Cognitive Computation
https://doi.org/10.1007/s12559-023-10189-6

# A Yolo-Based Model for Breast Cancer Detection in Mammograms

**Francesco Prinzi[1] · Marco Insalaco[2] · Alessia Orlando[2] · Salvatore Gaglio[3,4] · Salvatore Vitabile[1]**

Received: 28 February 2023 / Accepted: 1 August 2023


**Abstract**
This work aims to implement an automated data-driven model for breast cancer detection in mammograms to support physicians' decision process within a breast cancer screening or detection program. The public available CBIS-DDSM and the INbreast datasets were used as sources to implement the transfer learning technique on full-field digital mammography proprietary dataset. The proprietary dataset reflects a real heterogeneous case study, consisting of 190 masses, 46 asymmetries, and 71 distortions. Several Yolo architectures were compared, including YoloV3, YoloV5, and YoloV5-Transformer. In addition, Eigen-CAM was implemented for model introspection and outputs explanation by highlighting all the suspicious regions of interest within the mammogram. The small YoloV5 model resulted in the best developed solution obtaining an mAP of 0.621 on proprietary dataset. The saliency maps computed via Eigen-CAM have proven capable solution reporting all regions of interest also on incorrect prediction scenarios. In particular, Eigen-CAM produces a substantial reduction in the incidence of false negatives, although accompanied by an increase in false positives. Despite the presence of hard-to-recognize anomalies such as asymmetries and distortions on the proprietary dataset, the trained model showed encouraging detection capabilities. The combination of Yolo predictions and the generated saliency maps represent two complementary outputs for the reduction of false negatives. Nevertheless, it is imperative to regard these outputs as qualitative tools that invariably necessitate clinical radiologic evaluation. In this view, the model represents a trusted predictive system to support cognitive and decision-making, encouraging its integration into real clinical practice.

**Keywords** Breast cancer detection · Explainable AI · YoloV5 · Transfer learning · Proprietary dataset

## Introduction

Breast cancer is the most common worldwide tumor in the female population [1]. Previous randomized trials and incidence-based mortality studies have demonstrated a significant reduction in breast cancer mortality associated with participation in breast screening programs [2]. However, the problem of false positives and false negatives persists as a concern. Most of these errors can be attributed to dense breasts (masking effect), as well as human factors such as radiologist perception and erroneous decision-making behaviors. Additionally, the inherent imaging characteristics of tumors contribute to the issue, with benign masses often resembling malignant ones and malignant masses sometimes mimicking benign ones [3]. During the breast cancer diagnosis process, the physician aims to detect all the regions of interest (ROIs) in the whole mammogram: masses, calcifications, distortions, etc. Detection in the early stage of the disease is critical for planning new examinations, therapies, or lines of intervention. A missed detection, on the other hand, may result in irreversible injury to the patient. For

✉ Francesco Prinzi
francesco.prinzi@unipa.it

Marco Insalaco
marco.insalaco@community.unipa.it

Alessia Orlando
orlandoalessiamed@hotmail.it

Salvatore Gaglio
salvatore.gaglio@unipa.it

Salvatore Vitabile
salvatore.vitabile@unipa.it

[1] Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND), University of Palermo, Palermo, Italy

[2] Section of Radiology - Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND), University Hospital "Paolo Giaccone", Palermo, Italy

[3] Department of Engineering, University of Palermo, Palermo, Italy

[4] Institute for High-Performance Computing and Networking, National Research Council (ICAR-CNR), Palermo, Italy

Published online: 17 August 2023

this reason, breast cancer detection is the most complicated but also the most important task. Unfortunately, several proposed solutions in the literature do not aim to analyze the entire image, but rather limit detection to patch classification: the ROIs are first manually selected and cropped, and then the classifiers are trained to distinguish the crops. However, to support and imitate the physician's diagnostic process, an architecture capable of detecting all ROIs within the whole mammogram is required. Faster R-CNN, RetinaNet, and Yolo have encouraged the development of systems for breast cancer detection [4–7]. These frameworks certainly introduce two main difficulties: (1) the models have to learn the features of the whole mammogram, and the image resizing required for training may result in the loss of critical details; (2) since the model has to detect all ROIs among all patches of healthy tissue (i.e., non-ROIs), an unavoidable increase in the error rate must be faced. However, Yolo has proven to be an excellent tool in numerous scenarios, achieving higher accuracy and inference speed rates than its object detector competitors [8].

In [9], a comparison and evaluation of YoloV5 nano, small, medium, and large models using the CBIS-DDSM and INbreast datasets was performed. However, several aspects have not yet been considered. The issue of explainability was not addressed. Nevertheless, in critical domains like medical applications, ensuring model explainability is an essential prerequisite. Furthermore, it has not been examined whether the utilization of deeper architectures such as YoloV3 can enhance detection performance in the case of small datasets. Additionally, the potential advantages of incorporating a Transformer block into Yolo, considering their generalization capability, have not been investigated. In this work, a YoloV5-based model was proposed for breast cancer detection to support the physician's diagnostic process. A comparison between other feature extractors such as Darknet53 proposed in YoloV3 [10] and the Vision Transformer [11] was performed.

Given the need for large databases to facilitate deep training [12], the transfer learning (TL) technique was used. In fact, it has also recently been shown that training with small datasets by exploiting pre-trainings represents a future direction to provide a trusted system supporting cognitive and decision-making processes in the medical domain [13]. For this reason, the CBIS-DDSM [14] and INbreast [15] datasets were used as source datasets and a proprietary dataset as target. In contrast to CBIS-DDSM and INbreast, the proprietary dataset includes lesions that are more challenging to recognize, such as asymmetries and distortions, which hold significant clinical importance [16]. The proprietary dataset was acquired and annotated at the Radiology Section of the University Hospital "Paolo Giaccone" (Palermo, Italy). The workflow of the experiments performed is shown in Fig. 1.

However, despite the high performance of the deep learning models, their actual use is inhibited by their black-box nature, i.e., the internal logic is incomprehensible to users [17]. This has raised some critical issues about their use such as legal aspects, user acceptance, and trust [18, 19]. For this reason, in order to encourage the integration of these systems into real clinical practice, the problem of their explainability needs to be addressed. The gradient-free method Eigen-CAM [20] was used for saliency maps computation and compared with the occlusion sensitivity method. The saliency maps were employed to verify the learning model and to highlight the most important pixel involved in the prediction process. We believe that reporting regions in the form of heat maps can guide the physician's attention much more than ROIs prediction: ROIs are predicted and shown only above a certain confidence threshold, and the hardest-to-find regions may not exceed this threshold. In this way, the complicated, tedious, and exhausting process of mammogram evaluation can be supported by guiding the physician's attention to different ROIs.

The main contributions on the current manuscript are as follows:

- The first novelty falls within the field of explainable artificial intelligence (XAI). While data-driven methods have demonstrated high performance in various medical scenarios, their lack of transparency creates skepticism among both physicians and patients regarding these new technologies. This skepticism is particularly prominent in the development of clinical decision support systems (CDSS), where understanding the decision-making process and ensuring system reliability are crucial prerequisites for facilitating the diagnostic process. Conventional machine learning approaches are inadequate in meeting these demands and fail to provide justifications for the decisions made by the systems. Introducing explainability for breast cancer detection is of utmost importance due to the potential for early detection of invasive diseases in mammography screening. Quite frequently, these lesions may not be readily apparent and may fail to meet the confidence threshold established in Yolo to return the detection. Conversely, gradient-free XAI methods could remain unaffected by the final output and can provide valuable assistance in the diagnostic process, even in situations involving inaccurate or low confidence predictions. The saliency maps have been proposed as a valuable tool to enhance the predictions of YoloV5.
- A proprietary dataset was acquired during daily clinical sessions from the Radiology Section of the University Hospital "Paolo Giaccone" (Palermo, Italy) for model evaluation. Unlike CBIS-DDSM and INbreast, this dataset comprises a real clinical dataset containing numerous lesions that present greater complexity

