2 images made synthetically using both datasets, Table 3 for stage 3 images with lesions labeled as “masses and calcifications” using only the CBIS-DDSM dataset, and Table 4 for stage 4 images with normal images labeled for lesions using both datasets for the multiclass problem. Before the DCNN could use the dataset, the number of images in each class needed to be balanced to ensure that the model could learn from all classes equally. This was especially important for Tables 2 and 4, which required greater increases in normal images because there were fewer of them compared to the other classes.

Table 5 shows the outcomes of using the normalized dataset, including performance metrics such as accuracy, precision, recall, F1-score, and AUC. Table 6 shows the results of using the dataset with the CLAHE preprocessing technique and

Table 1 Using the technique of data augmentation to the training dataset of stage 1

Without data augmentation			
Class	Training set	Validation set	Test set
Normal	165	22	22
Tumor	82	11	11
With data augmentation			
Class	Training set	Validation set	Test set
Normal	4648	720	22
Tumor	4648	560	11

Table 2 Using the technique of data augmentation to the training dataset of stage 2

Without data augmentation			
Class	Training set	Validation set	Test set
Normal	209	20	22
Abnormal	2308	288	290
With data augmentation			
Class	Training set	Validation set	Test set
Normal	18,464	20	22
Abnormal	18,464	288	290

Table 3 Using the technique of data augmentation to the training dataset of stage 3

Without data augmentation			
Class	Training set	Validation set	Test set
Masses	1244	155	156
Calcifications	1064	133	134
With data augmentation			
Class	Training set	Validation set	Test set
Masses	24,880	155	156
Calcifications	24,880	133	134

different pretrained models, comparing the performance of each stage. The tables include overall test results as well as precision, recall, F1-score, and AUC results for each class. The ResNet-50 and EfficientNetB7 models achieved excellent results, with up to 99% accuracy attained when fine-tuning for stage 3. The VGG19 model also produced good results, especially when images were normalized with a range of 0–255, as shown in Tables 5 and 6.

5.3 Comparison with Previous Works

In this section, we compare our model with a few recent studies that were previously mentioned. The comparison is shown in Table 6, which displays the best results of our proposed models, namely, VGG-19, ResNet-50, ResNet-152, and EfficientNet-B7, compared to those of earlier studies with a common focus. It is noteworthy that EfficientNet-B7 performed the best in a binary classification between tumor images and healthy images. Additionally, Table 7 presents further information.

Table 4 Using the technique of data augmentation to the training dataset of stage 4

Without data augmentation			
Class	Training set	Validation set	Test set
Normal	209	22	22
Masses	1244	155	156
Calcifications	1064	133	134
With data augmentation			
Class	Training set	Validation set	Test set
Normal	19,904	20	22
Masses	19,904	155	156
Calcifications	19,904	133	134

6 Conclusions and Future Work

In conclusion, four pretrained deep convolutional neural networks (DCNNs) were compared for their effectiveness in classifying mammogram images using fine-tuning. The study utilized images from two datasets, Mini-MIAS and CBIS-DDSM, and ROIs were obtained by cropping images to help identify objects of interest with more accuracy. Transfer learning and fine-tuning were employed to improve the models' efficiency compared to state-of-the-art.

The four phases of the DCNNs were evaluated to determine their performance in an unrelated task, and ImageNet weights were incorporated to optimize the models. The third stage using improved ResNet-50 and EfficientNetB7 models generated remarkable results compared to state-of-the-art models. EfficientNetB7 is considered a better choice due to its high accuracy and efficiency, outperforming other models on various tasks.

However, the equipment used in this study had limitations, and the authors hope that future studies will use more specialized equipment to obtain quicker results and better comparisons. The authors also plan to compare the pathology of the images among normal, benign, and malignant and explore the possibility of applying the same networks to DCNN with a smaller input. Additionally, they plan to create a multimodal convolutional neural network and apply this classification task using different information from each DCNN to obtain various outcomes and integrate them for a final result that is more accurate and varied.

Acknowledgments César Eduardo Muñoz Chavez wants to thank CONACYT for the support of this research. Hermilo Sánchez-Cruz was partially supported by Universidad Autónoma de Aguascalientes, under grant PII22-5. Humberto Sossa thanks CONACYT and IPN under grants FORDECYT-PRONACES 6005 and SIP 20220226 for the financial support.

