Computer Methods and Programs in Biomedicine 221 (2022) 106884

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques

Asma Baccouche[a],[*], Begonya Garcia-Zapirain[b], Yufeng Zheng[c], Adel S. Elmaghraby[a]

[a] Department of Computer Science and Engineering, University of Louisville, Louisville, KY, 40292, USA
[b] eVida Research Group, University of Deusto, Bilbao, 4800, Spain
[c] University of Mississippi Medical Center, Jackson, MS, 39216, USA

ARTICLE INFO

Article history:
Received 14 January 2022
Revised 27 April 2022
Accepted 10 May 2022

Keywords:
Breast cancer
Detection
Classification
YOLO
Prior mammogram
Early diagnosis

ABSTRACT

Background and Objective: Computer-aided-detection (CAD) systems have been developed to assist radiologists on finding suspicious lesions in mammogram. Deep Learning technology have recently succeeded to increase the chance of recognizing abnormality at an early stage in order to avoid unnecessary biopsies and decrease the mortality rate. In this study, we investigated the effectiveness of an end-to-end fusion model based on You-Only-Look-Once (YOLO) architecture, to simultaneously detect and classify suspicious breast lesions on digital mammograms. Four categories of cases were included: Mass, Calcification, Architectural Distortions, and Normal from a private digital mammographic database including 413 cases. For all cases, Prior mammograms (typically scanned 1 year before) were all reported as Normal, while Current mammograms were diagnosed as cancerous (confirmed by biopsies) or healthy.

Methods: We propose to apply the YOLO-based fusion model to the Current mammograms for breast lesions detection and classification. Then apply the same model retrospectively to synthetic mammograms for an early cancer prediction, where the synthetic mammograms were generated from the Prior mammograms by using the image-to-image translation models, CycleGAN and Pix2Pix.

Results: Evaluation results showed that our methodology could significantly detect and classify breast lesions on Current mammograms with a highest rate of 93% ± 0.118 for Mass lesions, 88% ± 0.09 for Calcification lesions, and 95% ± 0.06 for Architectural Distortion lesions. In addition, we reported evaluation results on Prior mammograms with a highest rate of 36% ± 0.01 for Mass lesions, 14% ± 0.01 for Calcification lesions, and 50% ± 0.02 for Architectural Distortion lesions. Normal mammograms were accordingly classified with an accuracy rate of 92% ± 0.09 and 90% ± 0.06 respectively on Current and Prior exams.

Conclusions: Our proposed framework was first developed to help detecting and identifying suspicious breast lesions in X-ray mammograms on their Current screening. The work was also suggested to reduce the temporal changes between pairs of Prior and follow-up screenings for early predicting the location and type of abnormalities in Prior mammogram screening. The paper presented a CAD method to assist doctors and experts to identify the risk of breast cancer presence. Overall, the proposed CAD method incorporates the advances of image processing, deep learning and image-to-image translation for a biomedical application.

1. Introduction

Breast cancer is a malignant tumor that arises from the abnormal breast cells and it is one of the dangerous diseases that threaten women worldwide [1]. According to the American Cancer Society, over 279,000 cases were reported in the United States in 2020 and it is estimated that 43,600 women will die from breast cancer in 2021 [2]. The most common symptom for breast cancer is severe change in the breast structure and in the tissue appearance, which have been also noticed with a rapid formation of breast tumors and cell clusters [3].

Mammography screening is one of the effective medical imaging tools for early breast cancer detection and diagnosis, and it can lower rates of advanced and fatal breast cancer in its early stages [4]. To inspect for potential lesions, i.e. Mass, Calcification, Archi-

[*] Correspondence author.
E-mail address: asma.baccouche@louisville.edu (A. Baccouche).

https://doi.org/10.1016/j.cmpb.2022.106884

tectural distortion, radiologists rely on human visual understanding to detect and extract all diagnostic information from mammograms [5]. However, it has been proved that about 10% to 30% of cancer cases are missed on screening mammography, which generates a false negative rate up to 50% depending on the type of lesions and the breast density [6]. With the increase of subsequent follow-ups and screenings during the diagnosis period, it has been demonstrated that about 50% of Prior mammograms have lesions visible in retrospect [7]. Consequently, it made radiologists wonder whether normal Prior mammograms without clear signs of any type of lesions could actually contain hidden information indicating a future risk of tumor appearance [8].

For these reasons, computer-aided detection (CAD) technology has been introduced over the past 30 years to improve the precision of mammography interpretation [9]. Typically, CADs are developed to localize suspicious regions of lesions that exist in the screened mammograms. The CAD approach is usually based on extracting image characteristics such as, gray levels, texture, and shape to identify regions of interest (ROI) via simple machine learning techniques [10]. So far, these techniques have not lowered the high false positive rate, nor overcome the high variation of tumors shape, size and texture.

In recent studies, deep learning has shown interest in adopting advanced models that could extract sophisticated features to localize and identify breast tumors with an equal or higher performance of human interpretation [11, 12]. With the continuous increase of mammography data availability and the existing large computational computers, deep learning algorithms have been implemented to alleviate the radiologists' effort in reading and assessing mammography images [13]. Deep learning models have shown the ability to extract deep and high-level features from raw images without knowledge assistance, and they demonstrated remarkable success for objects detection and classification in mammography [14, 15]. Such models that were widely used in the literature are considered variation of the Convolutional Neural Networks (CNNs) model, i.e. R-CNNs, Fast CNNs and Faster R-CNNs models [16]. However, one single model called You-Only-Look-Once (YOLO) was suggested to conduct the detection and classification tasks simultaneously with low memory dependence and fast results, which made it convenient for CAD application [17, 18].

Although many developments have been carried out to improve the detection accuracy of breast lesions using the deep learning techniques, there are few efforts addressed to use Prior and follow-up mammograms to simulate the disease progression and avoid unnecessary screening or overdiagnosis that cost billions of dollars annually in healthcare spending [19].

In this study, we first propose using the YOLO-based model to simultaneously detect abnormal lesions on Current mammograms and classify them into Mass, Calcification, or Architectural Distortion. Second, we investigate potential performance of the trained model to localize and label abnormal regions in Prior mammograms that were reported as normal, but later diagnosed with abnormal findings at follow-up screening. To do that, our methodology uses image-to-image translation techniques to learn at a first stage an image mapping between pairs of mammograms that generates translated Prior mammograms to overcome misalignment between screenings, and at a second stage, it predicts location and nature of future lesions' appearance at early screening.

2. Background

Since its discovery in 1913, mammography has been considered an essential key for early detection and diagnosis of specious lesions. Mammography screening has helped radiologists identify breast cancer and several studies showed its impact for a significant reduction in mortality rate [20]. With the remarkable advances in computer vision and artificial intelligence to assist doctors for medical imaging analysis, many studies showed the effectiveness of CAD systems to automatically detect suspicious lesions from raw screened mammograms [21]. The introduction of neural network models changed the CAD's approach and substituted the use of hand-crafted features extraction with deep learning architectures that are capable of learning complex features at different scales [22].

Recent studies have attempted to develop CAD models to localize the existing lesions in a fast and precise way using the different neural networks. Ribli et al. [23] developed a CAD system using the Faster R-CNN model to detect and classify breast lesions of INbreast dataset into malignant or benign and obtained an AUC score of 0.95. Similarly, Peng et al. [24] suggested an automated mass detection approach that combined Faster R-CNN architecture and a multiscale-feature pyramid network. The work achieved a true positive rate of 0.93 and 0.95 respectively on CBIS-DDSM and INbreast datasets. The study yielded a detection accuracy of up to 90% and a classification accuracy of 93.5% on the DDSM dataset. In another work by Li et al. [25], a bilateral mass detection method was introduced using two networks: a registration network between left and right breasts and a Siamese-Faster-RCNN network to detect masses from pairs of registered mammograms. They reported results of a true positive rate of 0.88 on the INbreast dataset and 0.85 on a private dataset. Another attempt by Vivek et al. [26] used a Single Shot Detector (SSD) model presented in [27] to localize breast tumors in a first step and then segment and classify regions of interest. The work achieved a true positive rate of 0.97 on the INbreast dataset.

With the progress of deep learning architectures for object detection in mammography, the You-Only-Look-Once (YOLO) model has been introduced and shown success in achieving a fast and accurate detection and classification compared to state-of-the-art methods. This was manifested by Al-masni et al. [28] who developed a CAD system using the YOLO-based model and achieved a detection accuracy of 85.52% on the DDSM dataset. In addition, Hamed et al. [29] presented a YOLOV4-based CAD system with 2-path detection of masses in full and cropped mammograms and then classified them into benign and malignant. The system succeeded with an overall detection rate of 98% and classification accuracy of 95%. In the same context, Al-masni et al. [30] proposed a CAD system framework that detected breast masses in full images using the YOLO-based model with an overall accuracy of 99.7%. Accordingly, Baccouche et al. [31] recently proposed a YOLO-based fusion model to detect breast lesions and classify them into mass or calcification. The work achieved a detection accuracy rate of 98.1% on the INbreast dataset and 95.7% on the CBIS-DDSM dataset.

Early detection and diagnosis of breast cancer in mammography using the deep learning-based CAD systems can help prevent development of tumors by marking lesions, and thus it can effectively decrease death rate [32]. A retrospective study by Watanabe et al. [33] showed a potential area of improvement for radiologists' interpretation of screening mammograms for early detection using Artificial Intelligence. The studied CAD system succeeded to mark 30 (86%) of 35 missed micro-calcifications and 58 (73%) of 80 missed masses. In addition, missed malignant lesions were flagged as early as 70 months Prior to recall or diagnostic follow-up. In consequence, CAD systems could benefit from the change that occurred between Prior and Current mammographic exams. A recent work by Timp et al. [34] tried to improve the characterization of mass lesions by adding information about the tumor behavior over time. The authors presented a CAD program to detect temporal changes between two consecutive screening images using a regional registration method in order to localize lesions detected on the current views and their corresponding on the Prior views. After that, a Support Vector Machines (SVM) classifier was applied to show the effectiveness of temporal features. In a different study, Timp et al. [35] attempted to improve detection methods by including temporal information in the CAD system. A regional registration technique along feature space was used to map suspicious locations on the Current mammograms with a corresponding location on the Prior mammograms with 72% accuracy. Accordingly, Loizidou et al. [36] tried to increase the micro-calcification detection accuracy to 99.2% by adding temporal subtraction between mammogram pairs before applying SVMs classifier. In the same context, a recent study by Loizidou et al. [37] extended their previous work of breast micro-classification detection and classification by adding an image registration step of Prior mammograms before applying temporal subtraction of pairs. In a different work by Zheng et al. [38], follow-up digital mammography images were integrated together to develop a CAD method for breast cancer detection. All regional images were detected using the Haar features, local binary pattern and histogram of oriented gradient via the AdaBoost approach and then fed into a CNN to filter out the false positives cases.

With the advent of deep convolutional neural networks, image-to-image translation has been employed to solve many computer vision applications in medical imaging. Most of the recent applications, such as image synthesis and reconstruction, which build on image-to-image translation, are based on two fundamental architectures, called Pix2Pix and CycleGAN, depending on the image's data fashion, paired or unpaired datasets [39]. A recent application by Shen et al. [40] employed the Pix2Pix network for image-to-mask segmentation in mammography. Pix2pix was also employed by Liao et al. [41] to artificially remove artifacts in CT scans and the method showed improvement for clinical image reconstruction. Moreover, a CycleGAN was successfully employed by Modanwal et al. [42] to reconstruct and harmonize MRI images for breast cancer without requiring pairs of aligned images. The effectiveness of CycleGAN was adopted in a recent work by Baccouche et al. [43] that attempted to augment the mammography data by generating synthetic images between two unpaired mammography datasets using the CycleGAN model. A recent work by Hammami et al. [44] also enhanced the multi-organ detection performance by combining CycleGAN and YOLO.

Inspired by the reviewed works and their diagnosis results, we first attempt solving the task of detection and classification of three types of breast lesions (i.e., Mass, Calcification, Architectural distortion) on most recent screening mammograms using the YOLO-based fusion models. Second, we suggest replicating early-screened mammograms with healthy diagnosis and maintaining Prior shape and appearance while predicting suspicious findings that resemble the Current mammograms. We evaluated two state-of-the-art techniques for image-to-image translation, CycleGAN and Pix2Pix and compared their performance on predicting location and type of lesions on Prior mammograms at early screening.

