

## Article

# Comprehensive Analysis of Mammography Images Using Multi-Branch Attention Convolutional Neural Network

Ebtihal Al-Mansour, Muhammad Hussain , Hatim A. Aboalsamh  and Saad A. Al-Ahmadi 

Department of Computer Science, CCIS, King Saud University, Riyadh 11451, Saudi Arabia; hatim@ksu.edu.sa (H.A.A.); salahmadi@ksu.edu.sa (S.A.A.-A.)

\* Correspondence: mhussain@ksu.edu.sa

**Abstract:** Breast cancer profoundly affects women's lives; its early diagnosis and treatment increase patient survival chances. Mammography is a common screening method for breast cancer, and many methods have been proposed for automatic diagnosis. However, most of them focus on single-label classification and do not provide a comprehensive analysis concerning density, abnormality, and severity levels. We propose a method based on the multi-label classification of two-view mammography images to comprehensively diagnose a patient's condition. It leverages the correlation between density type, lesion type, and states of lesions, which radiologists usually perform. It simultaneously classifies mammograms into the corresponding density, abnormality type, and severity level. It takes two-view mammograms (with craniocaudal and mediolateral oblique views) as input, analyzes them using ConvNeXt and the channel attention mechanism, and integrates the information from the two views. Finally, the fused information is passed to task-specific multi-branches, which learn task-specific representations and predict the relevant state. The system was trained, validated, and tested using two public domain benchmark datasets, INBreast and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM), and achieved state-of-the-art results. The proposed computer-aided diagnosis (CAD) system provides a holistic observation of a patient's condition. It gives the radiologists a comprehensive analysis of the mammograms to prepare a full report of the patient's condition, thereby increasing the diagnostic precision.

**Keywords:** breast cancer; mammography; deep learning; multi-label classification; convolutional neural network (CNN)



**Citation:** Al-Mansour, E.; Hussain, M.; Aboalsamh, H.A.; Al-Ahmadi, S.A. Comprehensive Analysis of Mammography Images Using Multi-Branch Attention Convolutional Neural Network. *Appl. Sci.* **2023**, *13*, 12995. <https://doi.org/10.3390/app132412995>

Academic Editors: Zahid Mehmood Jehangiri, Mohsin Shahzad and Uzair Khan

Received: 20 October 2023  
Revised: 26 November 2023  
Accepted: 30 November 2023  
Published: 5 December 2023



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

## 1. Introduction

Breast cancer is a malignant transformation and proliferation of breast cells [1]. According to the American Cancer Society [2], it is ranked as the second most prevalent cancer among women in the United States. The survival of patients depends on whether it is diagnosed at an early stage [3]. Screening programs help to identify breast cancer at early stages to facilitate early treatment, which increases patients' survival rates. In contrast, delayed diagnoses allow the disease to spread, and the cancer can grow to a stage where treatment is no longer possible. Mammography is a breast-imaging technique that is usually used to detect abnormal tissues in the breast, thereby aiding the early diagnosis of abnormalities found in a patient.

Mammography screening involves many breast views, the most common being the craniocaudal (CC) and mediolateral oblique (MLO) views. Radiologists usually use these two views to observe breast tissues from different angles and detect abnormal tissues. A mammogram aids in identifying the breast density type according to the Breast Imaging Reporting and Data System (BI-RADS), the kind of abnormality (e.g., masses and calcifications), and the level of severity of the abnormality (benign or malignant) [4]. Several approaches for automated breast cancer diagnosis are available [5]. However, the existing studies mainly focus on a single view. The research presented in [6] demonstrated that

two views are more helpful in improving a diagnosis; however, this method requires special treatment to avoid redundant data existing in both views. Therefore, using a fusion technique helps to handle this case.

In addition, most studies model the diagnosis problem as a single-label classification, such as detecting the breast density type [7–10], identifying the masses as benign or malignant [11–14], diagnosing microcalcifications as benign or malignant [15,16], or classifying both masses and calcifications as benign or malignant [17,18]. Single-label classification ignores the interdependencies between different conditions; for instance, a breast with high density is more likely to be malignant than a breast with low density. In addition, single-label classification requires developing multiple methods, each focusing on one aspect of the problem. These issues can be overcome by using a method based on multi-label classification for identifying breast cancer in an initial phase. A method that formulates the diagnosis as a multi-label classification problem can help to diagnose a patient's condition comprehensively. It leads to better diagnosis results by considering additional aspects, such as the correlation between the density type, lesion type, and states of the lesions, which radiologists usually perform. It can assist radiologists in their decision making by providing a comprehensive report of the patient's condition, increasing the precision of their diagnosis.

Given the above discussion, we formulate a diagnosis as a multi-label classification using two views. Inspired by the success of advanced CNN and transformer models [19–25], we designed the proposed method using four modules: a feature extraction module, an attention module, a fusion module, and a multi-label classification module. First, the feature extraction module employs the state-of-the-art CNN and transformer models, such as the Swin transformer and ConvNeXt, and extracts features from two views, CC and MLO. The attention module concentrates on relevant features and suppresses irrelevant features. The information extracted from the two views is fused using the fusion module. Finally, the multi-label classification module takes the fused features as input and simultaneously predicts the density type, abnormality type, and severity levels.

The key contributions of this research paper are as follows:

- We propose a method based on the multi-label classification of two-view mammography images—with CC and MLO views—that diagnoses the patient's condition comprehensively.
- We employ channel attention for selectively emphasizing the most informative channels of the input feature maps while suppressing the less informative ones.
- We propose a multi-branch deep architecture, which takes the features from two views as input and performs multi-label classification.
- We thoroughly evaluated the proposed method on two public-domain benchmark datasets, INBreast and CBIS-DDSM.

This paper is organized as follows: Section 2 discusses the 'Related Work' on breast cancer diagnosis. Section 3 details the 'Proposed Method.' The 'Evaluation Method' is described in Section 4, while Section 5 covers the 'Experiments and Results.' Section 6 includes the 'Discussion,' followed by 'Limitations and Future Work' in Section 7. This paper concludes in Section 8 with the 'Conclusions,' and the 'Nomenclature' used throughout the paper is provided at the end.

## 2. Related Work

Significant research has contributed to developing and improving advanced CAD systems, especially for detecting and diagnosing breast cancer using mammography. Within this particular context, many studies have focused on solving related problems such as mass classification as non-cancerous (benign) or cancerous (malignant), the classification of microcalcifications as non-cancerous (benign) or cancerous (malignant), the classification of both microcalcifications and masses as non-cancerous (benign) or cancerous (malignant); classification as masses and microcalcifications; the classification of mammograms based on breast density; and the multi-label classification of mammograms. This section provides

an in-depth review of the recent state-of-the-art (SOTA) research and methods that address these problems.

### 2.1. Mass Classification as Benign or Malignant

A whole mammogram image or the regions of interest (ROIs) are classified as benign or malignant, and there are many methods for this purpose.

Some recent methods classify mammogram images containing masses as benign or malignant. Chen et al. [11] developed a method using two mammography views, MLO and CC, for both breasts, extracting spatial and frequency domain features. They utilized particle swarm optimization (PSO) and support vector machine (SVM) for feature selection and classification, achieving an AUC-ROC of 0.79 for two-view and 0.75 for four-view images. Das et al. [13] implemented adaptive contrast enhancement in mammogram images, followed by segmentation and artificial neural network classification, resulting in a high accuracy of 97.2%. Sun et al. [12] introduced a multi-dilated CNN that integrates multiple views and optimizes classification accuracy by modifying the cross-entropy cost function, achieving accuracies of 82.02% and 63.06%, respectively. Nagarajan et al. [14] employed bi-dimensional empirical mode decomposition and GLCM for feature extraction, leading to AUC-ROC values of 0.9 and 0.96. Ayana et al. [26] presented a novel model employing a transformer for feature extraction combined with transfer learning, tackling the issue of imbalanced data and achieving near-perfect classification accuracy. Yu, Xiang et al. [27] developed VGG19-DF with a dRVFL classifier, showing an average AUC of 0.93 and an accuracy of 81.71%.

### 2.2. Microcalcification Classification as Either Benign or Malignant

In order to classify microcalcifications, the research has considered full mammogram images or segmented ROIs. Some recent methods classify mammogram images containing macrocalcifications as benign or malignant.

George et al. [15] proposed a multi-scale connected chain graph method for classifying microcalcifications, achieving up to 90% accuracy. Mabrouk et al. [16] enhanced mammogram images using various mechanisms and integrated feature extractions followed by ANN, KNN, and SVM classification, resulting in an accuracy of 0.96. Gerbasi et al. [28] introduced DeepMiCa, a U-Net-based network for the segmentation and classification of microcalcifications, achieving an AUC of 95%. Sarvestani et al. [29] enhanced extracted ROIs using a fuzzy system and Gabor filtering, achieving a 93% accuracy rate in classifying microcalcifications.

### 2.3. Mass and Microcalcification Classification as Benign or Malignant

In the task of classifying masses and microcalcifications as benign or malignant, two approaches can be employed: either classifying the ROIs corresponding to segmented masses and microcalcifications or performing classification on the entire mammogram image to determine its benign or malignant nature.

Li et al. [17] enhanced the DenseNet architecture for classifying mammogram images, achieving a 94.55% accuracy rate. Mohanty et al. [18] proposed a method using block-based discrete-wavelet packet transform and principal component analysis, enhanced with a kernel extreme learning machine classifier, achieving accuracy rates above 99%. Jabeen et al. [30] developed an automated framework for breast cancer classification from mammogram images, employing a novel image enhancement technique and the EfficientNet-b0 model fine-tuned via deep transfer learning. The framework, which included advanced feature extraction and optimization using the Equilibrium-Jaya controlled Regula Falsi algorithm, was tested on the CBIS-DDSM and INBreast datasets, achieving notable accuracies of 95.4% and 99.7%, respectively. Chakravarthy et al. [31] combined deep learning with metaheuristic techniques to classify mammography images, achieving up to 97.36% accuracy. Azour et al. [32] utilized ensemble learning techniques with a combination of multiple deep-learning models, achieving an accuracy of approximately 82.4%.