Table 5 The values obtained as a result of the training are shown in this table. Each stage’s final result is observed along with each DCNN only with normalized images

		DCNN performance when using normalized images				
DCNN	Class	Accuracy	Precision	Recall	F1-score	AUC
Mini-MIAS						
	Normal		0.9545	0.9545	0.9545	
VGG19	Tumor	0.93939	0.9091	0.9091	0.9091	0.9318
	Overall		0.9091	0.9091	0.9091	
	Normal		0.9545	0.9545	0.9545	
ResNet-50	Tumor	0.9393	0.9091	0.9091	0.9091	0.9318
	Overall		0.9091	0.9091	0.9091	
	Normal		0.9130	0.9545	0.9333	
ResNet152	Tumor	0.9091	0.9	0.8181	0.8571	0.8863
	Overall		0.9	0.8181	0.8571	
	Normal		0.9565	1	0.9777	
EfficientB7	Tumor	0.9696	1	0.9091	0.9523	0.9545
	Overall		1	0.9091	0.9523	
CBIS-DDSM						
	Calcification		0.8875	0.8656	0.8787	
VGG19	Masses	0.8896	0.8875	0.9102	0.8987	0.8879
	Overall		0.8897	0.8896	0.8896	
	Calcification		0.9147	0.8805	0.8939	
ResNet-50	Masses	0.9068	0.9006	0.9294	0.9148	0.9050
	Overall		0.9006	0.9294	0.9148	
	Calcification		0.9076	0.8805	0.8939	
ResNet152	Masses	0.9034	0.9	0.9230	0.9113	0.9018
	Overall		0.9	0.9230	0.9113	
	Calcification		0.9166	0.9029	0.9097	
EfficientB7	Masses	0.9172	0.9177	0.9294	0.9235	0.9162
	Overall		0.9177	0.9294	0.9235	
	Normal		0.9829	0.9829	0.9965	
VGG19	Abnormal	0.9807	0.9444	0.7727	0.85	0.8846
	Overall		0.9802	0.9807	0.9798	
	Normal		0.9965	0.9965	0.9965	
ResNet-50	Abnormal	0.9935	0.9545	0.9545	0.9545	0.9755
	Overall		0.9545	0.9545	0.9545	
	Normal		0.9931	1	0.9965	
ResNet152	Abnormal	0.9935	0.9989	0.9090	0.9523	0.9545
	Overall		0.9090	0.9523	0.9931	
	Normal		0.9931	1	0.9965	
EfficientB7	Abnormal	0.9935	0.9989	0.9090	0.9523	
	Overall		0.9989	0.9090	0.9523	
	Normal		0.8636	0.8636	0.9266	0.9266

(continued)

Table 5 (continued)

		DCNN performance when using normalized images				
VGG19	Calcification	0.9253	0.8731	0.8731	0.8731	0.8888
	Masses		0.8974	0.8974	0.8974	0.8974
	Overall		0.8846	0.8846	0.884	0.9042
	Normal		0.9987	0.9545	0.9767	0.9772
ResNet-50	Calcification	0.9006	0.9055	0.8582	0.8812	0.8953
	Masses		0.8841	0.9294	0.9062	0.9038
	Overall		0.9014	0.9006	0.9004	0.9255
	Normal		0.9988	0.8636	0.9268	0.9318
ResNet152	Calcification	0.9006	0.8776	0.9104	0.8937	0.9074
	Masses		0.9090	0.8974	0.9032	0.9038
	Overall		0.9020	0.9006	0.9008	0.9143
	Normal		1	0.7722	0.8717	0.8863
EfficientB7	Calcification	0.9102	0.875	0.9402	0.9064	0.9195
	Masses		0.9337	0.9038	0.9185	0.9198
	Overall		0.9132	0.9102	0.91007	0.9086

Table 6 The values obtained as a result of the training are shown in this table. Each stage’s final result is observed along with each DCNN using the CLAHE technique

