Automated Breast Cancer Detection and Classification in Full Field Digital Mammograms Using Two Full and Cropped Detection Paths Approach

GHADA HAMED [ID], (Student Member, IEEE), **MOHAMMED MAREY,**
SAFAA ELSAYED AMIN, AND MOHAMED F. TOLBA [ID], (Senior Member, IEEE)
Department of Scientific Computing, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

Corresponding author: Ghada Hamed (ghadahamed@cis.asu.edu.eg)

ABSTRACT Breast cancer is one of the most severe diseases that threaten women's life results in increasing the death rate annually as confirmed by the World Health Organization. Breast cancer early detection is one of the main reasons behind reducing cancer severity. However, with the huge number of mammograms taken daily, the checking process conducted by radiologists becomes lengthy, tiring, and pruning to errors process. Hence, with the tremendous success achieved by utilizing CNNs in bioinformatics, the development of Computer-Aided Detection (CAD) systems has proved its necessity to solve the challenging cases for the biopsies missed by the ordinary checking leads to decreasing the false positive and negative rates. In this paper, we present a YOLOV4 based CAD system to localize lesions in full and cropped mammograms and then classify them to obtain their pathology type. The proposed method mainly consists of three phases that are applied on the full-field digital mammograms of the INbreast dataset. First, the mammograms are preprocessed to remove any extra artifacts and then cropped into small, overlapped slices. Second, masses are localized through two paths: the full mammograms and the cropped slices detection after configuring the YOLO-V4 model. Third, other feature extractors like ResNet, VGG, Inception, etc. are used to classify the localized lesions to compare their performance against YOLO. The proposed method proved using the experimental results the impact of utilizing YOLO-V4 as a detector with the 2-paths of detection of a full mammogram and the cropped slices in a trial to avoid any data loss by resizing the large-sized mammograms. Our system succeeds in detecting the masses' location with an overall accuracy of ≈98% which is more than the recently introduced breast cancer detection methods. Moreover, its ability to distinguish between benign and malignant tumors with an accuracy of ≈95%.

INDEX TERMS Breast cancer, breast mammograms, breast masses classification, lesions detection, You Only Look Once.

I. INTRODUCTION

Cancer is one of the most genetic and dangerous diseases that threaten people's life especially due to its late discovery. Cancer has more than 100 kinds and one of the most common types of cancer is breast cancer. According to World Health Organization (WHO), breast cancer is considered the most common and dangerous cancer type that spreads all over the world and annually leads to a larger death rate besides lung cancer. In the united states only, it is expected by WHO that in 2021 about 43,600 women will die from breast cancer.

Breast cancer is categorized under carcinomas which are formed by the body cells that cover most of the outside and inside surfaces of the body. Each cancer type is mainly represented in either benign or malignant tumors [1]. Hence, once the breast cancer appears in a specific part of the breast, it starts in invading the surrounding tissues nearby the suspected one very fast especially the malignant tumors. The benign tumors are sometimes greater than the malignant ones however they are not spread widely like the malignant ones

and once they are removed, they don't grow back. After many studies, it is confirmed that the noticeable increase in the death rate caused by breast cancer is due to the late detection through screening which results in almost all cases in bigger-sized tumors and consequently makes it difficult to be treated [2].

Also, with the huge number of mammograms screened daily, the process becomes hard, lengthy, complex and time consumable for radiologists and consequently in many cases, it pruned to errors [3]. Since radiologists till now miss between [10%-30%] of cancers either by taking decisions for some cases as benign despite their malignancy (false negative) or requiring additional screens due to their doubt about malignant tumors despite they are benign (false positive) [4], [5].

Here, comes the role of developing Computer-Aided Detection (CAD) systems to assist as a second reader with the radiologists' decisions to decrease both the false negatives or the false positives. The development and use of CAD systems are increased since 2001 and especially in the period from 2004 to 2008, which is increased by 91% either by private labs or hospitals [6]. Through the usage of CAD, it has been proved how it improves the process of cancer detection at earlier stages by decreasing the false positive and negative rates [7]–[10]. Not only this but also its use reduces the consumed time required by the radiologists to check the screened mammograms [11]. CADs are usually developed to localize the lesions existing in the screened breast mammogram [12], [13].

Mammography is considered one of the most common ways that can be applied to detect breast cancer in an earlier stage. The main reason behind making mammography the preferable choice for women is that it needs a very low dose of x-rays which makes it somehow safe unlike Magnetic Resonance Imaging (MRI) that needs a large dose of magnets and radio waves. Hence our proposed approach decides to work with mammographic datasets to detect the existing tumors.

Although many developments have been carried on to enhance the existing CADs for better detection accuracy, there are many challenges still acting as a barrier. These challenges are represented in either the developed CADs or the mammogram characteristics. The main problems in the already existing CADs are represented in the missing capability of applying them in the real-life systems used by hospitals or labs due to their long time for detection. Some CADs perform the detection in real-time but with lower accuracy. Hence, the good detection performance and the fast execution is somehow a trade-off problem. On the other hand, the screened mammograms are usually of large size and pixel depth which results in resizing them in almost published work to be ready for training a specific model. The resizing results in missing some of the important information that may exist in the screened mammograms and consequently it may affect the obtained accuracy [14].

Nowadays, the Convolutional Neural Network (CNN) has been proved its powerful capability to achieve a notice-able change in the obtained results compared with the other traditional methods [15]. The huge advances done on the deep learning models result in replacing the process of the handcrafted features with a deep learning-based model to automatically learn the essential required features that are more relevant in our case to the tumors we need to detect. Also, another important advantage for using deep learning models rather than the conventional ones is that their learning ability through more than one level of representation. So, the deep models can successfully learn different forms of data representations beginning from the raw data, to higher representation levels till reaching the highest level of representation to complete the learning process [16]–[18].

In this paper, we proposed a You Only Look Once (YOLO) based real-time CAD system using an introduced full mammogram cropping and merging algorithm that can localize masses in high accuracy and a very few seconds. The proposed system is composed of two paths. The first path is passing the full mammogram to our system after resizing it to localize masses. The second path is cropping the full mammogram into small slices without resizing the original mammogram and then detect any existing masses but in the sliced parts, then the sliced parts are finally merged to get any detected lesions missed from using the full mammogram in the first phase. The novelty in this work is represented in proposing a full detection and classification model based on YOLO-V4 to detect tumors in the breast using two approaches full and cropped detection paths then replace YOLO's classification layers with other features extractors to classify the localized lesions it obtains. By applying the proposed architecture because of the cropping idea, the shortage in the publicly available Full Field Digital Mammographic (FFDM) datasets is resolved. The proposed CAD system contains at its beginning an augmentation step different from the ordinary known augmentation methods that replace the small number of FFDMs with a large number of small cropped slices representing the full mammograms. Moreover, the proposed mammogram cropping idea succeeds in preserving the mammograms' high-resolution property by cropping each full mammogram into small slices to fit the YOLO model in a trial to avoid losing any data by resizing the full mammograms. The sliced overlapped parts result as well in catching some missed lesions not detected in the case of the full mammograms detection. Not only this but also the introduced cropping idea is represented as a third reader besides the proposed first layer full mammogram detection decision and the radiologists' decisions to confirm the localized lesions. To classify the detected tumors in the best manner, most of the feature extraction classifiers used by the most recent published work are applied versus YOLO to propose a comparative analysis on the performance of each in order to deduce the best classification model we can depend on it for obtaining the pathology type of the localized lesions.

The remaining of the paper is structured as follows: Section II presents some related work to the developed CADs responsible for breast cancer detection and classification.
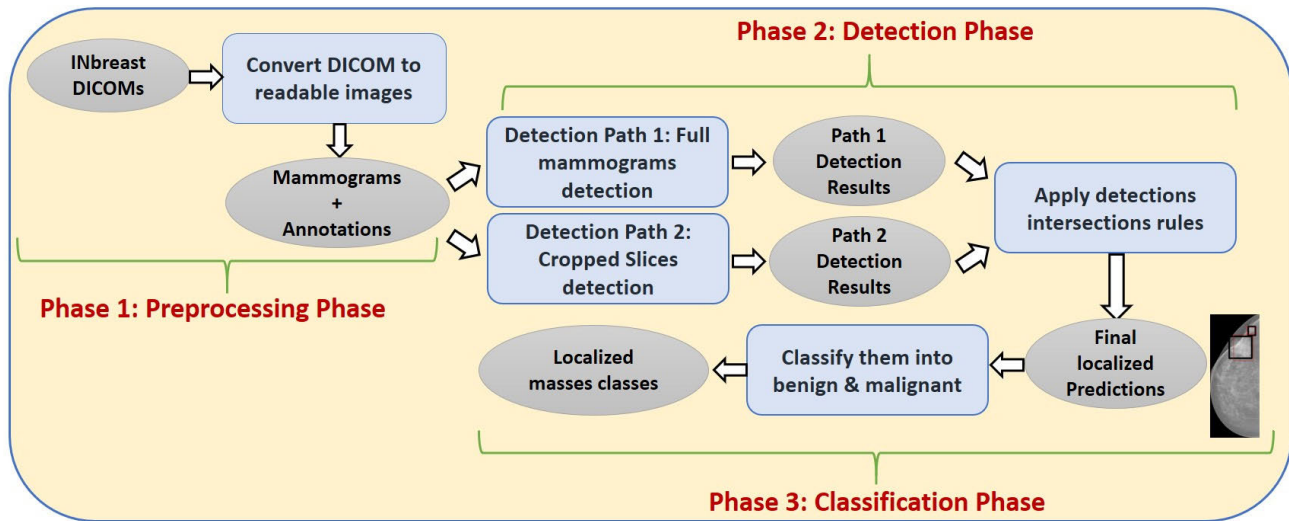
FIGURE 1. The overall architecture of YOLO-V4 based CAD to localize lesions and classify them.

Then, the proposed approach architecture is described in section III with introducing the details of the 2-paths detection of abnormal lesions in the breast and the localized masses classification phase. In section IV, all the conducted experiments are presented to show the proposed method performance and compare it versus other recent states of artwork in breast cancer localization and finally, the work conclusion is contained in section V.

