



ARTICLE

## Spatial Attention Integrated EfficientNet Architecture for Breast Cancer Classification with Explainable AI

Sannasi Chakravarthy<sup>1</sup>, Bharanidharan Nagarajan<sup>2</sup>, Surbhi Bhatia Khan<sup>3,7,\*</sup>,  
Vinoth Kumar Venkatesan<sup>2</sup>, Mahesh Thyluru Ramakrishna<sup>4</sup>, Ahlam Al Musharraf<sup>5</sup> and  
Khursheed Aurungzeb<sup>6</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, 638402, India

<sup>2</sup>School of Computer Science Engineering and Information Systems (SCORE), Vellore Institute of Technology, Vellore, 632014, India

<sup>3</sup>School of Science, Engineering and Environment, University of Salford, Manchester, M54WT, UK

<sup>4</sup>Department of Computer Science & Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, 562112, India

<sup>5</sup>Department of Management, College of Business Administration, Princess Nourah Bint Abdulrahman University, P. O. Box 84428, Riyadh, 11671, Saudi Arabia

<sup>6</sup>Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, P. O. Box 51178, Riyadh, 11543, Saudi Arabia

<sup>7</sup>Adjunct Research Faculty, Centre for Research Impact & Outcome, Chitkara University, Rajpura, 140401, India

\*Corresponding Author: Surbhi Bhatia Khan. Email: s.khan138@salford.ac.uk

Received: 04 April 2024 Accepted: 11 July 2024 Published: 12 September 2024

### ABSTRACT

Breast cancer is a type of cancer responsible for higher mortality rates among women. The cruelty of breast cancer always requires a promising approach for its earlier detection. In light of this, the proposed research leverages the representation ability of pretrained EfficientNet-B0 model and the classification ability of the XGBoost model for the binary classification of breast tumors. In addition, the above transfer learning model is modified in such a way that it will focus more on tumor cells in the input mammogram. Accordingly, the work proposed an EfficientNet-B0 having a Spatial Attention Layer with XGBoost (ESA-XGBNet) for binary classification of mammograms. For this, the work is trained, tested, and validated using original and augmented mammogram images of three public datasets namely CBIS-DDSM, INbreast, and MIAS databases. Maximum classification accuracy of 97.585% (CBIS-DDSM), 98.255% (INbreast), and 98.91% (MIAS) is obtained using the proposed ESA-XGBNet architecture as compared with the existing models. Furthermore, the decision-making of the proposed ESA-XGBNet architecture is visualized and validated using the Attention Guided GradCAM-based Explainable AI technique.

### KEYWORDS

EfficientNet; mammograms; breast cancer; Explainable AI; deep-learning; transfer learning



## 1 Introduction

Breast cancer (BC) also termed as mammary cancer, is a cruel form of cancer disease significantly affects women. BC causes breast cells to have abnormal growth and so it acquires the ability to spread progressively to other organs. Due to this, breasts are highly affected causing higher malformation among women [1]. This abnormal rupture of cells is done in an uncontrolled behavior. Based on the statistical survey of GLOBOCAN, around 19 million newer cancer cases have happened in 2020. Among these, exclusively for women, BC remains the most commonly diagnosed cancer globally. And from the above statistical survey, it is revealed that BC is not only a major source of death in developed countries but also it remains as a serious life-threatening health concern in developing nations [2]. There are several factors responsible for the incidence of breast cancer; some of them are age, gender, genetic mutations, family history, hormonal factors, and lifestyle choices [3]. Digital mammograms are obtained using mammographic procedures utilizing lower-dose X-rays as a source [4,5] used for initial cancer screening. MRI representing Magnetic Resonance Imaging offers detailed images of breast tissue and so exclusively useful for evaluating the extent of cancer in certain situations [6]. The next one is the biopsy which includes the procedure of collecting a smaller sample of tissues from the breast for examination under a microscope and this is the definitive methodology for diagnosing breast cancer [7]. In this way, the proposed research intends to design a Computer-Aided Diagnosis (CAD) based on earlier detection architecture of EfficientNet-B0 having a Spatial Attention Layer with XGBoost (ESA-XGBNet) for binary classification of mammograms.

For the application areas employing medical imaging and analysis, the initial challenge is the limited accessibility of mammogram inputs since deep learning requires a substantial amount of input images for providing effective classification performance [8]. Subsequently, the next one concerns feature engineering, that is, the deep learning model provides deep feature maps with some amount of redundancy [9]. To address these issues, the research intended to design a CAD framework using the EfficientNet-B0 model, Spatial Attention Mechanism, and XGBoost classifier. Thus, the research contributions to solving the mammogram classification are as follows:

- Mammograms from three distinct standard databases are fed as input to the proposed CAD framework. For these inputs, the work proposes an innovative framework that integrates the algorithms of deep learning and machine learning, employing pre-trained architecture for feature extrication and building a robust classifier model.
- The input mammograms are noise-removed using adaptive median filtering and contrast-enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE). The pre-processed mammograms are augmented sufficiently using appropriate data augmentation.
- The EfficientNet-B0 pre-trained model is enhanced with the spatial attention mechanism layer for computing attention weights for each spatial location.
- Flatten the above resultant feature vectors and feed as input to the XGBoost classifier.
- Finally, experimentation will be done for the classification of mammogram inputs as benign or malignant cases using the ESA-XGBNet framework.

The rest of the paper is organized as follows: [Section 2](#) describes the background and literature study, [Section 3](#) elaborates on the materials and methods regarding pre-processing, augmentation, and proposed transfer learning with attention mechanism (ESA-XGBNet), [Section 4](#) describes the experimental setup, results, performance comparison with existing works and analysis, and [Section 5](#) concludes the research with future extension of the proposed work.

## 2 Literature Review

The discussion on the previous sub-section states that breast tumors are serious life-threatening diseases that distress the daily lives of women all around the world. This cancer type is the most common invasive cancer among women globally, with an estimation of 2.3 million newer incidences and 685,000 mortalities annually as per the report of the World Health Organization, 2023 [10]. Earlier detection and accurate diagnosis are crucial for improving survival rates. The design of an AI-based robust CAD framework is highly helpful in earlier identification and timely interpretation of this cancer type. The CAD frameworks utilized both machine learning and deep learning techniques for providing better solutions to the classification problem [11]. The related works corresponding to the above classification problem are discussed further.