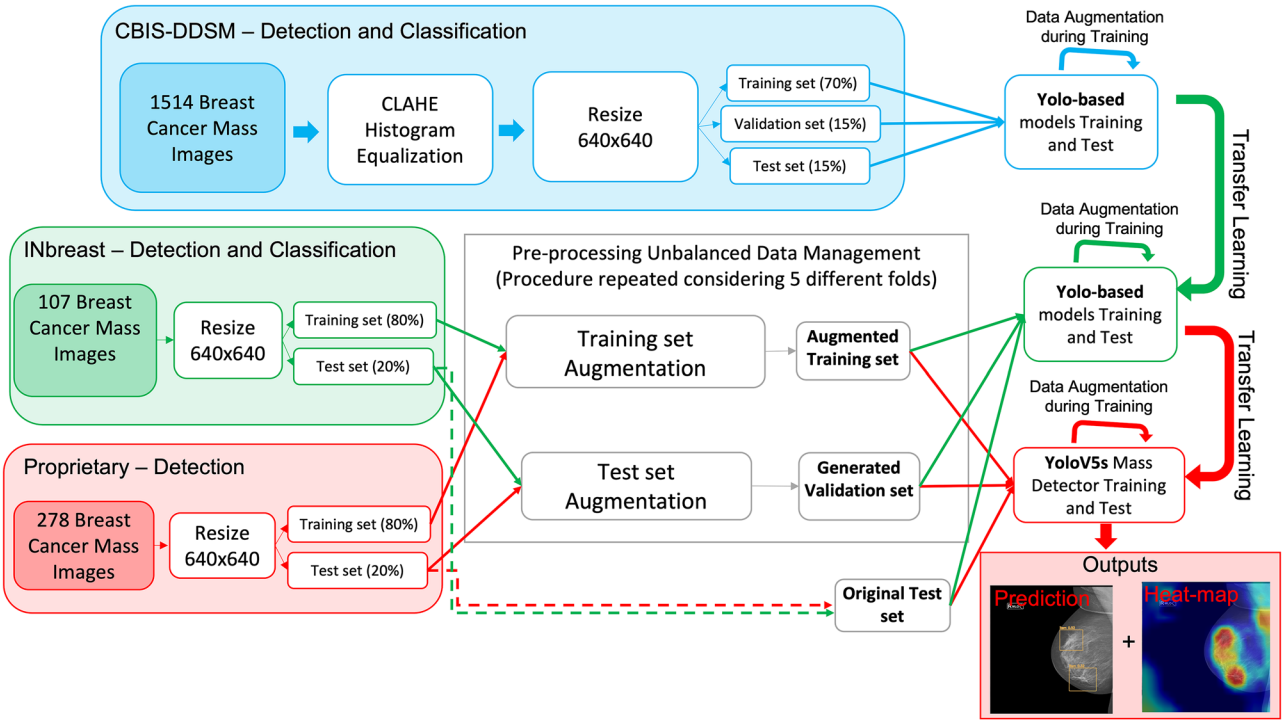


**Fig. 1** The overall architecture. The CBIS-DDSM dataset was used as source to evaluate several Yolo-based architectures (YoloV3, YoloV5 (n, s, m, l), and YoloV5-Transformer) on the INbreast target dataset. Then, the best trained architecture (YoloV5s) was used for mass detection on a proprietary dataset. A data augmentation procedure was performed before the training phase for class balancing as well as during the training. The output comprises bounding-box predictions and a heat map that highlights all the ROIs within the mammogram

in recognition, including asymmetries and distortions. These challenging cases hold an important clinical significance [16]. Furthermore, the training process involved the utilization of three datasets, enabling the final model to incorporate the knowledge acquired from CBIS-DDSM and INbreast datasets.

- The article presents a comparison of several Yolo-based models. In addition, we evaluated the integration of Transformers [11] inside Yolo. Transformers have had an enormous impact on large language models and computer vision tasks. However, the authors [11] acknowledge that Transformers lack certain inherent biases found in convolutional neural networks (CNNs), such as translation equivariance and locality. Consequently, Transformers may not generalize well when trained on limited amounts of data. This phenomenon is starting to be discussed in other studies [21]. In the context of mammograms and transfer learning, the generalizability of these findings remains uncertain.

This article is organized as follows: "Related Work" provides the related works on breast cancer classification both using patch-based classification and exploiting the whole mammogram. Section "Materials and Methods" describes the open-source CBIS-DDSM, the INbreast, and the proprietary datasets. The same section explains the three main used architectures of Yolo, their training, and the methods for saliency maps computation. Section "Results" shows the achieved results, and "Discussion", their discussion. Finally, in "Conclusions", the main conclusions are reported.

## Related Work

Given the incidence of breast cancer, many works have been proposed to support the physician's diagnostic process. Muduli et al. [22] and Mahmood et al. [23] have compared their own CNN architecture with state-of-the-art networks for malignant and benign ROIs classification. Soulami et al. [24] have also proposed a CNN, called CapsNet, to address the classification of ROIs. They showed that the classification of breast masses into normal, benign, and malignant is certainly more complex than a binary classification of masses into normal and abnormal. Also, Ragab et al. [25] have addressed breast cancer classification at patch-level, using AlexNet, GoogleNet, and ResNet-18-50-101 as feature extractors and a support vector machine as classifier. They also evaluated classification through deep feature fusion and a subsequent application of principal component analysis. Yu et al. [26] have explored several methods and CNN

architectures for tumor or normal ROIs classification. Two deep fusion models based on VGG16 were used to classify different patches extracted from the original ROI, to obtain the final prediction using a majority voting. In Agarwal et al. [27], a sliding window approach is used to scan the whole breast and extract all the possible cancer patches from the image. Several patch-based CNN (VGG16, ResNet50, and InceptionV3) were trained for breast cancer detection, that is the classification between positive and negative patches.

The aforementioned works train convolutional models that can distinguish ROIs, without dealing with recognizing them. However, at the breast screening stage, it is crucial to detect all ROIs and subsequently plan new lines of intervention. Jung et al. [7] used RetinaNet as object detector for the automatic localization of masses (both benign and malignant) in the whole mammogram. A dual-view deep convolutional neural network (DV-DCNN) for matching detected masses was proposed by AlGhamdi and Abdel-Mottaleb [28]. The authors used RetinaNet [29] for mass detection and the DV-DCNN architecture to determine if two patches from the craniocaudal (CC) and mediolateral oblique (MLO) views of the same breast represent the same mass, i.e., a positive pair. In [4] a Yolo-based Computer-Aided Diagnosis (CAD) was proposed for mass detection and classification, proving that the system works also where the masses exist over the pectoral muscles or dense regions. Aly et al. [5] define the evaluation process of screening mammograms as very monotonous, tiring, lengthy, costly, and significantly prone to errors for human readers. In fact, a YoloV3 model was proposed for mass detection and classification. They obtained the fairest and most accurate performance using an augmented dataset.

In this work, new feature extractors for breast cancer detection were considered. The YoloV5 architecture was compared with the previous YoloV3 model and considering also the Vision Transformer block. In addition, Eigen-CAM was used as explainable AI algorithm [30, 31] to provide a post hoc explanation. The Eigen-CAM method was compared with occlusion sensitivity. The generated saliency maps were used for two main reasons: (1) as explanatory debugging tool for preventing inadequate outputs [32, 33] and (2) to guide physicians' attention even on incorrect prediction scenarios.

## Materials and Methods

### Datasets

#### The CBIS-DDSM Dataset

The CBIS-DDSM dataset [14] is the curated version of the Digital Database for Screening Mammography (DDSM) dataset and is composed of scanned film mammograms. Focusing on masses, 1514 images with a total of 1618 lesions (850 benign and 768 malignant) were included. Of the total 1696 lesions, 78 were discarded due to a mismatch between the size of the image and its mask, generating ROIs that did not match a lesion.

#### The INbreast Dataset

The INbreast [15] dataset consists of 410 full-field digital mammograms (FFDM) classified into normal, benign, and malignant. Only the 107 positive images were selected, and lesions with Bi-Rads > 3 were considered malignant; the others were labeled as benign. Considering that some images contain multiple lesions, a total of 40 benign and 75 malignant ROIs were identified.