II. RELATED WORK

Recently, obvious progress has been achieved in the deep learning especially in extensively merging the use of various image processing techniques with other biological concepts to resolve critical issues like breast cancer detection. Numerous valuable attempts have been developed to obtain a successful computer-aided detection model to localize the existing lesions in a fast and accurate manner [19], [20].

Doi in [21] has proved the critical need for the development of different computer-aided detection systems with their impact on the early diagnosis of breast cancer which results in being one of the major research fields in diagnostic radiology and medical imaging. Besides Jiang *et al.* [22] and Chan *et al.* [23] that confirmed how the CADs can improve the performance of the masses detection process held by radiologists that results in presenting the CAD as a second eye for the doctors to check for breast cancer.

Jiang integrated a new dataset of breast mammograms named Film Mammography dataset number 3 (BCDR-F03). Then, they worked on the segmented tumors existing in breast mammograms by applying the GoogLeNet model and the AlexNet to classify them achieving an Area Under Curve (AUC) of 0.88 for GoogLeNet and 0.83 for AlexNet respectively [24]. Lotter *et al.* proposed a scanning model based on a CNN patch-based classifier to classify mammograms achieving an AUC of 0.92 [25].

Dhungel *et al.* succeeded in proposing a full detection and classification model to first localize masses in the mammograms of the INbreast dataset then segment them to be classified obtaining a True Positive Rate (TPR) of 0.95 ± 0.02 [26]. Ribli *et al.* used the fast R-CNN by selecting some INbreast mammograms to localize lesions and distinguish between the benign and malignant ones to get False Positives per Image (FPI) of 0.30 and TPR of 0.90 [27].

Nan *et al.* used the residual network of 22 paths to classify the benign detected lesions or the malignant ones achieving AUC of 0.843 for the malignant masses and 0.696 for the benign biopsied populations [28]. Rodriguez-Ruiz *et al.* using the experimental results proved how the performance of the breast cancer detection using their proposed stand-alone Artificial Intelligence (AI) model is very near to the radiologists' evaluation. Their system obtained an AUC of 0.840 while the human (radiologists) evaluation reached an average AUC of 0.814 [29]. Moreover, other different schemes for proposed CADs in these new surveys [30], [31] proved the great impact of AI on breast cancer early detection.

III. PROPOSED METHOD

In this section, the proposed architecture phases to detect lesions in mammograms and classify them are discussed, with the evaluation method applied to check for cancerous tumors in new mammograms. As shown in Fig. 1, the proposed architecture is composed of three phases. The first phase is the preprocessing phase that prepares the Digital Imaging and Communications in Medicine (DICOM) mammograms to be in the required format without any extra artifacts. The second phase is the detection phase responsible for configuring YOLO-V4 to localize masses in the abnormal mammograms through two detection paths. Finally, the third phase which is the classification phase applied by replacing YOLO-V4 classification layers with other feature extractors to distinguish
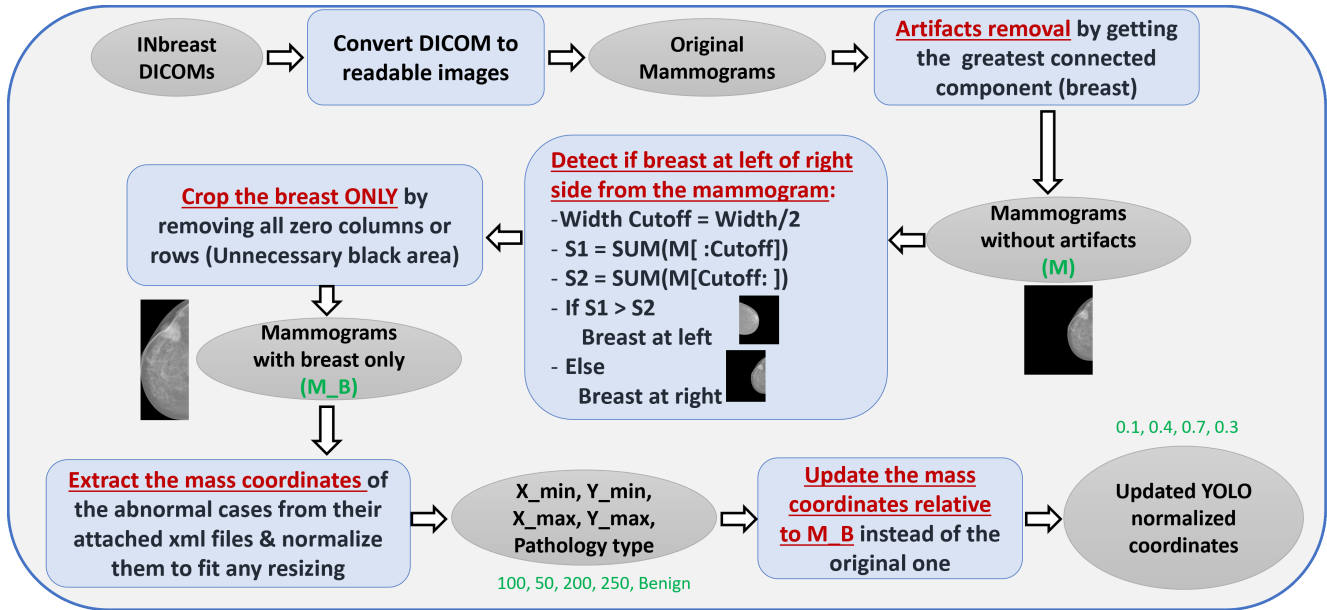
FIGURE 2. Phase 1: The mammograms preprocessing phase.

between benign or malignant detected lesions. Fig. 1 illustrates the overall architecture of the YOLO-V4 based detection model.

A. PHASE 1: PREPROCESSING PHASE

The mammograms are usually stored in form of Digital Imaging and Communications in Medicine (DICOM) format in large dimensions greater than those that fit in any deep learning method. Besides that, their ground truth annotations are not stored in a direct format. So, through this phase, the preprocessing steps shown in Fig. 2 are applied to prepare the mammographic datasets in the required format for masses checking.

The preprocessing step is considered an essential phase for both types of mammograms, either the scanned mammograms or FFDM. As clear from Fig. 3, the FFDM has better quality than the scanned mammogram since it is generated directly from the scanner in a digital format which results in reducing the noise quantity obtained due to the scanning process applied to convert the film mammogram to a scanned one.

So, the scanned mammograms are in a higher need of the preprocessing steps than FFDM. CADs obtain better results when they have applied on the full-field digital mammograms than the scanned film ones but generally more accurate results are obtained when the preprocessing steps are applied on either one of them. Since even in the case of the FFDM, sometimes labels exist at the top of some cases and even if the mammogram is free from any extra artifacts, some noise results in the obtained FFDM due to some dust obtained on the breast screening tool itself or any useless motion.

The first step as shown in Fig. 2 comes here to prepare the mammograms in the required format with their ground truth.
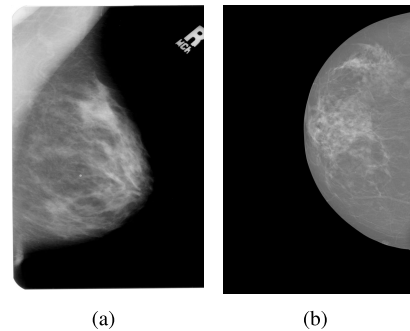

FIGURE 3. FFDM versus scanned mammograms: (a) A CBIS-DDSM sample mammogram which is scanned to be digitized, (b) An INbreast sample mammogram which is FFDM.

In the beginning, all DICOM files are converted into readable images in TIFF (Tagged Image File Format) and (Portable Graphics Format) PNG formats. Since the mammograms' pixels' values are usually fit between 14-bits and 16-bits contrast resolution, so in our work, the TIFF format is used to preserve the original values for pixels that in most datasets are greater than 256 colors. However, TIFF consumes a large space to store the mammograms, so the PNG is also used but by scaling the 14-bits or the 16-bits resolution into 8-bits resolution to compare between its manner versus using TIFF mammograms. Then, as shown in Fig. 4 in (a) any artifacts like the extra labels indicate the date, mammograms view, etc. are removed by extracting the greatest component (breast) to get a noise-free mammogram like the one shown in subfigure (b).

Moreover, in order to reduce any useless space in the mammogram leaving the breast only, some mammograms
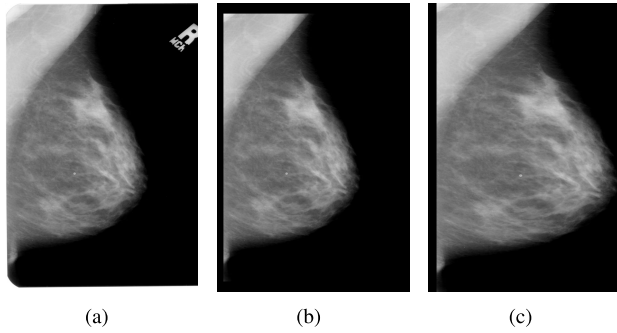
FIGURE 4. Preprocessing output: (a) A sample original mammogram (M), (b) M after removing any extra artifacts except the breast (M_Without_Labels), (c) M_Without_Labels after removing any continous black space (M_Without_Labels_Black_Cropped).

contain large black space as shown in Fig. 4 in (c) that is output from (b). So, in a trial to resize the mammogram naturally without losing any data, continuous black spaces in a mammogram are removed to reduce the original size to fit the breast only, and consequently, the coordinates of the lesion are updated related to the new mammogram. All coordinates are extracted from the XML files attached with each case and then normalized to [0, 1] to fit any size the original mammogram will be resized to. The full preprocessing steps are shown in Fig. 2.

