Article

Breast Cancer Detection and Localizing the Mass Area Using Deep Learning

Md. Mijanur Rahman [1], **Md. Zihad Bin Jahangir** [1], **Anisur Rahman** [1], **Moni Akter** [1], **MD Abdullah Al Nasim** [2,*], **Kishor Datta Gupta** [3] **and Roy George** [3]

[1] Department of Computer Science and Engineering, Southeast University, Dhaka 1215, Bangladesh; mijanur.rahman@seu.edu.bd (M.M.R.); 2018200000015@seu.edu.bd (M.Z.B.J.); 2018200000018@seu.edu.bd (A.R.); 2019000000090@seu.edu.bd (M.A.)
[2] Research and Development Department, Pioneer Alpha, Dhaka 1205, Bangladesh
[3] Department of Computer and Information Science, Clark Atlanta University, Atlanta, GA 30314, USA; kgupta@cau.edu (K.D.G.); george@cau.edu (R.G.)
[*] Correspondence: nasim.abdullah@ieee.org

Abstract: Breast cancer presents a substantial health obstacle since it is the most widespread invasive cancer and the second most common cause of death in women. Prompt identification is essential for effective intervention, rendering breast cancer screening a critical component of healthcare. Although mammography is frequently employed for screening purposes, the manual diagnosis performed by pathologists can be laborious and susceptible to mistakes. Regrettably, the majority of research prioritizes mass classification over mass localization, resulting in an uneven distribution of attention. In response to this problem, we suggest a groundbreaking approach that seeks to identify and pinpoint cancers in breast mammography pictures. This will allow medical experts to identify tumors more quickly and with greater precision. This paper presents a complex deep convolutional neural network design that incorporates advanced deep learning techniques such as U-Net and YOLO. The objective is to enable automatic detection and localization of breast lesions in mammography pictures. To assess the effectiveness of our model, we carried out a thorough review that included a range of performance criteria. We specifically evaluated the accuracy, precision, recall, F1-score, ROC curve, and R-squared error using the publicly available MIAS dataset. Our model performed exceptionally well, with an accuracy rate of 93.0% and an AUC (area under the curve) of 98.6% for the detection job. Moreover, for the localization task, our model achieved a remarkably high R-squared value of 97%. These findings highlight that deep learning can boost the efficiency and accuracy of diagnosing breast cancer. The automation of breast lesion detection and classification offered by our proposed method bears substantial benefits. By alleviating the workload burden on pathologists, it facilitates expedited and accurate breast cancer screening processes. As a result, the proposed approach holds promise for improving healthcare outcomes and bolstering the overall effectiveness of breast cancer detection and diagnosis.

Keywords: breast cancer detection; breast cancer localization; deep learning; MIAS mammography dataset; convolution neural network; machine learning

1. Introduction

Breast cancer continues to be the most prevalent disease affecting women in the United States, with a documented 44,130 cases in 2021 [1]. Furthermore, it is the primary factor contributing to cancer-related deaths. In order to lessen the severe consequences of this harmful illness, it is crucial to prioritize frequent mammography exams. These screenings are extremely important since they have the ability to identify cancerous growths at an early stage before they have a chance to spread to nearby tissues and organs. Mammography involves using X-ray imaging to examine changes in breast tissue. Typically, radiologists

are responsible for diagnosing breast cancer by analyzing aberrant masses and microcalcifications [2,3]. The large number of mammograms that radiologists must evaluate on a daily basis, along with the inherent difficulties in identifying problematic areas, creates a procedure that is filled with hurdles, high expenses, and the possibility of errors. Therefore, there is a strong need to develop inventive and sophisticated approaches that can serve as catalysts for the accurate and exact identification of breast cancer. Convolutional neural networks (CNN), a type of deep learning, have made significant advancements in mammography analysis. Recent studies [4] have demonstrated the promising potential of CNNs in addressing the complexities of this field. Several papers [5–11] have explored the application of deep learning in mammography analysis. Nevertheless, convolutional neural networks (CNN) require a significant quantity of training data, which can impose constraints. Image segmentation is essential for automatically detecting and outlining key characteristics in medical images, such as tumors, organs, arteries, and cells. U-Net [12], a convolutional neural network (CNN)-based encoder–decoder network, has gained prominence as a leading segmentation tool and is widely accepted as the standard for medical image segmentation [13] in the industry. Although U-Net and its derivatives, such as Connected-UNet [14] and AU-Net [15], have proven to be effective in identifying breast lumps with limited labeled training data, there are still difficulties in reliably identifying regions of interest within the image. Within the field of deep learning, there is a widely recognized and highly regarded strategy called YOLO (You Only Look Once) [16], which has gained significant attention, especially when used in conjunction with U-Net. The YOLO system is an object detection system that uses a single convolution neural network to accurately forecast the bounding boxes and class probabilities of objects in real-time images. This technique has demonstrated favorable results in various domains, particularly in the processing of complex medical images. By combining YOLO and U-Net, there is the potential to develop a very efficient approach for automatically detecting, localizing, and classifying breast lesions in mammography pictures. We used the downsample part of the U-Net with modifications for feature extraction from the mammogram image and the concept of predicting bounding boxes from YOLO for localization. However, instead of predicting bounding boxes, we predict a circle to better suit the shape of the lesions. The primary emphasis in breast mammography research using deep learning algorithms has been on detecting cancer, sometimes neglecting the crucial component of accurately identifying the location of tumors. Our suggested model improves the precision of localization, giving medical practitioners more precise data to make informed therapeutic decisions. The model aims to close this gap and provide a faster and more accurate strategy for detecting tumors by effectively managing both the tasks of locating and classifying them. Our study methodology involves a sequence of crucial steps designed to enhance the current understanding. Firstly, we conduct a thorough evaluation of the constraints and shortcomings identified in previous research approaches, which serves as a basis for proposing areas for enhancement. Afterward, we begin the careful gathering and preparation of the dataset, implementing strict methods to guarantee the accuracy and reliability of the data. Once we have designed the blueprint for our cutting-edge architecture, we move on with training the model using a meticulously chosen dataset, employing advanced techniques to optimize its performance. In the end, we evaluate the trained model through a comprehensive analysis, meticulously assessing its capabilities and efficacy, and verifying the precision of our methodology.