## 2.4. Multi-Label Classification of Mammograms

Few studies have addressed the multi-label classification of mammograms, with studies only [33] investigating this issue in recent years.

Chougrad et al. [33] introduced a CAD system for the multi-label classification of mammogram images. They employed VGG16-CNN with fine-tuning techniques and a label powerset classifier, demonstrating promising results across various datasets.

## 2.5. Analysis

The above studies indicate significant research addressing breast cancer detection in mammogram images from various perspectives and formulating different problems, such as classifying masses as non-cancerous or cancerous, microcalcifications as non-cancerous or cancerous, and masses and microcalcifications together as benign or malignant. These works achieved a favorable performance for the above-mentioned problems.

Only a few studies have focused on solving the problem of the multi-label classification of mammogram images [33], which simultaneously identifies the risk/density grade, abnormality type (e.g., mass or microcalcification), and state of the lesion (benign or malignant). The research presented in [19] adopted VGGNet and used transfer learning to fine-tune the model using ROIs before using the model as a feature extractor and multi-label classifier. Although this method provides favorable results and uses new techniques such as deep learning and transfer learning, it entails some limitations, such as using a simple CNN architecture and single-view ROIs as input. According to the study presented in [6], using multiple views can enhance prediction performance compared with using a single view. In addition, integrating residual learning into a CNN helps to overcome many challenges, such as vanishing gradients, overfitting, and complex correlations between labels [21].

Table 1 summarizes the existing research in the field. Jafari et al. [34] presented a CNN-based breast cancer detection method for mammography images, extracting features from various CNN models and selecting key features. Tested on the RSNA, MIAS, and DDSM datasets, it achieved the highest accuracy with an NN classifier: 92% for RSNA, 94.5% for MIAS, and 96% for DDSM.

**Table 1.** Comprehensive summary of the existing works.

| Paper                                      | Method                                                                                                                     | Dataset    | Performance                |
|--------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|------------|----------------------------|
| Mass classification as benign or malignant |                                                                                                                            |            |                            |
| Chen et al. [11], 2019                     | Fifty-nine features like shape, density, FFT, and DCT for feature extraction; PSO for feature selection; SVM as classifier | FFDM       | Sen. = 81<br>Spe. = 77     |
| Das et al. [13], 2020                      | Power-law transformation + shift invariant extrema characterization + ANN                                                  | MIAS, DDSM | ACC = 97.2<br>Sen. = 98.4  |
| Nagarajan et al. [14], 2019                | GLCM and GLRM from MBEMD and SVM/LDA                                                                                       | MGM        | ACC = 90<br>AUC = 0.92     |
| Ayana et al. [26], (2023)                  | Transformer                                                                                                                | DDSM       | AUC = $1 \pm 0$            |
| Yu, Xiang, et al. [27], (2023)             | CNN                                                                                                                        | DDSM       | AUC = 0.93<br>ACC = 81.71  |
| Sun et al. [12], 2019                      | CNN with dilated CONV layers                                                                                               | MIAS DDSM  | ACC = 63.06<br>ACC = 82.02 |

Table 1. Cont.

| Paper                                                             | Method                                                                                                                         | Dataset                       | Performance                                |
|-------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-------------------------------|--------------------------------------------|
| Microcalcification classification as benign or malignant          |                                                                                                                                |                               |                                            |
| George et al. [15]                                                | Topology, graph connectivity, multi-scale morphology, and KNN                                                                  | DDSM                          | ACC = 86.47<br>AUC = 0.899                 |
| Mabrouk et al. [16]                                               | HS, WT, ME, HE, Otsu, Shape, GLCM, invariant moment features, ANN, KNN, and SVM                                                | MIAS                          | ACC = 96<br>Sen. = 98                      |
| Gerbasi et al. [28], 2023                                         | UNet for segmentation + ResNet18                                                                                               | DDSM                          | AUC = 0.95                                 |
| Sarvestani et al. [29], 2023                                      | Fuzzy system + Gabor filtering for image enhancement + ANN for classification                                                  | DDSM                          | ACC = 93<br>Sen. = 95                      |
| Mass and microcalcification classification as benign or malignant |                                                                                                                                |                               |                                            |
| Li et al. [17], 2019                                              | DenseNet with the Inception structure.                                                                                         | FFDM                          | 94.55                                      |
| Mohanty et al. [18]                                               | BDWPT + PCA + WC-SSA-KELM                                                                                                      | MIAS                          | ACC = 99.28<br>Sen. = 99.44<br>AUC = 0.994 |
| Chakravarthy et al. [31], 2023                                    | Resnet18 + wKNN + PSO + DFOA + CSOA                                                                                            | MIAS                          | ACC = 84.35                                |
| F. Azour et al. [32], 2023                                        | VGG, Resnet, Inception-v3, DensNet, MobileNet, and EfficientNet                                                                |                               | ACC = 82.4                                 |
| Jabeen et al. [30], 2023                                          | Image enhancement, EfficientNet-b0, feature optimization and selection, and ML classifiers                                     | CBIS-DDSM and INBreast        | ACC = 95.4% and 99.7%                      |
| Jafari et al. [34], 2023                                          | Feature extraction from various CNNs, feature selection using mutual information, and classification with NN, kNN, RF, and SVM | RSNA, MIAS, and DDSM          | Acc = 92%, 94.5%, and 96%                  |
| Multi-label classification of mammograms                          |                                                                                                                                |                               |                                            |
| Chougrad et al. [33], 2020                                        | VGG16                                                                                                                          | DDSM, BCDR INBreast, and MIAS | Exact match: 0.822, 0.802 0.827, and 0.782 |

### 3. Proposed Method

We address the problem of simultaneously identifying the breast density type (according to BI-RADS), abnormality type (mass or calcification), and severity level/pathology (benign or malignant) from mammogram images. This is a multi-label classification problem. First, we formally define and formulate the problem and then present the details of the proposed method.

### 3.1. Problem Formulation

To screen a patient for breast cancer detection, two commonly used views of a patient's mammogram are the MLO and CC views. The problem is identifying the breast density, severity level/pathology, and findings from the two views. Based on the BI-RADS (Breast Imaging Reporting and Data System) guidelines, there are four density levels, BI-RADS I, BI-RADS II, BI-RADS III, and BI-RADS IV, which are used to classify breast density, where BI-RADS I represents the category with the lowest density, while BI-RADS IV corresponds to the category with the highest density. Additionally, there are two main abnormality types, masses, and calcifications, which are important to identify. Finally, the severity level/pathology means whether the case for the abnormality type is benign or malignant.

We represent an ROI as  $x \in \mathbb{R}^{m \times n}$ , where  $m$  and  $n$  represent the resolution of the ROI.  $x_{MLO}$  and  $x_{CC}$  stand for the ROIs extracted from the MLO view and the CC view, respectively. There are eight categories: BI-RADS I (1), BI-RADS II (2), BI-RADS III (3), BI-RADS IV (4), mass (5), calcification (6), benign (7), and malignant (8). The first four categories correspond to density types, the next two categories represent abnormality types, and the last two categories stand for severity levels (pathology). In view of this, the label for a pair of ROIs  $(x_{MLO}, x_{CC}) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$  corresponding to a patient is  $l = (l_d, l_f, l_p)$ , where  $l_d \in Y_d = \{0, 1\}^4$ ,  $l_f \in Y_f = \{0, 1\}^2$ , and  $l_p \in Y_p = \{0, 1\}^2$ ;  $Y_d$ ,  $Y_f$ , and  $Y_p$  are the label spaces of the density type, abnormality type/findings, and severity level/pathology, respectively, in one-hot encoding; and 0 means absent, and 1 means present. For example, if  $l = (l_d, l_f, l_p)$ , where  $l_d = [0\ 1\ 0\ 0]$ ,  $l_f = [1\ 0]$ , and  $l_p = [1\ 0]$ , then the density level is 2, the finding is a mass, and the case is benign. It is a multi-label classification problem. We need to design a mapping  $\varphi: \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \rightarrow Y_d \times Y_f \times Y_p$ , such that  $\varphi(x_{MLO}, x_{CC}) = (l_d, l_f, l_p)$ .

We employ deep learning techniques to design the mapping  $\varphi$  that extracts discriminative features from the input ROIs and associates them with three labels in an end-to-end manner. In the following subsections, we give the details of the deep-learning-based method for modeling  $\varphi$ .

### 3.2. Dataset Description

Our model utilizes two benchmark mammography datasets: the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [35] and INBreast [36]. Both datasets are publicly available and extensively annotated. Both cover various breast densities, abnormalities, and pathologies in different imaging views (MLO and CC). These datasets include a variety of breast densities (BI-RADS I-IV), abnormalities, and pathologies across the MLO and CC imaging views. They provide detailed annotations of ROIs and clinical findings such as masses, calcifications, and architectural distortions and categorize lesions as benign or malignant. For our research, we focused on cases where ROIs are present in both views, suitable for fusion methods while excluding cases with ROIs in only one view to ensure the relevance and comprehensiveness of our data.