#### The Proprietary Dataset

The dataset consists of 278 FFDMs containing a total of 307 lesions, annotated by expert radiologists dealing with the identification of abnormal regions. The images were acquired by a Fujifilm Full Field Digital at the Radiology Section of the University Hospital "Paolo Giaccone" (Palermo, Italy). Images have spatial resolution and pixel size of $5928 \times 4728$ and 50 μm, respectively. The image annotations were saved in grayscale softcopy presentation state (GSPS) format, compliant with the DICOM standard. All identified by radiologist ROIs were annotated by a circumscribed circle, and then the coordinates of the bounding-boxes used for Yolo input were calculated as the coordinates of the square circumscribed by the circle. The dataset used in our study was obtained from the real clinical practice at University Hospital "Paolo Giaccone" (Palermo, Italy). Specifically, the data was collected from the outpatient breast clinic, which specializes in second-level diagnostics. As a result, the acquired case series are heavily skewed towards more severe breast cancer lesions including distortions and asymmetries. Detecting and diagnosing distortions can be particularly challenging, as they are characterized by the presence of spicules radiating from a point, focal retractions, or straightening at the edges of the parenchyma [34]. Consequently, distortions are among the most commonly overlooked abnormalities [35]. Asymmetries refer to unilateral deposits of fibroglandular tissue that do not meet the criteria for being classified as masses. They can be further categorized as asymmetry, focal asymmetry, global asymmetry, or developing asymmetry. It has been estimated that around 20% of asymmetry cases are associated with malignancy, making them an important area of research [16]. The benign lesions represent 17.6% of the dataset (54 samples), and the 82.4% (253 samples) are malignant. The dataset reflects a real clinical scenario; in fact, it is composed of masses

(62%), asymmetries (%15), and distortions (23%). Given the large class imbalance, the proprietary dataset was used only for detection.

## Data Pre-Processing

For the CBIS-DDSM and INbreast datasets, the coordinates of the ROIs bounding box required for Yolo training were calculated considering the coordinates of the smallest rectangle containing the segmented lesion. Instead, the ROI coordinates for the proprietary dataset were computed from the square region that inscribes the circle containing the ROI. The CBIS-DDSM dataset has an acceptable size for deep learning architecture training. However, it is composed of scanned film mammograms, much noisier and less detailed than FFDM. For this reason, only for the CBIS-DDSM dataset, the contrast limited adaptive histogram equalization (CLAHE) was applied for image enhancement [23], with the following setting: 1 as contrast limit, $2 \times 2$ as grid size, followed by a $3 \times 3$ Gaussian filter. For all datasets, the gray levels were scaled in the range 0–255, and the images were resized to $640 \times 640$ using the Lanczos filter [36, 37]. The CBIS-DDSM dataset was splitted randomly considering 70% training, 15% validation, and 15% test set. Conversely, the INbreast and the proprietary datasets were split into training (80%) and test set (20%), respectively. Considering the small size of the two datasets and the unbalanced issue, the next "Data Augmentation" discusses data augmentation for class balancing and generation of the validation set "Techniques for Class Balancing Before the Training Phase", as well as the procedure to improve the training "Techniques Used During the Training Phase".

## Data Augmentation

### Techniques for Class Balancing Before the Training Phase

Due to the excessive imbalance classes for the INbreast and proprietary dataset, the minority class images (benign) of the training set were augmented. Although the main purpose of the work is to evaluate the detection performance on the proprietary dataset (regardless of lesion class), the following data augmentation procedure was applied to the proprietary dataset before the training phase. Figure 2 summarizes the transformation considered. In particular, 180° rotation and 180° rotation + flip upper-down (UD) were applied for benign images. The other transformations were applied during the training of Yolo, as discussed in the next subsection "Techniques Used During the Training Phase". In addition, according to [5], the remaining test dataset was augmented to obtain the validation set. In fact, flip UD, 180° rotation + flip UD, flip left-right (LR) and 180° rotation were applied on benign images, and Flip LR for malignant images. Considering the smaller difference between the classes, on INbreast, also, 180° rotations for malignant masses were considered [9].

This procedure resulted in the generation of a balanced validation set. In addition, the discussed procedure for INbreast and the proprietary datasets was repeated considering 5 different splitting of training and test sets (5-fold cross-validation).

### Techniques Used During the Training Phase

Transformations not considered in the previous step were performed during Yolo training. In particular, three different data augmentation configurations were chosen: low, medium, and high. In all cases, image translation, rotation, scale, shear, flip UD, flip LR, and also HSV augmentation were considered. In addition, although it is a common scenario for breast cancer, all three datasets contain few multi-lesion images. Therefore, to improve the model's capability to detect multiple lesions in the same image, the mosaic technique was used. The mosaic augmentation method consists of the generation of a $2 \times 2$ grid image, containing the considered image and three random images of the dataset. The mosaic technique improves training for two main reasons: (1) merging 4 images results in multiple ROIs in the

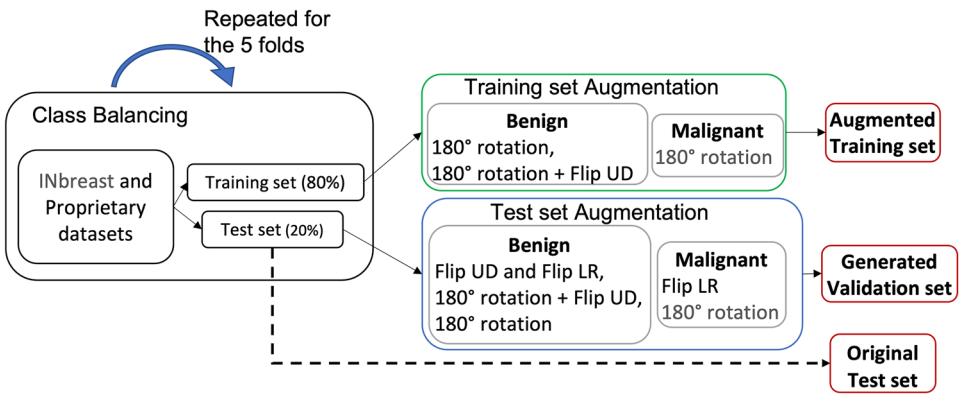


**Fig. 2** Transformations for class balancing and validation set creation. The procedure was repeated implementing the 5-fold cross-validation

same image, and the model improves in recognizing multiple ROIs simultaneously; (2) to achieve the same input size, the 4 merged images and their respective ROIs are downsized, improving the detection of smaller lesions.

Table 1 shows the parameter set for each configuration. The values reported HSV, translation, rotation, scale, and shear indicate the range considered for the random transformation. For flip and mosaic, the value indicates the probability of performing the transformation, so 0.5 is considered a higher level of augmentation because both augmented and non-augmented images are considered for training.

## Yolo Architectures Training

Like other single-stage object detectors, Yolo consists of three parts: backbone, neck, and head. The backbone part is a CNN that extracts and aggregates image features. The neck part allows for features extraction optimized for small, medium, and large object detection. In the end, the three feature maps for small, medium, and large object detection are given as input to the head part, thus composed of convolutional layers for the final prediction. Yolo requires that the image is divided into a grid, then makes a prediction for each grid cell. The prediction consists of a 6-tuple $y = (p_c, b_x, b_y, b_h, b_w, c)$, where $(b_x, b_y, b_h, b_w)$ identify coordinates $(x, y)$ and sizes $(height, width)$ of the predicted bounding-box, $p_c$ represent the probability that there is an object in the cell, and $c$ represent the predicted class. The mechanism of anchors is also used, to allow multiple object detection in the same grid cell. For this reason, the prediction is the 6-tuple discussed for each specified anchor. Each version of Yolo has its own peculiarities, which mainly concern the structure of the feature extractor, that is, the backbone.

### YoloV3 Model

YoloV3 is much deeper than the previous two versions and is more accurate but requires more time and data for training. In YoloV3, the Darknet53 was used as backbone [10]. Darknet53 is a hybrid approach between Darknet19 (used in YoloV2 [38]) and residual network elements (e.g., BottleNeck) [39], proposed to improve the Darknet19 and the efficiency of ResNet-101/152. The short-cut connections allow getting more fine-grained information, leading to better performance for small objects. The feature pyramids network (FPN) [40] is used as neck, allowing to learn objects of different sizes: it specializes in detecting large and small objects. In addition, the non-maximum suppression to select one bounding box out of many overlapping bounding boxes is used.

### YoloV5 Model

YoloV5 uses CSPDarknet53 as its backbone: it exploits the architecture Darknet53 proposed in [10] and employs a CSPNet [41] strategy to partition the feature map of the base layer into two parts and then merges them through a cross-stage hierarchy. In the neck part, PAnet [42] is used to generate the feature pyramids network (FPN) and allow the extraction of multi-scale feature maps. This structure allows the extraction of features optimized for small, medium, and large object detection. YoloV5 was released in nano, small, medium, large, and extra-large versions. The versions differ in the number of convolutional kernels used and thus the number of parameters. In this paper, a comparison between nano, small, medium, and large versions was performed.

