Breast Cancer Detection & Diagnosis in Full Field Digital Mammography with Deep Learning Approach

MCS

Author

Zeeshan Ahmed

Registration Number
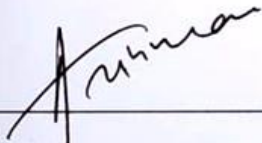
00000398016

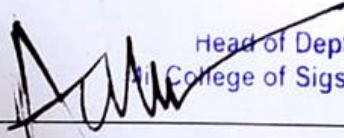Supervisor

Assoc Prof Dr. Ata Ur Rehman

A thesis submitted to the faculty of Software Engineering Department, Military College of Signals, National University of Sciences and Technology, Rawalpindi in partial fulfilment of the requirements for the degree of MS in Software Engineering
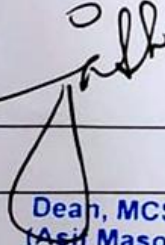
(September 2023)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Zeeshan Ahmed,** Registration No. **00000398016** of **Military College of Signals (MCS)** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: _____

Name of Supervisor: **Assoc Prof Dr. Ata Ur Rehman**

Date: _____

Signature (HoD): _____

Brig
Head of Dept of CSE
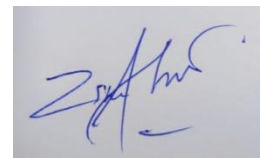College of Sigs (NUST)

Date: 2-1/11/23

Signature (Dean/Principal): _____

Date: 3/11/23

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

i

Declaration

I, Zeeshan Ahmed declare that this thesis titled " Breast Cancer Detection & Diagnosis in Full Field Digital Mammography with Deep Learning Approach" and the work presented in it are my own and has been generated by me as a result of my own original research.

 I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated

3. Where I have consulted the published work of others, this is always clearly attributed

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work

5. I have acknowledged all main sources of help

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Zeeshan Ahmed

00000398016 MSSE 28

Dedication

"In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to my parents, teachers and colleagues who supported me at each step of

the way.

Abstract

Breast cancer is a significant global health concern and its early detection is critical for improving the treatment outcomes. Full-field digital mammography (FFDM) has emerged as a valuable screening tool for breast cancer, with deep learning techniques offering promising avenues for enhanced detection and diagnosis. Although, research has focused on developing deep learning models for breast cancer screening or detection on a confined scope of lesions (primarily mass & calcification), there existed a significant gap in the literature regarding the detection of multiple breast cancer lesions or abnormalities. Addressing this gap, our research introduces an innovative methodology utilizing YOLOv8 deep object detector for the detection of six different types of breast cancer lesion types: Mass, Architectural Distortion, Asymmetry, Focal Asymmetry, Suspicious Calcification and Suspicious Lymph Node. We use Vindr-Mammo dataset in our research which provides an opportunity to work upon a broad spectrum of breast cancer lesions. We also employ a novel data augmentation approach of generating artifacts with synthetic lesions to enhance the sample space. Our model demonstrated 89.3% accuracy, 0.92 F1-score and 0.72 mAP. The proposed model is a pioneer effort that effectively and consistently detects a diverse spectrum of breast cancer lesions attesting its reliability in multi-lesion breast cancer detection tasks.

Acknowledgments

All praises to Allah for the strengths and His blessing in completing this thesis.

I would like to convey my gratitude to my supervisor Assoc Prof Dr. Ata Ur Rehman for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experiment and thesis works are major contributions to the success of this research. Also, I would thank my committee members; Asst Prof Dr. Muhammad Salman Khan, Prof Dr. Hammad Afzal and Assoc Prof Dr. Ihtesham Ul Islam for their support and guidance during the research.

Last, but not the least, I am highly thankful to my parents and my colleagues. They have always stood by my side and supported me. I would like to thank them for all their care, love and support through my times of stress and excitement.

Table of Contents

1. Introduction

1.1   Overview

Breast cancer characterized as an uncontrolled proliferation of cells within a specific region of the breast [1] is one of the most wide-spread cancer type, posing a significant global healthcare challenge with an estimated 2.2 million new cases every year [2]. As indicated by the projections, this trajectory is likely to escalate, with an estimated 19.3 million new cancer cases annually by 2025 [3]. Breast cancer detection at early stage greatly enhances the chances of successful treatment. Biennial screening has been observed to reduce breast cancer mortality rate by 30% [4].

Early and accurate detection of breast cancer is paramount for effective treatment and improved patient outcomes. Among standard imaging modalities employed for breast cancer detection, X-Ray (mammography), Ultrasound (US), Digital Breast Tomosynthesis (DBT) and Magnetic Resonance Imaging (MRI) are most prominent. Among these standards, mammography has emerged as the fundamental diagnostic tool for breast cancer screening due to its ability to detect abnormalities at an early stage and cheaper rates. A typical mammography exam is performed on the breast compressed between two plates and two views of each breast are taken; bilateral Craniocaudal (CC) and Mediolateral oblique (MLO).

Mammograms can be captured using different imaging techniques: Screen Film Mammography (SFM) and Full-Field Digital Mammography (FFDM). SFM is an old technique which involves the use of X-ray films to capture breast images with its limitations such as failure to visualize 10% to 20% breast cancers cases, suffer less contrast in images and inability to be digitalized. Whereas, FFDM employs digital detectors for image acquisition producing high-resolution X-ray images of the breast tissue, facilitating the identification of abnormalities indicative of breast cancer. FFDM offers distinct advantages over SFM, including improved image quality, faster acquisition and the ability to manipulate and enhance images for better visualization.

 Incorporation of deep learning methods over the last few years, have provided remarkable avenues to augment the accuracy and efficiency of breast cancer diagnosis, particularly when applied to mammograms. Despite the consistent advancements in the field, a noteworthy gap persists – the

comprehensive detection and classification of multiple lesion types within the mammography datasets. While previous research has made strides in detection and classification of limited lesions types primarily focusing on Mass and Calcification or their screening as benign or malignant, a broader spectrum of lesion types commonly encountered in clinical examinations has remained un-explored. To overcome this challenge, we used VinDr-Mammo [5], a large scale FFDM dataset consisting of four view exams with twenty thousand mammography images in DICOM format with precise annotations and breast level findings featuring multiple types of lesion or abnormalities.

1.2   Problem Statement

"Breast cancer is a major health concern globally, and early detection of the disease is critical for improving treatment outcomes. Full-field digital mammography (FFDM) has emerged as a valuable screening tool for breast cancer, with deep learning techniques offering promising avenues for enhanced detection and diagnosis. Although, research has focused on developing deep learning models for breast cancer detection, there remains a significant gap in the literature regarding the detection of breast lesion types or abnormalities. This lack of focus on breast lesion detection represents a significant challenge to the development of effective and accurate deep learning methods. As such, there is a pressing need for further research that addresses this gap to develop a robust, accurate and reliable deep learning model capable of accurately detecting different categories of breast cancer lesions or abnormalities."

1.3   Objectives

Main objective of this research are listed as follows:

- Deep Learning based Breast Cancer detection & diagnosis.
- To work upon a large-scale FFDM dataset.
- Pre-processing & analyzing the dataset.
- Development of a deep learning model to detect & classify multiple categories of the breast cancer abnormalities.

- Enhancement of dataset for finest results.
- Train and test the proposed model on training data without being exposed to the testing data and evaluate the test and validation data after successful training.
- Comparison with other deep learning models for finest accuracy/ F1 scores.
- Highlight the potential research areas requiring further probe on the subject.

1.4   Thesis Contribution

The aim of this research is to comprehensively detect and diagnose diverse lesion types, collectively representing a holistic range of abnormalities encountered in breast cancer pathology. For a meticulous integration of advanced deep learning technique YOLOv8 medium-sized model acclaimed for its proficiency in object detection, is proposed to determine the intricate lesion categories within FFDM images sourced from the Vindr-Mammo dataset. The main contributions of this study are as follows:

- **Comprehensive Breast Cancer Lesion Detection:** We propose a YOLOv8 model capable of accurately detecting six distinct lesion types in FFDM images i.e. Mass, Architectural Distortion, Asymmetry, Focal Asymmetry, Suspicious Calcification, and Suspicious Lymph Node. The proposed model is a pioneer effort which comprehensively detects a wide spectrum of abnormalities encountered in breast cancer pathology.

- **Breast Cancer Screening:** Proposed model not only identifies individual lesions but also performs breast cancer screening by determining the presence or absence of cancerous findings within FFDM images. This dual capability enhances the clinical utility of our model.

- **Performance Evaluation:** We perform a comprehensive performance evaluation of the proposed model with relevant deep learning models from existing literature to assess the effectiveness and superiority of our model in lesion detection and breast cancer screening tasks.

1.5   Thesis Organization

The thesis is structured as follows:

- Chapter 2 contains the literature reviewed in the thesis. The studies related to breast cancer detection and diagnosis with their proposed models and performance evaluations along with discussion on their short-comings and the existing gap in literature review.

- Chapter 3 contains the research methodology with discussion on the publicly available dataset and their limitations. The details of the VinDr-Mammo as selected dataset for the study and the pre-processing and data augmentations applied to it are also mentioned. Finally, the YOLOv8 model architecture and the training process are described in detail. This chapter also describes the experimental configuration of the hardware and software utilized for the experiment.

- Chapter 4 covers outcome of experimental results and performance evaluation. The detailed discussion with assessment of the results with respect to the related studies in the existing research arena is also provided.

- Chapter 5 marks the end of the document. The future work and conclusion are mentioned in this chapter.

2. Literature Review

2.1   Related works

The integration of Artificial Intelligence (AI) has not only reshaped medical diagnostics but also become an indispensable component of clinical decision-making, transforming the domain of Computer-Aided Detection (CAD) of breast cancer rapidly. A number of CAD systems have been developed to screen or detect breast cancers at earlier stages using deep learning driven techniques.

In a recent study [6], the authors developed a breast mass detection system using YOLOv7 and YOLOv8 deep object detectors. The study involved utilizing 1029 mammogram images from the VinDr-Mammo dataset for their experiment. Pre-processing included application of various image enhancement techniques like Contrast-Limited Adaptive Histogram Equalization (CLAHE), median filter and bilateral filter on same dataset separately each time and before training for drawing comparison of model performance on each technique. They achieved a mean average precision (mAP) of 0.65 with YOLOv8 trained on data preprocessed with median filter, outperforming YOLOv7 in detecting masses on mammograms.

In a complementary study [7], researchers focused on enhancing the detection rate of breast masses within mammograms by developing a technique called spatial based breast density enhancement for mass detection. This attempt was aimed to improve the detection rate of breast masses surrounded by tissues of varying density levels. This technique involved employing a modified YOLOv3 model for segmentation task. By optimizing exposure thresholds and intensity factors based on breast density, their study displayed 17.24% increased mAP with an overall mass segmentation accuracy of 94.41% and 96% accuracy in classifying benign and malignant masses. However, this research was conducted on only 112 images containing mass lesions from INbreast dataset, which exhibits extremely limited sample space and raises concerns to drive any reasonable conclusions.

In another notable research [8], the authors employ YOLOv3 model to tackle the multifaceted task of detecting and classifying suspicious masses and calcification lesions within mammogram

images. They used 2907 mammograms from CBIS-DDSM, 235 mammograms from the INbreast datasets and 487 mammograms from a private dataset in their experiment. To enhance precision and classification accuracy, a fusion approach was also recommended which suggests combining predictions on basis of confidence scores of the multiple Yolo based models which have been configured and trained differently to lower the error rate and enhance performance. They achieved an accuracy of 95.7%, 98.1%, and 98% for mass lesions and 74.4%, 71.8%, and 73.2% for calcification lesions on CBIS-DDSM, INbreast and private dataset, respectively.

