Article   Open Access

Deep Neural Network-Based Model for Breast Cancer Lesion Diagnosis in Mammography Images

Mohamed Amine Yakoubi [1] (ID), Nada Khiari [1] (ID), Amine Khiari [2] (ID), Ahlem Melouah [1] (ID)

[1] Laboratoire de Recherche en Informatique, Badji Mokhtar Annaba University, Annaba, Algeria
[2] Département d'informatique, Université Toulouse III – Paul Sabatier, Toulouse, France

Corresponding author: Mohamed Amine Yakoubi (amine.yakoubi@univ-annaba.dz)

Abstract

Deep learning has made identifying breast cancer lesions in mammography images an easy task in modern medicine, which has helped improve the diagnosis efficiency, sensitivity and accuracy by precisely identifying breast cancer from mammography images, contributing to timely detection and maintaining consistent performance. This paper presents the steps and strategies to develop a deep learning (DL) model to detect lesions in mammography images, based on U-Net architecture for precise segmentation, which has been developed for biomedical image segmentation, and incorporating ResNet34 as its encoder to extract features. Next, we employ the FastAI library, which simplifies and accelerates the model training tasks. For the data, studies and available resources lead us to INbreast, which is built with full-field digital mammograms contrary to other digitized mammograms. We obtained a high accuracy of 98% on the INbreast database, which is very challenging compared to state-of-the-art results.

Keywords

Deep learning; Mammography; Lesion detection; U-Net; ResNet34.

1   Introduction

The recent advances in computing power have helped computer scientists build models with high computing requirements, specifically for image processing (Fajrianti et al., 2022). From a collection of annotated images, models can be trained and produce better and faster results than a human domain expert (Schaudt et al., 2023). Radiologists need to screen and visualize hundreds of medical images. This task requires very high precision to detect all abnormalities. Clinical treatment can then be adopted. If we could automatize this task, or at least give the human expert a higher representation of the image, in which suspicious regions are highlighted, diseases would be easier to detect.

The recent advancement of artificial intelligence techniques, especially deep learning, continues to attract the medical imaging community's interest to improve the precision of cancer screening (Jiang et al., 2023). Due to statistics (Iacoviello et al., 2021), breast cancer is the leading cause of death in women between 20-50 years old. When patients are taken care of at an early stage of the disease, their vital prognosis improves. The process that patients undergo during diagnosis involves two different phases. First, X-ray images of each breast are taken from two different angles during a routine mammogram: examinations involve dual-view mammography (craniocaudal, CC, and mediolateral oblique, MLO) of each breast (Moshina et al., 2022), and one or two experts examine these electromagnetic waves for abnormalities. Afterwards, diagnostic evaluations are made for suspicious cases (Akhund et al., 2023). Despite the benefits, screening mammography results are threatened by a high risk of false positives. Screening mammograms needs to be evaluated by experienced readers. This procedure is monotonous, tiring, lengthy, costly and beyond all other consideration, prone to errors (Piva et al., 2023). In order to help radiologists, the development of computer-assisted detection and diagnostic CAD software has improved the predictive accuracy of screening mammography. CAD systems have been developed since the 1990s to ameliorate this problem (Dominguez et al., 2020). These applications review a mammogram and highlight any worrisome areas that the radiologist should examine. This system is expensive because of its costs and the necessity of maintenance. Health care systems in developing countries fail to cover these costs, thus offering equivalent functions using simpler software could help make those services more affordable and accessible.

Our work seeks to establish an effective deep learning model for detecting breast cancer lesions in mammography images. To achieve this, we combine the ResNet34 image classification model with the U-Net architecture for biomedical image segmentation. This method takes advantage of the strengths of U-Net in accurate and genuine segmentation (Bal-Ghaoui et al., 2023) and the ability of ResNet34 to reliably identify and localize lesions in mammography images, especially in situations with irregular forms and significant heterogeneity (Silalahi, 2021). The experimentation approach involved numerous stages, beginning with the search for a suitable dataset for the task at hand. Following a thorough review, we decided to use the INbreast dataset for transfer learning, because of its extensive and carefully curated mammography collection. To speed up the training process and boost efficiency, we employed the robust capabilities of the FastAI library, which is known for its user-friendly interface and outstanding functionality in deep learning tasks (Hubens et al., 2022). During the first model development experiment, we entailed various configurations of the U-Net architecture and incorporated transfer learning with INbreast. Preliminary results gave modest performance (accuracy = 0.586, sensitivity = 0.670 and specificity = 0.669), which required further examination for improvement.

With trust in the authenticity of the INbreast dataset, we looked into architectural refinement to unleash its full potential, exploiting nuanced insights to boost the model accuracy, sensitivity and specificity. The rest of this article is organized as follows: a detailed review of recent studies describing deep learning model used in breast cancer detection is discussed in Section 2. Section 3 consists in describing the database used and the model. Section 4 reviews its performance compared to existing similar experiments. At the end of the study, a conclusion and perspectives of our work are provided.

2 Related Works

Deep learning is a critical technology for future applications and is giving interesting solutions to medical analytic difficulties, because of its huge success in new inventions and filling the academic and medical gap. Among the medical fields, mammography has seen a great implementation of deep learning research. This section presents a detailed review of recent studies presenting deep learning models for lesion detection in mammography images. We present the studies grouped by the deep learning model topology used: FCNN, CNN, Multi-scale CNN, ResNet and Multi-view ResNet.

Liu et al. (2021) studied a deep learning (DL) model that integrated mammography and clinical variables to predict the malignancy of BI-RADS 4 microcalcifications in breast cancer screening (Figure 1).
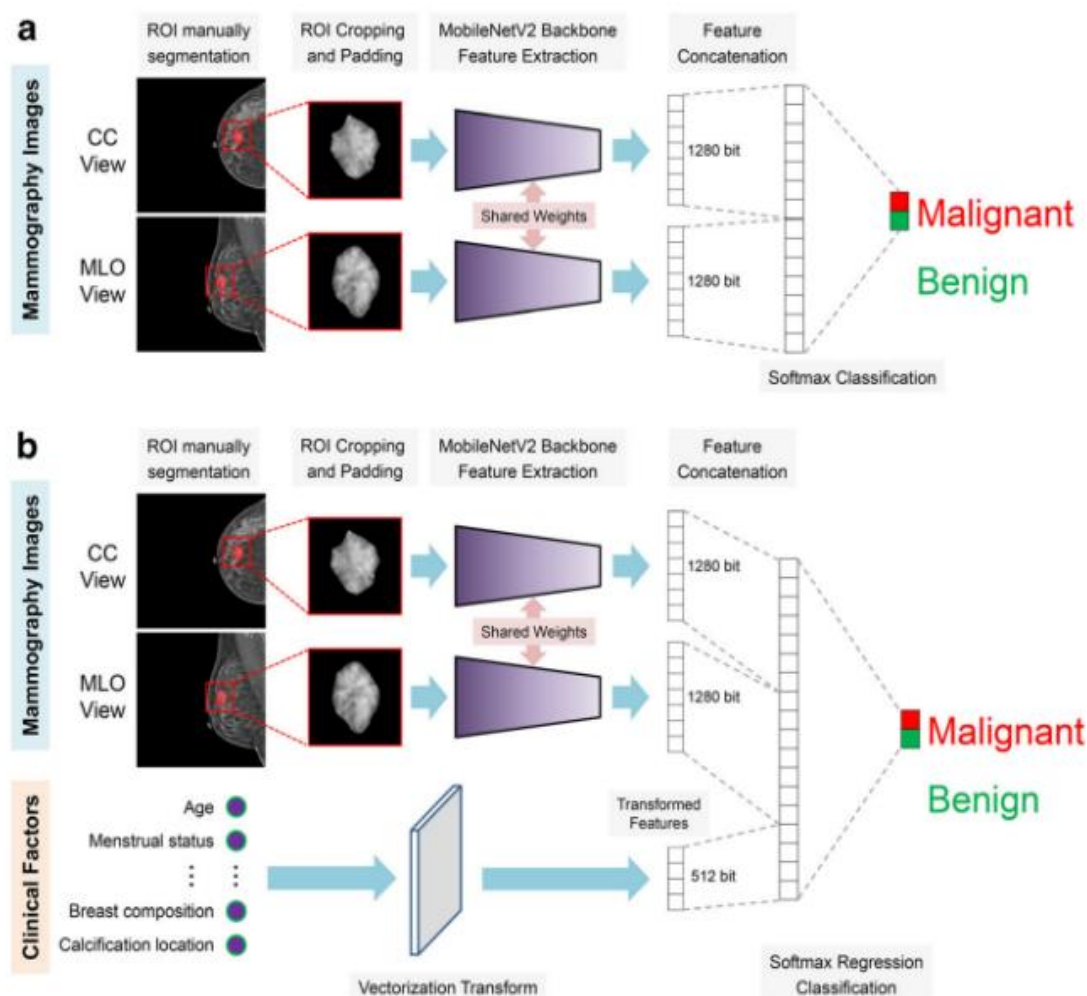

Figure 1. Schematic model illustration. Source: (Liu et al., 2021).

The researchers tested the model on a dataset consisting of mammograms taken from 71 patients, in which 71 lesions were detected: 37 malignant and 34 benign. The model is combined and incorporates mammography and clinical variables. Its evaluation was done with the area under the receiver operating characteristic curve (AUC) of 0.910, sensitivity of 85.3% and specificity of 91.9% in predicting malignant BI-RADS 4 microcalcifications in the testing dataset. The diagnostic performance of junior radiologists improved after AI assistance, with increased AUC and improved interobserver agreement.

As one of the remarkable limitations, as already cited in the paper, is that some patients' reports were excluded, which means that data cannot be considered representative of the existing population. Besides, the model utility in detection was somehow limited because of focusing on BI-RADS 4 mammographic microcalcification type.

