

EEF-Net: An Enhanced EfficientNet for breast tumor classification in mammograms

Dishant Padalia, Kush Vora, Darshil Mehta and Ninad Mehendale*

ARTICLE INFO

Keywords:
Neural Networks
Image Processing
Breast Cancer
Radiology

ABSTRACT

Breast cancer is one of the most common cancer types, and treatment largely depends on early detection. A person has a 13% (1 in 8 risks) of developing breast cancer at some time in their lives. Computer-aided diagnostic (CAD) systems powered by deep learning algorithms have enabled data analysis at high rates without compromising performance. Our work proposes an Enhanced EfficientNet (EEF-Net) for determining the severity of breast cancer through mammograms. EEF-Net is built on top of EfficientNet and has been fine-tuned to classify mammograms into three classes: benign, malignant, and healthy. The architecture was trained using the publicly accessible MIAS dataset, and a sophisticated image pre-processing pipeline was used to remove noise and other artifacts from the mammograms. The model achieved state-of-the-art results in the classification of breast cancer, achieving an accuracy of 97.14%, 98.67% sensitivity, 99.30% specificity, and 98.30% precision. EEF-Net will assist radiologists in mass screening patients with high precision and will minimize the radiologists' workload.

1. Introduction

Cancer mayoclinic.org is a wide range of diseases characterized by the uncontrollable division of abnormal cells with the ability to interfere and destroy normal human tissue. Cells die as they age or become damaged, and new cells supplant the existing ones. However, when this well-ordered system breaks down, abnormal or damaged cells grow and start reproducing. As a result, these damaged cells unite and produce tumors, which are tissue lumps. These tumors can be benign or cancerous. Cancer has the ability to spread throughout the body and is the world's second largest cause of mortality.

Breast cancer is one such form of cancer which affects both men and women but is far more common in women. In the year 2020 itself, there were 2.3 million who.int new cases of breast cancer with over 680,000 deaths globally. As of the end of 2020, 7.8 million women had been diagnosed with breast cancer in the previous five years, surpassing lung cancer and making it the world's most commonly diagnosed malignancy. Currently, over 3.8 million cancer.net women in the United States have been diagnosed with breast cancer. In the United States in 2022, it is expected that 287,850 cancer.net women will be diagnosed with invasive breast cancer and 51,400 women will be diagnosed with non-invasive breast cancer. Men account for 0.5-1% of all breast cancer cases. In the year 2022, an estimated 2,710 men in the United States will be diagnosed with metastatic breast cancer.

Breast cancer develops in the breast tissue and occurs when breast cells mutate (change) and multiply uncontrollably, resulting in a mass of tissue (tumor). Breast cancer, like other cancers, has the ability to infiltrate and grow in the tissue around the breast. There are various forms of breast cancer, but ductal carcinoma is the most frequent, account-

ing for more than 80% of all occurrences. This cancer develops in the milk ducts and spreads to surrounding breast tissue after breaking through the duct wall. Lobular cancer begins in the breast lobules (where milk is generated) and spreads to the surrounding breast tissue. This form of cancer accounts for 10% to 15% of all breast cancers. Inflammatory breast cancer is an uncommon and serious kind of cancer that seems to be an infection. Redness, swelling, pitting, and dimpling of the breast skin are common symptoms of inflammatory breast cancer. Obstructive cancer cells in their skin's lymph veins cause it. Breast cancers can form in other regions of the breast besides the milk ducts and lobules, however these types of cancer are less prevalent. Angiosarcoma, which develops in the cells that line blood or lymph vessels, and Phyllodes tumors, which begin in connective tissues and are usually benign, are two such examples.

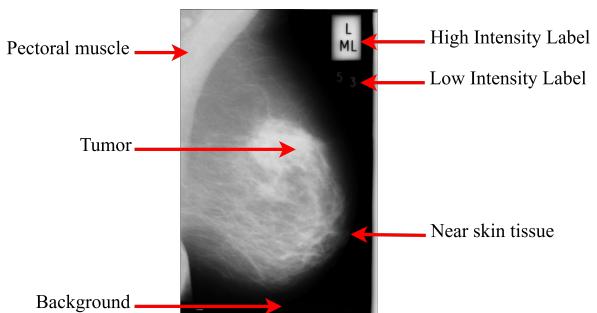


Figure 1: The mammogram above shows different types of artifacts, muscle/tissues, and tumor. On a mammogram, a lump or tumor will appear as a concentrated white area.

Breast cancer diagnostic tests include mammography, positron emission tomography (PET) scans, ultrasonography, and magnetic resonance imaging (MRI). However, mammography III (1976) is the most common and preferred approach because it makes use of the fact that a large portion of the breast contains fat tissue, which is generally transpar-

*Corresponding author (ninad@somaiya.edu)

✉ dishant.padalia@somaiya.edu (D. Padalia); kush.v@somaiya.edu (K. Vora); darshil05@somaiya.edu (D. Mehta)

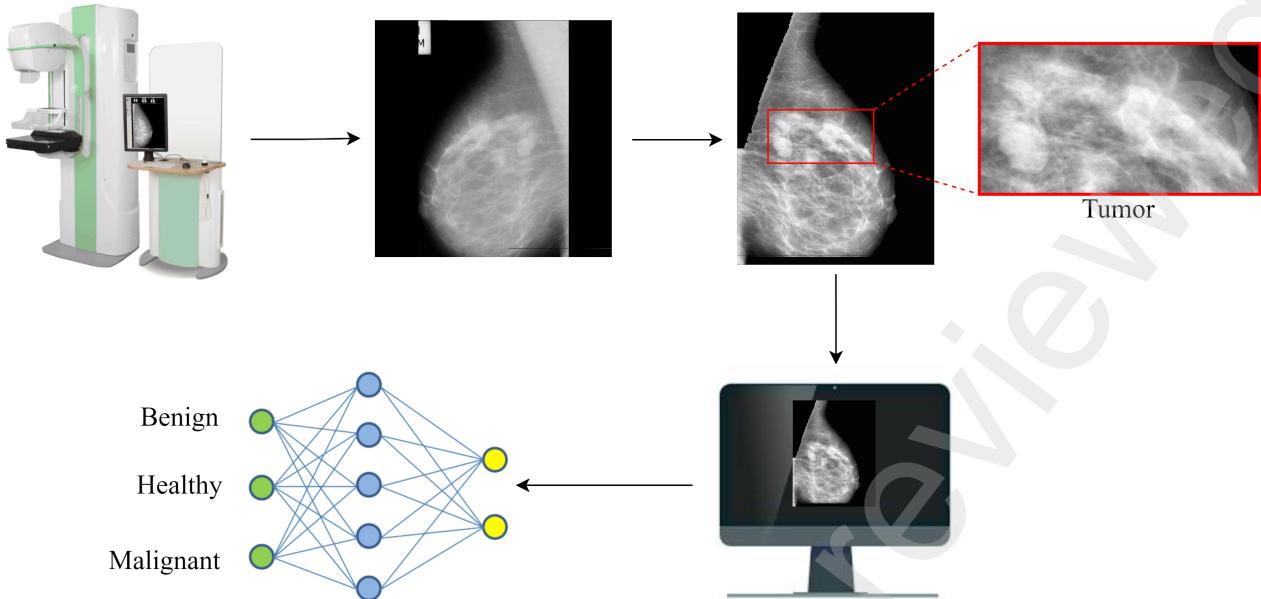


Figure 2: Concept diagram of the presented work. A mammographic machine is used to take the patient's breast X-ray (mammogram). The captured mammogram is then processed by a computer-aided diagnostic (CAD) system, which employs the EEF-Net to accurately and efficiently classify the image into three distinct classes.