3. Methods and materials

3.1. YOLO-based fusion model: overview

The main method is based on our recent work [31], where the YOLO architecture model for simultaneous detection and classification of breast lesions was proposed. We upgrade the work to evaluate the capability of our previous methodology on localizing suspicious regions from the entire breast mammograms and classifying the type of lesions into Mass, Calcification, or Architectural distortion.

YOLO is a deep learning network that where a single Convolutional Neural Networks (CNNs) architecture model simultaneously localizes the bounding boxes of objects and classifies their class labels from the entire images. The YOLO-based model has had four versions, but at the time our recent work was published, the latest version was YOLO-V3, which was adopted to detect different scaled objects using the DarkNet backbone framework.

As the previous work detailed, we employed a YOLO-based model in a different evaluation fashion. The basic model was initially trained using different configurations (i.e. target class labels). Then, each experiment was evaluated by selecting the best predicted bounding boxes within all augmented images (i.e. original and rotated images) having the highest confidence score. The technique was proved an effective way to determine the best representative images to precisely detect and classify breast lesions in each mammogram. After that, as shown in Fig. 1, the idea of YOLO-based fusion models was implemented in order to improve the final prediction results. Different predictions were joined to lower the final error rate and to combine models that were differently configured. In this work, we used the same notation by referring Model1 to the YOLO-based model that was trained and configured for one class either Mass, Calcification, or Architectural Distortion. Therefore, Model2 refers to the YOLO-based model that was configured for multiple classes training (i.e. all three classes together). Finally, the Fusion Model refers to the combined evaluation of Model1 and Model2 that was used to improve the overall detection performance. The final model should select predictions that were not within the single class predictions according to a threshold of 0.5, which showed satisfying results.

All models were developed and tested on the Current mammograms from the most recent screening, with either Mass, Calcification or Architectural Distortion lesions. Different from our previous work, we added a class label, 'Normal' for the current mammograms that were not diagnosed with abnormal findings during the follow-up screening. Our trained YOLO-based model on abnormal mammograms was applied on Normal mammograms to ensure that no bounding boxes were predicted, and consequently, classify the mammograms as Normal.

3.2. Image-to-image translation technique

Deep convolutional networks have been enormously improved to provide cutting-edge solutions to computer vision and they have given the ability to manipulate images for complex image-related tasks such as image synthesis, image reconstruction, image translation, etc. Recently, these tasks were significantly treated thanks to the discovery of Generative Adversarial Networks (GANs). A standard GAN comprises two models, a generator and a discriminator. These models compete against each other to produce fake data that is realistic enough to fool the discriminator. The architecture has known success in medical imaging applications [45] and many variants were introduced such as conditional GAN (cGAN), Wasserstein Generative Adversarial Network (WGAN), etc. Further work extended the idea to create multiple GANs that can serve for synthetic data augmentation, domain adaptation, and style transfer. This allowed using a pair of generators to learn mappings of images and a pair of discriminators to learn two different types of images. The idea emphasized the image-to-image translation that leverages external labeled dataset to reconstruct effectively the source domain images with additional characteristics of a target domain such as pixels, color distribution, shape, and texture. In this context, Pix2Pix and CycleGAN are two common models that were developed to apply image-to-image translation techniques. As shown in Fig. 2, similar to the standard GAN, the two models have the target of translating images between two domains, however the difference is that Pix2Pix model works with paired datasets but only accepts one image from source domain (A) but it corrects and updates the training using its corresponding image from a target domain (B). Differently, CycleGAN model works with un-

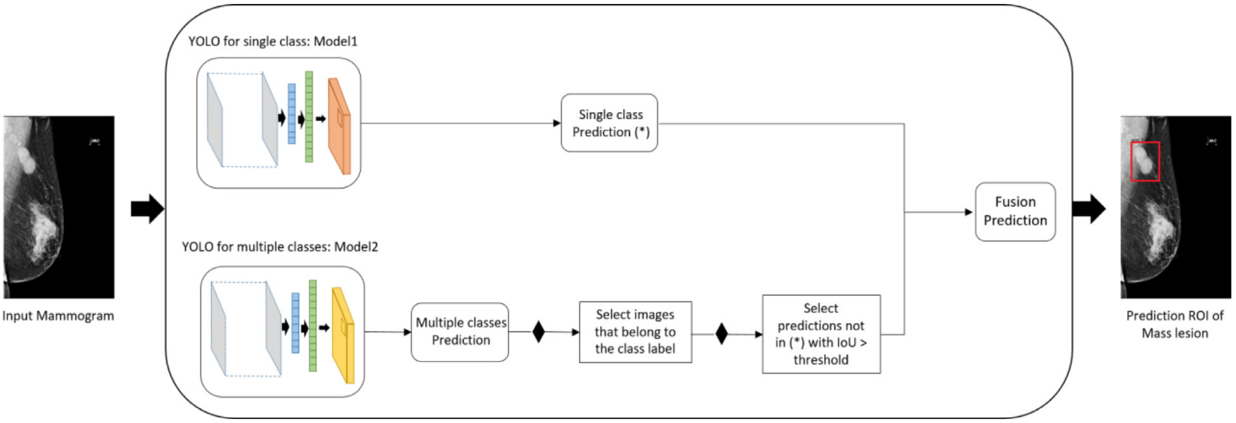
YOLO for single class: Model1
Single class Prediction (*)
Fusion Prediction
YOLO for multiple classes: Model2
Multiple classes Prediction
Select images that belong to the class label
Select predictions not in (*) with IoU > threshold
Input Mammogram
Prediction ROI of Mass lesion

Fig. 1. YOLO-based Fusion model - Example of mammogram with Mass lesion.

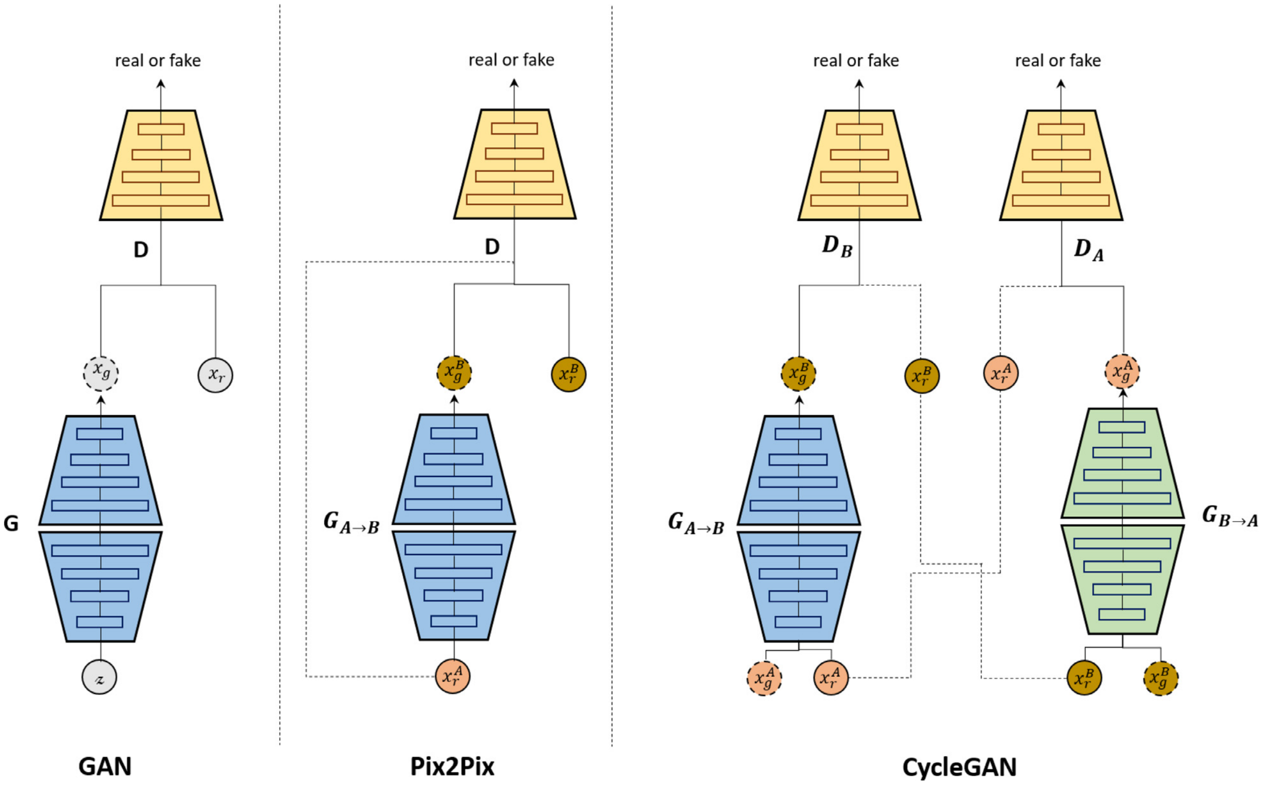

Fig. 2. Comparative scheme of standard GAN vs two variants for image-to-image translation: Pix2Pix and CycleGAN.

paired datasets, accepts two images, and performs a cyclic translation across domains to return new synthetic images.

In fact, Pix2Pix [46] is based on conditional GAN (cGAN) architecture to learn a mapping between images where the network is composed from a generator $G_{A\rightarrow B}$ and a discriminator D. The generator has an encoder-decoder structure and it tries to transfer special characteristics of an input image x_r^A to get an output image x_g^B. The discriminator uses PatchGAN architecture and it compares the input image to the generated image on one time and the input image to the corresponding image from the external dataset x_r^B another time to update the generator learning.

Moreover, the Cycle Generative Adversarial Network, called CycleGAN [47], was designed to learn mapping between images without the need to have correlations and one-to-one matches. The idea was built on the top of Pix2Pix architecture but with the use of two generators $G_{A\rightarrow B}$ and $G_{B\rightarrow A}$ for cycled images mapping and two discriminators D_A and D_B to distinguish between real and synthetic images. Additionally, the CycleGAN technique employs a cycle consistency for the generators to ensure a good reconstruction of the new image back to their original look. Consequently, the technique helps to capture both domains' features and style without mismatch.

3.3. Early detection and classification framework

In this work, we first apply and evaluate the YOLO technique on the Current mammograms to detect different breast lesions and classify them into Mass, Calcification, or Architectural Distortion, and the rest to Normal. Second, we consider two image-to-image techniques, Pix2Pix and CycleGAN, to learn mapping between Current mammograms and their corresponding Prior mammograms. As shown in Fig. 3, new synthetic Prior mammograms are generated to overcome the misalignment between the screenings due to temporal and texture changes. Next, the trained models on the first

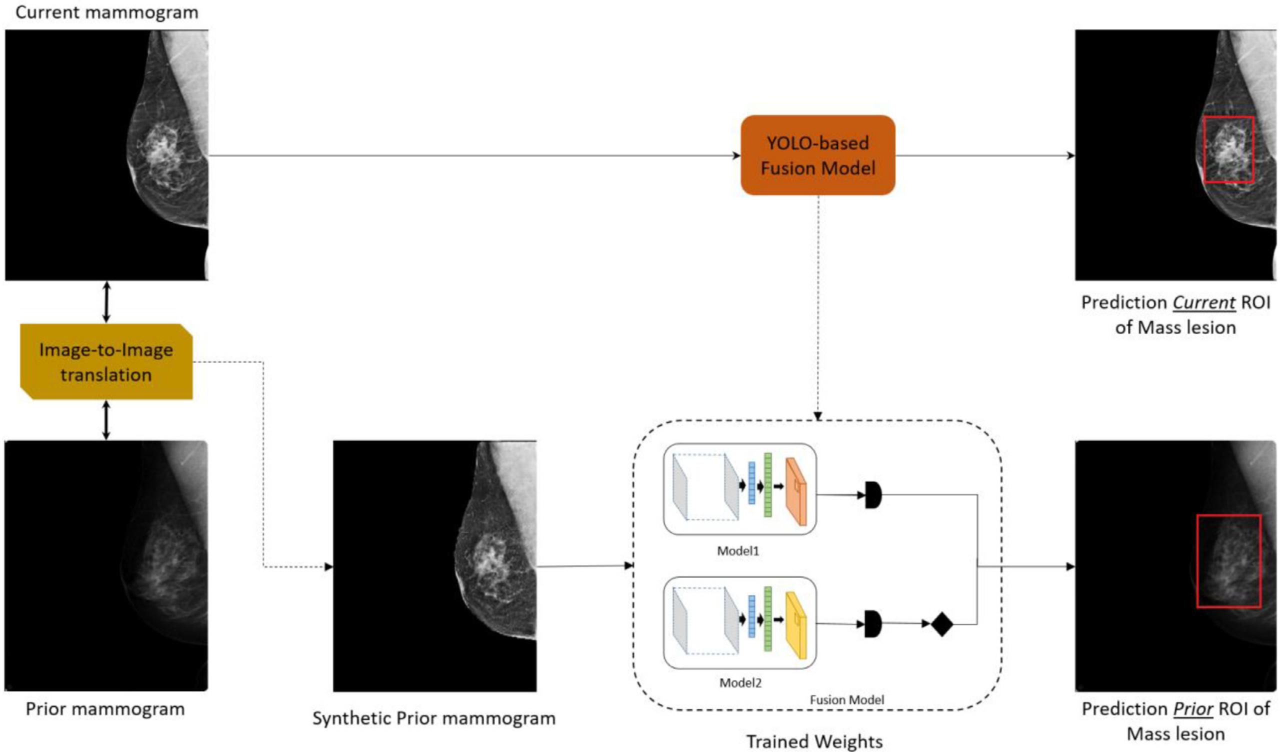
Current mammogram
YOLO-based Fusion Model
Prediction Current ROI of Mass lesion
Image-to-Image translation
Prior mammogram
Synthetic Prior mammogram
Model1
Model2
Fusion Model
Trained Weights
Prediction Prior ROI of Mass lesion