Several literature works utilized Artificial Neural Networks (ANN) for classification problems, exclusively for breast cancer. Next to this, Support Vector Machines (SVM) models were popularly and widely used for breast tumor classification. Patel et al. [12] proposed a model that integrates ANN and hybrid optimum feature selection algorithms for breast tumor classification in the year 2021. For evaluation, the work utilized the Mammographic Image Analysis Society mammogram data after suitable enhancement procedures. The paper attained a maximum of 99% accuracy in MIAS data classification. However, the work suffered from the problem of overfitting and limited data. Kumari et al. [13] proposed an innovative way of extracting features, Advanced Gray-Level Co-occurrence Matrix (AGLCM) was developed for extricating features from mammogram image sets. For preprocessing, the researchers utilized Contrast Limited Advanced Histogram Equalization (CLAHE) for enhancing the mammogram details. Texture-based features are derived for the classification phase. For this research work, the authors employed SVM, ANN, K-Nearest Neighbor (KNN), XGBoost, and Random Forest (RF) models for the classification phase. Their experimentation provided maximum classification performance for the combination of CLAHE, AGLCM, and XGBoost algorithms with around 95% accuracy. Their research has challenges in handling redundant handcrafted features and overfitting. The author of the work [14] introduced a newer and hybrid framework for the classification problem. They employed two hybridized optimizations, Harris Hawks and Crow Search Optimization with ANN and SVM algorithms. They experimented on the mammograms corresponding to the DDSM database. For this, they achieved the results of around 97% accuracy using the above hybrid optimization combination together with the ANN algorithm. They faced challenges in metaheuristic hybridization and tuning of machine learning models.

In addition to the above study, the works of [15–17] experimented with the effectiveness of radiomics feature extraction in medical images. And they revealed that a wider range of qualitative features such as information on texture, shape, and intensity characteristics of image inputs are acquired through radiomic feature extraction. This in turn helps in extracting valuable insights from the applied image inputs. However, higher dimensional feature representations with more redundant and irrelevant vectors are generated using the radiomic feature extrication [18]. This makes the above approach as inferior as compared with the deep feature extrication using CNN models. Over the past decades, the utilization of Convolution Neural Networks (CNN) for solving biomedical problems has rapidly increased for medical imaging and analysis, including breast cancer classification [19]. Out of this, the concept of transfer learning is an emerging one and is used by several researchers for their classification problems. Exclusively, in the application of breast tumor classification, recent research studies have reported an accuracy of around 98% using pretrained models namely Visual Geometry Group-19, Residual Neural Network, and Densely Connected Convolutional Networks.

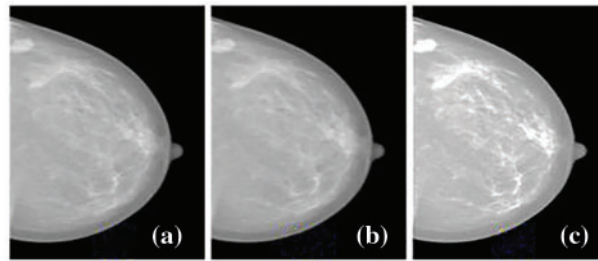
In addition to the above points, transfer learning architectures foster faster training since the architectures were already trained (pre-trained) on larger databases. And can be fine-tuned easily for deployment of breast cancer classification tasks. This represents that the number of malignant cases is lower than the number of benign cases for implementation. This can lead to biased model predictions supporting only the majority classes for classification reporting to overfitting problems. For tackling this problem, researchers have proposed several strategies, namely class-weighted loss functions, data augmentation techniques, and attention mechanisms. In detail, rotation, flipping, scaling, and cropping of image augmentation will increase the amount of minority class data (e.g., malignant class images) artificially. Thus, the data augmentation leads to the generation of newer synthetic samples from the available inputs. This facilitates the classification model to have good exposure to a more diversified range of input data. And reveals that the generalization ability of the classification model is improved thereby the risk of overfitting will be reduced exclusively in the presence of class imbalance. On the other hand, the dynamic focus on informative regions on the image inputs will be increased through the use of attention mechanisms. For the employed problem of breast cancer classification, the prioritizing of crucial or malignancy regions in the mammogram inputs is done using the attention modules. This prioritization of regions is done through the assignment of adaptive weights for different regions based on their severity presence. And thus the computation and assignment of adaptive weights using attention mechanisms will support in overcoming the issue of class imbalance. In this way, the approaches of data augmentation and attention mechanism will support the research effectively for mitigating the effects of class imbalance. Thus, the proposed work makes use of the above implementation for extensive enhancement of the overall performance of the classification framework.

### 3 Materials and Methods

The paper utilized input mammograms adopted from three distinct benchmark datasets: CBIS-DDSM (Curated Breast Imaging Subset of DDSM) [16], INbreast [17], and MIAS (Mammographic Imaging Analysis Society) [18] datasets. The first dataset is a larger dataset (starting from  $2000 \times 2000$  resolution) used for several breast cancer research problems. The second one, INbreast contains higher-quality mammographic information acquired and accumulated as Full-Field Digital Mammographic (FFDM) data with resolution of  $3000 \times 4000$ . The format of mammograms available in this dataset is DICOM. The third database, MIAS, is a popular and widely used mammographic dataset with a resolution of  $1024 \times 1024$ . In this way, the work utilized CBIS-DDSM (large and diverse), INbreast (high-quality mammograms), and MIAS (focusses more on microcalcification) datasets for the evaluation of the proposed work.

#### 3.1 Pre-Processing of Digital Mammograms

To perform experimentation using digital mammograms for earlier breast cancer detection, the mammograms are preprocessed before classification. The impulse noise in the mammograms is removed using adaptive median filtering approaches where the noise-distorted pixels are processed without disturbing the noise-free pixels. This filtering approach is employed to remove any noise spikes effectively and smooth out the mammograms while preserving the crucial image features. Afterward, the Contrast Limited Adaptive Histogram Equalization (CLAHE) approach is used for enhancing the mammogram contrast adaptively. CLAHE adapts the conventional histogram equalization technique but limits the contrast amplification in local regions. Herein, the parameters of `tileGridSize` and `clipLimit` are fine-tuned as  $4 \times 4$  and 2. The resultant sample preprocessed mammogram image taken from the INbreast dataset using the aforementioned procedure is illustrated in [Fig. 1](#).



**Figure 1:** Mammogram preprocessing (a) original image from inbreast dataset (b) adaptive median filtered output (c) CLAHE output

### 3.2 Augmentation of Mammogram Inputs

The preprocessed mammograms are partitioned and stratified with a ratio of 70:30 as training and testing data together with a cross-validation partition of ten. The data partitioning before augmentation ensures that the model's generalization ability is assessed accurately. To generate multiple amounts of mammograms with the existing images present in CBIS-DDSM, INbreast, and MIAS databases, the research makes use of augmentation. Here, the mammograms are processed with two operations involving six rotations with different degrees ( $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $234^\circ$ , and  $270^\circ$ ) and possible flipping (vertical and horizontal) of mammogram images. This facilitates the classification model to have good exposure to a more diversified range of input data. And so the generalization ability of the classification model is improved thereby the risk of overfitting will be reduced exclusively in the presence of class imbalance. The mammogram images present in the three original databases and how many mammogram images are augmented on the training data are comparatively summarized in Table 1. Here, it is revealed that the data augmentation procedure is not applied to the testing data in order to avoid data leakage and robust evaluation.