### 3.3. Two-View-Based Deep Model

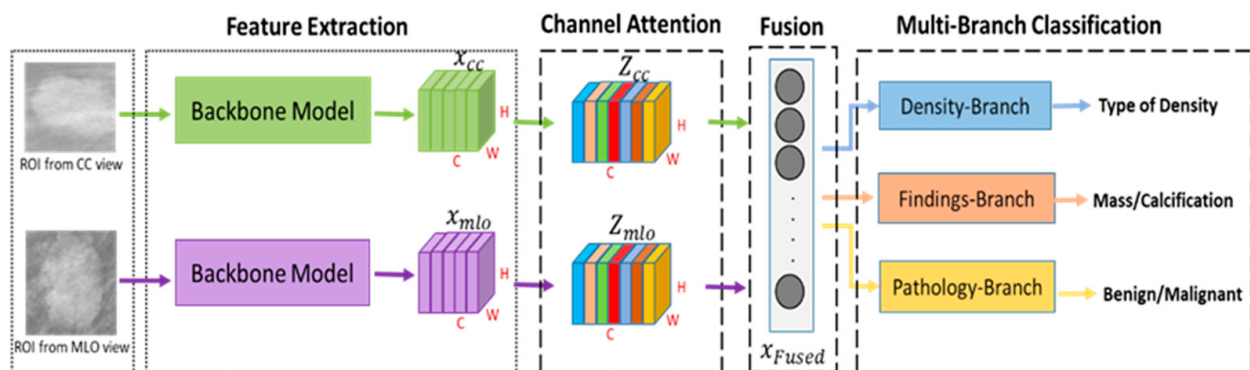
This section outlines the model architecture for simultaneously classifying mammography images into multiple labels, density, severity level, and abnormality type, using a channel-attention-based multi-task learning framework. It employs two branches, CC and MLO, with the Swin and ConvNeXT models as feature extractors.

The model consists of a feature extraction module and a fusion module. The feature extraction module, operating in the CC and MLO views, uses the Swin and ConvNeXT backbones to extract features crucial for all classification tasks, aiding in multiple-label prediction. A channel attention mechanism refines the focus on relevant features and reduces the less informative ones, ensuring selective emphasis on essential channels.

Subsequently, the fusion module merges features from both views, enhancing the multi-label classification by incorporating diverse information. Integrated features pass

through three fully connected layers, each dedicated to a specific classification task (density, severity level, or abnormality type). These layers function as classifiers, producing prediction labels for the input mammographic views, thus enabling simultaneous multi-label classification.

This architecture integrates channel attention, feature extraction, and fusion mechanisms, facilitating the learning of both shared and task-specific features, thereby improving efficiency in the multi-label classification of mammography images. This section provides an overview of the model's structure, with the subsequent sections detailing the data preprocessing, evaluation metrics, and deep learning model architecture. Figure 1 visually represents this model, highlighting its components and their interplay.



**Figure 1.** The overall architecture of the proposed system.

### 3.3.1. Details of the Model Architecture

This subsection delves into the design and functionality of our multi-label classification model for mammography images, building on the initial overview. We examine the feature extraction and fusion modules in detail, highlighting their roles in processing CC and MLO views for adequate classification. The model integrates channel attention and multi-task learning, focusing on pertinent features and learning shared and task-specific characteristics, enhancing breast cancer diagnosis via precise mammography classification.

### 3.3.2. Preprocessing

To prepare mammography images for model training, we first implement preprocessing, including resizing the ROIs to  $224 \times 224$  pixels, aligning them with the input requirements of our backbone architectures like Swin and ConvNeXT. The normalization of ROI pixel values, based on ImageNet dataset standards, ensures consistency in feature representation.

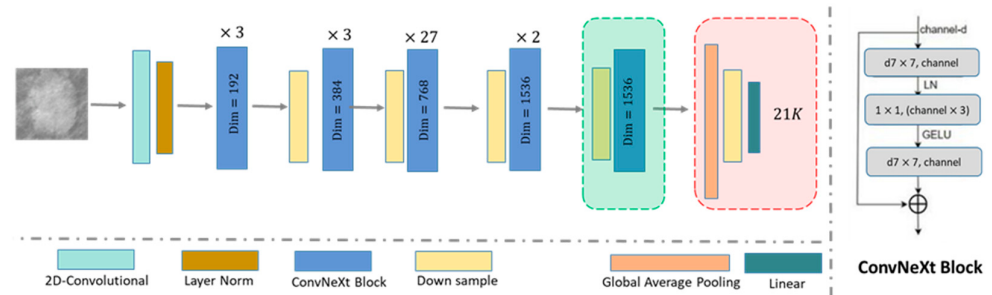
### 3.3.3. Backbone Model

Our proposed architecture uses two advanced models, ConvNeXT and the Swin transformer, for feature extraction in mammography image classification. These models were chosen for their superior performance in various computer vision tasks, marking them as leading solutions in CNN and transformer models.

The Swin transformer excels in capturing both local and global image features due to its hierarchical structure, shifted windows, and feature attention mechanism, leading to highly discriminative and informative representations. This feature makes it well-suited for feature extraction in multi-label classification.

The ConvNeXT model is known for its modularity, efficiency, and scalability, with a deep architecture that addresses gradient issues, providing stable and effective feature representations. We utilize pre-trained models on a 21 K image dataset. Their extracted features, refined through a channel attention mechanism, are combined to integrate features from both views, enhancing task-specific feature extraction for classification.

The depth of the Swin transformer and ConvNeXt models allows for the capture of distinct features crucial for each classification task, significantly boosting the model's performance in the multi-label classification of mammography images. Our final model uses ConvNeXt-L as a feature extractor, omitting its last fully connected layer and retaining the feature map from the penultimate ConvNeXt block (dimension 7,71536). The architecture, including omitted layers, is illustrated in Figure 2, with critical features indicated for clarity.

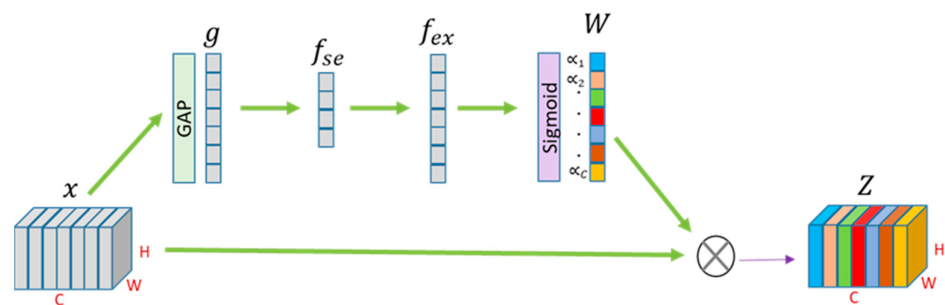


**Figure 2.** This diagram illustrates ConvNeXt-L, our backbone model, highlighting customized layers (red boundaries and shading) and key features in the penultimate layer (green box) for our mammography image classification task.

### 3.3.4. Chanel Attention Block

Not all feature channels of a feature map are equally important for the current task. Some channels may contain highly informative features that are directly relevant to the task, while others may contain less informative or redundant features that can potentially distract the network from focusing on the relevant information.

To address this, we incorporate a squeeze-and-excitation block after the last feature map generated by the backbone model. By doing so, we enable the model to adaptively weigh the importance of each channel, giving more attention to the informative channels and suppressing the less relevant ones. This dynamic attention mechanism assists the network in making better-informed decisions during the classification process. Figure 3 shows the channel attention block used.



**Figure 3.** Channel attention block.

The channel attention block takes the input feature map  $x \in \mathbb{R}^{H \times W \times C}$ , where  $C$  is the number of channels,  $H$  is the height, and  $W$  is the width. First, it applies a squeeze operation:

$$g = \text{GlobalAveragePool}(x) \quad (1)$$

$$f_{se} = \text{ReLU}(W_1 \cdot g + b_1) \quad (2)$$

where  $g \in \mathbb{R}^C$ ,  $f_{se} \in \mathbb{R}^{C/c'}$ ,  $W_1$  and  $b_1$  are the weights and biases of the FC layer, which squeezes  $g$  to  $f_{se}$  to incorporate interdependencies. Then, it applies the excitation operation:

$$w = \sigma(W_2 \cdot f_{se} + b_2) \quad (3)$$

where  $W_2$  and  $b_2$  are the weights and biases of the FC layer that adaptively recalibrate  $f_{se}$ , and  $w = [\alpha_1, \alpha_2, \dots, \alpha_C]^T$ , where  $\alpha_c$  signifies the channel-wise excitation factor for channel  $c$ . Once the channel-wise excitation factors  $w$  are computed, they are used to attend to the corresponding channels:

$$Z = w \odot x = [z_1, z_2, \dots, z_C] \quad (4)$$

### 3.3.5. Fusion Layer

Fusing features from both CC and MLO views enhances mammography classification in multi-label tasks. CC views offer a lateral perspective of breast tissue, while MLO views provide an oblique angle, capturing additional tissue. Merging these views gives classifiers a more comprehensive understanding of breast tissue, thus improving model performance.

For feature fusion, methods like concatenation, average-wise, and element-wise operations were considered. Following an ablation study, we chose average-wise operations for fusing CC and MLO view features. This involves computing the global average pooling for channel attention feature maps from both the CC and MLO branches, each resulting in a 1D vector. We then average these vectors to form the final fused feature representation. This approach aggregates relevant information from both views, balancing their differences for a robust feature representation for classification.

The dimension of the channel attention feature map from the backbone model's last layer is  $H \times W \times C$ , with  $H$  and  $W = 7$  and  $C = 1536$ . In this context, we denote the number of channels as  $D$  ( $D = 1536$ ) to differentiate from previous sections.