		DCNN Performance using CLAHE				
DCNN	Class	Accuracy	Precision	Recall	F1-score	AUC
Mini-MIAS						
	Normal		0.9545	1	0.9767	
VGG19	Tumor	0.9696	1	0.9166	0.9565	0.9583
	Overall		0.9166	0.9565	0.9545	
	Normal		0.8333	0.8695	0.9091	
ResNet-50	Tumor	0.9091	0.9091	0.9523	0.9302	0.8928
	Overall		0.9091	0.8333	0.8695	
	Normal		0.8333	0.9091	0.9130	
ResNet152	Tumor	0.9393	0.913	1	0.9545	0.9166
	Overall		1	0.8333	0.9091	
	Normal		0.9166	0.9166	0.9523	
EfficientB7	Tumor	0.9393	0.9523	0.9523	0.9523	0.9345
	Overall		0.9166	0.9166	0.9166	
CBIS DDSM						
	Calcification		0.8617	0.7910	0.8249	
VGG19	Masses	0.8448	0.8323	0.8910	0.8606	0.8410
	Overall		0.8459	0.8448	0.8441	
	Calcification		0.8759	0.8955	0.8856	
ResNet-50	Masses	0.8931	0.908	0.8910	0.8996	0.8932

(continued)

Table 6 (continued)

		DCNN Performance using CLAHE				
	Overall		0.90844	0.8910	0.8996	
	Calcification		0.9076	0.8805	0.8939	
ResNet152	Masses	0.9034	0.9	0.9230	0.9113	0.9018
	Overall		0.9	0.9230	0.9113	
	Calcification		0.8978	0.9179	0.9077	
EfficientB7	Masses	0.9137	0.9281	0.9102	0.9190	0.9140
	Overall		0.9281	0.9102	0.9190	
	Normal		0.9830	1	v	
VGG19	Abnormal	0.9839	1	0.7727	0.8717	0.8863
	Overall		0.7727	0.8717	0.9830	
	Normal		0.9989	0.9989	1	
ResNet-50	Abnormal	1	0.9989	0.9989	1	1
	Overall		0.9989	1	0.9989	
	Normal		0.9931	1	0.9965	
ResNet152	Abnormal	0.9935	1	0.9090	0.9523	0.9545
	Overall		0.9090	0.9523	0.9931	
	Normal		1	0.9989	1	
EfficientB7	Abnormal	1	0.9989	0.9989	1	1
	Overall		1	0.9989	1	
	Normal		1	0.9090	0.9523	0.95454
VGG19	Calcification	0.8557	0.856	0.7985	0.8262	0.8486
	Masses		0.8383	0.8974	0.8668	0.8621
	Overall		0.8573	0.8557	0.8554	0.8884
	Normal		1	0.9090	0.9523	0.9545
ResNet-50	Calcification	0.9166	0.9029	0.9029	0.9029	0.9149
	Masses		0.9177	0.9294	0.9235	0.9230
	Overall		0.9171	0.9166	0.9167	0.9308
	Normal		0.9090	0.9090	0.9090	0.9510
ResNet152	Calcification	0.8974	0.8931	0.8731	0.8830	0.8972
	Masses		0.8993	0.9166	0.9079	0.9070
	Overall		0.8973	0.8974	0.8973	0.9184
	Normal		1	0.9090	0.9523	0.9545
EfficientB7	Calcification	0.8974	0.9126	0.8582	0.8846	0.8982
	Masses		0.8734	0.9294	0.9006	0.8974
	Overall		0.8992	0.8974	0.8973	0.9167

Table 7 Comparison of this methodology to others found in the literature

Author	Database	Model	Accuracy	AUC
Auccahuasi et al.	MIAS	Custom CNN	94%	None
Alkhaleefah et al.	CBIS-DDSM, MIAS, BCDR,	Multiple fine-tuned models	96.49%	0.994%
Saber et al.	INBreast	VGG16 and VGG19	97.1%	0.988%
Qasim et al.	Mini-MIAS	Custom CNN	99.4%	
Shen et al.	CBIS-DDSM, INBreast	ResNet-50 and VGG16		0.91%
Khamparia et al.	DDSM	Modified VGG Alexnet, VGG16, VGG19, MVGG, MobileNet, and ResNet-50	94.3%	
Nasir Khan et al.	CBIS-DDSM, MIAS	VGGNet, GoogleNet, ResNet	92.29%	0.93%
Ragab et al.	DDSM, CBIS-DDSM	Custom AlexNet with SVM	87.2%	0.94%
Our proposed model	CBIS-DDSM, Mini-MIAS	VGG19, ResNet-50, ResNet151, and EfficientNetB7	99.98%	0.99%