2. Literature Review

The area of medical image processing has been much advanced by neural networks such as CNN and U-Net. Further advancing these advancements are object detection models like YOLO. In numerous medical image processing applications, such as object segmentation, detection, and classification, these models have been shown to be highly effective. Many studies have concentrated on finding and classifying breast masses in mammography pictures. In order to enhance the mass segmentation in these images, researchers from the University of Nevada have created two improved versions of the Connected-

UNets architecture: Connected UNets+, which includes residual skip connections, and Connected UNets++, which incorporates a modified encoder–decoder structure with residual skip connections [17]. An advanced technology called Mask-CNN (RoIAlign) deep learning has been developed to automatically detect, segment, and categorize breast lesions in mammography pictures. This state-of-the-art method was introduced by Jiménez Gaona et al. [18]. This novel methodology employs a very deep convolutional neural network, specifically the DenseNet architecture, to carry out classification and extract noteworthy characteristics. This sophisticated framework precisely detects and separates breast masses, representing a substantial advancement in the field of medical image processing. The event detection network (EDEN) [19] has demonstrated superior performance compared to the most advanced techniques in the field of survival analysis for breast cancer. EDEN, with a customized loss function and a time-aware long short-term memory network (LSTM), is capable of accurately labeling disease recurrence based on administrative claims. This offers medical professionals a potent instrument for prognostic evaluations and tailored treatment approaches, enhancing patient outcomes.

In ref. [20], Sajid et al. propose a novel approach to mammography breast cancer classification by combining user-defined features such as local binary patterns and the histogram of oriented gradients with CNN features. This revolutionary method, which starts a new era of outstanding medical image analysis, uses deep learning models to enhance the accuracy and efficacy of recognizing, isolating, and classifying breast masses. In medical image processing, CNNs, U-Nets, and YOLO are quite important algorithms. Among the many uses for these advanced models has been the precise identification and classification of breast masses. They are as helpful for other medical imaging techniques as mammography, including CT for lung cancer screening and MRI for tumor detection. With its vast array of advanced deep learning techniques, medical professionals' abilities are enhanced and disease detection and patient care are revolutionized.

In this paper [21], a novel approach to mass detection in digital mammograms that is independent of particular features is presented. This approach represents a significant change in the field of medical image analysis. This approach utilizes the complete image data, eliminating the conventional dependence on feature extraction approaches. The study presents an advanced system that incorporates two support vector machine (SVM) classifiers specifically developed to minimize the occurrence of false positives. The image data vectors are enhanced using a multi-resolution over-complete wavelet representation. These enhanced vectors are then inputted into the first support vector machine (SVM) to evaluate suspicious regions. The second support vector machine (SVM) categorizes the inputs into areas with masses and areas without masses, significantly decreasing the occurrence of incorrect identifications. A sophisticated voting system enhances the process of decision making by identifying areas that require further attention. When evaluated on mammograms from the USF-DDSM database, the suggested technique demonstrates a remarkable sensitivity rate of 80% and an exceptionally low false positive rate of 1.1 per picture.

Si and Jing [22] have created an advanced computer-aided detection and diagnosis (CAD) system that has significant potential for detecting and categorizing breast cancer on a large scale. The system utilizes a twin support vector machine (SVM) classifier and the dyadic wavelet approach to enhance the quality and diagnostic accuracy of mammography pictures. Through the efficient removal of undesired noise and the use of a segmentation technique, the region of interest (ROI) is precisely recovered, paving the way for subsequent research.

It is true that such remarkable performance indicates the possibility of the suggested approach to completely transform the diagnosis and detection of breast cancer. Eddaoudi et al. [23] presented a support vector machine and texture analysis-based mass detection approach. Three crucial phases make up the approach, and they all help to correctly classify ROIs in mammography pictures. In the initial stage, automatic initialization and contour detection utilizing snakes were employed to effectively delineate and split the pectoral muscle, a crucial step in isolating the target regions. Subsequently, in the second

stage, the ROI was further divided using a co-occurrence matrix and maximum thresholding Haralick characteristics. The final stage involved the application of a support vector machine (SVM) classifier to scrutinize the extracted characteristics and make insightful determinations regarding their affiliation with either normal or bulk regions. At 77% on average, the results showed promising performance. Still, the researchers saw significant improvements when using the classifier to pre-segmented mammograms, achieving an astounding average rate of 95% accuracy.

The abnormal detection classifier (ADC) is a state-of-the-art two-stage classifier designed by Jen and Yu [24] whose main goal is to identify abnormal mammograms. Understanding how important precise ROI detection is in mammograms, the researchers used basic image processing boosting techniques to successfully remove noise, non-breast regions, and spectral muscle, thus overcoming the obstacles related to ROI identification. This ground-breaking invention has enormous potential to improve the precision and effectiveness with which troublesome areas in mammography pictures are identified. Grey-level quantization was carefully used on all ROIs after the first processing stage to help recover a small but informative collection of significant features. We carefully assessed the ADC's performance with a dataset of 322 photos taken from the MIAS database. The results showed a remarkable 88% sensitivity and 84% specificity, which highlights the effectiveness and feasibility of the suggested ADC in precisely identifying and categorizing aberrant mammograms.

With their groundbreaking work, Ertosun and Rubin [25] presented a state-of-the-art visual search engine that uses deep learning to identify and localize breast masses in mammograms. Their technique consists of two essential parts: a highly developed deep learning classifier intended to precisely divide the whole image into separate mass and non-mass classes; and a sophisticated deep learning network-based regional probabilistic approach, which excels in precisely locating the masses within mammography images. With just an average of 0.9 false positives per image, the scientists claim that their unique method produces remarkable results. Moreover, when their technique is used for the given task, they claim an amazing accuracy rate of 85% in mass classification and localization.

In the field of medical image analysis, Jadoon et al. [26] put out a unique CNN-based method to divide mammograms into three different categories: benign, malignant, and normal. The discrete wavelet transform-based CNN (CNN-DW) and curvelet transform-based CNN (CNN-CT) are the two different variants of their novel approach. Especially, the effectiveness of using CNN features taken from mammograms to identify malignant diseases is demonstrated by this work. Extensive evaluations of the authors' suggested CNN-DW and CNN-CT methods on the IRMA dataset yielded very encouraging results. Impressively accurate at 81.83%, the CNN-DW method was surpassed by the CNN-CT, which obtained an even higher accuracy of 83.74%.