**Table 1:** Mammogram augmentation on the training data of three datasets

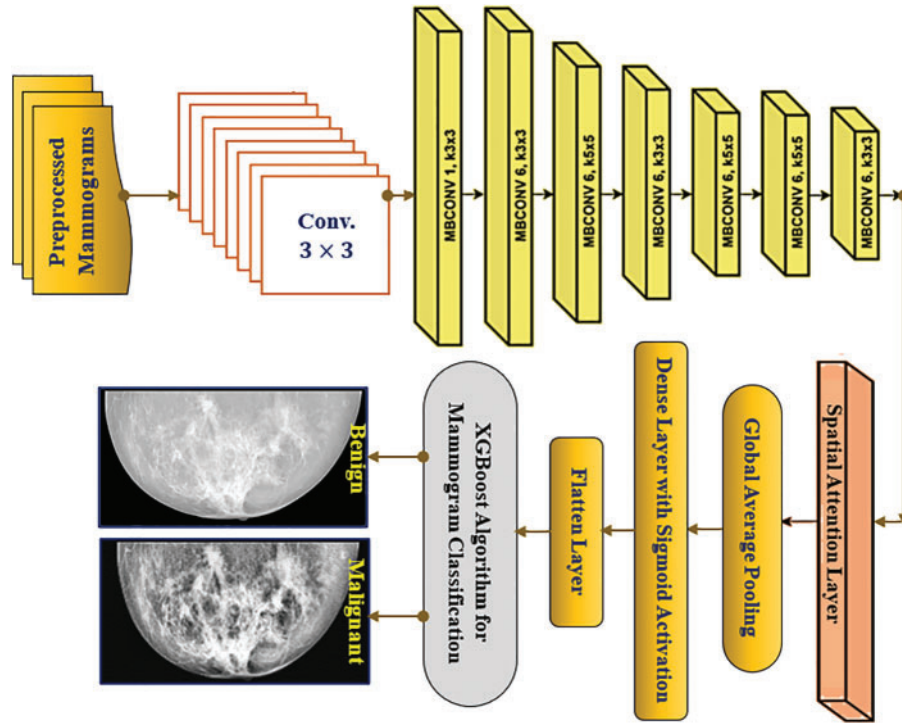
| Output targets | CBIS-DDSM dataset    |                       |                      | INbreast dataset     |                       |                      | MIAS dataset         |                       |                      |
|----------------|----------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|
|                | Original mammo grams | Training data         |                      | Original mammo grams | Training data         |                      | Original mammo grams | Training data         |                      |
|                |                      | Before augmen- tation | After augmen- tation |                      | Before augmen- tation | After augmen- tation |                      | Before augmen- tation | After augmen- tation |
| Benign         | 551                  | 386                   | 3088                 | 76                   | 53                    | 424                  | 56                   | 39                    | 312                  |
| Malignant      | 636                  | 445                   | 3560                 | 72                   | 50                    | 400                  | 51                   | 36                    | 288                  |

### 3.3 Proposed EfficientNet-B0 Integrated with Spatial Attention (ESA-XGBNet)

The research community is attempting to build several efficient CNN architectures through the balanced enhancement in width, depth, and resolution. Following this, the EfficientNet architecture provides better model performance with fewer parameters compared to conventional scaling mechanisms.



The architecture of the proposed ESA-XGBNet model is illustrated graphically in Fig. 2. The architecture of the proposed ESA-XGBNet model, depicted in Fig. 2, aims to implement EfficientNet-B0 augmented with a Spatial Attention Layer and an XGBoost model for the binary classification of mammograms, contributing to the early detection of breast tumors. The model begins with a stem convolution block responsible for initial mammogram processing, comprising convolution layers, normalization, and activation functions. Subsequent blocks consist of MBConv blocks, which are extensions of Mobile Inverted Residual Bottleneck blocks. These blocks integrate inverted residuals, linear bottlenecks, and depthwise separable convolutions, with the latter decomposing standard convolution operations into depthwise and pointwise convolutions to reduce computational costs. Additionally, the architecture employs a compound scaling approach that combines scaling factors of resolution, width, and depth in a principled manner to achieve effective and efficient model scaling.



**Figure 2:** Proposed ESA-XGBNet architecture for breast cancer problem

After MBConv blocks, a spatial attention layer is added to the architecture in order to take the generated feature maps from the network. After taking this, the spatial attention mechanism computes attention weights for each spatial location of the applied mammograms. For a feature map  $F$  representing dimensions of  $H \times W \times C$ , the global average pooling for getting a global feature vector ( $g_i$ ) is mathematically illustrated in Eq. (1).

$$g_i = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W F_{h,w,i} \quad (1)$$

Now, the attention weights will be obtained by passing the above global feature vector through a small dense layer ( $z_i$ ) with a sigmoid activation function ( $\sigma$ ). This can be written mathematically

considering  $W_{attn}$  as the weight parameters of the attention layer as in Eq. (2).

$$z_i = \sigma(W_{attn} \cdot g_i) \quad (2)$$

Now, spatial attention map ( $S_{h,w,i}$ ) is created by scaling the original feature map ( $F_{h,w,i}$ ) with the attention weights. This can be mathematically defined in Eq. (3).

$$S_{h,w,i} = z_i \cdot F_{h,w,i} \quad (3)$$

Afterward, the original feature map ( $F_{h,w,i}$ ) and the attention-enhanced feature map ( $S_{h,w,i}$ ) are summed for obtaining the final feature representation ( $O_{h,w,i}$ ) as given in Eq. (4).

$$O_{h,w,i} = F_{h,w,i} + S_{h,w,i} \quad (4)$$

The Softmax function, as described in Eq. (2), is utilized to determine learnable weights. This process involves multiplying the original feature maps by these calculated attention weights, as outlined in Eq. (3). Consequently, highly informative regions are emphasized while less informative regions are suppressed in the preprocessed mammogram images. This operation yields attention-weighted feature maps enriched with crucial information for breast tumor classification. Following this, as depicted in Fig. 2, a Global Average Pooling (GAP) layer is introduced adjacent to the attention-weighted feature map elements. This layer is responsible for reducing spatial dimensions, generating a single value per feature map channel, thereby extracting the most relevant information for each channel. Lastly, a flatten layer is incorporated, which takes the output of GAP as inputs for the XGBoost classification model. Consequently, the attention-driven, flattened feature vectors serve as representations of the applied mammogram image inputs, enhanced through the spatial attention mechanism.

### 3.4 XGBoost for Tumor Classification

XGBoost (Extreme Gradient Boosting) is an ensemble learning algorithm popularly used for solving several classification tasks. The algorithm combines the predictions from multiple weak learners (decision trees) to create stronger predictive decisions. Let  $X_{train}$  and  $X_{test}$  represent the mammogram-extracted features from EfficientNet integrated with spatial attention mechanisms. Here,  $X_{train}$  and  $X_{test}$  are the feature sets that correspond to training and testing data. Similarly,  $y_{train}$  and  $y_{test}$  are the output labels of the training and testing dataset. The XGBoost algorithm is an ensemble combination of decision trees, and the sequential addition of trees will be carried out in the training phases for reducing the residual errors. Specifically, each tree will be added with the intention of minimizing the overall objective function, which can be mathematically represented in Eq. (5).

$$Objective(XGBoost) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^k \Omega(f_k) \quad (5)$$

In Eq. (5),  $L$  represents the loss function,  $\hat{y}_i$  denotes the predicted label, and  $\Omega(f_k)$  is the regularization term for each tree. Here,  $\hat{y}_i$  can be mathematically given in Eq. (6).

