International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Breast Cancer Detection Using ML

Mr. Manish Kumar[1], Dr. Shallu Bashambu[2]

[1]Student, Maharaja Agrasar Institute of Technology
[2]Professor, Maharaja Agrasar Institute of Technology

ABSTRACT

The prompt and accurate diagnosis of breast lesions, including the distinction between cancer, non-cancer, and suspicious cancer, plays a crucial role in the prognosis of breast cancer. In this paper, we introduce a novel method based on feature extraction and reduction for the detection of breast cancer in mammography images. First, we extract features from multiple pre-trained convolutional neural network (CNN) models, and then concatenate them. The most informative features are selected based on their mutual information with the target variable. Subsequently, the selected features can be classified using a machine learning algorithm. We evaluate our approach using four different machine learning algorithms: neural network (NN), k-nearest Neighbor (kNN), random forest (RF), and support vector machine (SVM). Our results demonstrate that the NN-based classifier achieves an impressive accuracy of 92% on the RSNA dataset. This dataset is newly introduced and includes two views as well as additional features like age, which contributed to the improved performance. We compare our proposed algorithm with state-of-the-art methods and demonstrate its superiority, particularly in terms of accuracy and sensitivity. For the MIAS dataset, we achieve an accuracy as high as 94.5%, and for the DDSM dataset, an accuracy of 96% is attained. These results highlight the effectiveness of our method in accurately diagnosing breast lesions and surpassing existing approaches.

Keywords: breast cancer; convolutional neural network (CNN); computer aided diagnosis (CAD); feature selection; feature classification; mammography images

INTRODUCTION

Breast cancer is major cause of death in women around the world. According to WHO (World Health Organisation), breast cancer accounted for maximum deaths (2.26 million cases), worldwide in 2020 out of the 10 million cases of cancer. Breast cancer starts when cells in the breast begin to grow out of control. These accumulations of cells are called tumours and they can often be seen on an x-ray or felt as a lump. Breast cancer can spread when the cancer cells get into the blood or lymph system and are carried to other parts of the body making them prone to cancer. There are many different types of breast cancer and common ones include ductal carcinoma in situ (DCIS) and invasive carcinoma. The side effects of Breast Cancer are – Fatigue, Headaches, Pain and numbness (peripheral neuropathy), Bone loss and osteoporosis. There are two types of tumours. One is benign which is non-cancerous and the other one is malignant which is cancerous. Benign breast tumours are abnormal growths in the breast, but they do not spread outside. So, this means that they are not life threatening, but some types of benign tumours can increase a woman's risk of getting breast cancer. Different imaging tests are used for detecting breast cancer. Some of them are mammograms, breast ultrasound and breast MRI. A **mammogram** is nothing but an x-ray of

breast and it is used to look for any changes in the breast. A mammogram makes it easy to treat by finding and detecting breast cancer early, when the tumor is small and even before a lump can be felt.

Detection of breast cancer in its early stages using image processing techniques includes four parts. In the first part the digital images (mammograms) are pre-processed to remove any kind noise. Then in the second part the images undergo the segmentation process to enhance the tumor part. After this, in the third part, the important features in the segmented images are extracted. Finally, in the fourth part, with the help of the extracted features, the images are classified into normal, benign or malignant. Here, '**normal**' represents the breast with no tumor, '**benign**' represents the breast with non-cancerous tumor and '**malignant**' represents breast with cancerous tumor.

1.1 PROJECT OBJECTIVE

The objective of the project is todetect the initial phase tumors which shall not be prone to human error using image processing techniques such as image preprocessing, image segmentation, features extraction and selection and image classification.

Firstly the image pre-processing of the mammogram is carried out which helps in removing noise in the image, if any. Second the segmentation techniques were used with which the tumor part dilates in the breast and erodes the remaining parts. Along with the above two image processing techniques, feature extraction is also done using MATLAB. Finally the features extracted are used for classification of mammograms into normal, benign and malignant. The image classification process is done with python using about 200(approx.) images.

1.2PROJECT OUTLINE

This project report is presented over the five remaining chapters. Chapter 2 describes the causes of breast cancer. Chapter 3 presents the methodology which is used in the detection of breast cancer using digital image processing techniques. Chapter 4 explains the concepts of MATLAB and Python which were used in the project. Chapter 5 presents the simulation results of the detection of breast cancer using MATLAB and Python using various IMAGES. Finally, conclusions are drawn in chapter 6.

2. DETAILS ON BREAST CANCER

Breast cancer is a type of cancer that starts in the breast. Cancer starts when cells begin to grow out of control.Breast cancer cells usually form a tumour that can often be seen on an x-ray or felt as a lump. Breast cancer occurs almost entirely in women, but men can get breast cancer, too.

It's important to understand that most breast lumps are benign and not cancer (malignant). Non-cancerous breast tumours are abnormal growths, but they do notspread outside of the breast. They are not life threatening, but some types of benignbreast lumps can increase a woman's risk of getting breast cancer. Any breast lump orchange needs to be checked by a health care professional to determine if it is benign ormalignant (cancer) and if it might affect your future cancer risk.

2.1 WHERE BREAST CANCER STARTS

Breast cancers can start from different parts of the breast.

- Most breast cancers begin in the ducts that carry milk to the nipple (ductal cancers)
- Some start in the glands that make breast milk (lobular cancers)
- There are also other types of breast cancer that are less common like phyllodes tumour and angiosarcoma
- A small number of cancers start in other tissues in the breast. These cancers are called sarcomas and lymphomas and are not really thought of as breast cancers.

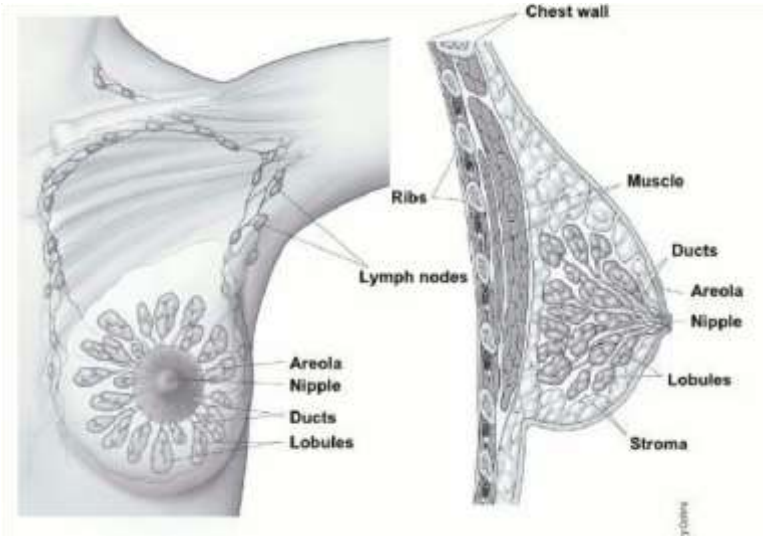


FIGURE 2.1 BREAST

Although many types of breast cancer can cause a lump in the breast, not all do.Many breast cancers are also found on screening mammograms, which can detect cancers at an earlier stage, often before they can be felt, and before symptoms develop.