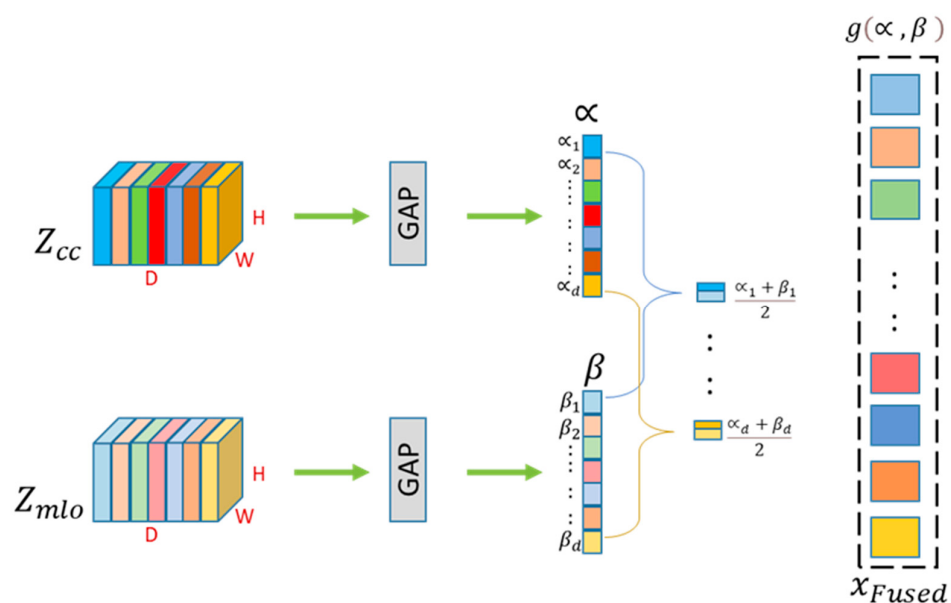
After applying GAP across  $H \times W$ , we obtain the vector with dimension  $1 \times 1536$ . The output of the fusion model is given below:

$$x_{Fused} = g(g_{GAP}(z_{cc}), g_{GAP}(z_{mlo})) \quad (5)$$

where  $x_{Fused} \in R^D$ ,  $z_{cc}, z_{mlo} \in R^{W \times H \times D}$ ,  $g_{GAP}(z_{cc}) = \alpha = [\alpha_1, \alpha_2, \dots, \alpha_d]^T$ ,  $g_{GAP}(z_{MLO}) = \beta = [\beta_1, \beta_2, \dots, \beta_d]^T$  and  $g(\alpha, \beta) = \left[ \frac{\alpha_1 + \beta_1}{2}, \dots, \frac{\alpha_d + \beta_d}{2} \right]$ .

The function is  $g_{GAP}$ , and  $g$  represents the global average pooling and point-wise average operation.

The diagram in Figure 4 shows the details of the fusion layer.



**Figure 4.** In the fusion block, GAP is calculated separately for the CC and MLO feature maps produced by attention block, followed by average CC and MLO feature correspondence to the same position.

### 3.3.6. Multi-Branch Classification

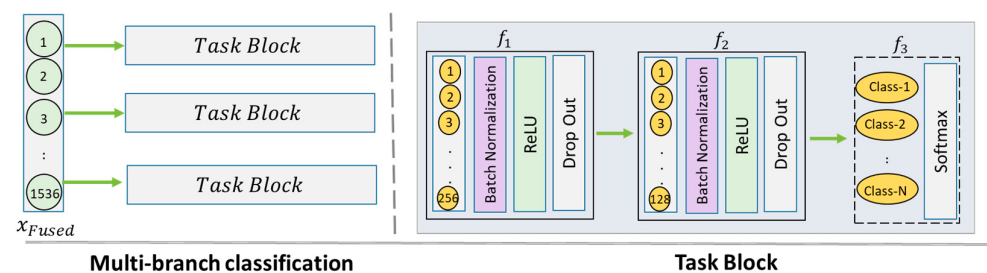
Our system classifies input ROIs into three distinct groups: density, severity level, and abnormality type. To enhance feature representation and task-specific learning, we integrated three branches of fully connected layers at the end of the fusion model.

Each branch in the multi-branch architecture focuses on learning features specific to a particular task, allowing the model to extract more effective task-specific features. In the post-fusion layer, we apply three parallel FC layers dedicated to density, severity level, and abnormality type, enabling distinct learning for each task.

The architecture includes a series of FC layers leading to a classification layer, with configurations determined via an ablation study. We use FC 256, FC 128, and dense layers with ReLU activation, batch normalization, and then the classification layer. This setup helps in learning task-specific features for efficient classification.

The final classification layer makes predictions for the three groups. Density classification uses four output neurons for BI-RADS classes, while abnormality type and severity level classifications use two output neurons each for classifying into mass or calcification and benign or malignant, respectively.

We employ the softmax activation function in each group's output layer for these tasks. This function transforms the network's output into class-specific probabilities, aiding in accurate class determination. The choice of softmax is guided by our multi-label classification needs and the nature of the tasks. Figure 5 illustrates the multi-branch classification block.



**Figure 5.** Multi-branch classification block. It takes a fused feature and passes it to three parallel task blocks. It presents a multi-branch neural network architecture with a unified task block structure replicated across three tasks: density, pathology, and severity level categories. The left side enumerates sequential task blocks, indicative of a deep and comprehensive processing approach. On the right is a breakdown of the task block. The output layer employs a softmax function, tailored with four classes for the density task and two classes for both pathology and severity levels.

In this block, we employ a series of mathematical equations to map the input data to the output, with the overarching goal of achieving multi-label classification. Specifically, we utilize these equations to model and predict probability distributions for three distinct branches: density (denoted as  $d$ ), severity level/pathology (denoted as  $p$ ), and abnormality type/findings (denoted as  $n$ ). The equations for each branch are presented as follows.

For a given branch,

$$b \ (b \in \{d, p, n\}) : P^b = \text{Softmax} \left( f_3^b \circ f_2^b \circ f_1^b (x_{Fused}) \right) \quad (6)$$

In this equation,  $P^b$  represents the probability distribution specific to branch  $b$ , and  $\{f_3^b, f_2^b, f_1^b\}$  denote the functions modeled with fully connected layers (FC1, FC2, and FC3) tailored to each branch. The output is a probability vector  $p^b$ , and the predicted class label  $l^b$  is determined as  $\max_{1 \leq i \leq k^b} p_i^b$ , where  $k^b$  is the number of classes associated with branch  $b$ .

Each function  $f_i^b$  follows a consistent structure, comprising a fully connected layer, batch normalization, ReLU activation, and dropout. The number of neurons and layer-specific parameters may vary across branches. Additionally, in each branch, we implement

a two-stage projection on the fusion features, initially reducing their dimension from  $D$  to  $D_1$ , and then further compressing them to  $D_2$ , which are variables determined according to the specific branch's requirements.

In each branch, we employ a two-stage projection on the fusion features, initially reducing their dimension from 1536 to 256, and then compressing them to a 128-dimensional space. These projections are followed by a classification layer specific to each task, which has varying numbers of classes (ranging from 2 to 4 depending on the task). This strategy yields the following benefits:

1. Dimensionality reduction: Reducing fusion features from 1536 to 256 and, further, to 128 dimensions decreases the data complexity, mitigating computational overhead and overfitting risks while retaining essential information.
2. Targeted feature learning: This helps the model learn crucial task-specific features by mapping them onto a lower-dimensional space, enhancing class discrimination.
3. Task-specific classification: Post-projection, a classification layer for each task, accommodating 2 to 4 classes, transforms features into class probabilities for precise task-specific classification.

The entire process is implemented using functions  $f_1$ ,  $f_2$ , and  $f_3$ , encompassing the operations described. This strategy integrates dimensionality reduction, task-driven feature learning, and specialized classification to optimize model performance across different tasks.

#### 4. Evaluation Method

This section describes the evaluation methods for our proposed model, including datasets, challenges, the evaluation protocol and metrics, and model training.

##### 4.1. Model Training

Our model, trained simultaneously across all branches, uses weighted cross-entropy loss for each branch, with an average calculated for the final loss. Key considerations include handling unbalanced multi-label data and utilizing three output branches for specific tasks.

We employed an RMSprop optimizer with dual learning rates for pre-trained ( $1 \times 10^{-4}$ ) and new layers ( $1 \times 10^{-3}$ ). The training involved 400 epochs, a batch size of 128, learning rate reduction on a plateau, dropout (factor 0.2), and weight decay ( $1 \times 10^{-6}$ ) for regularization. The stopping strategy had a patience of 40 epochs.

Pre-trained weights from the ImageNet-21K dataset were used for transfer learning. Data augmentation included random rotations, width and height shifts, horizontal flips, and zoom, coupled with normalization.

##### 4.2. Evaluation Protocol and Metrics

The datasets were split into 80:20 for training and testing, with a 10% validation set, using 5-fold cross-validation [37]. The evaluation metrics included mean average precision, F1-score, Hamming loss, coverage, ranking loss, and exact match [38,39].

The system was implemented using TensorFlow, Keras, and PyTorch in Anaconda Navigator (2022) on an Intel(R) Core(TM) i9-9900K CPU with a GPU with 32 GB memory and 64.0 GB RAM.

#### 5. Experiments and Results

This section outlines the experiments to evaluate our multi-label classification model using the CBIS-DDSM and INBreast datasets. We tested the model with various SOTA backbone models, assessing its performance using metrics like F1-score, mean average precision, and exact match. The focus was on classifying breast cancer and abnormalities in terms of density, severity level, and abnormality type. Additionally, an ablation study examined different fusion methods and configurations for a multi-branch block, culminating in a comprehensive assessment of our proposed fusion model.

### 5.1. Ablation Study

We conducted this study to analyze which fusion technique is more suitable and which configuration is the best for a multi-branch block.

#### 5.1.1. Which Fusion Technique Is Suitable?

We conducted this study to analyze and evaluate the model's performance when using different common fusion techniques. The goal was to determine which fusion method was more suitable for our model. We examined three fusion methods: concatenation, element-wise addition, and averaging. Table 2 shows the results of each technique.