Furthermore, the automated data-driven model proposed in a recent study [9] showcased the potential of transfer learning for breast cancer detection in mammograms. Leveraging CBIS-DDSM and INbreast datasets as sources to implement the transfer learning technique on a small-scale private FFDM dataset (190 mass images, 46 asymmetry images and 71 distortion images), the researchers experimented with various YOLO models including YOLOv3, YOLOv5, and YOLOv5. In addition, they generated saliency or heat maps using Eigen-CAM for model introspection and highlighting all the suspicious regions in the images. This technique was observed to affect a substantial reduce in false negatives but with an increase in false positives, but still effective to compare with the results obtained from YOLO based models. YOLOv5s (small) was finally found to be the most optimal model by achieving 0.49, 0.83 and 0.62 mAP values for DDSM, INbreast and private dataset respectively.

In another study [10], the authors used an innovative approach of YOLO-based models fusion with image-to-image translation for early breast cancer prediction. Their work encompasses two facets: the detection of breast cancer lesions with YOLO-based models fusion and the application of the same model on synthetic mammograms generated through image-to-image translation models: CycleGAN and Pix2Pix as the basis for early breast cancer prediction. The study thus showcases the potential of retrospective analysis for early detection and prediction of breast cancer. The experiment was performed on a private dataset with 833 mammogram images including three lesion types; Mass, Calcification and Architectural Distortion. The study utilized two mammograms of the same subject on temporal basis: current and prior. Their classification of breast lesions for current mammograms of the subjects achieved an accuracy rate of 93% for Mass lesions, 88% for Calcification lesions, and 95% for Architectural Distortion lesions. Accuracy rate for prior mammograms was achieved as 36% for Mass lesions, 14% for

Calcification lesions, and 50% for Architectural Distortion lesions. Normal mammograms were classified with an accuracy rate of 92% and 90% respectively on Current and Prior exams.

In another significant study [11], the researchers introduced a modified YOLOv5 model designed specifically for breast tumor detection and classification task. The preprocessing included removal of white borders, pectoral muscles and labels. Thereafter, CLAHE was applied for better contrast and erosion operation was performed for removal of tumor surrounding tissues to make the tumor part more prominent. They used CBIS-DDSM dataset divided into 60% for training, 30% for validation, and 10% for testing. Their proposed model out-performed YOLOv3 and faster RCNN, attaining 96.50% accuracy and 0.96 mAP. This study was mainly focused on improving the false positive rate (FPR) and false negative rate based limitations of previous studies.

Meanwhile, in a parallel investigation [12], the authors explored YOLO based deep learning approach for CAD and classification of breast lesions as benign or malignant. Their methodology encompassed a YOLO based detector for detection of breast cancer lesion from the mammogram images. Thereafter, engaging a standard feed-forward CNN architecture employed with ResNet-50 and InceptionResNet-V2 with subtle modifications for the classification tasks. They utilized DDSM and INbreast datasets for their experiment and achieved respective accuracies of 97.50% and 95.32%.

Another significant avenue of research lies in the classification of breast density and BIRAD score as benign or malignant, as demonstrated by a recent study [13]. This involved the selection of pre-trained CNN architectures including VGG, Resnet, Densenet, InceptionNet, and EfficientNet for feature extraction. This approach is based on four different views of mammogram images which are left CC, left MLO, right CC and right MLO which are fed into four different models each time with a slight difference in architectural organization of end layers used for concatenation and prediction tasks. They proposed View-specific Feature-Level Evidential Fusion (VS-LEF) model which extracted features from each mammogram as per left or right view separately. The subjective masses of each view were combined after processing through separate evidential layers. This was followed by combining the view-specific subjective masses at the last layer before prediction. The combination process was done using the Dempster's combination rule. They study employed VinDr-Mammo and mini-DDSM (a smaller

variant of DDSM) datasets for their experiment and achieved respective accuracies of 86% and 90% for breast density assessment and 97% and 87% for BIRADS score.

In a different study [14], the authors focused on detecting breast masses employing the Faster Region Convolutional Neural Network (RCNN). They worked on a privately acquired subset of the OPTIMAM mammography database (OMI-DB) containing approximately 80,000 FFDM images including 2145 cases depicting cancer. Notably, the OMI-H and OMI-G datasets nested within OMI-DB contain 2042 and 103 cancer images respectively. Several breast abnormalities such as masses, calcifications, architectural distortions, focal asymmetries or combinations of the above are contained within this dataset but the researchers only utilized those with mass lesions for their experiment. They implemented RCNN for detection of breast masses on OMI-H, and used transfer learning on OMI-G and INbreast datasets for the same. They obtained 0.93 recall on OMI-H dataset, 0.91 recall on OMI-G dataset and 0.99 recall for malignant and 0.85 recall for benign masses on the INbreast dataset.

Deep Convolutional Neural Networks (DCNNs) have played a pivotal role in detection of breast tumors in the field of mammography. In a different study [15], DCNNs such as Inception V4, ResNet-164, VGG-11, and DenseNet121 were used as base classifiers integrated into a fuzzy ensemble modeling technique. The decision scores from multiple models were adaptively combined through a weighted average function making an ensemble model to predict the image as benign or malignant. They used 1145 mammography images from Breast Cancer Digital Repository (not publicly accessible), MIAS, INbreast and DDSM datasets with normal, benign and malignant types. Whereas, the type of lesions or abnormalities worked upon were not specified in their research. This classifier fusion strategy proved to be beneficial in boosting the predicted accuracy of different approaches by achieving an exceptional accuracy of 99.32%.

In a similar study [16], the potential of feature learning techniques coupled with classification tasks was explored for identification of suspicious regions in mammograms as benign or malignant. Their dataset was comprised of 902 images collected from MIAS and a private dataset. This dataset was artificially inflated to a total number of 5568 images by employing methods such as sharpening, embossing, gaussian blur, rotation, edges detection, tilting and flipping as data augmentation. Thereafter a DCNN was developed utilizing transfer learning

from pre-trained weights from VGGNet-19, InceptionV3, Resnet152V2, Inception ResnetV2, EfficientNetB5 models. Their proposed model achieved 98% accuracy outperforming aforementioned models.

In another research [17], the researchers developed a customized CNN to conduct the classification of mammogram images with breast masses as benign, malignant or normal. Their proposed model consists of eight convolutionary, four max-pooling, and two fully connected layers. Three datasets were utilized for their rsearch: MIAS, DDSM, and a self-collected dataset. Their proposed model only accepts images of images of size 256×256 which is too small and a lot of loss in terms of appearance of cancer lesions like micro calcifications or asymmetries or distortions can be affected. This directly raises concerns on the clinical utility of their model in term of reliability. However, they reported achieving respective accuracies of 92.54%, 96.47%, and 95% with AUC scores of 0.85, 0.96, and 0.94 using MIAS, DDSM and private dataset.

Working on another avenue [18], the researchers introduced an approach to diagnose breast cancer with Autoencoder Generative Adversarial Network (AGAN) for data augmentation and a Convolutional Neural Network (CNN) for classification of the tumor as normal or abnormal. The researchers employed AGANs to augment images by generating additional representations of the mammogram images. The images generated by AGAN are than appended to the original set of mammograms and passed to the convolutional neural network, being the final classifier. They worked on DDSM dataset and achieved an accuracy of 89.17%.

Lastly, exploring the efficacy of class activation map methods in weakly supervised learning, this study [19] worked on detection of masses in mammography images. They investigated CAM, GradCAM, GradCAM++, XGrad-CAM, and LayerCAM by employing them within a weakly supervised learning framework. The study highlighted interplay between activation maps and detection outcomes leading to enhancement in detection rates based on the strategy of incorporating activation maps during training and testing phases. They used mammography images with mass lesion from VinDr-Mammo dataset and achieved an accuracy of 80.12% with an AUC of 0.87.

2.2   Discussion

These studies collectively highlight that research focus in the field of breast cancer detection has primarily centered on the development of deep learning models, specifically targeting a limited range of lesion types, primarily masses and calcifications. There have been isolated instances where a handful of researchers working on detecting of lesions on a broader spectrum. Such endeavors, to some extent have also been constrained by the availability of small datasets. Nevertheless, there still existed an opportunity to work on the detection of four variations of lesions with datasets like MIAS and INbreast.

However, the attention of the researchers has remained on the diagnostic aspect of these datasets with an emphasis on the classification with the overall outcomes as benign, malignant/ normal or abnormal. Unfortunately, the research community overlooked the intricate challenge of identifying distinct lesions or abnormalities, not fully capitalizing on this prospect.

It is worth noting that MIAS relies on an outdated SFM-based technology, which further underscores the underutilization of available resources. By leveraging a rich diversity of lesion types, a challenge existed thereby to pioneer development of a more refined breast cancer detection model, significantly contributing to the advancement of CAD systems.

However, it is imperative to acknowledge a critical limitation in some of these research works: the utilization of small sized mammogram images (resized to 256×256 etc). The use of such small image sizes raises concerns about the reliability and clinical utility of their models. In particular, it may result in a significant loss in the appearance of cancer lesions, making it extremely challenging to work upon subtle abnormalities like micro-calcifications, asymmetries or distortions correctly. This limitation highlights the need for larger and higher-resolution datasets to ensure that the developed models can reliably identify a broader spectrum of breast cancer lesions, ultimately improving their clinical applicability and diagnostic accuracy.

<div align="right">

Chapter 3

</div>

3. Research Methodology

3.1   Understanding Breast Cancer Lesions

This research is aimed at comprehensive detection and diagnosis of diverse lesion types or abnormalities generally encountered in breast cancer pathology. Understanding the characteristics of these different lesion types is fundamental for proceeding further towards our research objectives. Each lesion type possesses unique attributes that influence its appearance in mammograms and consequently, the effectiveness of detection algorithms. By defining these characteristics, we establish a foundation for accurate lesion identification and classification which enables us to develop a robust and comprehensive multi-lesion breast cancer detection system.

"Mass" has a three dimensional structure that occupies space within the breast tissues. It has completely or partially convex-outward borders [5]. The characteristics of a mass include having borders that are either completely or partially convex-outward. It appears denser in the center as compared to the outer areas if it is radio-dense.

When a potential mass is visible on only one mammographic projection, it's referred to as an "Asymmetry" until further mammography confirms its three-dimensional nature. They are characterized by the super-imposition of normal breast tissues.

"Focal Asymmetry" refers to an area of asymmetry in breast tissue with increased density that is not present on the opposite side of the breast. It raises concerns and requires further assessment to determine its origin. Whereas, "Global Asymmetry" represents a large amount of dense tissue spread over a substantial portion of the breast (at least one quadrant) [5]. It is reviewed relative to the corresponding area in the contralateral breast. There is no mass, architecture distortion or presence of calcifications associated with it.

 "Suspicious Calcification" refers to the presence of calcium deposits within the breast tissues. They are visible on mammograms as high intensity patches which can vary in appearance. Their specific patterns raise concern for potential malignancy: Amorphous means irregular calcifications, Coarse Heterogeneous translates to large and unevenly distributed calcifications,

Fine Pleiomorphic are small shaped variable sized calcifications, Fine Linear are thin linear calcifications and Fine Linear Branching means thin linear calcifications that display branches [5].

 "Architectural distortion" means usual structure of subject breast tissue has deformed with no observable mass [5]. They account for 12 to 45% of breast cancers missed during the screening process [38]. They may represent different development from benign to high-risk abnormalities.

