ISSN 1870-4069

Breast Anomaly Detection Using YOLO-Based Deep Learning CAD System in Digital Mammograms: A Second Opinion Tool for Breast Cancer Diagnosis (1st Stage)

Johan Uriel García-Pérez[1], Moisés Márquez-Olivera[1], Viridiana Hernández-Herrera[1], Laura Marrujo-García[2]

[1] Instituto Politécnico Nacional, CIITEC, Mexico

[2] Instituto Politécnico Nacional, CECyT 2, Mexico

johanuriel75@gmail.com, {vhernándezhe,mvmarquez,lmarrujog}@ipn.mx

Abstract. Since digital mammograms is currently the most widely used tool worldwide for breast cancer detection, it is important to develop a system capable of supporting clinical decision making by detecting breast abnormalities, and thus, in conjunction with a radiologist, provide a more accurate diagnosis. Therefore, we propose to use an artificial intelligence (AI) algorithm based on convolutional neural networks (CNN) to detect breast abnormalities in digital mammograms. The analysis was performed on an integrated database of 1,213 digital mammograms obtained from a public database. Using a Hold-Out 70-30 validation method, the database was divided into two sections: training set (849 images) and validation set (364 images). Three training sessions were conducted, and each one was set up with 100 epochs: the first one consisted of batch size = 4, obtaining a maximum accuracy of 68 % in the semantic segmentation mode. The second training was performed under semantic segmentation, with batch size = 8, achieving an accuracy up to 91%. The third one was performed with batch size = 8 in segmentation mode, to the pre-processed database with a brightness and contrast enhancement filter, obtaining a maximum accuracy of 69 %. This allows us to conclude that the CNN is able to identify abnormalities in breast tissue. In addition, an increase in the accuracy and sensitivity of the CNN was observed when batch size increased, by making conditions under network was trained were factors that influenced the extraction of information from the mammograms.

Keywords: Breast cancer, abnormalities detection, computer aided detection, deep learning, you only look once, YOLO v8.

1 Introduction

Cancer is the leading cause of death in the world [1], and breast cancer is placed as most incidence as well as one of the most common death causes as far as cancer is concerned in women both in México, United states and worldwide [2, 3]. Is it estimated that, in 2020, about 2,261,419 new breast cancer cases in women were registered which represented approximately 24.5 % of the total new cancer cases worldwide in women (Fig. 1-a), furthermore, for this same year, of the total new registered breast cancer cases, 684,996 deaths were attributed to breast cancer, which represented 15.5 % of the total cancer deaths worldwide only in women (Fig. 1-b) [4, 5].

Breast cancer occurs when breast cells begin to grow out of control, forming a mass or conglomerate named tumour, which may be cancerous (malignant) or benign [6, 7]. Owing to early stages of breast cancer presents subclinically, currently within different clinical studies for breast cancer detection, screening mammography is one of the most common used tools for early breast cancer detection [8, 9].

Conventional screed-based mammography has given way to digital mammography, resulting in many benefits, including a simplified workflow and improved performance in certain patient subgroups, as well as the revolution in breast cancer care was witnessed by the introduction of mammography as an optimized radiographic imaging modality for the breast [10]. Mammography is a standard screening method for early breast cancer detection, however, it is really difficult for radiologist to provide accurate predictions for early detection, as it is complicated to interpret expertly due to various factors [11, 12].

In medicine, Artificial Intelligence (AI) has got two main application branches: physical and virtual, where virtual components are represented by Machine Learning (ML) or Deep Learning (DL), considering DL is developed by mathematical algorithms focused on improving learning through experience [13].

ML is dedicated to research and implement methods aiming to provide computers the ability to learn how to solve problems with explicit programming solutions, while DL is defined by multiple non-linear transformation modules combination, successfully modifying input information, achieving an internal data representation at multiple complexity and abstraction levels [14, 15].

This has made DL algorithms gain attention due to their considerable success because they can automatically learn feature representations and making feature extraction can be achieved from data without the need for prior definition by human experts, allowing this data-driven approach defining more abstract features, turning it more informative and generalisable [12, 16].

Deep learning remains within ML domain, and it is a special class of Artificial Neural Networks (ANN) that resembles the multi-layered human cognition system [8], besides, there is Convolutional Neural Network (CNN) another class of ANN that has become dominant in different computer vision tasks, including radiology [17], emphasizing in mammographic classification [18] because of its examination, recognition or image classification capability [19].

In recent years, CNNs have been applied in digital medical images classification for breast cancer detection and prediction because CNN application in breast cancer screening has gotten a significant advantage over traditional methods with respect to

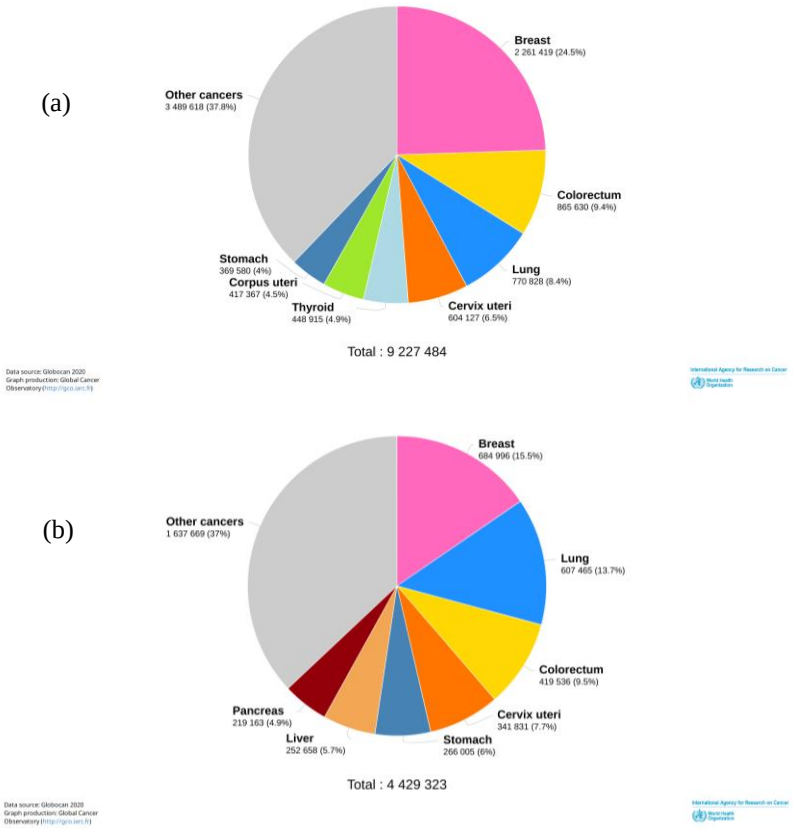

Fig. 1. Graphical schemes of 2020 mortality and incidence cancer estimations in female where a) shows the incidence number of new cases and b) the deaths number of breast cancer (GLOBOCAN 2020) [4].

the time taken to perform each test, where conventionally methods can take too long to analyse one piece of data at the time [19].

Because of breast cancer can be properly treated if an early diagnosis is correctly determined, making it feasible to have got screening methods for detecting early breast cancer signs [20], allowing AI to be an interesting factor for its application to support the detection of masses or microcalcifications present in breast tissue that can be visualized by mammography which can assist the physician in making a breast cancer diagnosis [21].

Initially in health care, computers were used in clinical image for administrative work such as image acquisition and storage, until now, they have become indispensable components in work environment, which includes the use of Computer Aided Diagnosis (CAD) systems [22], that is an AI for which is used to assist radiologist and cut back workload [23].