2.2 TYPES OF BREAST CANCER

There are many different types of breast cancer and common ones include ductal carcinoma in situ (DCIS) and invasive carcinoma. Others, like phyllodes tumours and angiosarcoma are less common.

Once a biopsy is done, breast cancer cells are tested for proteins called estrogen receptors, progesterone receptors and HER2. The tumour cells are also closely looked at in the lab to find out what grade it is. The specific proteins found and the tumour grade can help decide treatment options.

2.3 HOW BREAST CANCER SPREADS

Breast cancer can spread when the cancer cells get into the blood or lymph system and are carried to other parts of the body.

The lymph system is a network of lymph (or lymphatic) vessels found throughout the body that connects lymph nodes (small bean-shaped collections of immune system cells). The clear fluid inside the lymph vessels, called lymph, contains tissue byproducts and waste material, as well as immune system cells. The lymph vessels carry lymph fluid away from the breast. In the case of breast cancer, cancer cells can enter those lymph vessels and start to grow in lymph nodes.

Most of the lymph vessels of the breast drain into:

- Lymph nodes under the arm (auxiliary nodes)
- Lymph nodes around the collar bone (supraclavicular [above the collar bone] and infraclavicular [below the collar bone] lymph nodes)
- Lymph nodes inside the chest near the breast bone (internal mammary lymph nodes).

If cancer cells have spread to your lymph nodes, there is a higher chance that the cells could have travelled through the lymph system and spread (metastasized) to other parts of your body. The more lymph nodes with breast cancer cells, the more likely it is that the cancer may be found in other organs. Because of this, finding cancer in one or more lymph nodes often affects your treatment plan. Usually, you will need surgery to remove one or more lymph nodes to know whether the cancer has spread.

Still, not all women with cancer cells in their lymph nodes develop metastases, and some women with no cancer cells in their lymph nodes develop metastases later.

2.4 HOW COMMON IS BREAST CANCER

Breast cancer is the most common cancer in American women, except for skin cancers. The average risk of a woman in the United States developing breast cancer sometime in her life is about 13%. This means there is a 1 in 8 chance she will develop breast cancer. This also means there is a 7 in 8 chance she will never have the disease.

2.5 BREAST CANCER SIGNS AND SYMPTOMS

- Swelling of all or part of a breast (even if no lump is felt)
- Skin dimpling (sometimes looking like an orange peel)
- Breast or nipple pain
- Nipple retraction (turning inward)
- Nipple or breast skin that is red, dry, flaking or thickened
- Nipple discharge (other than breast milk)
- Swollen lymph

2.6 MAMMOGRAMS

Mammograms are low-dose x-rays of the breast. Regular mammograms can help find breast cancer at an early stage, when treatment is most successful. A mammogram can often find breast changes that could be cancer years before physical symptoms develop. Results from many decades of research clearly show that women who have regular mammograms are more likely to have breast cancer found early, are less likely to need aggressive treatment like surgery to remove the breast (mastectomy) and chemotherapy, and are more likely to be cured.

Mammograms are not perfect. They miss some cancers. And sometimes a woman will need more tests to find out if something found on a mammogram is or is not cancer. There's also a small possibility of being diagnosed with a cancer that never would have caused any problems had
it not been found during screening. (This is called *overdiagnosis*.)

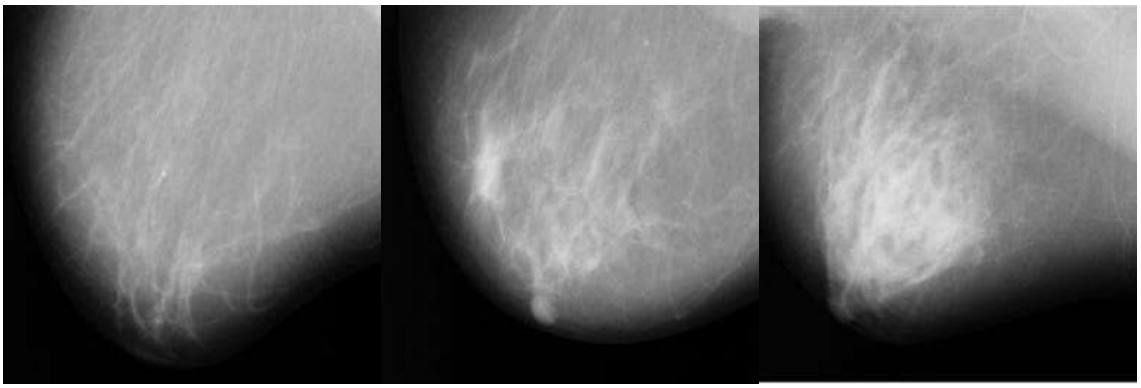


FIGURE 2.2 MAMMOGRAMS

There are two types of mammograms. A screening mammogram is used to look for signs of breast cancer in women who don't have any breast symptoms or problems. X-ray pictures of each breast are taken, typically from 2 different angles.Mammograms can also be used to look at a woman's breast if she has breast symptoms or if a change is seen on a screening mammogram. When used in this way, they are called diagnostic mammograms. They may include extra views (images) of the breast that aren't part of screening mammograms. Sometimes diagnostic mammograms are used to screen women who were treated for breast cancer in the past.

In the past, mammograms were typically printed on large sheets of film. Today, digital mammograms are much more common. Digital images are recorded and saved as files in a computer.

Literature Review:

Deep Learning in Breast Cancer Detection

Deep learning methods have revolutionized medical imaging, particularly in breast cancer detection, by automatically learning hierarchical feature representations from images. CNNs, in particular, have emerged as powerful tools for classifying mammographic images into malignant and benign categories. Several studies have explored different CNN architectures and feature extraction techniques to enhance classification accuracy.

CNN Architectures for Mammography Analysis

Multiple CNN architectures have been employed for breast cancer detection, each leveraging different methodologies for feature extraction and classification:

1. **AlexNet**: Introduced in 2012, AlexNet was among the first deep CNN architectures used for image classification. It employs multiple convolutional layers with rectified linear unit (ReLU) activation functions to extract features from images. Despite its success in general image recognition tasks, AlexNet has shown moderate performance in mammography classification due to its relatively shallow depth.
2. **ResNet50**: ResNet50 utilizes residual connections to mitigate the vanishing gradient problem in deep networks. This architecture has demonstrated improved performance in medical image classification, including mammography, by allowing the training of deeper networks.
3. **EfficientNet**: A more recent development, EfficientNet uses a compound scaling approach to optimize depth, width, and resolution, achieving state-of-the-art performance in medical image analysis with lower computational costs.
4. **MobileNet**: Designed for resource-constrained environments, MobileNet employs depthwise separable convolutions to reduce computational complexity while maintaining high classification accuracy. It is particularly suitable for mobile and embedded AI applications in healthcare.
5. **ConvNeXt**: This architecture enhances traditional CNN designs by incorporating elements inspired by transformer networks. It captures diverse features through parallel branches, improving robustness in image classification tasks.