Tsai et al. (2022) presented a deep neural network (DNN) model for BI-RADS classification of screening mammography (Figure 2).
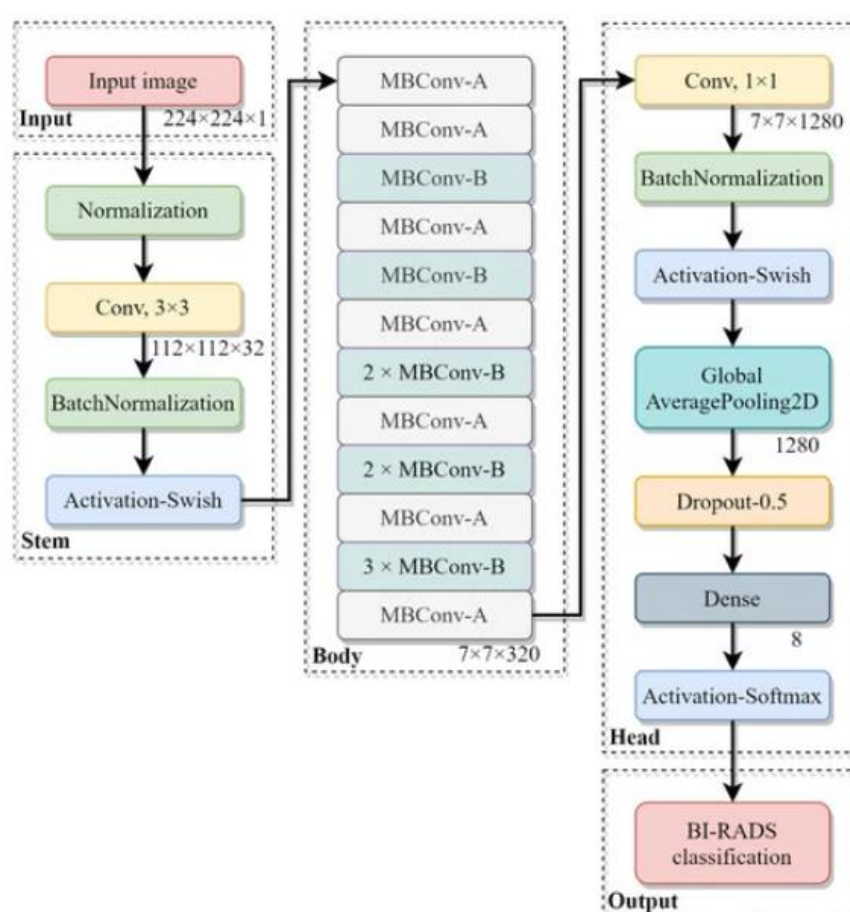

Figure 2. *Schematic illustration of BI-RADS classification model. Source: (Tsai et al., 2022).*

The proposed model was trained using block-based images segmented from a mammogram dataset of Taiwanese women. It achieved an overall accuracy of 94.22%, an average sensitivity of 95.31%, an average specificity of 99.15%, and an area under curve (AUC) of 0.9723. The work shared almost the same limitations as the previous paper, especially in restricting the model utility by focusing on only BI-RADS categories.

Ribli et al. (2018) created a Faster R-CNN model based on computer-aided detection in order to ensure whether a mammogram is malignant or benign (Figure 3).
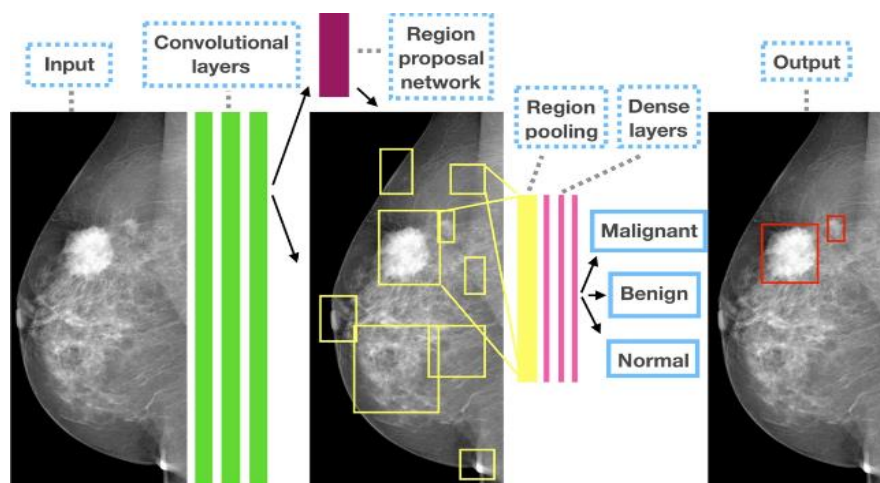

Figure 3. *Outline of Faster R-CNN model. Source: (Ribli et al., 2018).*

The proposed model represented a layering branch of convolutional layers known as a region proposal network (RPN), found above the last convolutional layer of the original network. It experienced training on two different databases: the public Digital Database for Screening Mammography (DDSM) with 2620 digitized screening mammography examinations and a dataset from the Semmelweis University in Budapest that contains 847 FFDM images of 214 examinations, besides tested on the public INbreast dataset with 115 FFDM cases with pixel-level ground truth annotations and histological proof for cancers. After the testing phase, the system performance reached the highest AUC score reported on the INbreast dataset with a fully automated system based on a single model at that time, with AUC = 0.95. The limitation of the model comes from the small size of the publicly available pixel-level annotated dataset that affected the classification performance.

Meanwhile, Kooi et al. (2017) presented a convolutional neural network model trained on a large dataset of mammographic lesions compared with some of the most recent ones in the mammography CAD system at the time, using a custom-created feature set (Figure 4). The purpose was to build a system which is able to read mammograms independently.
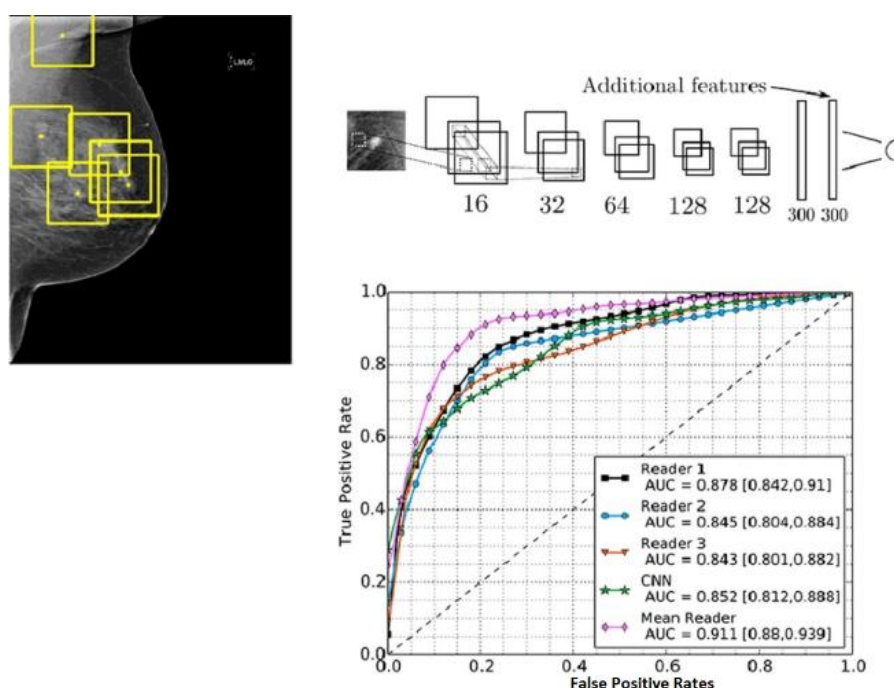

Figure 4. Graphical model abstract. Source: (Kooi et al., 2017).

Hologic Selenia digital mammography equipment was utilized to record the mammograms, which were obtained from a major screening programme in the Netherlands. In order to accelerate the process, images went from 70 to 200 microns, and to fend off information loss and bias, raw images were used and a log transform was exclusively proffered. The model used 44,090 mammographic views, which have been divided into 39,872 views to train the model and 4218 to validate it. As for testing, 18,182 images of 2064 patients with 271 annotated malignant lesions were taken. This model was compared with three experienced readers at a patch level, and it was shown that the human readers and CNN had the same capacity. As a result, this model outperformed the state-of-the art CAD system; therefore, it has considerable potential to advance the field of research. However, some CNN researchers proved its limitation for this task, because it may not generalize well to different imaging modalities or populations if the training data are not representative.

Hadad et al. (2017) applied a cross-modal fine-tuning approach on a convolutional neural network to identify masses in breast MRI images after training it on mammography X-ray images. The small size of the training set and the adoption of this strategy were justified by the lack of domain-specific pre-trained

models that could be applied without training the network. Cross-modal transfer learning has a positive effect on improving the classification performance in this case, producing a robust model able to fix issues from another domain or from a different imaging modality. The strategy used in cross-modal transfer learning is fine-tuning. As learning rates of other layers increase, certain network weights are fixed, and the decision layer is adjusted to the desired output size.

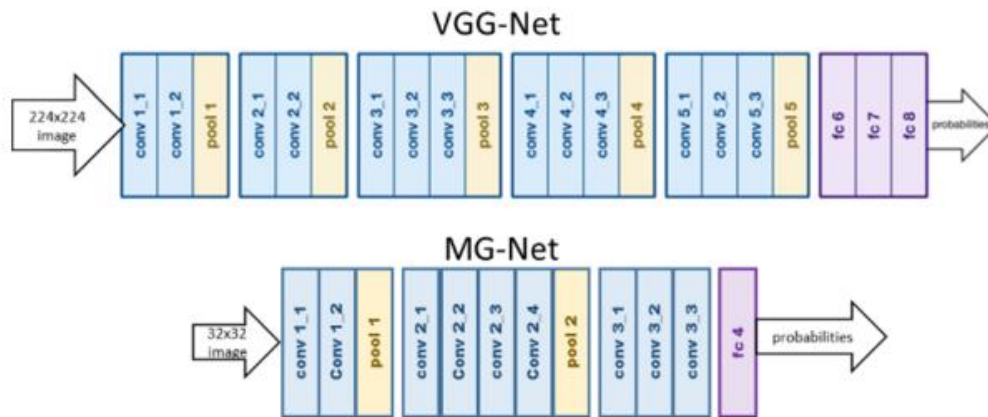Figure 5 describes the VGG-Net and MG-Net architectures.