B. PHASE 2: MASSES DETECTION PHASE

It is concluded from the most recent work proposed for breast cancer detection that the full mammograms due to their large sizes, they are resized to small dimensions to fit the one of the deep learning voracious method [32]–[37]. The resizing issue is mainly summarized in a reduction in the image quality and the loss of some significant information that may feature the existent masses especially when the resizing occurred by a large ratio. For the sake of this, in this paper two paths are introduced to detect tumors which are the full mammogram detection and the cropped slices detection. YOLO has 4 versions and its latest version (YOLO-V4) is one used here for learning how to localize tumors in mammograms. The 3 versions are compared before on the INbreast dataset in [37] and based on the conducted comparative analysis, it is found that YOLO-V4 will perform better as will be proved in the experimental results. So, YOLO-V4 has been configured to be applied after adopting both detection paths as follows:

1) PATH 1 OF FULL MAMMOGRAMS DETECTION (FMD-PATH): MASSES LOCALIZATION IN FULL MAMMOGRAMS

As shown in Fig. 5, two strategies are applied to be compared and then obtain the better one to continue the localization process.

In the first strategy, the mammograms are labeled as benign, malignant, and normal mammograms, i.e. 2 classes (benign & malignant) besides the negative samples of the normal mammograms. While in the second strategy, the mammograms are labeled as normal and abnormal, i.e. 1 class (mass
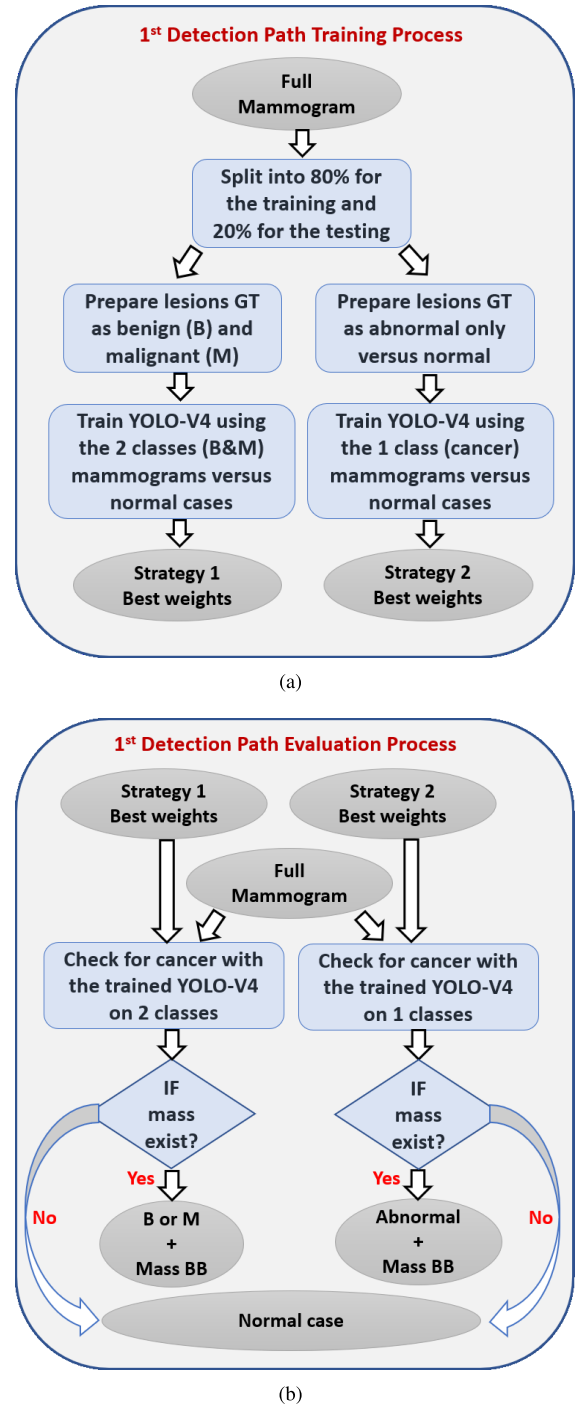

FIGURE 5. Phase 2: The first path of the lesions localization in mammograms: (a) 1st Detection Path Training Process; (b) 1st Detection Path Evaluation Process.

whatever its pathology type) besides the normal ones. In both strategies, the dataset is divided into 80% for the training and 20% for the testing.

2) PATH 2 OF CROPPED MAMMOGRAMS DETECTION (CMD-PATH): MASSES LOCALIZATION IN CROPPED SLICED MAMMOGRAMS

The idea behind this path is to make further analysis on a given case to get any missed masses not obtained by detection
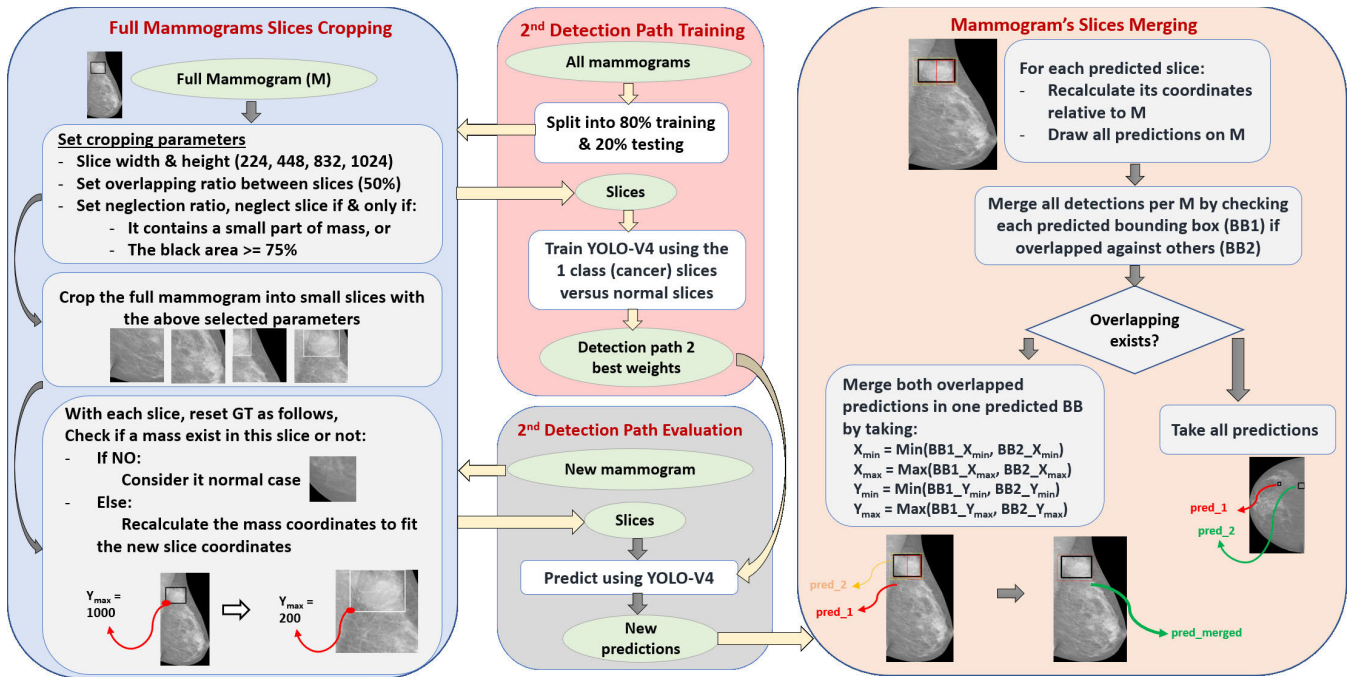
FIGURE 6. Phase 2: The second path of the lesions localization in mammograms.

path 1. The second detection path includes three main steps. First, each full large-sized mammogram is cropped into small slices. Second, these small slices are prepared to input to YOLO-V4 for training to be able to check for any existing lesions in newly taken mammograms and also to be used for the model detection evaluation. The detection process is performed on these slices to act with them as new small mammograms contain lesions (abnormal) or not (normal). Then, the third and final main step is post-processing each mammogram's slices by merging them together to get the existent masses bounding boxes on the concatenated full mammogram. In this path, the mass is either obtained in the abnormal cases without knowing its malignancy type or nothing detected in the case of the normal mammograms. The full cropping algorithm is detailed in Fig. 6.

This phase contains at the beginning a different kind of data augmentation that is represented in cropping each full mammogram into several small slices with a specific overlapping ratio. The idea behind the cropping step is not only to replace the full mammogram resizing with a small focused slice but also overcoming the small size of the existing FFDM samples by creating new samples from slicing each mammogram which results in the existence of larger training and testing datasets using unconventional data augmentation method.

3) DETECTION INTERSECTION RULES OF FMD-PATH AND CMD-PATH

Both introduced detection paths are utilized here with each other to be able to check any existing lesions in the sample evaluated mammogram. The commonly used approach to

localize lesions in a mammogram is resizing it to a smaller size, then apply on it a specific algorithm or model to get masses in abnormal cases. In this paper, this approach is applied through the FMD-Path but due to the missed lesions in some cases, the CMD-Path is proposed which enhances the detection performance since it succeeds in getting most of the lesions detected by the FMD-Path in addition to others not obtained by the FMD-Path. Moreover, usually the CADs till now act as a second reader besides the radiologists' decisions so using the proposed method, we introduced a third supporter to the final decision which represented in the CMD-Path proved using the experimental results its ability to act as a third reader to a specific mammographic case. Based on the obtained results, the FMD-Path is usually the main reference to our decision since it looks for the mammogram as one shot with all the properties and features included in it. Then, the necessity for applying the CMD-Path appears in the case of the dense mammograms or the mammograms containing tumors but not obvious since their features and color intensities are similar to the breast. So, results obtained from both paths of detection are combined as follows to get suspected areas in mammograms:

1) FMD-Path: The full mammogram is checked as an input to our CAD to get "FMD-Path Detections Results" using the best weights obtained from experiment 2.
2) CMD-Path: The full mammogram is first preprocessed to be cropped into slices of size 1024×1024, then the mammogram's slices are prepared to input the CAD system for lesions localization using the best weights
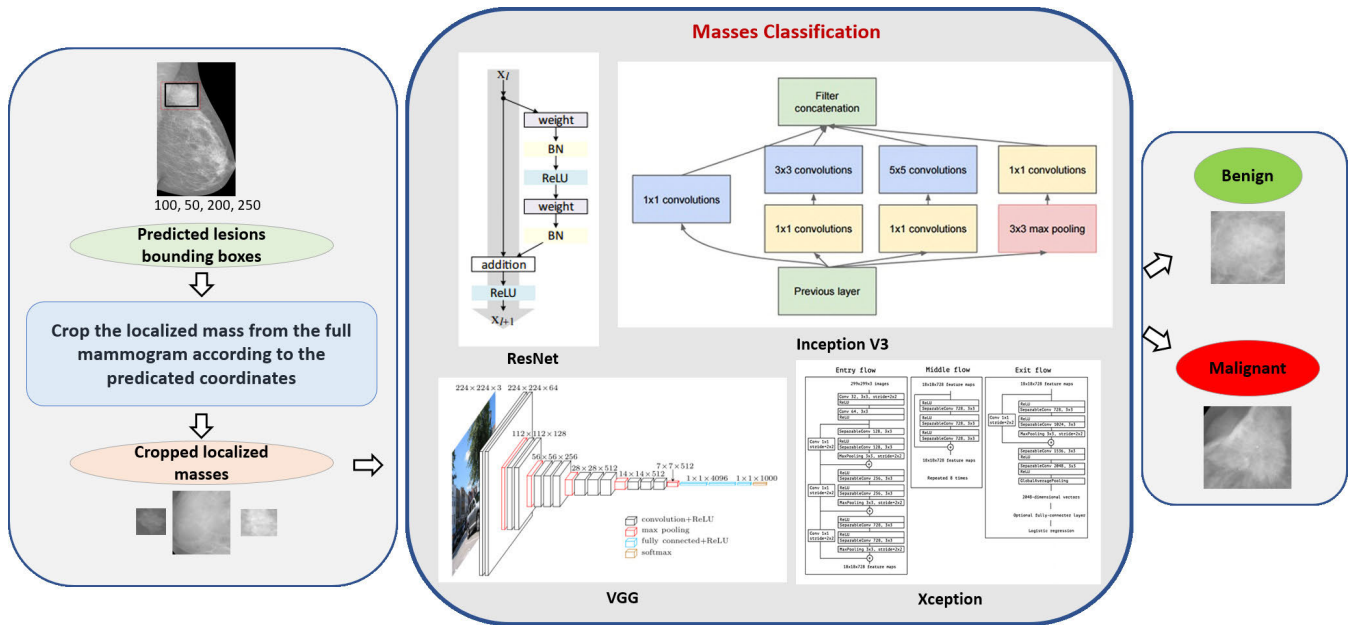
FIGURE 7. Phase 3: The localized masses classification using different features extractors other than YOLO.

obtained from experiment 3. Then, the output slices are post-processed to be merged to get the full mammogram with the full detected masses if any ("CMD-Path Detections Results").

3) For the results obtained from both paths, one or more of the following rules will be applied as follows:

- Rule 1: Get any intersected localized lesions between FMD-Path and CMD-Path results to act as a confirmed cancerous case, i.e. "Confirmed localized masses" = "FMD-Path Detections Results" ∩ "CMD-Path Detections Results".
- Rule 2: If no results are obtained from path 2 (CMD-Path) but there are "FMD-Path Detections Results", then take it.
- Rule 3: If no results are obtained from path 1 (FMD-Path) but there are "CMD-Path Detections Results", then the confidence score (CS) of the obtained masses will be checked to take only those that their CF ≥ 0.5.

C. PHASE 3: MASSES CLASSIFICATION PHASE

Since both the second strategy of the first detection path (FMD-Path) or the second detection path (CMD-Path), the output in case of the abnormal mammograms is the bounding box of the tumor without differentiating if it is benign or malignant. So, to be able to propose a complete detection and classification model and also to be able to compare the proposed work performance with others, the obtained tumors must be classified into benign or malignant. So, this classification phase shown in Fig. 7 comes to continue classifying the detected masses obtained by applying either strategy 2 of the first detection path or the second one. In this phase,

the vector of coordinates representing the bounding box of the existent masses is used to crop the mass area into separate images. Then, YOLO-V4 role of classifying objects after localizing them is replaced by other feature extractors. Multiple classifiers are used to distinguish between the benign and malignant masses to conclude the best one that can achieve the best classification result.

IV. EXPERIMENTAL RESULTS
A. EVALUATION METRICS
Since the proposed model is mainly composed of two phases: masses localization followed by the classification of the detected masses. There are some evaluation metrics used to evaluate the accuracy of the lesions detection and others used to evaluate the masses classification performance.

1) DETECTION EVALUATION METRICS
We used the Intersection Over Union (IOU) metric to evaluate how much the model is accurate in localizing tumors in breast mammograms as follows:

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \qquad (1)$$

In our work, we consider the mass detected correctly if the IOU is greater than or equal to 0.5 and otherwise the detection is neglected.

2) CLASSIFICATION EVALUATION METRICS
Each localized mass is passed through a specific classifier to output its malignancy type then it is compared against the given ground truth to get one of four options. The first is the benign mass that is classified as benign (True Negative), the second is the benign mass classified as malignant
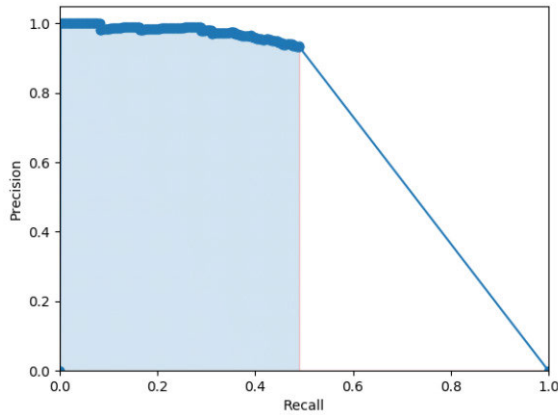
FIGURE 8. mAP example.

FIGURE 9. Examples from the most widely used public datasets of breast mammograms. (a) MIAS Mammogram Example, its ID is MIAS_SPIC_mdb202rl_MLO; (b) DDSM Mammogram Example, its ID is A_1389_1.LEFT_CC; (c) CBISDDSM Mammogram Example, its ID is Test_P_00194_RIGHT_CC; (d) INbreast Mammogram Example, its ID is 50997515.

(False Positive), the third is the malignant mass classified as malignant (True Positive) and finally, the malignant mass classified as benign (False Negative). These four metrics are the main components of the confusion matrix that is one of the significant metrics used to evaluate the classification accuracy. We measured True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) and used them to calculate the following:

- Precision:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- Overall accuracy

$$Overall\ accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (4)$$

- Mean Average Precision (mAP) The average precision (AP) is the calculated area under the curve of the recall with the precision shown in Fig. 8 The mAP is the mean of the AP that is computed for the benign and the malignant classes.

B. DATASETS
We used in our experiments the INbreast [38] that is considered one of the few benchmark datasets that contain FFDMs and with highly accurate ground truth for the attached lesions with each case. INbreast is considered a public dataset that can be requested from [38]. It contains 410 FFDM mammograms collected from 115 patients. The INbreast contains both Mediolateral Oblique (MLO) and Craniocaudal (CC) views for normal and abnormal cases. Each mammogram exists in DICOM format with an XML file that contains some necessary information about the case and the ground truth annotations of any lesions existing in case of abnormal mammograms. All mammogram pixels' values are of 14-bits contrast resolution and have two dimensions which are either
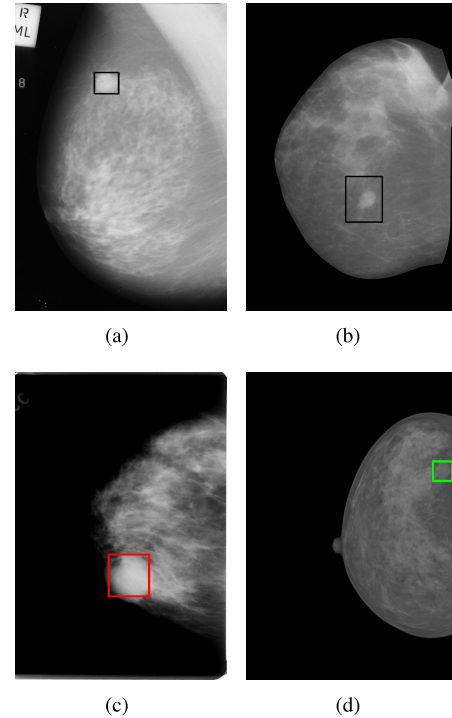
3328×4084 or 2560×3328. The full mammograms assigned to the testing set are 22 normal and 22 abnormal mammograms which represented 20 % of the overall samples. The number of lesions included in the abnormal cases is 24 since some cases contain more than one mass.

The INbreast is considered the only public dataset that contains fully field digital mammograms. Also, it is characterized by the accurate lesions annotations that represent the masses' ground truth since they are validated and confirmed by two specialists. There are other public mammographic datasets and the most popular and commonly used ones are the Mammographic Image Analysis Society Digital Mammogram Database (MIAS), the Digital Database for Screening Mammography (DDSM) dataset, and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM).

The MIAS is the oldest dataset containing digitized mammograms which have lower quality due to the scanning process and the digitizing tools used to prepare them as obvious in subfigure (a) in Fig. 9.

The DDSM dataset has a large set of mammograms, however as noticed from the given ground truth and confirmed in [40], the lesions annotations are not accurate as shown in Fig. 9 (b). Even if the DDSM is used in the training process, avoid using it in the detection and the classification evaluation process due to its low precision. The CBIS-DDSM is a modified and updated version from the DDSM which is

considered nowadays the most recent mammographic dataset that contains correct ground truth annotations for scanned mammograms as the sample given in subfigure (c) in Fig. 9. Fig. 9 shows a sample from each dataset and as shown the FFDM mammogram of the INbreast in the last subfigure (d) is the one that owns the best quality and the most accurate given ground truth that fits the actual existent mass as a result of the absence of any scanning or digitizing process.