Machine Learning Classifiers for Mammography Analysis

While CNNs are widely used for feature extraction, different machine learning classifiers have been explored for final decision-making:

- **Neural Networks (NNs)**: When combined with CNN-extracted features, NNs have demonstrated superior classification performance, achieving up to 96% accuracy on benchmark datasets.
- **k-Nearest Neighbors (k-NN)**: While computationally efficient, k-NN classifiers generally exhibit lower accuracy compared to NNs.
- **Random Forest (RF)**: RF classifiers have been used to aggregate CNN-extracted features, offering competitive accuracy with robust performance.

- **Support Vector Machines (SVMs)**: Despite being effective in traditional medical image classification, SVMs tend to underperform compared to deep learning-based classifiers in complex mammographic analysis.

Performance Comparison with State-of-the-Art Methods

Recent studies have benchmarked CNN-based methods against traditional machine learning approaches. The best CNN-based models have outperformed classical techniques such as SVM and k-NN, particularly when using feature selection and ensemble learning strategies. Performance metrics such as accuracy, sensitivity, precision, and F1-score indicate that deep learning-based approaches provide more reliable breast cancer detection outcomes.

Cross-Dataset Validation

One major challenge in CNN-based breast cancer detection is the generalization ability of models across different datasets. Studies have shown that models trained on one dataset (e.g., RSNA) may exhibit reduced performance when tested on another dataset (e.g., MIAS or DDSM). This highlights the need for diverse training data and domain adaptation techniques to enhance model robustness.

METHODOLOGY

Datasets

A. The main dataset for this project is the radiological society of north america (RSNA) dataset from a recent Kaggle competition [22]. The dataset contains 54,713 images in dicom format from roughly 11,000 patients. For each patient, there are at least four images from different laterality and views. For each subject, two different views CC and MLO, and images from left and right laterality were provided. The images are of various sizes and formats, including jpeg and jpeg2000, and different types, such as monochrome-1 and monochrome-2. The dataset provides additional features some of which can be used for classification purposes: age, implant, BIRADS, and density. We base our work on this dataset, but since this dataset is new, it has not been used in any published research yet. Hence, for comparison purposes, we use two other well-known datasets MIAS and DDSM. This dataset is imbalanced as only 2 percent of the images are from cancer patients, which makes any classification method biased. To compensate for this, we use all positive cases and only 2320 images from negative cases. Figure 1 depicts two sample images from this dataset for cancer and normal cases. of the images are from cancer patients, which makes any classification method biased. To compensate for this, we use all positive cases and only 2320 images from negative cases. Figure 1 depicts two sample images from this dataset for cancer and normal cases.

B. The mammographic image analysis society (MIAS) [23] dataset is a well-known and widely used dataset for the development and evaluation of CAD systems for BC detection. It consists of 322 mammographic images, with each image accompanied by a corresponding ground truth classification of benign or malignant tumors. The dataset is particularly valuable for researchers interested in developing machine learning algorithms for BC detection, as it includes examples of both normal and abnormal mammograms, as well as a range of breast densities and lesion types. Figure 2 depicts two sample images from this dataset for cancer and normal cases.

C. The digital database for screening mammography (DDSM) [24] includes 55,890 images, of which 14% are positive, and the remaining 86% are negative. Images were tiled into 598 × 598 tiles, which

were then resized to 299 × 299. A subset of this dataset which is for positive cases and is called CBIS-DDSM, has been annotated and the region of interest has been extracted by experts. In this research, we do not use the CBIS-DDSM and use the original DDSM dataset as we are classifying the images from normal subjects and cancer patients. Figure 3 depicts two sample images from this dataset.

Table 1. This table shows the description of three datasets.

Dataset	Number of Images	Image Types	Image size
RSNA	**54,713**	**Variable**	**Variable**
MIAS	**322**	**PGM**	**1024*1024**
DDSM	**55,890**	**JPEG**	**598*598**

Models

A. AlexNet [25] is a deep CNN architecture that was introduced in 2012 and achieved a breakthrough in computer vision tasks such as image classification. It consists of eight layers, including five convolutional layers and three fully connected layers. The first convolutional layer uses a large receptive field to capture low-level features such as edges and textures, while subsequent layers use smaller receptive fields to capture increasingly complex and abstract features. AlexNet was the first deep network to successfully use the rectified linear unit (ReLU) activation functions, which have since become a standard activation function in deep learning. It also used dropout regularization to prevent overfitting during training. AlexNet's success on the ImageNet dataset, which contains over one million images, demonstrated the potential of deep neural networks for image recognition tasks and paved the way for further advances in the field of computer vision.

B. ResNet50 [26] is a deep CNN architecture that uses residual connections to enable learning from very deep architectures without suffering from the vanishing gradient problem. It consists of 50 layers, including convolutional layers, batch normalization layers, ReLU activation functions, and fully connected layers. ResNet50 also uses a skip connection that bypasses several layers in the network, allowing it to effectively learns both low-level and high-level features.

C. EfficientNet [27] is a family of deep CNN architectures that were introduced in 2019 and have achieved state-of-the-art performance on a range of computer vision tasks. EfficientNet uses a compound scaling method to simultaneously optimize the depth, width, and resolution of the network, allowing it to achieve high accuracy while maintaining computational efficiency. EfficientNet consists of a backbone network that extracts features from input images and a head network that performs the final classification. The backbone network uses a combination of mobile inverted bottleneck convolutional layers and squeeze-and-excitation (SE) blocks to capture both spatial and channel-wise correlations in the input. The head network uses a combination of global average pooling and fully connected layers to perform the final classification.

D. MobileNet [28] is a deep learning architecture suitable for efficient and accurate analysis of medical images, specifically in the context of BC diagnosis. With its emphasis on computational efficiency, MobileNet can effectively extract features from mammography images, enabling the detection of subtle patterns or abnormalities associated with breast cancer. By utilizing depthwise separable convolutions, MobileNet optimizes memory consumption and computational load, making it ideal for resource-constrained environments. The integration of the ReLU6 activation function further enhances efficiency and compatibility with medical imaging devices. Overall, MobileNet offers a valuable

solution for BC analysis, providing accurate results while operating efficiently on limited computational resources.