Nowadays, many AI algorithms are already being used in medical field (Figure 2) [22] but the potential use that can be given to AI in breast cancer diagnosis, extends to

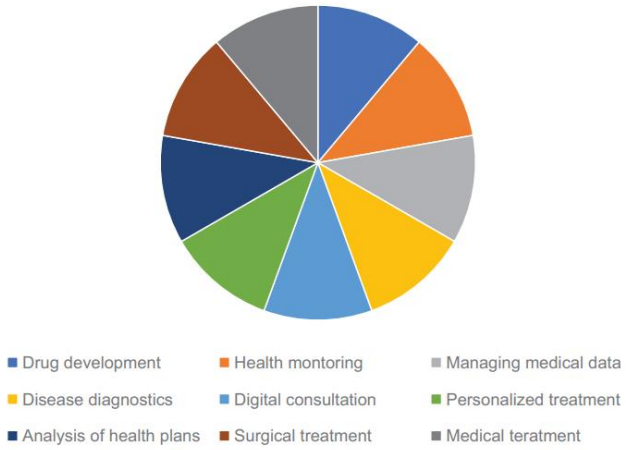

Fig. 2. Artificial Intelligence applications in health care [22].

support modalities in image interpretation and histopathology [21] because early detection can potentially improve the prognosis of breast cancer and significantly reduce mortality in women [24], therefore, CAD systems play an important role either for Computer Aided Detection (CADe), that focus on locating suspicious lesions; or Computer Aided Diagnosis (CADx), focusing on determining whether a previously detected lesion is benign or malignant [25].

Moreover, considering the current different techniques that provide digital results for breast cancer detection, DL opens a new way to be implemented in digital mammogram analysis, because it is able to integrate it for different tasks such as: injuries segmentation and classification; image generation and reconstruction; cancer risk prediction; and therapy response prediction and evaluation, where results have shown similar or better results by DL algorithm than radiologist results [23].

1.1 State-of-the-Art

Sundries researchers have proposed different CAD systems which may help for breast cancer detection or diagnosis using digital mammograms. Recent studies [26] implement an autonomous diagnosing cancer system using an integration method including CNN and image texture attribute extraction, applying a customised nine-layered CNN for categorizing in CNN stage, reaching a specificity and accuracy of 97.8 % and 98 % respectively for this method tested on MIAS (Mammographic Image Analysis Society) repository, moreover 98.8 % and 97.9 % were reached when tested on DDSM (Digital Database for Screening Mammography) repository.

On the other hand, a breast cancer image detection and a model based on convolutional and deconvolutional neural network (CDNN) was proved [27] testing the algorithm on a common dataset for ROI (Region Of Interest) segmentation, showing the model automatic classification performance improving of breast cancer which may provide a new idea about using medical diagnosis assisted by artificial intelligence. CNN method [20] was proposed to boost automatic breast cancer identification by

analysing hostile ductal carcinoma tissue zones using a dataset of 275,000 images, founding the model successful due to 87 % accuracy achieved which is approximately 9 % more than ML reached, resulting as a probably option for reducing human mistakes.

Nevertheless, Rehman et al. [11] reached a 97 % score with a 2.35 and 99 % true positive ratio with 2.45 false positives per image by the Fully Connected Depthwise Separable Convolutional Neural Network (FC-DSCNN) computer-vision-based tested model on 35688 DDSM images and 2885 PINUM images respectively.

Ortíz-Rodríguez et al. [28] used image processing techniques to develop imaging biomarkers through mammographic analysis for breast cancer detection in early stages, training and testing a generalized regression ANN to classify malignant and benign tumours with a 95.83 % of accuracy reached.

Salama et al. [29] proposed a breast cancer image segmentation and classification framework including different pre-trained models, applied to MIAS, DDSM, and Curated Breast Image Subset of DDSM (CBIS-DDSM) for benign and malignant classification, showing U-Net model and InceptionV3 model the best results with 98.87 % of accuracy achieved.

Muduli et al. [30] tested a five learnable layers of four convolutional layers and a fully connected layer CNN model to facilitate automatically extraction of prominent features from mammograms and ultrasounds datasets, achieving a 96.55 %, 90.68 %, and 91.28 % accuracy in MIAS, DDSM and INbreast datasets respectively, moreover a 100 % and 89.73 % accuracy were achieved from BUS-1 and BUS-2 datasets respectively.

Agarwal et al. [31] state a patch-based CNN method for automated mass detection in Full Field Digital Mammograms (FFDM), training the model using CBIS-DDSM and INbreast datasets where InceptionV3 showed the best performance.

Since the detection and diagnosis of abnormalities in digital mammograms analysis is a still challenging task for radiologist because of necessity of analysing and identifying a “small” number of cancers that depends mostly on manual segmentation (which may take too long), computer equipment or operator [8], and even other different factors as: normal breast tissue variable appearance, overlapping tissue structures which may hide injuries in breast density tissues hindering mass detection [32, 33], breast radiographic complex structure and radiologist fatigue or distraction, that contribute to radiologist difficulties with misdiagnostic interpretations on mammograms [34, 35].

In this this paper we propose a semi-automatic system in which, the radiologist will be able to give a diagnosis by using CAD system as a support tool for abnormalities detection and then, just in few seconds, the radiologist can apply some different filters to the mammogram, analyse it, and finally, to get preliminary results that allows the radiologist to make a diagnosis, considering the AI algorithm provided results.

The use of AI algorithms in medical environment suggests that specialist can enhance the diagnosis 20 % more than a diagnosis made only by radiologist. This means the use of CAD systems will highly reduce misdiagnosis, improving diagnostic accuracy and sensitivity, decreasing radiologist workflow, visual fatigue due to mammography reading rates, and avoiding manual segmentation for greater productivity without impact diagnostic opinion [8, 36, - 39].

2 Materials and Methods

Convolutional Neural Networks employ the convolution operation as one of their layers, which perform similar operations to image processing filters [40, 41]. Convolutional layer consists of filters and image maps [42] taking an image and a small logistic regression, passing the logistic regression over the whole image [43]. Using a CNN with fewer parameters might improve significantly the time it takes to learn by reading the image "chunk-by-chunk" with the aim of allows to convolution to extract features from the input image preserving the spatial relationship between pixels [44, 42].

In CNN, the convolution operation is very similar to Gaussian and Sobel filters in image processing, because a kernel slides across an image analysing nearby pixels multiplying the weights with each aligned pixel, element-wise across the filter to finally add a bias value to the output [41, 44]. The amount the kernel shifts between pixels is called "stride" [44]. The develop procedure for the proposed CAD system for breast abnormalities detection was divided into 4 stages which included image and data acquisition, sorting and type approval; image pre-processing; model training; and finally, CAD system evaluation.

2.1 Dataset

In this study, digital mammograms were collected from Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [45] to train and validate proposed CAD system. CBIS-DDSM is an open access database which includes 10,239 images from 6,775 studies. CBIS-DDSM was analysed in order to rule out incomplete images (e.g. masks), or missing data corresponding to coordinates. Later, the resulting 1,213 images with respective data coordinates, were split according to *Hold-Out* method in a 70:30 ratios for training and validation packages, respectively.

Then, once split the dataset, is was standardized ensuring that 70 percent of training images contained the same number of images from Craniocaudal (CC) and Mediolateral (MLO) views, as well as the same number of images with and without anomalies, collecting masses according to Breast Imaging Reporting and Data System (BI-RADS), a system that allows to standardized terminology, systematize mammographic reports, and lesions categorizing, stablishing suspicion degree [46]. In addition, data number concerning to anomalies coordinates must correspond to training set of images number. Due to all images acquired from DDSM were stablished to 640x640 pixels, no more resizing processes were required. The same procedure was replicated for validation set.

2.2 Image Pre-Processing

After building the database, an adequate image pre-processing filter was carried out to improve contrast and brightness between the regions of interest (ROI) and the other sections of the mammography. In addition, performing an adequate segmentation of the mammograms will allow us to optimise the resources acquired, making image processing more efficient for the algorithm training. For the above reasons, an

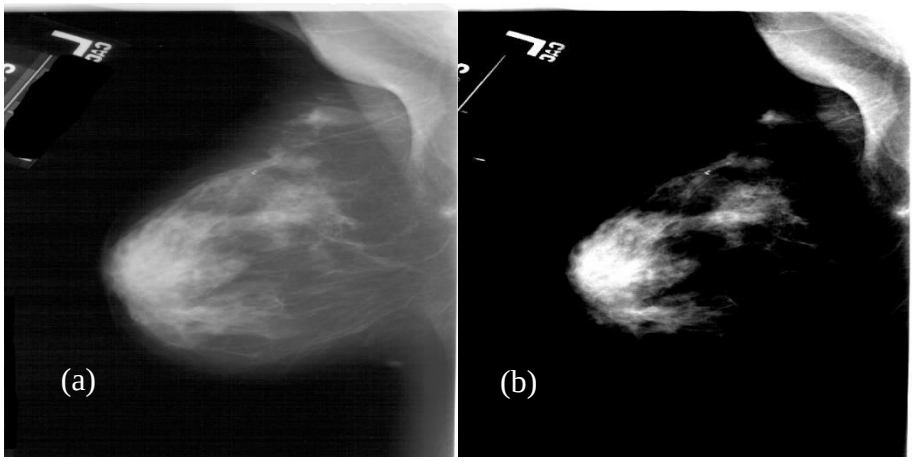

Fig. 3. Mammography comparison of left breast in MLO view where (a) is the raw mammogram from CBIS-DDSM, and (b) is the same mammogram with brightness and contrast enhancement filter.

enhancement filter image was developed in python to discard most of the fatty tissue, allowing fibrous tissue visualisation where any abnormality may be contained (Fig. 3).