"Suspicious Lymph Node" or Axillary lymph node receives lymph from vessels that drain the arm, walls of the thorax, breast and the upper walls of the abdomen. They have many features [5] which include loss or disruption of central fatty hilum which is a lighter central area seen in mammography. Loss of pericapsular fat line which is a thin layer of fat around the lymph node and loss of this fat line is considered suspicious. Irregular outer margins are lymph nodes with irregular, indistinct or ill-defined outer margins. Hyperattenuating refer to lymph nodes that appear denser or more opaque on imaging than usually expected. Calcified indicates the presence of calcifications within a lymph node making it suspicious.

"Skin Thickening" is defined as thickness becoming more than 2 mm which may be localized or spread out [5]. "Skin Retraction" is skin pulled in abnormally whereas, "Nipple Retraction" is nipple pulled inwards [5].

3.2   Public Datasets

The field of mammography interpretation has attracted significant attention by the research community; however, there still exists a paucity of publicly available datasets. Prominent among the commonly used datasets are: Mammographic Image Analysis Society (MIAS) [20], Digital Database for Screening Mammography (DDSM) [21], INbreast [22], and Chinese Mammography Database (CMMD) [23] datasets. Though most of these datasets include annotations or Regions of Interest (ROI) of breast abnormalities, their limited sample size and type of lesions/ abnormalities within the exams may not optimally harness the potential of contemporary deep learning networks.

DDSM in particular, is most frequently used for deep learning studies due to its substantial volume of examinations. However, it is worth noting that DDSM is a digitalized rendition of SFM which inherently suffers from loss of data in the process. It is pertinent to highlight that the

image acquisition mode preferred for Clinical Decision Support (CDS) tools in clinical practice predominantly aligns with FFDM [5].

Table 1: Publicly available Mammography Datasets.

	Name	Year	Type	Images	Annotation	Lesion Types
1	MIAS [20]	1994	SFM	322	Yes	Mass, Calcification, Asymmetry and Architectural Distortions
2	DDSM [21]	1996	SFM	10,480	Yes	Mass & Calcification
3	INBreast [22]	2012	FFDM	410	Yes	Mass, Calcification, Asymmetry and Architectural Distortions
4	CMMD [23]	2016	FFDM	3728	No	Mass & Calcification
5	VinDr-Mammo [5]	2022	FFDM	20,000	Yes	Mass, Calcification, Asymmetry, Focal Asymmetry, Global Asymmetry, Architectural Distortion, Suspicious Lymph Node, Nipple Retraction, Skin Thickening and Skin Retraction

3.3 VinDr-Mammo Dataset

In order to undertake the challenge of working on a broader spectrum of lesion types, VinDr-Mammo dataset [5] was employed. This dataset contains 5000 examinations, each consisting of four associated mammography images, featuring two views for each breast: Mediolateral-Oblique (MLO) & Craniocaudal (CC). Total 20,000 FFDM mammography images in DICOM format were collected from mammography exams conducted during 2018 to 2020 with the Picture Archiving and Communication System (PACS) from two hospitals "Hospital 108" and

"Hanoi Medical University Hospital" in Vietnam. The dataset presents a diverse spectrum of lesions: Mass, Architectural Distortion, Asymmetry, Global Asymmetry, Focal Asymmetry, Suspicious Calcification, Skin Thickening, Nipple Retraction, Skin Retraction, Suspicious Lymph Node.

Table 2: Distributions of various lesions in Vindr-Mammo dataset.

Lesion Type	Objects	Distribution
Mass	1226	50.92%
Suspicious Calcification	543	22.56%
Asymmetry	97	4.03%
Focal Asymmetry	269	11.18%
Global Asymmetry	26	1.08%
Architectural Distortion	119	4.94%
Skin Thickening	57	2.37%
Skin Retraction	18	0.75%
Nipple Retraction	37	1.54%
Suspicious Lymph Node	57	2.37%

Breast level findings/ annotations of the lesions have been documented in two CSV files ("breast_level_annotations.csv and finding_annotations.csv). The images are pre-arranged into subfolders as per the respective study identifiers, each of which contains four distinct images corresponding to 4 different views of the breast. The names of sub-folder and images are reserved after the respective study and image identifiesr. The information of the annotations is provided for each image whereas redundancy also exists due to presence of two or more lesions in the same image.

3.4   Excluding Lesions with Inadequate Sample Space

Some abnormalities in VinDr-Mammo dataset [5] such as Skin Retraction, Global Asymmetry and Nipple Retraction possess extremely limited sample space (<40), raising concerns regarding the reliability of any results drawn from them. Consequently, the decision was taken to exclude them from the study. Skin Thickening has fifty seven samples but it shares objects with Skin Retraction and Nipple Retraction in some images which when excluded reduces its sample size to forty four images, hence qualifying it for exclusion as well. Furthermore, the images marked by extremely small bounding box annotations, rendering them ineffective to the learning process, were also deemed unsuitable for inclusion in the study. These images were generally associated with Calcifications, however a few also formed part of Architectural Distortions.
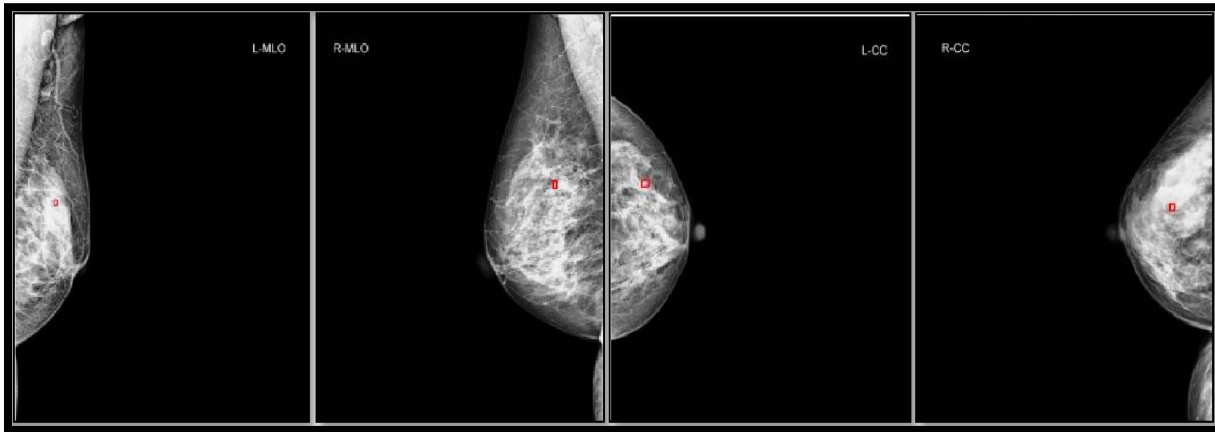

Figure 1: Examples of four different views of mammogram images from VinDr-Mammo dataset marked by extremely small bounding box annotations.

3.5   Splitting Dataset into Train, Test & Validation Sets

Finally, the data was partitioned randomly into distinct subsets of training, test, and validation sets. To thoroughly assess the models performance and its capacity to generalize, the validation and test sets consist of 10% random images with an inter-mix of multiple lesion types including 30% non-cancer images as backgrounds in validation and test sets each. Table 3 shows the dataset distribution of 80:10:10 split ratio for training, validation and testing sets respectively.

Table 3: Distribution of train, test and validation sets used in the study.

Dataset	Images	Backgrounds	Ratio
Train	1827	400	80%
Validation	206	63	10%
Test	206	63	10%
Total	2239	526	100%

3.6   Data Pre-Processing

To improve the quality of images preprocessing of medical images becomes of prime importance. Figure 2 illustrates the data pre-processing pipeline used in this experiment, which consists of several key steps to prepare the DICOM images for lesion detection and analysis as follows:

- DICOM to PNG Conversion

- YOLO Label Generation

- Brightness & Contrast Adjustment

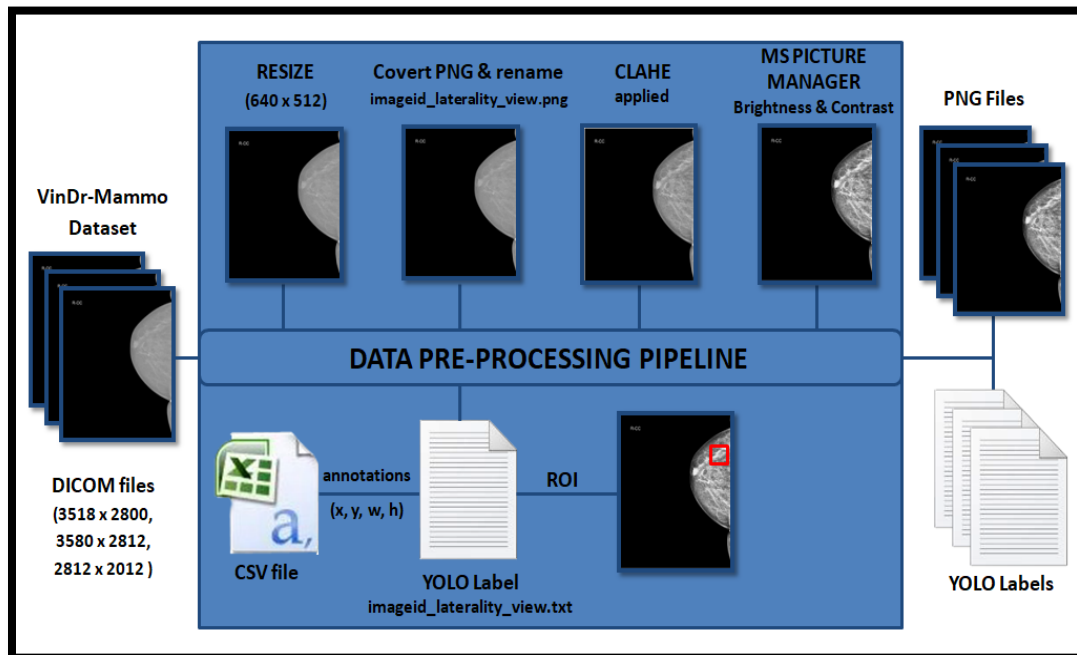- Data Augmentation: Synthetic Lesion Generation

Figure 2: Data Pre-Processing Pipeline.

3.6.1 DICOM to PNG Conversion

While working with medical images, a crucial step involves converting DICOM images into the more widely accessible PNG format. DICOM images in VinDr-Mammo dataset [5] consist of three different sizes: 3518 x 2800, 3580 x 2812 and 2812 x 2012. They were subsequently resized to 640 x 512 and converted in the PNG format following suitable intensity scaling. All the images used in this experiment were renamed to a standard convention as "imageIdentifier_laterality_view.png" to include more information.

3.6.2 YOLO Label Generation

Annotations associated with medical images are crucial for supervised learning tasks. The YOLO label generation method was employed to create bounding box labels for the lesions and relevant anatomical structures. The given annotations in the aforementioned CSV file against each image identifier were read and converted in the YOLO format. This format creates a bounding box for each ROI. Every bounding box is defined by its "x, y, w, h" values after its class number, where "x" and "y" represent the center coordinates with "w" as width and "h" as height, all relative to the concerned image. The naming convention of YOLO standard was followed and all labels were named after their respective images. Empty labels were also created for the non-cancer images designated as backgrounds.

3.6.3 Brightness & Contrast Adjustment

To address the challenge of inadequate brightness and contrast within medical images, we developed a customized image preprocessing pipeline: the integration of Contrast Limited Adaptive Histogram Equalization (CLAHE) and Microsoft Picture Manager brightness and contrast adjustment provided consistent and uniform image enhancement across the dataset, ensuring an optimal level of visual clarity for crucial details.