Since the CBIS-DDSM is considered the most accurate dataset contains scanned mammograms, we conducted some of the experiments to compare the performance of the introduced CAD system on it versus the INbreast FFDM mammograms. For example, when the FMD-Path using strategy 2 is applied on the CBIS-DDSM scanned mammograms, it achieves mAP of 64.67% which is 97.86% for the FFDM mammograms using the same settings and parameters. After interpreting the failed cases, we found that there are a lot of FP cancerous cases. This is in most cases due to the quality of the mammograms reduced due to digitizing the scanned version compared with the FFDM as shown by the given examples in Fig. 9. Also, in most FP scanned mammograms, the pixels' values are concentrated at very large values which results in false localized lesions. These are the main reasons which direct us to complete our experiments on FFDMs besides that there is another essential reason represented in the labs and the hospitals' dependency nowadays on diagnosing FFDM instead of scanned ones thanks to the great development that occurred in the existing medical equipment.

C. IMPLEMENTATION DETAILS

We have implemented the masses classification phase by configuring some of the popular and commonly used classification models using Keras that is based on Tensorflow as a backend. To preprocess the INbreast and extract the ground truth annotations or to post-process the mammogram slices to get the full detections, we used Matlab and Python 3.7 as our development environments. Besides using C++ programming languages to update and compile YOLO on Ubuntu 14.04 operating system. The proposed YOLO-V4 based CAD system is implemented on an Intel Core (TM) i7-9700K desktop processor 8 cores up to 4.9 GHz Turbo 300 series with 16 GB RAM and GIGABYTE GeForce RTX 2080 Ti overclocked 11G graphics card.

YOLO-V4 has some essential parameters that shall be changed based on the problem and the used dataset type. These parameters are set by specific values based on several experiments that were carried on with various values in a try to interpret them and how they affect the training process and the evaluated detected performance. Based on this, there are a set of parameters that will have fixed values in all experiments since it is proved its significance on the lesions detection and others have more than one value. The parameters with fixed values are:

- Training iterations number = 4000. Since it is mentioned in [41] and confirmed by results as well, that 2000 iter-

ations are sufficient to train only one class. So, since the maximum number of classes we have is 2, the number of iterations is set by 4000.
- Learning Rate (LR): Different values are used for learning rate as 0.001 & 0.0001 and by the experimental results it is found that starting with 0.001 is better than 0.0001.

The parameters that vary based on each experiment are:

- The number of classes (C): It is either two, i.e. benign & malignant versus normal mammograms, or one, i.e. abnormal versus normal mammograms.
- The Model input size: The network size can be set any value multiple of 32 as 448×448 or 608×608 or 832×832, etc. As the input mammogram size increased, the training process becomes better and the detection accuracy is enhanced. However, as the input mammogram size increased, the subdivisions can be occurred in a batch increased and consequently the training process becomes slower. So, in the main experiments, we usually resized the mammograms to 448×448 to perform fast experiments and then the experiment that obtained the best results was updated to resize its mammograms to larger dimensions.
- Steps: It is specified to change the LR value during the training process automatically after a specific number of iterations to avoid overfitting. For example, if it is set by 4000, 5000, this means that the first time for the LR change to be changed is at step number 4000 and then changed to another value at iteration 5000. The large LR allows the model to learn the general features of the localized masses in the training set in a faster manner. However, as the neural network sees a large amount of data, the LR needs to be decreased over time to result in changing the weights less aggressively.
- Scales used for learning rate change as follows:0.1 and then 0.1. This means that at iteration 4000 the LR is multiplied by 0.1 and then by 0.1 at iteration 5000.

D. EXPERIMENTS

1) EXPERIMENT 1: FMD-PATH USING STRATEGY 1
In this experiment, all mammograms were resized to smaller sizes to fit YOLO. Two trials are done in this experiment, the first full detection trial (F1) without including normal mammograms in the training set and the evaluation test. While the second full detection trial (F2) includes normal mammograms, i.e. only abnormal cases with benign and malignant lesions to obtain the results in Table. YOLO-V4 is configured using these values:

- The number of classes (C): 2.
- The network input size: 448×448.
- Steps: 3200 & 3600.
- Scales: 0.1 & 0.1.

As shown in Table. 1, the trial that doesn't include normal cases achieves better results by nearly 3%. The reason behind this is that the existence of normal mammograms in F2 makes

TABLE 1. The mass detection & classification results of FMD-Path using strategy 1 where F in the Trial # is donated for the FULL mammograms detection trial, B is donated for the benign class and M is donated for malignant class and Prec. is donated for the precision metric.

Trial #	B:TP/FP	M:TP/FP	B AP (%)	M AP (%)	Missed Masses (FN)	Detected Normal (FP)	Prec. (%)	Recall (%)	mAP (%) @IOU= 0.5
F1	4/0	15/2	77.29	93.75	4	-	90	79	85.52
F2	7/1	11/5	90.73	75.20	5	2	75	75	82.96

TABLE 2. The mass detection results of FMD-Path using strategy 2 where S is donated for the scales option.

Trial #	Scales	Mass: TP/FP	Missed Masses (FN)	Detected Normal (FP)	Prec. (%)	Recall (%)	mAP (%) @IOU= 0.5
F3	S1	17/4	3	0	81	71	84.89
F4	S2	22/0	2	0	100	92	97.86

the model not be able to distinguish between the normal cases and abnormal cases, which results in missing biopsies more than those missed in F1, i.e. treating some abnormal cases as normal cases due to the similarity between some benign cases and the normal ones.

2) EXPERIMENT 2: FMD-PATH USING STRATEGY 2

The goal of this experiment is to evaluate the proposed YOLO-V4 detection model but when all lesions are grouped into only class (abnormal case) versus the normal (mass-free) cases. So, all masses whatever their types are annotated as abnormal cases. The training and the testing set become balanced here by adding the same number of normal mammograms with the abnormal ones. In all trials YOLO-V4 is configured using these values:

- The number of classes (C): 1 (Mass) with negative samples representing the normal cases.
- The network input size: 448 × 448.
- Steps: S1 option (1500,2000,2500,3000) - S2 option (3200,3600).
- Scales: SC1 option (0.1,0.1,10.0,0.1) which is used with S1 steps - SC2 option (0.1,0.1) which is used with S2 steps.

As shown in Table 2, the full detection trial 4 (F4) is much better than full detection trial 3 (F3) and this is due to the way we changed the learning rate. Since the model is pre-trained before on images with general categories so, it doesn't have any knowledge about breast cancer and its features. For the sake of this, the learning rate shouldn't be decremented or changed at the beginning steps of the training process since we are starting with zero information and so the learning rate needs to be high then decremented at further steps as achieved by steps of S2 with scales of SC2.

On the other hand, when we compared the trials (F1 & F2) executed to detect lesions using strategy 1 with the trials (F3 & F4), we found that specializing the detector to learn how to localize cancer in the breast mammogram whatever its pathology type is better than learning how to localize benign and malignant lesions by nearly ĩ2%. So, by applying strategy 2 to train YOLO-V4 by its CSPDarknet53 backbone, 97.8% of the lesions in the testing set are detected with IOU greater than or equal to 0.5.

The best results obtained from all the conducted trials is from trial F4 which is using full mammograms and two values for the LR scale at late steps such that 22 lesions out of 24 in 22 mammograms are localized completely correct with mAP of 97.8%.

3) EXPERIMENT 3: CMD-PATH

The objective behind this experiment is to evaluate the mammogram cropping idea on the masses detection performance in the full mammograms. In this experiment, the slices with masses are treated as abnormal cases but act like zooming area in the mammogram and the slices without masses are treated as abnormal cases. However, since the mass region is considered very small relative to the whole mammogram, so the slices obtained with masses (abnormal area) are very small relative to those without masses (normal area). As shown in Table 3, this phase plays an important role in supporting the proposed method by a different kind of data augmentation to increase the used samples in either the training or the testing samples by slicing each mammogram into a different number of overlapped slices. Moreover, to reserve data balancing during training, we select random slices without masses and slices of the normal mammograms with the same number of the slices with masses, since the number of the slices with masses is the smallest one. For example, in the case of 224 × 224, we selected randomly 1840 samples from the set of slices without masses generated from the abnormal cases and 1840 samples from the cropped slices obtained from the normal cases besides mainly the 1840 slices with masses. Regarding the testing set, we should not know which slice contains mass in the real scenarios, so all slices shall be checked and hence in the evaluation phase, we took all the mammograms' cropped slices either from the normal or the abnormal mammograms.

In this experiment, the same 80% and 20% mammograms, representing the training set and the testing set respectively, are cropped into small slices using different values for the following parameters:

- The number of classes (C): 1 (Mass) with negative samples representing the normal cases.
- Original full mammograms dimensions: 3328 × 4084 or 2560 × 3328.
- Cropped slice dimensions: 224 × 224, 448 × 448, 832 × 832, 1024 × 1024.
- Overlapping ratio: The slices are overlapped by 0.5 in both width and height.
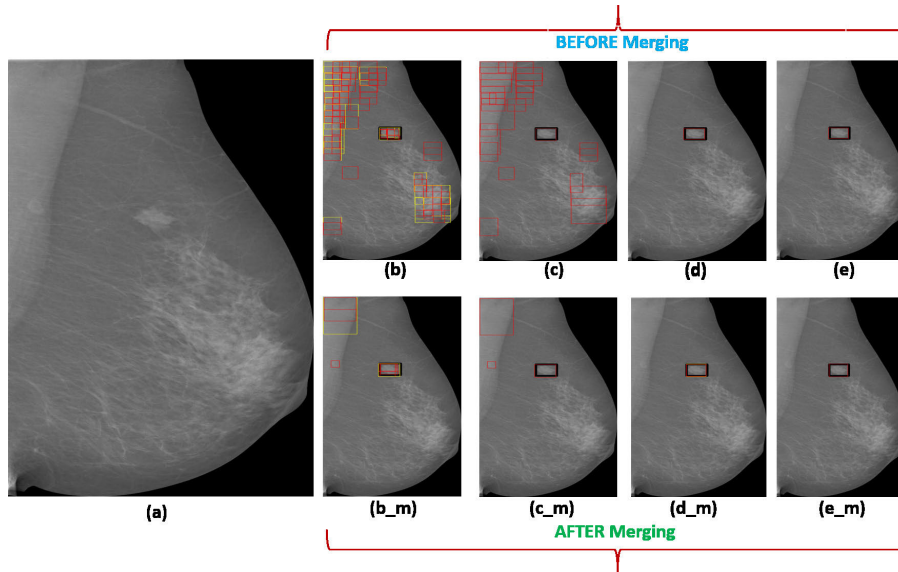
FIGURE 10. The detection output using different slice size, (a) shows the original mammogram of ID 22427751, (b-e) show the detection output using slice size of 224, 448, 832, 1024, respectively, (b_m - e_m) show the detection output using slice size of 224, 448, 832, 1024, respectively but after the postprocessing merging step.