### YoloV5-Transformer

In contrast to convolutional networks, Transformer are able to model the relationships among various small patches in the image. The Transformer block assumes the image is split into a sequence of patches, where each patch is flattened to a vector. These flattened image patches are used to create lower-dimensional linear embeddings and fed into a Transformer encoder, composed by a multi-head attention to find local and global dependencies in the image. It has been shown that the introduction of a Transformer block to convolutional networks can improve efficiency and overall accuracy [43]. In YoloV5, the Transformer block was embedded in the penultimate layer of the backbone, that is, among the three convolutional layers preceding the spatial pyramid pooling layer.

### Models Training

Considering the small size of both INbreast and proprietary datasets, training a deep architecture such as Yolo may harm the reliability of the trained models. Therefore, despite it being composed of scanned film mammograms, the CBIS-DDSM is employed as source dataset for initial training. The

**Table 1** Setting for data augmentation during the training phase

| Level | H,S,V | Translation | Rotation | Scale | Shear | Flip (UD, LR) | Mosaic |
|---|---|---|---|---|---|---|---|
| Low | 0.0, 0.0, 0.0 | 0.1 | 5.0 | 0.1 | 5.0 | (0.5, 0.5) | 0.0 |
| Med | 0.007, 0.35, 0.2 | 0.3 | 10.0 | 0.3 | 5.0 | (0.5, 0.5) | 1.0 |
| High | 0.015, 0.7, 0.4 | 0.3 | 20.0 | 0.3 | 10.0 | (0.5, 0.5) | 0.5 |

above setup allows the TL technique on the INbreast and proprietary target datasets. Considering that both source and target datasets are labeled, the performed TL was inductive transfer learning [44]. Since Yolo simultaneously solves a regression task to predict bounding box coordinates, and two classification tasks to predict objectiveness and class score, two different loss functions were employed. For regression, complete Intersection over Union (IoU) loss was used; for classification, binary cross-entropy with logits loss function was used in both cases.

### Performance Evaluation

The results obtained were presented considering the most common indexes for object detection tasks such as precision, recall, and average precision. The average precision (AP) is defined as the area under the precision-recall curve. The IoU was set to 0.5. For CBIS-DDSM and INbreast datasets, AP was calculated for detecting malignant (M AP) and benign (B AP) lesions separately, as well as the mean of the two classes (mAP).

## Models Explanation: Eigen-CAM

Examining trained models is essential before incorporating them into actual clinical practice. As a result, our system produces prediction explanations as the second output to fulfill this requirement. Saliency maps have the capability to reveal the pixels or regions that played a significant role in the decision-making process of the system. This effectively highlights all potential ROIs to the physician. Several gradient-based methods such as CAM [45], Grad-CAM [46], and GradCAM++ [47] have been proposed to implement interpretability and transparency of deep learning models. In particular, they are class discriminative visualization methods and require the class probability score for the gradient computations. However, gradient-based methods suffer from this problem: backpropagating any quantity requires additional computational overhead and assumes that classifiers produced correct decisions, and whenever a wrong decision is made, all mentioned methods will produce wrong or distorted visualizations [20]. For this reason, the localization accuracy of the above methods remains weak, especially in the case of incorrect predictions. In addition, while traditional CNNs provide class distributions for each sample, YOLO's output includes bounding box coordinates, object presence probabilities in each cell, and class distributions. These issues often make the output non-differentiable and impractical to implement gradient-based algorithms. As a result, many object detection studies employing Yolo rely on Eigen-CAM for architecture interpretation [48–50]. Eigen-CAM is preferred due to its gradient-free nature and principal components use from the extracted feature maps. It should be noted that gradient-based methods, which rely on output and activation maps, can produce distorted visualizations when predictions are incorrect. To address these issues, this study presents Eigen-CAM for saliency map computation and compares it with the occlusion sensitivity method.

Eigen-CAM is a gradient-free method that computes and visualizes the principal components of the learned features/representations from the convolutional layers, resulting in intuitive and compatible with all the deep learning models. In Eigen-CAM, it is assumed that all relevant spatial features learned over the hierarchy of the CNN model will be preserved during the optimization process, and non-relevant features will be regularized or smoothed out. The Eigen-CAM is computed considering the input image $I$ of size $i \times j$ projected onto the last convolutional layer $L = K$ and is given by $O_{L=K} = W_{L=K}^{T} I$. The matrix $O_{L=K} = U\Sigma V^{T}$ is factorized using the singular value decomposition to obtain the principal components. The activation map is given by the projection on the first eigenvector $L_{Eigen-CAM} = O_{L=K} V_1$, where $V_1$ is the first eigenvector in the V matrix. Similar to Eigen-CAM, Occlusion sensitivity can be linked to image detection tasks, and it is gradient-free and independent of the specific architecture used. It assesses changes in activations resulting from occluding different regions of the image [51].

The saliency maps have been proposed as a valuable tool to enhance the predictions of YoloV5, which can assist physicians in the diagnostic process, especially when the model fails to make accurate predictions. YoloV5 only provides predictions if they surpass a certain confidence threshold. The purpose of saliency maps is to identify all ROIs and mitigate false negative issues. It has been observed that many cancer types progress to an invasive stage due to the failure of early prediction also with preliminary signs. Therefore, in contrast to YoloV5's predictions, saliency maps offer all potential ROIs, even with low confidence. This inevitably leads to an increase in false positives. Considering this, physicians receive two outputs: firstly, the conventional YoloV5 output that balances precision and recall, providing only ROIs that exceed a certain confidence level. In addition, saliency maps propose all potential ROIs, which may serve as early cancer indications, even if their probability of being lesions (i.e., not exceeding the threshold) is low. Thus, a simple predictive model transforms into a decision-support system, as physicians receive not only a definitive decision

**Table 2** Comparison of the nano, small, medium, and large architectures of YoloV5 on the CBIS-DDSM dataset, considering all default hyperparameters

| Model | B AP | M AP | Precision | Recall | mAP |
|---|---|---|---|---|---|
| n | 0.257 | 0.479 | 0.473 | 0.408 | 0.368 |
| s | 0.257 | 0.518 | 0.447 | 0.427 | 0.387 |
| m | 0.280 | 0.514 | 0.489 | 0.403 | 0.397 |
| l | 0.239 | 0.488 | 0.491 | 0.377 | 0.364 |

**Table 3** Performance of YoloV5 small version, considering the equalized CBIS-DDSM dataset, Adam optimizer, and the three data augmentation configurations

| Hyps | B AP | M AP | Precision | Recall | mAP |
|---|---|---|---|---|---|
| Equal | 0.300 | 0.501 | 0.487 | 0.408 | 0.400 |
| Adam+equal | 0.321 | 0.555 | 0.487 | 0.464 | 0.438 |
| aug-low | 0.241 | 0.49 | 0.46 | 0.394 | 0.366 |
| aug-med | 0.337 | 0.549 | 0.497 | 0.487 | 0.433 |
| aug-high | 0.361 | 0.634 | 0.566 | 0.482 | 0.498 |

but also suggestions of lesions that the system recommends paying attention to.

## Results

The experiments were performed in Google Colaboratory Pro, using Python 3 environment. The PyTorch implementation proposed by Ultralytics [52] was exploited, and the Weights & Biases platform [53] was used to monitor the training process. The trainings were performed for 100 epochs and 16 as batch. The validation mAP was used for model selection, considering the best model as a weighted combination of mAP@0.5, mAP@0.5:0.95 metrics, respectively 0.9 and 0.1.

### CBIS-DDSM Results and Data-Augmentation Improvements

The CBIS-DDSM dataset was used to evaluate the optimal YoloV5 architecture and for hyperparameters optimization, considering the nano, small, medium, and large versions. Then, it was exploited as source dataset to implement inductive TL and improve the generalization capabilities on INbreast and proprietary FFDM images. For this reason, given the huge amount of hyperparameters, an initial analysis was performed using all the proposed default values for each model. Table 2 shows the achieved results for each version of YoloV5. The nano and large versions have a lower mAP than the small and medium versions. Conversely, the small model, compared with the medium model, results in a more balanced precision and recall pair, while it contains about one-third of its parameters. Therefore, all subsequent experiments were carried out only considering the small model.