2.3 You Only Look Once (YOLO) Model Training

Mass detection in breast tissue is a critical task for CAD systems [39], nevertheless, instead of developing our own deep model, an existing model will firstly use and adapted to solve our problem. Therefore, a Darknet-19 classifier model which forms the basis of real-time object detection system named YOLO [47, 48], was selected. YOLO is one of the state-of-the-art deep learning techniques [49, 47, 50] that uses a single convolutional network to whole image by dividing the input image into sub-regions and predicts multiple bounding boxes with their respective class probabilities for each region [49, 50].

YOLO is a unified model which original structure consist of 24 convolution layers, followed by 2 fully connected layers and trains on full images and directly optimizes detection performance [51, 50]. For this proposed CAD system, the YOLOv8 model released in 2023 by Ultralytics [52], was selected for abnormalities detection in breast tissue.

First, for running YOLOv8, a new environment was created to stablish a specific space to this model. The new environment was created using Anaconda prompt, where all Ultralytics packages and other needed libraries were installed. Moreover, other hyperparameters were modified within the model to focus the model on solving the abnormalities detection problem and use it the most efficient way. Once done, the resulted pre-processed Data Base of 6,065 digital mammograms (RDDSM) was used for model learning stage, where 70 % of RDDSM was used for training set realising a three-times training of 100 epochs and 0.5 confidence each.

Initial training consisted of batch size = 8 in detection mode. Second training was carried out in semantic segmentation [53] mode with batch size = 4, and final training

Table 1. Comparison between trained models to training time and testing time.

Model	Training time/epoch (s)	Testing time/image (s)
YOLO8-S4	600	4
YOLO8-S8	780	6

Table 2. Conditions for training YOLOv7 and YOLOv8 models.

Model	epoch	mode	confidence
YOLO7-S8	100	segment	0.5
YOLO8-D8	100	detect	0.5

was held in semantic segmentation mode with batch size = 8. This to visualise the difference between the performance of each model conditions in identifying abnormalities with different breast tissue conditions through mammography. The idea of using different bath sizes, it is to support the training stage, considering that a bigger batch size can accelerate it, but also requires more GPU memory which is limited to 4 GB provided by RTX 3050 Ti GPU from computer where model is trained. Using a lower bath size results in lower memory consumption but training speed could be affected.

2.4 CAD System Evaluation

Once all the previous stages were done, and proposed model were validated and tested, finally the CAD system was tested by running each of the proposed models over 20 randomly selected mammograms from mini-MIAS database to determine which of them have got the best performance on determining abnormalities trough breast density tissue. The criteria of sensitivity and accuracy evaluation are presented in equations 1-2 respectively [39, 49]:

$$Sensitivity\ (ST) = \frac{TP}{TP+FN}, \tag{1}$$

$$Accuracy\ (AC) = \frac{TP+TN}{TP+FN+TN+FP}, \tag{2}$$

where TP and FN, TN, and FP, correspond to true positive, false negative, true negative and false positive evaluated cases, respectively.

For determining TP, FN, TN, and FP classification, TP cases were considered when CAD system reached a score over 95 % on tested mammogram and it corresponded to mini-MIAS database selected coordinates; FN were considered when there were no tumours detected by CAD system, but mini-MIAS mammograms presented any tumour; TN were classified when CAD system no detect any tumours and the mini-MIAS mammograms no presented any tumour too; and FP were considered when CAD system predicted a tumour in a place that was discordant with mini-MIAS database coordinates or there were no present tumours in mini-MIAS database, to finally compare time taken while training and image prediction (Table 1).

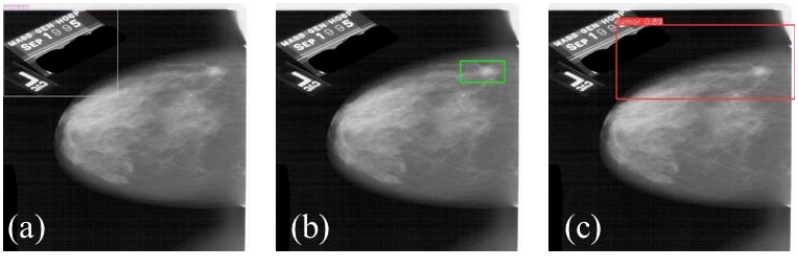

Fig. 4. Validation comparison where (a) is the segmentation mode with YOLOv7 model detecting the mammogram label inside pink bounding box; (b) is the original image with ROI visualized as a green bounding box; and (c) is the YOLOv8 detection where tumour is inside wide red bounding box.

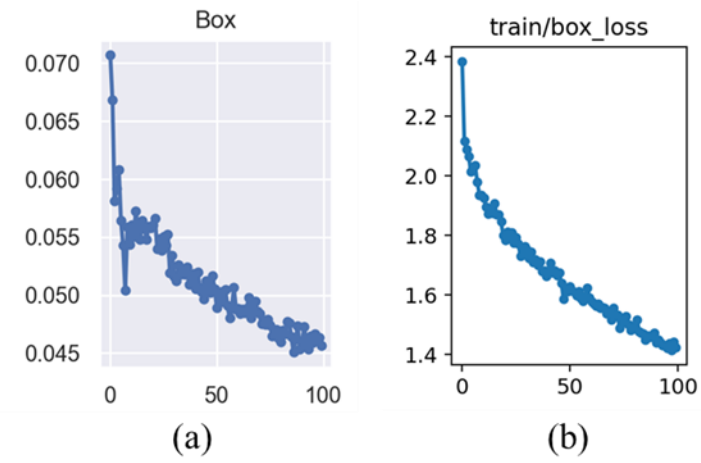

Fig. 5. Box loss training in 100 epochs for (a) YOLO7-S8 model and (b) YOLO8-D8 model.

YOLO8-D8, YOLO8-S4 and YOLO8-S8, correspond to detection mode with batch size = 8, segmentation mode with batch size = 4 and segmentation mode with batch size = 8, respectively.

3 Results

To present the results and compare the evolution between models, the YOLOv7 model was used to train under the same conditions as the YOLOv8 model, just changing to segment mode (Table 2).

Where YOLO7-S8 and YOLO8-D8 corresponds to YOLOv7 model in segmentation mode with batch size = 8, and YOLOv8 in detection mode with batch size = 8, respectively. The metric results are shown below for training and validation stages, this will allow to compare the accuracy obtained in every processed image by both models. In validation process is observed that YOLOv8 model has a higher accuracy than YOLOv7 model.

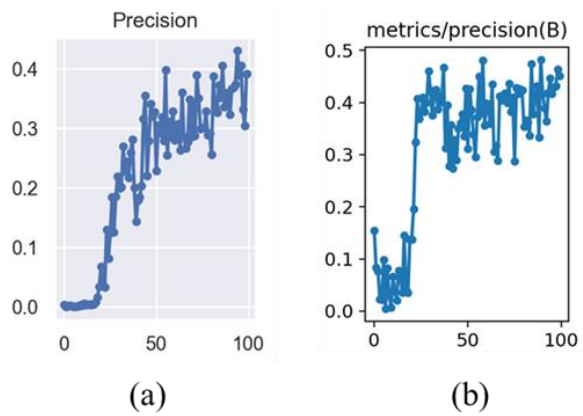

Fig. 6. Precision metrics over 100 epochs for each (a) YOLO7-S8, and (b) YOLO8-D8 models.