References

- Villarreal-Garza, C., Aguila, C., Magallanes-Hoyos, M.C., Mohar, A., Bargalló, E., Meneses, A., Cazap, E., Gomez, H., López-Carrillo, L., Chávarri-Guerra, Y., Murillo, R., Barrios, C.: Breast cancer in young women in Latin America: an unmet, growing burden. *Oncologia*. **18**(12), 1298–1306 (2013). <https://doi.org/10.1634/theoncologist.2013-0321>
- Villarreal-Garza, C., Mesa-Chavez, F., Plata de la Mora, A., Miaja-Avila, M., Garcia-Garcia, M., Fonseca, A., de la Rosa-Pacheco, S., Cruz-Ramos, M., García Garza, M.R., Mohar, A., Bargallo-Rocha, E.: Prospective study of fertility preservation in young women with breast cancer in Mexico. *J. Natl. Compr. Cancer Netw*, 1–8 (2021). <https://doi.org/10.6004/jnccn.2020.7692>
- Chávarri-Guerra, Y., Villarreal-Garza, C., Liedke, P.E., Knaul, F., Mohar, A., Finkelstein, D.M., Goss, P.E.: Breast cancer in Mexico: a growing challenge to health and the health system. *Lancet. Oncol*. **13**(8) (2012). [https://doi.org/10.1016/S1470-2045\(12\)70246-2](https://doi.org/10.1016/S1470-2045(12)70246-2)
- Houssein, E.H., Emam, M.M., Ali, A.A., Suganthan, P.N.: Deep and machine learning techniques for medical imaging-based breast cancer: a comprehensive review. *Expert Syst. Appl*. **167** (2021). <https://doi.org/10.1016/j.eswa.2020.114161>
- Pisano, E.D., Yaffe, M.J.: Digital mammography. *Radiol*. **234**(2), 353–362 (2005). <https://doi.org/10.1148/radiol.2342030897>
- Wang, J., Zhu, H., Wang, S.H., Zhang, Y.D.: A review of deep learning on medical image analysis. *Mobile. Netw. Appl*. **26**, 351–380 (2021). <https://doi.org/10.1007/s11036-020-01672-7>
- Dong, S., Wang, P., Abbas, K.: A survey on deep learning and its applications. *Comput. Sci. Rev*. **40** (2021). <https://doi.org/10.1016/j.cosrev.2021.100379>
- Zhang, X., Dahu, W.: Application of artificial intelligence algorithms in image processing. *J. Vis. Commun. Image Represent*. **61**, 42–49 (2019). <https://doi.org/10.1016/j.jvcir.2019.03.004>
- Hassabis, D., Kumaran, D., Summerfield, C., Botvinick, M.: Neuroscience-inspired artificial intelligence. *Neuron*. **95**(2), 245–258 (2017). <https://doi.org/10.1016/j.neuron.2017.06.011>