The application of CLAHE begins by accentuating local contrast variations, effectively enhancing the quality of intricate details present within the images. This pivotal enhancement

proves particularly beneficial in instances where minute details might otherwise remain indistinct. To further improve the image quality Microsoft Picture Manager was utilized to operate on an image-by-image basis. This effected meticulous adjustments to both brightness and contrast further enhancing precise breast level details.
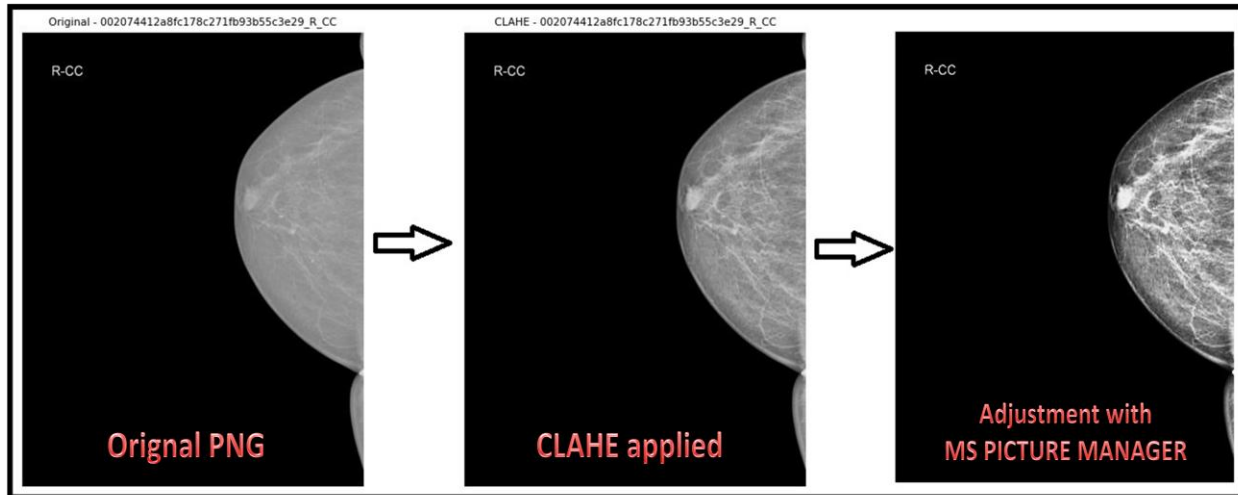

Figure 3: An example of a mammogram image before and after applying CLAHE & MS Picture Manager brightness & contrast adjustment.

3.6.4 Data Augmentation with Synthetic Lesion Images

Augmentation of training data with synthesized artifacts can be useful for various purposes such as to tackle the challenge of class imbalance or to avoid overfitting on over-sampled minority class. There are several ways to create synthetic lesions in medical images used in different studies. Generative Adversarial Networks (GAN) is most prominent technique that produces artificially generated low to high resolution images to augment datasets for training purposes. This technique has been utilized in these studies [24, 25, and 26]. But, training GANs requires a large number of images with respective ground truth annotations which might not be always available, particularly in our case.

Image Transformation is another technique which effects pixel level transformations on medical images to simulate lesions. This includes morphological operations, noise addition and intensity modulation to modify the appearance of tissue to resemble lesions. This study [27] employs these techniques for creating artifacts with calcifications and architectural distortions for data augmentation purposes.

In this study, we used an alternative approach to create samples with synthetic lesions known as ROI-Based Processing [28]. Image processing tools such as Matrix Laboratory (Matlab), GNU Image Manipulation Program (GIMP), Adobe Photoshop or ImageJ can be utilized for this purpose. It involves selecting specific regions within non cancer images and changing their pixel values with the respective ROIs masks containing lesions from cancer images. After inserting the subject lesion, adjustments such as scaling, rotation, brightness, contrast and transparency of adjacent pixels were affected to ensure the lesion appears consistent with the surrounding tissue.

Advantages of this approach are that it doesn't necessarily require a large number of training examples to begin with and is also compatible with all resolutions and image sizes. We created synthetic lesions of Asymmetry, Architectural Distortion, Suspicious Lymph node (30% each) and Focal Asymmetry (15%) to artificially augment the low sample space of images containing these lesions types in VinDr-Mammo dataset.

It is pertinent to mention that the use of these artifacts is limited to the scope of this experiment only and not for any clinical or medical purposes whatsoever. The image identifiers of all artifacts containing synthetic lesions were also renamed after the type of lesion with the naming convention mentioned in part 3.4.1 to avoid any ethical concerns.

Computer Vision Annotation Tool (CVAT) [29] was used to annotate the synthetic artifacts and generate their respective YOLO format labels. These artifacts and their respective labels were included in the training set.
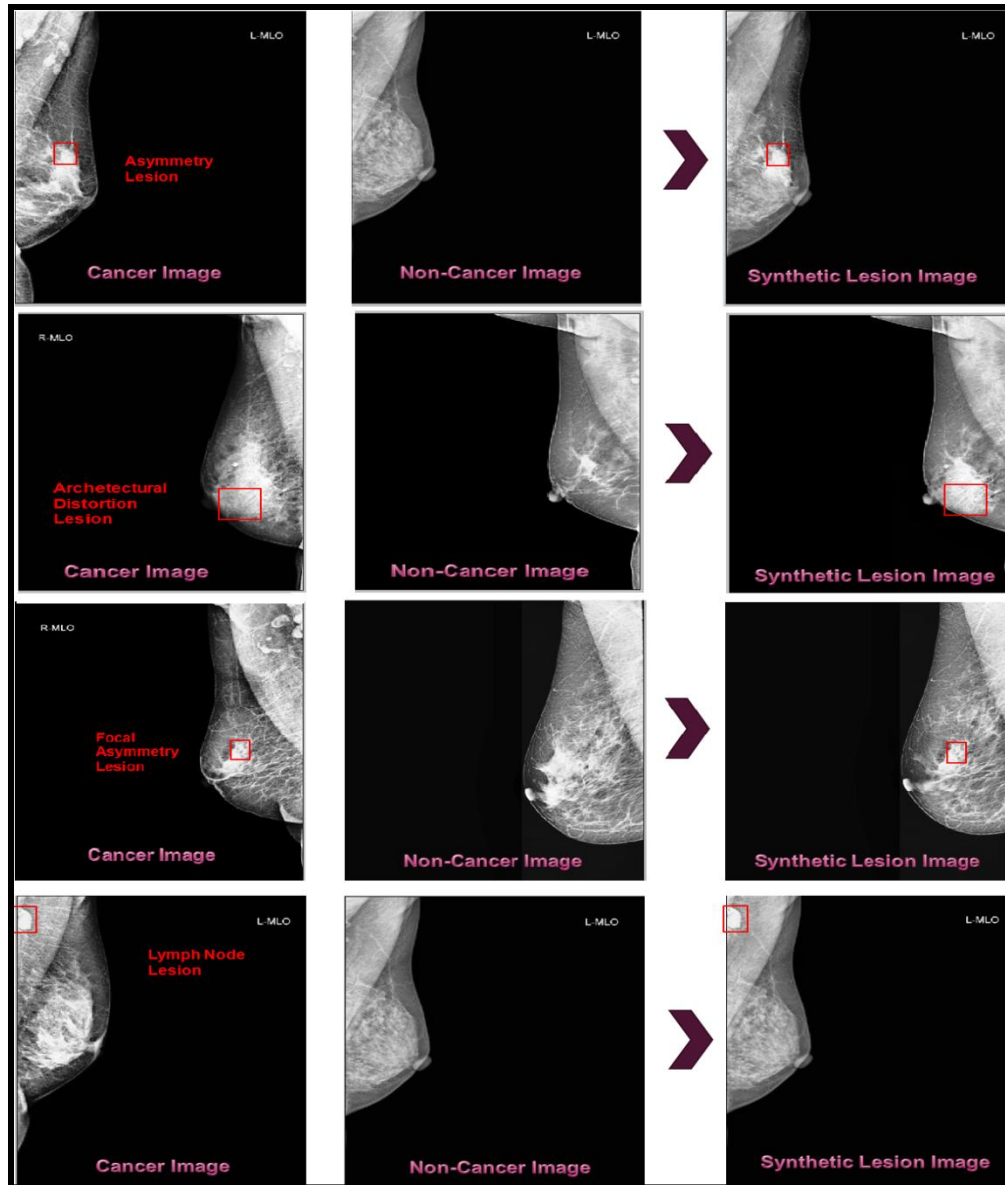
Figure 4: Creating synthetic lesion artifacts with Asymmetry in 1st row, Architectural Distortion in 2nd row, Focal Asymmetry in 3rd row and Lymph Node in 4th row.

3.7   YOLO Introduction

The YOLO model was introduced by Joseph Redmon and colleagues in their publication at Computer Vision and Pattern Recognition Conference (CVPR) 2016 [39] which marked a groundbreaking milestone in the realm of object detection. YOLO's name means "You Only Look Once," conveying its innovative approach in the name. Unlike previous methods that required multiple passes or a two-step process, YOLO achieved real time practical results in

object detection tasks with single network pass. It eliminated the need for sliding windows followed by extensive classifier runs, providing a more efficient solution. YOLO also introduced a simpler output mechanism relying solely on regression for predicting object probabilities and box coordinates, which sets it apart from earlier methods like fast RCNN [40], which used different outputs for regression and classification tasks.

The versatility of YOLO has led to its widespread application across various domains. In the context of video analysis, it has proven instrumental in action recognition within surveillance, sports analysis, and human-computer interaction [41, 42, 43, 44]. In agriculture, YOLO's capabilities have been harnessed for detecting and classifying crops, pests and diseases, thereby enhancing precision agriculture and automating farming processes [45, 46, 47]. Additionally, it has found utility in security, facial recognition [48, 49] and biometrics tasks

YOLO's impact extends to autonomous vehicle systems, enabling swift detection and tracking of a variety of objects including cars, humans, motorcycles etc [50, 51, 52, 53, 54, 55]. YOLO has also been employed by various security systems to monitor real time video feed, facilitating the quick detection of dubious activities, offensive gestures/ behaviours, adherence of social distancing and facial mask compliance in times of COVID pandemic [56, 57]. Furthermore, YOLO based systems have also been used in defect or anomaly detection for the sole aim of quality control (QC) in industry level manufacturing and production systems [58, 59, 60].

In the domain of traffic control systems, the applications of YOLO are numerous. Vehicle number plate recognition, detection of traffic signs and monitoring of traffic flow/ padestrians has significantly contributed to the development of smart traffic management systems [61, 62]. These models have also found utility in national parks and forests for detection of fires and monitoring of different animals specially the endangered species, thereby supporting their care and conservation efforts and active management of ecosystem [63]. Beyond these applications, YOLO has made a substantial impact in robotics and object detection from drones [64, 65, 66, 67].

Exceptional object detection capabilities of YOLO have proven themselves to be priceless in the medical arenas as well. The successfully employment in tumor detection, classification and

segmentation of diseases and identification of various medicines have lead to remarkable improvement in CAD systems resulting in more efficient treatment outcomes [68, 69, 70, 71]. For this reason, YOLO's latest variant "YOLOv8" was selected for our research experiment, owing to its exceptional performance and proven success.

3.8   YOLOv8 Architecture

YOLO deep object detector has the ability to make predictions in a single forward pass through the network, which makes it significantly faster than two-stage detectors like Faster RCNN. For this reason we selected its latest variant's medium size model "YOLOv8m" for our experiment.