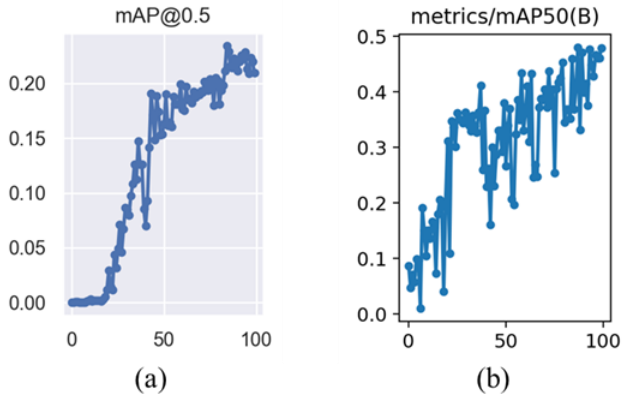

Fig. 7. mAP metrics over 100 epochs for each (a) YOLO7-S8 and (b) YOLO8-D8 models, showing the confidence increasing with every epoch realised.

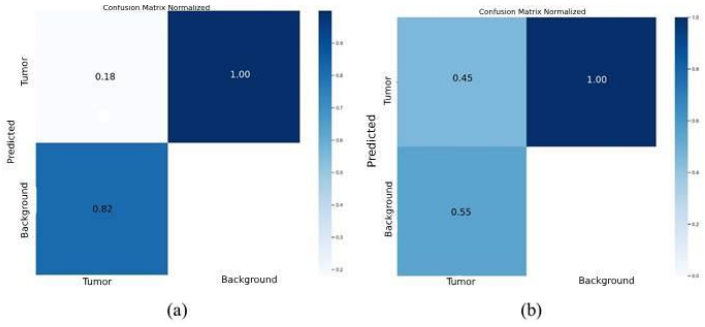

Fig. 8. Confusion matrices for (a) YOLO7-S8 and (b) YOLO8-D8 models, with 18 % accuracy and 45 % accuracy, respectively.

Even when both models have got a significant error when making prediction, the bounding box in YOLOv8 model is more accurate than YOLOv7 model (Fig. 4), as

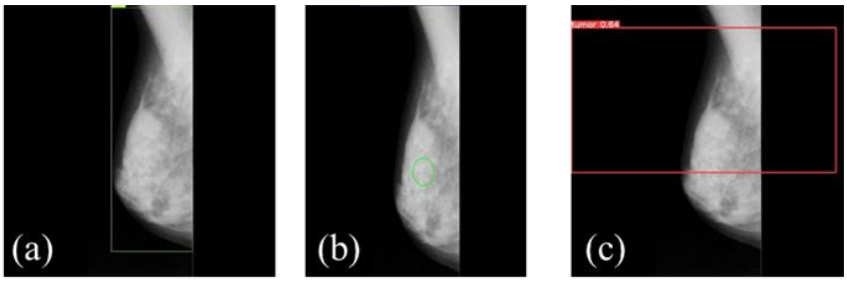

Fig. 9. Mini-MIAS database predictions for (a) YOLO7-S8 detecting all breast and pectoralis muscle inside green bounding box; (b) the original image with ROI visualized as a green bounding circle; and (c) the YOLO8-D8 prediction where tumour is inside a pretty wide red bounding box.

YOLOv7 model gets confused with mammography label, however, in YOLOv8 model although tumour is within bounding box, it is very wide, making less accurate as a support to radiologist. It was noted in training stage that errors decreased as the number of epochs performed increased (Fig. 5).

The box loss may be attributed to the model learning highly significant features through each iteration, whether it extracts values related to pixels within the segmented area, considers neighbouring pixels to compare with those found outside the segmentation boundaries, and considers features within the ROI boundary to inherit those features to the next layers.

The precision and mAP (mean Average Precision) were observed in both models, noting that both accuracy and average accuracy remain quite dispersed. Precision in YOLOv7 model, first epochs it keeps increasing, and as it approaches 50 epochs, it starts to disperse highly significantly, which may be attributed to low feature extraction trough convolution layers, which prevented it from determining where any tumours were located.

On the other hand, for model 8, initially the accuracy was mostly close, although it was low, but as the epochs were performed, a significant increase was obtained, although it was still quite sparse (Fig. 6). Nevertheless, mAP was slightly less dispersed in YOLOv7 model, however, it has quite significant increases and losses, contrary to YOLOv8 model, which has a totally dispersed mAP and even significant information loss after reaching its peak (Fig. 7).

It is observed how box and accuracy vary markedly from one model to another when certain hyperparameters are modified. Furthermore, in confusion matrices (Fig. 8) was noticed, even when YOLOv7 model was in segmentation mode, and epochs were the same, YOLOv8 model shown to be more efficient at screening in mammograms, reaching up to 0.45 % accuracy detection which is 27% more accurate than the 0.18 % reached by YOLOv7 model.

Finally, when testing for mini-MIAS database predictions, results were lower than expected, as both models presented slightly the same pattern at the validation stage, with YOLO8-D8 model being relatively closer to detecting the tumour, as it was within the delimited area, but this was too large, leaving a fairly wide error margin, as opposed to YOLO7-D8 model, which simply delimited the entire region belonging to the breast and pectoralis muscle (Fig. 9).

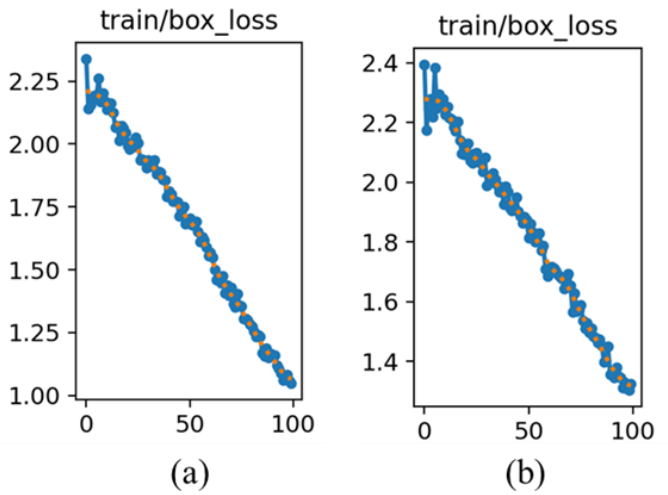

Fig. 10. Box loss during training in 100 epochs for (a) YOLO8-S4 model and (b) YOLO8-s8 model.

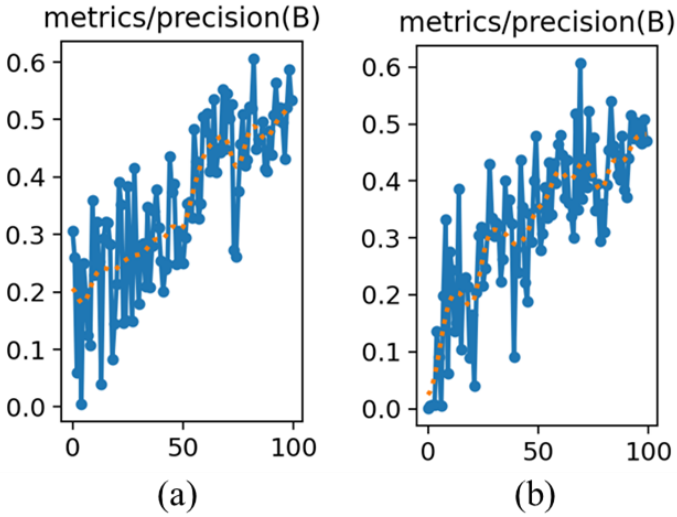

Fig. 11. Precision graphs over 100 epochs for each (a) YOLO8-S4, and (b) YOLO8-S8 models. Orange dots show the increasing accuracy of each model, being lower the YOLO8-S8 accuracy at starting point.

3.1 Comparison of Training Results

Analysing the previous results, it is possible to observe a significant efficiency with respect to the YOLOv8 model, which indicates that this model is able to extract features that allow it to differentiate the anomaly from the breast tissue.