Table 3 shows that the histogram equalization specified in the data pre-processing section improves the model performance. In addition, the Adam optimizer using 0.001 as learning rate outperforms the default stochastic gradient descent (SGD) optimizer with learning rate of 0.01. Therefore, experiments to evaluate the impact of data augmentation were carried out using the equalized dataset and Adam optimizer. Table 3 shows how the results improve as data augmentation increases. The extensive data augmentation employed emphasizes the necessity for substantial amounts of data when training this deep architecture, confirming the choice of using the CBIS-DDSM dataset to perform TL on INbreast and proprietary datasets.

### Inbreast Results and Transfer Learning Evaluation

Exploiting the optimized hyperparameters for the CBIS-DDSM dataset, YoloV3 and YoloV5-Transformer models were also trained on the CBIS-DDSM dataset, to implement the TL technique on the INbreast target dataset. Table 4 shows the achieved results. Considering the dataset size, the performance was calculated in 5-fold cross validation, and mean and standard deviation were reported for each metric. The best training protocol for the CBIS-DDSM, that is, Adam optimizer, high data augmentation, and 16 as batch, was used for all the experiments. In addition, INbreast was also trained from scratch to show the difference in accuracy with and without TL. The YoloV5s model outperforms its previous version YoloV3 and also the YoloV5-Transformer. YoloV3 contains a feature extractor with more parameters than YoloV5s and Transformer (about 61 vs. 7 million) and therefore needs a larger amount of data for their training. In addition, the YoloV5-Transformer version showed lower performance while it has a comparable number of parameters to YoloV5s. Comparing YoloV5s training from scratch and with TL on the INbreast, an increase of 0.061 mAP and 0.119 of B AP was calculated. The imbalance of the dataset clearly reflects the model performance: the benign lesions detection rate, which is the minority class, is lower than malignant lesions for each considered model.

**Table 4** 5-fold results for the three used architectures on INbreast dataset (*Tr* is for Transformer; *NoTL* is the training without transfer learning)

| Model | B AP | M AP | Precision | Recall | mAP |
|---|---|---|---|---|---|
| YoloV3 | $0.585 \pm 0.093$ | $0.890 \pm 0.036$ | $0.785 \pm 0.012$ | $0.695 \pm 0.104$ | $0.738 \pm 0.061$ |
| YoloV5s-Tr | $0.642 \pm 0.060$ | $0.894 \pm 0.054$ | $0.799 \pm 0.118$ | $0.742 \pm 0.146$ | $0.771 \pm 0.048$ |
| YoloV5s-NoTL | $0.652 \pm 0.051$ | $0.890 \pm 0.047$ | $0.835 \pm 0.059$ | $0.713 \pm 0.770$ | $0.771 \pm 0.038$ |
| YoloV5s | $0.771 \pm 0.131$ | $0.898 \pm 0.069$ | $0.854 \pm 0.097$ | $0.729 \pm 0.100$ | $0.835 \pm 0.098$ |

**Table 5** 5-Fold results on the proprietary dataset, considering the training with and without transfer learning

| Model | Precision | Recall | mAP |
|---|---|---|---|
| YoloV5s no-TL | 0.665 ± 0.054 | 0.541 ± 0.043 | 0.561 ± 0.053 |
| YoloV5s TL | 0.726 ± 0.110 | 0.591 ± 0.063 | 0.621 ± 0.035 |

### Proprietary Dataset Results and Transfer Learning Evaluation

The YoloV5s model was the most accurate for the two open-source datasets and was used for lesion detection on proprietary dataset. The trained model using the CBIS-DDSM as source dataset and INbreast as target dataset was the checkpoint to start training on the proprietary dataset. For this reason, the model trained on the proprietary dataset brings the knowledge learned on CBIS-DDSM and INbreast. Figure 3 shows the difference in validation mAP calculated during training with and without transfer learning. In particular, higher initial mAP, faster mAP growth in the early epochs, and higher mAP asymptote was calculated using transfer learning [54]. The result was confirmed in the test set, with an mAP of 0.561 and 0.61 without and with transfer learning, respectively. Table 5 shows the results computed within the 5-Fold Cross Validation strategy.

### Explainability Results

To evaluate the performance using XAI methods, we conducted a manual analysis on a proprietary dataset subset consisting of 50 images and 56 lesions. No healthy images were considered. Our focus was evaluating the differences in false positives and false negatives using two XAI techniques: Eigen-CAM and occlusion sensitivity. Through a qualitative analysis, the generated saliency maps do not exhibit complete overlap as shown in Figs. 4 and 5. However, a poor overlap between saliency maps calculated through different methods has been widely shown in the literature [55–57]. More specifically, it has been observed that when considering occlusion sensitivity, the regions linked to lesions appear to be slightly illuminated compared to Eigen-CAM, where they are more prominently highlighted. In addition, the quantitative analysis showed the superiority of Eigen-CAM for this object detection task in mammography. Table 6 summarizes the results. In the selected subset, the Yolo model correctly detected 41 lesions, but missed 15 lesions (false negatives) and incorrectly identified 19 non-existent lesions (false positives). However, when we employed Eigen-CAM, we observed better results. Out of the 56 lesions, 52 were correctly detected, reducing the false negatives to just 4. However, the use of Eigen-CAM led to an increase in false positives, with a total of 34. On the other hand, the occlusion sensitivity method did not perform as well as Eigen-CAM, showing an increase in false negatives to 20 and false positive of 55.

## Discussion

### Performance and Transfer Learning Importance

The proposed work for breast cancer detection introduces several novelties and advantages. Three different datasets were considered. The CBIS-DDSM is the largest and therefore the most appropriate for deep training. However, it is composed of scanned film mammograms, resulting in images that are notably distinct from the FFDM images. Conversely, the INbreast and the proprietary FFDM datasets can be considered a good benchmark for testing Yolo on real clinical practice images. For this reason, the CBIS-DDSM dataset was used to obtain an optimized pre-training