YOLOv8 architecture utilizes a few key components to perform object detection tasks. The input image is first resized to default size that the network expects which is 640 x 512 and passed to the "Backbone" which is a series of convolutional layers to extract relevant features from the input image. It is similar to the one in YOLOv5 with minor changes in "CSPLayer" referred now as "C2f module" (Cross Stage Partial Bottleneck with Two Convolutions). It combines various high level features with their contextual information for improvement purposes [28]. The Spatial Pyramid Pooling Fusion (SPPF) layer and the subsequent convolution layers process features at a variety of scales, while the Upsample layers increase the resolution of the feature maps [29]. The "Head" consists of multiple convolutional layers followed by a series of fully connected layers responsible for predicting bounding boxes, objectness scores and class probabilities for the detected objects. It maps the high-dimensional features to the output bounding boxes and object classes [29].

The final layer of the network predicts objects at a lower resolution so this information is combined with earlier feature maps to account for objects of different sizes. The activation function employed to calculate the probability as objectness score for detecting an object is "Sigmoid". Class probabilities are predicted by the softmax function, giving the detected object's likelihood to belong to a specific class [28].

YOLO predicts multiple bounding boxes and their class probabilities for each grid cell in the feature map. Class probabilities are than assigned to these boxes. To improve accuracy, the overlapping bounding boxes are removed using Non-Maximum Suppression (NMS) [28]. The

box with the highest class probability is kept for each object. A list of bounding boxes with respective class labels and confidence scores is formed as output.

A significant advancement in YOLOv8 is its transition to anchor-free detection featuring a decoupled head, enabling the independent processing of objectness, classification, and regression tasks. This architectural innovation enhances the model's overall accuracy as each branch can concentrate on its specific task. Another improved feature in YOLOv8 is anchor free box. This involves predicting the object's center rather than an offset based on anchor boxes [30]. This innovation reduces the number of box predictions, subsequently accelerating the complex post-inference step of NMS. YOLOv8 also introduced architectural enhancements such as adjustments in convolutional layers and kernel sizes.

YOLOv8 utilizes loss functions of CIoU and DFL as the box loss and binary cross-entropy as classification loss [30]. These functions have demonstrated notable enhancements in detection task particularly when smaller objects are of concern. Furthermore, YOLOv8 offers versatility in its usage by extending usage with simple commands through the command line interface. It can also be conveniently installed as a PIP package. Additionally, it features seamless integration options for labeling, training, and deployment, making it a versatile choice for various applications.

Figure 5 displays overall YOLOv8 architecture which is designed to be quick and proficient while still achieving high rates of detection accuracy. One limitation of YOLO is that it still might struggle with detecting small objects compared to some other methods. A recently released feature of YOLOv8 is its support for utilizing 1280 image size, which might overcome this limitation. The YOLOv8 GitHub repository [31] provides an open opportunity to peer review the detailed code implementation.
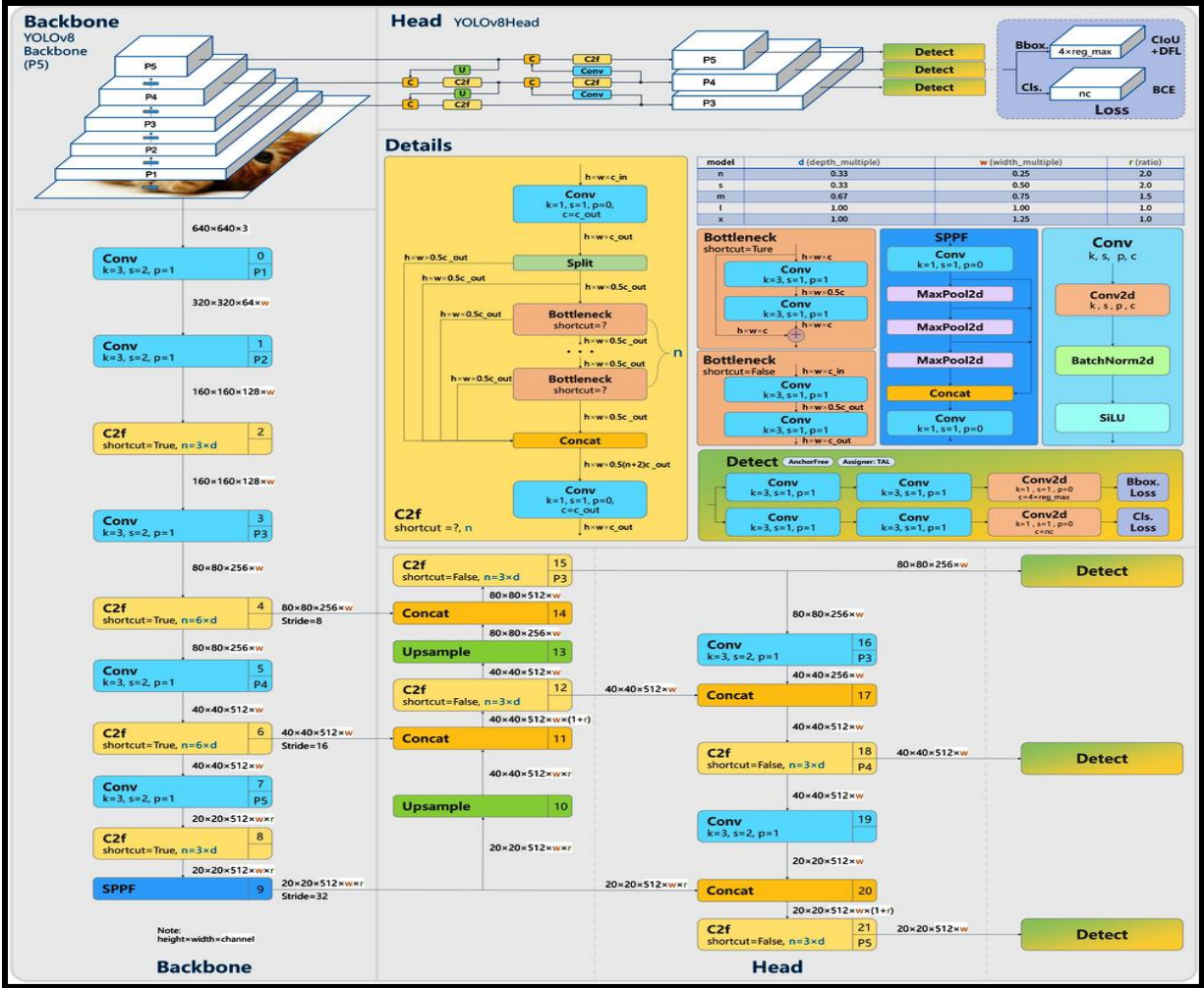
Figure 5: YOLOv8 Architecture. Layers with respective types with relevant kernel size, number of channels and other parameters etc are displayed [30].

3.9 Training YOLOv8m

Our experiment was conducted on the configurations/ hyper parameter settings as mentioned in Table 3. In each training iteration, a batch of 32 training mammograms each comprising input images and corresponding ROI labels is fed into the YOLOv8m network.

In the training process of a batch, YOLOv8m first affects data augmentations on the mammogram images through albumentations library. This library introduces various transformations to them augmenting their diversity and complexity. These include blur operations such as "Blur" and "MedianBlur" applied with a low probability of 0.01, utilizing a blur limit to simulate different levels of image blurring. Furthermore, a grayscale conversion

"ToGray" is occasionally applied (p=0.01) while CLAHE is used (p=0.01) to adjust image contrast within a specified clip limit and tile grid size.
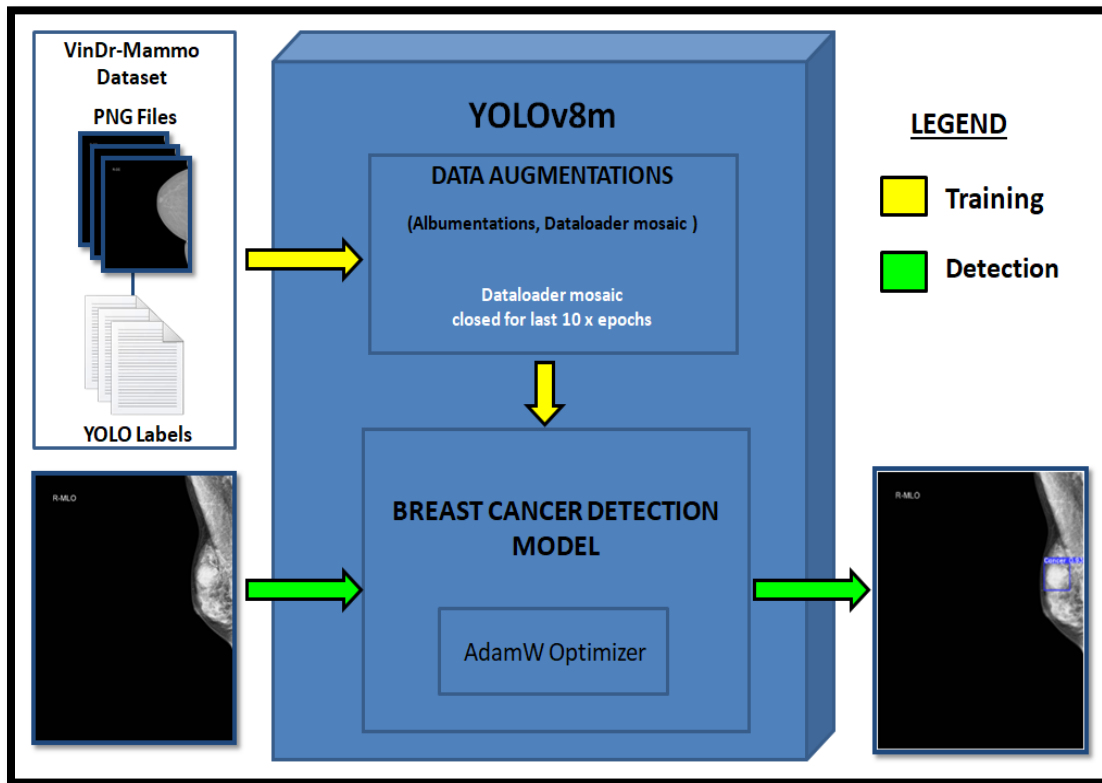

Figure 6: An illustration of YOLOv8m model's training process and subsequent breast cancer detection with VinDr-Mammo dataset.

Moreover, a specific augmentation technique known as "mosaic augmentation" is also employed during training. This technique involves four images concatenated to form a single input enabling the learning of distinct breast cancer lesions in various locations against the surrounding pixels. This process though promotes robustness of the model by offering it exposure to objects in diverse contexts but eventually degrades performance if performed throughout the training routine [30]. For performance optimization purpose, YOLOv8 deliberately deactivates it for last 10 training epochs.