10. Bagchi, S., Huong, A.: Signal processing techniques and computer-aided detection systems for diagnosis of breast cancer – a review paper. *Ind. J. Sci. Technol.* **10**(3) (2017). <https://doi.org/10.17485/ijst/2017/v10i3/110640>
11. Batchu, S., Liu, F., Amireh, A., Waller, J., Umair, M.: A review of applications of machine learning in mammography and future challenges. *Oncologia.* **99**(8), 483–490 (2021). <https://doi.org/10.1159/000515698>
12. Mohanty, A.K., Senapati, M.R., Beberta, S., Lenka, S.K.: Texture-based features for classification of mammograms using decision tree. *Neu. Comput. Appl.* **23**(3–4), 1011–1017 (2013). <https://doi.org/10.1007/s00521-012-1025-z>
13. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: a review. *Neurocomputing.* **187**, 27–48 (2016). <https://doi.org/10.1016/j.neucom.2015.09.116>
14. Janiesch, C., Zscheck, P., Heinrich, K.: Machine learning and deep learning. *Electr. Mark.* **31**, 685–695 (2021). <https://doi.org/10.1007/s12525-021-00475-2>
15. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature.* **521**(7553), 436–444 (2015). <https://doi.org/10.1038/nature14539>
16. Fourcade, A., Khonsari, R.H.: Deep learning in medical image analysis: a third eye for doctors. *J. Stomatol. Oral. Maxillofac. Surg.* **120**(4), 279–288 (2019). <https://doi.org/10.1016/j.jormas.2019.06.002>
17. Mamoshina, P., Vieira, A., Putin, E., Zhavoronkov, A.: Applications of deep learning in biomedicine. *Mol. Pharm.* **13**(5), 1445–1454 (2016). <https://doi.org/10.1021/acs.molpharmaceut.5b00982>
18. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al.: A guide to deep learning in healthcare. *Nat. Med.* **25**(1), 24–29 (2019). <https://doi.org/10.1038/s41591-018-0316-z>
19. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data.* **4** (2017). <https://doi.org/10.1038/sdata.2017.177>
20. Suckling, J.P.: The mammographic image analysis society digital mammogram database. *Digital. Mammo.*, 375–386 (1994)
21. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning. *ICANN.* **11141**, 270–279 (2018). <https://doi.org/10.48550/arXiv.1808.01974>
22. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Member, S., Xiong, H., He, Q.: A comprehensive survey on transfer learning. *Proc. IEEE.* **109**, 43–76 (2019). <https://doi.org/10.48550/arXiv.1911.02685>
23. Falconi, L.G., Perez, M., Aguilar, W.G., Conci, A.: Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. *Adv. Sci. Technol. Eng. Syst.* **5**(2), 154–165 (2020). <https://doi.org/10.25046/aj050220>
24. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010). <https://doi.org/10.1109/TKDE.2009.191>
25. Rampun, A., et al.: Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network. *Med. Image Anal.* **57**, 1–17 (2019). <https://doi.org/10.1016/j.media.2019.06.007>
26. 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.* **12**(8), 1812 (2022). <https://doi.org/10.3390/diagnostics12081812>
27. Wei, B., Han, Z., He, X., Yin, Y.: Deep Learning Model Based Breast Cancer Histopathological Image Classification, pp. 348–353. *Int. Conf. Cloud. Comput. Big. Data. Anal.* (2017). <https://doi.org/10.1109/ICCCBDA.2017.7951937>
28. Aucahuasi, W., Delrieux, C., Sernaqué, F., Flores, E., Moggiano, N.: Detection of Microcalcifications in Digital Mammography Images, Using Deep Learning Techniques,

- Based on Peruvian Casuistry, pp. 1–4. E-Health. Bioeng. Conf (2019). <https://doi.org/10.1109/EHB47216.2019.8969906>
29. Alkhaleefah, M., Shang-Chih, M.A., Chang, Y.L., Huang, B., Chittam, P.K., Achhannagari, V.P.: Double-shot transfer learning for breast cancer classification from x-ray images. *Appl. Sci.* **10**(11), 3999 (2020). <https://doi.org/10.3390/app10113999>
 30. Charan, S., Khan, M.J., Khurshid, K.: Breast Cancer Detection in Mammograms Using Convolutional Neural Network, pp. 1–5. *iCoMET* (2018). <https://doi.org/10.1109/ICOMET.2018.8346384>
 31. Saber, A., Sakr, M., Abo-Seida, O.M., Keshk, A.: Tumor detection and classification in breast mammography-based on fine-tuned convolutional neural networks. *Int. J. Comput. Inf.* **9**(1), 74–84 (2022). <https://doi.org/10.21608/IJCI.2021.103605.1063>
 32. Qasim, K.R., Ouda, A.J.: An accurate breast cancer detection system based on deep learning CNN. *MLU.* **20**(1), 984–990 (2020). <https://doi.org/10.37506/mlu.v20i1.499>
 33. Montaha, S., Azam, S., Rafid, A.K.M.R.H., Ghosh, P., Hasan, M.Z., Jonkman, M., De Boer, F.: BreastNet18: a high accuracy fine-tuned VGG16 model evaluated using ablation study for diagnosing breast cancer from enhanced mammography images. *Biology.* **10**(12), 1347 (2021). <https://doi.org/10.3390/biology10121347>
 34. Allugunti, V.R.: Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *Int. J. Eng. Comput. Sci.* **4**(1), 56–49 (2022). <https://doi.org/10.33545/26633582.2022.v4.i1a.68>
 35. Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W.: Deep learning to improve breast cancer detection on screening mammography. *Sci. Rep.* **9**(1), 12495 (2019). <https://doi.org/10.1038/s41598-019-48995-4>
 36. Khamparia, A., Bharati, S., Podder, P., Gupta, D., Khanna, A., Phung, T.K., Thanh, D.N.H.: Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimens. Syst. Signal. Process.* **32**(2), 747–765 (2021). <https://doi.org/10.1007/s11045-020-00756-7>
 37. Hameed, Z., Zahia, S., Garcia-Zapirain, B., Aguirre, J.J., Vanegas, A.M.: Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors.* **20**(16), 4373 (2020). <https://doi.org/10.3390/s20164373>
 38. Khan, H.N., Shahid, A.R., Raza, B., Dar, A.H., Alquhayz, H.: Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. *IEEE Access.* **7**, 165724–165733 (2019). <https://doi.org/10.1109/ACCESS.2019.2953318>
 39. Ragab, D.A., Sharkas, M., Marshall, S., Ren, J.: Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ.* **7**, 6201 (2019). <https://doi.org/10.7717/peerj.6201>
 40. Xi, P., Shu, C., Goubran, R.: Abnormality Detection in Mammography using Deep Convolutional Neural Networks, pp. 1–6. *MeMeA* (2018). <https://doi.org/10.1109/MeMeA.2018.8438639>
 41. Hepsa, P.U., Özel, S.A., Yazıcı, A.: Using Deep Learning for Mammography Classification, pp. 418–423. *UBMK* (2017). <https://doi.org/10.1109/UBMK.2017.8093429>
 42. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks For-Large-Scale Image Recognition, pp. 1–14. *ICLR* (2015). <https://doi.org/10.48550/arXiv.1409.1556>
 43. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition, pp. 770–778. *CVPR* (2016). <https://doi.org/10.48550/arXiv.1512.03385>
 44. Tan, M., Le, Q.V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, pp. 6105–6114. *ICML* (2019). <https://doi.org/10.48550/arXiv.1905.11946>
 45. Géron, A.: *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems*, vol. 2. Shroff Publishers (2019)
 46. Chetlur, S., Woolley, C., VanderMersch, P., Cohen, J., Tran, J., Catanzaro, B., Shelhamer, E.: cuDNN: Efficient Primitives for Deep Learning. *ArXiv* (2014). <https://doi.org/10.48550/arXiv.1410.0759>

47. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going Deeper with Convolutions, pp. 1–9. CVPR (2015). <https://doi.org/10.1109/CVPR.2015.7298594>
48. Zuiderveld, K.: Contrast Limited Adaptive Histogram Equalization, pp. 474–485. Graphics Gems (1994). <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
49. Sharma, J., Rai, J.K., Tewari, R.P.: Identification of Pre-processing Technique for Enhancement of Mammogram Images, pp. 115–119. MedCom (2014). <https://doi.org/10.1109/MedCom.2014.7005987>
50. Iswardani, A., Hidayat, W.: Mammographic image enhancement using digital image processing technique. IJCSIS. **16**(5), 222–226 (2018). <https://doi.org/10.48550/arXiv.1806.11496>
51. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big. Data. **6**, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>
52. Oza, P., Sharma, P., Patel, S., Adedoyin, F., Bruno, A.: Image augmentation techniques for mammogram analysis. J. Imaging. **8**(5), 141 (2022). <https://doi.org/10.3390/jimaging8050141>
53. Oyelade, O.N., Ezugwu, A.E.: A deep learning model using data augmentation for detection of architectural distortion in whole and patches of images. BSPC. **65**, 102366 (2021). <https://doi.org/10.1016/j.bspc.2020.102366>
54. Wang, J., Perez, L.: The Effectiveness of Data Augmentation in Image Classification Using Deep Learning (2017). <https://doi.org/10.48550/arXiv.1712.04621>
55. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. Information. **11**(2), 125 (2020). <https://doi.org/10.3390/info11020125>
56. Russakovsky, O., Deng, J., Su, H., et al.: ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. **115**, 211–252 (2015). <https://doi.org/10.3390/info11020125>
57. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, pp. 248–255. CVPR (2009). <https://doi.org/10.1109/CVPR.2009.5206848>
58. Kingma, D.P., Lei Ba, J.: ADAM: a method for stochastic optimization. ICLR. (2015). <https://doi.org/10.48550/arXiv.1412.6980>
59. Badr, E.A., Joun, C., Nasr, G.E.: Cross Entropy Error Function in Neural Networks: Forecasting Gasoline Demand, pp. 381–384. In FLAIRS-02 Proceedings (2002)
60. Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. JMLR. **15**(1), 1929–1958 (2014)
61. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. ICML. **37**, 448–456 (2015). <https://doi.org/10.48550/arXiv.1502.03167>
62. Ng, A.Y.: Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance. ICML (2004). <https://doi.org/10.1145/1015330.1015435>