ent to the low doses of radiation employed in mammography. It is a low-dose x-ray technique that enables one to examine the internal structure of the breast. In addition, it can detect very subtle changes in tissue architecture, as well as minor calcium deposits, which may indicate the formation of a tumor in some situations. Mammograms today can detect lesions as small as 1 centimeter (approximately half an inch) in diameter. Breast cancer typically causes no pain in its initial stages when it is curable, which is why screening is vital for early detection and treatment. Different artifacts and muscle/tissues are shown in Figure 1.

Manual diagnosis of breast cancer from mammograms is an exigent task and requires many years of experience in the field of radiology. To overcome this predicament, deep learning and artificial intelligence methods are used in medical imaging to diagnose diseases from scans or images. Convolutional Neural Networks Suzuki (2017) are a powerful feature extractor; therefore, using them to classify medical images can avoid tortuous feature engineering. Our work proposes an enhanced EfficientNet Tan and Le (2019) (EEF-Net), which classifies a breast mammogram into three classes: benign, malignant, and healthy. EEF-Net uses EfficientNet as the backbone, which has been trained on the ImageNet Deng, Dong, Socher, Li, Li and Fei-Fei (2009) dataset consisting of 1.2 million images, and a few layers have been added on top of this model to fine-tune it for this particular task. Our network has achieved state-of-the-art results in the classification of the stage of breast cancer, achieving accuracy, sensitivity, specificity, and a precision score of 97.14%, 98.67%, 99.3%, and 98.3%, respectively. Figure 2.

represents the concept diagram of the presented system. It demonstrates how the proposed architecture can be used in a real time setting to generate accurate detection of breast cancers.

2. Literature Review

The ability to classify a tumor on a mammogram as benign or malignant is critical in assisting experts in determining the severity of breast cancer and the best course of treatment, with the advent of machine learning algorithms and deep learning techniques, several computer systems that aid experts in reliably identifying breast cancer have been developed.

Q. Abbas et al. Abbas (2016) developed a computer-aided diagnostic (CAD) system that distinguishes benign from malignant mammographic masses. The system is divided into four phases. Each mammogram is examined, and speed-up robust features (SURF) and local binary pattern variance (LBPV) descriptors are extracted. The deep invariant features (DIFs) are then built in supervised and unsupervised forms using a multilayer deep-learning architecture. A fine-tuning step is incorporated to determine the features, and the final decision is made using the softmax linear classifier. On a dataset of 600 masses (300 benign and 300 malignant), the CAD system attained an accuracy of 91.5%. The work presented by Prabhpreet Kaur et al. Kaur, Singh and Kaur (2019) uses K-means clustering to select Speed-Up Robust Features (SURF). In addition, a new layer is added at the classification level, which carries out the training and

testing of the deep neural network and Multi-Support Vector Machine (MSVM). The results of the proposed system using K-means clustering with MSVM were compared with a system using the decision tree model. The quantitative comparison of the two approaches showcases that the deep learning-based technique is better than other state-of-the-art techniques. A. Papadopoulos et al. Papadopoulos, Fotiadis and Likas (2005) proposed a novel computer-based technique for classifying microcalcification clusters. The proposed architecture was trained on the Nijmegen and MIAS datasets for classifying benign and malignant breast cancer. Preprocessing techniques such as background removal and contrast enhancement were performed, leading to better image classification. After preprocessing the images, features were extracted using Artificial Neural Networks (ANN). PCA and enhanced PCA were incorporated to reduce the feature vector size. Neural networks and Support Vector Machines were used to classify the images, yielding an accuracy of 73% and 79%, respectively.

The convolution neural network (CNN) made a remarkable advance when GoogleNet employed it to diagnose cancer with an accuracy of 89%, whereas human pathologists could only attain an accuracy of 70%. This resulted in the application of CNNs in various medical domains. CNN models have achieved state-of-the-art metrics in diagnosing breast cancer. S. Charan et al. Charan, Khan and Khurshid (2018) employed convolution neural networks (CNN) to classify breast mammograms as normal or abnormal. The MIAS dataset was divided into seven classes, six representing the type of abnormality, and was split into a 70-30 ratio for training and testing. The images were preprocessed by resizing them to 224 x 224, morphological closing was applied for noise removal, and masking was performed at the end to set the background pixels to 0. The above-preprocessing steps were performed to remove artifacts from the mammograms. A 7-layer CNN architecture was trained on original images as well as the preprocessed dataset. The architecture trained on the preprocessed images performed better and learned faster than the one trained on the original images and achieved an overall accuracy of 65% on the MIAS dataset. The CNN-based system presented by F. Ting et al. Ting, Tan and Sim (2019) achieved a sensitivity, accuracy, and specificity of 89.47%, 90.50%, and 90.71%, respectively. P. Hepsag et al. Hepsağ, Özel and Yaz (2017) observed an increase in accuracy from 65% to 85% after performing region of interest (ROI) extraction and training the CNN architecture on these masks. Transfer learning (TL) with convolutional neural networks increased the performance of several CNN architectures. Transfer learning learns on a new task by leveraging prior knowledge of similar tasks. It has made a significant contribution to medical image analysis by overcoming the problem of data paucity while also saving time and hardware resources. A. Saber et al. Saber, Sakr, Abo-Seida, Keshk and Chen (2021) proposed a deep learning model based on the transfer learning technique to detect breast cancer from a breast mammogram. In the proposed work, various preprocessing steps

such as noise removal, improving contrast, and non-breast region removal were performed to determine the area with cancer. This was followed by data augmentation to increase the number of training samples. Several pre-trained CNN viz. VGG-19, VGG-16, InceptionV3, ResNet50, and others, were trained on the preprocessed dataset. The VGG-16 model obtained the best results with an accuracy of 98.96%, sensitivity of 97.83%, specificity of 99.13%, and area under the curve (AUC) of 0.995.