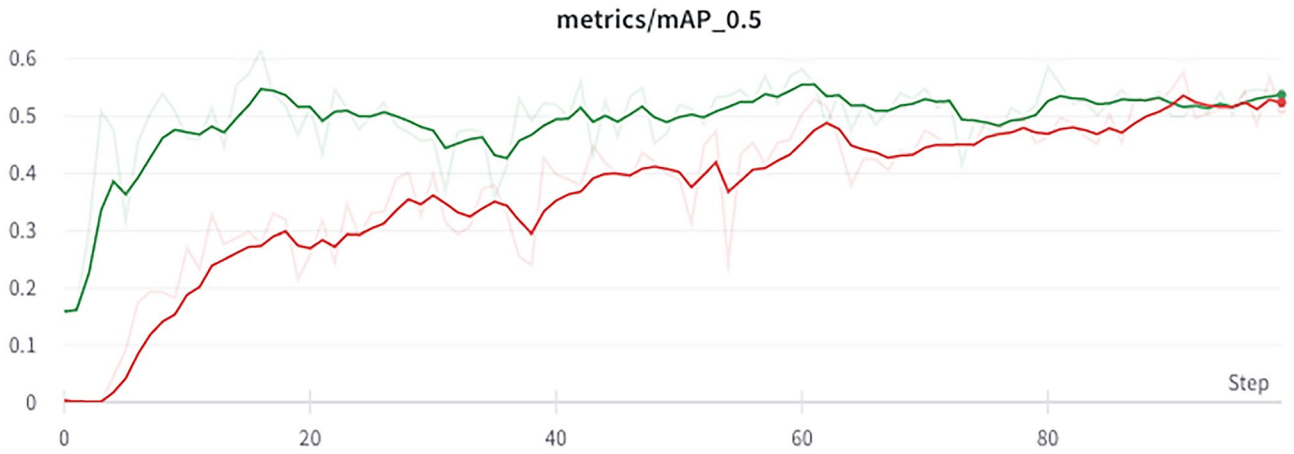


**Fig. 3** Training performance with (green) and without (red) transfer learning on the proprietary dataset

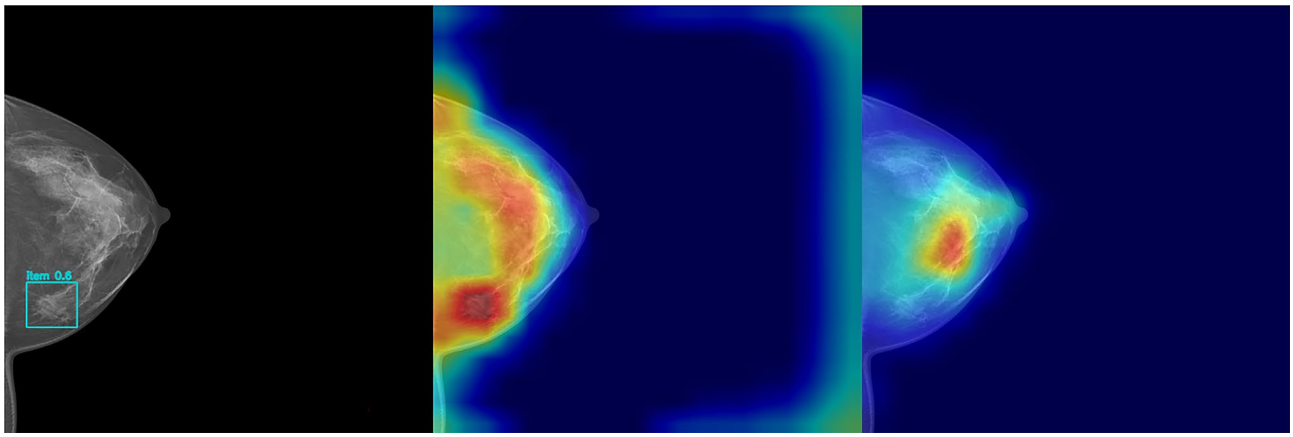


**Fig. 4** Example of a bounding-box prediction on the left and the respective saliency map on the right. The ROI is correctly predicted with a confidence index of 0.6. However, other suspicious areas are also highlighted on the saliency map

compared with the common COCO dataset (that is the benchmark for Yolo). In fact, the COCO dataset is used for the recognition of objects, cars, people, etc., on real-life images. In each case, with a significantly different distribution than breast cancer in mammograms. Then, for all experiments, the transfer learning technique was exploited using the CBIS-DDSM as source dataset, and different Yolo architectures were compared. Considering that Yolo architectures evolve to improve both accuracy and inference speeds, it was not obvious to find YoloV5 more accurate than YoloV3. Moreover, among the various versions of YoloV5, the small version was the most accurate, also compared with the YoloV5s-Transformer. The performance obtained on the proprietary dataset was lower than on INbreast. However, our dataset contains three times the number of lesions, allowing for a more accurate evaluation of the models. Also, although both are datasets for breast cancer analysis, it is natural that the distributions, and consequently the training, differ. In fact, INbreast was acquired with a MammoNovation Siemens FFDM machine with a pixel size of 70 μm and our dataset with a Fujifilm FFDM with a pixel size of 50 μm. The spatial resolution is also very different: for INbreast 3328×4084 or 2560×3328 and for the proprietary dataset 5928×4728. Moreover, the main difference lies in the heterogeneity of the datasets. In fact, for INbreast, the 107 considered abnormalities are only masses, with 2 asymmetries. In contrast, our dataset is mainly composed of masses (62%), but also of asymmetries (15%) and distortions (23%). The presence of these types of lesions, which account for 38% of our dataset, poses an additional challenge for accurate detection. In fact, according to BI-RADS [58], the term architectural distortion (AD) is used when normal architecture is distorted with a non-definite visible mass. AD is not always a sign of cancer and may represent different benign processes and high-risk lesions [59], and it is responsible for 12 to 45% of breast cancer missed during screening [60].

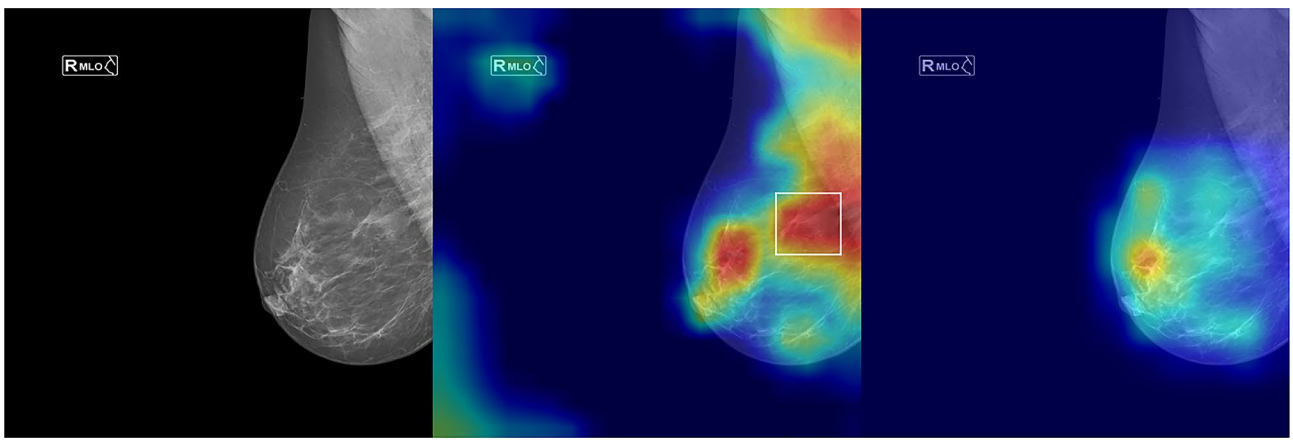


**Fig. 5** Example of wrong prediction on the left and the respective saliency map on the right. Despite the error, the saliency map calculated via Eigen-Cam provides several suspicious ROIs, as well as the miss-detected lesion (marked with the white bounding-box)

**Table 6** Performance variation through the use of saliency maps

| Model | Lesions # | TP | FP | FN |
|---|---|---|---|---|
| Yolo-based | 56 | 41 | 19 | 15 |
| Eigen-CAM | 56 | 52 | 34 | 4 |
| OS | 56 | 36 | 55 | 20 |

Asymmetries are areas of fibroglandular tissue visible on only one mammographic projection, mostly caused by the superimposition of normal breast tissue. There are different types of asymmetries: for example, the developing asymmetry has a 15% risk of malignancy [61], and the global symmetry instead is mostly a normal variant. Although this introduces a significant level of complexity, it moves the system towards the real-world clinical scenario. For this reason, the achieved results are encouraging and demonstrate that breast cancer detection can be addressed without reducing the task to patch classification.

## Comparison

An accurate comparison with other studies is complex because of different datasets, pre-processing, and training protocols. However, Table 7 shows some similar works. In [62], OPTIMAM dataset (OMI-H), composed of about 5300 mammograms, was used as source dataset to perform TL on INbreast dataset. Using the faster R-CNN architecture, they obtained an AUC-ROC of 0.79 and 0.95 for benign and malignant lesion detection. YoloV1 was used in [4], resulting in 99.5 and 99.9 for benign and malignant lesion detection in the DDSM dataset. Yolo9000 (e.g., YoloV2) is used in [63]: in contrast to our system, localization and classification performance were evaluated separately on the INbreast dataset. In particular, first, the lesions are localized, and then only the localized ones are classified, resulting in a detection accuracy of 97.2 and a classification accuracy of 95.3. The most similar work to ours in terms of evaluation protocol and workflow was proposed by Aly et al. [5]. Using YoloV3, they obtained an AP of 94.2 and 84.6 for benign and malignant detection, respectively. However the reported best results are computed using a higher image spatial resolution (832×832 vs. our 640×640), and the results were reported in 5-fold cross-validation only for 448×448 spatial resolution. In fact, comparing our result on the best fold with their result on 608×608 images, we obtained an AP of 88.5 (vs. their 87.5) and 92.2 (vs. their 80.8) for benign and malignant detection, respectively, illustrated in Aly et al. [5], increasing the image size proves beneficial for the learning process. However, the disparity between experiments conducted with sizes of 448×448 vs. 608×608 is quite substantial, but it diminishes significantly when considering the size of 832×832. This finding suggests that larger image sizes may yield slightly improved results, while the increased complexity of models and the associated optimization could pose a considerably increased computational cost.

## Explainability Discussion

Despite the encouraging performance, the system must be both accurate and trusted by physicians for its integration into real clinical practice. Therefore, an introspection and explanation of the trained model were conducted via Eigen-CAM. Figures 4 and 5 show two generated saliency maps via Eigen-CAM and occlusion sensitivity methods. In particular, the former image represents a correct prediction and the latter an incorrect prediction. In Fig. 4, the Eigen-CAM heat-map results most brightly around the predicted lesion, but it is suggested that the physician should also pay attention to other areas of the image. In Fig. 5, instead, the model makes an error in prediction (missed detection). In this figure, the advantage of using a gradient-free method can be seen. In fact, the generated Eigen-CAM heat map identifies several salient areas that demand the physician's attention.