E. ConvNeXt [29] is an architecture that enhances the representational capacity of CNNs by leveraging parallel branches to capture diverse and complementary features, leading to improved performance on challenging visual recognition tasks. It has demonstrated excellent performance on various computer vision tasks, including image classification, object detection, and semantic segmentation. Its ability to capture complex relationships between features has made it a popular choice for tasks requiring a high-level understanding of visual data.

In this paper, we propose a method based on the extraction and concatenation of features obtained from various CNN models. The extracted features are then reduced such that only good features are selected and then used for the classification of normal and cancerous images. **Figure 4** illustrates the block diagram of the proposed system. As one can see, the images from different datasets are first preprocessed, and then features are extracted through different CNN models. The extracted features are reduced and then classified into two: cancer and no cancer. The details for each block are as follows:

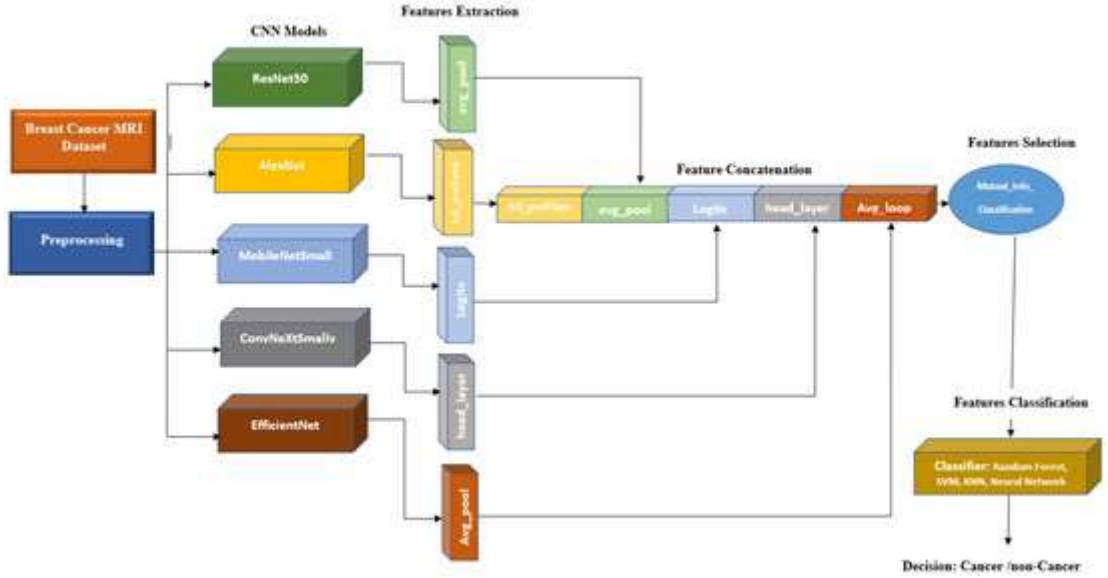


Figure 4. Block diagram of the proposed system.

Preprocessing: In this research, the images obtained from various datasets exhibit variations in sizes and resolutions.

Normalization:

The RSNA dataset consists of images in various formats, including 12 and 16 bits per pixel. Additionally, it has two different photometric interpretations known as MONOCHROME1 and MONOCHROME2. The former represents grayscale images with ascending pixel values from bright to dark, while the latter represents grayscale images with ascending pixel values from dark to bright. To ensure consistency within the RSNA dataset, we convert all MONOCHROME1 images to MONOCHROME2.

In order to standardize the pixel values across the RSNA dataset, intensity normalization is performed. This involves scaling the pixel values to the range of 0 to 255, which is equivalent to 8 bits per pixel. By applying this normalization process, the pixel values across the dataset become more consistent and comparable.

On the other hand, the DDSM and MIAS datasets already have pixel values within the range of 0 to 255, eliminating the need for additional normalization. Therefore, the pixel values in these datasets are deemed suitable, and no further adjustment is required.

F. 2.

Region of Interest Selection:

To select the region of interest, we initially apply a global thresholding method to the image. Subsequently, we extract the contour of the largest object present in the image, which corresponds to the breast area. Utilizing this contour, we generated a mask that enables us to crop the image and isolate the specific region of interest for further analysis.

3.

Image Alignment:

In breast cancer datasets, there are two distinct laterality categories: left and right. To enhance consistency and improve accuracy in analysis, we align all laterality labels to the left side. This process involves horizontally flipping all left breast images to create a uniform orientation throughout the datasets. By standardizing the laterality representation, we ensure a consistent and reliable dataset for further research and analysis purposes.

G. B.

Feature extraction: For feature extraction, we exploit the features computed by pre-trained CNN models described in **Section 2.2**. For each model, the features are extracted from the last layer before the last fully connected (FC) layer as the output of the final FC layer has been trained for 1000 classes of the ImageNet dataset, and hence, we skip this layer and extract the features from the last layer before the final FC layer. **Table 2** depicts the layer before the final FC layer and the number of features extracted for each CNN model used in this paper.

Table 2. This table shows the CNN models used in the proposed method along with the layer name where the features have been extracted and the number of features extracted from each model.

CNN Models	Layer Name	Number of Features
ResNet50	avg_pool	2048
AlexNet	fc8_preflatten	4096
MobileNetSmall	Logits	**1000**
EfficientNet	avg_pool	**1280**

Feature concatenation: The 1-dimensional (1D) features extracted in the previous step are concatenated to form a single 1D feature vector. Note that for each CNN model, we have extracted features from two different views CC and MLO. Hence, 10 1D vectors are concatenated here. This forms a vector with a size of 18,384 For the RSNA dataset that we use as the basis of our research, we have an additional useful feature for the patient age. **Figure 5** depicts the distribution of the age feature provided by the RSNA dataset for both cancer and non-cancer subjects. As can be observed, age can also be considered a valuable feature. We can also simply normalize and add age to our feature vector to have 18,385 features in total.