**Table 2.** Performance results for CBIS-DDSM with different fusion methods.

| Fusion Method | F1%                | mAP%               | EM%                |
|---------------|--------------------|--------------------|--------------------|
| Concatenation | 94.29% $\pm$ 0.011 | 90.54% $\pm$ 0.016 | 86.69% $\pm$ 0.016 |
| Multiply      | 93.97% $\pm$ 0.01  | 89.8% $\pm$ 0.013  | 85.6 % $\pm$ 0.023 |
| Average       | 94.7% $\pm$ 0.01   | 91.3% $\pm$ 0.017  | 86.9% $\pm$ 0.019  |

The table compares three fusion techniques: concatenation, element-wise, and average-wise, using metrics like F1-score, mAP, and EM. Average-wise outperformed others in all metrics, indicating its effectiveness in integrating features from two views. Consequently, we selected average-wise as our preferred fusion method.

#### 5.1.2. How Many Hidden Layers in Multi-Branch Block

We conducted a study on how many deep layers we needed for our model. The table below shows the model's performance for a multi-branch block in a fusion model with different numbers of hidden layers. In this study, we examined up to five layers. Table 3 shows that the highest F1-score, mAP, and EM were achieved with two fully connected layers, with scores of 94.72% and 91.33 for F1-score, mAP, and EM, respectively.

**Table 3.** The performance of a multi-branch block fusion model with different configurations.

| # Layers | # of Neurons in Each Layer | F1%   | mAP%  | EM%   |
|----------|----------------------------|-------|-------|-------|
| 1        | 256                        | 94.44 | 90.69 | 86.69 |
| 2        | 256 and 128                | 94.72 | 91.33 | 86.86 |
| 3        | 512, 256, and 128          | 94.27 | 90.5  | 86.45 |
| 4        | 1024, 512, 256, and 128    | 94.24 | 90.4  | 86.53 |
| 5        | 2048, 1024, 512, and 256   | 94.48 | 90.9  | 86.69 |

This study recommends using two hidden layers to provide a good balance and trade-off between the model's performance and complexity, as it achieved the best performance among all evaluated metrics.

#### 5.1.3. Which Backbone Model Is Better?

The question was which of the backbone models used in the experiments performs the best among selected SOTA pre-trained CNN and transformer models? We used the CBIS-DDSM dataset to test those two models. As shown in Table 4, ConvNeXt was selected as a backbone model for our proposed model as it obtained the best result. This decision was made because ConvNeXt outperformed the Swin transformer model in terms of F1-score, RL, and Cov.

**Table 4.** The performance of different backbone models on the DDSM dataset.

| Model            | F1%   | HL   | mAP% | RL   | Cov  | EM%  |
|------------------|-------|------|------|------|------|------|
| ConvNeXt         | 0.910 | 0.07 | 0.85 | 0.12 | 4.11 | 0.78 |
| Swin transformer | 0.90  | 0.07 | 0.85 | 0.13 | 4.12 | 0.78 |

#### 5.1.4. The Effect of Fusion

Fusing features from both CC and MLO views enhances mammography classification, as they capture different breast tissue aspects. CC views offer a lateral perspective, while MLO views include additional tissue through an oblique angle. This fusion provides a more comprehensive tissue analysis, potentially improving model performance. Our experiments using SOTA backbone models compared the efficacy of single- and dual-view approaches for multi-label classification. The results in Table 5 show that dual-view fusion surpasses single-view classifiers in performance on the CBIS-DDSM dataset.

**Table 5.** The influence of the overall performance when comparing the single view and two views of the CBIS-DDSM dataset.

| View        | Model                   | F1%    | mAP%   | EM%    |
|-------------|-------------------------|--------|--------|--------|
| Single view | Swin                    | 90%    | 85%    | 78%    |
|             | ConvNeXt                | 91.0%  | 85%    | 78%    |
| Dual view   | Swin                    | 94.18% | 90.15% | 86.28% |
|             | ConvNeXt                | 94.54% | 90.88% | 87.27% |
|             | ConvNeXt with attention | 94.72% | 91.33% | 86.86% |

#### 5.1.5. The Effect of the Attention Module

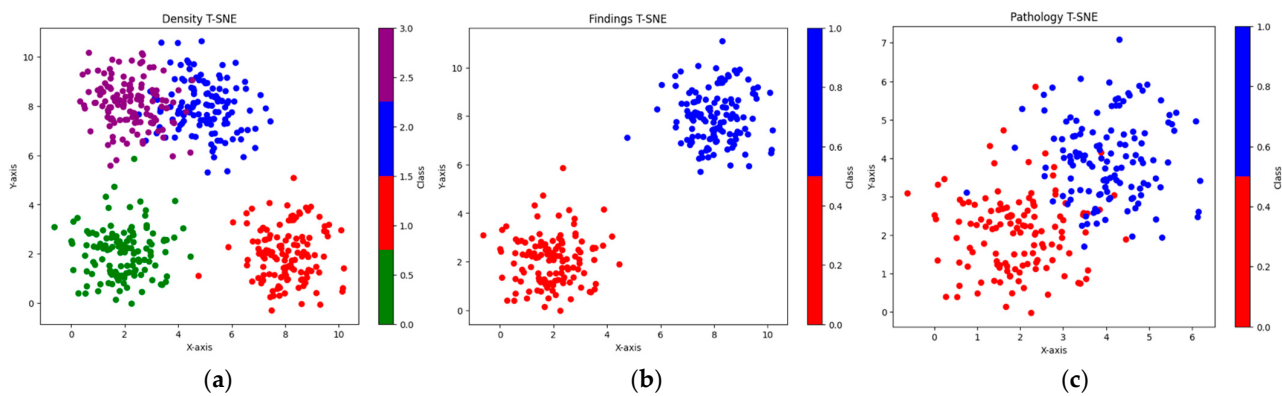
In deep learning models, not all channels in the input image contribute equally to classification, as some may contain irrelevant or noisy data. To address this, we used a channel attention mechanism in our backbone model and fusion process. This experiment demonstrated that channel attention positively impacts performance. Table 6 shows that integrating this mechanism into the ConvNeXt fusion model improves most evaluation metrics. The F1-score increased from 94.54% to 94.72%, and the HL decreased from 0.0364 to 0.0355, while mAP and RL showed slight improvements.

**Table 6.** This table shows the impact of incorporating the attention mechanism into the fusion model.

| Model    | Chanel Attention | F1%    | HL     | mAP%   | RL    | Cov   | EM%    |
|----------|------------------|--------|--------|--------|-------|-------|--------|
| ConvNeXt | No               | 94.54% | 0.0364 | 90.88% | 0.071 | 3.636 | 87.27% |
| ConvNeXt | Yes              | 94.72% | 0.0355 | 91.33% | 0.070 | 3.657 | 86.86% |

#### 5.1.6. The Analysis of Features

In this section, we assess the performance of the proposed model on the test data by analyzing the distribution of the features learned by each classification branch. Figure 6 shows how the features are discriminative for each label. This indicates that the proposed model can extract the discriminative features with less overlapping.



**Figure 6.** Visualization of the distribution of features (a), (b), and (c) indicate the distribution of the features among density, abnormality type (findings), and severity level (pathology), respectively.

## 6. Discussion

We employed the two-views (CC and MLO) technique to construct and evaluate a proposed multi-label classification of breast cancer into eight labels corresponding to three groups, simultaneously, i.e., density (I, II, III, or IV), abnormality type/findings (mass or calcification), and severity level/pathology (benign or malignant). The CBIS-DDSM benchmark dataset was used to decide which technique is more suitable for this task; for example, it was used to evaluate the effect of different configurations, backbone models, fusion methods, and so on. The SOTA backbone model was used as the core component of the model to enable the automatic extraction of the features without human intervention.

By relying on both views, the system can detect and identify the abnormalities that might not exist and are visible in a single view. Furthermore, certain irregularities might only be seen in one view and not the other; accordingly, using both views raises the chance of identifying the abnormalities and decreases the possibility of missing and neglecting any potential concerns and important information.

We examined the SOTA CNN and transformer models, including ConvNeXt and the Swin transformer, and concluded that ConvNeXt was the best model for our proposed CAD system.

After evaluating the performance of each model, we found that ConvNeXt was the most appropriate option for our specific tasks. ConvNeXt outperformed other architectures due to its use of depth-wise separable convolutions. These convolutions required fewer learnable parameters than traditional convolutional layers, making the network more effective and reducing overfitting. In addition, it can be adaptive to the data by dynamically building networks with a varying number of layers, depending on the complexity of the data. This makes the architecture more adaptable and flexible to several tasks and allows it to extract important features more efficiently.

Moreover, ConvNeXt utilizes multi-scale processing. This allows for extracting the features from multiple levels of abstraction. This helps the model to extract and capture extra related information and enhance the performance of the introduced system.

Three different fusion techniques were examined. The average-wise method was selected based on the results as it gave the best performance among the methods.

As the proposed system classifies the input ROIs into three non-overlapped groups, integrating three branches into the end of the backbone model assists in enhancing the performance as it helps improve the feature representation and task-specific learning. Each branch of the multi-branch architecture focuses on learning features specific to a particular task, leading to more effective feature representation and task-specific learning. This helps the model capture more task-related features, yielding a better performance on that task, leading to improved performance.

On the other hand, incorporating a channel attention mechanism for each task in the classification layer of a deep neural network has a slightly positive impact on the

model's performance. Channel attention leads the model to focus on the significant relevant channels and suppress noninformative ones.

Incorporating channel attention into ConvNeXt enables the selective enhancement of the most informative channels in the feature maps. It leads to a more effective representation of features and, eventually, results in enhanced model performance.

### 6.1. Performance Comparison with the SOTA Methods

As shown in the related work section, few studies have addressed the multi-label classification of mammograms, with only [33] investigating this issue in recent years.