These results signal a major development in the field of medical image analysis and highlight the potential of their CNN-based classification approach as a useful tool in the early detection and diagnosis of breast cancer. Even with the increasing amount of study being conducted on breast mammography using deep learning algorithms, there is still a clear difference in the overemphasis on image categorization rather than precise tumor location. Effective diagnosis and therapy planning depend on accurate localization; hence, this attention imbalance frequently results in less-than-ideal therapeutic results. Remarkably, earlier studies in this field have focused mostly on cancer diagnosis and have given far less attention to accurately locating anomalies in mammography pictures. Localization deserves further attention in order to guarantee the precision and dependability of the diagnostic procedure. In our suggested model, we seek to improve mass localization precision, a crucial step in the diagnostic pipeline, therefore providing medical practitioners with more accurate and relevant data for improved clinical decision making. Taking care of this critical gap, our model offers a complete solution for rapid and precise tumor identification by taking on both localization and classification tasks simultaneously. By means of this

all-encompassing strategy, we want to enhance diagnostic effectiveness and efficiency, thus enhancing patient care and health results.

3. Methodology

This section presents a comprehensive overview of the methodologies and techniques employed in the present research, encompassing four pivotal subsections: dataset and preprocessing, model architecture, training, and evaluation metrics. The dataset utilized in this study was procured from a reputable and dependable source, specifically, https://www.kaggle.com, accessed on 31 May 2024. To ascertain the veracity and reliability of the data, a comprehensive preprocessing phase was rigorously conducted, aimed at refining and enhancing the dataset's quality for subsequent analyses. Figure 1 provides a visual representation of the workflow utilized in our methodology.
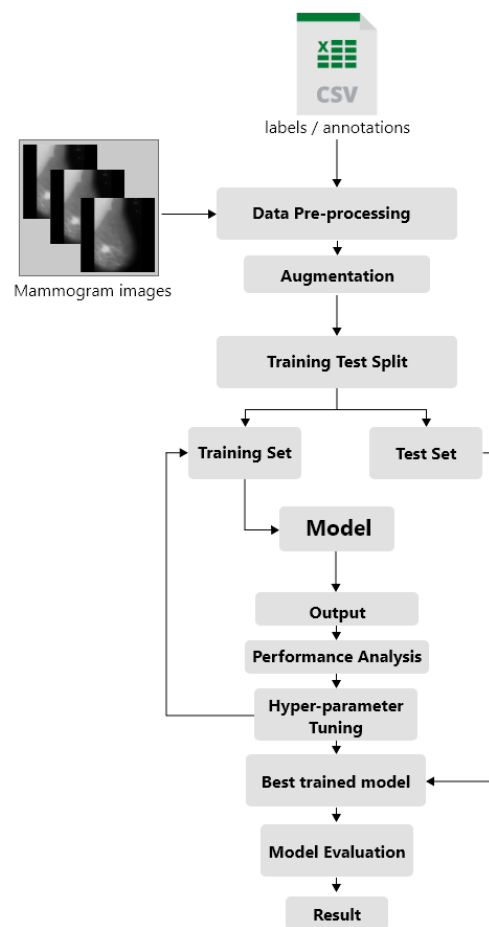

Figure 1. A concise visual depiction of the methodology employed in the study.

3.1. Dataset and Preprocessing

In our experimental investigations, we leveraged the publicly accessible Mammographic Image Analysis Society (MIAS) datasets [27], comprising a total of 330 images that were thoughtfully categorized into multiple classes. The MIAS dataset, compiled in the late 1980s and early 1990s for research in mammographic image analysis, features images captured using standard mammography X-ray equipment of that era. The original dimensions of each image in the dataset were 1024 × 1024 pixels, but for the purpose of consistency and efficient processing, all images were uniformly resized to 512 × 512 pixels. This resizing decision aimed to optimize our model's efficiency by standardizing image dimensions, which benefits from reduced computational load and faster processing times during both training and inference stages. Additionally, specific preprocessing techniques were applied

to enhance mammographic features, ensuring accurate classification and localization tasks. To ensure coherence and accuracy in subsequent analyses, we appropriately resized the 'x' and 'y' image coordinates, along with the radius of the abnormality area outlined in the labels file. Sample images from the dataset are visually depicted in Figure 2.
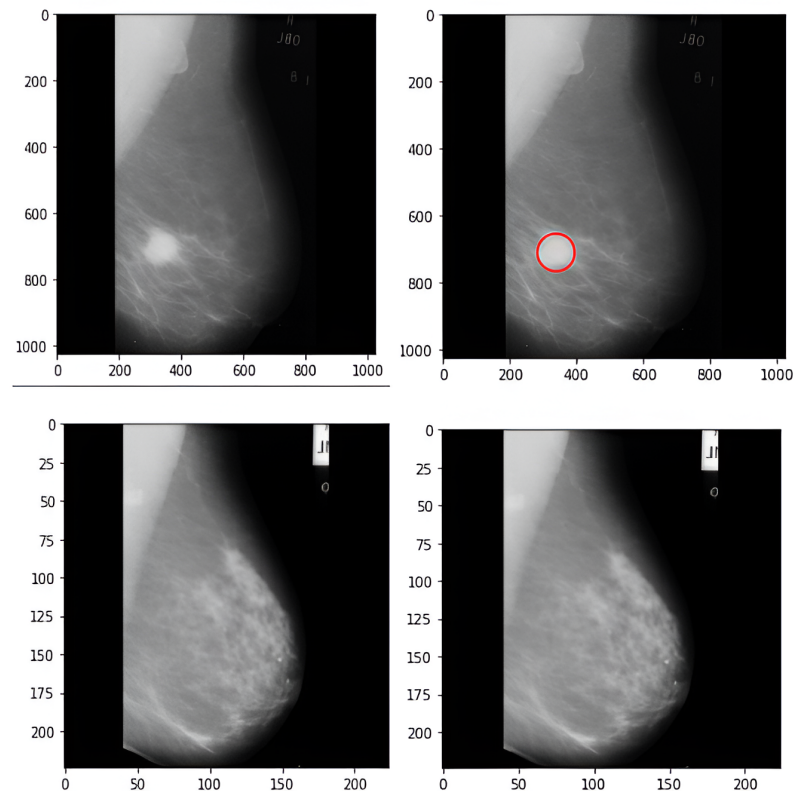

Figure 2. Illustrative sample of mammogram image data showcasing the complexity and diversity of the dataset used in the study. The red circle shows the localization result.