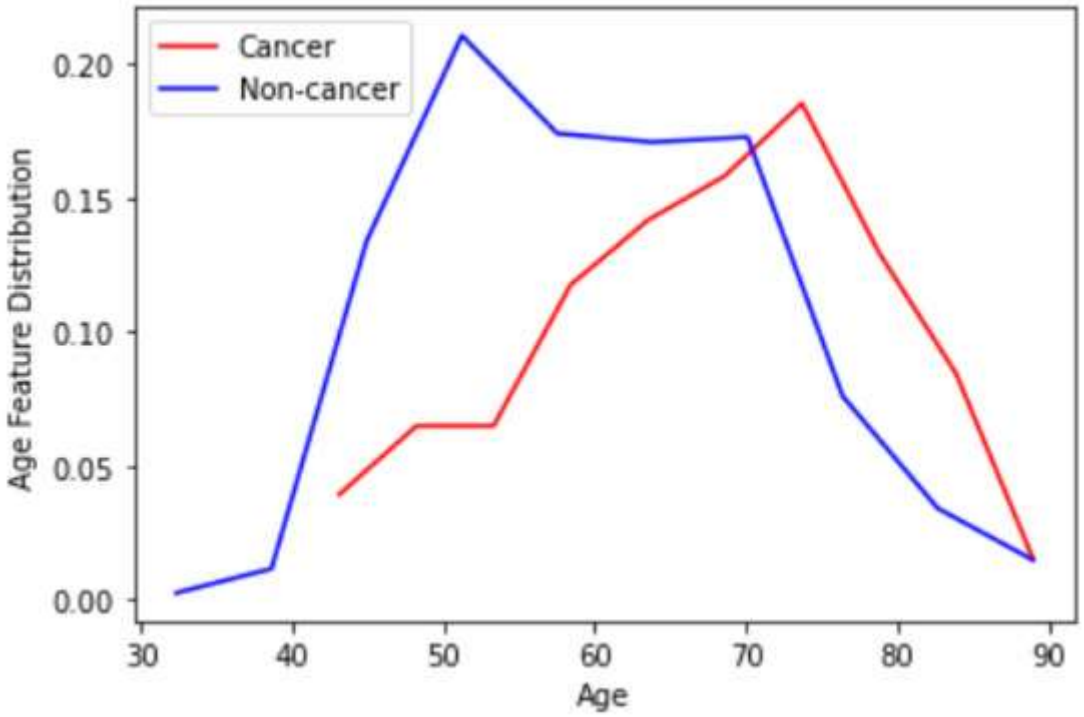


Figure 5. This figure shows the distribution of age for cancer and noncancer subjects in the RSNA dataset. Feature selection: The majority of the features are redundant and do not carry any useful information and only increase the complexity of the system. **Figure 6** illustrates 2 samples of good and weak features. As one can see from the figure, in the case of weak features, the distribution of the feature for normal and cancerous subjects are similar showing that there is no useful information in this feature and the calculated mutual information between them is zero. For the case of good features, normal and cancerous subjects have obviously different distributions showing that these features carry useful information, although small, that can improve the performance of classifiers used in the next step. To compute mutual information we use the method in [**30**]. We empirically found a 0.02 threshold gives us the best results. Note that we have also adopted feature selection based on mutual information empirically and after using various feature selection methods. The number of features for each dataset before and after feature selection is presented in **Table 3**.

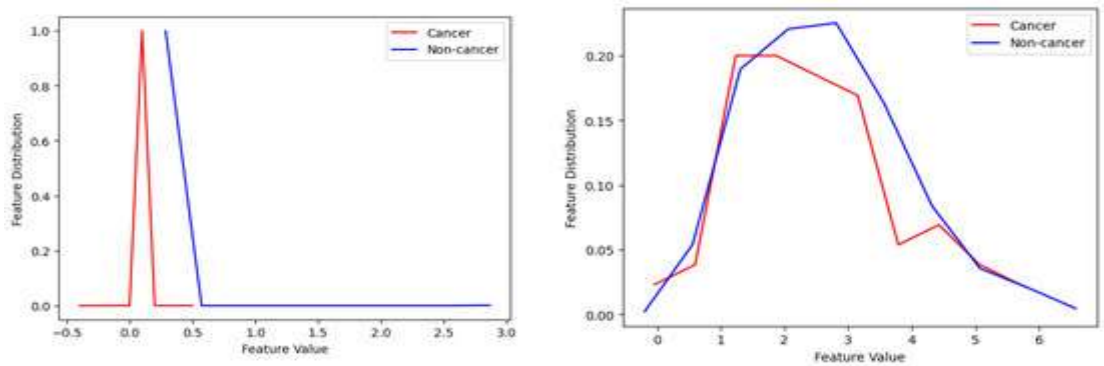


Figure 6. These figures show distributions of (a) a good feature and (b) a weak feature extracted using a pre-trained CNN model. for cancer and noncancer subjects in the DDSM dataset. The mutual information computed for these two features is 0.035 and zero, respectively.

Table 3. The total number of features obtained from each dataset before and after feature selection.

Dataset	Before Feature Selection	After Feature Selection
RSNA	18,385	452
MISA	9192	212

Feature classification: After selecting the best features, we need to classify them. For this purpose, we tried multiple machine learning algorithms such as k-NN, random forest (RF), SVM, and NN. In our study, we utilize an RF algorithm with specific parameters to enhance breast cancer detection. We construct an ensemble of 100 trees, setting the minimum number of samples required to split a node as 2. Additionally, we limit the maximum number of features considered for each tree to 5 and the maximum tree depth to 4. These parameter settings are chosen to optimize the performance of our model and improve the accuracy of breast cancer detection in our X-ray image datasets.

In our SVM classifier implementation, we utilize a linear kernel and set the regularization parameter "C" to a value of 1. The linear kernel allows us to learn a linear decision boundary, while the "C" parameter balances the trade-off between training accuracy and the complexity of the decision boundary.

In the k-NN classifier, we set k = 5, and for the NN classifier, we used two fully connected (FC) layers with a hidden layer including 96 neurons and a single-neuron classification layer. For the classification layer, we use a sigmoid activation function that classifies non-cancer cases from cancerous ones.

Results and Discussion

This section showcases the results obtained from the three datasets introduced in **Section 2.1** using the models described in **Section 2.2**, as well as a combination of all datasets as illustrated in **Figure 4**. For each dataset, we employed k-fold cross-validation with k = 10. This means that the method was trained and tested 10 times, with 90% of the data allocated for training and 10% for testing in each iteration.

4.1. Evaluation Metrics [31]

To assess the performance of our experiments, we utilize various evaluation metrics.

- True positives (TP): Instances where the predicted class and actual class are both positive. This indicates that the classifier accurately classified the instance with a positive label.
- False positives (FP): Instances where the predicted class is positive but the actual class is negative. This means that the classifier incorrectly classified the instance with a positive label. In the context of breast abnormality classification, an FP response corresponds to a type I error according to statisticians. For example, it could refer to a calcification image being classified as a mass lesion or a benign mass lesion being classified as a malignant mammogram in the diagnosis.
- True negatives (TN): Instances where the predicted class and actual class are both negative. This indicates that the classifier correctly classified the instance with a negative label.
- False negatives (FN): Instances where the predicted class is negative but the actual class is positive. This means that the classifier incorrectly classified the instance with a negative label. In the context of breast abnormality classification, an FN response is considered a type II error. For instance, it could refer to a mass mammogram being classified as calcification or a malignant mass lesion being

classified as a benign mammogram in the diagnosis. Type II errors are particularly significant in their consequences.

- Accuracy: This metric represents the overall number of correctly classified instances. In the case of the abnormality classifier, accuracy signifies the correct classification of image patches containing either mass or calcification. Similarly, accuracy shows the correct classification of image patches as either malignant or benign in the pathology classifier.