This batch than goes through a sequence of operations including convolutions, activations, pooling and non-linear transformations as mentioned above, processing the images and generating predictions for the lesion bounding boxes and cancer class probabilities. The convergence is measured by comparing calculated predictions to the actual ground truth

annotations and losses are computed. The loss quantifies the disparity between predicted values and actual values encompassing localization accuracy, confidence scores and classification correctness.
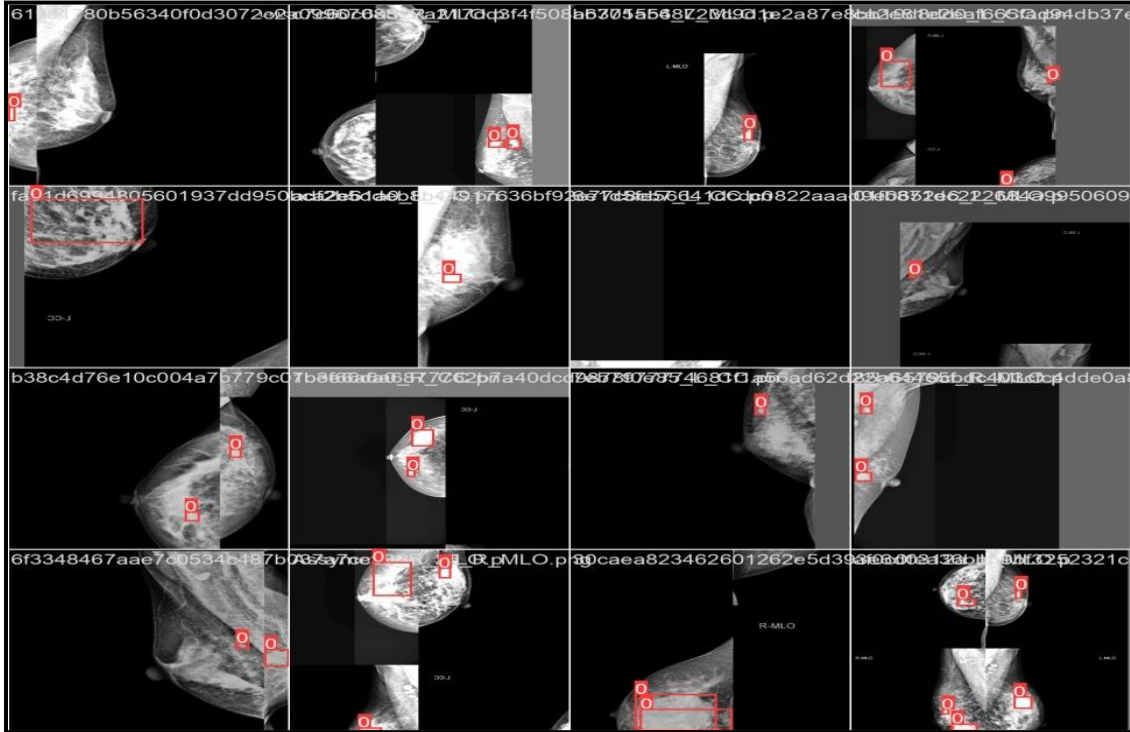

Figure 7: Visualization of a training batch in YOLOv8 displaying albumentations and mosaic augmentation in effect.

Following the loss computation, the AdamW optimizer comes into play. AdamW [32], which is also referred to as Adam with decoupled weight decay (also known as L2 regularization), has gained popularity as the preferred optimizer for training a variety of deep neural network models. The primary benefit attributed to AdamW is its ability to enhance the generalization capabilities compared to the original Adam optimizer [33]. This improvement in generalization brings AdamW to a level of effectiveness comparable to that of SGD with momentum [34] particularly in context of image classification tasks. It adjusts the network's weights and biases with respect to the calculated gradients of the loss.

Before the "t – th" iteration (where t ≥ 1) and the weights are denoted as θ_{t-1}, with an initial learning rate represented as $\alpha \in R$, $\beta \in R$ as the factor of momentum, $\lambda \in R$ as the weight decay

26

value, 1st and 2nd moment vectors as m_t and v_t and ϵ as the smoothing term to prevent division by zero than the update process of the AdamW optimizer is given by following equation [35]:

$$\boldsymbol{\theta_t = (1 - \alpha\lambda)\theta_{t-1} \frac{\alpha m_t}{\sqrt{v_t}+\epsilon}}$$

The step size of parameter updates is determined by the learning rate of 0.001. AdamW's inclusion of weight decay helps prevent overfitting by slightly penalizing large parameter values. This process of forward pass, loss calculation, gradient computation and parameter updates is repeated for each of the 58 steps within the batch.

The network iteratively learns from the samples refining its parameters to minimize the loss. After all these steps are completed, the batch's parameters are updated and the process repeats for the next batch. This iterative training, performed across 300 x epochs allows YOLOv8m to learn and generalize well with respect to multiple lesions of breast cancer present in the mammogram images of training data.

3.10 Experimental Configuration

The experiments were conducted using Google Colab Pro operating with CPU Intel Xeon 2.30GHz and 35.2 GB RAM. TPU was utilized to accelerate the computation process. Python-3.10.12 and torch-2.0.1 were used. The computations were made for training, validation and test sets with aforementioned adjustments to VinDr-Mammo dataset [5]. The study considered a range of hyperparameters to enhance the model's performance including batch size, learning rate and a choice of optimization function. The hyperparameters, as given in Table 4 were determined based on their impact on model performance.

Table 4: Hyperparameter configurations of Yolov8m model.

Hyperparameter	Value
Batch Size	32
Learning Rate	0.001

Epochs	300
Steps per epochs	58
Weight Decay	0.0005
Momentum	0.9
Optimizer	AdamW

4. Results

4.1   Performance Evaluation

With the aforementioned configurations using YOLOv8 model, we performed multiple experiments to explore optimal parameters. Figure 8 shows the graphs of box loss, cls loss, dfl loss, precision, recall and mAP during the training and validation processes.
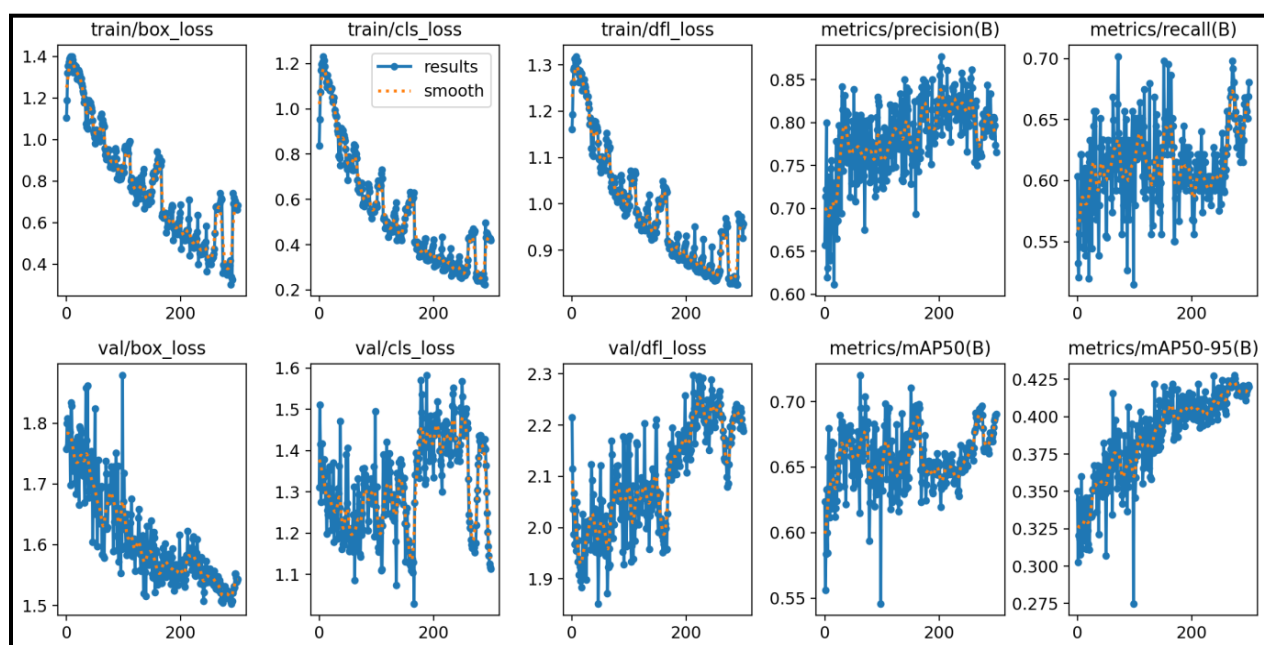

Figure 8: Graphs of box loss, cls loss, dfl loss, precision, recall and mAP during training and validation processes.

The evaluation results show the capability of the YOLOv8 deep object detector to detect multiple breast cancer lesions by demonstrating 89.3% accuracy, 95.4% precision, 88.8% recall, 0.92 F1-score, 0.76 MCC value and 0.72 mAP. These overall results are based on image wise screening which means if an image contained mass and calcification lesions in which mass was detected correctly and calcification was not detected, it was classified as Cancer. The individual detection accuracies of different lesions are presented in Table 5. These accuracies were measured based on correct predictions of the respective ground truths or ROI based detection as per respective occurrences within images. Precision Recall Curve of test data at mAP50 is given by Figure 9.

Table 5: The detection accuracies of different lesions types present in the test set.

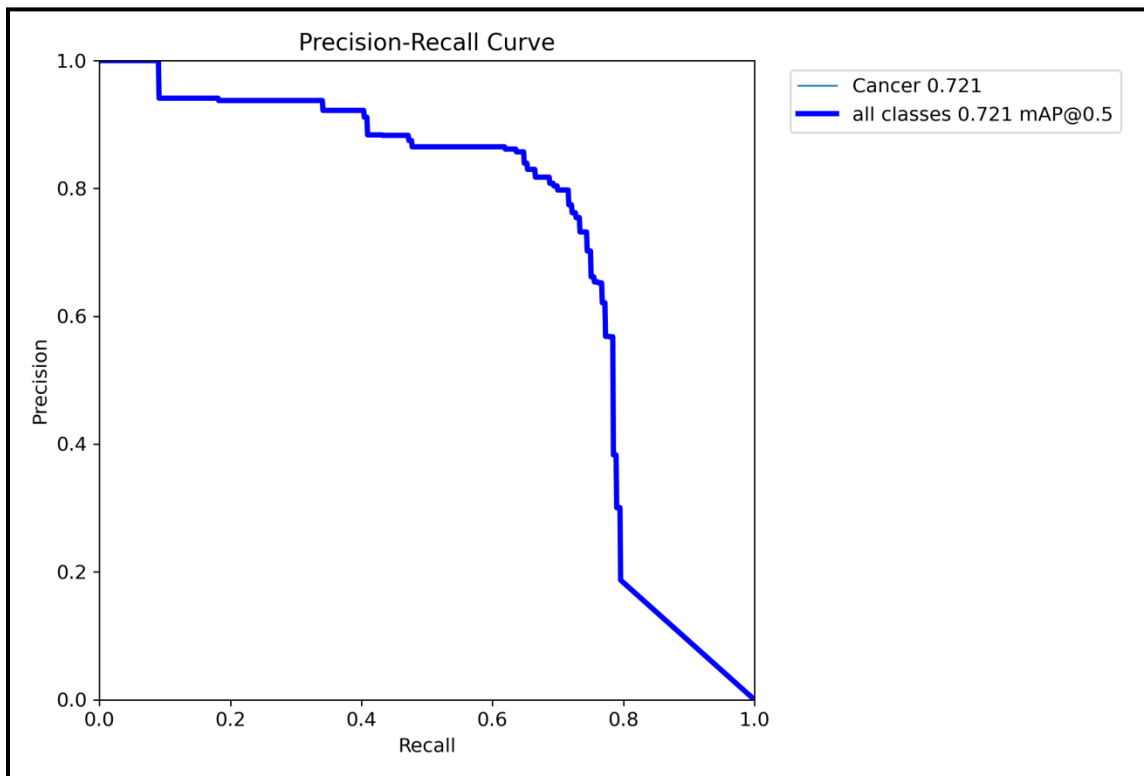| Lesion Type	Accuracy
Mass	88%
Asymmetry	83.33%
Suspicious Calcification	91.43%
Architectural Distortion	82.35%
Focal Asymmetry	83.33%
Suspicious Lymph Node	85.71%