Fig. 3. Framework for early detection and classification on Prior mammograms – Example of Prior mammogram with normal diagnosis and Current mammogram with Mass lesions (red bounding boxes).

step are used to predict the location and type of breast lesions on the translated Prior mammograms. Predicting the bounding boxes for suspicious lesions of "future cancers" in Prior mammograms is challenging. Therefore, we integrate all diagnostic information into one framework that explores possible evidence of invisible patterns for indicating the risk of "future cancer". Inference models are directly applied on translated Prior mammograms and evaluation was carried out using true bounding boxes' positions and class labels of their corresponding Current mammograms.

4. Results

All experiments using the proposed methods were conducted on a PC with the following specifications: Intel(R) Core (TM) i7-8700K processor with 32 GB RAM, 3.70 GHz frequency, and one NVIDIA GeForce GTX 1090 Ti GPU. Python 3.6 was used for conducting all experiments.

4.1. Dataset

In this study, we used a collection of private dataset from the University of Connecticut Center (UCHC), named *UCHC DigiMammo* (UCHCDM) database [48]. The dataset contains screening mammograms of 230 patients, where each case had an initial screening, called Prior exam, and a second follow-up screening between 1 to 6 years, called the Current exam, and a sample is displayed in Fig. 4.

Each screening in the dataset acquires two different views, CC and MLO. All images were saved with the Digital Imaging and Communications in Medicine (DICOM) format, and were annotated by expert radiologists in a description text file with corresponding pathology of a mammographic finding (i.e. Mass, Calcification, Architectural Distortion, Normal), as detailed in Table 1a and Table 1b. Pixel-level ground-truth images were also provided separately where suspicious locations were circulated. A total of 413 mammograms are considered separately for Current and Prior exams, and they have an average size of 2950 × 3650 pixels.

Table 1a
Overall Data Distribution – Current and Prior Exams.

Total number of patients with pairs and pathology	230
Total number of Mammograms	833
Total number of Mammograms with pathology	826
Total number of Prior Mammograms (Normal)	413
Total number of Current Mammograms	413

Table 1b
Detailed Data Distribution – Current Exams.

Number of Current Mammograms with Mass Lesions	181
Number of Current Mammograms with Calcification Lesions	116
Number of Current Mammograms with Architectural Distortion Lesions	74
Number of Current Mammograms without Lesions (Normal)	42

4.2. Data preparation

All mammograms were collected using a digital X-ray mammography tool that compressed and stored the images in DICOM format. Therefore, we applied some preprocessing steps using the denoising and the histogram equalization methods to all original images to improve the quality prior to training process. Due to large sizes of original DICOM images, all mammograms were down-sampled using a bi-cubic interpolation over a 4 × 4 neighborhood. In our experiments, we used image' sizes of 448 × 448 pixels (i.e. divisible by 32 according to DarkNet backbone architecture of YOLO-V3), which can fit in our GPU memory. Finally, all training images were normalized to have the intensity values in the range of [0, 1]. Samples of original and preprocessed images are illustrated below in Fig. 5.

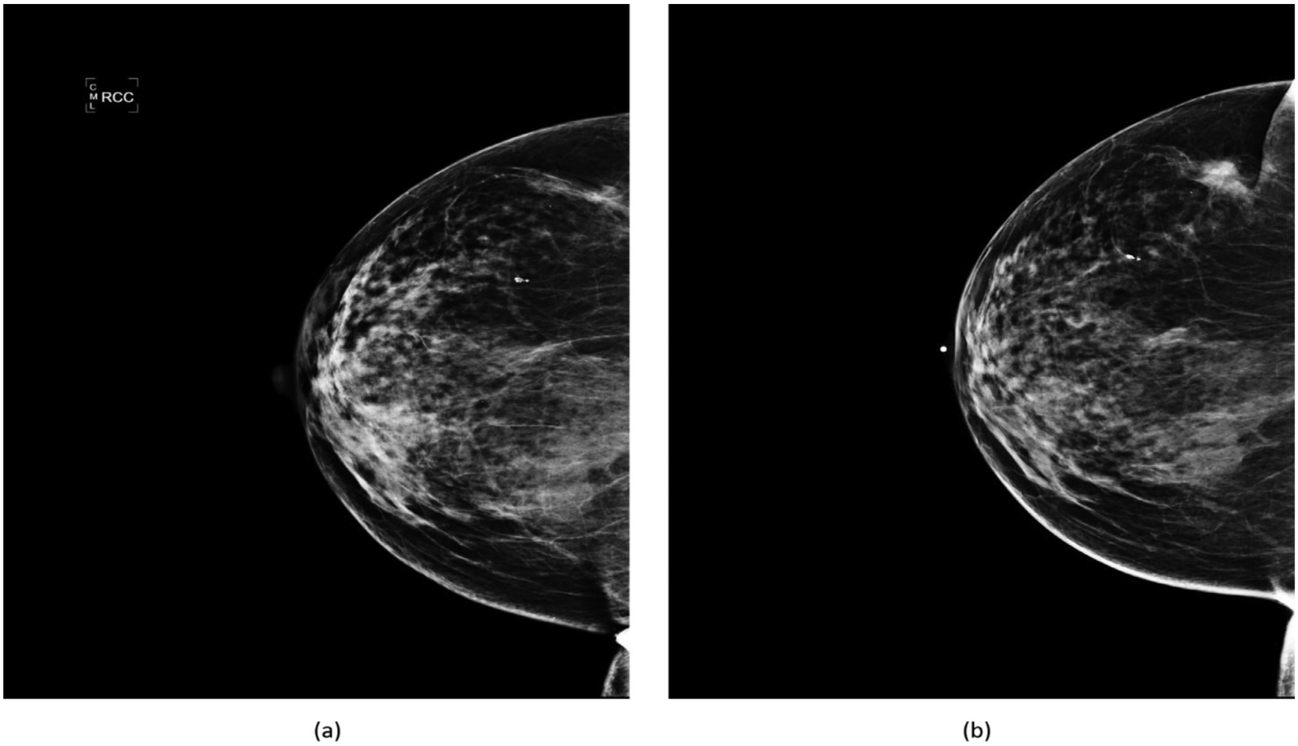

Fig. 4. Sample of Prior and Current mammograms screenings (2.5 years) for Case# 31, Right CC View. (a) Prior exam with Normal mammogram (i.e. No diagnosis). (b) Current exam with Mass present.

Deep learning models require a large number of data to ensure a fast learning convergence and generalized inference. However, medical datasets suffer from a shortage of annotated images because it is hard to collect and label medical images. To solve this problem, data augmentation techniques were mainly suggested to increase the dataset's size by rotating or flipping instances. In this paper, we rotated original images four times with the angles $\Delta\theta = \{0^\circ,\ 90^\circ,\ 180^\circ,\ 270^\circ\}$. Consequently, a total number of 1,652 mammograms were generated for UCHCDM dataset to train and test the model. Original samples from each class are shown below in Fig. 6.

4.3. Evaluation metrics and experiments settings

In this study, we used object detection and classification metrics to measure the performance of YOLO-based models. To evaluate the detection of breast lesions' location in the mammograms and their type, we first measured the intersection over union (IoU) score between each detected box and its corresponding ground-truth (i.e. (x, y, h, w) coordinates and class label), and then we verified if it exceeded a confidence score threshold of 0.35. Eq. (1) details the IoU score formula.

$$IoU\ score = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{1}$$

After that, we reported a final objective measure, called detection accuracy rate, which considered the predicted class probability of true detected boxes. Inspired by the work of Samuelson et al. [49] and recently adapted in the work [31], we computed the number of true detected images within lesions' type (i.e. Mass, Calcification, Architectural distortion) and Normal images over the total number of mammograms used, as defined in Eq. (2).

$$Detection\ accuracy = \frac{True\ detected\ cases}{Total\ number\ of\ cases} \tag{2}$$

Hence, the suggested measure allows removing all cases that have a lower IoU score (i.e. low detection precision) before reporting the final detection accuracy rate. Therefore, the predicted boxes that had confidence probability scores equal or greater than the confidence score threshold were only considered. We measured the detection accuracy rate overall and separately for each class label to evaluate the performance of the YOLO-based model.

Additionally, we particularly reported the Current mammograms prediction results using the area under curve (AUC) that reflects the performance of the model and the trade-off between the true positive rate and false positive rate for each target class label. We used three additional metrics called precision, recall, and sensitivity that are computed using the TP, FP, and FN that are defined per predicted class to represent the number of true positive, false positive, and false negative predictions, respectively as shown in Eq. (3), Eq. (4), and Eq. (5).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Sensitivity = 1 - FNR = 1 - \frac{FN}{FN + TP} \tag{5}$$

Experiments for the image-to-image techniques that were conducted using the CycleGAN and Pix2Pix models were trained accordingly on unpaired and paired datasets images. The CycleGAN model was based on the available tutorial in Keras webpage (https://keras.io/examples/generative/cyclegan). The architecture model has two generators and two discriminators networks. The generator network consists of two down-sampling blocks with filter sizes [128, 256], nine residual blocks with filter size 256, and two up-sampling blocks with filter sizes [128, 64]. The discriminator network is based on four down-sampling blocks with filter sizes [64, 128, 256, 512]. For Pix2Pix model, we similarly used two