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i) \quad (6)$$

In Eq. (6), the term  $f_k(x_i)$  represents the  $k^{th}$  tree prediction for the  $i^{th}$  sample. As given in Eq. (5), the objective function supports the training phase for minimizing errors and controlling the model complexity. The outline of the algorithm for mammogram classification using the proposed ESA-XGBoost architecture is given below.

---

**Algorithm 1:** Algorithm of the proposed ESA-XGBNet for mammogram classification
 

---

**Step 1:** *Loading and Preprocessing of Mammograms*

- Loading mammogram images with labels from CBIS-DDSM, INBreast, and MIAS datasets.
- Preprocessing of mammogram images including data augmentation.

**Step 2:** *Dataset Partitioning*

- Partition of preprocessed mammogram data into training and testing sets.

**Step 3:** *Loading of pretrained transfer learning model*

- Load an EfficientNet-B0 architecture with weights pretrained on the larger ImageNet database.

**Step 4:** *Inclusion of Spatial Attention Layer*

- Adding of spatial attention layer after the last layer of the EfficientNet-B0 architecture.
- Applying global average pooling for getting a global feature vector as given below

$$g_i = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W F_{h,w,i} \quad (\text{Eq. (1)})$$

- Now, the attention weights will be obtained by passing the above global feature vector through a small dense layer ( $z_i$ ) with a sigmoid activation function as depicted below.

$$z_i = \sigma(W_{att} \cdot g_i) \quad (\text{Eq. (2)})$$

- Creation of a spatial attention map ( $S_{h,w,i}$ ) by scaling the original feature map ( $F_{h,w,i}$ ) with the attention weights as depicted below.  $S_{h,w,i} = z_i \cdot F_{h,w,i}$  (Eq. (3))

**Step 5:** *Combining Original and Attention-Enhanced Features*

- Summing of the original feature map ( $F_{h,w,i}$ ) and the attention-enhanced feature map ( $S_{h,w,i}$ ).
- Obtaining the final feature representation ( $O_{h,w,i}$ ) as depicted below.  $O_{h,w,i} = F_{h,w,i} + S_{h,w,i}$  (Eq. (4))

**Step 6:** *Flatten the feature vectors for XGBoost inputs.***Step 7:** *Training of XGBoost Algorithm as illustrated in Eqs. (5) and (6).***Step 8:** *Make predictions using the test data***Step 9:** *Evaluation and Performance Analysis***Step 10:** *Interpretability using Explainable Artificial Intelligence (XAI)*


---

## 4 Experimental Results and Discussion

The experimentation and its outcomes attained using the proposed ESA-XGBNet architecture evaluated with three distinct standard mammographic databases will be discussed in this section. The proposed CAD framework is evaluated using the mammograms of CBIS-DDSM, INbreast, and MIAS datasets after preprocessing. The mammographic images after CLAHE-processed is partitioned and stratified with a ratio of 70:30 as training and testing data together with a cross-validation partition of ten. The data partitioning before augmentation ensures that the model's generalization ability is assessed accurately. Thus, the testing set should remain unchanged to accurately reflect real-world scenarios where the model encounters new, unaltered data. The adopted EfficientNet-B0 architecture takes the input mammograms of size  $224 \times 224$ . The weights that are trained using the ImageNet database are passed to the model for further learning on input mammograms to capture deep features. And the architecture includes the fine-tuning of hyperparameters with a batch size of 32, a learning rate of 0.004 with ten epochs, loss function as binary cross-entropy, and the stochastic gradient descent as optimizer. For the XGBoost algorithm, the number of estimators is fine-tuned as 100, and the maximum depth as 3. The ten-fold cross-validated results are obtained for the experimentation. Performance metrics such as Specificity (Sp), Sensitivity (Sen), Accuracy (Acc), F1 score, Precision (Pr) [19–21], and Kappa [5] are employed for comparative analysis. An ablation study refers to a systematic experimental investigation used for understanding the individual contributions



of different modules toward the overall performance. Six ablation tests are conducted initially to validate the effectiveness of the proposed CAD model, examining the impact of EfficientNet-B0 architecture, spatial attention mechanism, Decision Tree algorithm, and XGBoost model on the binary classification problem. The outcomes of these ablation tests for mammograms from the CBIS-DDSM database are summarized in Table 2. Herein, Test\_1 indicates that the mammogram classification is performed only using the existing EfficientNet-B0 model, Test\_2 and Test\_3 denote that the mammogram classification is performed using the Decision Tree (DT) and XGBoost (XGB) algorithm applied with the features extracted using the EfficientNet-B0 model respectively. Test\_4 indicates that the mammogram classification is performed using the EfficientNet-B0 architecture integrated with the Spatial Attention mechanism. Test\_5 and Test\_6 indicate that the mammogram classification is done using the combination of EfficientNet-B0+Spatial Attention extracted features with DT and XGB algorithms, respectively. From this experimentation, it is revealed that the mammogram classification using the XGBoost algorithm applied with the extracted features of the EfficientNet-B0+Spatial Attention mechanism contributes more towards the improvement of overall classification performance.

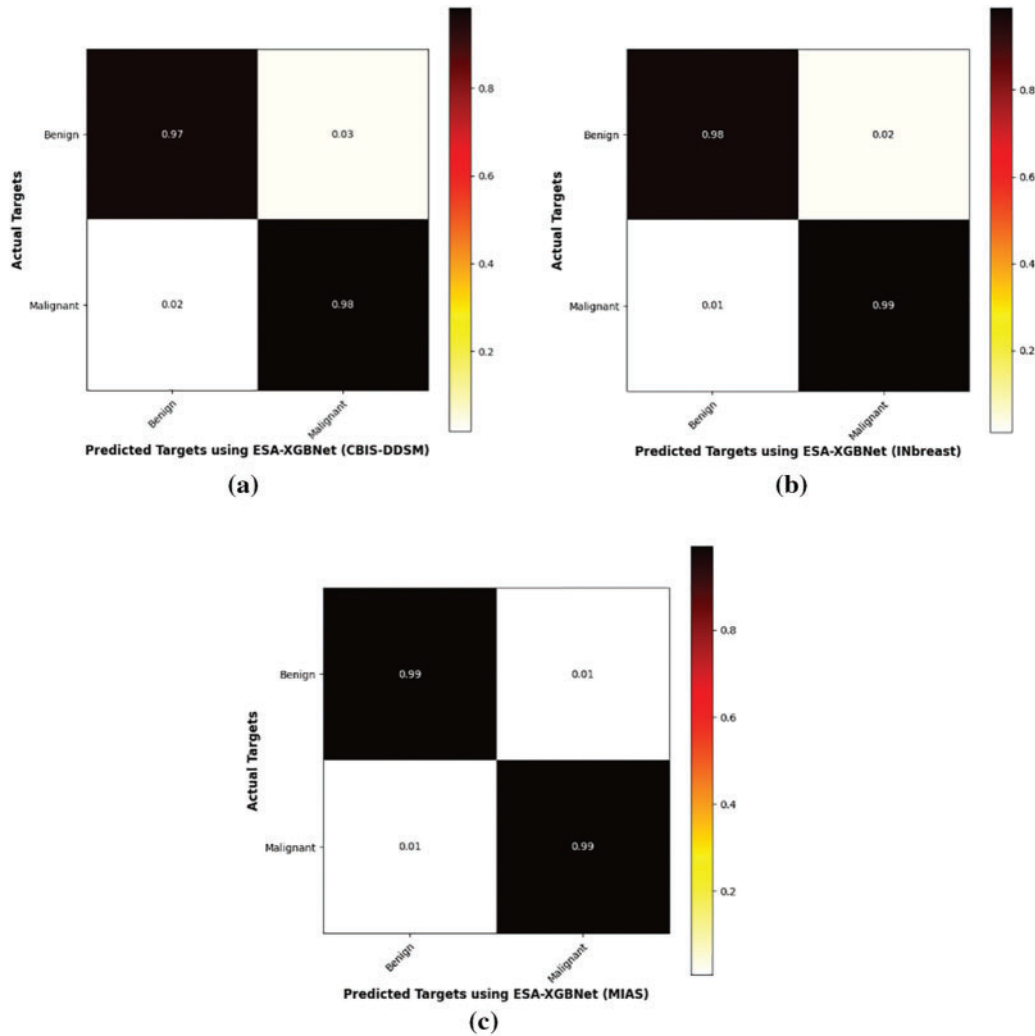
**Table 2:** Ablation study to validate the effect of the modules used in the proposed framework with CBIS-DDSM database