The multi-label classification of mammogram images proposed by Chougrad et al. [33] simultaneously classifies a mammogram into its abnormality type/findings (mass/microcalcifications), severity level/pathology (benign/malignant), and density class (I–IV). They used ROIs as input for the deep learning module, transfer learning to initiate the VGG16-CNN weights with a fine-tuning technique, and a label powerset classifier for classification. The proposed algorithm considers the correlation between labels. They evaluated their method using the CBIS-DDSM [35], BCDR [40], INBreast [36], and MIAS [41] datasets with multiple metrics.

On the other hand, the introduced system outperformed the SOTA methods as we used the fusion method and SOTA backbone model in addition to the multi-branch and channel attention techniques. It achieved a higher performance in all metrics on both the CBIS-DDSM and INBreast datasets. It is observed that fusing two view inputs improves the overall performance across all performance metrics when compared with a single view. This is because several irregularities and features might be visualized better in one view than in the other, and combining the features from two views leads to a comprehensive assessment of the breast situation. In addition, utilizing the SOTA ConvNeXt performs better than using other backbone models.

Table 7 gives the performance metrics for the proposed and existing methods on the INBreast and CBIS-DDSM datasets.

**Table 7.** The comparison of the proposed method with SOTA method on CBIS-DDSM and INBreast datasets.

| Reference            | Method   | F1                | HL                | mAP               | RL                | Cov               | EM                |
|----------------------|----------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| CBIS-DDSM            |          |                   |                   |                   |                   |                   |                   |
| Chougrad et al. [33] | VGG16    | 93.5% $\pm$ 0.019 | 0.047 $\pm$ 0.022 | 89.5% $\pm$ 0.017 | 0.087 $\pm$ 0.025 | 3.895 $\pm$ 0.320 | 82.2% $\pm$ 0.041 |
| Proposed method      | ConvNeXt | 94.7% $\pm$ 0.01  | 0.036 $\pm$ 0.007 | 91.3% $\pm$ 0.017 | 0.07 $\pm$ 0.012  | 3.66 $\pm$ 0.095  | 86.9% $\pm$ 0.019 |
| INBreast             |          |                   |                   |                   |                   |                   |                   |
| Chougrad et al. [33] | VGG16    | 94.2% $\pm$ 0.102 | 0.042 $\pm$ 0.092 | 88.7% $\pm$ 0.140 | 0.082 $\pm$ 0.125 | 3.723 $\pm$ 0.147 | 82.7% $\pm$ 0.092 |
| Proposed method      | ConvNeXt | 95.1% $\pm$ 0.016 | 0.032 $\pm$ 0.013 | 92.8% $\pm$ 0.025 | 0.065 $\pm$ 0.021 | 3.55 $\pm$ 0.174  | 88.9% $\pm$ 0.035 |

For the CBIS-DDSM dataset, Chougrad et al. [33], who applied the VGG16 architecture, attained an F1-score of 0.935, a Hamming loss (HL) of 0.047, a mean average precision (mAP) of 0.895, a ranking loss (RL) of 0.087, a coverage (Cov) of 3.895, and an exact match (EM) of 0.822. The proposed method, which utilized the ConvNeXt architecture with a fusion technique, outperformed the existing method, achieving a higher F1-score of 0.947, a lower HL of 0.036, a higher mAP of 0.913, a lower RL of 0.07, a lower Cov of 3.66, and a higher EM of 0.869.

For the INBreast dataset, Chougrad et al.'s method [33] achieved the following performance metrics when testing their proposed method using the INBreast dataset: an F1-score of 0.935, a hamming loss of 0.047, a mean average precision of 0.895, a ranking loss of 0.087, a coverage of 3.895, and an exact match of 0.822. The proposed method outperformed their method, achieving a higher F1-score of 95.13, a lower HL of 0.032, a higher mAP of 0.928, a lower RL of 0.06565, a lower Cov of 3.557143, and a higher EM of 0.88857.

These results indicate the superiority of the proposed method compared with the existing methods for both the CBIS-DDSM and INBreast datasets.

On the CBIS-DDSM dataset, our proposed method achieved improvement for all matrices. For the F1-score, the proposed method improved by 1.33% compared with Chougrad et al.'s method [33]. For HL, the proposed method significantly improved by 23.4% compared with Chougrad et al.'s method [33]. The proposed method showed a 4.60% improvement in mAP compared with Chougrad et al.'s [33] method. For RL, the proposed method significantly improved by 16.09% compared with Chougrad et al.'s method [33]. Regarding Cov, the proposed method showed a slight enhancement of 0.60% compared with Chougrad et al.'s [33] method. The proposed method improved by 5.06% in EM compared with Chougrad et al.'s method [33].

On the other hand, regarding the INBreast dataset, the proposed system outperformed the existing method in the INBreast dataset, with an improvement of 0.99% in the F1-score, 23.1% in Hamming loss, 4.64% in mAP, 20% in ranking loss, 4.45% in the coverage score, and 7.42% in exact match.

Overall, the proposed method achieved notable improvements compared with the previous methods. The results in Table 7 indicate that the proposed method using the ConvNeXt architecture with two views performs better than the method using both the VGG16 and Efficientnetb3 architectures among all the performance metrics. The higher F1-score, EM, and mAP, and the lower HL, Cov, and RL show that the proposed method achieves better accuracy and precision in predicting the presence of abnormalities in mammograms and classifying breast density.

In addition, the ROC curves shown in Figure 7, for each category—density, abnormality type (case), and severity level (pathology)—demonstrate the model's classification effectiveness. For 'Density,' the model's ability to differentiate between multiple density classes yielded an AUC score of 0.91, demonstrating its discriminative power. The 'Abnormality type (case)' category exhibited near-perfect classification with an AUC of 0.99, indicative of the model's exceptional accuracy in case determination. Similarly, the 'severity level (pathology)' category showed an AUC of 0.96, reflecting the model's high proficiency in identifying pathological features. These AUC values, significantly exceeding the 0.5 threshold of random chance, underscore the model's potential to provide reliable and accurate diagnostic assistance in mammographic analysis.

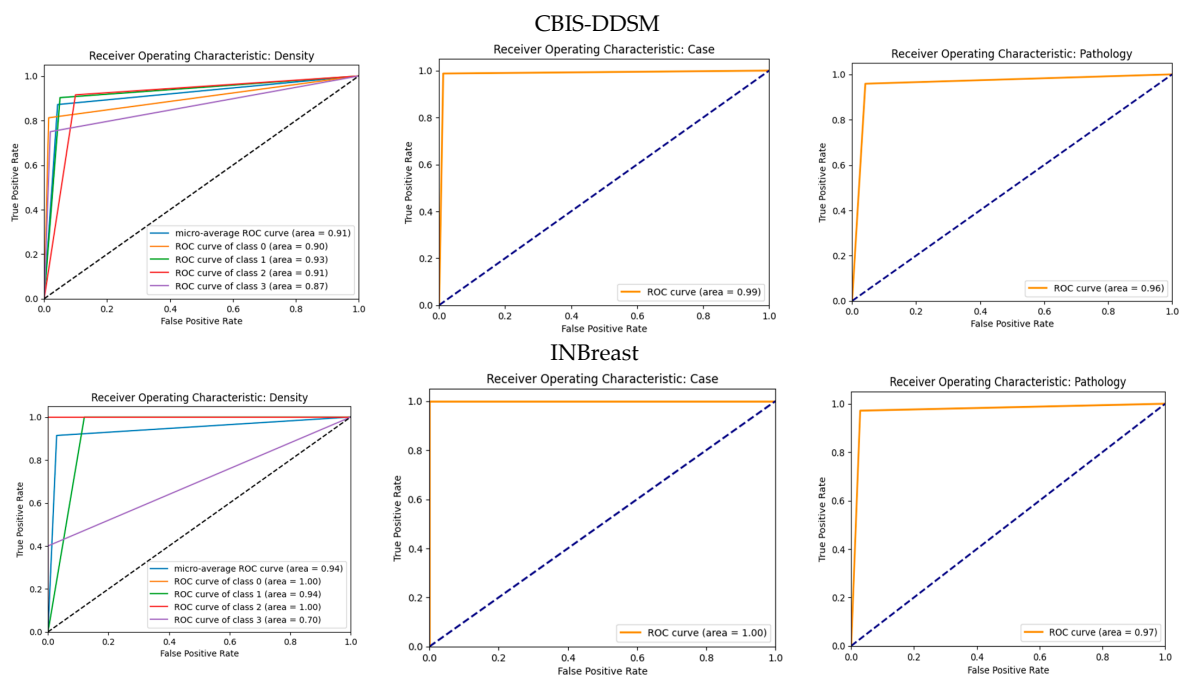


Figure 7. ROC curves for each case.

### 6.2. Comparative Analysis with Recent Deep Networks

Our research advances the multi-label classification of mammograms by integrating state-of-the-art techniques. It contrasts with previous studies like that of Chougrad et al. [33], which primarily relied on the simpler CNN model with the VGG16 architecture. Our model employs the ConvNeXt architecture, utilizing depth-wise separable convolutions. This design requires fewer learnable parameters, reducing the risk of overfitting and proving more adaptability in dynamically building networks based on data complexity, a clear advantage over the more rigid structure of VGG16.

Unlike traditional CNN models like VGG16, our approach incorporates advanced techniques like residual learning and transformer mechanisms. It also incorporates a channel attention mechanism based on squeeze-and-excitation, focusing the model on the most significant features and suppressing the less important ones in the input feature maps. A fusion method is utilized to integrate features from both CC and MLO views, providing a holistic analysis and a step forward from traditional single-view analyses, leading to a more complete representation of mammograms.

The dual-view technique (with CC and MLO views) significantly enhances our model's ability to detect abnormalities that might be visible in one view but not the other. It is a notable improvement over single-view analysis. Our exploration of various fusion methods revealed the average-wise method as the most effective. Incorporating a multi-branch architecture and channel attention techniques leads to more effective feature representation and task-specific learning. These innovative approaches contrast with traditional single-branch architectures, enhancing our model's accuracy and precision in predicting abnormalities.