Figure 5. *Network architectures of VGG-Net and MG-Net. Source: (Hadad et al., 2017)*

Two datasets were used in the experiment: the mammography (MG) dataset with full-field digital visuals, which were taken from 282 patients, and the MRI dataset including dynamic contrast-enhanced sequences with data of 123 patients. Both datasets were reviewed by a breast radiologist, identifying mass of lesions and delineating their limitations with the assignment of each lesion with a BI-RADS score. The experiments involved three types of evaluation: training a CNN from scratch using MR images (achieved accuracy of 0.94 and AUG of 0.98), fine-tuning a VGG-Net CNN (it reached accuracy of 0.90 and AUG of 0.95) and a fine-tuning a CNN on MG-Net (with accuracy of 0.93 and AUG of 0.97). As already mentioned in the original paper, the model faced some limitations while working on small sizes of datasets.

Lotter et al. (2017) proposed a multi-scale convolutional neural network model trained with a curriculum learning strategy. It proceeded in two stages, starting with training a patch classifier, followed by a full image classification (Figure 6).
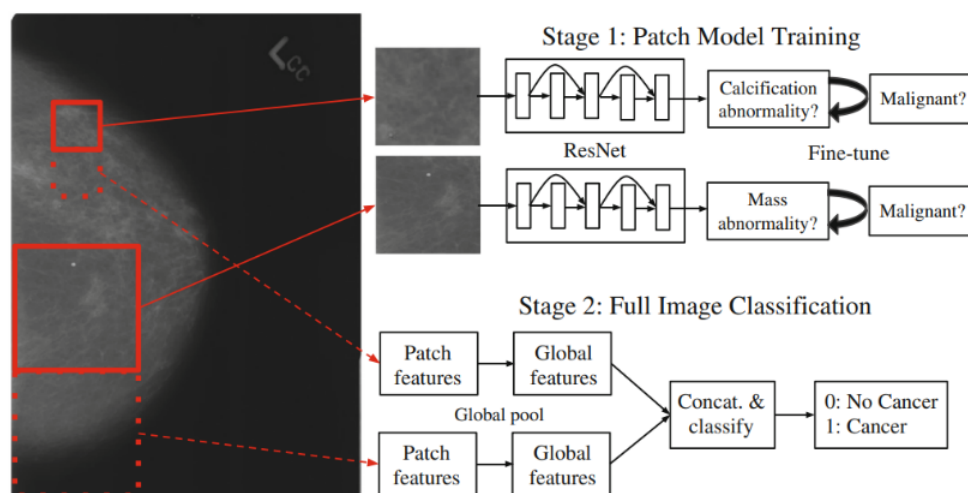

Figure 6. *Schematic of multi-scale CNN and curriculum learning. Source: (Lotter et al., 2017).*

The model was evaluated on the Digital Database for Screening Mammography. That was separated into 87% for training, 5% for validation and 8% for testing. This choice was made to increase the amount of

data to be trained, while ensuring a suitable interval for the last results. The experiment started with training segmentation masks of lesions from mammograms used to train CNN-based patch classifiers, and after that, the learned features were used to initialize a scanning-based model that rendered a judgment on the entire image and is trained end-to-end on outcome data. The results showed that the model achieved AUROC of 0.92.

De Moor et al. (2018) worked on a convolutional neural network model with modified U-Net architecture to detect and segment malignant lesions from digital mammography images. The contribution consisted in doubling the filter numbers of each block and applying batch normalization (Figure 7).
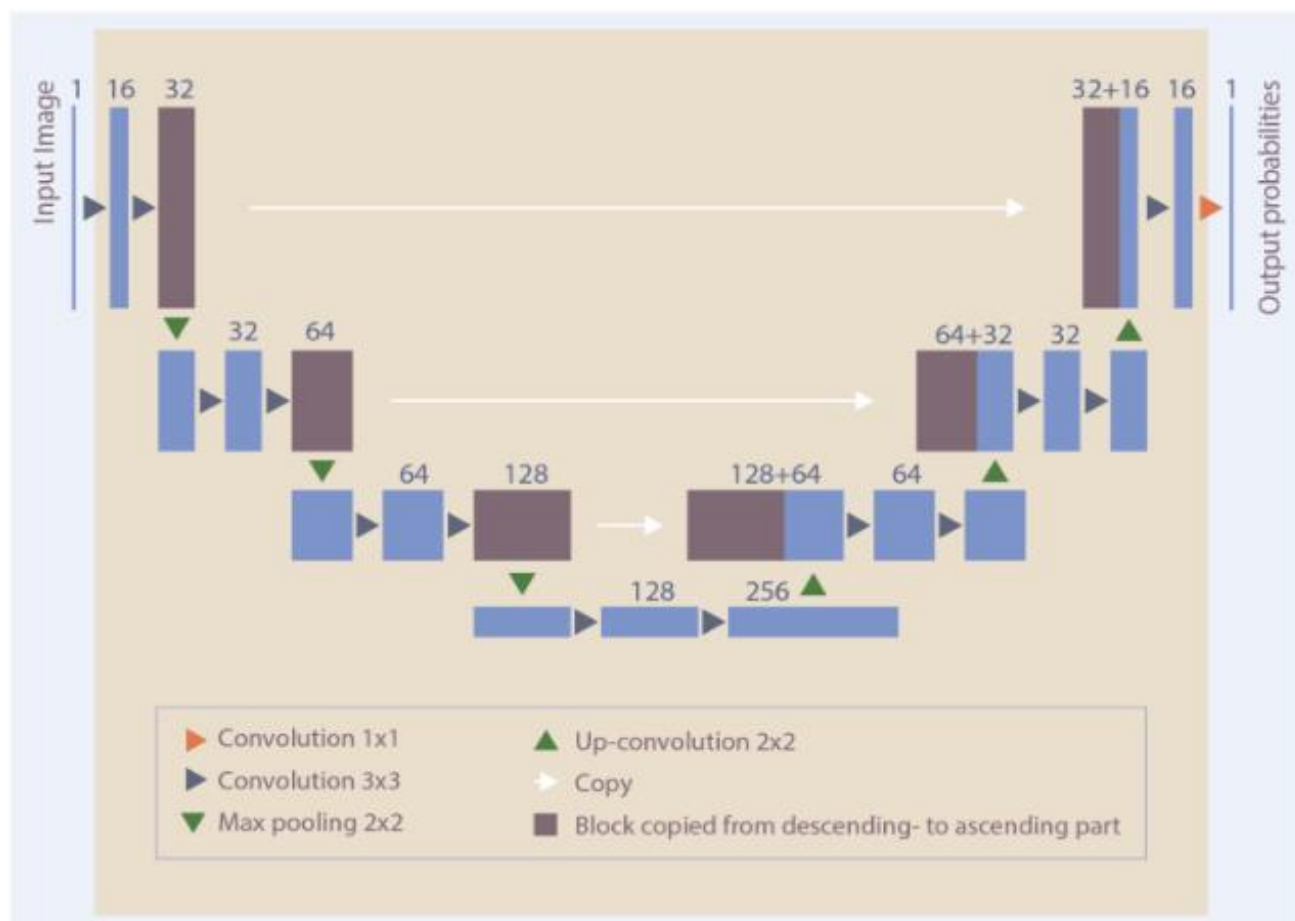

Figure 7. Diagram describing U-Net architecture with study contribution. Source: (de Moor et al., 2018).

The database used contained 7196 digital mammography examinations with 28,294 images, which were divided in a random way into training (50%), validation (10%) and testing (40%). Its performance was evaluated with free receiver operating characteristic (FROC) analysis, which gave a result of sensitivity of 0.98 and false-positive rate per image of 7.81 in the examination-based FROC. The main limitation of the model is that it only studied soft tissue lesions.