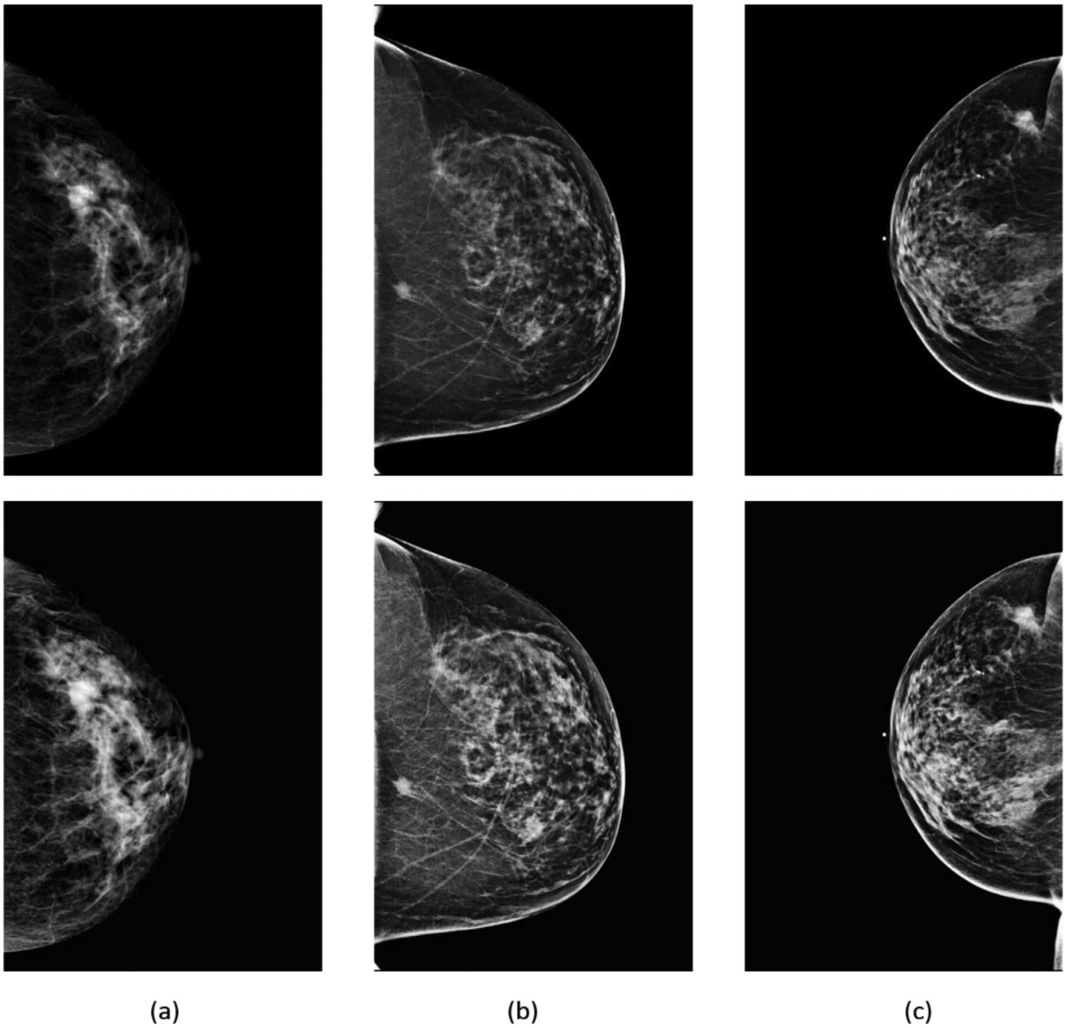

Fig. 5. Samples from Current exams of original (upper row) and preprocessed mammograms (bottom row). (a) Case# 9: Left CC View. (b) Case# 14: Left CC View. (c) Case# 31: Right CC View.

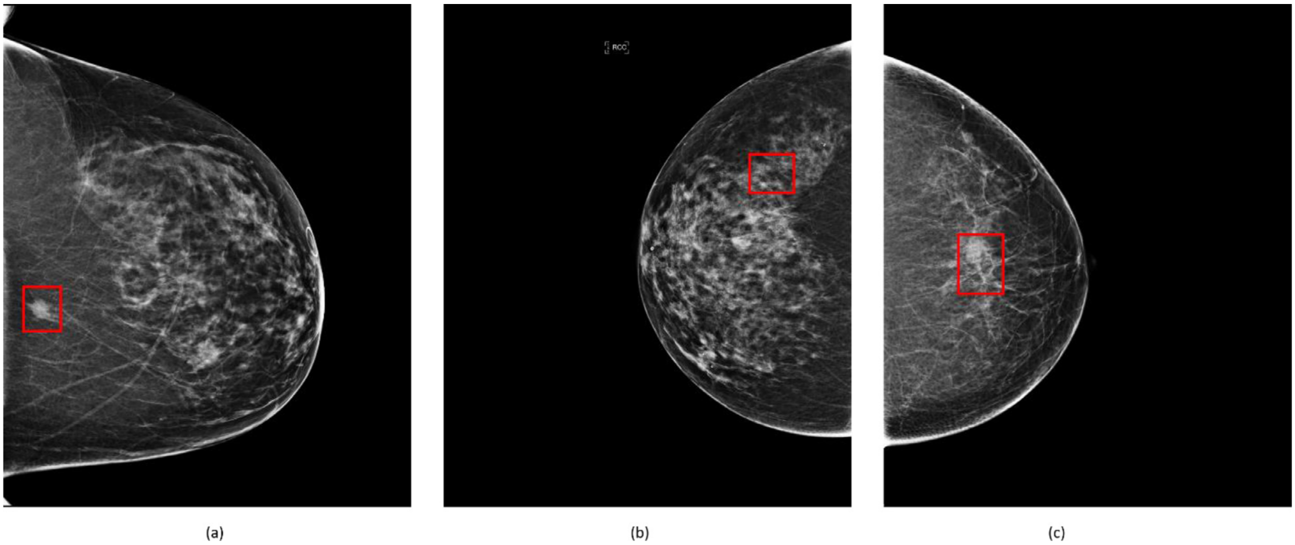

Fig. 6. Samples of original mammograms with red bounding boxes of ground-truth. (a) Case# 14: Current exam with Mass present, Left CC View. (b) Case# 220: Current exam with Calcification present, Right CC View. (c) Case# 27: Current exam with Architectural Distortion present, Left CC View.

Table 2
Cross Validation Folds: Data distribution across class labels.

Data	Breast Lesions			Normal	Overall
	Mass	Calcification	Architectural Distortion		
Training	578	373	243	130	1,324
Testing	144	92	60	32	328
Total	722	465	303	162	1,652

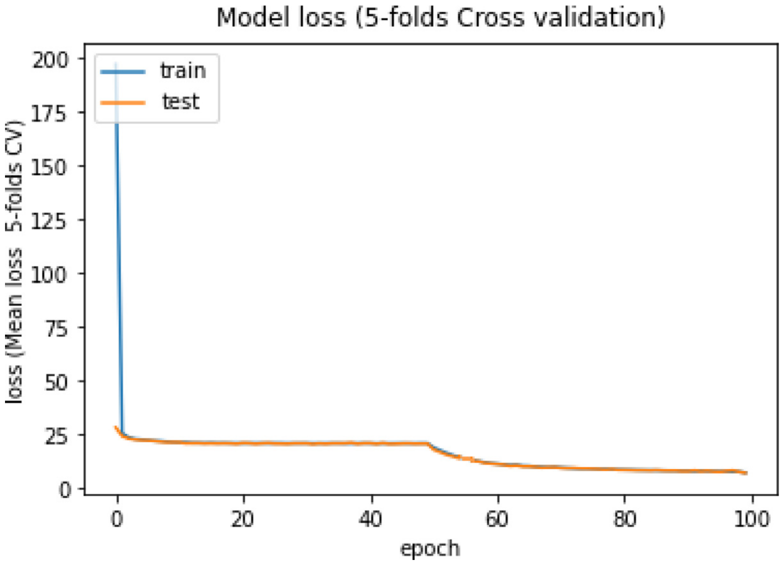

Fig. 7. Learning curve plot between Training and Testing sets.

generators and two discriminators networks. The generator network contains seven down-sampling and up-sampling blocks with filter sizes [64, 128, 256, 512, 512, 512, 512]. We used the same discriminator network from the CycleGAN architecture model. Hence, the two models were trained and evaluated on 100 epochs and optimized using the Adam technique with learning rate of 0.0002 and beta score of 0.5.

To ensure the model robustness, we performed a 5-fold cross validation by training and testing the model using different test sets of random mammograms. Consequently, the entire dataset was randomly divided into equal 5 folds of 1,324 training images (80%) and 328 testing images (20%) with respect to the imbalanced classes as detailed in Table 2. Finally, we reported the average of results over all the folds.

For all experiments using the YOLO-based model, we set the learning rate to be 0.001, the batch size to be 8, and the number of epochs to be 100. The loss function combined the bounding box regression loss, the class label loss, and the confidence loss. All functions were based on cross-entropy and they were scaled to handle the imbalance of class labels on each batch. In addition, the early stopping method was used for the second half of iterations to dynamically reduce the learning rate by 10% every 10 epochs in case of constant loss function value. In order to prevent overfitting, all models were initialized by weights from a pre-trained model on a large public dataset, Microsoft COCO. Then, the models were re-trained and new layers were fine-tuned on our mammography dataset. As a consequence, we only monitored the learning curve with the loss function that was iteratively dropped and optimized during the epochs. As shown below in Fig. 7, there was no overfitting observed during the learning.

4.4. Evaluation of YOLO-based model on current mammograms

First part of the study considered only Current mammograms from the most recent screening exams. The YOLO-based models were trained differently over the Current views of the UCHCDM dataset. We varied the models according to the input dataset and the target class. Hence, Model1 was configured for single classes and Model2 was configured for mixed classes. Finally, the Fusion Model was designated to combine Model1 and Model2 for each target class according to the approach described in [31]. Table 3 shows quantitative comparison of the detection accuracy rate and count that were reported using the 5-fold cross validation as $\mu \pm \sigma$, where μ and σ refer to the mean and standard deviation, respectively.

Results show the advantage of the adapted Fusion Model and confirm its highest results overall and for each class label. Fusion Model had the highest score of 95% for Architectural Distortion lesions and a score of 92% overall. Moreover, results in Table 3 show the ability of YOLO architecture to detect and classify the breast lesions with a maximum accuracy rate of 93% for mammograms with Mass lesions, 88% for mammograms with Calcification lesions, and 95% for mammograms with Architectural Distortion lesions. Appropriately, Normal mammograms were also correctly classified with a maximum accuracy score of 94% where no bounding boxes were detected. All experiments had similar inference time with a maximum value of 0.62 seconds per image.

Additionally, to get a better understanding of the models' performance, we generated the free-response receiver operating characteristic (FROC) curves to illustrate the number of false positives per image (FPI) for each target class label. Plots of the FROCs between Average sensitivity and the average number of false positives are shown in Fig. 8 that specifically compares between Model1, Model2, and the Fusion Model.

By varying the threshold and the range of false positive between 0.05 and 0.20 overall, we could achieve an average sensitivity between 0.7 and 0.95 for all cases. Fig. 8 clearly shows that the Fusion model had the highest performance compared to the other evaluated models. It is observed that the proposed model could obtain an average sensitivity of more than 0.90 with an average FPI

Table 3
Comparison performance for different models across labeled classes on Test sets.

Models	Results	Breast Lesions			Normal	Overall	Inference time per image (sec)
		Mass	Calcification	Architectural Distortion			
	Count	113	74	49	25	261	
Model1	$\mu \pm \sigma$	79% ± 0.09	80% ± 0.05	82% ± 0.03	78% ± 0.01	79% ± 0.04	0.60
	Count	110	79	51	28	268	
Model2	$\mu \pm \sigma$	76% ± 0.03	86% ± 0.04	85% ± 0.02	87% ± 0.05	82% ± 0.03	0.62
	Count	135	81	57	30	303	
Fusion Model	$\mu \pm \sigma$	93% ± 0.118	88% ± 0.09	95% ± 0.06	94% ± 0.11	92% ± 0.09	0.62