### 6.3. Quantitative Analysis of Proposed Model

In this subsection, we provide the qualitative analysis of our model's performance, as illustrated in Figure 8. This analysis complements our quantitative findings and gives insight into the practical application of our method.

Figure 8 is structured into two columns, each representing a progressive increase in breast tissue density from left to right. Column 1 corresponds to the lowest density (BI-RADS I), while column 2 represents the highest density (BI-RADS IV). Within each column, rows 1 and 2 illustrate the differences between benign and malignant calcifications, and rows 3 and 4 distinguish between benign and malignant masses. This arrangement demonstrates the algorithm's capability to analyze and accurately differentiate breast abnormalities across tissue densities, showcasing its robustness and precision in lower- and higher-density scenarios.

Additionally, our model's inference efficiency is noteworthy, processing each image in an average of 286 milliseconds on advanced hardware. This speed is crucial for rapid and accurate breast tissue analysis in clinical settings. The combination of our model's diagnostic precision demonstrated in the annotated images and its quick processing time solidifies its potential as an effective tool in medical image analysis.

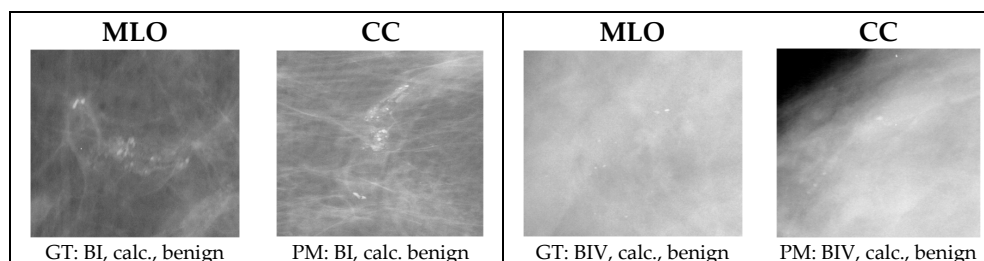
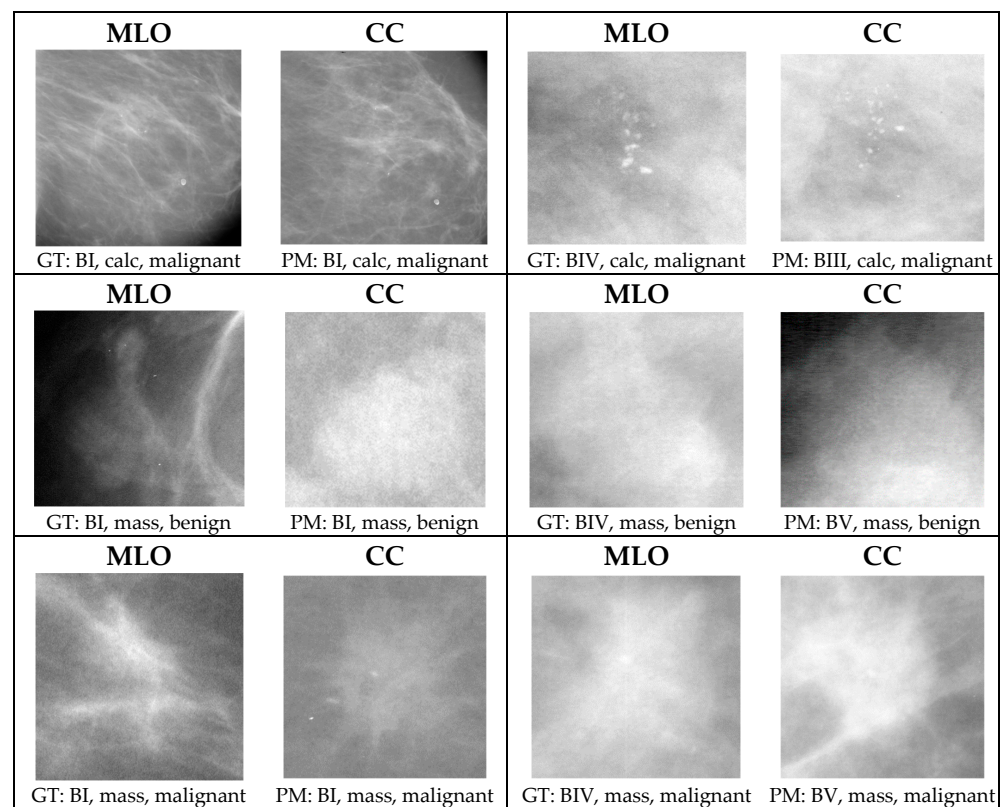


Figure 8. Cont.



**Figure 8.** The ground truth (GT) labels and the labels predicted with the proposed model (PM) of each pair from two views.

## 7. Limitations and Future Work

The current work concentrates only on masses and calcifications in mammograms. It does not predict other abnormalities like asymmetries and architectural distortions. Furthermore, it is not able to reveal the regions in mammograms that play key roles in decision making, i.e., the interpretability of a decision.

In future studies, we aim to expand the scope of the algorithm by including abnormalities like asymmetries and architectural distortions and explore various fusion techniques for integrating CC and MLO views and different channel attention mechanisms, including SKNet, to enhance its performance in a multi-label classification system.

Additionally, we plan to implement spatial attention mechanisms for revealing the regions in mammogram images that play crucial roles in decision making and improving accuracy by utilizing complete contextual information. Future work will also investigate multi-label classification algorithms for more effective breast cancer diagnosis and risk assessment, particularly those centered on problem transformation and adaptation techniques.

The benchmark datasets, which we used to develop and evaluate the proposed system, are annotated according to the fourth edition of BI-RADS. However, the fifth edition of BI-RADS is available now, and there is a need to annotate the datasets according to this edition and evaluate the system's performance.

## 8. Conclusions

This research paper presented an innovative deep-learning-based model precisely designed to utilize the power of dual mammogram views: the craniocaudal (CC) and mediolateral oblique (MLO) views. This model's main objective is to diagnose comprehensively by simultaneously classifying mammograms based on their density, severity level/pathology, and abnormality type/findings. To achieve this, our model incorporates the state-of-the-art ConvNeXt as its backbone model. The design of this model is based on

techniques like residual learning and transformer mechanisms, setting a solid foundation for advanced deep learning techniques. We utilized a channel attention mechanism based on squeeze-and-excitation to improve the ability of the model to concentrate on the most significant features and suppress the less important ones in the input feature maps. We employed an average-element-wise fusion method to consolidate the features' importance from both the CC and MLO views. This fusion method operates as a new layer within the model, seamlessly integrating the information extracted from the two views. This collaborative data merging ensures that no essential details are neglected, and the model acquires a holistic understanding of the mammogram image. Recognizing the diverse nature of the classification tasks, we introduced a multi-task/multi-branch architecture. This architecture tailors the feature-learning process to the unique requirements of each task: density, abnormality type, and lesion severity level. These tasks each have their distinct path within the architecture, facilitating more effective feature representation and task-specific learning. This enables our model to provide more accurate diagnoses for specific medical aspects. Employing multi-label learning helps enhance the model's ability to learn task-specific features and improves the model's performance. The proposed method was evaluated using benchmark datasets, the CBIS-DDSM and INBreast datasets, and outperformed SOTA. The proposed model is limited to masses and microcalcifications, and its extension to include other abnormalities will be the subject of future work.

**Author Contributions:** Conceptualization, E.A.-M. and M.H.; data curation, E.A.-M.; formal analysis, E.A.-M. and S.A.A.-A.; funding acquisition M.H.; methodology, E.A.-M. and M.H.; project administration, M.H. and H.A.A.; resources, M.H. and H.A.A.; software, E.A.-M.; supervision, M.H. and H.A.A.; validation, E.A.-M.; visualization, E.A.-M.; writing—original draft, E.A.-M.; writing—review and editing, M.H. and S.A.A.-A. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors extend their appreciation to the Deputyship for Research and Innovation of the Ministry of Education in Saudi Arabia for funding this research work under project no. IFKSUOR3-482-2.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Public-domain datasets were used for the experiments. The CBIS-DDSM dataset is available at: <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM> (accessed on 29 November 2023). The InBreast dataset is available at: <https://biokeanos.com/source/INBreast>, <https://www.kaggle.com/datasets/martholi/inbreast>. (accessed on 29 November 2023).