In addition, the saliency maps depicted in Figs. 4 and 5 indicate that the activations primarily concentrate on the breast region. Any minimal activations observed outside this area (in the Eigen-CAM maps) can be attributed to artifacts and are not considered confounding factors for the physician. It is possible to speculate that the slight activations at the black edges of the images might assist in aligning the coordinates of the bounding boxes predicted in the opposite area of the image, where only the background is present. The obtained saliency maps are class-independent as confirmed by clinical literature findings, where mammography is typically employed as a screening examination aimed at identifying certain abnormalities. On the other hand, other examination modalities, such as MRI, are more informative

**Table 7** Comparison between the proposed and other breast cancer detection works, considering the INbreast dataset. (*Det*, detection; *Cls*, classification; *Acc*, accuracy; *AP*, average precision; → is for TL from dataset1 to dataset2)

| Paper | Architecture | Dataset | Performance |
|---|---|---|---|
| [62] | Faster R-CNN | Optimam → INbreast | AUC B: 0.79; M: 0.95 |
| [4] | YoloV1 | DDSM | AUC B: 99.5; M: 99.9 |
| [63] | YoloV2 | DDSM & INbreast | Det. Acc: 97.2; Cls Acc (AUC): 95.3 |
| [5] | YoloV3 | INbreast | AP B: 94.2; M: 84.6 |
| Our | YoloV5s | CBIS-DDSM → INbreast | AP B:0.771 ± 0.131; M:0.898 ± 0.069 |

for characterization purposes and are thus considered secondary examinations [12, 64].

Based on these findings, Eigen-CAM proves to be the method more suitable with respect to occlusion sensitivity for generating saliency maps in object detection tasks. Despite the unavoidable increase in false positives, the reduction in false negatives was significant. This reduction is particularly important from a clinical perspective, as it enables early diagnosis and facilitates the scheduling of further examinations by ruling out the growth of invasive lesions. Considering these factors, we believe that saliency maps should complement, rather than replace, the outputs of the Yolo model. In fact, Yolo's predictions resulted strict with a small number of false positives, while Eigen-CAM's predictions are more conservative with a minimal number of false negatives. Above all, these outputs should be seen as a qualitative tool that always requires clinical radiologic evaluation. For this reason, it is the responsibility of the physician to determine which areas necessitate additional examination.

## Conclusions

In this work, a Yolo-based model was proposed for breast cancer detection. Although the CBIS-DDSM dataset is composed of scanned film mammograms, the use of the transfer learning technique improves the models' generalization capabilities when Yolo is fine-tuned with FFDM images (INbreast and proprietary datasets). The results obtained on the INbreast dataset were exploited to train YoloV5 on the proprietary dataset. The performance obtained are very encouraging, also considering the heterogeneity of the proprietary dataset, which is composed of particularly difficult-to-recognize lesions such as asymmetries and distortions. In addition, the use of the saliency maps makes the internal process of deep learning models transparent and encourages the integration of our model within a clinical decision support system. In fact, the gradient-free Eigen-CAM method highlights all the suspicious ROIs, also in incorrect prediction scenarios. For this reason, it represents the enhanced output of our model. The proposed model represents a trusted predictive system to support cognitive and decision-making and control processes in the clinical practice. In addition, the XAI results pave the way for a prospective study in which the diagnostic performance of physicians is evaluated with and without the support of both Yolo and Eigen-CAM outputs, using an external data cohort. This represents a step towards the integration of data-driven systems into real clinical practice.

**Funding** Open access funding provided by Università degli Studi di Palermo within the CRUI-CARE Agreement. This work was partially supported by the University of Palermo Grant EUROSTART, CUP B79J21038330001, Project TRUSTAI4NCDI.

**Data Availability** Data will be made available on reasonable request.

## Declarations

**Ethical Approval** Retrospective data collection was approved by the local ethics committee.

**Consent to Participate** The requirement for evidence of informed consent was waived because of the retrospective nature of our study.

**Conflict of Interest** The authors declare that they have no conflict of interest.



## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209–249. https://doi.org/10.3322/caac.21660.
2. Duffy SW, Tabár L, Yen AM-F, Dean PB, Smith RA, Jonsson H, Törnberg S, Chen SL-S, Chiu SY-H, Fann JC-Y, Ku MM-S, Wu WY-Y, Hsu C-Y, Chen Y-C, Svane G, Azavedo E, Grundström H, Sundén P, Leifland K, Frodis E, Ramos J, Epstein B, Åkerlund A, Sundbom A, Bordás P, Wallin H, Starck L, Björkgren A, Carlson S, Fredriksson I, Ahlgren J, Öhman D, Holmberg L, Chen TH-H. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. Cancer. 2020;126(13):2971–2979. https://doi.org/10.1002/cncr.32859.
3. Ekpo EU, Alakhras M, Brennan P. Errors in mammography cannot be solved through technology alone. Asian Pac J Cancer Prev: APJCP. 2018;19(2):291. https://doi.org/10.22034/APJCP.2018.19.2.291.
4. Al-Masni MA, Al-Antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, Valarezo E, Choi M-T, Han S-M, Kim T-S. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning Yolo-based cad system. Comput Methods Programs Biomed. 2018;157:85–94. https://doi.org/10.1016/j.cmpb.2018.01.017.
5. Aly GH, Marey M, El-Sayed SA, Tolba MF. Yolo based breast masses detection and classification in full-field digital mammograms. Comput Methods Programs Biomed. 2021;200:105823. https://doi.org/10.1016/j.cmpb.2020.105823.
6. Baccouche A, Garcia-Zapirain B, Olea CC, Elmaghraby AS. Breast lesions detection and classification via Yolo-based fusion models. Comput Mater Contin. 2021;69:1407–1425. https://doi.org/10.32604/cmc.2021.018461.
7. Jung H, Kim B, Lee I, Yoo M, Lee J, Ham S, Woo O, Kang J. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PloS one. 2018;13(9):0203355. https://doi.org/10.1371/journal.pone.0203355.