Based on the mammogram scans, the severity of the abnormality present among the tissues is divided into two classes: benign and malignant. Based on the mammogram scans, the severity of the abnormality present among the tissues is divided into two classes: benign and malignant.

The dataset employed in this study exhibits class imbalance (Table 1), wherein certain classes contain a lesser number of samples compared to others. This imbalance poses a challenge for machine learning models, as they may exhibit a bias towards the majority class, compromising the overall performance. To mitigate this issue and enhance the model's robustness, data augmentation techniques were diligently applied. These techniques involve the augmentation of the original dataset by performing various operations such as zoom-in, zoom-out, and flip, thereby increasing the dataset's size and rebalancing the class distribution. Through the application of data augmentation, the initial dataset of 330 mammography images was expanded, resulting in a new dataset comprising 750 images (Figure 3). Of these augmented images, 600 (88%) were designated for the training of the machine learning model, while the remaining 150 (12%) were reserved for evaluating its performance during testing.

Table 1. Dataset details showcasing the number of images per class in the study.

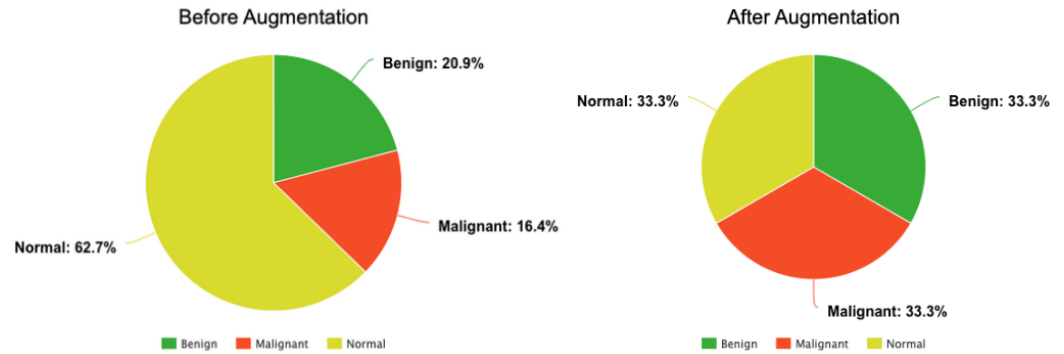| Class	Number of Images
Benign	69
Malignant	54
Normal	207
Total	330

Figure 3. Comparison of dataset class ratios before and after augmentation, highlighting the impact of data augmentation techniques on balancing the distribution of classes within the dataset.

3.2. Model Architecture

The architecture proposed in Figure 4 draws inspiration from two well-established models, namely U-Net and YOLO. While the downsampling component bears resemblance to U-Net, it incorporates distinct filters to cater to the specific requirements of the task at hand. Additionally, the localization section is inspired by YOLO is renowned for its prowess in object detection. In our approach, we adopt a unique perspective, treating the entire image as a window for analysis. To facilitate this, the training dataset's x, y, and radius values are appropriately transformed. Specifically, we normalize the x, y, and radius values by dividing them by the image size, which in this case is 512×512 pixels.
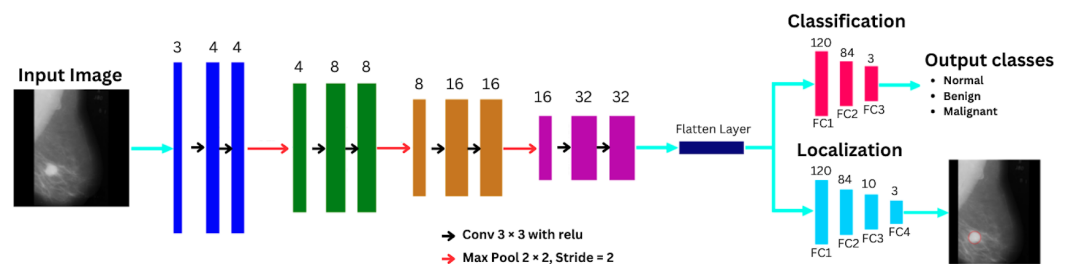

Figure 4. Model architecture illustrates the proposed model's intricate structure, with a clear distinction between the contracting route on the left and the classification and localization segment on the right.

Upon referring to Figure 4, the intricacies of the network architecture become visually evident, showcasing a thoughtful division into two distinctive components: the contracting route on the left and the classification and localization segment on the right. The contracting route adheres to the well-established blueprint of convolutional neural networks, where a sequence of two 3×3 convolutions is iteratively applied. Subsequent to each convolutional layer, a rectified linear unit (ReLU) activation function is strategically employed, contributing to the network's nonlinear capabilities. A downsampling operation is then executed using a 2×2 max pooling technique with a stride of 2, effectively reducing the spatial dimensions of the feature maps. Notably, at each downsampling stage, the number of feature

channels undergoes a doubling process, fostering the creation of a compacted-height layer after four iterations.

Subsequent to the stratified layered structure described earlier, the network's trajectory progresses into the following key components: the categorization phase and the localization module. In the classification step, three fully connected layers are adeptly integrated, orchestrating the intricate interplay of learned features to generate the desired class output, efficiently identifying the breast mass into one of the defined categories (normal, malignant, or benign). On the other hand, the localization aspect constitutes a critical element of the architecture, consisting of four interconnected layers meticulously engineered to yield the pivotal x, y, and radius coordinates.

The model at hand exhibits a substantial complexity, with a total of 7,904,590 trainable parameters seamlessly integrated into its design. Notably, the model does not encompass any non-trainable parameters, indicating a comprehensive reliance on learnable features and eliminating the need for fixed or static elements. The model's input size is measured at 1.00 megabytes, signifying the magnitude of data that it can assimilate and process in a single instance. The forward/backward pass, on the other hand, is an essential computational activity for training and optimizing the model; it uses up a lot of resources, with an estimated size of 108.76 MB. The model's parameter size, which includes all of the trainable parameters, is 30.15 megabytes, which is the total amount of RAM used for this purpose. Consequently, the model's memory footprint is somewhat large, reaching an estimated 139.91 MB when all of its components are considered. Making educated decisions to improve performance and resource consumption is made possible by these comprehensive measures, which cover computational complexity and memory requirements. They serve as important guides for strategic model deployment and optimization.