TABLE 3. The effect of the full mammograms cropping step on augmenting the existing original samples.

Data	Training Samples				Testing Samples			
Slice size	Slices with masses	Slices without masses	Normal slices	Selected	Slices with masses	Slices without masses	Normal slices	Selected
224×224	1840	22,110	22,948	1840	481	5952	8487	ALL
448×448	645	5223	5798	645	157	1425	2189	ALL
832×832	315	1308	1801	315	73	377	674	ALL
1024×1024	214	726	1211	214	56	211	434	ALL

TABLE 4. The mass detection results of **CMD-Path** where C in the Trial # is donated for the CROPPED slices detection trial.

Trials		Abnormal				Normal		mAP (%) @IOU=0.5	
C#	Slice size	TP	FP	TP + FP	FN	TN	FP	Slices	Full
C1	224×224	1	0	21	0	2	20	43.36%	11.15%
C2	448×448	13	2	6	1	9	13	64.63%	54.01%
C3	832×832	10	5	5	2	13	9	78.76%	69.53%
C4	1024×1024	17	0	3	2	18	4	88.99%	89.13%

- Neglection ratio: Any slice is neglected from the training and the evaluation sets if either the mass is not fully included in the slice or the ratio of the non-zero pixels to the overall slice area is smaller than 25%.
- Steps: 3200 & 3600.
- Scales: 0.1 & 0.1.

As proved by the results in Table 4 and shown in the given example in Fig. 10, when the slice size increases the normal cases detected as positive (normal FP) decreases. Also, it is observed that the number of the abnormal cases detected true increased with the bigger slices' sizes.

The interpretation of these obtained results is that as the slice size increases, the almost bounding box covered the mass becomes appeared in the cropped slice, i.e. completely

included. So, the model has trained more accurately by using nearly complete features of the full shape of mass, color, size, etc., and consequently, the TP of the abnormal cases increases, and the number of normal cases detected as abnormal decreases as well. While on the other hand when the slice size decreases, the number of the normal regions detected as positive (FP) increases because the mass in most cases is divided on more than 1 slice due to the small-sized slices. Consequently, it is trained on the colors only more than the full shape and characteristics of the mass. Hence with the small-sized slices, the FP increases since any bright region will be predicted as mass whether it is actually a tumor region or not.

4) EXPERIMENT 4: FMD-PATH USING STRATEGY 2 SUPPORTED WITH CMD-PATH

In this experiment, both paths of detection evaluated in experiments 2 and 3 are combined here using the specified
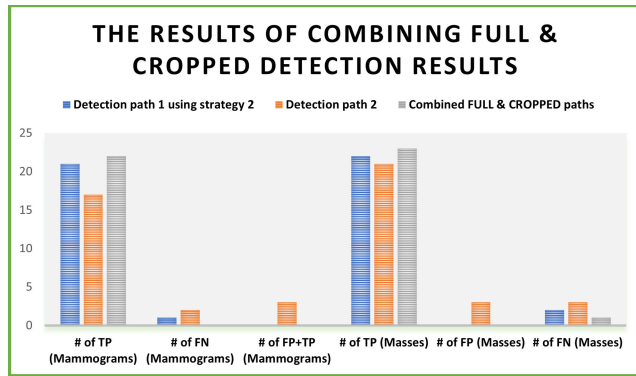
FIGURE 11. The results of combining FMD-Path using strategy 2 with CMD-Path versus each individual path.

TABLE 5. The mass classification results of the localized lesions where V.A. and T.A. are donated for the validation and the testing accuracy respectively and Prec. is donated for precision metric.

Classifier	TN	FP	TP	FN	V.A.	T.A.	Prec.	Recall	F1 Score
YOLO-V4	4	2	15	0	0.90	0.90	0.88	1.0	0.94
Inception V3	6	2	14	0	0.95	**0.91**	0.88	1.0	0.94
VGG16	3	5	14	0	0.89	0.77	0.74	1.0	0.85
ResNet50	0	8	14	0	0.74	0.64	0.64	1.0	0.78
Xception	2	6	14	0	0.89	0.73	0.7	1.0	0.82
Inception ResNetV2	3	5	14	0	0.89	0.77	0.74	1.0	0.85

FIGURE 12. The detection output of some test samples with their detected bounding boxes and the confidence probability (CP) of each localized lesion, (a-d) show the ground truth boxes of the existent masses, (e-h) show detection results of the first detection path that is working on full mammograms using strategy 2, (i-l) show detection results of the second detection path that is working on cropped slices to avoid resizing, (m-p) show detection results of both paths after applying the specified detection rules.

detection rules set to check any existing lesions in the same evaluated mammogram. As shown in Fig. 11, the proposed idea of applying both paths together on the same cases using the combination detection rules, associates in obtaining a higher number of true abnormal mammograms (TP mammograms) and a lower number of both missed abnormal cases (FN mammograms) and false abnormal regions as well (FP mammograms). The output of some test cases is shown in Fig. 12.

5) EXPERIMENT 5: MASSES CLASSIFICATION PHASE
This is the experiment of replacing YOLO features extractor layers with other different feature extractors to distinguish between either the benign or the malignant masses. Each pretrained classifier is fine-tuned on the masses of the training mammograms and then evaluated on the 22 localized lesions extracted from experiment 4 as shown in Table 5.

Since in the case of YOLO-V4, all the validation set must be first detected then classified. So, due to the small number of the existing samples, we used the testing set as the validation set to train better using larger samples. While for
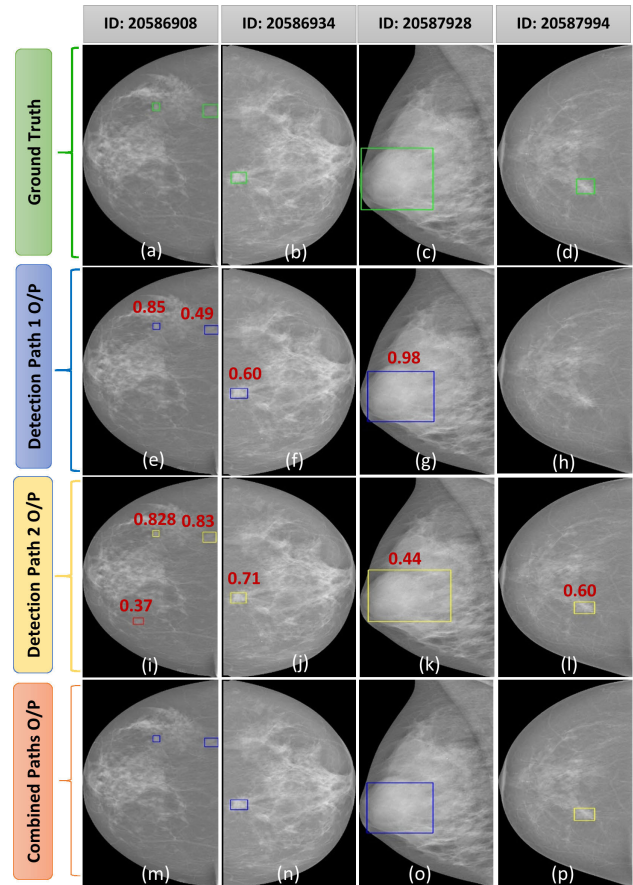
the remaining classifiers, we are able to divide them into training and validation sets. Since the training set is already augmented in the detection phase using the cropping idea based on the two-layer detection approach. Then, all masses are cropped from the training set used in the detection, which helped us in dividing them into 60% and 20% for the training and validation sets respectively, leaving the testing set as it is to introduce a fair comparison between the used classifiers and those obtained from YOLO-V4.