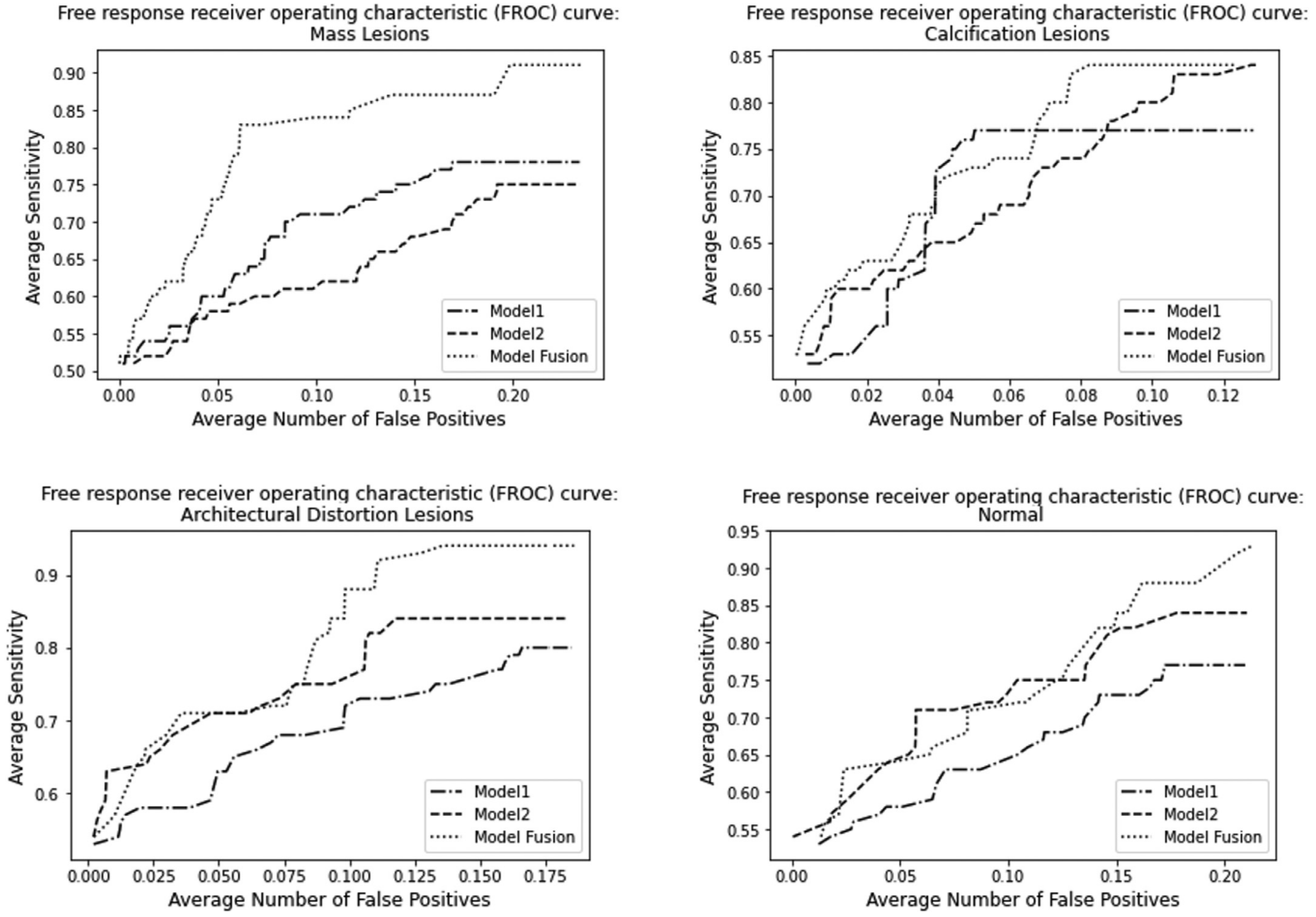

Fig. 8. FROC curve plots of the YOLO based proposed Detection and Classification models per class label on Test sets.

of 0.20 for Mass lesions, an average sensitivity of more than 0.85 with an average FPI of 0.12 for Calcification lesions, and an average sensitivity of more than 0.90 with an average FPI of 0.175 for Architectural Distortion lesions. Accordingly, Normal cases in Current views were evaluated using the FROC analysis and a false positive was considered when no detection should be occurred in a non-cancerous case but it was missed by the model. It is to notice that we could obtain an average sensitivity of around 0.95 with an average FPI of 0.20.

Finally, we analyzed the performance results with a particular focus on the classification task that was conducted by the YOLO-based Fusion model. Table 4 explores the calculated classification metrics by each class label, where we achieved the highest sensitivity of 94.11% on the cancer cases with Architectural Distortion and a sensitivity of 92.09% on the non-cancerous cases.

Additionally, Fig. 9 illustrates a visual comparison of the trade-off between the false positive rate (FPR) and the true positive rate (TPR) according to the ROC curve plot between the different cases. We observed a highest AUC score of 0.95 for the Mass and the Architectural Distortion cases, and an AUC score of 0.96 for the Normal cases. The low results for the Calcification lesions could be explained with the fact that this type of breast lesions do not have standard shape and location and they are often small and randomly distributed which can limit the automatic detection [31].

Moreover, Fig. 10 illustrates the confusion matrix for the classification of the true detected bounding boxes applied on the Current mammograms, where three types of lesions are presented with the Normal cases (i.e. correct prediction without detected lesions). Clearly, the prediction error for different classes is low with a high rate of 6.2% corresponding to the Normal class label and 7.6% corresponding to the Calcification cases. The distribution of error within classes could be explained by the inability of YOLO-based model to detect and distinguish some different types of lesions having similar shapes such as Calcification and Architectural

Table 4
Performance results for Detection and Classification on Test sets.

Class Label	Accuracy	Precision	Recall	Sensitivity	AUC
Mass	0.94	0.94	0.94	0.93	0.95
Calcification	0.93	0.88	0.88	0.88	0.94
Architectural Distortion	0.98	0.95	0.95	0.94	0.95
Normal	0.98	0.94	0.94	0.92	0.96

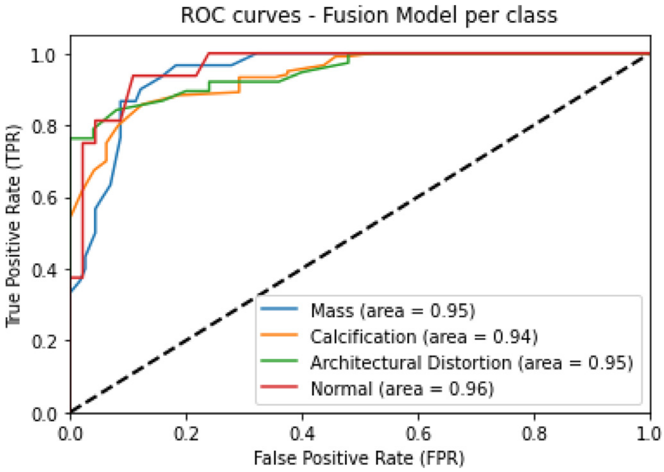

Fig. 9. ROC curve plots of the proposed YOLO-based Fusion Model per class label on Test sets.

Distortion that often have irregular shape in challenging positions within the breast.

4.5. Evaluation of YOLO-based model on prior mammograms

Second part of the study focused on using the pairs of mammograms, including current views and their Prior screening exams in order to provide an early detection and classification of lesions on the Prior screening exams. All Prior mammograms were not annotated with diagnosis and thus were considered Normal (i.e. non-cancerous, corresponding to 0s in “Experts prediction” row in Tables 5a, 5b, and 5c). In this part, we introduce a retrospective approach to look back at the Prior mammograms and try to explore any patterns of breast lesions before waiting on a follow-up screening.

Our methodology is based on joining the learned mapping between the temporal views and a trained model on Current mammograms that were annotated by experts. First, the pairs of datasets were prepared using the same configuration, and two image-to-image translation models were trained between the two datasets to determine the images mapping. Consequently, synthetic mammograms from Prior screening exams were generated to resemble the Current mammograms and preserve the general texture of the Prior mammograms.

After that, the YOLO-based model that was trained and validated previously on the Current mammograms, was saved and used for inference on Prior mammograms. Experiments were evaluated using only the Fusion Model that showed the highest performance in Section 4.4. We first evaluated the performance using the original Prior mammograms without image-to-image translation and later compared to the Prior mammograms that were translated using the CycleGAN and Pix2Pix techniques.

Table 5a, 5b and 5c present results of early prediction on Prior mammograms that are reported using the 5-fold cross validation as $\mu \pm \sigma$, where μ and σ refer to the mean and standard deviation, respectively.

We considered a true prediction where the location and type of breast lesions were correctly captured using the inference model, retrospectively on non-cancerous screening views at t=0 years. The inference evaluation was concluded using the ground-truth labels of the Current views that were generated by experts later at t=[1, 6] years. Consequently, all predictions should be fairly compared to 0 predictions (i.e. all were missed) by experts at t=0.

Results in Table 5a, 5b and 5c show the count and percentage of mammograms for each class and overall, that were correctly

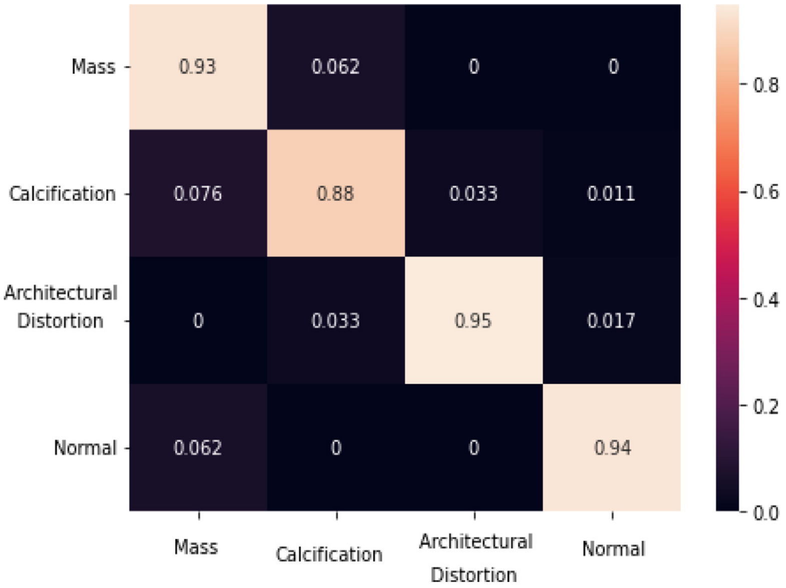

Fig. 10. Confusion matrix of prediction results for Current Mammograms.

Table 5a
Inference results of YOLO Fusion model on Test sets of original Prior Mammograms.

Results for Prior Mammogram Prediction	Breast Lesions			Normal	Overall	Inference time per image (sec)
	Mass	Calcification	Architectural Distortion			
True prediction	33	16	19	26	94	
$\mu \pm \sigma$	22% ± 0.09	17% ± 0.07	31% ± 0.06	81% ± 0.02	28% ± 0.06	0.62
Experts prediction	0	0	0	0	0	
False prediction	111	76	41	6	234	
$\mu \pm \sigma$	77% ± 0.08	82% ± 0.16	68% ± 0.03	18% ± 0.17	71% ± 0.03	

Table 5b
Inference results of YOLO Fusion model on Test sets of Prior Mammograms using CycleGAN for image-to-image translation.

Results for Prior Mammogram Prediction	Breast Lesions			Normal	Overall	Inference time per image (sec)
	Mass	Calcification	Architectural Distortion			
True prediction	32	10	22	26	91	
$\mu \pm \sigma$	22% ± 0.02	10% ± 0.08	36% ± 0.06	81% ± 0.02	27% ± 0.07	0.63
Experts prediction	0	0	0	0	0	
False prediction	112	82	38	6	237	
$\mu \pm \sigma$	77% ± 0.07	89% ± 0.03	63% ± 0.13	18% ± 0.07	72% ± 0.02	

Table 5c
Inference results of YOLO Fusion model on Test sets of Prior Mammograms using Pix2Pix for image-to-image translation.

Results for Prior Mammogram Prediction	Breast Lesions			Normal	Overall	Inference time per image (sec)
	Mass	Calcification	Architectural Distortion			
True prediction	52	13	30	29	124	
$\mu \pm \sigma$	36% ± 0.01	14% ± 0.01	50% ± 0.02	90% ± 0.06	37% ± 0.1	0.63
Experts prediction	0	0	0	0	0	
False prediction	92	79	30	3	204	
$\mu \pm \sigma$	63% ± 0.08	85% ± 0.08	50% ± 0.01	9% ± 0.03	62% ± 0.12	

predicted at Prior views and considered for an early detection and classification. All true predictions presented two scenarios; one for all correct prediction on both Current mammograms and their corresponding Prior views from the first exams (i.e. t=0), and another scenario for only correct prediction on Prior mammograms even though their corresponding Current views were not correctly predicted.

It is observed that the highest results were reported by the YOLO-based model that was inferred on synthetic Prior mammograms by Pix2Pix technique, where a total number of 52 mammograms (36% ± 0.01) were accurately anticipated. We also noticed a high percentage of 36% ± 0.01 was shown for Mass lesions, 14% ± 0.01 for Calcification lesions, and 50% ± 0.02 for Architectural Distortion lesions. In addition, 90% ± 0.06 of Normal mammograms were accordingly classified on Prior exam screenings. The inference time per each configuration was reported with a maximum value of 0.63 seconds per image.

Consequently, the Pix2Pix model indicates the most effective technique for image-to-image translating mammograms from Prior to Current appearance in order to help increase the number of correct detection and categorization of breast lesions at t=0. An overall true prediction rate of 37% was reported using the proposed methodology that reveals the success of our suggested framework to help an early diagnosis without the urgent need of a follow-up screening that might occur a late stage for breast cancer.

We also reported the false prediction rate that counted the missed cases on Prior views by the inference model. The reported numbers could be explained by the fact that we did not train the model on Prior views as they were annotated by experts as being Normal at t=0.

Although the gold standard of the retrospective comparison we presented is 0 predictions at t=0, we also noticed a drop of 9% on the false prediction using the synthetic Prior images that were generated by Pix2Pix model for image-to-image translation with an overall value of 62%.