In the classification task of mammogram scans, CNN-based systems outperformed machine learning-based systems; however, a few difficulties remain. These challenges include semantic feature ignorance, missing patches in reduced contrast mammography images, and segmentation ambiguity. The work proposed by S. Malebary et al. Malebary and Hashmi (2021) addresses these issues by presenting architecture that ensembles various machine learning and deep learning models like K-means clustering, Long Short-Term Memory (LSTM) network of Recurrent Neural Network (RNN), CNN, random forest, and various boosting techniques to classify the breast mass as benign, malignant or healthy. This research study utilized two datasets, MIAS and DDSM, containing 322 and 2620 images, respectively. The proposed BMC (Breast Mass Classification) system achieved 96% and 95% accuracy for DDSM and MIAS datasets.

The presence of artifacts, labels, and pectoral muscle in breast scans hamper learning of various deep learning algorithms. To overcome this problem, a number of preprocessing strategies that segment or remove these artifacts have been introduced. V. Singh et al. Singh, Srivastava and Srivastava (2018) introduced a content-based image retrieval technique for digital mammography. This study refines the dataset by performing three significant steps: label suppression, pectoral muscle removal, and smoothing and extraction of region of interest (ROI). Label suppression was performed by removing the isolated pixels surrounded by eight black pixels. In addition, a region-growing algorithm was implemented with a very low threshold to obtain the segmented pectoral muscle. The generated mask (segmented muscle) is then used to remove the pectoral muscle, followed by smoothing. Finally, feature sets that contain statistical, shape, wavelet, Gabor and GLCM features were used as the feature vectors corresponding to the image. These feature vectors were ranked in a non-decreasing order based on the Euclidean distance similarity measure. D. Zebari et al. Zebari, Zeebaree, Abdulazeez, Haron and Hamed (2020) proposed a method for segmenting the breast-boundary region and pectoral muscle region detection. Thresholding was used for primary segmentation, followed by morphological operations and masking to correct the over-estimation of the boundary by deleting the small objects. A Histogram of Oriented Gradient (HOG) was employed to increase the segmentation accuracy. The presented work was implemented over the mini-MIAS, INbreast and BCDR datasets. An accuracy of 99.31% was obtained for predicting the boundary of the breast region, whereas 98.64% was the accuracy for

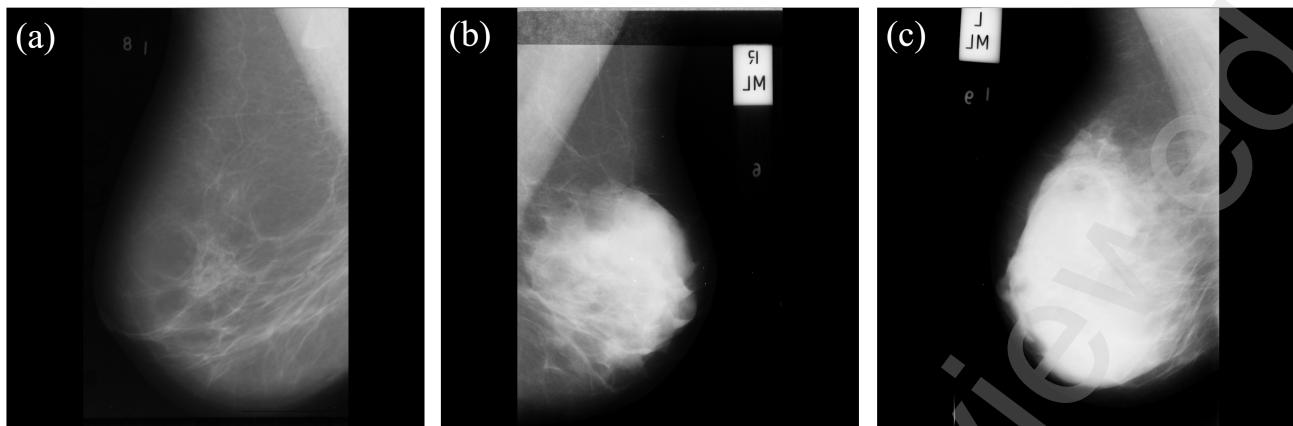


Figure 3: Image samples of the MIAS dataset used in the proposed work. (a) mammogram of a patient with no breast cancer, (b) mammogram of a patient with benign tumor, (c) mammogram of a patient with a malignant tumor.

the prediction of pectoral muscle segmentation. M. Mustra et al. Mustra and Grgic (2013) claim that segmentation of the breast skin-air interface performed on scanned mammograms is a more complicated process than digital mammograms. First, the left breast Medio-Lateral Oblique (MLO) images were flipped horizontally, followed by thresholding with a minimum and maximum intensity equal to 20% and 90% of the maximum intensity, respectively. In order to get rid of artifacts (orientation tags/low-intensity labels) morphological operations were performed with a square mask of 101 pixels. Hough transform and the Sobel edge detection filter was used to detect the most extended straight lines of the tissue edge.

3. Materials and methods

3.1. MIAS dataset

An association of UK research groups interested in analyzing mammograms has created a database of digital mammograms called the MIAS (The Mammographic Image Analysis Society) dataset. The dataset contains 330 images, of which 207 are healthy with no signs of breast cancer, 69 cases of benign tumors, and 34 of malignant tumors. Every image in the database is arranged in pairs (total of 165 pairs), each representing a single patient's left and right mammogram and all the images have been centered in the matrix with a size of 1024 x 1024 pixels and are stored in Portable Gray Map (PGM) format. In addition to the mammograms, the dataset includes information on the character of background tissue, the type of abnormality present, the severity of abnormality, x and y coordinates of the center of abnormality, and the radius (in pixels) of a circle enclosing the cancerous area. Using the center's coordinates and circle's radius enclosing the abnormality, one can extract the region of interest (ROI) and use it for various deep learning and statistical use cases. This work uses the dataset to develop a computer aided diagnostic system (CAD) which can classify a mammographic scan as malignant or benign. Sample images from all the three classes can be seen in Fig. 3.

3.2. Data Preparation

The mammograms in the MIAS dataset contain several artifacts that can hamper the deep learning model's learning capability. Hence, it is necessary to remove these artifacts to minimize negative results generated by the model. The proposed study leverages various pre-processing steps, viz. flipping, cropping, removing labels from the mammogram, segmenting the pectoral muscle, etc. to make the images in the dataset uniform and free from superfluous artifacts. Figure 4. depicts the entire pre-processing pipeline and shows the output image generated after each step described below.

3.2.1. Flipping and cropping

The dataset contains 329 images, each representing the left and right mammogram of a patient. All the left-facing mammograms were flipped to the right to maintain images with a similar orientation. Apart from the direction of the mammogram, images varied in terms of size. All the images were resized to 1024 x 800 pixels. In images with a width less than 800 pixels, the image was padded by adding black pixels, and the larger images were cropped to maintain the 1024 x 800 image shape.