| Modules                     | Test_1 | Test_2 | Test_3 | Test_4 | Test_5 | Test_6 |
|-----------------------------|--------|--------|--------|--------|--------|--------|
| EfficientNet-B0 model       | ✓      | ✓      | ✓      | ✓      | ✓      | ✓      |
| Spatial attention mechanism | ×      | ×      | ×      | ✓      | ✓      | ✓      |
| Decision tree algorithm     | ×      | ✓      | ×      | ×      | ✓      | ×      |
| XGBoost algorithm           | ×      | ×      | ✓      | ×      | ×      | ✓      |
| Accuracy (%)                | 83.614 | 86.493 | 87.265 | 88.493 | 94.110 | 97.585 |

The comparative analysis of the proposed architecture involves experimenting with various models using the EfficientNet-B0 backbone and different classifiers. Initially, a simple EfficientNet-B0 model is used for end-to-end classification. Then, the EfficientNet-B0 model is employed for feature extraction, followed by classification using Naïve Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), and Decision Tree (DT) classifiers. The experimentation proceeds by integrating the EfficientNet-B0 model with the Spatial Attention (ESA) mechanism, resulting in architectures such as ESA-NB, ESA-KNN, ESA-SVM, ESA-DT, and ESA-XGBNet models. The confusion matrix demonstrating the test results obtained for ESA-XGBNet using three mammogram datasets is illustrated in Fig. 3.

Tables 3–5 present the performance of various CNN architectures and the proposed ESA-XGBNet model applied to mammograms from CBIS-DDSM, INbreast, and MIAS databases. EfficientNet-B0 demonstrates competent classification performance, with accuracies ranging from 83.614% to 87.227% and average precision and F1 scores around 84.5% and 85.6%. Incorporating deep features from EfficientNet-B0 into different machine learning models improves classification results significantly. Support Vector Machines (SVM) yield robust accuracies ranging from 86.317% to 90.732%, with precision and F1 scores in the 87–90% range, owing to their effectiveness in capturing complex relationships among feature vectors. Decision Trees (DT) and XGBoost classifiers also demonstrate enhanced performance, with accuracies ranging from 86.493% to 93.069% for XGBoost. XGBoost particularly excels due to its ensemble learning, distributed computing, and combining

multiple weak-learner predictions, resulting in improved sensitivity, specificity, precision, and F1 score compared to other models.



**Figure 3:** Confusion matrix test results obtained for ESA-XGBNet using three mammogram datasets (a) CBIS-DDSM (b) INbreast (c) MIAS datasets

**Table 3:** Classification performance of the proposed ESA-XGBNet for CBIS-DDSM database

| CNN architectures   | Sen (%) | Spc (%) | Acc (%) | Prc (%) | F1 score (%) |
|---------------------|---------|---------|---------|---------|--------------|
| EfficientNet-B0     | 84.91   | 82.12   | 83.614  | 84.57   | 84.74        |
| EfficientNet-B0-NB  | 84.97   | 84.47   | 84.737  | 86.33   | 85.64        |
| EfficientNet-B0-KNN | 85.23   | 85.07   | 85.159  | 86.83   | 86.02        |
| EfficientNet-B0-SVM | 86.67   | 85.90   | 86.317  | 87.65   | 87.16        |
| EfficientNet-B0-DT  | 86.81   | 86.13   | 86.493  | 87.84   | 87.32        |

(Continued)

**Table 3 (continued)**

| CNN architectures                                   | Sen (%)      | Spc (%)      | Acc (%)       | Prc (%)      | F1 score (%) |
|---|--------------|--------------|---------------|--------------|--------------|
| EfficientNet-B0-XGB                                 | 87.53        | 86.96        | 87.265        | 88.57        | 88.05        |
| EfficientNet with spatial attention mechanism (ESA) | 88.77        | 88.17        | 88.493        | 89.65        | 89.21        |
| ESA-NB  | 89.03        | 88.93        | 88.985        | 90.28        | 89.65        |
| ESA-KNN   | 91.19        | 90.29        | 90.775        | 91.55        | 91.37        |
| ESA-SVM   | 92.64        | 92.33        | 92.495        | 93.31        | 92.97        |
| ESA-DT  | 94.01        | 94.22        | 94.110        | 94.95        | 94.48        |
| ESA-XGBNet  | <b>97.03</b> | <b>98.21</b> | <b>97.585</b> | <b>98.44</b> | <b>97.73</b> |

**Table 4:** Classification performance of the proposed ESA-XGBNet for INbreast database

| CNN architectures                                   | Sen (%)      | Spc (%)      | Acc (%)       | Prc (%)      | F1 score (%) |
|---|--------------|--------------|---------------|--------------|--------------|
| EfficientNet-B0                                     | 85.07        | 85.53        | 85.304        | 84.78        | 84.92        |
| EfficientNet-B0-NB                                  | 86.81        | 86.62        | 86.712        | 86.01        | 86.41        |
| EfficientNet-B0-KNN                                 | 87.38        | 87.72        | 87.556        | 87.08        | 87.23        |
| EfficientNet-B0-SVM                                 | 88.54        | 89.36        | 88.964        | 88.75        | 88.64        |
| EfficientNet-B0-DT                                  | 89.70        | 90.46        | 90.090        | 89.91        | 89.80        |
| EfficientNet-B0-XGB                                 | 90.86        | 91.56        | 91.216        | 91.07        | 90.96        |
| EfficientNet with spatial attention mechanism (ESA) | 92.59        | 92.65        | 92.624        | 92.27        | 92.43        |
| ESA-NB  | 93.17        | 93.75        | 93.468        | 93.39        | 93.28        |
| ESA-KNN   | 94.33        | 94.85        | 94.595        | 94.55        | 94.44        |
| ESA-SVM   | 95.49        | 95.94        | 95.721        | 95.71        | 95.60        |
| ESA-DT  | 96.64        | 97.04        | 96.847        | 96.87        | 96.76        |
| ESA-XGBNet  | <b>97.80</b> | <b>98.68</b> | <b>98.255</b> | <b>98.60</b> | <b>98.21</b> |