Dhungel et al. (2017) made a multi-view deep residual neural network (mResNet) model, created for fully automated classification of mammograms (Figure 8).
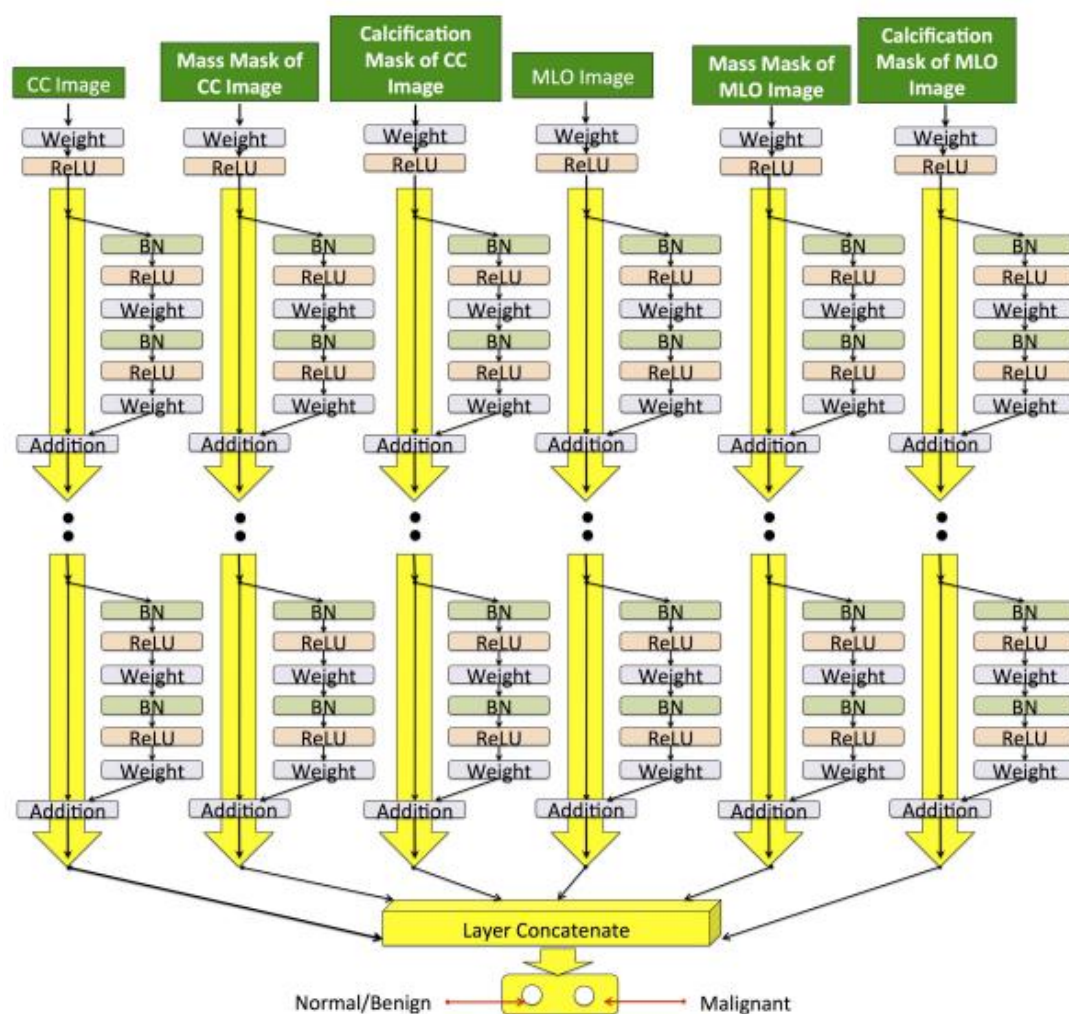
Figure 8. *Diagram illustrating model' classification and mask generation process. Source: (Dhungel et al., 2017).*

The model was formed by a sequence of outputs of a group of associated deep residual networks, which had six input images, including unregistered craniocaudal (CC) and mediolateral oblique (MLO) mammogram views and associated automatically detected lesions. The publicly available INbreast dataset, which comprises 116 cases containing 410 images, was used for the experiment, and a five-fold cross-validation technique was used to split the dataset. As a result, it produced an AUC of 0.8. The primary limitation of the work is that due to concerns with proprietary intellectual property, a number of algorithms and results were not accessible in the open literature.

Al-Antari et al. (2018) presented a deep learning model using YOLO for mass detection, FrCn for mass segmentation and CNN for mass classification (Figure 9).
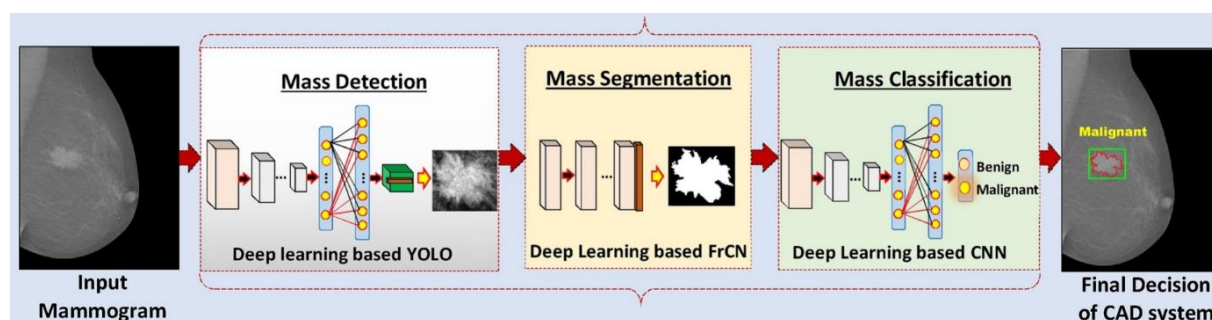

Figure 9. *Diagram illustrating model process. Source: (Al-Antari et al., 2018).*

Their model used the INbreast database of X-ray mammography to test its performance. For training, a large amount of annotated dataset was required, so data augmentation was used to increase the size of the dataset, giving a total of 896 mammograms. The evaluation results of the proposed CAD system via four-fold cross-validation tests showed a mass detection accuracy of 98.96%, a mass segmentation accuracy of 92.97% and a mass classification accuracy of 95.64%. The only limitation is that it requires a large dataset to function properly.

Al-Masni et al. (2018) proposed a novel CAD system based on ROI-based convolutional neural network (CNN) which is called You Only Look Once (YOLO). This model passes through four phases: preprocessing, feature extraction, mass detection and finally mass classification (Figure 10).
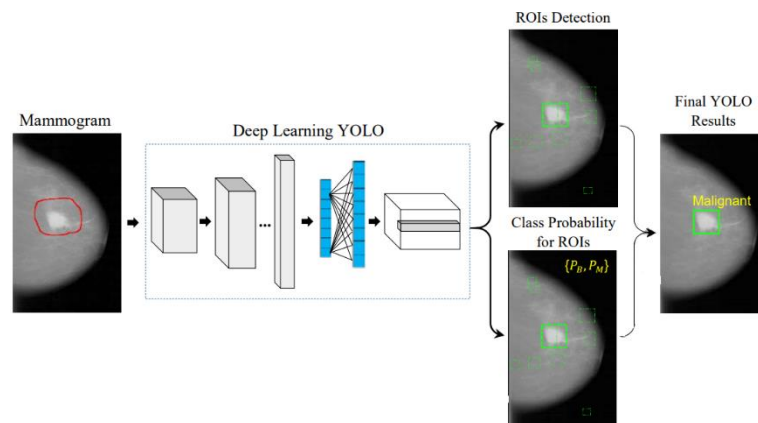

Figure 10. Diagram describing YOLO process. Source: (Al-Masni et al., 2018)

The Digital Database for Screening Mammography (DDSM) was used for training and testing the proposed model, its small size required the augmentation technique to increase the training data. As a result, this CAD system had an accuracy of mass location detection of 99.7% and classification accuracy of 97%. Just like the previous work (this experiment was done by the same team), the model requires a large dataset to function effectively.

Platania et al. (2017) used convolutional neural network model called BC-DROID for automated breast cancer detection and diagnosis (Figure 11).
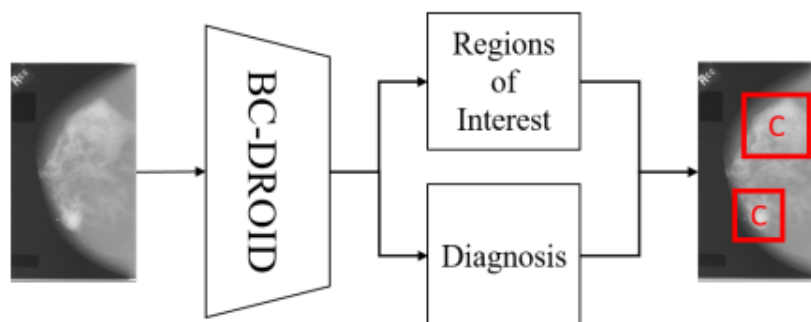

Figure 11. Diagram depicting BC-DROID process. Source: (Platania et al., 2017).

Image Retrieval in Medical Applications (IRMA) was used, plus some additional metadata, which were provided by the original DDSM dataset. The deep neural network model has two principal components. The first one is a stack of 3x3 convolutional layers, to extract the features. The second component is three fully connected layers, which work on predicting the output. This model achieved a detection accuracy of up to 90% and a classification accuracy of 93.5% (AUC of 92.315%). It is considered the first work enabling both automated detection and diagnosis of these areas in one step from full mammogram images. The research revealed that the model struggles with detecting objects that are very small or very close together, because of the fixed spatial resolution characteristic.

Following is a summarized table of the listed works, which highlights the key aspects of each study, including the type of model, datasets, performance metrics and notable limitations (Table 1).

Table 1. Key aspects of each mentioned study.

Study	Model type	Dataset	Performance	Limitations
Liu et al. (2021)	Combined DL model	71 patients, 71 lesions	AUC: 0.910, sensitivity: 85.3%, specificity: 91.9%	Not representative of the population, focused on BI-RADS 4 microcalcifications
Tsai et al. (2022)	DNN-based model	Taiwanese women mammograms	Accuracy: 94.22%, sensitivity: 95.31%, specificity: 99.15%, AUC: 0.9723	Restricted to BI-RADS categories
Ribli et al. (2018)	Faster R-CNN	DDSM, Semmelweis University, INbreast	Highest AUC on INbreast: 0.95	Small size of publicly available pixel-level annotated dataset affecting classification performance
Kooi et al. (2017)	CNN	44,090 mammographic views, 18,182 images for testing	Comparable to human readers	CNNs may not generalize well to different imaging modalities or populations
Hadad et al. (2017)	CNN with cross-modal transfer	Mammography X-ray and MRI datasets	Accuracy: 0.94 (training), 0.90 (VGG-Net fine-tuning), 0.93 (MG-Net fine-tuning)	Small dataset sizes
Lotter et al. (2017)	Multi-scale CNN	DDSM	AUROC: 0.92	No specific limitations mentioned
de Moor et al. (2018)	Modified U-Net CNN	7,196 digital mammography examinations, 28,294 images	Sensitivity: 0.98, false positive rate per image: 7.81	Focused only on soft tissue lesions
Dhungel et al. (2017)	Multi-view deep residual network	INbreast dataset, 116 cases, 410 images	AUC: 0.8	Proprietary algorithms and results not publicly accessible
Al-Antari et al. (2018)	YOLO, FrCn, CNN	INbreast database	Detection accuracy: 98.96%, segmentation accuracy: 92.97%, classification accuracy: 95.64%	Requires large dataset for proper functioning
Al-Masni et al. (2018)	ROI-based CNN (YOLO)	DDSM	Mass location detection accuracy: 99.7%, classification accuracy: 97%	Requires large dataset for proper functioning
Platania et al. (2017)	CNN (BC-DROID)	IRMA, additional DDSM metadata	Detection accuracy: 90%, classification accuracy: 93.5%, AUC: 92.315%	Struggles with very small or closely located objects due to fixed spatial resolution characteristic

After reviewing different methods applied to detecting lesions in mammography images, it can be concluded that there is a wide variety of methods that have been used. We cannot confirm any of the methods as being the most suitable, because each one has its own advantages and limitations, and the information available is very variable in quantity and quality. Also, there are many public and private datasets, thus comparing their performance is quite difficult, since each study used a different dataset. Besides, the large number of applications presented shows the interest of researchers in this field.