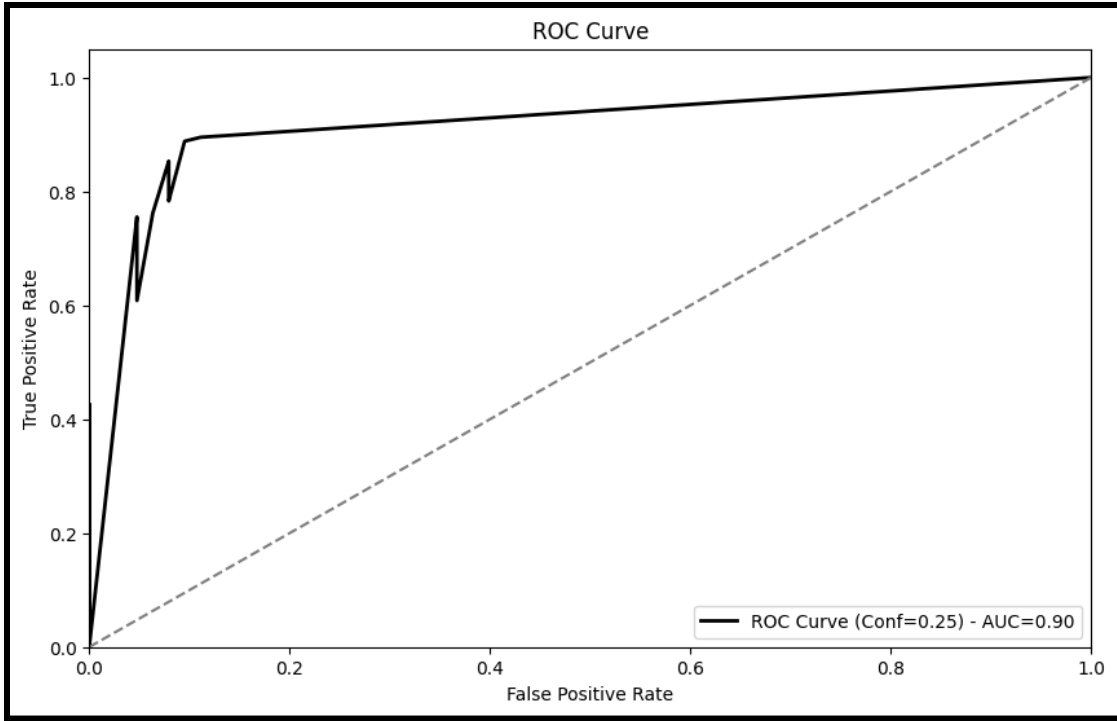Figure 9: Precision Recall Curver at mAP50.

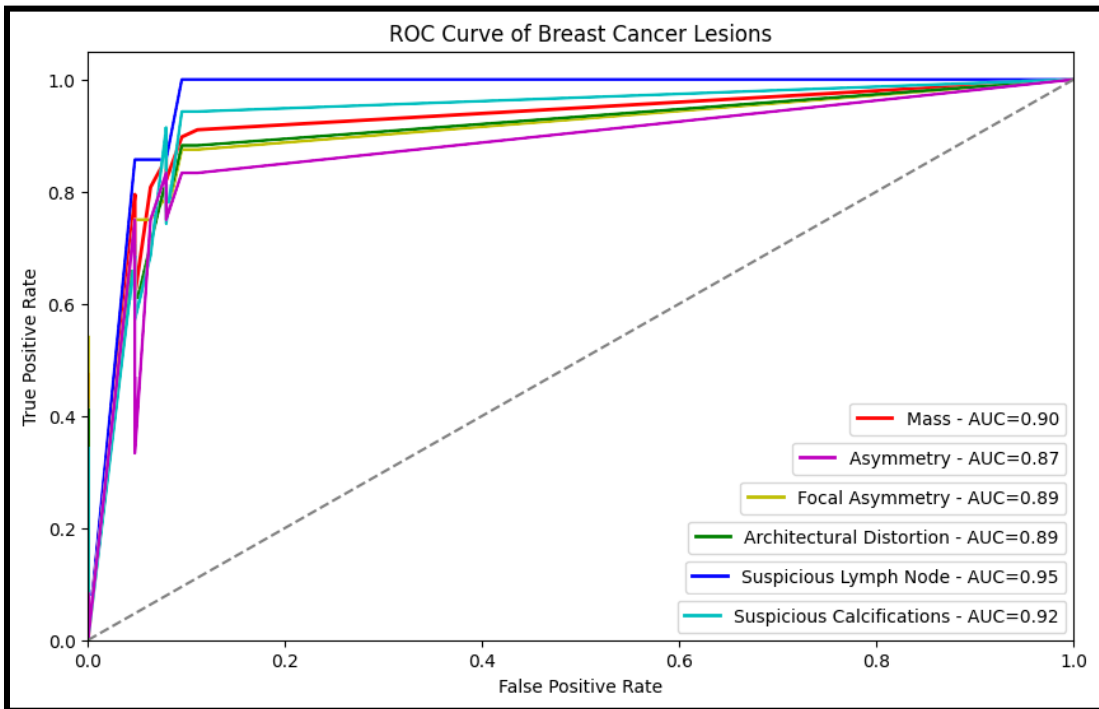Figure 10: ROC curve of the overall screening process.

Figure 11: ROC curves of Suspicious Calcification, Suspicious Lymph Node, Architectural Distortion, Focal Asymmetry, Asymmetry and Mass lesions.

4.2 Comparison with Related Works

An accurate comparison with other studies is a complex phenomenon due to various reasons such as difference in base datasets, pre-processing techniques applied configuration of hyperparamenters and various other training protocols and environments. However, Table 7 shows some similar works where the studies [6, 17] also worked on VinDr-Mammo dataset but only for mass detection. This study [9] worked on detection Mass, Asymmetry & Architectural Distortions Mass and similarly, study [10] worked on detection of Calcification & Architectural Distortions. Each of them employed detection efforts on three types of lesion but their respective datasets suffer from low sample space with regards to each lesion type. The dataset used in this study [11] is relatively larger in size but only contains Mass & Calcification lesions. In comparison, we worked on a large scale dataset containing six diverse types of lesion i.e. Mass, Architectural Distortion, Asymmetry, Focal Asymmetry, Suspicious Calcification & Suspicious Lymph Node. Our data augmentation approach of generating synthetic artifacts contributed greatly in improving out experimental efforts and obtaining consistent detection rate. To the best of our knowledge, our research is the only existing endeavor in the field where a diverse spectrum of lesion types is explored and successfully worked upon achieving reliable results. Detailed comparison of our research with some related works in provided in Table 6.

Table 6: Comparison of our research with some related works.

Paper	Technique	Lesions Detection	Dataset	Results
Our Work	YOLOv8	Mass, Architectural Distortion, Asymmetry, Focal Asymmetry, Suspicious Calcification & Suspicious Lymph Node	VinDr-Mammo	89.3% accuracy, 0.72 mAP
[6]	YOLOv8	Mass	VinDr-Mammo	0.65 mAP
[19]	Weakly Supervised Learning	Mass	VinDr-Mammo	80.12% accuracy
[9]	YOLOv5	Mass, Asymmetry & Architectural Distortions	CBIS-DDSM, INbreast, Private Dataset of 307 images	0.49 mAP 0.83 mAP 0.62 mAP
[10]	YOLO-based fusion model	Mass, Calcification & Architectural Distortions	Private Dataset of 833 x images	Mass: 93% accuracy, Calcification: 88% accuracy Architectural Distortion: 95% accuracy
[11]	modified YOLOv5	Mass & Calcification	CBIS-DDSM	96.5% accuracy, 0.96 mAP

5. Future Work & Conclusions

5.1 Future Work

There are still many arenas to explore given the present research in the field of breast cancer detection and diagnosis. Our work is a humble effort to explore diverse type of lesions but it only includes six different types of lesions at best. One avenue to work upon is to collect more mammography images of distinct lesion types such as nipple retraction, skin retraction, skin thickening and global asymmetry to develop a more comprehensive multi-lesion detection model. Multiple YOLO based models working in an ensemble as used in this study [12], can be exploited to accomplish this challenge. Whereas, the generalizability of our model may also be refined by using above mentioned approach or by enhancing the given dataset. This could be done by incorporating Mass, Calcification, Asymmetry and Architectural Distortions images from other public datasets into VinDr-Mammo improving its sample space. The mammograms for the other lesion types may also be privately acquired. Yolov8 1280 image size model can be also exploited to work upon large resolution images for result improvement purposes.

5.2 Conclusion

The integration of AI in the medical arenas with promising deep learning techniques specially in the field of cancer detection on mammograms has demonstrated significant potential to enhance CAD based breast cancer diagnosis systems. Yet, the challenge of detecting a wide range of lesion types within mammography datasets persisted until now. By leveraging the prowess of one-shot deep object detection technique in the form of YOLOv8 with VinDr-Mammo dataset and employing a progressive data augmentation approach, this study achieved a meticulous accuracy of 89.3%, F1-score of 0.92 and mAP of 0.72. These results not only confirm the effectiveness of the YOLOv8 model in detecting multiple breast cancer lesions but also position this research as a pioneering effort in comprehensively addressing a range of abnormalities encountered in breast cancer pathology.

In essence, this research while contributing to the advancement of breast cancer detection also serves as a stepping stone towards more comprehensive and robust diagnostic methodologies. YOLOv8 offers an active end-to-end approach for accurately identifying cancer lesions within a mammogram image. This offers a promising avenue for future research and makes it an appealing choice in potential clinical applications in improving breast cancer diagnosis to enhance patient outcomes.

References

1. Shaikh, K., Krishnan, S., & Thanki, R. (2021). *Artificial Intelligence in Breast Cancer Early Detection and Diagnosis.* Springer.

2. Sung, H. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians,* 71, 209–249.

3. Jedy-Agba, E., Curado, M. P., & Ogunbiyi, O. (2012). Cancer incidence in Nigeria: a report from population-based cancer registries. *Cancer Epidemiology,* 36(5), e271–e278.

4. Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics. Retrieved from https://github.com/ultralytics/ultralytics

5. Nguyen, H. T., Nguyen, H. Q., Pham, H. H., Lam, K., Le, L. T., Dao, M., & Vu, V. (2022). VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *medRxiv preprint,* posted March 10, 2022.

6. Mahoro, E., & Akhloufi, M. A. (2023). Breast masses detection on mammograms using recent one-shot deep object detectors. In *2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)* (pp. 979-8-3503-3849-2/23/$31.00). IEEE. DOI: 10.1109/BioSMART58455.2023.1016203.

7. Razali, N. F., Isa, I. S., Sulaiman, S. N., Karim, N. K. A., Osman, M. K., & Soh, Z. H. C. (2023). Enhancement technique based on the breast density level for mammogram for computer-aided diagnosis. *Bioengineering,* 10(2), 153.

8. Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., & Elmaghraby, A. S. (2021). Breast Lesions Detection and Classification via YOLO-Based Fusion Models. *Computers, Materials & Continua,* Advance online publication. DOI: 10.32604/cmc.2021.018461.

9. Prinzi, F., Insalaco, M., Orlando, A., et al. (2023). A Yolo-Based Model for Breast Cancer Detection in Mammograms. *Cogn Comput.* https://doi.org/10.1007/s12559-023-10189-6.

10. Baccouche, A., Zapirain, B. G., Zheng, Y., & Elmaghraby, A. S. (2022). Early Detection and Classification of Abnormality in Prior Mammograms Using Image-to-Image Translation and YOLO Techniques. *Computer Methods and Programs in Biomedicine,* Science Direct.