Acc=(TP+TN)(TP+TN+FP+FN)Acc=TP+TNTP+TN+FP+FN

- Sensitivity or Recall: This metric represents the proportion of positive image patches that are correctly classified. In the abnormality type classifier, sensitivity indicates the fraction of image patches that are truly mass lesions and are correctly classified. Similarly, the abnormality pathology classifier shows the fraction of truly malignant image patches that are correctly classified. Given the significance of type II errors, this metric is valuable for evaluating performance.

Sn=(TP+FN)Sn=TPTP+FN

- Precision: This metric reflects the proportion of positive predictions that are correctly categorized. It is calculated using the following formula:

Pr=(TP+FP)Pr=TPTP+FP

- F1 Score: This measure combines the impact of recall and precision using the harmonic mean, giving equal penalties to extreme values. It is commonly calculated using the formula:

$F-Score$=(2×Sn×xPr)(Sn+Pr)F−Score=2×Sn×xPrSn+Pr

4.2. Performance Evaluation of the Proposed Model for Different Classifiers

Table 4 presents a comparison of performance metrics for different CNN models using the RSNA dataset. Among the individual CNN models, EfficientNet consistently outperforms the other models in terms of accuracy, sensitivity, precision, AUC, and F-Score. Its superior performance can be attributed to its architecture, which enables it to capture relevant features and make accurate predictions on the RSNA dataset. EfficientNet proves to be the most effective choice among the individual models for accurately classifying medical images in the RSNA dataset. From the last row of the table, one can see that the proposed concatenation scheme, significantly improves all performance metrics, for instance, the achieved accuracy is 6 percent more than the best CNN model, i.e., EfficientNet.

Table 4. Performance comparison of the proposed method for different CNN models and Concat. Model with the NN classifier for RSNA dataset.

CNN Models	Acc	Sn	Pr	AUC	F-Score
AlexNet	81%	84%	87%	0.82	0.86
Resnet50	84%	90%	86%	0.89	0.88
MobileNetSmall	77%	85%	81%	0.81	0.83
ConvNexSmall	79%	87%	83%	0.83	0.85
EfficientNet	86%	92%	88%	0.92	0.90
Concat. Model	**92%**	**96%**	**92%**	**0.96**	**0.94**

Table 5 presents a summary of the results obtained using the kNN classifier with k = 5. The findings indicate a significant decline in performance compared to the NN model. Specifically, without feature concatenation, the highest accuracy is achieved with AlexNet, which is 8 percent lower than the accuracy of the same model with the NN classifier, and 13 percent lower than the best-performing EfficientNet model with the NN classifier. Additionally, the accuracy of the concatenated model is also 14 percent lower compared to the concatenated model with the NN classifier.

Table 5. Performance comparison of the proposed method for different CNN models and Concat. Model with the kNN classifier for RSNA dataset

CNN Models	Acc	Sn	Pr	AUC	F-Score
AlexNet	73%	70%	72%	0.70	0.71
Resnet50	72%	75%	71%	0.73	0.73
MobileNetSmall	64%	71%	67%	0.68	0.69
ConvNexSmall	66%	74%	70%	0.71	0.72
EfficientNet	71%	78%	74%	0.76	0.76
Concat. Model	**78%**	**81%**	**79%**	**0.82**	**0.80**

Table 6 displays the results obtained from the RF classifier. It demonstrates that the accuracy of the concatenated Model is equivalent to that of the KNN classifier, but falls short compared to the NN. Among the individual models, EfficientNet exhibits the most favorable performance metrics, while mobileNetSmall exhibits the least favorable performance.

Table 6. Performance comparison of the proposed method for different CNN models and Concat. Model with the RF classifier for RSNA dataset.

CNN Models	Acc	Sn	Pr	AUC	F-Score
AlexNet	71%	67%	69%	0.68	0.68
Resnet50	69%	70%	67%	0.71	0.68
MobileNetSmall	60%	67%	63%	0.64	0.65
ConvNexSmall	62%	69%	65%	0.67	0.67
EfficientNet	73%	74%	70%	0.75	0.72
Concat. Model	**78%**	**79%**	**77%**	**0.80**	**0.78**

Table 7 displays the results of the proposed method using the SVM classifier. It is evident from the table that SVM exhibits the lowest accuracy among all four investigated methods. Specifically, the accuracy of

the SVM-based method is 19 percent lower than that of the NN-based method. Furthermore, in comparison to the KNN and RF-based systems, the accuracy of the concatenated model decreased by 5 percent.

Table 7. Performance comparison of the proposed method for different CNN models and Concat. Model with the SVM classifier for RSNA dataset.

CNN Models	Acc	Sn	Pr	AUC	F-Score
AlexNet	62%	61%	63%	0.62	0.62
Resnet50	64%	66%	63%	0.65	0.64
MobileNetSmall	60%	63%	59%	0.60	0.61
ConvNexSmall	62%	65%	61%	0.63	0.63
EfficientNet	68%	70%	66%	0.68	0.68
Concat. Model	**73%**	**75%**	**72%**	**0.74**	**0.73**

4.3. Comparison of the Proposed System with State-of-the-Art Methods

Based on the findings presented in Tables 4–7, it is evident that the NN classifier achieves the highest level of performance. Therefore, we employed the suggested approach using the NN classifier as the benchmark to compare it with the existing methods. To the best of our knowledge, the RSNA dataset has not been utilized in any previously published papers. Consequently, for the purposes of this section, we conducted a comparison of our proposed model against existing methods using the MIAS and DDSM datasets and summarized the results in Table 8. Table 8. Performance comparison of our proposed model vs. methods using the MIAS and DDSM Method.

Method	Dataset	Number of Images	ACC	Sn	Pr
SVM & Hough [32]	MIAS & InBreast	322&206	86.13%	80.67%	**92.81%**
LQP & SVM [33]	MIAS	95	94%	NA	NA
GMM & SVM [34]	Mini-MIAS dataset	90	92.5%	NA	NA
KNN [35]	Mini-MIAS	120	92%	NA	NA
Voting Classifier [36]	MIAS	322	85%	NA	NA
CNN-4d [37]	Mini-MIAS	547	89.05%	90.63%	83.67%
CNN [38]	DDSM	10,480	93.5%	NA	NA
CNNs [39]	DDSM	11,218	85.82%	82.28%	86.59%
Our Method + NN	RSNA	54,713	92%	96%	92%
Our Method + NN	MIAS	322	**94.5%**	**96.32%**	91.80%
Our Method + NN	DDSM	55,890	**96%**	**94.70%**	**97%**