3.2.2. Removing artifacts and redundant pixels

Each mammogram has a label in the upper right corner. It is crucial to remove these labels before training the model on the mammogram. This was accomplished by locating the start and end coordinates of the brightest portion in the image's upper right corner. Between the start and end points, all pixels were set to zero. Furthermore, the black bars around the photos were trimmed to reduce redundant pixels.

3.2.3. Image Enhancement using CLAHE

Varying light intensity and camera lens contamination cause blurry photos, making feature extraction challenging. This can be avoided by employing neighborhood image processing techniques to improve the image. Contrast Limited Adaptive Histogram Equalization Zuiderweld (1994) is one

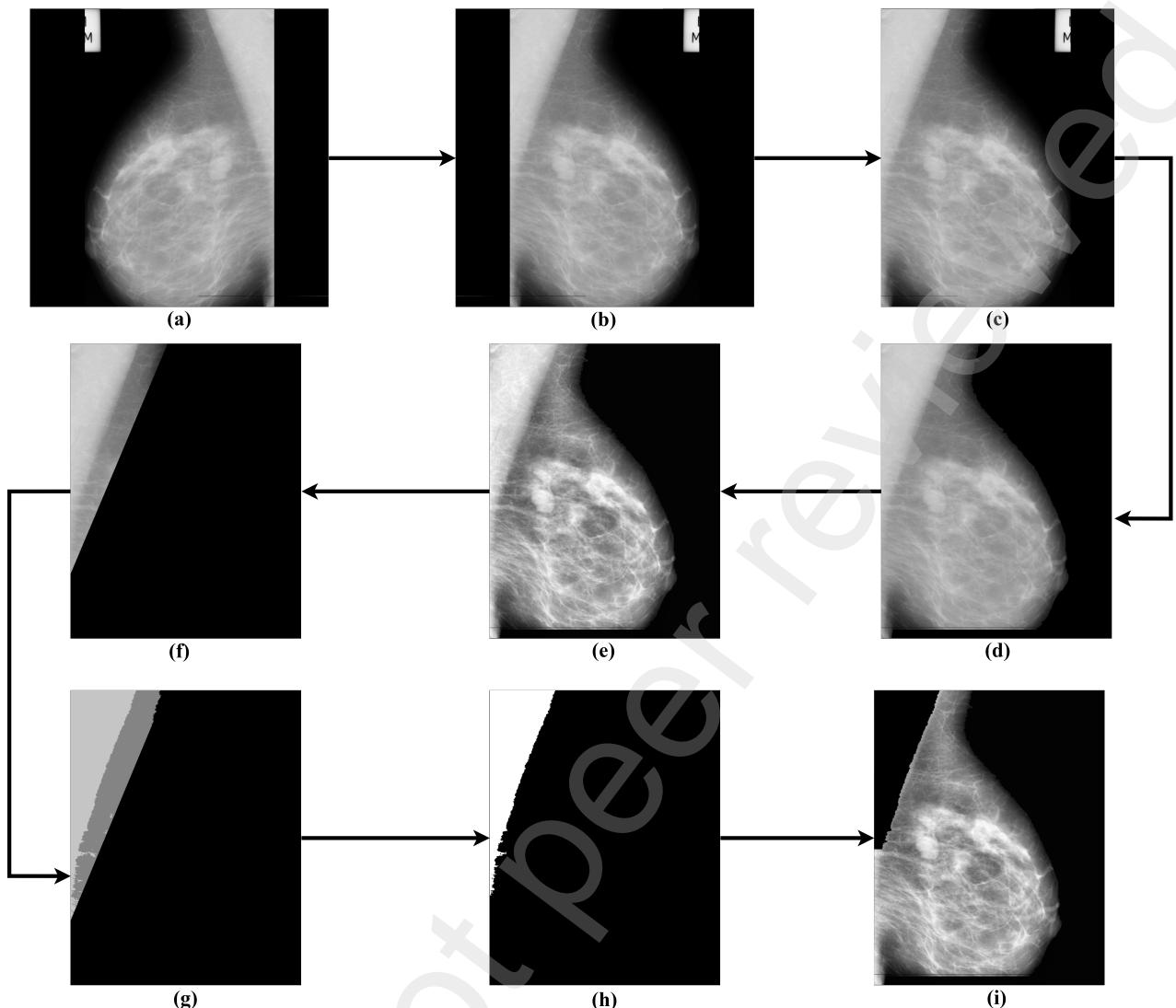


Figure 4: Output generated after each pre-processing step. (a) original mammogram in the MIAS dataset, (b) flipping to the right orientation, (c) cropping of left vertical black bar and redundant pixels, (d) removal of artifacts and duct tape marks, (e) image enhancement using CLAHE, (f) extraction of ROI, (g) applying k-means to segment the pectoral muscle from other parts of the image, (h) extraction of pectoral muscle, (i) removal of pectoral muscle from enhanced image.

such approach (CLAHE). It enhances pixels based on their immediate neighbors. To achieve the best contrast for the photos, different grid widths for histogram equalization were tried. To create the best contrast, the clip and grid sizes were set to 2 and 8, respectively.

3.2.4. Removing pectoral muscle

All the images have a right-side orientation, and hence the pectoral muscle is always at the top-left corner of the mammogram. The density of the pectoral muscle is similar to the density of the cancer tissue; thus, getting rid of the pectoral muscle becomes necessary. The pectoral muscle is removed by carrying out the following steps.

The first step for removing the pectoral muscle was crop-

ping the region of interest (ROI). This was achieved by finding the end of the bright area. After finding this point, a polygon was bounded by the point [800,0] (the pectoral muscle is always above this point), the top-right corner, the bottom-left corner and the bottom-right corner of the mammogram. Removing this polygon from the original mammogram leaves us with a triangular region in the top left, which contains the pectoral muscle. Figure x. shows the bounding polygon on a mammogram.

The K-Means Hartigan and Wong (1979) clustering algorithm is an unsupervised algorithm that separates the area of interest from the background. Based on the K-centroids, it clusters or partitions the given data into K-clusters or sections. This algorithm is used when the data is unlabeled (i.e.