**Table 5:** Classification performance of the proposed ESA-XGBNet for MIAS database

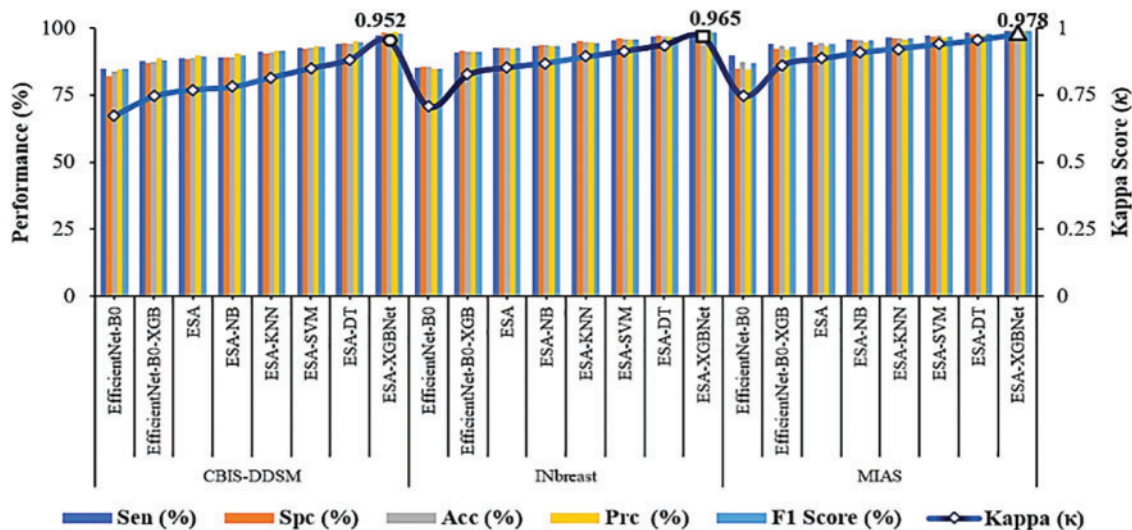
| CNN architectures                                   | Sen (%) | Spc (%) | Acc (%) | Prc (%) | F1 score (%) |
|---|---------|---------|---------|---------|--------------|
| EfficientNet-B0                                     | 89.87   | 84.82   | 87.227  | 84.36   | 87.03        |
| EfficientNet-B0-NB                                  | 90.69   | 86.31   | 88.396  | 85.78   | 88.17        |
| EfficientNet-B0-KNN                                 | 91.50   | 87.80   | 89.564  | 87.23   | 89.31        |
| EfficientNet-B0-SVM                                 | 92.32   | 89.29   | 90.732  | 88.70   | 90.47        |
| EfficientNet-B0-DT                                  | 93.14   | 90.77   | 91.900  | 90.19   | 91.64        |
| EfficientNet-B0-XGB                                 | 93.95   | 92.26   | 93.069  | 91.71   | 92.82        |
| EfficientNet with spatial attention mechanism (ESA) | 94.77   | 93.75   | 94.237  | 93.25   | 94.00        |

(Continued)

**Table 5 (continued)**

| CNN architectures | Sen (%)      | Spc (%)      | Acc (%)       | Prc (%)      | F1 score (%) |
|-------------------|--------------|--------------|---------------|--------------|--------------|
| ESA-NB            | 95.59        | 95.24        | 95.405        | 94.81        | 95.20        |
| ESA-KNN           | 96.41        | 95.98        | 96.184        | 95.62        | 96.01        |
| ESA-SVM           | 97.22        | 96.73        | 96.963        | 96.43        | 96.80        |
| ESA-DT            | 98.04        | 97.47        | 97.741        | 97.24        | 97.66        |
| ESA-XGBNet        | <b>98.86</b> | <b>98.93</b> | <b>98.910</b> | <b>98.86</b> | <b>98.79</b> |

As illustrated in Fig. 2, the further direction of research is the inclusion of a spatial attention mechanism to the aforementioned experimental setup to enhance the breast tumor classification. Thus, the performance of the EfficientNet-B0 with Spatial Attention Mechanism (ESA) model is found to be more competent than the standalone architecture as given in Tables 3–5. In this way, the classification accuracies of 88.493% (CBIS-DDSM), 92.624% (INbreast), and 94.237% (MIAS) are obtained for breast tumor classification. The supreme classification performance of accuracies –97.585% (CBIS-DDSM), 98.255% (INbreast), and 98.91% (MIAS), precision scores –98.44% (CBIS-DDSM), 98.60% (INbreast), and 98.86% (MIAS), and F1 scores of 97.73% (CBIS-DDSM), 98.21% (INbreast), and 98.79% (MIAS), respectively. All the above-attained results are further validated using the Statistical Kappa ( $\kappa$ ) validation. The overall performance comparison and validation of attained results of the proposed ESA-XGBNet model are graphically illustrated in Fig. 4. From this graphical comparison, it is evident that the proposed ESA-XGBNet architecture outperforms others with the kappa validation of 0.952 (CBIS-DDSM), 0.965 (INbreast), and 0.978 (MIAS), respectively.

**Figure 4:** Performance comparison and validation of attained results–ESA-XGBNet

Even though a classification model discriminates well in identifying abnormalities, it provides a trade-off between sensitivity and specificity values Fig. 5 illustrates a plot between True Positive Rate (TPR) and False Positive Rate (FPR) for the employed classification architectures applied with three datasets. Thus, the plot of Fig. 5 illustrates how well a classification model discriminated between

benign and malignant cases. Accordingly, a robust classification model will have a curve that rises steeply toward the top-right corner of the plot. In Fig. 5, the markers in each performance curve refer to the twelve employed CNN models as listed in the order of Tables 3–5 for three datasets. Accordingly, the performance curve of the proposed ESA-XGBNet steps towards the top-right for all the employed databases.