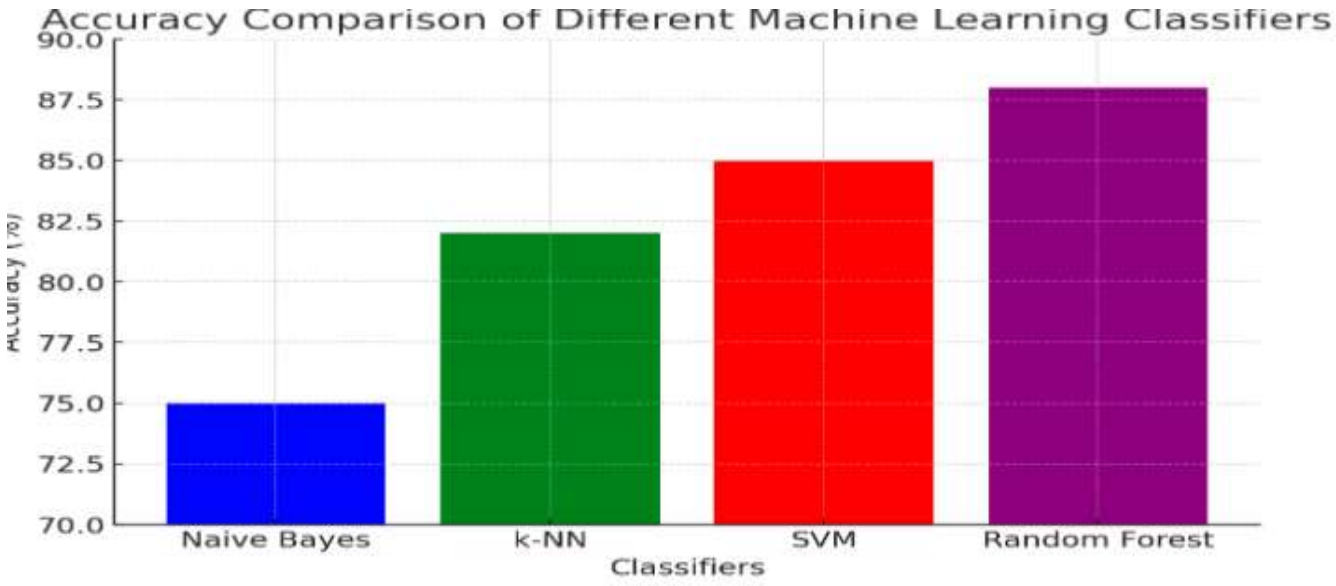
Upon examining Table 8, it is evident that our proposed model has exhibited superior performance compared to state-of-the-art algorithms in terms of accuracy and sensitivity across both the MIAS and DDSM datasets. While the method described in [32] demonstrated slightly better precision for the MIAS dataset, our algorithm outperformed it in the remaining two performance metrics. 4.4. Cross-Dataset Validation So far, we have trained and tested the proposed method on the same dataset. However, it is crucial to evaluate the ability of a model trained on one dataset to perform well on different datasets or images collected from diverse machines and under varying image collection standards. In this subsection,

we assess the performance of our method when trained on one of three datasets: RSNA, MIAS, and DDSM, and subsequently tested on images from a different dataset. The results of these experiments are summarized in Table 9.

Table 9. Performance of the proposed model with cross-dataset validation, i.e., trained and tested with different datasets.

Train Dataset	Test Dataset	ACC	Sn	Pr
RSNA	MIAS	79.13%	82.67%	80.81%
RSNA	DDSM	74%	77.50%	76%
MIAS	RSNA	76.5%	78.80%	78%
MIAS	DDSM	80.70%	82%	82.80%
DDSM	RSNA	72%	75.50%	76%
DDSM	MIAS	79%	80%	79.87%

Since the RSNA dataset comprises images of various types and resolutions, crossvalidating it with another dataset yields slightly lower performance metrics. Specifically, when the method is trained on either the MIAS or DDSM dataset and tested on RSNA images, the achieved performance is slightly reduced. Figure 1 visually depicts the resemblance between RSNA and MIAS images compared to RSNA and DDSM images, further supporting the observation that cross-validation between RSNA and MIAS datasets leads to higher accuracy compared to cross-validation involving RSNA and DDSM datasets. These findings are also supported by the results presented in Table 9



Conclusions

We have developed a novel method to address the accurate diagnosis of breast cancer in mammography images. Our approach involves the extraction and selection of features from multiple pre-trained CNN models, followed by classification using various machine learning algorithms: kNN, SVM, RF, and NN. The results obtained for different datasets demonstrate the effectiveness of our proposed scheme. Our findings indicate that the NN-based classifier yielded the best performance in our experiments. Notably, we achieved impressive accuracies of 92%, 94.5%, and 96% for the RSNA, MIAS, and DDASM datasets, respectively. These results surpass those of existing methods, underscoring the superiority of our approach in terms of accuracy and sensitivity. In terms of future work, we envision several directions to enhance

our method. Firstly, exploring advanced deep learning techniques, such as attention mechanisms, could further improve the model's performance. Secondly, investigating the integration of additional clinical and genomic data could potentially enhance the accuracy and predictive capabilities of our system. Lastly, conducting rigorous validation on larger-scale datasets from multiple healthcare institutions would provide more robust evidence of the method's effectiveness and generalizability.

References

1. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Cancer statistics for the year 2020: An overview. Int. J. Cancer 2021, 149, 778–789. [CrossRef]
2. Lei, S.; Zheng, R.; Zhang, S.; Wang, S.; Chen, R.; Sun, K.; Zeng, H.; Zhou, J.; Wei, W. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Commun. 2021, 41, 1183–1194. [CrossRef]
3. Marks, J.S.; Lee, N.C.; Lawson, H.W.; Henson, R.; Bobo, J.K.; Kaeser, M.K. Implementing recommendations for the early detection of breast and cervical cancer among low-income women. Morb. Mortal. Wkly. Rep. Recomm. Rep. 2000, 49, 35–55.
4. Du-Crow, E. Computer-Aided Detection in Mammography; The University of Manchester: Manchester, UK, 2022.
5. Evans, A.; Trimboli, R.M.; Athanasiou, A.; Balleyguier, C.; Baltzer, P.A.; Bick, U.; Herrero, J.C.; Clauser, P.; Colin, C.; Cornford, E.; et al. Breast ultrasound: Recommendations for information to women and referring physicians by the European Society of Breast Imaging. Insights Imaging 2018, 9, 449–461. [CrossRef]
6. Schueller, G.; Schueller-Weidekamm, C.; Helbich, T.H. Accuracy of ultrasound-guided, large-core needle breast biopsy. Eur. Radiol. 2008, 18, 1761–1773. [CrossRef] [PubMed]
7. Shi, X.; Liang, C.; Wang, H. Multiview robust graph-based clustering for cancer subtype identification. IEEE/ACM Trans. Comput. Biol. Bioinform. 2022, 20, 544–556. [CrossRef]
8. Wang, H.; Jiang, G.; Peng, J.; Deng, R.; Fu, X. Towards Adaptive Consensus Graph: Multi-view Clustering via Graph Collaboration. IEEE Trans. Multimed. 2022, 1–13. [CrossRef]
9. Wang, H.; Wang, Y.; Zhang, Z.; Fu, X.; Zhuo, L.; Xu, M.; Wang, M. Kernelized multiview subspace analysis by self-weighted learning. IEEE Trans. Multimed. 2020, 23, 3828–3840. [CrossRef] 10. Wang, H.; Yao, M.; Jiang, G.; Mi, Z.; Fu, X. Graph-Collaborated Auto-Encoder Hashing for Multi-view Binary Clustering. arXiv 2023, arXiv:2301.02484.
10. Bai, J.; Posner, R.; Wang, T.; Yang, C.; Nabavi, S. Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review. Med. Image Anal. 2021, 71, 102049. [CrossRef]
11. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Trans. Neural Netw. Learn. Syst. 2022, 33, 6999–7019. [CrossRef] [PubMed]
12. Zuluaga-Gomez, J.; Al Masry, Z.; Benaggoune, K.; Meraghni, S.; Zerhouni, N. A CNN-based methodology for breast cancer diagnosis using thermal images. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 2021, 9, 131–145. [CrossRef]
13. Ero ˘glu, Y.; Yildirim, M.; Çinar, A. Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. Comput. Biol. Med. 2021, 133, 104407. [CrossRef] [PubMed]