Additionally, Fig. 11 illustrates the confusion matrix for the classification of the true detected bounding boxes on the Prior mammograms, where three types of lesions are presented with the Normal cases (i.e. correct prediction without detected lesions). It is clear that prediction error for different classes is low with a maximum rate of 3.5%.

4.6. Retrospective analysis for the early detection and classification

In this part, we investigate the follow-up exam time (i.e. originally between 1 to 6 years) of the true early prediction results for each class label at Section 4.5. Fig. 12 illustrate retrospectively that our suggested methodology is capable to anticipate the presence of breast lesions that were originally diagnosed at a later exam (i.e. t>0). In particular, Mass lesions were predicted beforehand but later detected by experts and radiologists within 2 to 3.5 years. The comparative figure also represents the latest follow-up exam time of results from using the image-to-image translation techniques versus the original mammograms.

Moreover, Fig. 13 illustrates a comparison of inference results of the YOLO-based model with and without image-to-image translation across the different classes and overall. We can visually conclude that the Pix2Pix translation method had the best performance overall, which could also be explained by the fact the Pix2Pix model was trained between paired images, compared to the CycleGAN model that used unpaired images. Hence, the Pix2Pix model is more efficient than the CycleGAN for the particular task of image-to-image translation considering the advantage of image alignment it presents between the paired datasets.

It is also to notice that the two image-to-image translation methods yielded to better results than using original images (i.e. No translation) for all different lesions except for the Calcifications.

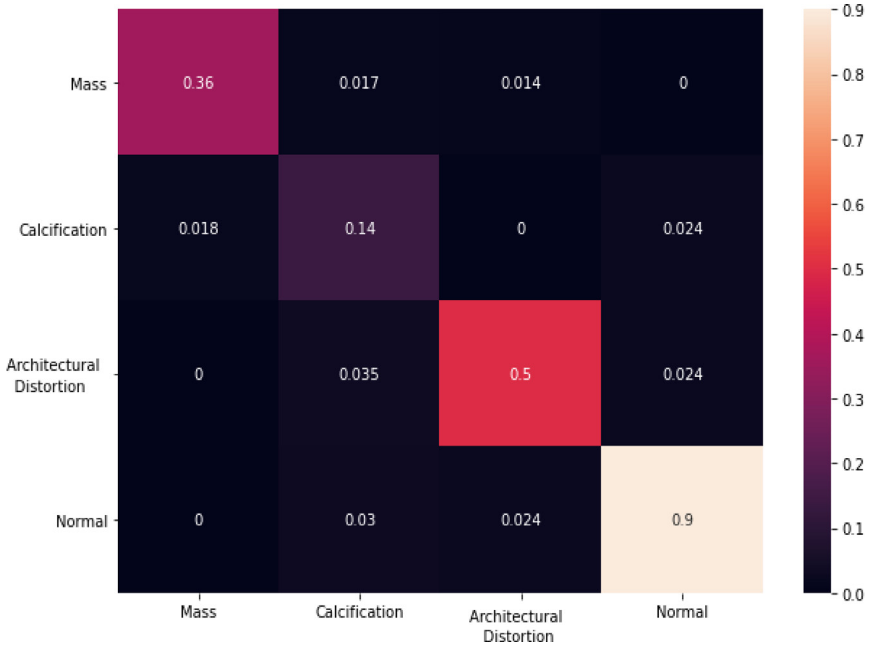

Fig. 11. Confusion matrix of prediction results for Prior Mammograms.

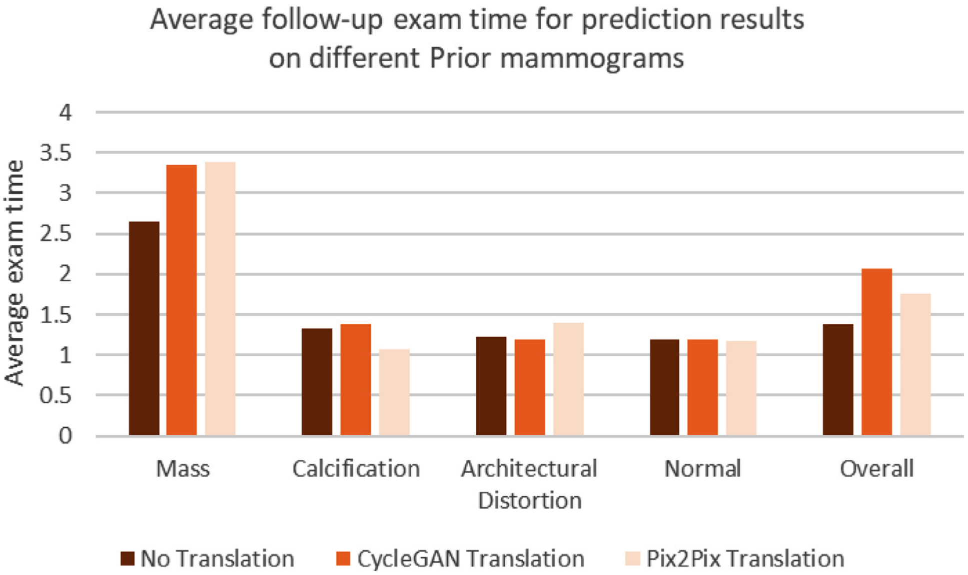

Fig. 12. Comparison of Mean follow-up exam time for prediction results across classes with and without image-to-image translation Pix2pix and CycleGAN on Prior mammograms.

This particularly can be explained with the difference between the types of breast lesions in terms of shape, size and texture. It is commonly known that calcifications do not appear in standard shape and location and they can be bilateral, thick learn, clustered, pleomorphic and vascular, etc. [31, 50]. Due to their irregular size and position, replicating such of abnormities using image-to-image synthesis did not help the detection and identification of calcifications lesions in prior views. The calcifications are often small and clustered and they require a smooth pixel distribution, however the x-rays images could be degraded and the breast calcifications could be hard to be identified at early state [51].

Furthermore, we compared results of the early detection and classification across the Prior exam's time that varied between 1 to 6 years. Fig. 14 provides a visual observation of the percentage of correctly predicted Prior mammograms for each class label using the best-reported experiment (i.e. using the Pix2Pix translation). It is clear that the follow-up exam time of 1 year had the highest rate of predicted images. This emphasizes the success of our methodology to early localizing and identifying lesions that are often considered the hardest to diagnose. Another observation is that our methodology captured the Mass lesions that had follow-up requests of later than 3 years, which might be too late to diagnose patients with Mass breast lesion.

Finally, two samples of mammograms that were taken from different patients, including Prior exam views and their corresponding Current exam views are shown below in Fig. 15. Two cases of results are demonstrated: 1) When both Current and Prior mammograms were correctly predicted, and 2) When only the Prior mammograms were correctly predicted. It seems that when the model failed to predict lesions in some Current mammograms, their corresponding Prior mammograms were successfully predicted using the inference model. Predicted bounding boxes were

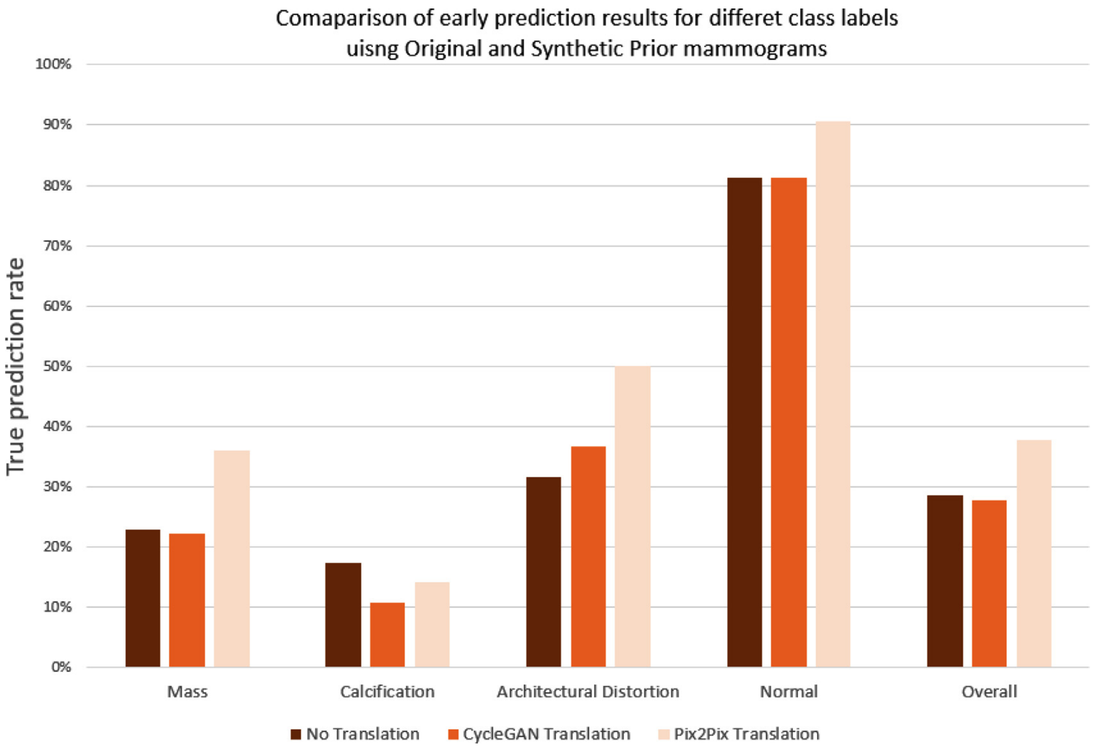

Fig. 13. Comparison performance of YOLO Fusion model across classes with and without image-to-image translation.

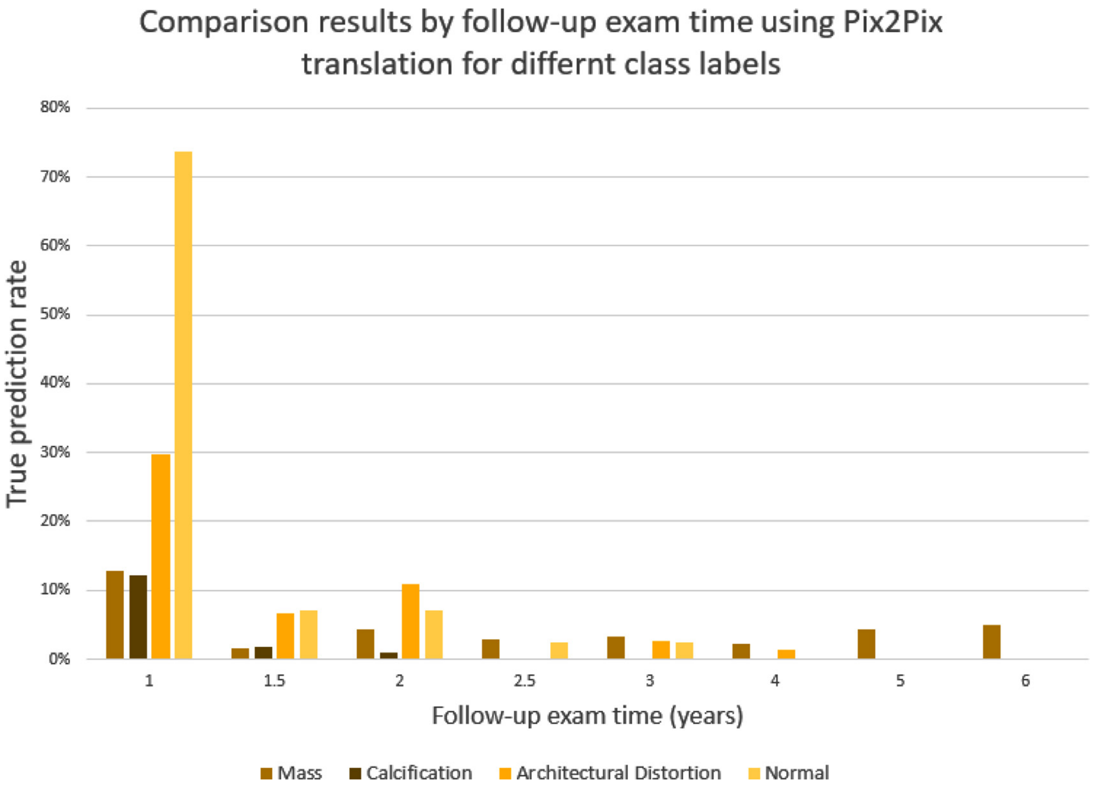

Fig. 14. Comparison performance of YOLO Fusion model and Pix2Pix for image-to-image translation for different classes across follow-up exam time (years).

slightly different between views but they exceeded the threshold score, and this could be explained with the different quality of the acquired images and the type of the detected lesions.