11. Mohiyuddin, A., Basharat, A., Ghani, U., Peter, V., Abbas, S., Naeem, O. B., & Rizwan, M. (2022). Breast Tumor Detection and Classification in Mammogram Images Using Modified YOLOv5 Network. *Hindawi, Computational and Mathematical Methods in Medicine.*

12. Al-Antari, M. A., Han, S.-M., & Kim, T.-S. (2020). Evaluation of Deep Learning Detection and Classification towards Computer-Aided Diagnosis of Breast Lesions in Digital X-ray Mammograms. *Computer Methods And Programs In Biomedicine.*

13. Anonymous. (2023). Multi-View Deep Evidential Fusion Neural Network For Assessment Of Screening Mammograms. *ICLR 2023*.

14. Agarwal, R., Díaz, O., Yap, M. H., Llado, X., & Marti, R. (2020). Deep Learning for Mass Detection in Full Field Digital Mammograms. *Computers in Biology and Medicine*.

15. Altameem, A., Mahanty, C., Poonia, R., Saudagar, A., & Kumar, R. (2022). Breast Cancer Detection in Mammography Images Using Deep Convolutional Neural Networks and Fuzzy Ensemble Modeling Techniques. *MDPI, Basel, Switzerland*.

16. Mahmood T, Li J, Pei Y, Akhtar F,Rehman MU, Wasti SH. (2022). Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. *PLoS ONE*.

17. Gnanasekaran, V. S., Joypaul, S., Sundaram, P. M., & Chairman, D. D. (2020). "Deep Learning Algorithm for Breast Masses Classification in Mammograms." IET Image Processing.

18. Swiderski, B., Gielata, P. O., Olszewski, P., Osowski, S., & Kołodziej, M. (2021). "Deep Neural System for Supporting Tumor Recognition of Mammograms Using Modified GAN." Expert Systems with Applications.

19. Sampaio, V., & Cordeiro, F. R. (2022). A Study on Class Activation Map Methods to Detect Masses in Mammography Images using Weakly Supervised Learning. *Proceedings Of The 19th National Meeting On Artificial And Computational Intelligence*.

20. Suckling, J. (1994). The Mammographic Image Analysis Society digital mammogram database. *Digital Mammography,* pp. 375–386.

21. Bowyer, K. (1996). The digital database for screening mammography. In *Third international workshop on digital mammography,* vol. 58, pp. 27.

22. Moreira, I. C. (2012). INbreast: toward a full-field digital mammographic database. *Academic Radiology,* 19, 236–248.

23. Cui, C., Li, L., Cai, H., Fan, Z., Zhang, L., Dan, T., Li, J., & Wang, J. (2021). The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast. *The Cancer Imaging Archive.* DOI: https://doi.org/10.7937/tcia.eqde-4b16.

24. Swiderski, B., Gielata, P. O., Olszewski, P., Osowski, S., & Kołodziej, M. (2021). Deep Neural System for Supporting Tumor Recognition of Mammograms Using Modified GAN. *Expert Systems with Applications*.

25. Alyafi, B.; Diaz, O.; Martí, R. (2020). DCGANs for realistic breast mass augmentation in x-ray mammography. In *Medical Imaging 2020: Computer-Aided Diagnosis;* Hahn, H.K.; Mazurowski, M.A., Eds.; SPIE: Bellingham, WA, USA, 2020; p. 68.

26. Wu, E.; Wu, K.; Lotter, W. (2020). Synthesizing lesions using contextual GANs improves breast cancer classification on mammograms. *arXiv 2020,* arXiv:2006.00086.

27. Walsh, R., & Tardy, M. (2022). A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer. *MDPI, Basel, Switzerland*.

28.   MathWorks. (2023). ROI-Based Processing.
       https://www.mathworks.com/help/images/roi-based-processing.html

29.   Center for Visual Artificial Intelligence. (2023). CVAT - Computer Vision Annotation
       Tool. https://www.cvat.ai/.

30.   Juan Terven and Diana Cordova-Esparza. (2023). A Comprehensive Review of YOLO:
       From YOLOv1 and Beyond. *arXiv:2304.00501*.

31.   Jocher, G. (2023). Ultralytics Documentation. https://docs.ultralytics.com/.

32.   Solawetz, J. (2023). Whats New In Yolov8. https://blog.roboflow.com/whats-new-in-
       yolov8/.

33.   Jocher, G., Chaurasia, A., & J. Qiu. (2023). YOLO by Ultralytics.
       https://github.com/ultralytics/ultralytics.

34.   Loshchilov, I., Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint
       arXiv:1711.05101*.

35.   Kingma, D.P., Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint
       arXiv:1412.6980*.

36.   Sutskever, I., Martens, J., Dahl, G., Hinton, G. (2013). On the importance of initialization
       and momentum in deep learning. In *International conference on machine learning,* pp.
       1139–1147. PMLR.

37.   Lei Guan. (2023). Weight Prediction Boosts the Convergence of AdamW.
       arXiv:2302.00195.

38.   Rangayyan RM, Banik S, Desautels J. (2010). Computer-aided detection of architectural
       distortion in prior mammograms of interval cancer. *J Digit Imaging,* 23(5), 611–31.

39.   J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-
       time object detection," in Proceedings of the IEEE conference on computer vision and
       pattern recognition, pp. 779–788, 2016.

40.   R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on
       computer vision, pp. 1440–1448, 2015.

41.   Shinde, S., Kothari, A., & Gupta, V. (2018). YOLO-based human action recognition and
       localization. Procedia Computer Science, 133, 831–838.

42.   Ashraf, A. H., Imran, M., Qahtani, A. M., Alsufyani, A., Almutiry, O., Mahmood, A.,
       Attique, M., & Habib, M. (2022). Weapons detection for security and video surveillance
       using CNN and YOLO-v5s. CMC-Computer Materials Continua, 70, 2761–2775.

43.   Zheng, Y., & Zhang, H. (2022). Video analysis in sports by lightweight object detection
       network under the background of sports industry development. Computational
       Intelligence and Neuroscience, 2022.

44.   Ma, H., Celik, T., & Li, H. (2021). YOLO: Detection and classification based on facial
       expressions. In Image and Graphics: 11th International Conference, ICIG 2021, Haikou,
       China, August 6–8, 2021, Proceedings, Part I 11 (pp. 28–39). Springer.

45. Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-v3 model. Computers and Electronics in Agriculture, 157, 417–426.

46. Wu, D., Lv, S., Jiang, M., & Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Computers and Electronics in Agriculture, 178, 105742.

47. Lippi, M., Bonucci, N., Carpio, R. F., Contarini, M., Speranza, S., & Gasparri, A. (2021). A YOLO-based pest detection system for precision agriculture. In 2021 29th Mediterranean Conference on Control and Automation (MED) (pp. 342–347). IEEE.

48. Yang, W., & Jiachun, Z. (2018). Real-time face detection based on YOLO. In 2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII) (pp. 221–224). IEEE.

49. Chen, W., Huang, H., Peng, S., Zhou, C., & Zhang, C. (2021). YOLO-face: A real-time face detector. The Visual Computer, 37, 805–813.

50. Lan, W., Dang, J., Wang, Y., & Wang, S. (2018). Pedestrian detection based on YOLO network model. In 2018 IEEE International Conference on Mechatronics and Automation (ICMA) (pp. 1547–1551). IEEE.

51. Hsu, W.-Y., & Lin, W.-Y. (2021). Adaptive fusion of multi-scale YOLO for pedestrian detection. IEEE Access, 9, 110063–110073.

52. Benjumea, A., Teeti, I., Cuzzolin, F., & Bradley, A. (2021). YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. arXiv preprint arXiv:2112.11798.

53. Dazlee, N. M. A. A., Khalil, S. A., Abdul-Rahman, S., & Mutalib, S. (2022). Object detection for autonomous vehicles with sensor-based technology using YOLO. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 129–134.

54. Liang, S., Wu, H., Zhen, L., Hua, Q., Garg, S., Kaddoum, G., Hassan, M. M., & Yu, K. (2022). Edge YOLO: Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 23(12), 25345–25360.

55. Li, Q., Ding, X., Wang, X., Chen, L., Son, J., & Song, J.-Y. (2021). Detection and identification of moving objects at busy traffic road based on YOLOv4. The Journal of the Institute of Internet, Broadcasting and Communication, 21(1), 141–148.

56. Kumar, P., Narasimha Swamy, S., Kumar, P., Purohit, G., & Raju, K. S. (2021). Real-time, YOLO-based intelligent surveillance and monitoring system using Jetson TX2. In Data Analytics and Management: Proceedings of ICDAM (pp. 461–471). Springer.

57. Bhambani, K., Jain, T., & Sultanpure, K. A. (2020). Real-time face mask and social distancing violation detection system using YOLO. In 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC) (pp. 1–6). IEEE.

58. Li, J., Su, Z., Geng, J., & Yin, Y. (2018). Real-time detection of steel strip surface defects based on improved YOLO detection network. IFAC-PapersOnLine, 51(21), 76–81.

59. Ukhwah, E. N., Yuniarno, E. M., & Suprapto, Y. K. (2019). Asphalt pavement pothole detection using a deep learning method based on YOLO neural network. In 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 35–40). IEEE.

60. Du, Y., Pan, N., Xu, Z., Deng, F., Shen, Y., & Kang, H. (2021). Pavement distress detection and classification based on YOLO network. International Journal of Pavement Engineering, 22(13), 1659–1672.

61. Chen, R.-C., et al. (2019). Automatic license plate recognition via sliding-window Darknet-YOLO deep learning. Image and Vision Computing, 87, 47–56.

62. Dewi, C., Chen, R.-C., Jiang, X., & Yu, H. (2022). Deep convolutional neural network for enhancing traffic sign recognition developed on YOLOv4. Multimedia Tools and Applications, 81(26), 37821–37845.

63. Roy, A. M., Bhaduri, J., Kumar, T., & Raj, K. (2023). Wildect-YOLO: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection. Ecological Informatics, 75, 101919.

64. Kulik, S., & Shtanko, A. (2020). Experiments with neural net object detection system YOLO on small training datasets for intelligent robotics. In Advanced Technologies in Robotics and Intelligent Systems: Proceedings of ITR 2019 (pp. 157–162). Springer.

65. Dos Reis, D. H., Welfer, D., De Souza Leite Cuadros, M. A., & Gamarra, D. F. T. (2019). Mobile robot navigation using an object recognition software with RGBD images and the YOLO algorithm. Applied Artificial Intelligence, 33(14), 1290–1305.

66. Sahin, O., & Ozer, S. (2021). YoloDrone: Improved YOLO architecture for object detection in drone images. In 2021 44th International Conference on Telecommunications and Signal Processing (TSP) (pp. 361–365). IEEE.

67. Chen, C., Zheng, Z., Xu, T., Guo, S., Feng, S., Yao, W., & Lan, Y. (2023). YOLO-based UAV technology: A review of the research and its applications. Drones, 7(3), 190.

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

69. Nie, Y., Sommella, P., O'Nils, M., Liguori, C., & Lundgren, J. (2019). Automatic detection of melanoma with YOLO deep convolutional neural networks. In 2019 E-Health and Bioengineering Conference (EHB) (pp. 1–4). IEEE.

70. Ünver, H. M., & Ayan, E. (2019). Skin lesion segmentation in dermoscopic images with a combination of YOLO and GrabCut algorithm. Diagnostics, 9(3), 72.

71. Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of RetinaNet, SSD, and YOLOv3 for real-time pill identification. BMC Medical Informatics and Decision Making, 21, 1–11.