From the resulting box loss graphs of each training for segmentation task, it can be noticed that, in both cases, the YOLO8-S4 model and the YOLO8-S8 model showed a

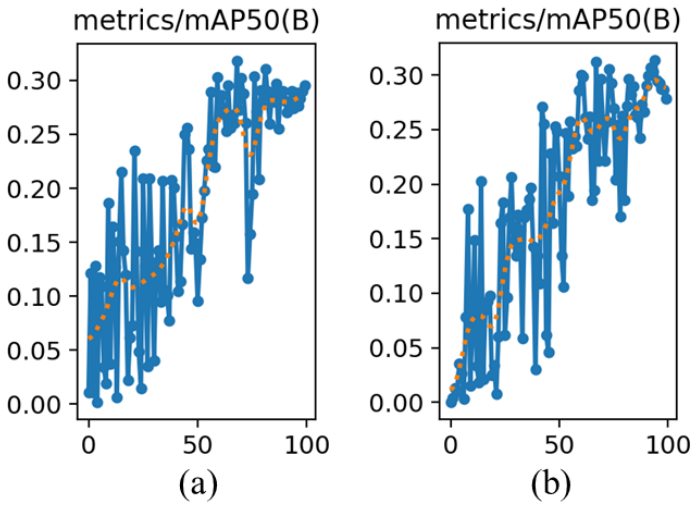

Fig. 3. mAP metrics over 100 epochs for each (a) YOLO8-S4 and (b) YOLO8-S8 models. Orange dots show the increasing confidence with every epoch realised and accuracy behaviour.

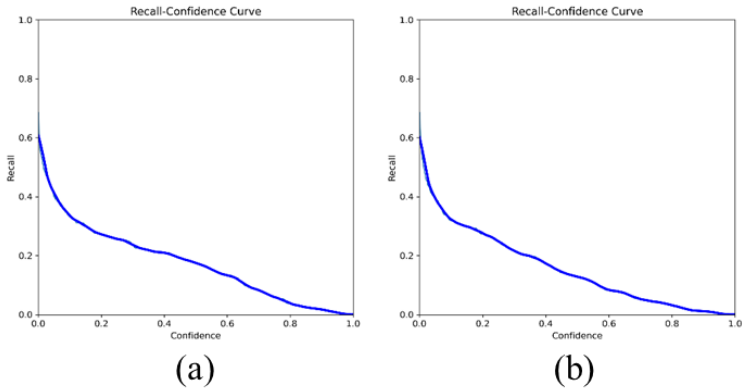

Fig. 4. Confidence function with (a) recall approximately to 60 % and confidence up to 65 % for YOLO8-S4 model and (b) 60 % recall and 65 % confidence for YOLO8-S8 model.

very low dispersion, moreover, this low dispersion behaves in a constant way during the 100 training epochs (Fig. 10), although, initially, the YOLO8-S8 model showed a higher dispersion in the first epochs.

This could imply the loss information was much lower for both models when training was carried out in segmentation mode.

As for the accuracy results (Fig. 11), it could be noted a quite significant dispersion in both models, denoting that their maximum accuracy is even below their maximum peak accuracy. In both models a slight dispersion decreasing can be seen when they are close to 100 epochs, but in the YOLO8-S8 model, the dispersion is much more noticeable. It can be observed that at beginning YOLO8-S8 has got an even lower

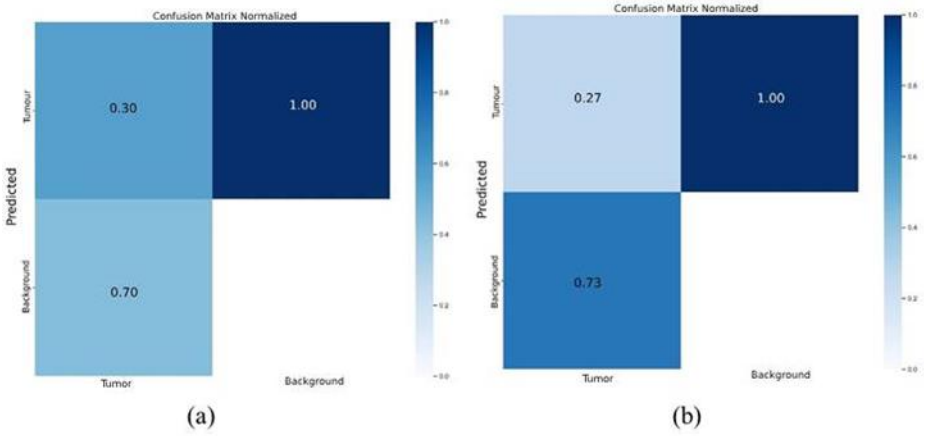

Fig. 14. Confusion matrices for (a) YOLO8-S4 and (b) YOLO8-S8 models, with 30 % accuracy and 27 % accuracy, respectively.

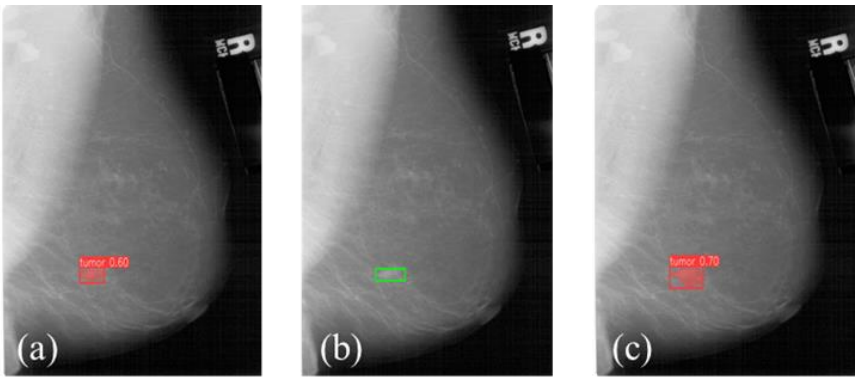

Fig. 155. Score validation comparison where (a) YOLO8-S4 reaching 60 % of detection showed inside red bounding box; (b) is the original image with ROI visualized as a green bounding box; and (c) YOLO8-S8 with 70 % of detection where tumour is inside red bounding box.

accuracy than the YOLO-S4 model (orange dots). This would imply that the extraction of highly significant features is quite complicated, which could be due to the complexity of the interpretation of the anomaly through the superimposition of the breast tissue on the mammograms. For the case of mAP, in the YOLO8-S4 model, the dispersion that exists during the learning process is observable, however, when approaching epoch 50, there is a more uniform gradual increase, but subsequently there is a quite noticeable decrease in accuracy approximately between epochs 70-80.

In the YOLO8-S8 model, the increase, although also very dispersed, is more uniformly increasing, with no significant decrease in accuracy over the 100 epochs (Fig. 12). Comparing the confidence-recall curves (Fig. 13), it is seen that difference between one model and other are too low (only 1 %), which could reflect a similar behaviour at the time of making detections in the mammograms.

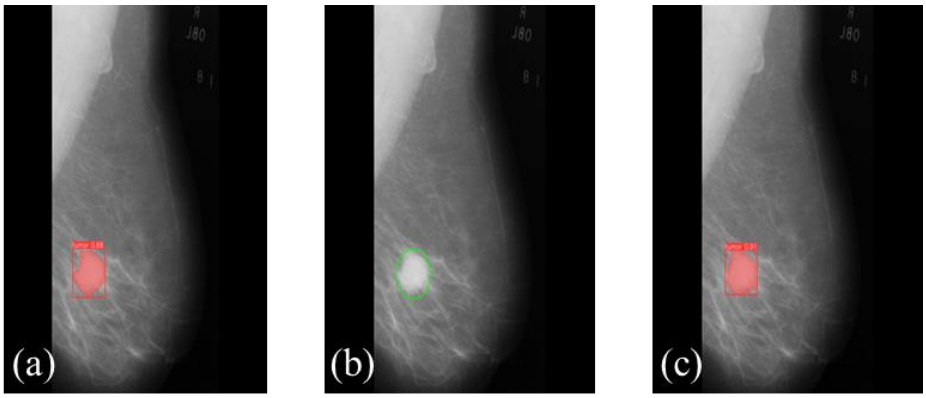

Fig. 16. Evaluation test for (a) YOLO8-S4 reaching 68 % of detection showed inside red bounding box; (b) is the original image with ROI visualized as a green bounding circle; and (c) YOLO8-S8 with 91 % of detection where tumour is segmented inside red bounding box.