Moreover, Normal mammograms were shown in the last row accordingly where correct predictions were demonstrated for both screening and for only Predicted mammograms.

Finally, a comparison of latest studies and similar methods were reported against our proposed methodology. For a complete and fair comparison, only works that were applied for Mass lesions detection were reported and compared in Table 6. Comparing both detection accuracy rate and inference time with the other works that were evaluated on the public datasets CBIS-DDSM, INbreast,

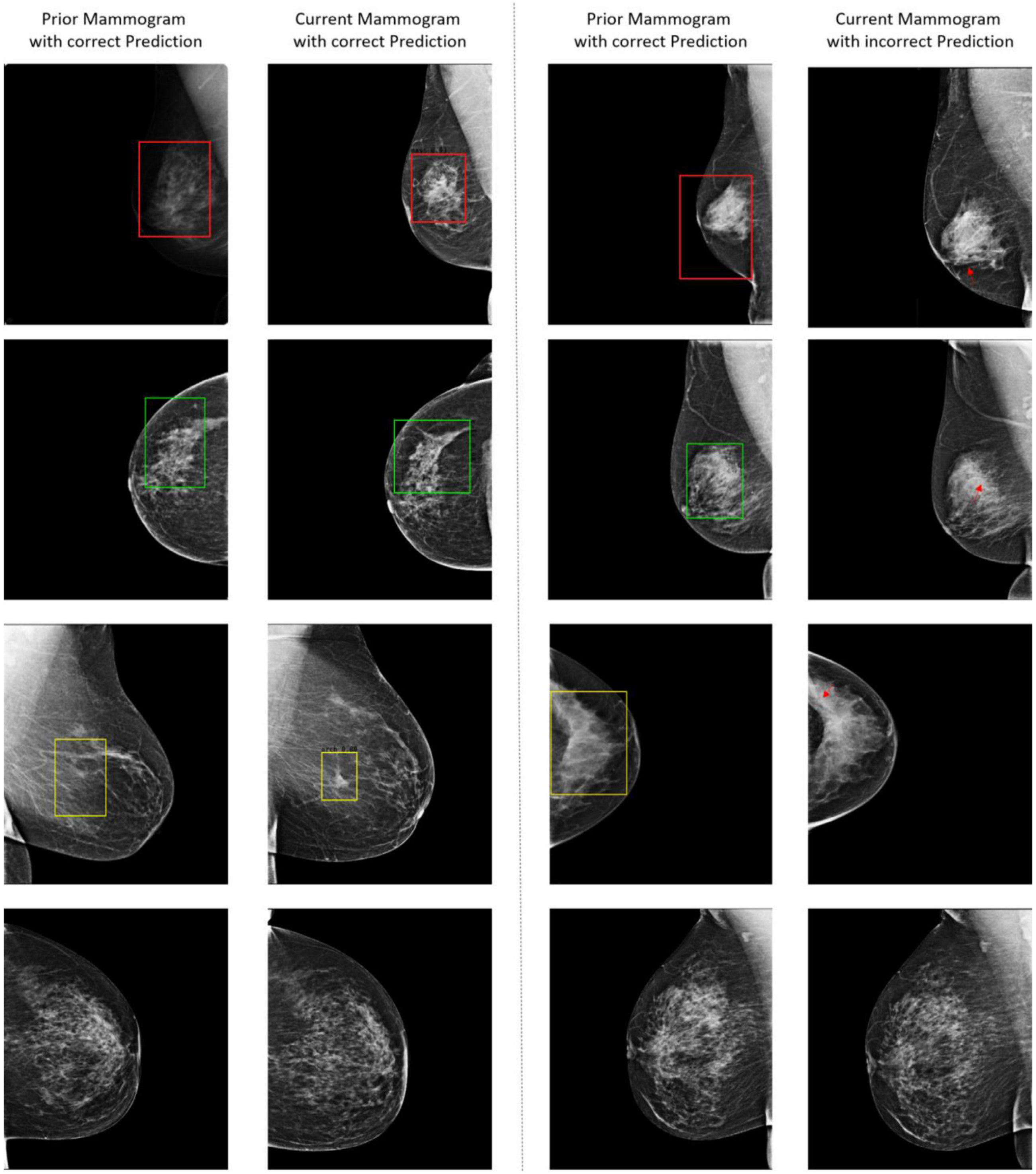
Prior Mammogram with correct Prediction
Current Mammogram with correct Prediction
Prior Mammogram with correct Prediction
Current Mammogram with incorrect Prediction

Fig. 15. Example results on Prior mammograms vs Current mammograms using the YOLO-based Fusion model that was inferred on the translated Prior images from Pix2Pix method across classes: Mass (red bounding boxes, first row), Calcification (green bounding boxes, second row), and Architectural Distortion (yellow bounding boxes, third row). Red arrows point to the ground truth location. Last row belongs to the Normal class.

and MIAS, our YOLO-based fusion models achieved overall better than previous works. Our recent work was considered having the best trade-off between the detection accuracy and testing time comparing to the work by Peng et al. [24] that had a better inference time of 0.134 second per image but it only had a detection accuracy rate of 93.45% on the CBIS-DDSM dataset. Accordingly, the work by Al-Antari et al. [18] had faster inference time of 0.025 seconds per image on Mass detection for the INbreast dataset, but our results exceeded their detection accuracy rate of only 97.27% where we reported a detection accuracy rate of 98.1% on the CBIS-DDSM dataset [31]. Additionally, it is fair to mention that all experiments in the related works were conducted using different configurations and preprocessing techniques, which may show different performance on public datasets. We also compared the work of Zheng et al. [38] that similarly conducted the detection and classification tasks on the UCHCDM dataset. Although the surveyed work

Table 6
Comparison of Mass detection with other works.

Reference	Year	Method	Dataset	Detection accuracy rate (%)	Inference time per image (sec)
Al-masni et al. [28]	2017	YOLO	DDSM	85.25	NA
Dhungel et al. [22]	2017	Cascade Deep Learning and Random Forest	INbreast	96	39
Al-masni et al. [30]	2018	YOLO	DDSM	99.7	NA
Zheng et al. [38]	2018	Detection: 3 cascading detectors (Haar, LBP, and HOG) Classification: VGG-19	UCHCDM	92.8	0.62 0.88
Agarwal et al. [13]	2019	CNN patch classifier and mass probability map (MPM)	CBIS-DDSM INbreast	82 98	NA
Peng et al. [24]	2020	Faster R-CNN	CBIS-DDSM INbreast	93.45 95.54	0.134
Al-Antari et al. [18]	2020	YOLO	INbreast	97.27	0.025
Singh et al. [27]	2020	Single Shot Detector (SSD)	INbreast	97	NA
Aly, G. et al. [17]	2021	YOLO	INbreast	89.5	0.009
Lbachir et al. [52]	2021	Mean-shift, standard deviation filter and fast scanning algorithm	MIAS CBIS-DDSM	96.53 92	NA
Isfahani et al. [53]	2021	Growth regional method	MIAS	92	NA
Silalahi et al. [54]	2021	CNN (VGG + ResNet)	INbreast	90	NA
Baccouche et al. [31]	2021	YOLO-based Fusion Models	CBIS-DDSM INbreast Private	95.7 98.1 98	0.55 0.58 0.52
Proposed Methodology	2022	YOLO-based Fusion Models (Current mammograms)	UCHCDM	92.09	0.62

Table 7
Comparison of Early detection with other works.

Reference	Year	Method	Dataset	Class Label	Early Detection accuracy rate (%)	Inference time per image (sec)
Watanabe et al. [33]	2019	cmAssist – Custom deep learning networks	Private	Mass	27	NA
Loizidou et al. [36]	2019	Temporal subtraction	Custom	Calcification	20	NA
Proposed Methodology	2022	YOLO-based Fusion Models + Pix2Pix translation (Prior mammograms)	UCHCDM + Synthetic dataset	Mass Calcification Architectural Distortion Normal Overall	36 ± 0.01 14 ± 0.01 50 ± 0.02 90 ± 0.06 37 ± 0.10	0.63

achieved a better overall performance of 92.8% than our reported results, the tasks were not simultaneous and required separate inference time of 0.62 seconds per image for the detection method and 0.88 seconds per image on the classification.

Summing up, the comparative works in Table 6 have been conducted to either extract the location of breast mass lesions or to detect and then classify the abnormality of the breast lesion. We mainly compared works that were based on the YOLO model [28, 30, 18, 17], however we analyzed other similar deep learning models such as CNN [13, 54], Faster R-CNN [24], and SSD [27]. Besides, we included latest machine learning related works such as random forest [22], feature extraction [38], fast scanning algorithm [52], and growth regional method [53]. Our work was previously applied on two public datasets and it has currently demonstrated the same efficiency on the Current views of the private dataset UCHCDM.

Furthermore, we compared the effort of similar works on conducting an early detection of breast lesions against our paper's contribution. Table 7 shows two recent works that had the closest similarity on integrating Prior mammograms views to predict the location and type of abnormal lesions. It is clear that our work surpassed the work of Watanabe et al. [33] that was able to accurately detect and distinguish Mass lesions with an early detection accuracy rate of 27%. However, our proposed methodology had a lower early detection accuracy rate on Calcification lesions where they had 20% on a custom dataset that was generated using the temporal subtraction technique. Genuinely, all the reviewed works were assessed on private datasets and the reported results could be distinctive compared to our study's outcome. All comparable works did not measure the testing time but our proposed method achieved an inference time of only 0.63 seconds per image.

5. Discussion and conclusion

In this study, we have proposed using the YOLO architecture model to detect and classify suspicious lesions in mammograms. Following our recent work [31], we have shown the advantage of using a YOLO-based fusion model to correctly localize and identify three different types of lesions: Mass, Calcification, and Architectural Distortion. The proposed framework was furthermore developed to integrate the Prior mammograms from all used follow-up screenings and provide an early detection and classification on initial screened mammograms. The work emphasized the ability of a possible retrospective prediction on Prior mammograms that were diagnosed as Normal but at a later stage, they were reported with a clear presence and progress of abnormal findings.

Similar methodologies addressed the problem and used pairs of mammograms to enhance the CAD systems' results on Current mammograms by including temporal features for a regional registration [35], or adding temporal subtraction between pairs to an SVM classifier [36] or a CNN model [37]. However, our study employed one single model that was trained and tested on Current views, and next inferred on their corresponding Prior views. We have emphasized the performance of our proposed methodology by directly applying the saved YOLO-based fusion model differently on original and synthetic Prior mammograms that were generated using the image-to-image translation techniques. Two state-of-the-art models, CycleGAN and Pix2Pix, were trained and validated between the pairs of mammograms (Prior, Current) to create new translated Prior mammograms that can overcome the misalignment between the two screenings due to temporal and texture changes.

Performance results showed a high early detection accuracy rate of 36% ± 0.01 for Mass lesions, 14% ± 0.01 for Calcification lesions, and 50% ± 0.02 for Architectural Distortion lesions. In addition, 90% ± 0.06 of Normal mammograms were accordingly classified based on Prior exam screenings. Quantitative results were reported on two scenarios: both Current and Prior mammograms were correctly predicted, and only Prior mammograms were correctly predicted when their corresponding Current views were mispredicted.

The reported outcome forms a promising performance of the YOLO architecture model to capture missed lesions in former screening views that were clearly present in their latest screening views. Although the early detection and classification results for abnormal lesions were not significantly high (i.e. true prediction rate less than 60%), the percentage of the correctly predicted Prior mammograms is high enough to make a clinical impact for breast cancer.

Fig. 14 demonstrated that in particular our methodology could early detect locations of Mass lesions much earlier than the expert diagnoses using follow-up screenings within 3 to 6 years, which is relatively late for breast cancer diagnosis.

Limitations of this work may occur in preparing the right format of the training configuration for the YOLO-based model. Another challenge presented by a long training of image-to-image techniques due to the high number of hyperparameters. CycleGAN and Pix2Pix models were separately trained and evaluated aside to generate synthetic data and the training took on average 3 hours.

This paper provided an assessment of the YOLO-based fusion model for breast lesion detection and classification on mammography with a low error rate. Moreover, a new framework was presented for a retrospective early detection and classification of abnormality in mammograms and assist radiologists with assured diagnoses for each type of lesions.