3.3. Training

The training process of our proposed model involved a number of carefully performed processes, flawlessly performed on our local computer. In order to obtain the best possible training performance, we utilized a GTX1050 GPU (NVIDIA Corporation, Dhaka, Bangladesh) with 4 GB of GPU RAM. This allowed us to take advantage of its strong computing abilities to quickly and effectively optimize our model. The dataset used for training was carefully divided into two different sets: a training set consisting of 600 photos and a separate testing set of 150 images. Throughout the training phase, we deliberately utilized a batch size of 30, a carefully selected decision that expedited the optimization of the model's parameters. The model was trained for 300 epochs, a wise choice that allowed sufficient time for the model to progressively understand and assimilate the complex patterns and fundamental characteristics present in the training dataset. We established a learning rate of 2×10^{-5}, which struck a careful equilibrium. This rate allowed the model to make steady progress in its learning while avoiding any negative effects on its stability during the training phase. To determine these optimal hyperparameters, such as batch size and learning rate (alpha), we experimented with various values and found that these specific settings best fit our model.

$$\text{Cross Entropy Loss}(y, \hat{y}) = -\sum_i y_i \log(\hat{y}_i) \tag{1}$$

where y is the true label distribution, $\hat{y}$ is the predicted label distribution, and i iterates over all classes.

$$\text{Mean Squared Error}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

where y is the true label, $\hat{y}$ is the predicted label, and n is the number of samples.

In the pursuit of optimizing our model's performance, we judiciously incorporated two distinct loss functions: the cross-entropy loss Equation (1) and the mean squared error (MSE) loss Equation (2). Each of these loss functions played a pivotal role in fine-tuning the model's classification and localization capabilities, respectively. Subsequently,

to holistically gauge the model's overall performance, we summed the individual losses from the classification and localization components, thus yielding the final composite loss. Inspired by the principles of YOLO, our model's localization mechanism approached the challenge by treating the entire image as a single window. The dataset was prepared to align with this approach, as the x, y, and radius values were appropriately divided by the image size of 512 × 512 (512 being the dimension of both height and width).

3.4. Evaluation Metrics

In our endeavor to comprehensively assess the prowess of our proposed model, we employed a battery of rigorous evaluation metrics, each instrumental in scrutinizing distinct facets of its performance. These evaluation criteria encompassed accuracy, precision, recall, F1 score, ROC curve, and R squared error, collectively constituting a comprehensive assessment toolkit.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Accuracy, a fundamental metric in classification tasks, serves as a pivotal gauge for assessing the model's proficiency in accurately predicting both positive and negative events. The calculation of accuracy Equation (3) involves the formulation of a ratio representing the total number of accurately predicted instances (both true positives and true negatives) over the entire dataset's population. True positives (TP) denote the correctly classified positive instances, while true negatives (TN) refer to the accurately identified negative instances. On the other hand, false positives (FP) correspond to the instances falsely classified as positive, and false negatives (FN) represent the erroneously identified negative instances.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

In essence, the precision Equation (4) serves as a powerful measure of the model's capability to precisely identify and correctly classify positive instances from the pool of predicted positive results. By evaluating the precision metric, we gain valuable insights into the model's ability to accurately and reliably identify the exact locations of tumors, thus illuminating its precision in detecting and classifying tumor regions with a high degree of accuracy.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

In the context of our research, recall Equation (5) assumes critical significance as it specifically plays a pivotal role in assessing the model's capacity to effectively detect positive instances. It quantifies the percentage of true positive results that the model correctly identifies, thereby offering valuable insights into detecting and capturing all tumor regions, underscoring its accuracy in identifying the presence of tumors in a comprehensive manner.

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

The harmonic mean of recall and precision is the F1 score. Symbolically, the F1 score Equation (6) is computed as a single statistic that incorporates recall and precision. By leveraging the F1 score, we gain valuable insights into the trade-offs between recall and precision, enabling a thorough comparison of various models with varying precision-recall characteristics.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (7)$$

The R-squared error Equation (7) calculates the degree to which the expected and actual values agree. It serves as a critical measure to assess the performance of the regression model utilized for tumor localization. By evaluating the degree to which the anticipated

tumor radius values align with the actual ones, the R-squared error provides a valuable indication of the model's accuracy in predicting the tumor size.

Confusion Matrix: A confusion matrix is a table that compares a dataset's anticipated and actual values to assess a classification model's effectiveness. The true positive rate (TPR) against false positive rate (FPR) trade-off at various categorization levels is depicted graphically by the receiver operating characteristic (ROC) curve. In our research, the ROC curve was utilized to assess how well our model performed when applied to various criteria.

4. Result

To evaluate the effectiveness of our model, we used several evaluation indicators for both outcomes. In Table 2, we used accuracy, precision, recall, and F1 score as metrics to evaluate the classification part. We used the R-squared value for the localization result as the evaluation metric, shown in Table 3. The results of our examination revealed exceptional performance across all evaluation metrics for the categorization output. The performance of our model was exceptional, as seen by its high accuracy, precision, recall, and F1 score. This indicates its ability to accurately classify cases of breast cancer. Likewise, our model performed exceptionally well in producing accurate localization results, as indicated by its high R-squared value. This demonstrates the model's efficacy in accurately identifying the specific region impacted by cancer in breast images. In addition to the evaluation measures described earlier, we provided a thorough study of the model's performance using a confusion matrix and an ROC curve. The confusion matrix presented a detailed representation of the quantities of true positive, true negative, false positive, and false negative samples. The ROC curve allows for a simultaneous evaluation of the true positive rate and the false positive rate, covering a range of classification criteria. Collectively, these assessment methods provide more clarity and precision about the resilience and precision of our model for detecting breast cancer and localizing cancerous areas.

Table 2. Classification results.

Metric	Training Data	Test Data
Accuracy	99.7%	93.0%
Precision	99.6%	93.2%
Recall	99.9%	94.7%
F1 Score	99.6%	93.2%
AUC	99.9%	96.0%

Table 3. Localization results.