8. Darma IWAS, Suciati N, Siahaan D. A performance comparison of balinese carving motif detection and recognition using YOLOv5 and mask R-CNN. In: 2021 5th International Conference on Informatics and Computational Sciences (ICICoS), 2021;pp. 52–57. https://doi.org/10.1109/ICICoS53627.2021.9651855.
9. Prinzi F, Insalaco M, Gaglio S, Vitabile S. Breast cancer localization and classification in mammograms using YoloV5. In: Esposito A, Faundez-Zanuy M, Morabito FC, Pasero E, editors. Applications of artificial intelligence and neural systems to data science. Smart innovation, systems and technologies. Vol. 360. Singapore: Springer; 2023. https://doi.org/10.1007/978-981-99-3592-5_7.
10. Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767. 2018. https://doi.org/10.48550/arXiv.1804.02767.
11. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020. https://doi.org/10.48550/arXiv.2010.11929.
12. Prinzi F, Orlando A, Gaglio S, Midiri M, Vitabile S. ML-based radiomics analysis for breast cancer classification in DCE-MRI. In: Applied Intelligence and Informatics: Second International Conference, AII 2022, Reggio Calabria, Italy, September 1–3, 2022, Proceedings. 2023;pp. 144–158. https://doi.org/10.1007/978-3-031-24801-6_11. Springer
13. Chugh G, Kumar S, Singh N. Survey on machine learning and deep learning applications in breast cancer diagnosis. Cognit Comput. 2021;pp. 1–20. https://doi.org/10.1007/s12559-020-09813-6.
14. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific Data. 2017;4(1):1–9. https://doi.org/10.1038/sdata.2017.177.
15. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. Acad Radiol. 2012;19(2):236–48. https://doi.org/10.1016/j.acra.2011.09.014.
16. Abdelrahman L, Al Ghamdi M, Collado-Mesa F, Abdel-Mottaleb M. Convolutional neural networks for breast cancer detection in mammography: a survey. Comput Biol Med. 2021;131. https://doi.org/10.1016/j.compbiomed.2021.104248.
17. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv (CSUR). 2018;51(5):1–42. https://doi.org/10.1145/3236009.
18. Lipton ZC. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue. 2018;16(3):31–57. https://doi.org/10.1145/3236386.3241340.
19. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. Xai-explainable artificial intelligence Science robotics. 2019;4(37):7120. https://doi.org/10.1126/scirobotics.aay7120.
20. Muhammad MB, Yeasin M. Eigen-CAM: class activation map using principal components. In: 2020 International Joint Conference on Neural Networks (IJCNN), 2020;pp. 1–7. https://doi.org/10.1109/IJCNN48605.2020.9206626.
21. Zhu H, Chen B, Yang C. Understanding why ViT trains badly on small datasets: an intuitive perspective. arXiv preprint arXiv:2302.03751. 2023.
22. Muduli D, Dash R, Majhi B. Automated diagnosis of breast cancer using multi-modal datasets: a deep convolution neural network based approach. Biomed Signal Process Control. 2022;71. https://doi.org/10.1016/j.bspc.2021.102825.
23. Mahmood T, Li J, Pei Y, Akhtar F, Rehman MU, Wasti SH. Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. Plos one. 2022;17(1):0263126. https://doi.org/10.1371/journal.pone.0263126.
24. Soulami KB, Kaabouch N, Saidi MN. Breast cancer: classification of suspicious regions in digital mammograms based on capsule network. Biomed Signal Process Control. 2022;76. https://doi.org/10.1016/j.bspc.2022.103696.
25. Ragab DA, Attallah O, Sharkas M, Ren J, Marshall S. A framework for breast cancer classification using multi-DCNNs. Comput Biol Med. 2021;131. https://doi.org/10.1016/j.compbiomed.2021.104245.
26. Yu X, Pang W, Xu Q, Liang M. Mammographic image classification with deep fusion learning. Sci Rep. 2020;10(1):1–11. https://doi.org/10.1038/s41598-020-71431-x.
27. Agarwal R, Diaz O, Lladó X, Yap MH, Martí R. Automatic mass detection in mammograms using deep convolutional neural networks. J Med Imaging. 2019;6(3):031409. https://doi.org/10.1117/1.JMI.6.3.031409.
28. AlGhamdi M, Abdel-Mottaleb M. DV-DCNN: dual-view deep convolutional neural network for matching detected masses in mammograms. Comput Methods Programs Biomed. 2021;207. https://doi.org/10.1016/j.cmpb.2021.106152.
29. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell. 2020;42(2):318–27. https://doi.org/10.1109/TPAMI.2018.2858826.
30. Montavon G, Samek W, Müller K-R. Methods for interpreting and understanding deep neural networks. Digit Signal Process. 2018;73:1–15. https://doi.org/10.1016/j.dsp.2017.10.011.
31. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012.
32. Kulesza T, Burnett M, Wong W-K, Stumpf S. Principles of explanatory debugging to personalize interactive machine learning. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. 2015;pp. 126–137. https://doi.org/10.1145/2678025.2701399.
33. Pocevičiūtė M, Eilertsen G, Lundström C. In: Holzinger, A., Goebel, R., Mengel, M., Müller, H. (eds.) Survey of XAI in digital pathology. 2020;pp. 56–88. Springer, Cham. https://doi.org/10.1007/978-3-030-50402-1_4.
34. Durand MA, Wang S, Hooley RJ, Raghu M, Philpotts LE. Tomosynthesis-detected architectural distortion: management algorithm with radiologic-pathologic correlation. Radiographics. 2016;36(2):311–21. https://doi.org/10.1148/rg.2016150093.
35. Oyelade ON, Ezugwu AE-S. A state-of-the-art survey on deep learning methods for detection of architectural distortion from digital mammography. IEEE Access. 2020;8:148644–76. https://doi.org/10.1109/ACCESS.2020.3016223.
36. Al-Dhabyani W, Gomaa M, Khaled H, Aly F. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int J Adv Comput Sci Appl. 2019;10(5):1–11. https://doi.org/10.14569/IJACSA.2019.0100579.
37. Kyono T, Gilbert FJ, van der Schaar M. MAMMO: a deep learning solution for facilitating radiologist-machine collaboration in breast cancer diagnosis. arXiv preprint arXiv:1811.02661. 2018. https://doi.org/10.48550/arXiv.1811.02661.
38. Redmon J, Farhadi A. Yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;pp. 7263–7271. https://doi.org/10.1109/CVPR.2017.690.
39. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016;pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
40. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings

of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;pp. 2117–2125. https://doi.org/10.48550/arXiv.1612.03144.
41. Wang C-Y, MarkLiaoH-Y, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H. CSPNet: a new backbone that can enhance learning capability of CNN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2020;pp. 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203.
42. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018;pp. 8759–8768. https://doi.org/10.1109/CVPR.2018.00913.
43. Wu B, Xu C, Dai X, Wan A, Zhang P, Yan Z, Tomizuka M, Gonzalez J, Keutzer K, Vajda P. Visual transformers: token-based image representation and processing for computer vision. arXiv preprint arXiv:2006.03677. 2020. https://doi.org/10.48550/arXiv.2006.03677.
44. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data. 2016;3(1):1–40. https://doi.org/10.1186/s40537-016-0043-6.
45. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;pp. 2921–2929. https://doi.org/10.1109/CVPR.2016.319.
46. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). 2017;pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.
47. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018;pp. 839–847. https://doi.org/10.1109/WACV.2018.00097.
48. Tan Q, Xie W, Tang H, Li Y. Multi-scale attention adaptive network for object detection in remote sensing images. In: 2022 5th International Conference on Information Communication and Signal Processing (ICICSP). 2022;pp. 218–223. https://doi.org/10.1109/ICICSP55539.2022.10050627. IEEE.
49. Li W, Huang L. YOLOSA: object detection based on 2D local feature superimposed self-attention. Pattern Recognition Letters. 2023;168:86–92. https://doi.org/10.1016/j.patrec.2023.03.003.
50. Qiu M, Christopher LA, Chien S, Chen Y. Attention mechanism improves YOLOv5x for detecting vehicles on surveillance videos. In: 2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2022;pp. 1–8. https://doi.org/10.1109/AIPR57179.2022.10092237. IEEE.
51. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. 2014;pp. 818–833. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53.
52. Ultralytics: YoloV5 Ultralytics Github. 2022. (Last accessed 24-Jan-2023). https://github.com/ultralytics/yolov5.
53. wandb: Weights & Biases. 2022. (Last accessed 24-Jan-2023). https://github.com/wandb/wandb.
54. Torrey L, Shavlik J. Chapter 11 transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. 2010;pp. 242–264. https://doi.org/10.4018/978-1-60566-766-9.
55. Zhang J, Chao H, Kalra MK, Wang G, Yan P. Overlooked trustworthiness of explainability in medical AI. medRxiv. 2021.
56. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digital Health. 2021;3(11):745–50. https://doi.org/10.1016/S2589-7500(21)00208-9.
57. Bodria F, Giannotti F, Guidotti R, Naretto F, Pedreschi D, Rinzivillo S. Benchmarking and survey of explanation methods for black box models. arXiv preprint arXiv:2102.13076. 2021.
58. ACR: American college of radiology et.al: ACR BI-RADS Atlas: breast imaging reporting and data system. Reston, VA: American College of Radiology 2014. 2013;pp. 37–78.
59. Babkina TM, Gurando AV, Kozarenko TM, Gurando VR, Telniy VV, Pominchuk DV. Detection of breast cancers represented as architectural distortion: a comparison of full-field digital mammography and digital breast tomosynthesis. Wiad Lek. 2021;74(7):1674–9. https://doi.org/10.36740/WLek202107121.
60. Rangayyan RM, Banik S, Desautels J. Computer-aided detection of architectural distortion in prior mammograms of interval cancer. J Digit Imaging. 2010;23(5):611–31. https://doi.org/10.1007/s10278-009-9257-x.
61. Arian A, Dinas K, Pratilas GC, Alipour S. The breast imaging-reporting and data system (BI-RADS) made easy. Iran J Radiol. 2022;19(1). https://doi.org/10.5812/iranjradiol-121155.
62. Agarwal R, Díaz O, Yap MH, Lladó X, Martí R. Deep learning for mass detection in full field digital mammograms. Comput Biol Med. 2020;121:103774. https://doi.org/10.1016/j.compbiomed.2020.103774.
63. Al-Antari MA, Han S-M, Kim T-S. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. Comput Methods Programs Biomed. 2020;196. https://doi.org/10.1016/j.cmpb.2020.105584.
64. Militello C, Rundo L, Dimarco M, Orlando A, Woitek R, D'Angelo I, Russo G, Bartolotta TV. 3D DCE-MRI radiomic analysis for malignant lesion prediction in breast cancer patients. Acad Radiol. 2022;29(6):830–40. https://doi.org/10.1016/j.acra.2021.08.024.