14. Huang, Q.; Yang, F.; Liu, L.; Li, X. Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis. Inf. Sci. 2015, 314, 293–310. [CrossRef]
15. Huang, Q.; Huang, Y.; Luo, Y.; Yuan, F.; Li, X. Segmentation of breast ultrasound image with semantic classification of superpixels. Med. Image Anal. 2020, 61, 101657. [CrossRef]
16. Zhou, J.; Luo, L.; Dou, Q.; Chen, H.; Chen, C.; Li, G.; Jiang, Z.; Heng, P. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J. Magn. Reson. Imaging 2019, 50, 1144–1151. [CrossRef]
17. Yurttakal, A.H.; Erbay, H.; Ikizceli, T.; Karaçavu¸s, S. Detection of breast cancer via deep convolution neural networks using MRI images. Multimed. Tools Appl. 2019, 79, 15555–15573. [CrossRef]
18. Rahman, A.S.; Belhaouari, S.B.; Bouzerdoum, A.; Baali, H.; Alam, T.; Eldaraa, A.M. Breast mass tumor classification using deep learning. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2 February 2020; pp. 271–276.
19. Sun, L.; Wang, J.; Hu, Z.; Xu, Y.; Cui, Z. Multi-view convolutional neural networks for mammographic image classification. IEEE Access 2019, 7, 126273–126282. [CrossRef]
20. Heravi, E.J.; Aghdam, H.H.; Puig, D. Classification of Foods Using Spatial Pyramid Convolutional Neural Network. InCCIA 2016, 288, 163–168. [CrossRef]
21. Carr, C.; Kitamura, F.; Partridge, G.; Kalpathy-Cramer, J.; Mongan, J.; Andriole, K.; Lavender Vazirabad, M.; Riopel, M.; Ball, R.; Dane, S.; et al. RSNA Screening Mammography Breast Cancer Detection, Kaggle 2022. Available online: https: //kaggle.com/competitions/rsna-breast-cancer-detection (accessed on 1 December 2022).
22. Suckling, J.; Parker, J.; Dance, D.; Astley, S.; Hutt, I.; Boggis, C.; Ricketts, I.; Stamatakis, E.; Cerneaz, N.; Kok, S.; et al. Mammographic Image Analysis Society (Mias) Database v1. 21. Available online: https://www.repository.cam.ac.uk/handle/18 10/250394 (accessed on 2 February 2023).
23. Heath, M.; Bowyer, K.; Kopans, D.; Kegelmeyer, P.; Moore, R.; Chang, K.; Munishkumaran, S. Current status of the digital database for screening mammography. Digit. Mammogr. Nijmegen 1998, 1998, 457–460. [CrossRef]
24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Commun. ACM 2017, 60, 84–90. [CrossRef]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 1 July 2021; pp. 10096–10106.
27. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv 2017, arXiv:1704.04861.
28. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, New Orleans, LA, USA, 19–20 June 2022; pp. 11976–11986.
29. Ross, B.C. Mutual Information between Discrete and Continuous Data Sets. PLoS ONE 2014, 9, e87357. [CrossRef] [PubMed]
30. Azour, F.; Boukerche, A. An efficient transfer and ensemble learning based computer aided breast abnormality diagnosis system. IEEE Access 2022, 11, 21199–21209. [CrossRef]

31. Rampun, A.; Scotney, B.W.; Morrow, P.J.; Wang, H.; Winder, J. Breast density classification using local quinary patterns with various neighbourhood topologies. J. Imaging 2018, 4, 14. [CrossRef]
32. Vijayarajeswari, R.; Parthasarathy, P.; Vivekanandan, S.; Basha, A.A. Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. Measurement 2019, 146, 800–805. [CrossRef]
33. Arafa, A.A.A.; El-Sokary, N.; Asad, A.; Hefny, H. Computer-aided detection system for breast cancer based on GMM and SVM. Arab. J. Nucl. Sci. Appl. 2019, 52, 142–150. [CrossRef]
34. Diaz, R.A.; Swandewi, N.N.; Novianti, K.D. Malignancy determination breast cancer based on mammogram image with k-nearest neighbor. In Proceedings of the 2019 1st International Conference on Cybernetics and Intelligent System (ICORIS), Denpasar, Indonesia, 22 August 2019; Volume 1, pp. 233–237.
35. Agrawal, S.; Rangnekar, R.; Gala, D.; Paul, S.; Kalbande, D. Detection of breast cancer from mammograms using a hybrid approach of deep learning and linear classification. In Proceedings of the 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, India, 5 January 2018; pp. 1–6.
36. Li, B.; Ge, Y.; Zhao, Y.; Guan, E.; Yan, W. Benign and malignant mammographic image classification based on convolutional neural networks. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing, New York, NY, USA, 26 February 2018; pp. 247–251.
37. Platania, R.; Shams, S.; Yang, S.; Zhang, J.; Lee, K.; Park, S.J. Automated breast cancer diagnosis using deep learning and region of interest detection (bc-droid). In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Boston, MA, USA, 20 August 2017; pp. 536–543.
38. Swiderski, B.; Kurek, J.; Osowski, S.; Kruk, M.; Barhoumi, W. Deep learning and non-negative matrix factorization in recognition of mammograms. In Proceedings of the Eighth International Conference on Graphic and Image Processing (ICGIP 2016), Tokyo, Japan, 8 February 2017; Volume 10225, pp. 53–59.