Evaluation Metrics	Training Data	Test Data
R Squared value	99.1%	97.6%

4.1. Confusion Matrix

According to our model evaluation, we provide the confusion matrix for both our training and test data in Figure 5. The confusion matrix displays a three-part structure, where the rows and columns are precisely arranged to represent the three unique classes: normal, benign, and malignant. More precisely, the initial row and column represent the normal class, the second row and column represent the benign class, and the third row and column represent the malignant class.

Our model achieved flawless accuracy in categorizing the normal and malignant classes for the training data, with only one false positive and no false negatives. Nevertheless, it exhibited one incorrect result indicating the innocuous class. The model achieved high accuracy in classifying all three classes of the test data; however, it did produce some incorrect positive and negative predictions for the normal and benign classes. Our model accurately identified 49 images of the normal class, with 4 images incorrectly classified as

positive and 4 images incorrectly classified as negative. For the benign class, the model correctly identified 33 images with no false positives or false negatives. Lastly, for the malignant class, the model correctly identified 49 images, with 1 false positive and 2 false negatives.

<table>
<tr><td colspan="5" align="center">Test Set</td></tr>
<tr><td>TARGET

OUTPUT</td><td>Normal</td><td>Benign</td><td>Malignant</td><td>SUM</td></tr>
<tr><td>Normal</td><td>49
35.51%</td><td>4
2.90%</td><td>0
0.00%</td><td>53
92.45%
7.55%</td></tr>
<tr><td>Benign</td><td>0
0.00%</td><td>33
23.91%</td><td>0
0.00%</td><td>33
100.00%
0.00%</td></tr>
<tr><td>Malignant</td><td>2
1.45%</td><td>1
0.72%</td><td>49
35.51%</td><td>52
94.23%
5.77%</td></tr>
<tr><td>SUM</td><td>51
96.08%
3.92%</td><td>38
86.84%
13.16%</td><td>49
100.00%
0.00%</td><td>131 / 138
94.93%
5.07%</td></tr>
</table>

<table>
<tr><td colspan="5" align="center">Test Set</td></tr>
<tr><td>TARGET

OUTPUT</td><td>Normal</td><td>Benign</td><td>Malignant</td><td>SUM</td></tr>
<tr><td>Normal</td><td>49
35.51%</td><td>4
2.90%</td><td>0
0.00%</td><td>53
92.45%
7.55%</td></tr>
<tr><td>Benign</td><td>0
0.00%</td><td>33
23.91%</td><td>0
0.00%</td><td>33
100.00%
0.00%</td></tr>
<tr><td>Malignant</td><td>2
1.45%</td><td>1
0.72%</td><td>49
35.51%</td><td>52
94.23%
5.77%</td></tr>
<tr><td>SUM</td><td>51
96.08%
3.92%</td><td>38
86.84%
13.16%</td><td>49
100.00%
0.00%</td><td>131 / 138
94.93%
5.07%</td></tr>
</table>

Figure 5. Confusion matrix for the training and test sets.

4.2. ROC Curve

Depicted in Figure 6 is the receiver operating characteristic (ROC) curve, which offers significant insights regarding the model's ability to differentiate between the test data. Through the analysis of area under the curve (AUC) values, it is possible to assess the efficacy of the model with respect to each class in isolation. The remarkable AUC value of 0.98 for class 0 (normal) indicates that the model possesses a strong capability to accurately classify normal classes from the other classes. In a similar, the AUC for class 1 (benign) was an exceptional 0.99, showing the model's exceptional ability to accurately categorize benign instances. Moreover, in regard to class 2 (malignant), the model achieved a tremendous AUC of 0.99, providing additional evidence of its ability to accurately classify malignant samples. The AUC numbers we obtained show that our model is very good at finding breast cancer across all three classes, with the malignant class performing the best. Also, the high AUC values show that our model does a good job of telling the difference between the rates of true positives and fake positives for each class.
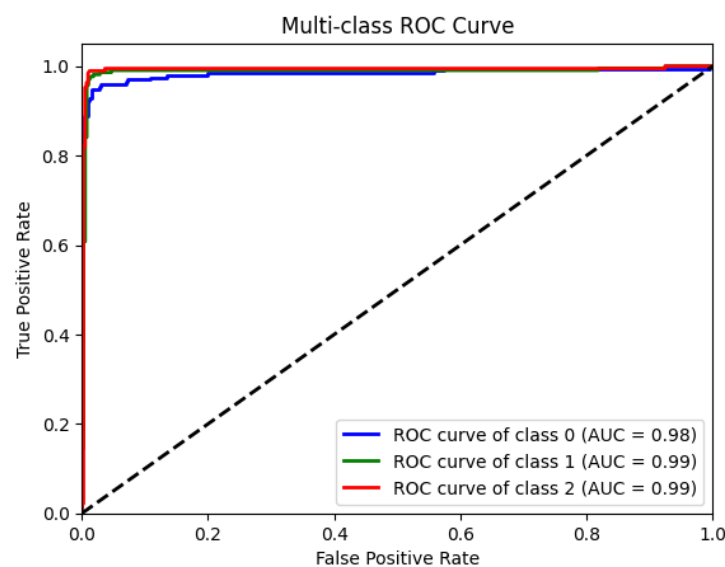

Figure 6. Multi-class ROC curve.

5. Discussion

A big part of making sure that AI models work well and quickly is testing them. Researchers and people who use these models can both learn from this process what they can do better and what they need to change. The more researchers try their models, the more they learn about how well they work, which helps them figure out how to improve them. It is very important to look at the models when trying to find and spot breast cancer. It lets us see how well and consistently our model can find cancerous areas and tell the difference between cases that are dangerous, cases that are not dangerous, and cases that are normal.

5.1. Classification

Our main study goal is to divide breast cancer cases into three groups: those that are cancerous, those that are not cancerous, and those that are normal. We used a wide range of success metrics, as shown in Table 2, to make this happen. Several metrics were looked at, and each one gave us information about a different part of the model's success. Some of these measures are precision, recall, R-squared, area under the curve (AUC), and accuracy.

The model was tested and found to be very accurate, correctly identifying 93.0% of breast cancer patients and showing good discriminatory ability with an AUC of 98.6%. In addition, our model had a high recall of 94.7%, which means it could correctly spot a lot of breast cancer patients. Also, it shows a precision of 93.2%, which means that it correctly classified the cases it found. The model was also able to tell the difference between true positive and fake positive rates, which shows that it can accurately find people with breast cancer.