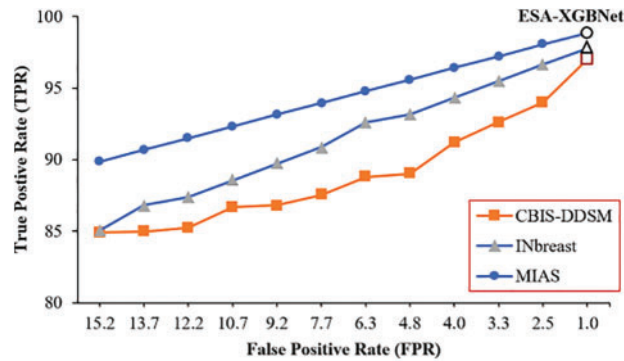


Figure 5: TPR vs. FPR plot for the proposed models applied for three datasets

Grad-CAM is a common procedure to visualize where a convolutional neural network architecture is looking for target prediction. Sample mammogram (INbreast and CBIS-DDSM) visualization are given in Fig. 6 which interprets where our proposed model focused for its target prediction. Here, sample benign and malignant labeled mammograms (Fig. 6a1,a2) from the INbreast and CBIS-DDSM databases are taken. The superimposed visualizations (Fig. 6b1,b2) and their corresponding heat maps (Fig. 6c1,c2) are generated with the correct prediction labels of Benign and Malignant.

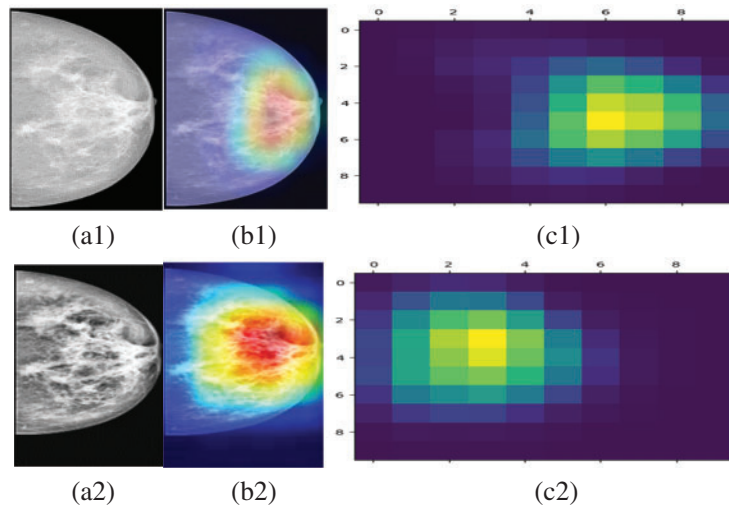


Figure 6: (a1) Preprocessed mammogram (INbreast) with a label of Benign (b1) respective superimposed visualization (c1) attention guided Grad-CAM heat maps with the correct prediction of benign label (a2) preprocessed mammogram (CBIS-DDSM) with a label of malignant (b2) respective superimposed visualization (c2) attention guided grad-CAM heat maps with the correct prediction of malignant label

As a final point of research, the comparison of existing and recent works against the proposed ESA-XGBNet architecture is summarized in Table 6. Herein, the research works were applied with diversified and different mammogram databases. And the robust classification performance of breast tumors is obtained apparently using the proposed CAD framework as depicted in Fig. 2. However, the work has the potential limitation of added complexity due to the inclusion of a spatial attention layer with EfficientNet-B0 architecture and this will be taken care in our future extension.

**Table 6:** Performance comparison with the recent works

| Literature works       | Mammogram data  | Classification accuracies |
|------------------------|-----------------|---------------------------|
| Falconi et al. [19]    | INbreast        | 90.9%                     |
| Ragab et al. [20]      | DDSM            | 79%                       |
|                        | CBIS-DDSM       | 87.2%                     |
| Baccouche et al. [21]  | CBIS-DDSM       | 85.38%                    |
|                        | Private Dataset | 96.08%                    |
| Muduli et al. [22]     | MIAS            | 96.55%                    |
|                        | DDSM            | 90.68 %                   |
|                        | INbreast        | 91.28%                    |
| Surendiran et al. [23] | DDSM            | 93.3%                     |
| Li et al. [24]         | DDSM            | 94.7%                     |
| Yang et al. [25]       | MIAS            | 95.3%                     |
|                        | INbreast        | 96.52%                    |
|                        | DDSM            | 91.2%                     |
| Proposed ESA-XGBNet    | CBIS-DDSM       | 97.585%                   |
|                        | INberast        | 98.255%                   |
|                        | MIAS            | 98.910%                   |

## 5 Conclusion and Future Direction

The research intended to propose a robust early-detection CAD framework for a worldwide societal disease, breast cancer. The input considers diversified mammogram images adopted from three distinct datasets: CBIS-DDSM, INbreast, and MIAS databases. The methodology involves substantial experimentation beginning with the choice of dataset and ending with a detailed performance analysis. Initially, the research preprocessed the input mammograms for their better representation in balancing contrast and exposure. After noise removal using an adaptive median filter and mammogram enhancement using the Contrast Limited Adaptive Histogram Equalization approach, the preprocessed mammogram images are then fed into the EfficientNet-B0 architecture. Here, several experiments have been conducted with the intention of obtaining robust performance for breast tumor classification. In addition to this, five machine learning models namely Naïve Bayes, K-Nearest Neighbor, Support Vector Machines, Decision Trees, and XGBoost algorithms are adopted exclusively for the classification stage. With the aim of improving the classification performance, a spatial attention layer is included in the EfficientNet-Bo model and so the weighted attention deep feature vectors are derived for ML-based classification. In this way, the robust classification performance of 97.585% (CBIS-DDSM), 98.255% (INbreast), and 98.910% (MIAS) accuracy scores are obtained



using the proposed ESA-XGBNet architecture as compared with the existing models. Furthermore, the decision-making of the proposed ESA-XGBNet architecture is interpreted and validated using the kappa metric and Attention Guided GradCAM-based Explainable AI technique. The future direction of the proposed architecture will include the segmentation of tumors and utilizing breast ultrasound (BUS) clinical imageries with U-Net-based models to improve the robustness of breast cancer diagnosis. The preprocessing steps that we employed in our study have necessitated the exclusion of images that did not meet the required standards such as image quality, and annotations. In addition, demanding high computational requirements of training deep learning models, the study opted for a subset of the dataset that allowed us to conduct thorough experiments within a reasonable timeframe and resource allocation. This subset was sufficient to demonstrate the effectiveness of our proposed model. Furthermore, while the initial number of images was 1187 (CBIS-DDSM dataset), data augmentation significantly increased the effective size of our training set. This augmentation provided the model with diverse examples, enhancing its generalization capabilities without compromising the integrity of the test set. However, an extensive experimentation will be carried out using all the images of CBIS-DDSM dataset. And investigating the complexity of the proposed work with different attention mechanisms applied for distinct medical problems.

**Acknowledgement:** This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Funding Statement:** This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Sannasi Chakravarthy and Bharanidharan Nagarajan; data collection: Surbhi Bhatia Khan and Ahlam Al Musharraf; methodology: Surbhi Bhatia Khan and Ahlam Al Musharraf; analysis and interpretation of results: Vinoth Kumar Venkatesan and Khursheed Aurungzeb; draft manuscript preparation: Mahesh Thyluru Ramakrishna. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data that support the findings of this study are openly available at: <https://www.cancerimagingarchive.net/collection/cbis-ddsm/>, <http://peipa.essex.ac.uk/info/mias.html>, <https://paperswithcode.com/dataset/inbreast> (accessed on 18 June 2024).