Table 3. ST and AC comparison results for YOLO8-S4 and YOLO8-S8 models trough TP, FN, TN, and TP calculations.

Model	TP	FN	TN	FP	ST	AC
YOLO8-S4	15	2	1	1	0.88	0.84
YOLO8-S8	18	1		1	0.95	0.90

Furthermore, we can note the confidence of both models is moderately strong in learning, which implies both models can continue learning and could be potential candidates for future applications.

3.2 Validation Results

Contrary to what would be expected, considering the results provided by the YOLO8-D8 model, this occasion both models presented a similar behaviour during the training stage, however, in the confusion matrices (Fig. 14), it can be observed that the accuracy of both models is very close to each other, but in both cases, it is below the 45 % previously obtained with the YOLO8-D8 model, which would mean an information loss during learning stage.

As YOLO8-S4 and YOLO8-S8 models managed to obtain 30 % and 27 % of accuracy, respectively, they are apparently less efficient, even YOLO8-S8 a little bit less.

Considering the results previously observed, the YOLO8-S4 model stood out slightly in confidence and accuracy terms, with respect to the curves analysed, however, a totally opposite performance was achieved when the validation of the models in RDDSM was carried out, being the YOLO8-S8 model which managed to obtain an average of 10 % higher detecting tumours, than the other model, reaching up to 70 % of detection, with respect to the 60 % achieved by the YOLO8-S4 model (Fig. 15).

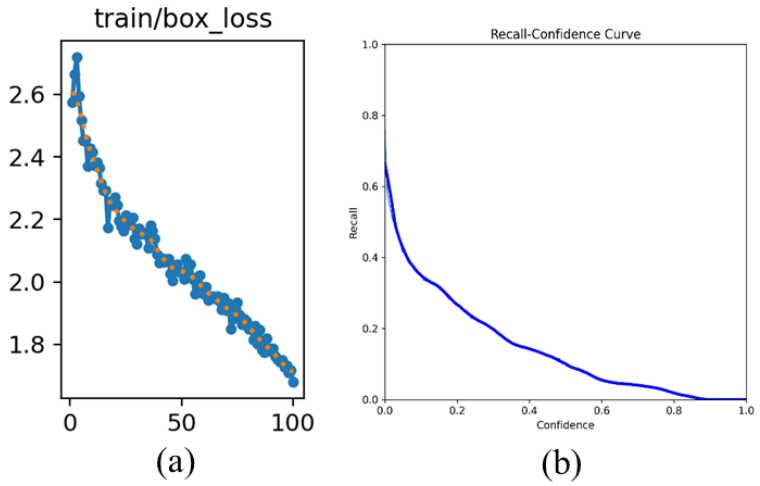

Fig. 17. YOLO8-S8 results trained on RSDDM with the enhancement filter applied where (a) represents the box loss metrics and orange dots shows the loss behaviour, and (b) represents the recall-confidence curve for the same model applied to RDDSM filtered.

3.3 CAD System Evaluation

Once previous processes were finalized, best weights were selected and used for further predictions in mini-MIAS database. YOLO8-S8 model was even higher than in the validation stage, as it achieved a score of 91 % on the mini-MIAS database, 23 % higher than the 68 % achieved by the YOLO8-S4 model (Fig. 16). Even though both models showed a lower accuracy than the previously YOLO7-S8 and YOLO8-D8 tested models, the YOLO8-S8 model, which was even less accurate, scored the highest of all models.

Finally, the comparison between the proposed models in segmentation mode is shown below (Table 3). Where TP, FN, TN, and FP, correspond to true positive, false negative, true negative and false positive evaluated cases, respectively. Moreover, St and AC corresponds to sensitivity and accuracy respectively.

As in the previous results, the YOLO8-S8 model showed a better performance during the 20 runs on mini-MIAS database, achieving a ST and AC of 95 % and 90 %, respectively. This difference can also see when individually evaluating TP, FN, TN, and FP, where YOLO8-S8 model had 18 hits in TP, one error in FN and one error in FN.

RDDSM brightness and contrast enhancement filter assessment. After testing the proposed models on the selected databases and comparing the results between each one, the model with the best performance during the training and evaluation stages was selected to be, subsequently, trained under the same conditions on the RDSSM, now with the brightness and contrast enhancement filter.

Initially, with YOLO8-S8, training stage was set at 100 epochs, with 0.5 confidence and batch size = 8. Each epoch had an approximately duration of 11 minutes. Once finished training stage, some troubles were found. In the first training results, it could be seen that box loss curve decreased steadily and with little dispersion, although in the

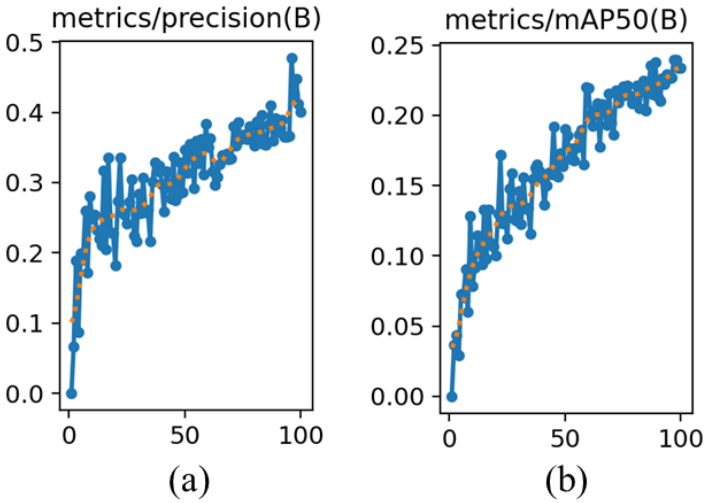

Fig. 18. (a) Precision graph and (b) mAP metrics for model applied to RDDSM filtered. Orange dots shows the increasing of accuracy foe every epoch.

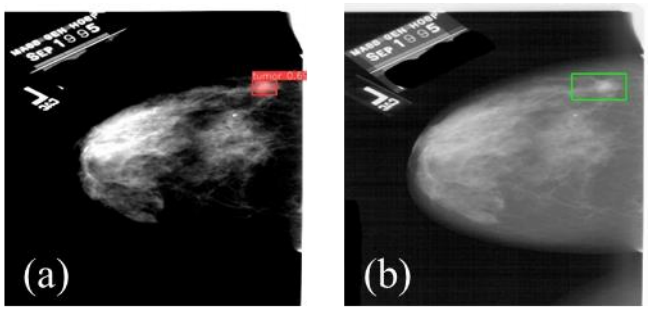

Fig. 19. Comparison results between (a) model validation in filtered mammogram from RSDDM with tumour detected inside rex bounding box with 69 %, and (b) original mammogram with tumour identified inside green bounding box.

initial periods, the decrease was slightly faster (Fig. 17-a). The confidence-recall curve (Fig. 17-b) shows similar confidence to the previous models (Fig. 13), although it is slightly higher, with a 1 % difference, reaching 66 % compared to 65 % for the previously tested models.

Subsequently, for the accuracy and mAP metrics, in both cases the initial accuracy was considered low, although in the early stages it increases rapidly. In the case of precision (Fig 18-a), there is a mostly significant dispersion before the first 50 stages, later it starts to decrease, but there is no stability when approaching the final epochs.

As for mAP, when approaching the final epochs, there seems to be a smaller dispersion, however, during the rest of the learning process, the dispersion is notorious, especially before the first 50 epochs (Fig. 18-b). With these results, it could be predicted that the model would behave similarly to that applied to the unfiltered RDDSM. The

Table 4. Metrics comparison between each model.

Model	Confusion Matrix	Box loss	Precision	mAP
YOLO7-S8	0.18	0.045	0.4	0.23
YOLO8-D8	0.45	1.4	0.48	0.48
YOLO8-S4	0.30	1.00	0.58	0.28
YOLO8-S8	0.27	1.4	0.49	0.29
YOLO8-S8F	0.24	1.4	0.4	0.24

result in the evaluation stage is shown below. Initially, system was not able to detect any tumours, but as more tests were performed on different mammograms, the system was able to identify some tumours (Fig. 19), reaching up to 69 % in detected mammograms, which could be a low accuracy compared to previous models, (Fig. 15). Even when model was tested on mini-MIAS database, this had pronounced difficulty in screening, achieving only 6 hits out of the 20 mammograms chosen for evaluation.