3   Proposed Model

A deep neural network is a neural network which contains more than two layers. It treats given data in a complex procedure by applying sophisticated mathematical modelling. Simonyan & Zisserman (2014) affirmed that that network depth is very important, and deep networks naturally integrate low/mid/high level features (Zeiler and Fergus, 2014) with an end-to-end multilayer classification besides the possibility of augmentation of feature levels according to the network depth. In the convergence stage, while the depth increases, the accuracy gets saturated and then quickly decreases. This is known as vanishing gradient. The gradients ensure the calculation of the loss function, which easily shrinks to zero after several applications of the chain rule. This change is not provided by overfitting, it is due to the additional layers, which leads to increasing the training error rate as reported by He and Sun (2015) and Srivastava et al. (2015).

For this reason, a deep residual learning framework was created with layers that apply learning residual functions, which offers easier training for the networks that have additional layers compared to the old ones, besides giving the model the ability to acquire accuracy regardless of the depth, and the flexibility to any space increases because of the parameter space to explore. The name ResNet34 refers to the fact that its total number of weighted layers is 34 (Figure 12).
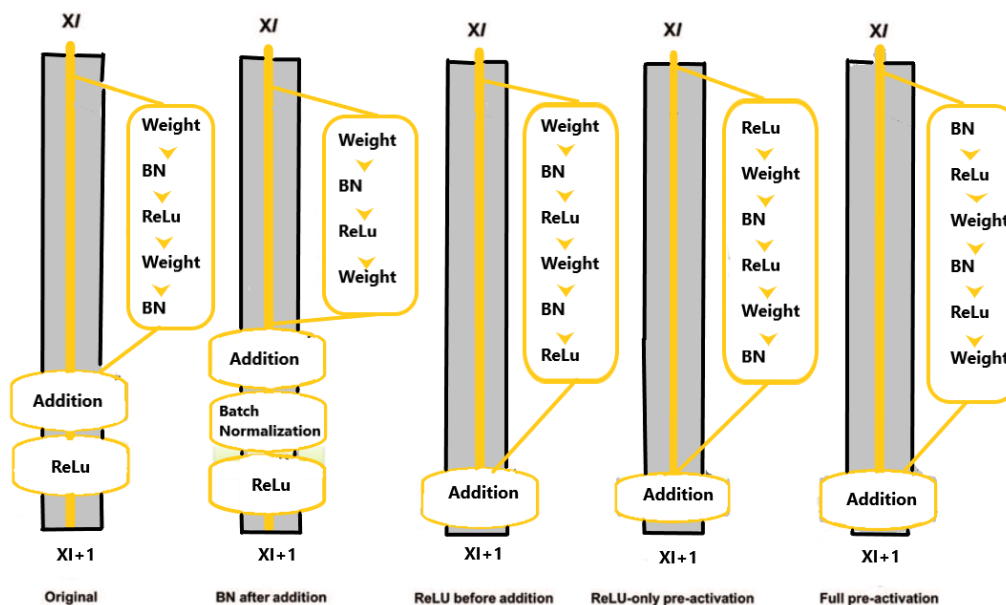

Figure 12. *Basic architecture of ResNet34.*

Meanwhile, ResNet34 architecture was inspired by the VGG networks: it consists of one convolutional and pooling step, followed by 4 layers of the same attitude with 3x3 filters (Figure 13), based on these rules:

- The layers have an equal number of filters for the same output feature map size.
- To maintain the time complexity per layer if the feature map size is cut in half, there must be twice as many filters (He et al., 2016).

A downsampling is performed by convolutional layers that have a stride of 2. The network ends with a global average pooling layer and a 1000-way fully-connected layer with softmax (Figure 12).
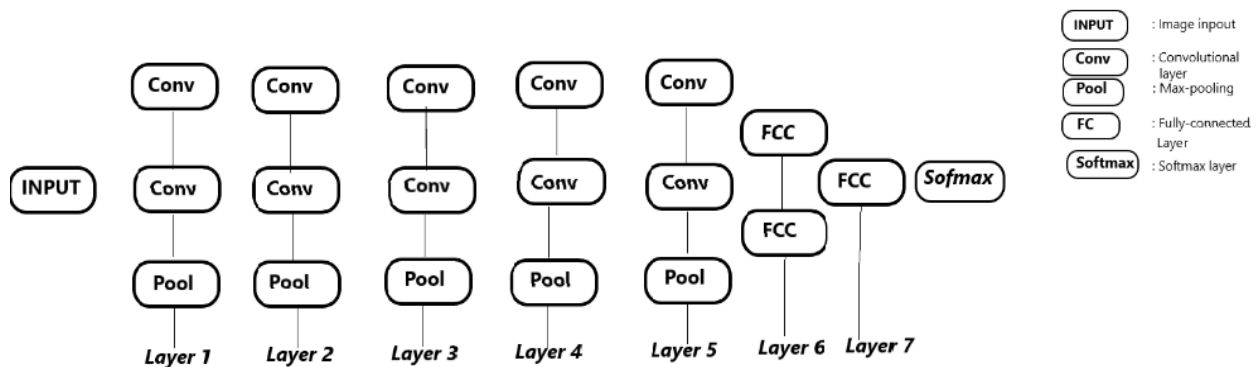

Figure 13. *Basic VGG-Net architecture.*

As we work on building a deep learning model to detect lesions in mammography images, we believe that ResNet34 suits the purpose: it won the 1st place in the tasks of ImageNet detection, ImageNet localization, COCO detection and COCO segmentation (He et al., 2016).

3.1　Architecture

Deep convolutional networks have made great success in visual recognition, but it has been limited by the size of datasets and networks. For this case, U-Net appears to give a hand to some types of deep neural networks that demand numerous annotated samples to provide successful training; thanks to data augmentation with elastic deformations, working with a limited number of annotated samples is sufficient.

In biomedical image processing, each pixel is class-labelled; following this, Ciresan et al. (2012) trained a network for pixel prediction in a sliding window. The U-Net team observed some weaknesses: the network worked slowly and there was a trade-off between localization accuracy and the use of context. Based on this issue, U-Net was created.

In this work, we used the U-Net network. As mentioned above, it was developed by Olaf Ronneberger, Philipp Fischer and Thomas Brox for biomedical image segmentation (Ronneberger et al., 2015). Its architecture is composed of two paths: firstly, the contraction path (encoder) extracts features holding the information of what is in the image using 3x3 convolutionals, each one followed by a linear unit (Relu) and a 2x2 max pooling layer with stride 2 for downsampling. During this step, the feature map is reduced. Secondly, the expansion path (the decoder), is for recovering the feature map size for the segmentation image, using up-convolution, which is an upsampling followed by 2x2 convolutions and 3x3 convolutions. U-Net won the ISBI cell tracking challenge in 2015. Also, this network is fast, segmentation of a 512x512 image lasts less than a second using a recent GPU, as shown by Ronneberger et al. (2015).