During our examination of the confusion matrices for our test data, we observed instances of both false positives and false negatives in the benign and normal classes. The malignant class exhibited the highest level of performance, as indicated by the ROC curve, which demonstrated high levels of sensitivity and specificity for all three classes. The results indicate that our system has outstanding performance in accurately detecting breast cancer, a crucial factor for early detection and improved patient outcomes.

Compared to other models, as Table 4 illustrates, our model shows superior accuracy and AUC. For instance, the method by Eddaoudi et al. [23] achieved an accuracy of 95%, while our approach achieved 93.0% accuracy and a much higher AUC of 98.6%, indicating better overall performance. Although Eddaoudi et al.'s method is highly accurate, it primarily relies on texture features, which may limit its ability to generalize across diverse datasets. Our model's higher AUC suggests better overall performance and robustness due to its ability to extract features using deep learning techniques. The sensitivity of our model also surpasses that of Jen and Yu [24], which reported 88% sensitivity or recall, compared to our model's 94.7%. Furthermore, Jen and Yu's method focuses on detecting abnormal mammograms, providing high sensitivity but potentially suffering from higher false positives due to less specificity. In contrast, our model's combination of U-Net and YOLO contributes to more accurate classification and localization, enhancing its practical applicability in clinical settings.

5.2. Localization

Our model does more than just classification; it also focuses on finding mass areas. This localization result can help a lot when it comes to making accurate diagnoses and treatment choices because it helps find specific cancerous spots. The R-squared number in Table 3 was used to judge how well our model's localization worked. We were amazed that our algorithm was able to accurately locate cancerous breast tissue, as shown by its amazing R-squared value of 97.6%. This good result shows that our model might be useful for helping doctors give more accurate diagnoses and treatment plans to women with breast cancer. In the end, our evaluation results show that our model is useful and reliable for the job of classifying and localizing breast cancer. Our model's high AUC, precision, recall, and R-squared value show that it can exactly find breast cancer cases and pinpoint

areas that are cancerous. These results hold a lot of hope for finding breast cancer early and giving patients better care, which will ultimately lead to better outcomes.

Table 4. Comparison of our model and other models.

Reference	Classification	Localization	Method
[21]	-	Sensitivity 80%	Support vector machine (SVM)
[22]	-	Sensitivity 89%	Support vector machine (SVM)
[23]	Accuracy 95%	-	Support vector machine (SVM)
[24]	Sensitivity 88%	Specificity 84%	Abnormality detection classifier (ADC)
[25]	Accuracy 85%	Sensitivity 85%	Deep learning classifier with regional probabilistic
[26]	Accuracy 85%	-	CNN-DW and CNN-CT
[28]	Accuracy 75% AUC 83%	-	EfficientNet
[29]	Accuracy 88%	-	Deep CNN
[30]	-	Sensitivity 88%	CNN
Our approach	Accuracy 93.0% AUC 98.6% Precision 93.2% Recall 94.7%	R-squared value 97.6% Sensitivity 96.81%	Our proposed model

Compared to Ertosun and Rubin [25], who achieved 85% sensitivity in localization using deep learning classifiers, our model's R-squared value of 97.6% indicates significantly higher precision in pinpointing cancerous regions. Ertosun and Rubin's probabilistic approach is valuable for uncertainty estimation but may not achieve the same precision due to its probabilistic nature. Additionally, our model's approach of combining classification with localization provides a comprehensive solution, as opposed to models like those proposed by Campanini et al. [21] and Si and Jing [22], which focus mainly on detection with sensitivity rates of 80% and 89%, respectively. Campanini et al.'s SVM-based approach is effective but can struggle with feature selection, while Si and Jing's twin SVM-based CAD system can be computationally intensive. Our dual-output model offers a more holistic solution by integrating detection and precise localization, improving overall diagnostic accuracy and efficiency.

Our model shows promise for finding and localizing breast cancer. Additional studies could help make it work even better, which could lead to more accurate and useful diagnoses and treatment plans for people with breast cancer.

6. Limitations and Future Work

Despite our model's excellent performance in the evaluation, we must consider several drawbacks. First and foremost, it relies solely on the MIAS dataset, which may limit its usefulness to larger populations and other imaging settings. Second, for this kind of study, the sample size of 330 images is somewhat small, which might impact the robustness and generalizability of the model. Finally, the dataset mostly covers particular geographical areas, which might affect how well the model works with people with different demographics or imaging requirements. Future work can utilize larger and more varied datasets to enhance the model's validation and clinical situational performance. Creating apps that support various imaging equipment would expand the model's applicability and practicality in real-world situations.

7. Conclusions

In the end, our study developed a novel and useful model for finding breast cancer and figuring out where it is located. Our model takes the best parts of two well-known models—U-Net and YOLO—and blends them into a single, useful output that can be used in a variety of clinical situations. Our evaluation measures showed that our model was very good at finding breast cancer and pinpointing where the cancer is. This can help find the disease earlier, make sure it is correctly diagnosed, and improve patient outcomes. The good results of our study show that our proposed model has a lot of promise to improve the accuracy and effectiveness of breast cancer diagnosis and treatment steps. Our model's unique dual output can help with planning a test, making plans for surgery, and giving radiation therapy. Our model can also help find situations where tests might not be necessary. This means that patients will be less uncomfortable, save money, and feel less anxious.