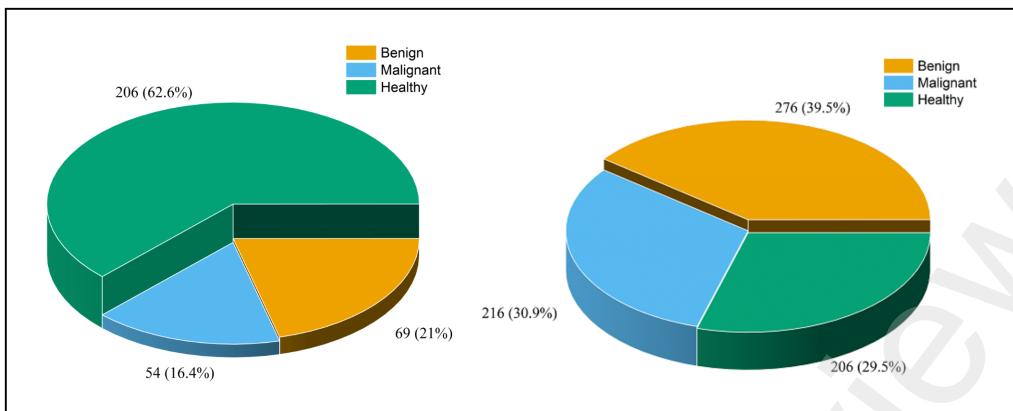


Figure 5: The distribution of classes in the MIAS dataset before and after augmentation. The initial collection contained 329 pictures, many of which were healthy mammograms. The augmented dataset contains 698 photos with evenly distributed classes.

data without defined categories or groups). We performed the K-means algorithm on the extracted triangular region to cluster three different regions. One of the three-segmented regions represents the pectoral muscle, while the other two represent a different muscle.

To extract the pectoral muscle, we apply the region growing algorithm. A seed point (starting point of the region growing algorithm) is obtained by traversing the image and finding the coordinate with maximum pixel intensity. All the points connected to the seed point are combined, and a binary image is returned. The width of the pectoral muscle decreases as we go down. This observation is used to find the point where the width of the bright intensity pixels (after a steady decrease) see an increase. The region below this point is set to zero (black).

In the final step, the extracted pectoral muscle is subtracted from the original enhanced image (one obtained after performing the CLAHE operation) to give the resultant image without the pectoral muscle.

3.2.5. Resizing and Augmentation

As a part of the last pre-processing step, the images were resized to 224 x 224. However, the dataset was unbalanced and contained a disproportionate number of samples from the ‘healthy’ class. Of 329 sample images, 206 were healthy (class 1), 54 malignant (class 2), and 69 benign (class 3). Deep learning models often over-classify the larger class(es) because of their higher prior probability when there is a class imbalance in the training data. As a result, instances belonging to the smaller class(es) are more frequently misclassified than those belonging to the larger class (es). To overcome this, 4-fold augmentation was performed on the images from the benign and malignant classes. Data augmentation not only helps overcome the class imbalance problem but also increases the number of training samples. Rotation was performed on the original images at angles of 5, 15, and 25 degrees. These three newly created images, along with the original image, were used for training. The new augmented dataset has 698 images – 206 healthy, 216 malignant, and

276 benign. Class distribution before and after data augmentation can be observed in Figure 5.

3.2.6. Training Data Preparation

The dataset was divided into train and test sets for training on the deep learning model and evaluating the model’s performance on unseen data. After preprocessing and augmentation, the dataset consists of 698 mammograms, each having one of three classes, benign, malignant or healthy . These 698 images were split into an 80:20 ratio for the number of training to testing images.

3.3. Model architecture

A Convolutional Neural Network O’Shea and Nash (2015) (ConvNet/CNN) is a deep learning system that can take an input image, give priority to distinct aspects/objects in the image (learnable weights and biases), and differentiate one from the other. CNN requires far less pre-processing than other classification techniques. Furthermore, although filters in primitive techniques are hand-crafted, ConvNets may acquire these features with sufficient training. The design of a ConvNet is similar to the connecting structure of neurons in the human brain and was inspired by the Visual Cortex layout. CNN’s role is to compress the pictures into a more manageable format while keeping the features required to provide an accurate prediction.

Convolutional Neural Networks employ various layers with differing functionality. Convolutional Layer, Pooling Layer, and Fully-Connected Layer are the three main types of layers. A convolutional layer is the primary building element of ConvNet. It consists of a series of filters and kernels whose parameters must be learned during training. The filters are usually smaller in size than the image itself. Each filter interacts with the image to generate an activation map. The pooling layer is inserted between consecutive convolutional layers, and its role is to gradually lower the spatial size of the representation, reducing the number of parameters and computation in the network and thereby controlling overfitting. In a fully connected layer, the neurons have com-

plete connections to all the activations in the prior layers. In the presented study, various transfer learning models and numerous models were developed from scratch and evaluated to find a model which performs the best on this specific task. In machine learning, transfer learning refers to the use of a previously learned model on a new task. In transfer learning, a computer leverages past experience to enhance prediction about a new task. In the proposed work, several state-of-the-art architectures, viz. VGG19 Simonyan and Zisserman (2014) , DenseNet169 Huang, Liu, Maaten and Weinberger (2017) , InceptionV3 Szegedy, Vanhoucke, Ioffe, Shlens and Wojna (2016) , ResNet50 He, Zhang, Ren and Sun (2016) , ResNet101, ResNet152 and EfficientNets are fine-tuned and evaluated on the MIAS dataset. A qualitative comparison between these architectures can be seen in Table 1.

This research presents an enhanced EfficientNet Tan and Le (2019) (EEF-Net) model for the multi-class classification of breast mammography images. The EfficientNet architecture was proposed by Mingxing. Tan and Quoc V. Leand of Google AI. In general, the models are either too deep, or have a very high resolution. Increasing some attributes initially improves the model, but it rapidly saturates, resulting in a model with large nuber of parameters that is inefficient. However, in an EfficientNet, they are scaled more methodically (everything is progressively raised). Furthermore, EfficientNet employs a compound coefficient ϕ that is used to uniformly scale network's width, depth, and resolution. The base of the EfficientNet is based on an inverted bottleneck residual blocks of MobileNetV2 Sandler, Howard, Zhu, Zhmoginov and Chen (2018) and MnasNet along with squeeze-and-excitation blocks.

The first part of EfficientNet is called the stem, which comprises seven layers, and the top consists of three layers. The stem and top are the same for all the EfficientNet models, and all the experimenting and modifications happen in the middle layers. The first layer in the stem is the input layer and it takes an input image of dimension 224 x 224 x 3 (RGB image). The next layer is a Rescaling layer that rescales the input values to a new range; consequently, the next layer is a Normalization layer that helps to standardize the data into a specific spectrum; this is followed by a Zero Padding layer, this aids in adding zeros around the input matrix and is commonly used to modify the image size as per the requirement of the model. The last three layers of the stem are the Convolutional layer, Batch Normalization layer and Activation layer. Batch normalization keeps the mean output close to 0 and the standard deviation (SD) of the output close to 1. Swish Ramachandran, Zoph and Le (2017) (SiLu), ie. Sigmoid Linear Units activation function is used in the EfficientNet architecture. The activation of the SiLU is computed by the sigmoid function multiplied by its input. The Swish activation function was proposed by Google Brain in 2017 and is a gated version of the sigmoid activation function. It outperforms ReLU Agarap (2018) activation in very deep neural networks and has non-monotonic property, i.e. for all $x < 0$, swish is neither a decreasing function nor an increasing function.