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

## Nomenclature

The following is a list of symbols and terms used in this manuscript along with their definitions for reference.

|            |                                                        |
|------------|--------------------------------------------------------|
| $x$        | Input feature map                                      |
| $g$        | GlobalAveragePool function                             |
| $W_1, b_1$ | Weights and biases of the first fully connected layer  |
| $f_{se}$   | Squeezed feature map after ReLU activation             |
| $W_2, b_2$ | Weights and biases of the second fully connected layer |
| $f_{ex}$   | Excited feature map                                    |
| $\sigma$   | Sigmoid activation function                            |
| $W$        | Channel-wise excitation factors                        |
| $Z$        | Feature map with applied attention                     |

|                   |                                                                                        |
|-------------------|----------------------------------------------------------------------------------------|
| $z_i$             | $i^{th}$ channel of $Z$                                                                |
| $\alpha_i$        | Channel-wise excitation factor for channel $i$                                         |
| $x^a$             | Weighted feature map                                                                   |
| $z_{cc}, z_{mlo}$ | Feature maps for CC and MLO views.                                                     |
| $g_{GAP}$         | GlobalAveragePool operation                                                            |
| $x_{Fused}$       | Fused feature map                                                                      |
| $b$               | Label for breast cancer classification (density d, pathology p, and normal n)          |
| $p^b$             | Probability of class $b$ after applying softmax                                        |
| $D$               | Number of channels                                                                     |
| $\alpha, \beta$   | Vectors representing global average pooled features for CC and MLO views, respectively |
| $N$               | Total number of instances                                                              |
| $y, \hat{y}_n$    | Ground truth label and predicted label for the $n^{th}$ instance                       |
| $F1_{score}$      | Harmonic mean of precision and recall                                                  |
| $MAP$             | Mean average precision                                                                 |
| $r(x_n, y)$       | Rank of label $y$ for $n^{th}$ instance                                                |
| $HL$              | Hamming loss                                                                           |
| $\Delta$          | Symmetric difference operator                                                          |
| $RL$              | Ranking loss                                                                           |
| $\overline{y_n}$  | Complement of the set $y_n$                                                            |
| $Coverage$        | Average count of labels to examine for all reference labels                            |
| $EM$              | Exact match                                                                            |

## References

1. National Cancer Institute (United States). Available online: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (accessed on 9 February 2015).
2. American Cancer Society. Available online: <https://www.cancer.org/healthy/find-cancer-early/womens-health/cancer-facts-for-women.html> (accessed on 1 August 2019).
3. American Cancer Society. Available online: <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html> (accessed on 8 August 2019).
4. Allison, K.H.; Abraham, L.A.; Weaver, D.L.; Tosteson, A.N.A.; Nelson, H.D.; Onega, T.; Geller, B.M.; Kerlikowske, K.; Carney, P.A.; Ichikawa, L.E.; et al. Trends in breast tissue sampling and pathology diagnoses among women undergoing mammography in the US: A report from the breast cancer surveillance consortium. *Cancer* **2015**, *121*, 1369–1378. [CrossRef] [PubMed]
5. Ramadan, S.Z. Methods Used in Computer-Aided Diagnosis for Breast Cancer Detection Using Mammograms: A Review. *J. Healthc. Eng.* **2020**, *2020*, 9162464. [CrossRef] [PubMed]
6. Abdelhafiz, D.; Yang, C.; Ammar, R.; Nabavi, S. Deep convolutional neural networks for mammography: Advances, challenges and applications. *BMC Bioinform.* **2019**, *20*, 281. [CrossRef] [PubMed]
7. Ahn, C.K.; Heo, C.; Jin, H.; Kim, J.H. A novel deep learning-based approach to high accuracy breast density estimation in digital mammography. In *Medical Imaging 2017: Computer-Aided Diagnosis*; SPIE: Bellingham, WA, USA, 2017; Volume 10134, pp. 691–697.
8. Ionescu, G.V.; Fergie, M.; Berks, M.; Harkness, E.F.; Hulleman, J.; Brentnall, A.R.; Cuzick, J.; Evans, D.G.; Astley, S.M. Prediction of reader estimates of mammographic density using convolutional neural networks. *J. Med. Imag.* **2019**, *6*, 031405. [CrossRef] [PubMed]
9. Wu, N.; Geras, K.J.; Shen, Y.; Su, J.; Kim, S.G.; Kim, E.; Wolfson, S.; Moy, L.; Cho, K. Breast density classification with deep convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 6682–6686.
10. Mohamed, A.A.; Berg, W.A.; Peng, H.; Luo, Y.; Jankowitz, R.C.; Wu, S. A deep learning method for classifying mammographic breast density categories. *Med. Phys.* **2018**, *45*, 314–321. [CrossRef] [PubMed]
11. Chen, X.; Zargari, A.; Hollingsworth, A.B.; Liu, H.; Zheng, B.; Qiu, Y. Applying a new quantitative image analysis scheme based on global mammographic features to assist diagnosis of breast cancer. *Comput. Methods Programs Biomed.* **2019**, *179*, 104995. [CrossRef]
12. 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]
13. Das, P.; Das, A. Shift invariant extrema based feature analysis scheme to discriminate the spiculation nature of mammograms. *ISA Trans.* **2020**, *103*, 156–165. [CrossRef]
14. Nagarajan, V.; Britto, E.C.; Veeraputhiran, S.M. Feature extraction based on empirical mode decomposition for automatic mass classification of mammogram images. *Med. Nov. Technol. Devices* **2019**, *1*, 100004. [CrossRef]
15. George, M.; Chen, Z.; Zwiggelaar, R. Multiscale connected chain topological modelling for microcalcification classification. *Comput. Biol. Med.* **2019**, *114*, 103422. [CrossRef]

16. Mabrouk, M.S.; Afify, H.M.; Marzouk, S.Y. Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques. *Ain Shams Eng. J.* **2019**, *10*, 517–527. [\[CrossRef\]](#)
17. Li, H.; Zhuang, S.; Li, D.-A.; Zhao, J.; Ma, Y. Benign and malignant classification of mammogram images based on deep learning. *Biomed. Signal Process. Control* **2019**, *51*, 347–354. [\[CrossRef\]](#)
18. Mohanty, F.; Rup, S.; Dash, B.; Majhi, B.; Swamy, M.N.S. An improved scheme for digital mammogram classification using weighted chaotic salp swarm algorithm-based kernel extreme learning machine. *Appl. Soft Comput.* **2020**, *91*, 106266. [\[CrossRef\]](#)
19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
20. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
21. Chen, Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; Feng, J. Dual path networks. In *NIPS*; Curran Associates Inc.: Red Hook, NY, USA, 2017.
22. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
23. 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, Las Vegas, NV, USA, 26 June–1 July 2016.
24. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
25. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. *arXiv* **2017**, arXiv:1709.01507.
26. Ayana, G.; Dese, K.; Dereje, Y.; Kebede, Y.; Barki, H.; Amdissa, D.; Husen, N.; Mulugeta, F.; Habtamu, B.; Choe, S.-W. Vision-Transformer-Based Transfer Learning for Mammogram Classification. *Diagnostics* **2023**, *13*, 178. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Yu, X.; Ren, Z.; Guttery, D.S.; Zhang, Y.-D. DF-dRVFL: A novel deep feature based classifier for breast mass classification. *Multimed. Tools Appl.* **2023**, 1–30. [\[CrossRef\]](#)
28. Gerbasi, A.; Clementi, G.; Corsi, F.; Albasini, S.; Malovini, A.; Quaglini, S.; Bellazzi, R. DeepMiCa: Automatic segmentation and classification of breast MicroCalcifications from mammograms. *Comput. Methods Programs Biomed.* **2023**, *235*, 107483. [\[CrossRef\]](#)
29. Sarvestani, Z.M.; Jamali, J.; Taghizadeh, M.; Dindarloo, M.H.F. A novel machine learning approach on texture analysis for automatic breast microcalcification diagnosis classification of mammogram images. *J. Cancer Res. Clin. Oncol.* **2023**, *149*, 6151–6170. [\[CrossRef\]](#)
30. Jabeen, K.; Khan, M.A.; Balili, J.; Alhaisoni, M.; Almujaally, N.A.; Alrashidi, H.; Tariq, U.; Cha, J.-H. BC2NetRF: Breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection. *Diagnostics* **2023**, *13*, 1238. [\[CrossRef\]](#)
31. Chakravarthy, S.S.; Bharanidharan, N.; Rajaguru, H. Deep Learning-Based Metaheuristic Weighted K-Nearest Neighbor Algorithm for the Severity Classification of Breast Cancer. *IRBM* **2023**, *44*, 100749. [\[CrossRef\]](#)
32. Azour, F.; Boukerche, A. An Efficient Transfer and Ensemble Learning Based Computer Aided Breast Abnormality Diagnosis System. *IEEE Access* **2022**, *11*, 21199–21209. [\[CrossRef\]](#)
33. Chougrad, H.; Zouaki, H.; Alheyane, O. Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing* **2020**, *392*, 168–180. [\[CrossRef\]](#)
34. Jafari, Z.; Karami, E. Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. *Information* **2023**, *14*, 410. [\[CrossRef\]](#)
35. Lee, R.S.; Gimenez, F.; Hoogi, A.; Miyake, K.K.; Gorovoy, M.; Rubin, D.L. A curated mammography data set for use in computer-aided detection and diagnosis re- search. *Sci. Data* **2017**, *4*, 170177. [\[CrossRef\]](#)
36. Moreira, I.C.; Amaral, I.; Domingues, I.; Cardoso, A.; Cardoso, M.J.; Cardoso, J.S. INbreast: Toward a full-field digital mammographic database. *Acad. Radiol.* **2012**, *19*, 236–248. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Solla Sara, A.; Levin, E.; Fleisher, M. Accelerated Learning in Layered Neural Networks. *Complex Syst.* **1988**, *2*, 3.
38. Multi-Label Classification by Exploiting Local Positive and Negative Pairwise Label Correlation—ScienceDirect, (n.d.). Available online: <https://www.sciencedirect.com/science/article/pii/S0925231217301571> (accessed on 11 May 2018).
39. Multi-Label Learning Based on Label-Specific Features and Local Pairwise Label Correlation—ScienceDirect, (n.d.). Available online: <https://www.sciencedirect.com/science/article/pii/S0925231217313462> (accessed on 11 May 2018).
40. Lopez, M.G.; Posada, N.; Moura, D.C.; Pollán, R.R.; Valiente, J.M.F.; Ortega, C.S.; Solar, M.; Diaz-Herrero, G.; Ramos, I.M.A.P.; Loureiro, J.; et al. BCDR: A breast cancer digital repository. In Proceedings of the 15th International Conference on Experimental Mechanics, Porto, Portugal, 22–27 July 2012.
41. Suckling, J.; Parker, J.; Dance, D. The Mammographic Image Analysis Society Digital Mammogram Database. In *Excerpta Medica; International Congress Series; Excerpta Medica Foundation*: Amsterdam, Netherlands, 1994; Volume 1069, pp. 375–378.

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