As our purpose is to develop a deep learning model for detecting lesions in mammography images, we find ResNet34 to be a suitable choice as the encoder within the selected U-Net architecture. Thus, introducing and developing the U-Net structure with ResNet34 as the encoder (Figure 14) is considered a novelty in detecting lesions in mammography images. Additionally, using tools such as the FastAI library and employing transfer learning with the INbreast dataset adds more innovation to our model. This approach has demonstrated its efficacy in medical segmentation tasks, as evidenced by papers specializing in lung image segmentation (Biyyala, 2023), pneumothorax segmentation in chest X-rays (Abedalla et al., 2020) and optical coherence tomography layer segmentation (Yojana and Rani, 2023). Encouraged by these successes mainly using the U-Net architecture, we embarked on a study to apply this approach to a new problem domain, achieving remarkable results when compared to recent works.
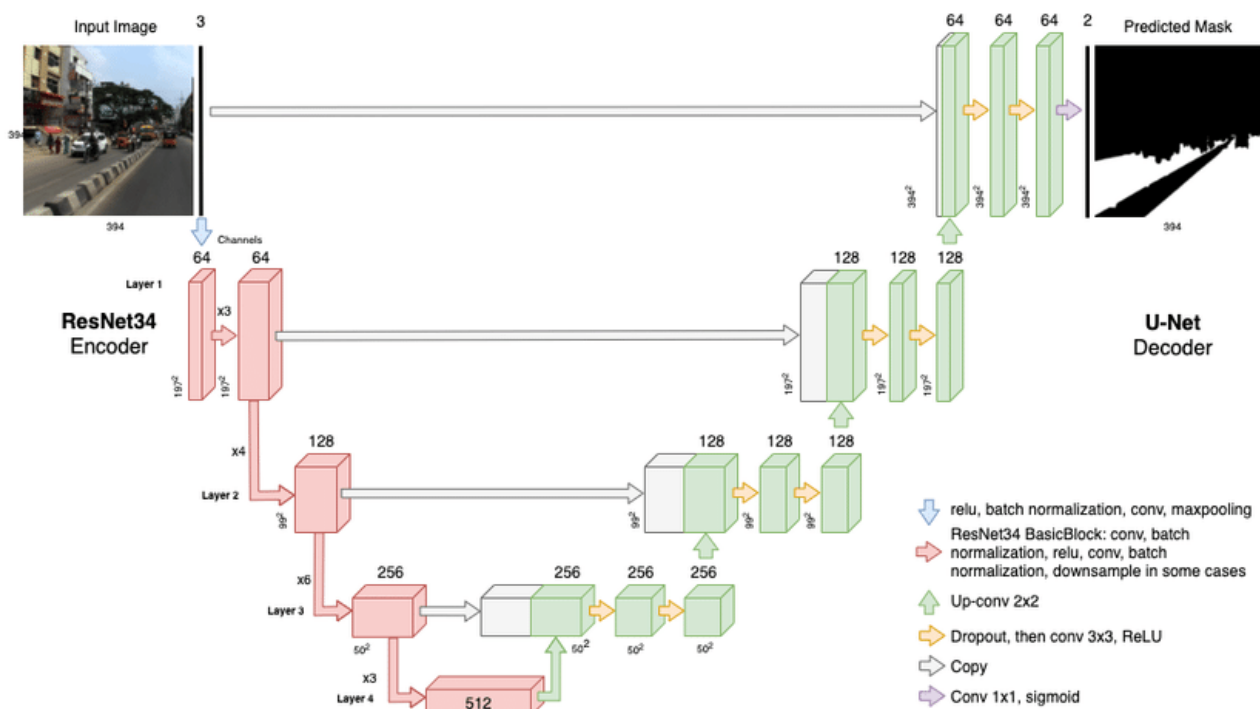
Figure 14. *U-Net architecture + ResNet34 implemented with PyTorch for road segmentation.*
Source: (Salcedo et al., 2022).

3.2 Dataset

Breast screening uses an X-ray test called mammogram, which detects malignant lesions that are impossible to see or feel. The earlier the condition is found, the better the patient's chance to avoid breast removal (mastectomy) or chemotherapy. This mammogram needs to be taken at a special clinic or mobile breast screening unit. This is done by a female health practitioner. The breasts are X-rayed one at a time, by placing them on the X-ray machine and gently but firmly compressing with a clear plate. Two X-rays are taken of each breast separately at different angles. Currently used computer-aided detection approaches are based on machine learning for classification on top of manually and precisely designed features that describe an X-ray image. Breast cancer does not always have readily observable symptoms because of the tiny tumour portions that appear during mammography screening, which makes the classification generally unclear. For example, a full-field digital mammography image is typically 4000x3000 pixels, although the region of concern for cancer may be as little as 100x100 pixels (Shen et al., 2019). That is why numerous studies have focused solely on the classification of annotated lesions.

The dataset used was obtained at the Breast Centre in CHSJ, Porto, under permission of both the hospital's ethics committee and the National Committee of Data Protection. It was selected between April 2008 and July 2010, the acquisition equipment was Siemens MammoNovation full-field digital mammography, with a solid-state detecting tool of amorphous selenium, pixel size of 70 microns, and 14-bit contrast resolution (Moreira et al., 2012). The image matrix was 3328x4084 or 2560x3328 pixels, depending on the patient's breast size. Images were saved in the DICOM format.

A total of 115 cases were taken, where 90 had two images, namely the mediolateral oblique view and the craniocaudal view (Figure 15), of each breast and the other 25 cases were of women who had had a mastectomy and two views of only one breast were included. The dataset contains a total of 410 images, 8 of the 91 cases with two views of the organ; besides, it was taken in separate sessions. The sample cases included: normal mammograms, mammograms with masses, mammograms with calcifications, architectural distortions (Figure 16), asymmetries and images with multiple findings (Moreira et al., 2012). While radiologists were examining and comparing the two breasts and the two views, they were marking

assessments about the abnormalities found and the recommended actions to be taken in each case. For this purpose, the American College of Radiology established the Breast Imaging Reporting and Data System (BI-RADS) scale to standardize the terminology of the mammographic report, the assessment of findings and the recommended actions to be taken.
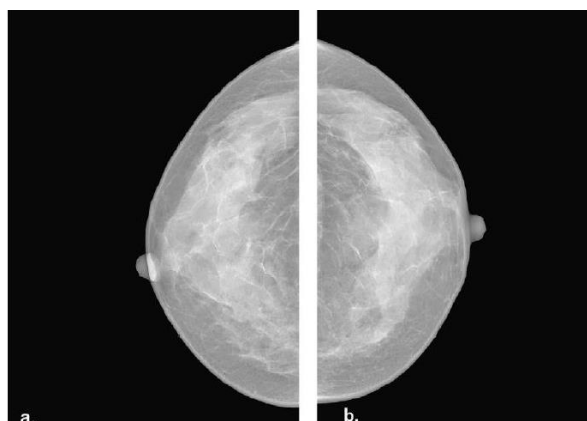

Figure 15. (*a) Craniocaudal (CC) view of the right breast; (b) CC view of the left breast.
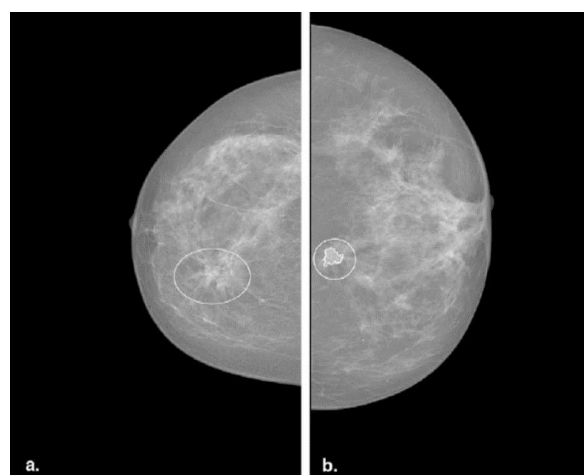Source: (Moreira et al., 2012).*

Figure 16. *Annotation examples: (a) distortion (b) spiculated region. Source: (Moreira et al., 2012).*

Based on the level of suspicion, abnormalities were placed into one of six BI-RADS categories, as shown in Table 2. This dataset was chosen after studying many other databases. Its robustness depends on it being full-field digital mammograms, contrary to digitized mammograms; it presents a tremendous variation of cases, and is considered publicly available with precise annotations. Thus, the coverage of such technological development presents a decisive step to develop future CADs.

Table 2. *BI-RADS assessment categories. Source: (Moreira et al., 2012).*

Category	Description
0	Needs additional image evaluation and/or prior mammograms for comparison.
1	Negative.
2	Benign finding(s).
3	Probably benign finding(s). Short-interval follow-up is suggested.
4	Suspicious anomaly. Biopsy should be considered.
5	Highly suggestive of malignancy. Appropriate action needs to be taken.
6	Biopsy-proven malignancy.

4   Experiment

4.1   Library

FastAI is a research lab whose objective is to make AI accessible by offering an easy-to-use library built on top of PyTorch, as well as exceptionally good tutorials. After much research, we chose the FastAI library because it is considered a high-level library, which requires only a few lines of code to manipulate a model. Besides, it provides an implementation of the latest state-of-the-art techniques extracted from research papers, achieving state-of-the-art results for nearly any problem.

4.2   Evaluation metrics

Accuracy (ACC), sensitivity (SEN), specificity (SPE) and F1-score were chosen to describe the model performance. Accuracy (ACC) is determined by dividing the number of right predictions by the total. It is defined by the following equation:

$$Accuracy = \frac{True_{Positives} + True_{negatives}}{True_{Positives} + True_{negatives} + False_{Positives} + False_{negatives}} \tag{1}$$

Sensitivity (SEN), also known as recall or true positive rate, measures the ability of a classification model to correctly identify positive instances among all actual positive instances. It is defined by the following equation:

$$Sensitivity\ (recall) = \frac{True_{Positives}}{True_{Positives} + False_{negatives}} \tag{2}$$

Specificity (SPE) measures the ability of a classification model to correctly identify negative instances among all actual negative instances. It is defined by the following equation:

$$Specificity = \frac{True_{negatives}}{True_{Negatives} + False_{Positives}} \tag{3}$$

The F1-score metric is the harmonic mean of precision and recall (sensitivity), which provides a balance between precision and recall, computing the correct predictions made by the model. It is defined by the following equation:

$$F1\text{-}score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \tag{4}$$

4.3   Results

The experiment started by data upload, which was of different types, making us search for the appropriate method to deal with each type separately. We also had to convert the data to an easier and more convenient format in order to obtain data that are in line with the following approaches. After that, we entered the learning phase, starting with installing libraries and obtaining some methods that helped accomplish the task. Next, we trained the model without transfer learning and then with transfer learning to see what these changes would bring to the results.