As screening mammography has been considered an essential tool for breast cancer that has been acknowledged to lead to a significant reduction of mortality rate, CAD systems have tried to redress its outcome and lower the number of missed detection on screening. Therefore, the contribution of this paper could be utilized to screen Prior mammograms and detect those with the highest abnormal risk of breast cancer. Consequently, it will provide a warning signal for radiologists to forecast and anticipate the cancer progress.

Further investigation might be required to assess the future risk's region and analyze the signal's texture and surrounding contours, which facilitate understanding of the abnormality in order to develop a cost-effective clinical application.

Author contributions

A.B conceived the idea, developed and implemented the methods. B.G-Z and Y.Z helped with formulating and validating the experiments and analysis. Y.Z provided the data. B.G-Z, Y.Z and A.S.E supervised the project. A.B wrote the paper. All authors reviewed and edited the manuscript.

Declaration of Competing Interest

The authors have no conflict of interest to disclose.

Acknowledgments

The authors would especially like to express their gratitude to Dr. Clifford Yang and his team (University of Connecticut Health Center) for providing the UCHC DigiMammo (UCHCDM) database. A special thanks to Cristian Castillo and the all radiologists Dr. Kictzia Yigal Larios, Dr. Raquel Balbás at FUCAM A.C., Dr. Guillermo Peralta and Dr. Néstor Piña at Cancer Center Tec100 by MRC International for providing the Private dataset from the National Institute of Cancerology (INCAN).

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (3) (2021) 209–249.

[2] American Cancer Society: Cancer Facts and Figures 2020, American Cancer Society, Atlanta, Ga, 2020 Last accessed September 25, 2020.

[3] B.J. Svensson, E.S. Dylke, L.C. Ward, D.A. Black, S.L. Kilbreath, Screening for breast cancer–related lymphoedema: self-assessment of symptoms and signs, Support. Care Cancer 28 (7) (2020) 3073–3080.

[4] S.W. Duffy, L. Tabár, A.M.F. Yen, P.B. Dean, R.A. Smith, H. Jonsson, … T.H.H. Chen, Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women, Cancer 126 (13) (2020) 2971–2979.

[5] N Houssami, CI Lee, DSM Buist, D. Tao, Artificial intelligence for breast cancer screening: opportunity or hype? Breast 36 (2017) 31–33.

[6] D.B. Woodard, A.E. Gelfand, W.E. Barlow, J.G. Elmore, Performance assessment for radiologists interpreting screening mammography, Stat. Med. 26 (2007) 1532–1551.

[7] L.J. Warren Burhenne, S.A. Wood, C.J. D'Orsi, S.A. Feig, D.B. Kopans, K.F. O'Shaughnessy, … R.A Castellino, Potential contribution of computer-aided detection to the sensitivity of screening mammography, Radiology 215 (2) (2000) 554–562.

[8] P.C. Brennan, Z. Gandomkar, E.U. Ekpo, K. Tapia, P.D. Trieu, S.J. Lewis, … K.K. Evans, Radiologists can detect the 'gist'of breast cancer before any overt signs of cancer appear, Sci. Rep. 8 (1) (2018) 1–12.

[9] R. Benny, T.A. Anjit, P. Mythili, An overview of microwave imaging for breast tumor detection, Progress In Electromagnetics Research B 87 (2020) 61–91.

[10] D.Q. Zeebaree, H. Haron, A.M. Abdulazeez, D.A. Zebari, Trainable model based on new uniform LBP feature to identify the risk of the breast cancer, in: 2019 International Conference on Advanced Science and Engineering (ICOASE), IEEE, 2019, pp. 106–111.

[11] Wu N, Phang J, Park J, et al. Deep neural networks improve radiologists' performance in breast cancer screening [published online October 7, 2019]. IEEE Trans Med Imaging.

[12] Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis, Radiol Artif Intell 1 (4) (2019) e180096.

[13] R. Agarwal, O. Diaz, X. Lladó, M.H. Yap, R. Martí, Automatic mass detection in mammograms using deep convolutional neural networks, Journal of Medical Imaging 6 (3) (2019) 031409.

[14] L. Shen, L.R. Margolies, J.H. Rothstein, E. Fluder, R. McBride, W. Sieh, Deep learning to improve breast cancer detection on screening mammography, Sci. Rep. 9 (1) (2019) 1–12.

[15] T. Schaffter, D.S. Buist, C.I. Lee, Y. Nikulin, D. Ribli, … Y. Guan, DM DREAM Consortium, Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms, JAMA network open 3 (3) (2020) e200265-e200265.

[16] J. Latif, C. Xiao, A. Imran, S. Tu, Medical imaging using machine learning and deep learning algorithms: a review, in: 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, 2019, pp. 1–5.

[17] G.H. Aly, M. Marey, S.A. El-Sayed, M.F. Tolba, YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms, Comput. Methods Programs Biomed. 200 (2021) 105823.

[18] M.A. Al-Antari, S.M. Han, T.S. Kim, Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms, Comput. Methods Programs Biomed. 196 (2020) 105584.

[19] Mei-Sing Ong, Kenneth D Mandl, National expenditure for false positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year, Health a_airs 34 (4) (2015) 576–583 2015.

[20] S.W. Duffy, O.W. Morrish, P.C. Allgood, R. Black, M.G. Gillan, P. Willsher, … F.J. Gilbert, Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer, Eur. J. Cancer 88 (2018) 48–56.

[21] S.J.S. Gardezi, A. Elazab, B. Lei, T. Wang, Breast cancer detection and diagnosis using mammographic data: Systematic review, Journal of medical Internet research 21 (7) (2019) e14464.

[22] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (2017) 114–128.

[23] D. Ribli, A. Horváth, Z. Unger, P. Pollner, I. Csabai, Detecting and classifying lesions in mammograms with deep learning, Sci. Rep. 8 (1) (2018) 1–7.

[24] J. Peng, C. Bao, C. Hu, X. Wang, W. Jian, W. Liu, Automated mammographic mass detection using deformable convolution and multiscale features, Med. Biol. Eng. Comput. 58 (7) (2020) 1405–1417.

[25] Y. Li, L. Zhang, H. Chen, L. Cheng, Mass detection in mammograms by bilateral analysis using convolution neural network, Comput. Methods Programs Biomed. 195 (2020) 105518.

[26] V.K. Singh, H.A. Rashwan, S. Romani, F. Akram, N. Pandey, M.M.K. Sarker, … J. Torrents-Barrena, Breast tumor segmentation and shape classification in

mammograms using generative adversarial and convolutional neural network, Expert Syst. Appl. 139 (2020) 112855.

[27] S.Y. Siddiqui, I. Naseer, M.A. Khan, M.F. Mushtaq, R.A. Naqvi, D. Hussain, A. Haider, Intelligent Breast Cancer Prediction Empowered with Fusion and Deep Learning, CMC-COMPUTERS MATERIALS & CONTINUA 67 (1) (2021) 1033–1049.

[28] M.A. Al-masni, M.A. Al-antari, J.M. Park, G. Gi, T.Y. Kim, P. Rivera, … T.S. Kim, Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 1230–1233.

[29] G. Hamed, M. Marey, S.E. Amin, M.F. Tolba, Automated Breast Cancer Detection and Classification in Full Field Digital Mammograms Using Two Full and Cropped Detection Paths Approach, IEEE Access 9 (2021) 116898–116913.

[30] M.A. Al-Masni, M.A. Al-Antari, J.M. Park, G. Gi, T.Y. Kim, P. Rivera, … T.S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, Comput. Methods Programs Biomed. 157 (2018) 85–94.

[31] A. Baccouche, B. Garcia-Zapirain, C.C. Olea, A.S. Elmaghraby, Breast lesions detection and classification via yolo-based fusion models, Comput. Mater. Contin. 69 (2021) 1407–1425.

[32] A. Kumar, S. Mukherjee, A.K. Luhach, Deep learning with perspective modeling for early detection of malignancy in mammograms, Journal of Discrete Mathematical Sciences and Cryptography 22 (4) (2019) 627–643.

[33] A.T. Watanabe, V. Lim, H.X. Vu, R. Chim, E. Weise, J. Liu, … C.E. Comstock, Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography, J. Digit. Imaging 32 (4) (2019) 625–637.

[34] S. Timp, C. Varela, N. Karssemeijer, Temporal change analysis for characterization of mass lesions in mammography, IEEE Trans. Med. Imaging 26 (7) (2007) 945–953.

[35] S. Timp, N. Karssemeijer, Interval change analysis to improve computer aided detection in mammography, Med. Image Anal. 10 (1) (2006) 82–95.

[36] K. Loizidou, G. Skouroumouni, C. Nikolaou, C. Pitris, A new method for breast micro-calcification detection and characterization using digital temporal subtraction of mammogram pairs, in: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2019, pp. 1–4.

[37] K. Loizidou, G. Skouroumouni, C. Nikolaou, C. Pitris, An automated breast micro-calcification detection and classification technique using temporal subtraction of mammograms, IEEE Access 8 (2020) 52785–52795.

[38] Y. Zheng, C. Yang, A. Merkulov, Breast cancer screening using convolutional neural network and follow-up digital mammography, Computational Imaging III, 10669, International Society for Optics and Photonics, 2018.

[39] S. Kaji, S. Kida, Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging, Radiological physics and technology 12 (3) (2019) 235–248.

[40] T. Shen, C. Gou, J. Wang, F.Y. Wang, Collaborative Adversarial Networks for Joint Synthesis and Segmentation of X-ray Breast Mass Images, in: 2020 Chinese Automation Congress (CAC), IEEE, 2020, pp. 1743–1747.

[41] H. Liao, Z. Huo, W.J. Sehnert, S.K. Zhou, J. Luo, Adversarial sparse-view CBCT artifact reduction, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2018, pp. 154–162.

[42] G. Modanwal, A. Vellal, M. Buda, M.A. Mazurowski, MRI image harmonization using cycle-consistent generative adversarial network, Medical Imaging 2020: Computer-Aided Diagnosis, 11314, International Society for Optics and Photonics, 2020.

[43] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, A.S. Elmaghraby, Connected-UNets: a deep learning architecture for breast mass segmentation, NPJ Breast Cancer 7 (1) (2021) 1–12.

[44] M. Hammami, D. Friboulet, R. Kechichian, Cycle GAN-Based Data Augmentation for Multi-Organ Detection in CT Images Via Yolo, in: 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 390–393.

[45] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Med. Image Anal. 58 (2019) 101552.

[46] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[47] Y. Hiasa, et al., Cross-modality image synthesis from unpaired data using CycleGAN, in: International workshop on simulation and synthesis in medical imaging, Springer Cham, 2018, pp. 31–41.

[48] Y. Zheng, C. Yang, A. Merkulov, M. Bandari, Early breast cancer detection with digital mammograms using Haar-like features and AdaBoost algorithm, in: SPIE Proceedings, 9871, Sensing and Analysis Technologies for Biomedical and Cognitive Applications 2016, 2016 98710D.

[49] F. Samuelson, C. Abbey, Using relative statistics and approximate disease prevalence to compare screening tests, The International Journal of Biostatistics 12 (2) (2016) 104.

[50] P.L.A. Hernández, T.T. Estrada, A.L. Pizarro, M.L.D. Cisternas, C.S. Tapia, Calcificaciones mamarias: descripción y clasificación según la 5. a edición BI-RADS, Revista chilena de radiología 22 (2) (2016) 80–91.

[51] P.T. Rajendran, V. Krishnapillai, S. Tamanang, K.K. Chelliah, Comparison of image quality criteria between digital storage phosphor plate in mammography and full-field digital mammography in the detection of breast cancer, The Malaysian journal of medical sciences: MJMS 19 (1) (2012) 52.

[52] I.A. Lbachir, I. Daoudi, S. Tallal, Automatic computer-aided diagnosis system for mass detection and classification in mammography, in: Multimedia Tools and Applications, 80, 2021, pp. 9493–9525.

[53] Z.N. Isfahani, I. Jannat-Dastjerdi, F. Eskandari, S.J. Ghoushchi, Y. Pourasad, Presentation of novel hybrid algorithm for detection and classification of breast cancer using growth region method and probabilistic neural network, Computational Intelligence and Neuroscience 2021 (2021).

[54] A.R.J. Silalahi, Breast Cancer Lesion Detection and Classification in mammograms using Deep Neural, IOP Conference Series: Materials Science and Engineering, 1115, IOP Publishing, 2021.