With the analysed results of the models applied to the raw RDDSM, the expectations increased for the evaluation of the model applied to the same database with the brightness and contrast enhancement filter, since in the images, the fibrous tissue was mostly visible, which would mean that the neural network would present fewer difficulties by not having to process irrelevant information such as areas of fatty tissue. However, the results showed that the model extracted fewer features, as in some cases, it failed to make a detection in the images, and in others, the percentage achieved was considered low.

Finally, in order to further show the performance comparison between the models, the metrics resulting from each training for each model are shown below. Where YOLO7-S8 is the YOLOv7 model in segmentation task with batch size = 8; YOLO8-D8 is YOLOv8 model in detection task with batch size = 8; YOLO8-S4 is YOLOv8 model in segmentation task with batch size = 4; YOLO8-S8 is YOLOv8 model in segmentation task with batch size = 8; and YOLO8-S8F is YOLOv8 model in segmentation task with batch size = 8 trained in filtered images.

4 Discussion

The performances obtained in each of the models during the training stages are found to be variable and, in some cases, quite scattered in terms of accuracy and learning rate. Although during the evaluation stage they showed a better performance, and even in the test phase where up to 91 % of assertiveness was achieved, this would imply that the model can be optimised by improving the images with pre-processing to avoid the overlapping of the breast tissue [49], as it was observed, the application of a brightness and contrast filter was counterproductive in terms of the observer and what was interpreted by the neural network.

Another method is by modifying the hyperparameters of the CNN and performing an in-depth analysis on the use of neurons within the network, in order to turn off those that are not being used and reduce the processing time during training and predictions.

Considering the RDDSM results, and analysing the learning curves of the models, it is possible to realise an increase in learning by augmenting data with which the network will be trained, as DL models need a considerable amount of data [49, 39] in order to be able to extract the most features through each iteration.

One of the differences most remarkable are between YOLO7-S8 and YOLO8-D8, where YOLO8-D8 model show relative better results than another model. This may due to YOLO topology, since in YOLOv8 CSPLayer was changed to C2f module, combining high-level features with contextual information to improve detection accuracy, moreover, the anchor-free used model which allows each branch to focus on its task improving model overall accuracy [52], and even YOLO does not require a complex pipeline once looked the image [49].

Instance segmentation showed better results than instance detection. Instance segmentation works by separating an example from belonging class by separating individually and comparing each labelled pixel values between segmented classes to non-segmented classes and it is useful where too many objects of same class are present and need to be differentiated. This probably affects the way model learn features, because object detection uses the pixel space given to a specific object.

Another drawback observed was the confusion of the model tested with the filtered database with the model tested with the raw images. This confusion could be due to the intensity of the brightness in the pixels related to the tumour, the dense breast tissue and the pectoral muscle [49], causing the comparison between pixel values to be depreciated as they could contain the same values and prevented the neural network from extracting any highly significant features.

5 Conclusion

In this paper, the evaluation of the Darknet-19 YOLO V8 model is presented, testing different possibilities that a CAD system can offer for mass detection. The selected model achieved up to 91 % of maximum assertiveness when tested on a publicly available database, however, it is necessary to test it on a properly constructed database in order to evaluate the model on more recent cases using more recent digital mammograms.

While it is considerable that the model needs to be trained on a much more robust database, what has been demonstrated so far provides a guideline for using CAD systems as medical assistants that can provide a second opinion, or function as medical decision support, and with the detection times determined, a decrease in workflow is foreseen, moreover, with the detection times determined, a reduction in the workflow is foreseen, as well as a reduction in the workload in hospitals whose staff capacity is affected and an improvement in the radiologist's performance by reducing visual fatigue, as a specific region is previously obtained that can be subjected to analysis, as a result of the prediction of the proposed system, avoiding the specialist in question from performing a complete reading of the mammography

6 Future Work

Future work will initially focus on data augmentation for better training of the proposed model, as well as adequate pre-processing of the images to deal with the problems of intensity and overlapping of the breast tissue. The development of a graphical interface for a doctor-computer interface is also foreseen, facilitating access to the diagnostic tool when performing mammography analysis in the medical imaging section.

7 Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgments. We are pleased to mention the Centro de Investigación e Innovación Tecnológica from Instituto Politécnico Nacional (CIITEC-IPN), where we have been able to make use of the facilities, allowing us to carry out this research. We would also like to thank the Consejo Nacional de Ciencia y Tecnología (CONACyT) for the financial support granted, making possible the development of this research and future work.

References

1. Organización Mundial de la Salud.: Cólera. Boletín epidemiológico, Sistema nacional de vigilancia epidemiológica, Sistema único de información, Dirección general de epidemiología, vol. 39, no. 41, pp. 3–9 (2022)
2. American Cancer Society: How common is breast cancer? Breast Cancer Statistics (2024) https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html
3. Instituto Mexicano del Seguro Social: Epidemiología del cáncer de mama. IMMS (2022) https://www.gob.mx/imss/articulos/epidemiologia-del-cancer-de-mama-318014
4. Global Cancer Observatory: Cancer today (2024) gco.iarc.fr/today/en
5. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249 (2021) doi: 10.3322/caac.21660
6. American Society of Clinical Oncology: Cáncer de mama: Introducción. American Society of Clinical Oncology Journals (2012) https://www.cancer.net/node/18093
7. Instituto Nacional del Cáncer: Prevención del cáncer de seno (mama) (PDQ®) – versión para pacientes (2013) cancer.gov/espanol/tipos/seno/paciente/prevencion-seno-pdq
8. Ting, F. F., Tan, Y. J., Sim, K. S.: Convolutional neural network improvement for breast cancer classification. Expert Systems with Applications, vol. 120, pp. 103–115 (2019) doi: 10.1016/j.eswa.2018.11.008
9. Fuchsjäger, M., Morris, E., Helbich, T.: Breast imaging: Diagnosis and intervention, Medical Radiology (2022) doi: 10.1007/978-3-030-94918-1
10. Rehman, K. U., Li, J., Pei, Y., Yasin, A., Ali, S., Mahmood, T.: Computer vision-based microcalcification detection in digital mammograms using fully connected depthwise separable convolutional neural network. Sensors, vol. 21, no. 14, pp. 4854 (2021) doi: 10.3390/s21144854

11. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., Aerts, H. J.: Artificial intelligence in radiology. Nature Reviews Cancer, vol. 18, no. 8, pp. 500–510 (2018)
12. Hamet, P., Tremblay, J.: Artificial intelligence in medicine. Metabolism, vol. 69, pp. S36–S40 (2017) doi: 10.1016/j.metabol.2017.01.011
13. Galveia, J. N., Travassos, A., Quadros, F. A., da-Silva-Cruz, L. A.: Computer aided diagnosis in ophthalmology: deep learning applications. Classification in BioApps: Automation of Decision Making, pp. 263–293 (2017) doi: 10.1007/978-3-319-65981-7_10
14. Cottrell, S. S.: A simple method for finding the scattering coefficients of quantum graphs. Journal of Mathematical Physics, vol. 56, no. 9 (2015) doi: 10.1063/1.4931082
15. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., Farhan, L.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, vol. 8, no. 1 (2021) doi: 10.1186/s40537-021-00444-8
16. Yamashita, R., Nishio, M., Gian-Do, R. K., Togashi, K.: Convolutional neural networks: an overview and application in radiology. Insights into Imaging, vol. 9, no. 4, pp. 611–629 (2018) doi: 10.1007/s13244-018-0639-9
17. Abdelrahman, L., Al-Ghamdi, M., Collado-Mesa, F., Abdel-Mottaleb, M.: Convolutional neural networks for breast cancer detection in mammography: A survey. Computers in biology and medicine, vol. 131, p.p. 104248 (2021) doi: 10.1016/j.compbiomed.2021.10424
18. Desai, M., Shah, M.: An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN). Clinical eHealth, vol. 4, pp. 1–11 (2021) doi: 10.1016/j.ceh.2020.11.002
19. Alanazi, S. A., Kamruzzaman, M. M., Islam-Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., Siddiqi, M. H.: Boosting breast cancer detection using convolutional neural network. Journal of Healthcare Engineering, vol. 2021, pp. 1–11 (2021) doi: 10.1155/2021/5528622
20. Houssami, N., Kirkpatrick-Jones, G., Noguchi, N., Lee, C. I.: Artificial intelligence (AI) for the early detection of breast cancer: A scoping review to assess AI's potential in breast screening practice. Expert Review of Medical Devices, vol. 16, no. 5, pp. 351–362 (2019) doi: 10.1080/17434440.2019.1610387
21. Amisha, Malik, P., Pathania, M., Kumar-Rathaur, V.: Overview of artificial intelligence in medicine. Journal of Family Medicine and Primary Care, vol. 8, no. 7, pp. 2328 (2019) doi: 10.4103/jfmpc.jfmpc_440_19
22. Balkenende, L., Teuwen, J., Mann, R. M.: Application of deep learning in breast cancer imaging. Seminars in Nuclear Medicine, vol. 52, no. 5, pp. 584–596 (2022) doi: 10.1053/j.semnuclmed.2022.02.003
23. Vállez, N., Bueno, G., Déniz, O., Dorado, J., Seoane, J. A., Pazos, A., Pastor, C.: Breast density classification to reduce false positives in cade systems. Computer Methods and Programs in Biomedicine, vol. 113, no. 2, pp. 569–584 (2014) doi: 10.1016/j.cmpb.2013.10.004
24. Sechopoulos, I., Teuwen, J., Mann, R.: Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. Seminars in Cancer Biology, vol. 72, pp. 214–225 (2021) doi: 10.1016/j.semcancer.2020.06.002
25. Melekoodappattu, J. G., Dhas, A. S., Kandathil, B. K., Adarsh, K. S.: Breast cancer detection in mammogram: Combining modified CNN and texture feature based approach. Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 9, pp. 11397–11406 (2022) doi: 10.1007/s12652-022-03713-3
26. Wang, Y., Yang, F., Zhang, J., Wang, H., Yue, X., Liu, S.: Application of artificial intelligence based on deep learning in breast cancer screening and imaging diagnosis. Neural Computing and Applications, vol. 33, no. 15, pp. 9637–9647 (2021) doi: 10.1007/s00521-021-05728-x

27. Ortiz-Rodriguez, J. M., Guerrero-Mendez, C., Martinez-Blanco, M. R., Castro-Tapia, S., Moreno-Lucio, M., Jaramillo-Martinez, R., Solis-Sanchez, L. O., Martínez-Fierro, M. L., Garza-Veloz, I., Moreira-Galvan, J. C., Barrios-Garcia, J. A.: Breast cancer detection by means of artificial neural networks. Advanced Applications for Artificial Neural Networks (2017) doi: 10.5772/intechopen.71256
28. Salama, W. M., Aly, M. H.: Deep learning in mammography images segmentation and classification: Automated CNN approach. Alexandria Engineering Journal, vol. 60, no. 5, pp. 4701–4709 (2021) doi: 10.1016/j.aej.2021.03.048
29. Muduli, D., Dash, R., Majhi, B.: Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network-based approach. Biomedical Signal Processing and Control, vol. 71, pp. 102825 (2022) doi: 10.1016/j.bspc.2021.102825
30. Agarwal, R., Diaz, O., Lladó, X., Yap, M. H., Martí, R.: Automatic mass detection in mammograms using deep convolutional neural networks. Journal of Medical Imaging, vol. 6, no. 03, pp. 1 (2019) doi: 10.1117/1.jmi.6.3.031409
31. Muralidhar, G. S., Haygood, T. M., Stephens, T. W., Whitman, G. J., Bovik, A. C., Markey, M. K.: Article commentary: Computer-aided detection of breast cancer — have all bases been covered? Breast Cancer: Basic and Clinical Research, vol. 2, pp. BCBCR.S785 (2008) doi: 10.4137/bcbcr.s785
32. Mert, A., Kılıç, N., Bilgili, E., Akan, A.: Breast cancer detection with reduced feature set. Computational and Mathematical Methods in Medicine, vol. 2015, pp. 1–11 (2015) doi: 10.1155/2015/265138
33. Warren-Burhenne, L. J., Wood, S. A., D'Orsi, C. J., Feig, S. A., Kopans, D. B., O'Shaughnessy, K. F., Sickles, E. A., Tabar, L., Vyborny, C. J., Castellino, R. A.: Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology, vol. 215, no. 2, pp. 554–562 (2000) doi: 10.1148/radiology.215.2.r00ma15554
34. Touami, R., Benamrane, N.: Microcalcification detection in mammograms using particle swarm optimization and probabilistic neural network. Computación y Sistemas, vol. 25, no. 2 (2021) doi: 10.13053/cys-25-2-3429
35. Taylor, P.: Computer aided detection. Breast Cancer Research, vol. 2, no. S2 (2000) doi: 10.1186/bcr227
36. Khoo, L. A. L., Taylor, P., Given-Wilson, R. M.: Computer-aided detection in the United Kingdom national breast screening programme: prospective study. Radiology, vol. 237, no. 2, pp. 444–449 (2005) doi: 10.1148/radiol.2372041362
37. Aslam, M. A., Aslam, Cui, D.: Breast cancer classification using deep convolutional neural network. Journal of Physics: Conference Series, vol. 1584, no. 1, pp. 012005 (2020) doi: 10.1088/1742-6596/1584/1/012005
38. Al-antari, M. A., Al-masni, M. A., Choi, M. T., Han, S. M., Kim, T. S.: A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. International Journal of Medical Informatics, vol. 117, pp. 44–54 (2018) doi: 10.1016/j.ijmedinf.2018.06.003
39. Ketkar, N., Moolayil, J.: Convolutional neural networks. Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch, pp. 197–242 (2021) doi: 10.1007/978-1-4842-5364-9_6
40. Teoh, T. T., Rong, Z.: Convolutional neural networks. Artificial Intelligence with Python, pp. 261–275 (2022) doi: 10.1007/978-981-16-8615-3_16
41. Manaswi, N. K.: Convolutional neural networks. Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition with TensorFlow and Keras, pp. 91–96 (2018) doi: 10.1007/978-1-4842-3516-4_6
42. Skansi, S.: Convolutional neural networks. Introduction to Deep Learning: from logical calculus to artificial intelligence, pp. 121–133 (2018) doi: 10.1007/978-3-319-73004-2_6
43. Prakash-Kolla, B., Kanagachidambaresan, G. R.: Programming with TensorFlow: Solution for Edge Computing Applications pp. 45–51 (2021) doi: 10.1007/978-3-030-57077-4

44. Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., Munishkumaran, S.: Current status of the digital database for screening mammography. Digital Mammography: Nijmegen, pp. 457–460 (1998) doi: 10.1007/978-94-011-5318-8_75
45. Aibar, L., Santalla, A., Criado, M. L., González–Pérez, I., Calderón, M., Gallo, J., Parra, J. F.: Clasificación radiológica y manejo de las lesiones mamarias. Clínica e Investigación en Ginecología y Obstetricia, vol. 38, no. 4, pp. 141–149 (2011) doi: 10.1016/j.gine.2010.10.016
46. Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., Acharya, U. R.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, vol. 121, pp. 103792 (2020) doi: 10.1016/j.compbiomed.2020.103792
47. Redmon, J., Farhadi, A.: YOLO9000: Better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)
48. Al-masni, M. A., Al-antari, M. A., Park, J., Gi, G., Kim, T., Rivera, P., Valarezo, E., Choi, M., Han, S., Kim, T.: Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. Computer Methods and Programs in Biomedicine, vol. 157, pp. 85–94 (2018) doi: 10.1016/j.cmpb.2018.01.017
49. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
50. Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B.: A review of yolo algorithm developments. Procedia Computer Science, vol. 199, pp. 1066–1073 (2022) doi: 10.1016/j.procs.2022.01.135
51. Terven, J., Córdova-Esparza, D., Romero-González, J.: A comprehensive review of yolo architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, vol. 5, no. 4, pp. 1680–1716 (2023) doi: 10.3390/make5040083
52. Li, B., Shi, Y., Qi, Z., Chen, Z.: A survey on semantic segmentation. In: IEEE International Conference on Data Mining Workshops (2018) doi: 10.1109/icdmw.2018.00176