We started with the basic version of U-Net for breast cancer lesion detection. After testing, we noticed some limitations (Figure 17) that led us to search for a better solution. As mentioned above, we settled for ResNet34 as an encoder for U-Net, and that contributed to enhancing the model performance (Figure 18).
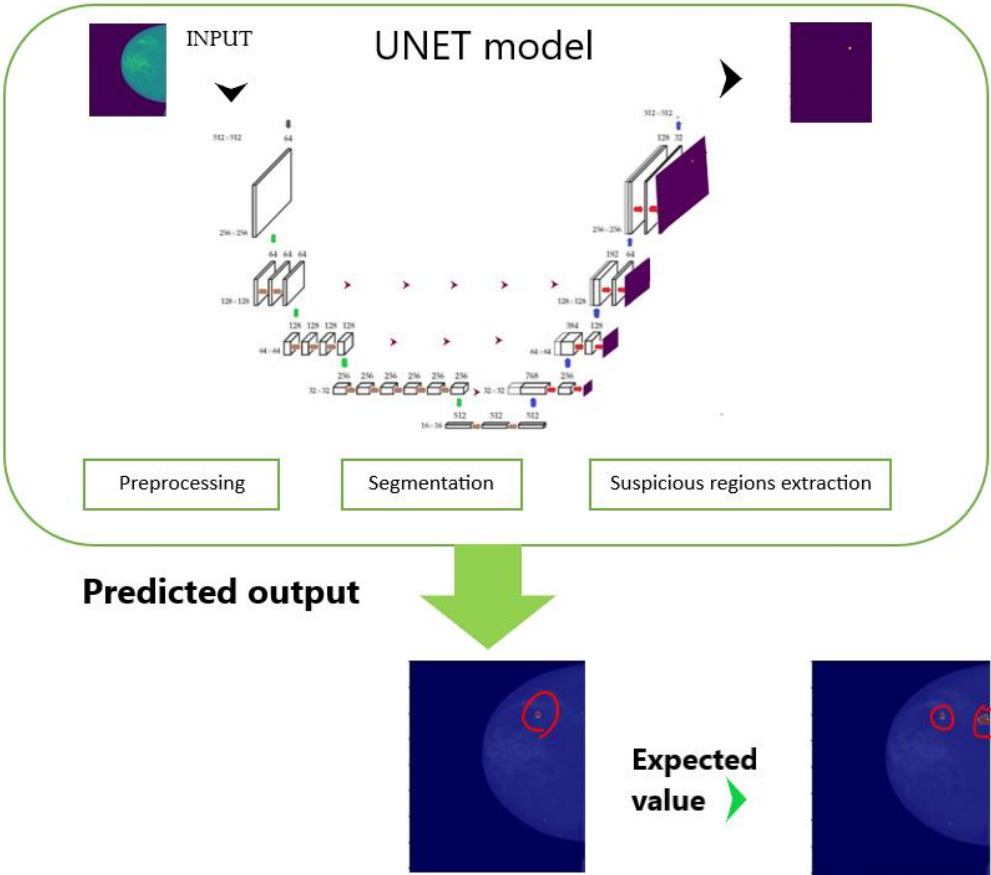
Figure 17. *Procedural workflow of U-Net, besides its performance (segmentation and mask extraction), in which some limitations occurred.*
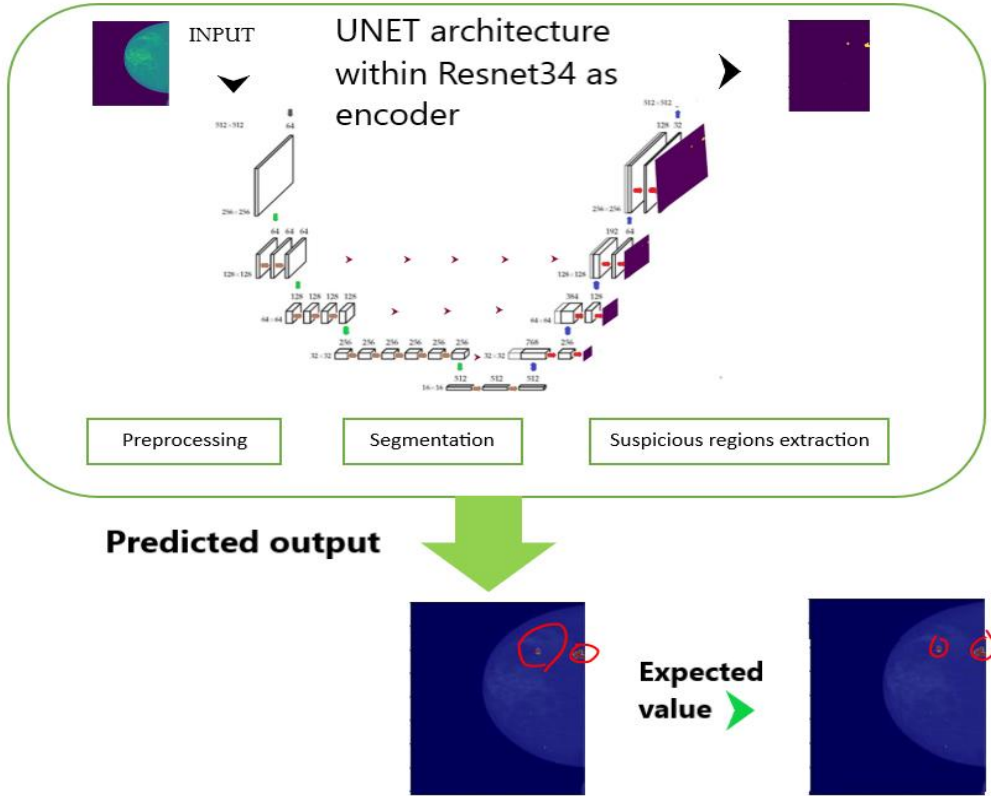

Figure 18. *Model procedural workflow and performance using U-Net with ResNet34 as encoder, which helped fix the previous error.*

We would like to note that the validation set and the training set were created by providing the validation set with 20% of the prepared dataset images. As a result, we got 275 images for training and 68 for validation. Our model achieved an accuracy of 0.891 with a validation loss that went down to 0.089 without transfer learning. However, transfer learning with the pretrained U-Net encoder on the Imagenet dataset achieved an accuracy of 0.9890 and a validation loss of 0.031. Moreover, we noticed that the pretrained model had made a remarkable difference on both the result and the network learning speed, which performed better than the non-pretrained model. The results are summarized in Table 3.

Table 3. Model performance.

Method	Environment	Dataset	Training results	Validation results
U-Net	Google colaboratory	INbreast	Accuracy = 0.717, Sensitivity = 0.801, Specificity = 0.730, F1-score = 0.775	Accuracy = 0.586, Sensitivity = 0.670, Specificity = 0.669, F1-score = 0.603
U-Net with ResNet34 as encoder	Google colaboratory	INbreast	Accuracy = 0.979, Sensitivity = 0.983, Specificity = 0.974, F1-score = 0.981	Accuracy = 0.986, Sensitivity = 0.970, Specificity = 0.969, F1-score = 0.983

Furthermore, we studied the impact of the following hyperparameters on the model performance: epoch and loss function. Table 4 shows the obtained results.

To identify the best epochs for optimal performance, metrics were monitored across different epochs of the training process. Table 4 describes the task of choosing the best one in which the model performs the best. Here, the 10th epoch stands out as the best one based on the provided metrics. There, the model showcased the lowest loss function value = 0.21, the highest values of accuracy = 0.98, sensitivity = 0.97, specificity = 0.96 and F1-score = 0.95. That indicates that it had converged well and effectively minimized errors during training. This finding contributed to speeding up the training process while preventing overfitting that could be caused by using more epochs.

Table 4. Model performance in different epochs, according to evaluation metrics.

Epoch	Loss function	Accuracy	Sensitivity	Specificity	F1-score
5	0.22	0.97	0.93	0.91	0.93
10	0.21	0.98	0.97	0.96	0.95
50	0.23	0.96	0.95	0.96	0.94
100	0.23	0.95	0.89	0.94	0.92
150	0.29	0.93	0.87	0.91	0.92
200	0.24	0.94	0.89	0.92	0.91

4.4 Comparative study

Table 5 lists the performance of our model on both the training set and the validation set, compared to some previous studies on breast cancer detection, using the INbreast dataset (80% for training and 20% for validation) and accuracy as an evaluation metric in order to reveal the contribution of this work.

Table 5. Comparison between our model and recent studies.

Method	Evaluation dataset	Training set (80%)	Validation set (20%)
This study	INbreast	Accuracy = 0.979, sensitivity = 0.983, specificity = 0.974, F1-score = 0.981	Accuracy = 0.986, sensitivity = 0.970, specificity = 0.969, F1-score = 0.983
DNN-based model (Tsai et al., 2022)	INbreast	Accuracy = 0.962, sensitivity = 0.950, specificity = 0.97 F1-score = 0.949	Accuracy = 0.952, sensitivity = 0.943, specificity = 0.981, F1-score = 0.944
Automated deep learning empowered breast cancer diagnosis technique (Escorcia-Gutierrez et al., 2022)	INbreast	Accuracy = 0.940, sensitivity = 0.973, specificity = 0.964, F1-score = 0.970	Accuracy = 0.978, sensitivity = 0.950, specificity = 0.959, F1-score = 0.951
Full-field digital mammography-based deep learning (Liu et al., 2021)	INbreast	Accuracy = 0.939, sensitivity = 0.901, specificity = 0.930, F1-score = 0.900	Accuracy = 0.956, sensitivity = 0.853, specificity = 0.919, F1-score = 0.890

As Table 5 shows, we tested different DL models, which were similar to our project, to determine the robustness of these models, and especially ours. We picked 80% and 20% of the INbreast dataset respectively for training (275 images) and validation (68). The results show that our model performed better than others, with an accuracy of 98%, a sensitivity of 98% and a specificity of 98% while training, and an accuracy of 97%, a sensitivity of 97% and a specificity of 96% in the validation test. The DNN-based model achieved an accuracy of 96%, a sensitivity of 95% and a specificity of 97% in training, and an accuracy of 95%, a sensitivity of 94% and a specificity of 98% in the validation test. However, the automated deep learning-empowered model reached a 94% accuracy, 97% sensitivity and a specificity of 96% in training, while in validation it reached 97%, 95% and another 95% in accuracy, sensitivity and specificity respectively. Last but not least, an accuracy of 93%, a sensitivity of 90% and a specificity of 93% was the performance of the full-field deep learning model in the training test with an accuracy of 95%, a sensitivity of 0.85% and a specificity of 91% in the validation test. Overall, there was not a significant difference between our model and other DL models except for the training timing, thanks to the FastAI library, which remarkably improved testing timing. These results give the privilege to our model to be considered a well-performing deep learning model in detecting lesions in mammography images. The obtained accuracy can be explained by the high discriminative features extracted using the U-Net model, reinforced by a data augmentation technique that provides more samples in order to deal with the limited labelled data of the INbreast database of mammograms. Moreover, the time consumption of the proposed method is not high since 10 epochs were sufficient for the convergence of the model.