Based on experiment 1 of the FMD-Path using strategy 1 in which we merged the detection and the classification tasks together, a larger number of FNs is obtained than the trial based on two separate detection and classification phases of experiment 2 for the FMD-Path using strategy 2. Where, experiment 1 of the FMD-Path using strategy 1 achieves 85.5% for the detection accuracy and 90.0% for classifying the already truly detection lesions using merged detection and classification phases. While experiment 2 of the FMD-Path using strategy 2 achieves 97.8% for the detection accuracy
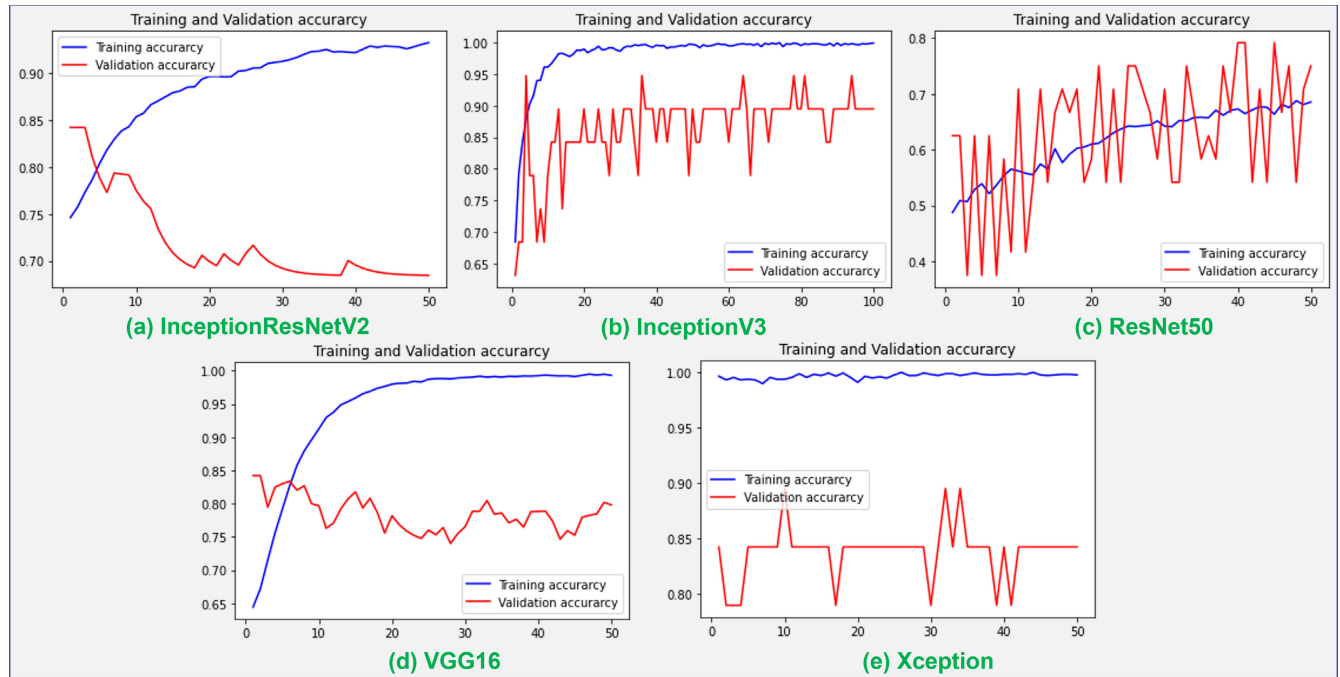
FIGURE 13. Classification training and validation Accuracy.

which are classified later with 91.0% when YOLO's classification layers are replaced by other classifiers. Hence, the main and prior task in breast cancer is detecting correctly any cancerous cases if exist. So, we completed the classification process on the resultant localized lesions from experiment 2 of the FMD-Path using strategy 2 because it obtains a larger number of TPs and a smaller number of FNs than merged both processes. The main reason behind this is that as the number of correctly detected lesions (abnormal cases) increases and the number of correctly normal cases increases, the process of cancer detection at earlier stages will be improved. Consequently, the number of truly detected cancerous cases increases which results in early treatment and reduces the deaths rate as well. Moreover, the larger number of TP and confirmed localized cases will be classified to get their pathology type.

As shown in Table 5, Inception Net v3 is the one that achieves near or better results compared with those obtained from YOLO-V4 when it is used to detect benign versus malignant lesions due to the new updates added in the net of Inception V3 like the label smoothing that is added to the loss function to avoid overfitting caused when the network becomes too confident about a specific class. Also, the batch normalization added in Inception V3 in the auxiliary classifiers enhances the features learning process resulting in better overall accuracy compared with the other classifiers. The training and the validation accuracy of each classifier is shown in Fig. 13.

E. COMPARATIVE ANALYSIS
The proposed approach is compared against other recent proposed detection and classification methods in Table 6.

After the comparative analysis and the conducted experiments, the proposed system proved its effectiveness against others that achieves high accuracy with fast speed to check breast cancer. The proposed model contributes with the following:

1) An automated full computer-aided detection and classification model that succeeds in localizing lesions in abnormal mammograms with accuracy reached 98%.
2) The proposed model is trained on normal and abnormal mammograms such that the normal cases detected as normal and the abnormal are detected with bounding boxes representing the lesions that may exist in the breast.
3) Replacing YOLO-V4's features extraction layers with other features extraction classifiers to obtain the best results achieved by the Inception V3 model for distinguishing between the benign and the malignant tumors which is nearly 95%.
4) Proposing two paths detection system represented in FMD-Path and CMD-Path which results in a robust model since it succeeds in reducing either the false positives or the false negatives decisions taken by radiologists because it acts as a second reader. Not only this by using the 2-paths proposed approach, but it also acts as a second and a third mammogram reader after the radiologists to confirm any existing lesions if exist.
5) The CMD-Path succeeds in obtaining any missed masses from FMD-Path especially the small-sized masses because this path is applied on the mammograms with preserving their original features and resolution. It serves as a confirmation radiologist for

TABLE 6. Comparison between the proposed work and another recent introduced breast mass detection and classification methods.

Paper	Main detection building block	Classification method	Dataset	Inference Time	TPR at FPI or Detection Accuracy	IOU	Classification Accuracy	Nvidia GPU
[36]	Faster-RCNN	-	Private datasets & INbreast	4 Secs	B: 0.99 at 0.03 M: 0.85 at 1.0	0.1	No classification	GPU with 12 GB
[37]	YOLOV3	Inception V3	INbreast	< 9 ms	0.92 at 0.086 ACC=89.5%	0.5	Validation = - Testing = 95.5%	RTX 2080Ti
[35]	YOLO-V2	InceptionResNet V2	DDSM & INbreast	0.025 Secs with 71 FPS	ACC= INbreast: 97.27% DDSM: 99.17%	0.5	Validation = - Testing = INbreast: 95.32% DDSM: 97.50%	GTX 1080
[32]	YOLO-V1 & FrCN	CNN	INbreast	3 Secs	98.96%	0.5	Validation = - Testing = 95.64%	GTX 1080
[42]	Morphological segmentation	FCM classifier	MIAS	Average 0.9 Secs	89%	-	Validation = - Testing = 80%	-
Proposed	2 paths detection based on YOLO-V4	Inception V3	INbreast	< 11 ms	ACC=97.86%	0.5	Validation = 95.0% Testing = 91.0%	RTX 2080Ti

checking any localized lesions by the FMD-Path using the combination detection rules.

6) The CMD-Path is used here as a novel augmentation method that increases the number of either the training or the testing samples which can be used in any other proposed model instead of the ordinary augmentation methods such as rotation, flipping, etc. It augments the input data by cropping the original image into small overlapped slices which introduces two advantages: increasing the training samples and creating new meaningful samples from the already existing ones.

7) The proposed 2-paths detection model is a real-time detection and classification system that checks any existing masses in the screened mammograms by generating the bounding boxes representing tumors and their pathology types in less than 10 seconds.

V. CONCLUSION

In this paper, we have proposed a YOLO-V4 based CAD system to localize any suspected cancerous area in the breast and classify them into benign or malignant if existing with high accuracy. The proposed model overcomes the issue of resizing the breast mammograms into smaller sizes to fit CNNs for cancer localization by proposing a mammograms cropping idea that acts as a third part detector. Nowadays, due to the huge number of mammograms taken daily and with the necessity to early discover breast cancer to reduce the death rate, the CADs play an essential role to be a second reader for the screened mammograms to decide with the radiologists one. Here, we proposed a 2-paths system that can confirm the detected cases two times without the radiologists' time. The two paths are applied together to avoid any dependency that may result in slowing the diagnosis process. The first path is responsible to take the full mammogram after resizing it into a smaller size to check any existing masses. While the second path is responsible to do the same detection process but in

a different manner by cropping the full mammogram in its original size into small overlapped slices and then each slice is treated as a separate case that the model shall search in for tumors. This path detects lesions in the cropped slices of the full mammogram, then the obtained slices are merged to get the complete bounding box of existing cancer. The second path in the proposed architecture especially succeeds in localizing the small tumors missed from the first detection path. The proposed model is a real-time detector that can detect and classify any existing masses in maximum few seconds and it also improves the detection accuracy of the already existing CADs proposed recently by reaching 98% for detection. Moreover using different experiments, it is proved that it is better to use YOLO-V4 for detecting only cancerous areas whatever their types. For the sake of this, YOLO-V4 role of the feature extraction is replaced by more than one classifier that is fine-tuned on breast cancer images which results in depending on the Inception-V3 as the best classifier since it is the one that obtains the best results among the other 6 classifiers we used to classify the detected lesions from the combined proposed 2 paths.

REFERENCES

[1] V. Chaurasia, S. Pal, and B. B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *J. Algorithms Comput. Technol.*, vol. 12, no. 2, pp. 119–126, 2018.

[2] R. Roslidar, A. Rahman, R. Muharar, M. R. Syahputra, F. Arnia, M. Syukri, B. Pradhan, and K. Munadi, "A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection," *IEEE Access*, vol. 8, pp. 116176–116194, 2020.

[3] E. U. Ekpo, M. Alakhras, and P. Brennan, "Errors in mammography cannot be solved through technology alone," *Asian Pacific J. Cancer Prevention: APJCP*, vol. 19, no. 2, p. 291, 2018.

[4] K. K. Evans, R. L. Birdwell, and J. M. Wolfe, "If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening," *PLoS ONE*, vol. 8, no. 5, 2013, Art. no. e64366.

[5] G. Hamed, M. Abd El-Rahman Marey, S. El-Sayed Amin, and M. F. Tolba, "Deep learning in breast cancer detection and classification," in *Proc. Joint Eur.-US Workshop Appl. Invariance Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 322–333.

[6] V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A. J. Frangos, and J. H. Sunshine, "How widely is computer-aided detection used in screening and diagnostic mammography?" *J. Amer. College Radiol.*, vol. 7, no. 10, pp. 802–805, Oct. 2010.

[7] M. Sato, M. Kawai, Y. Nishino, D. Shibuya, N. Ohuchi, and T. Ishibashi, "Cost-effectiveness analysis for breast cancer screening: Double reading versus single + CAD reading," *Breast Cancer*, vol. 21, no. 5, pp. 532–541, 2014.

[8] X. Bargalló, G. Santamaría, M. del Amo, P. Arguis, J. Ríos, J. Grau, M. Burrel, E. Cores, and M. Velasco, "Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program," *Eur. J. Radiol.*, vol. 83, no. 11, pp. 2019–2023, Nov. 2014.