EfficientNet comprises a total of 238 layers. The layers in the model are divided into three modules, as shown in Figure 6. The first module incorporates a Depthwise Convolutional layer, Batch Normalization layer and activation function. Depthwise convolution is a variant of convolution that convolves each input channel with a separate kernel. This module is used as a starting point for the sub-blocks. The second module consists of 7 layers and is used as a starting point for the first sub-block of all seven main blocks except the 1st one. The third module comprises the Global Average Pooling layer, rescaling layer, and two convolutional layers; this module is connected as a skip connection to all the sub-blocks. These modules are combined to form the core of EEF-Net architecture. There are seven main blocks between the stem and top of the model, each utilizing the above modules. Additional layers are annexed to Efficient-Net architecture to customize and enhance the model for the specific research problem. The top three layers of the model are detached, and the output of block seven is flattened using the Flatten layer; it collapses the spatial dimensions of the input into the channel dimension. This flattened output is passed to a fully-connected dense layer with 512 units and ReLU activation function, which then is passed to a Batch Normalization layer and then a Dropout layer with a dropout rate of 20 percent. Dropout is a technique where randomly selected neurons are ignored during training, preventing the overfitting of the model. Another two similar sets of Dense, Batch Normalization and Dropout layers are added with 256 and 128 units, respectively. Lastly, the output is passed to the output layer with three neurons (because there are three classes) having a softmax activation function.

Apart from EEF-Net, CNN-LSTM also performed remarkably on the MIAS dataset. A CNN-LSTM is developed by combining CNN layers on the front, LSTM layers on the top, and a Dense layer in the output. In this architecture, CNN is employed for deep feature extraction, and LSTM is used for detection using the extracted feature. In this architecture, the stem comprises three sets of Convolutional, Max Pooling, and Batch Normalization layers followed by three fully connected Dense layers; lastly, these layers are connected to two LSTM layers that interpret the features extracted by the CNN layers. The output of the final LSTM layer is passed to a dense output layer with three units and a softmax activation function for classification. This model achieved an accuracy of 96.43% on the test set.

3.4. Model training

TensorFlow Abadi, Barham, Chen, Chen, Davis, Dean, Devin, Ghemawat, Irving and Isard (2016) framework was used to train and evaluate the performance of the deep learning architectures presented in this study. It is an open-source library developed by Google mainly for deep learning use cases. It is a complete suite of tools and resources which can be used to build and deploy machine learning-powered applications/systems. The proposed study was performed on a Windows 10 machine with 96GB of DDR4 RAM. The system has a Quadro RTX 500 GPU with CUDA 11.2 and

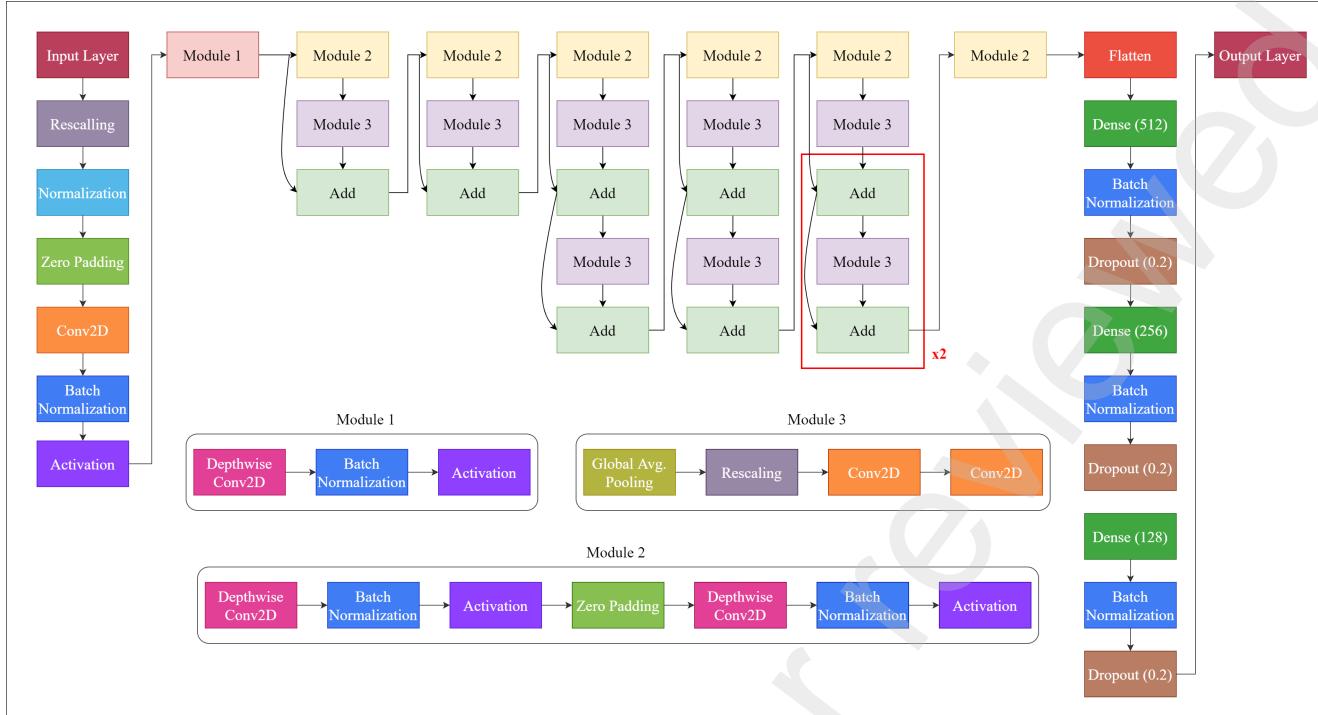


Figure 6: Architecture diagram of the EEF-Net. The stem consists of seven layers, in order, Input layer, rescaling layer, normalization layer, Zero-padding layer, convolutional layer, batch normalization layer and activation layer. The central part of the model is developed using the three modules. The top consists of three sets of dense, batch normalization and dropout layers. Finally, this output is given to the output layer with softmax activation for classification.

CudNN 8.1. Two machines with the above specifications were used to perform the data pre-processing and the model training.

The images of the MIAS dataset were pre-processed using the techniques mentioned in section 4.1. Post pre-processing, these mammograms were resized to 224 x 224 and were augmented to overcome the class imbalance problem. The EEF-Net was trained for a total of 200 epochs with early stopping. Early stopping allows one to stop training the model, when the performance stops improving on a given metric.