Author Contributions: Data curation, A.R. and M.A.; formal analysis, M.Z.B.J.; investigation, M.Z.B.J.; methodology, M.Z.B.J.; supervision, M.M.R., M.A.A.N., K.D.G., and R.G.; visualization, M.Z.B.J. and A.R.; writing—original draft, M.Z.B.J.; writing—review and editing, M.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The dataset can be accessed on the Kaggle website at https://www.kaggle.com/datasets/kmader/mias-mammography/data, accessed on 5 January 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. DeSantis, C.E.; Ma, J.; Gaudet, M.M.; Newman, L.A.; Miller, K.D.; Goding Sauer, A.; Jemal, A.; Siegel, R.L. Breast Cancer Statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 438–451. [CrossRef] [PubMed]
2. Elter, M.; Horsch, A. CADx of mammographic masses and clustered microcalcifications: A review. *Med. Phys.* **2009**, *36*, 2052–2068. [CrossRef] [PubMed]
3. Jiang, Y.; Nishikawa, R.M.; Schmidt, R.A.; Metz, C.E.; Giger, M.L.; Doi, K. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad. Radiol.* **1999**, *6*, 22–33. [CrossRef] [PubMed]
4. Zaheer, R.; Shaziya, H. GPU-based empirical evaluation of activation functions in convolutional neural networks. In Proceedings of the 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 19–20 January 2018; pp. 769–773.
5. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
6. Brinker, T.J.; Hekler, A.; Enk, A.H.; Berking, C.; Haferkamp, S.; Hauschild, A.; Weichenthal, M.; Klode, J.; Schadendorf, D.; Holland-Letz, T.; et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur. J. Cancer* **2019**, *119*, 11–17. [CrossRef] [PubMed]
7. Assiri, A.S.; Nazir, S.; Velastin, S.A. Breast tumor classification using an ensemble machine learning method. *J. Imaging* **2020**, *6*, 39. [CrossRef] [PubMed]
8. Manickavasagam, R.; Selvan, S.; Selvan, M. CAD system for lung nodule detection using deep learning with CNN. *Med. Biol. Eng. Comput.* **2022**, *60*, 221–228. [CrossRef] [PubMed]
9. Tandon, Y.K.; Bartholmai, B.J.; Koo, C.W. Putting artificial intelligence (AI) on the spot: Machine learning evaluation of pulmonary nodules. *J. Thorac. Dis.* **2020**, *12*, 6954. [CrossRef] [PubMed]
10. Bechelli, S. Computer-Aided Cancer Diagnosis via Machine Learning and Deep Learning: A comparative review. *arXiv* **2022**, arXiv:2210.11943.
11. Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer diagnosis using deep learning: A bibliographic review. *Cancers* **2019**, *11*, 1235. [CrossRef] [PubMed]
12. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
13. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
14. Baccouche, A.; Garcia-Zapirain, B.; Castillo Olea, C.; Elmaghraby, A.S. Connected-UNets: A deep learning architecture for breast mass segmentation. *NPJ Breast Cancer* **2021**, *7*, 151. [CrossRef] [PubMed]

15. Sun, H.; Li, C.; Liu, B.; Liu, Z.; Wang, M.; Zheng, H.; Feng, D.D.; Wang, S. AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. *Phys. Med. Biol.* **2020**, *65*, 055005. [CrossRef] [PubMed]

16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

17. Sarker, P.; Sarker, S.; Bebis, G.; Tavakkoli, A. Connectedunets++: Mass segmentation from whole mammographic images. In Proceedings of the International Symposium on Visual Computing, San Diego, CA, USA, 3–5 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 419–430.

18. Jiménez Gaona, Y.; Rodriguez-Alvarez, M.J.; Espino-Morato, H.; Castillo Malla, D.; Lakshminarayanan, V. Densenet for breast tumor classification in mammographic images. In Proceedings of the International Conference on Bioengineering and Biomedical Signal and Image Processing, Gran Canaria, Spain, 19–21 July 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 166–176.

19. Dumas, E.; Hamy, A.S.; Houzard, S.; Hernandez, E.; Toussaint, A.; Guerin, J.; Chanas, L.; de Castelbajac, V.; Saint-Ghislain, M.; Grandal, B.; et al. EDEN: An Event DEtection Network for the annotation of Breast Cancer recurrences in administrative claims data. *arXiv* **2022**, arXiv:2211.08077.

20. Sajid, U.; Khan, R.A.; Shah, S.M.; Arif, S. Breast cancer classification using deep learned features boosted with handcrafted features. *Biomed. Signal Process. Control.* **2023**, *86*, 105353. [CrossRef]

21. Campanini, R.; Dongiovanni, D.; Iampieri, E.; Lanconelli, N.; Masotti, M.; Palermo, G.; Riccardi, A.; Roffilli, M. A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Phys. Med. Biol.* **2004**, *49*, 961. [CrossRef] [PubMed]

22. Si, X.; Jing, L. Mass detection in digital mammograms using twin support vector machine-based CAD system. In Proceedings of the 2009 WASE International Conference on Information Engineering, Taiyuan, China, 10–11 July 2009; Volume 1, pp. 240–243.

23. Eddaoudi, F.; Regragui, F.; Mahmoudi, A.; Lamouri, N. Masses detection using SVM classifier based on textures analysis. *Appl. Math. Sci.* **2011**, *5*, 367–379.

24. Jen, C.C.; Yu, S.S. Automatic detection of abnormal mammograms in mammographic images. *Expert Syst. Appl.* **2015**, *42*, 3048–3055. [CrossRef]

25. Ertosun, M.G.; Rubin, D.L. Probabilistic visual search for masses within mammography images using deep learning. In Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA, 9–12 November 2015; pp. 1310–1315.

26. Jadoon, M.M.; Zhang, Q.; Haq, I.U.; Butt, S.; Jadoon, A. Three-class mammogram classification based on descriptive CNN features. *BioMed Res. Int.* **2017**, *2017*, 3640901. [CrossRef] [PubMed]

27. Suckling, J. The mammographic images analysis society digital mammogram database. In Proceedings of the Exerpta Medica, International Congress Series, York, UK, 10–12 July 1994; Volume 1069, pp. 375–378.

28. Gengtian, S.; Bing, B.; Guoyou, Z. EfficientNet-Based Deep Learning Approach for Breast Cancer Detection With Mammography Images. In Proceedings of the 2023 8th International Conference on Computer and Communication Systems (ICCCS), Guangzhou, China, 21–24 April 2023; pp. 972–977.

29. Nalifabegam, J.; Ganeshbabu, C.; Askarali, N.; Natarajan, A.; Maheshwari, P. Cancer Classification Revolution: Employing Advanced Deep CNNs for Multi-Class Detection of Breast Irregularities. In Proceedings of the 2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 9–10 December 2023; Volume 1, pp. 1–4.

30. Pourasad, Y.; Zarouri, E.; Salemizadeh Parizi, M.; Salih Mohammed, A. Presentation of novel architecture for diagnosis and identifying breast cancer location based on ultrasound images using machine learning. *Diagnostics* **2021**, *11*, 1870. [CrossRef] [PubMed]