[9] E. L. Henriksen, J. F. Carlsen, I. M. Vejborg, M. B. Nielsen, and C. A. Lauridsen, "The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: A systematic review," *Acta Radiol.*, vol. 60, no. 1, pp. 13–18, Jan. 2019.

[10] C. Romero, A. Almenar, J. M. Pinto, C. Varela, E. Muñoz, and M. Botella, "Impact on breast cancer diagnosis in a multidisciplinary unit after the incorporation of mammography digitalization and computer-aided detection systems," *Amer. J. Roentgenol.*, vol. 197, no. 6, pp. 1492–1497, Dec. 2011.

[11] T. Onega, E. J. Aiello Bowles, D. L. Miglioretti, P. A. Carney, B. M. Geller, B. C. Yankaskas, K. Kerlikowske, E. A. Sickles, and J. G. Elmore, "Radiologists' perceptions of computer aided detection versus double reading for mammography interpretation," *Academic Radiol.*, vol. 17, no. 10, pp. 1217–1226, Oct. 2010.

[12] G. Hamed, M. Abd El-Rahman Marey, S. El-Sayed Amin, and M. F. Tolba, "The mass size effect on the breast cancer detection using 2-levels of evaluation," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.* Cham, Switzerland: Springer, 2020, pp. 324–335.

[13] M. A. S. A. Husaini, M. H. Habaebi, S. A. Hameed, M. R. Islam, and T. S. Gunawan, "A systematic review of breast cancer detection using thermography and neural networks," *IEEE Access*, vol. 8, pp. 208922–208937, 2020.

[14] G. Hamed, M. Marey, S. Amin, and M. Tolba, "Comparative study and analysis of recent computer aided diagnosis systems for masses detection in mammograms," *Int. J. Intell. Comput. Inf. Sci.*, vol. 21, no. 1, pp. 33–48, Feb. 2021.

[15] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] N. I. R. Yassin, S. Omran, E. M. F. El Houby, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review," *Comput. Methods Programs Biomed.*, vol. 156, pp. 25–45, Mar. 2018.

[18] I. Hirra, M. Ahmad, A. Hussain, M. U. Ashraf, I. A. Saeed, S. F. Qadri, A. M. Alghamdi, and A. S. Alfakeeh, "Breast cancer classification from histopathological images using patch-based deep learning modeling," *IEEE Access*, vol. 9, pp. 24273–24287, 2021.

[19] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, Jul. 2016.

[20] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Comput. Methods Program. Biomed.*, vol. 127, pp. 248–257, Apr. 2016.

[21] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Med. Imag. Graph.*, vol. 31, nos. 4–5, pp. 198–211, Jun. 2007.

[22] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," *Academic Radiol.*, vol. 6, no. 1, pp. 22–33, 1999.

[23] H.-P. Chan, E. Charles, P. Metz, K. L. Lam, Y. Wu, and H. Macmahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms," *Arbor*, vol. 1001, p. 48109, Nov. 1990.

[24] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in *Proc. 5th Int. Conf. Bioinf. Comput. Biol.*, Jan. 2017, pp. 59–62.

[25] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale CNN and curriculum learning strategy for mammogram classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 169–177.

[26] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Med. Image Anal.*, vol. 37, pp. 114–128, Apr. 2017.

[27] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.

[28] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, and S. Jastrzęb, "Deep neural networks improve Radiologists' performance in breast cancer screening," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1184–1194, Apr. 2020.

[29] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, T. H. Helbich, M. Chevalier, T. Tan, T. Mertelmeier, M. G. Wallis, I. Andersson, S. Zackrisson, R. M. Mann, and I. Sechopoulos, "Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists," *JNCI: J. Nat. Cancer Inst.*, vol. 111, no. 9, pp. 916–922, Sep. 2019.

[30] A. Rodríguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S. H. Heywang-Köbrunner, I. Sechopoulos, and R. M. Mann, "Detection of breast cancer with mammography: Effect of an artificial intelligence support system," *Radiology*, vol. 290, no. 2, pp. 305–314, Feb. 2019.

[31] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, and T. Back, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020.

[32] M. A. Al-Antari, M. A. Al-Masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *Int. J. Med. Inform.*, vol. 117, pp. 44–54, Sep. 2018.

[33] H. Jung, B. Kim, I. Lee, M. Yoo, J. Lee, S. Ham, O. Woo, and J. Kang, "Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0203355.

[34] M. A. Al-masni, M. A. Al-antari, J.-M. Park, G. Gi, T.-Y. Kim, P. Rivera, E. Valarezo, M.-T. Choi, S.-M. Han, and 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.*, vol. 157, pp. 85–94, Apr. 2018.

[35] M. A. Al-antari, S.-M. Han, and 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.*, vol. 196, Nov. 2020, Art. no. 105584.

[36] R. Agarwal, O. Díaz, M. H. Yap, X. Lladó, and R. Martí, "Deep learning for mass detection in full field digital mammograms," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103774.

[37] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO based breast masses detection and classification in full-field digital mammograms," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105823.

[38] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. Cardoso, J. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Academic Radiol.*, vol. 19, no. 2, pp. 236–248, 2012.

[39] *The INbreast Dataset*. [Online]. Available: http://medicalresearch. inescporto.pt/breastresearch/index.php/Get_INbreast_Database

[40] S. Singh and K. Bovis, "An evaluation of contrast enhancement techniques for mammographic breast masses," *IEEE Trans. Inf. Technol. Biomed.*, vol. 9, no. 1, pp. 109–119, Mar. 2005.

[41] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. [Online]. Available: http://arxiv.org/abs/2004.10934

[42] P. B. Chanda and S. K. Sarkar, "Detection and classification of breast cancer in mammographic images using efficient image segmentation technique," in *Advances in Control, Signal Processing and Energy Systems*. Singapore: Springer, 2020, pp. 107–117.

GHADA HAMED (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in scientific computing from the Faculty of Computers and Information Sciences, Ain Shams University, Cairo, Egypt, in 2012 and 2017, respectively, and the M.Sc. degree in DNA-based steganography. She is currently pursuing the Ph.D. degree with the Department of Scientific Computing, Faculty of Computers and Information Sciences, Ain Shams University. Since 2013, she has been working as a teaching assistant. Since 2017, she has been working as an assistant lecturer and a researcher in scientific computing. Since 2018, she has started to go through the research in breast cancer diagnosis and how to utilize the computer science field with the medical field to contribute with added value CADs. She has participated in many research projects related to image processing like video resolution enhancement, DNA steganography, and operating systems development. Since 2013, she coauthored about 13 research articles in refereed journals, such as *Computer Methods and Programs in Biomedicine* (Elsevier) besides other international conferences. Her research interests include the merge of neural networks, algorithms analysis and design, deep learning methodologies, and computer vision with bioinformatics. She is a Reviewer in some distinguished international journals, such as *Multimedia Systems* and the *International Journal of Intelligent Computing and Information Sciences*.
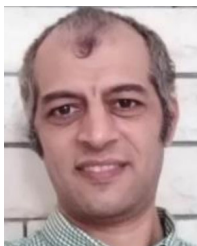
MOHAMMED MAREY received the Ph.D. degree from INRIA-Rennes, France, in 2011. In May 2008, he visited the Artificial Intelligence Group of Mobile Robots, Beckman Institute, University of Illinois at Urbana-Champaign, USA. In May 2009, he also visited the Humanoid Research Group, National Institute of Advanced Industrial Science and Technology (AIST), Japan. Since 2017, he has been the Director of the Information Technology Research and Consultancy Center, Ain Shams University. Since 2021, he has been a member of the research council at Ain Shams University. He is currently an Associate Professor of scientific computing with the Faculty of Computer and Information Sciences, Ain Shams University. His research interests include utilization of methodologies and technologies of artificial intelligence and computer vision in the fields of medical and robotic applications. These include action recognition, autonomous robotics control, robotics motion planning, task planning, robotic kinematic, redundancy, and singularities of the manipulator, vehicles, humanoid, and medical robots.

SAFAA ELSAYED AMIN received the M.Sc. degree in pure mathematics and computer sciences from the Faculty of Sciences, Ain Shams University, and the Ph.D. degree in computer and information sciences from Heinz Nixdorf Institute, Paderborn University, Germany. She is currently an Associate Professor with the Department of Scientific Computing, Faculty of Computer and Information Sciences, Ain Shams University. She has supervised over 15 M.Sc. and Ph.D. students with the faculties of computers in Egyptian universities. Her research interests include biomedical, image processing, and nanorobotics. She has more than 40 publications in these areas. She shared in many computers sciences local and international conferences. She was a Reviewer in some international journals, such as the *International Journal of Intelligent Computing and Information Systems Sciences*.

MOHAMED F. TOLBA (Senior Member, IEEE) has been a Professor of scientific computing with Ain Shams University, since 1984. He was the Vice President from Ain Shams University, from 2002 to 2006, and the Dean of the Faculty of Computers and Information Sciences, from 1996 to 2002. He has supervised more than 90 M.Sc. and 50 Ph.D. degrees at Ain Shams University and other Egyptian Universities. He is currently a consultant to different local and international organizations for IT. He has more than 220 publications in the fields of AI, image processing, pattern recognition, OCR, scientific computing, and simulation and modeling. He was a member of the International Association for Science and Technology for Development (IASTED), Canada, from 1995 to 2007, the International Society for Computers and their Applications (ISCA), USA, from 1998 to 2007, the Advisory Committee of Strengthening Science and Technology Researchers Project—STRP Ministry of Scientific Research, from 2006 to 2009, and the Committee for Evaluation of Egyptian Space Program of the National Authority for Remote Sensing and Space Sciences—Ministry of Scientific Research. He has been also a member of the Association for Computing Machinery (ACM), USA, since 2000, the Software Engineering Competence Center (SECC), since 2004, the Information Technology Academic Collaboration (ITAC), since 2005, and the E-Learning Committee Board, since 2008. He is also the honorary chairman of many international conferences and the chairman of several IT sector committees in Egypt.

● ● ●