5 Conclusion

In this work, we explored deep learning advances in the medical field, especially in breast cancer diagnosis. We chose the INbreast database to explore approaches and methods. Image segmentation was investigated as a method for lesion detection in this problem. We used a well-known architecture in the biomedical field: U-Net, and ResNet34 as the encoder of the network. This article describes some steps and methodologies that were taken for the implementation of the system as well as the chosen approaches.

A large part of the work consisted of data preparation for the learning algorithm and using the FastAI neural network library for training the neural network. It is worth mentioning that this work provides an efficient method for breast cancer segmentation and classification based on the U-Net architecture, reinforced by data augmentation techniques to deal with the limited number of labelled samples. Our tool aims to assist radiologists with mammographic interpretation in clinical works and ameliorate mammogram interpretation efficiency as well. As a consequence, it leads to a reduction in the radiologist's workload and helps in the case of shortage of radiologists. In future work, the generative adversarial network (GAN) can be added to the proposed method to improve the results.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: M.A.Y.: Conceptualization, Methodology, Writing – original draft, Validation, Writing – review & editing. N.K.: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. A.K.: Investigation, Software, Validation, Writing – review & editing. A.M.: Supervision, Validation, Writing – review & editing.

Data Availability: The authors used a previously published dataset from (Moreira et al., 2012).

References

Abedalla, A., Abdullah, M., Al-Ayyoub, M., & Benkhelifa, E. (2020). 2ST-UNet: 2-Stage Training Model using U-Net for Pneumothorax Segmentation in Chest X-Rays. In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. https://doi.org/10.1109/ijcnn48605.2020.9207268

Akhund, J., Memon, S., Qamber, J., Rashid, I., Elhaj, A., & Alshammari, A. (2023). Ultrasound vs. Mammography in Evaluating Suspicious Breast Lesions. *Biological & Clinical Sciences Research Journal*, 2023(1), 288. https://doi.org/10.54112/bcsrj.v2023i1.288

Al-Antari, M. A., Al-Masni, M. A., Choi, M., Han, S., & Kim, T. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics*, 117, 44–54. https://doi.org/10.1016/j.ijmedinf.2018.06.003

Al-Masni, M. A., Al-Antari, M. A., Park, J., Gi, G., Kim, T., Rivera, P., Valarezo, E., Choi, M., Han, S., & Kim, T. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer Methods and Programs in Biomedicine*, 157, 85–94. https://doi.org/10.1016/j.cmpb.2018.01.017

Bal-Ghaoui, M., Alaoui, M. H. E. Y., Jilbab, A., & Bourouhou, A. (2023). U-Net transfer learning backbones for lesions segmentation in breast ultrasound images. *International Journal of Electrical and Computer Engineering*, 13(5), 5747–5754. https://doi.org/10.11591/ijece.v13i5.pp5747-5754

Biyyala, V. (2023). Lung Image Segmentation: Custom U-Net with ResNet Encoder Architecture. Preprint on ResearchGate. https://doi.org/10.13140/RG.2.2.20038.04162

Ciresan, D., Giusti, A., Gambardella, L., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, (pp. 1–9). NISP.

De Moor, T., Rodriguez-Ruiz, A., Mérida, A. G., Mann, R., & Teuwen, J. (2018). Automated soft tissue lesion detection and segmentation in digital mammography using a u-net deep learning network. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1802.06865

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). Fully automated classification of mammograms using deep residual neural networks. In *14th International Symposium on Biomedical Imaging (ISBI 2017)*, (pp. 310–314). IEEE. https://doi.org/10.1109/isbi.2017.7950526

Dominguez, A. R., Puga, H., Rodrıguez, M. O., & Gasca, I. G. (2020). CAD of Breast Cancer: A Decade-Long Review of Techniques for Mammography Analysis. *Research in Computing Science* 149(5), 115–124.

Escorcia-Gutierrez, J., Mansour, R. F., Beleno, K., Jimenez-Cabas, J., Perez, M., Madera, N., & Velasquez, K. (2022). Automated Deep learning empowered breast cancer diagnosis using biomedical mammogram images. *Computers, Materials & Continua*, 71(3), 4221–4235. https://doi.org/10.32604/cmc.2022.022322

Fajrianti, E. D., Pratama, A. A., Nasyir, J. A., Rasyid, A., Winarno, I., & Sukaridhoto, S. (2022). High-Performance Computing on Agriculture: Analysis of corn leaf disease. *International Journal on Informatics Visualization*, 6(2), 411–417. https://doi.org/10.30630/joiv.6.2.793

Hadad, O., Bakalo, R., Ben-Ari, R., Hashoul, S., & Amit, G. (2017). Classification of breast lesions using cross-modal deep learning. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017),* (pp. 109–112). IEEE. https://doi.org/10.1109/isbi.2017.7950480

He, K., & Sun, J. (2015). Convolutional neural networks at constrained time cost. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5353–5360). IEEE. https://doi.org/10.1109/cvpr.2015.7299173

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (pp. 770–778). IEEE. https://doi.org/10.1109/cvpr.2016.90

Hubens, N., Mancas, M., Gosselin, B., Preda, M., & Zaharia, T. (2022). FasterAI: a lightweight library for neural networks compression. *Electronics*, 11(22), 3789. https://doi.org/10.3390/electronics11223789

Iacoviello, L., Bonaccio, M., De Gaetano, G., & Donati, M. B. (2021). Epidemiology of breast cancer, a paradigm of the "common soil" hypothesis. *Seminars in Cancer Biology*, 72, 4–10. https://doi.org/10.1016/j.semcancer.2020.02.010

Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep Learning for Medical Image-Based Cancer diagnosis. *Cancers*, 15(14), 3608. https://doi.org/10.3390/cancers15143608

Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., Heeten, A. D., & Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis*, 35, 303–312. https://doi.org/10.1016/j.media.2016.07.007

Liu, H., Chen, Y., Zhang, Y., Wang, L., Luo, R., Wu, H., Wu, C., Zhang, H., Tan, W., Yin, H., & Wang, D. (2021). A deep learning model integrating mammography and clinical factors facilitates the malignancy prediction of BI-RADS 4 microcalcifications in breast cancer screening. *European Radiology*, 31(8), 5902–5912. https://doi.org/10.1007/s00330-020-07659-y

Lotter, W., Sorensen, G., & Cox, D. (2017). A multi-scale CNN and curriculum learning strategy for mammogram classification. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 169–177). Springer. https://doi.org/10.1007/978-3-319-67558-9_20

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast. *Academic Radiology*, 19(2), 236–248. https://doi.org/10.1016/j.acra.2011.09.014

Moshina, N., Bjørnson, E., Holen, Å., Larsen, M., Hansestad, B., Tøsdal, L., & Hofvind, S. (2022). Standardised or individualised X-ray tube angle for mediolateral oblique projection in digital mammography? *Radiography*, 28(3), 772–778. https://doi.org/10.1016/j.radi.2022.03.002

Piva, I. C., Gonçalves, G., de Melo, J. D. A. C., & Huhn, A. (2023). *The implementation of clinical quality control in mammography exams: Challenges for professionals and managers*. Seven Editora. https://doi.org/10.56238/devopinterscie-279

Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K., & Park, S. J. (2017). Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID). In *Proceedings of the 8th ACM international conference on bioinformatic* (pp. 536–543). ACM. https://doi.org/10.1145/3107411.3107484

Ribli, D., Horváth, A., Unger, Z., Pollner, P., & Csabai, I. (2018). Detecting and classifying lesions in mammograms with Deep Learning. *Scientific Reports*, 8(1), 4165. https://doi.org/10.1038/s41598-018-22437-z

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Salcedo, E., Jaber, M., & Carrión, J. R. (2022). A novel road maintenance prioritisation system based on computer vision and crowdsourced reporting. *Journal of Sensor and Actuator Networks*, 11(1), 15. https://doi.org/10.3390/jsan11010015

Schaudt, D., Von Schwerin, R., Hafner, A., Riedel, P., Späte, C., Reichert, M., Hinteregger, A., Beer, M., & Kloth, C. (2023). Leveraging human expert image annotations to improve pneumonia differentiation through human knowledge distillation. *Scientific Reports*, 13(1), 9203. https://doi.org/10.1038/s41598-023-36148-7

Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports*, 9(1), 12495. https://doi.org/10.1038/s41598-019-48995-4

Silalahi, A. R. J. (2021). Breast cancer lesion detection and classification in mammograms using deep neural. In *IOP Conference Series: Materials Science and Engineering (Volume 1115).* IOP Publishing. https://doi.org/10.1088/1757-899X/1115/1/012018

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations 2015*. ICLR. https://doi.org/10.48550/arXiv.1409.1556

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. *arXiv (Cornell University)*.
https://doi.org/10.48550/arxiv.1505.00387

Tsai, K., Chou, M., Li, H., Liu, S., Hsu, J., Yeh, W., Hung, C., Yeh, C., & Hwang, S. (2022). A High-Performance Deep Neural Network Model for BI-RADS classification of screening mammography. *Sensors*, 22(3), 1160.
https://doi.org/10.3390/s22031160

Yojana, K., & Rani, L. T. (2023). OCT layer segmentation using U-NET semantic segmentation and RESNET34 encoder-decoder. *Measurement: Sensors*, 29, 100817. https://doi.org/10.1016/j.measen.2023.100817

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision – ECCV 2014*, (pp. 818-833). Springer. https://doi.org/10.1007/978-3-319-10590-1_53