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] F. Ye *et al.*, “Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer,” *Mol. Cancer*, vol. 22, no. 1, 2023, Art. no. 105. doi: [10.1186/s12943-023-01805-y](https://doi.org/10.1186/s12943-023-01805-y).
- [2] M. Arnold *et al.*, “Current and future burden of breast cancer: Global statistics for 2020 and 2040,” *The Breast*, vol. 66, pp. 15–23, 2022.

- [3] L. Wilkinson and T. Gathani, "Understanding breast cancer as a global health concern," *Br. J. Radiol.*, vol. 95, no. 1130, 2022, Art. no. 20211033. doi: [10.1259/bjr.20211033](https://doi.org/10.1259/bjr.20211033).
- [4] H. Hussein *et al.*, "Supplemental breast cancer screening in women with dense breasts and negative mammography: A systematic review and meta-analysis," *Radiology*, vol. 306, no. 3, 2023, Art. no. e221785.
- [5] S. R. Sannasi Chakravarthy and H. Rajaguru, "SKMAT-U-Net architecture for breast mass segmentation," *Int. J. Imaging Syst. Technol.*, vol. 32, no. 6, pp. 1880–1888, 2022. doi: [10.1002/ima.22781](https://doi.org/10.1002/ima.22781).
- [6] S. R. Sannasi Chakravarthy and H. Rajaguru, "A systematic review on screening, examining and classification of breast cancer," in *2021 Smart Technol., Commun. Robot. (STCR)*, Sathyamangalam, India, 2021, pp. 1–4. doi: [10.1109/STCR51658.2021.9588828](https://doi.org/10.1109/STCR51658.2021.9588828).
- [7] A. J. A. D. Freitas *et al.*, "Liquid biopsy as a tool for the diagnosis, treatment, and monitoring of breast cancer," *Int. J. Mol. Sci.*, vol. 23, no. 17, 2022, Art. no. 9952. doi: [10.3390/ijms23179952](https://doi.org/10.3390/ijms23179952).
- [8] M. Al Moteri, T. R. Mahesh, A. Thakur, V. Vinoth Kumar, S. B. Khan and M. Alojail, "Enhancing accessibility for improved diagnosis with modified EfficientNetV2-S and cyclic learning rate strategy in women with disabilities and breast cancer," *Front. Med.*, vol. 11, Mar. 2024. doi: [10.3389/fmed.2024.1373244](https://doi.org/10.3389/fmed.2024.1373244).
- [9] Z. Jiang *et al.*, "Identification of novel cuproptosis-related lncRNA signatures to predict the prognosis and immune microenvironment of breast cancer patients," *Front. Oncol.*, vol. 12, 2022. doi: [10.3389/fonc.2022.988680](https://doi.org/10.3389/fonc.2022.988680).
- [10] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA Cancer J. Clin.*, vol. 73, no. 1, pp. 17–48, 2023. doi: [10.3322/caac.21763](https://doi.org/10.3322/caac.21763).
- [11] S. Arooj *et al.*, "Breast cancer detection and classification empowered with transfer learning," *Front. Public Health*, vol. 10, 2022, Art. no. 924432. doi: [10.3389/fpubh.2022.924432](https://doi.org/10.3389/fpubh.2022.924432).
- [12] J. J. Patel and S. K. Hadia, "An enhancement of mammogram images for breast cancer classification using artificial neural networks," *IAES Int. J. Artif. Intell.*, vol. 10, no. 2, 2021, Art. no. 332. doi: [10.11591/ijai.v10.i2.pp332-345](https://doi.org/10.11591/ijai.v10.i2.pp332-345).
- [13] L. K. Kumari and B. N. Jagadesh, "A robust feature extraction technique for breast cancer detection using digital mammograms based on advanced GLCM approach," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 8, no. 30, 2022, Art. no. e3. doi: [10.4108/eai.11-1-2022.172813](https://doi.org/10.4108/eai.11-1-2022.172813).
- [14] S. Thawkar, "Feature selection and classification in mammography using hybrid crow search algorithm with Harris hawks optimization," *Biocybernet. Biomed. Eng.*, vol. 42, no. 4, pp. 1094–1111, 2022. doi: [10.1016/j.bbe.2022.09.001](https://doi.org/10.1016/j.bbe.2022.09.001).
- [15] P. Rajpurkar *et al.*, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*.
- [16] Lee *et al.*, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, no. 1, pp. 1–9, 2017.
- [17] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso and J. S. Cardoso, "Inbreast: Toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, 2012. doi: [10.1016/j.acra.2011.09.014](https://doi.org/10.1016/j.acra.2011.09.014).
- [18] J. Suckling, J. Parker, D. Dance, S. Astley, and I. Hutt, "Mammographic image analysis society (MIAS) database v1. 21, (2015)," Accessed: Mar. 28, 2021. [Online]. Available: <https://www.repository.cam.ac.uk/handle/1810/250394>
- [19] L. Falconí, M. Pérez, W. Aguilar, and A. Conci, "Transfer learning and fine tuning in mammogram bi-rads classification," in *2020 IEEE 33rd Int. Symp. on Comput. Based Med. Syst. (CBMS)*, Rochester, MN, USA, 2020, pp. 475–480. doi: [10.1109/CBMS49503.2020.00096](https://doi.org/10.1109/CBMS49503.2020.00096).
- [20] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *Bioinformatics Genom.*, vol. 7, no. 5, 2019, Art. no. e6201. doi: [10.7717/peerj.6201](https://doi.org/10.7717/peerj.6201).
- [21] A. Baccouche, B. Garcia-Zapirain, and A. S. Elmaghraby, "An integrated framework for breast mass classification and diagnosis using stacked ensemble of residual neural networks," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 12259. doi: [10.1038/s41598-022-15632-6](https://doi.org/10.1038/s41598-022-15632-6).

- [22] D. Muduli, R. Dash, and B. Majhi, "Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach," *Biomed. Signal Process. Control*, vol. 71, 2022, Art. no. 102825. doi: [10.1016/j.bspc.2021.102825](https://doi.org/10.1016/j.bspc.2021.102825).
- [23] B. Surendiran, P. Ramanathan, and A. Vadivel, "Effect of BIRADS shape descriptors on breast cancer analysis," *Int. J. Med. Eng. Inform.*, vol. 7, no. 1, pp. 65–79, 2015. doi: [10.1504/IJMEI.2015.066244](https://doi.org/10.1504/IJMEI.2015.066244).
- [24] H. Li, J. Niu, D. Li, and C. Zhang, "Classification of breast mass in two-view mammograms via deep learning," *IET Image Process.*, vol. 15, no. 2, pp. 454–467, 2021. doi: [10.1049/ipr2.12035](https://doi.org/10.1049/ipr2.12035).
- [25] C. Yang, D. Sheng, B. Yang, W. Zheng, and C. Liu, "A dual-domain diffusion model for sparse-view CT reconstruction," *IEEE Signal Process. Lett.*, vol. 31, pp. 1279–1283, 2024.