An optimizer is a function that is used to adjust the weights and the learning rate of a neural network, reducing the total loss and helping converge to minimia. In the presented study, Adam Kingma and Ba (2014) was used as the optimizer. Adam is a stochastic gradient descent approach that is based on the adaptive estimate of first- and second-order moments. A learning rate (an amount that the weights are updated during training) of 0.001 gave the most optimum convergence for the cost function. Since EEF-Net predicted three different classes: healthy, malignant and benign, the categorical cross-entropy was used as the loss function. The equation of the loss function is as follows:

Categorical Cross Entropy :

$$Loss = - \sum_{i=1}^N y_i \cdot \log \hat{y}_i$$

The softmax activation function is used in the output layer to predict the class of the image. It computes the probabilities of each class (the total probability equals one), and the label with the highest probability is chosen as the output. The equation of the activation function is as follows:

Softmax Activation Function :

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$$

3.5. Performance Metrics

Different metrics are used to evaluate the model's performance or quality, and these metrics are known as performance metrics. These performance metrics allow us to see how well our model performed with the given data. By tuning the hyper-parameters, we can improve the model's performance. Each deep learning model strives to generalize well on previously unseen data, and performance metrics aid in determining how well the model generalizes on the new dataset. The proposed study uses the following four metrics to evaluate the CNN architectures: accuracy, precision, recall, and F1 score.

3.5.1. Accuracy

Accuracy is the number of instances correctly classified by a learning model on a particular dataset. It is calculated by dividing the total number of correct predictions by the

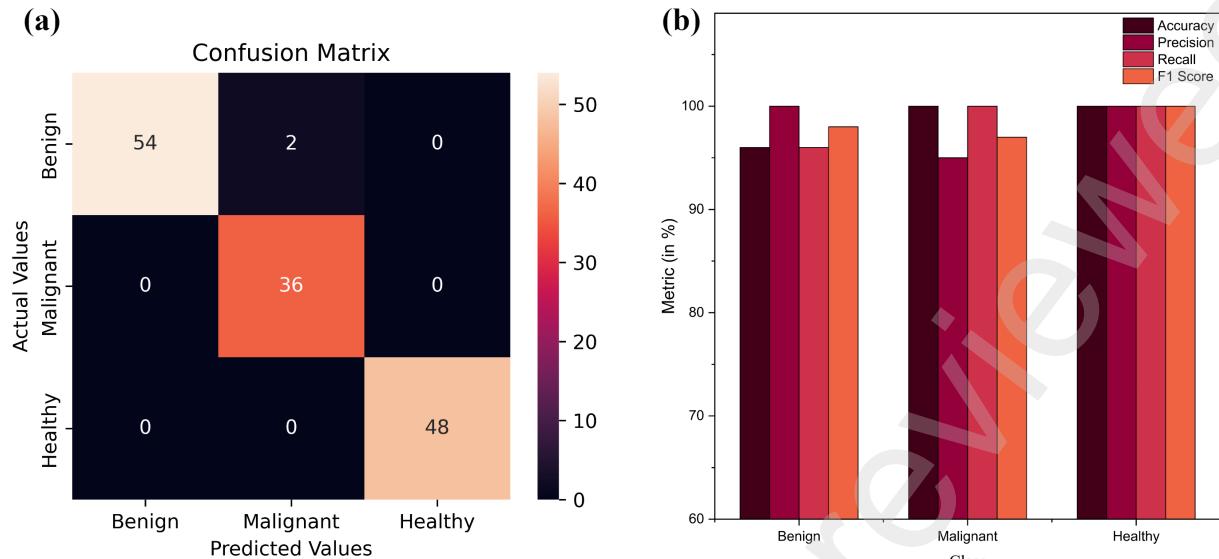


Figure 7: Results achieved after evaluating the EEF-Net on the testing set. (a) confusion matrix of the three different classes, (b) performance of the architecture on the three different classes. The architecture misclassified 2 benign mammograms as malignant.

total number of input samples.

3.5.2. Precision

Precision is also known as positive predictive value and is the quality of a positive prediction made by the model. It is calculated by dividing the number of true positives (TP) by the total number of positive predictions.

3.5.3. Recall

Recall, also known as sensitivity or true positive rate is the fraction of positives that are correctly classified by the model.

3.5.4. F1 Score

The F1-score combines the precision and recall of a classifier by taking their harmonic mean.

4. Results and Discussion

A number of models are trained and assessed on 140 mammograms (test dataset) in the presented work using the metrics listed in section 5. Several state-of-the-art architectures, viz. VGG19 Simonyan and Zisserman (2014) (Visual Geometry Group), DenseNet Huang et al. (2017) , ResNets He et al. (2016) (Residual Networks), Efficient-Nets Tan and Le (2019) , and InceptionNet Szegedy et al. (2016) were trained and fine-tuned on the MIAS dataset. Table 1 shows that the EEF-Net fared better than alternative architectures, with an accuracy of 97.14%. The CNN-LSTM architecture likewise performed well, with an overall accuracy of 96.43%. On the other hand, architectures containing

residual skip connections (ResNets) fared poorly because of a large number of trainable parameters, causing the model to overfit on the train data. Architectures such as VGG19, InceptionV3, and DenseNet169 illustrated satisfactory performance.

The testing set includes 140 mammograms, 56 of which are benign, 36 of which are malignant, and 48 of which are healthy. 138 of the 140 mammograms are correctly classified by EEF-Net. The confusion matrix is depicted in Figure 7.a. The architecture misidentified two mammograms (having benign tumor) as having a malignant tumor. It accurately classified all the healthy and malignant mammograms. Figure 7.b. represents the class-wise accuracy, recall, precision, and F1 score.

Table 2. shows that our performance is better than other proposed architectures. EEF-Net achieved an accuracy, sensitivity, specificity, and F1 score of 97.14%, 98.67%, 99.30%, and 98.30%, respectively. Our architecture performs better than other architectures due to the extensive pre-processing pipeline, which eliminates noise, different low and high-intensity labels, and pectoral muscles, which can hinder the models' performance.

5. Conclusion

Breast cancer frequently begins too small to be felt. As it grows, it has the potential to spread across the breast or to other regions of the body. Breast cancers can be diagnosed early when they are small and have not disseminated to lymph nodes, allowing for breast conservation and a reduction in mortality from breast cancer. Thus, breast screening

Model	No of param	Accuracy	Class	Precision	Recall	F1-Score
VGG19	33,038,147	92.86	Benign	91	89	90
			Malignant	82	86	84
			Healthy	100	98	99
EfficientNet	49,487,410	96.43	Benign	93	100	97
			Malignant	100	89	94
			Healthy	100	100	100
InceptionV3	48,185,891	90.71	Benign	88	89	88
			Malignant	83	81	82
			Healthy	100	100	100
Resnet50	75,136,643	79.29	Benign	69	88	77
			Malignant	67	39	49
			Healthy	100	100	100
Resnet101	94,207,107	72.86	Benign	60	93	73
			Malignant	20	3	5
			Healthy	100	100	100
Resnet152	109,919,875	76.43	Benign	64	73	68
			Malignant	48	39	43
			Healthy	100	98	99
DenseNet169	54,558,019	96.43	Benign	95	96	96
			Malignant	94	92	93
			Healthy	100	100	100
CNN-LSTM	40,392,195	96.43	Benign	95	96	96
			Malignant	94	92	93
			Healthy	100	100	100
EEF-Net	36,330,918	97.14	Benign	100	96	98
			Malignant	95	100	97
			Healthy	100	100	100

Table 1

Class-wise performance of different models on the MIAS dataset. With an accuracy of 97.14%, the EEF-Net outperformed other architectures.

Method	Accuracy	Sensitivity	Specificity	F1-score
Charan et al. (2018)	65	-	-	-
Papadopoulos et al. (2005)	73	-	-	-
Hepsağ et al. (2017)	85	-	-	-
Ting et al. (2019)	90.5	89.47	90.71	-
Abbas (2016)	91.5	92	84.2	-
Shan (2021)	92	96	93	-
Malebary and Hashmi (2021)	95	97	97	98
Kaur et al. (2019)	96.9	-	-	-
EEF-Net	97.14	98.67	99.3	98.3

Table 2

Comparison of the EEF-Net with other proposed architectures.

plays a significant role in breast cancer prognosis. However, radiologists manually examine mammograms to determine and estimate the severity of a malignancy, which involves specialized knowledge and often a significant amount of time. Computer-aided diagnostic systems powered by deep learning algorithms assist radiologists in detecting tumors and provide a high-precision solution to this problem. The presented study proposes the EEF-Net, a novel and optimum solution that accurately predicts the severity of tumors. The presented architecture has produced state-of-the-art results and can be used by radiologists to mass screen patients, enhancing efficacy and minimizing radiologists' workload.

Authors' contributions:

Conceptualization was done by Ninad Mehendale and Dishant Padalia. All the experiments/code executions were performed by Dishant Padalia and Kush Vora. The formal analysis was performed by Dishant Padalia and Darshil Mehta. Manuscript writing- original draft preparation was done by Dishant Padalia and Kush Vora. Review and editing were done by Ninad Mehendale, Dishant Padalia, and Kush Vora. Visualization work was carried out by Darshil Mehta.

Acknowledgments:

The authors would like to thank the colleagues.

Funding information:

No funding was involved in the present work.

Consent for publication:

Authors have taken all the necessary consents for publication from participants wherever required.

Compliance with ethical standards

Conflict of interest:

The authors declare that they have no conflict of interest.

Consent to participate:

This article does not contain any studies with animals or humans performed by any of the authors. Informed consent was not required as there were no human participants. All the necessary permissions were obtained from Institute Ethical committee and concerned authorities.

Ethics approval:

All authors consciously assure that the manuscript fulfills the following statements:

1. This material is the authors' own original work, which has not been previously published elsewhere.
2. The paper is not currently being considered for publication elsewhere.
3. The paper reflects the authors' own research and analysis in a truthful and complete manner.
4. The paper properly credits the meaningful contributions of co-authors and co-researchers.

5. The results are appropriately placed in the context of prior and existing research.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. **TensorFlow**: a system for Large-Scale machine learning, in: 12th USENIX symposium on operating systems design and implementation (OSDI 16), pp. 265–283.
- Abbas, Q., 2016. DeepCAD: A computer-aided diagnosis system for mammographic masses using deep invariant features. *Computers* 5, 28.
- Agarap, A.F., 2018. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375 .
- cancer.net, . Breast Cancer Statistics. URL: <https://www.cancer.net/cancer-types/breast-cancer/statistics>.
- Charan, S., Khan, M.J., Khurshid, K., 2018. Breast cancer detection in mammograms using convolutional neural network, in: 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1–5.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255.
- Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* 28, 100–108.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
- Hepsağ, P.U., Öznel, S.A., Yaz, A., 2017. Using deep learning for mammography classification, in: 2017 International Conference on Computer Science and Engineering (UBMK), pp. 418–423.
- Huang, G., Liu, Z., Maaten, L.V.D., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
- III, J.C.B., 1976. Mammography: a contrary view. *Annals of Internal Medicine* 84, 77–84.
- Kaur, P., Singh, G., Kaur, P., 2019. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. *Informatics in Medicine Unlocked* 16, 100151.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .
- Malebary, S.J., Hashmi, A., 2021. Automated breast mass classification system using deep learning and ensemble learning in digital mammogram. *IEEE Access* 9, 55312–55328.
- mayoclinic.org, . Cancer - Symptoms and causes. URL: <https://www.mayoclinic.org/diseases-conditions/cancer/symptoms-causes/syc-20370588>.
- Mustra, M., Grgic, M., 2013. Robust automatic breast and pectoral muscle segmentation from scanned mammograms. *Signal processing* 93, 2817–2827.
- O'Shea, K., Nash, R., 2015. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458 .
- Papadopoulos, A., Fotiadis, D.I., Likas, A., 2005. Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artificial intelligence in medicine* 34, 141–150.
- Ramachandran, P., Zoph, B., Le, Q.V., 2017. Swish: a self-gated activation function. arXiv preprint arXiv:1710.05941 7, 5.
- Saber, A., Sakr, M., Abo-Seida, O.M., Keshk, A., Chen, H., 2021. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. *IEEE Access* 9, 71194–71209.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520.
- Shan, W., 2021. Double adaptive weights for stabilization of moth flame

- optimizer: balance analysis, engineering cases, and medical diagnosis. Knowledge-Based Systems 214, 106728.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .
- Singh, V.P., Srivastava, S., Srivastava, R., 2018. Automated and effective content-based image retrieval for digital mammography. Journal of X-ray Science and Technology 26, 29–49.
- Suzuki, K., 2017. Overview of deep learning in medical imaging. Radiological physics and technology 10, 257–273.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.
- Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, pp. 6105–6114.
- Ting, F.F., Tan, Y.J., Sim, K.S., 2019. Convolutional neural network improvement for breast cancer classification. Expert Systems with Applications 120, 103–115.
- who.int, . Breast Cancer. URL: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- Zebari, D.A., Zeebaree, D.Q., Abdulazeez, A.M., Haron, H., Hamed, H.N.A., 2020. Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. Ieee Access 8, 203097–203116.
- Zuiderveld, K., 1994. Contrast limited adaptive histogram equalization. Graphics gems , 474–485.