

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

8-2022

Computer aided diagnosis system for breast cancer using deep learning.

Asma Baccouche
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>



Part of the [Artificial Intelligence and Robotics Commons](#), [Biomedical Commons](#), [Categorical Data Analysis Commons](#), [Computer and Systems Architecture Commons](#), [Data Science Commons](#), [Diagnosis Commons](#), and the [Statistical Models Commons](#)

Recommended Citation

Baccouche, Asma, "Computer aided diagnosis system for breast cancer using deep learning." (2022). *Electronic Theses and Dissertations*. Paper 3931.
<https://doi.org/10.18297/etd/3931>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

COMPUTER AIDED DIAGNOSIS SYSTEM FOR BREAST CANCER USING DEEP
LEARNING

By

Asma Baccouche

M.Sc. Computer Science, Department of Computer Science
and Engineering, University of Louisville, Kentucky, 2019

A Dissertation

Submitted to the Faculty of the
J. B. Speed School of Engineering of the University of
Louisville of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy in Computer Science and Engineering

Department of Computer Science and Engineering
University of Louisville
Louisville, Kentucky

August 2022

COMPUTER AIDED DIAGNOSIS SYSTEM FOR BREAST CANCER USING DEEP
LEARNING

By

Asma Baccouche

M.Sc. Computer Science, Department of Computer Science
and Engineering, University of Louisville, Kentucky, 2019

A Dissertation Approved on

July 25, 2022

By the Following Dissertation Committee:

Adel S. Elmaghraby, Ph.D., Co-Advisor

Maria Begona Garcia Zapirain, Ph.D., Co-Advisor

Daniel Sierra-Sosa, Ph.D., Co-Advisor

Monica Gentili, Ph.D.

Juw Won Park, Ph.D.

Hui Zhang, Ph.D.

DEDICATION

This dissertation is dedicated to my parents

Mr. Aroussi Baccouche

and

Mrs. Monia Ben cheikha

and my siblings

Ms. Afef Baccouche

and

Mr. Mehdi Baccouche

whose good examples have taught me to go on every adventure for the things I aspire to achieve

and finally, to my grandmother's soul Fatma Ben cheikha

for her endless kindness and love

ACKNOWLEDGMENTS

I would like to thank Dr. Adel Elmaghraby, my PhD dissertation advisor for his guidance and his precious remarks during the duration of my studies to make this experience within the program more comfortable and interesting. I cannot thank him enough for his support and encouragement to accomplish this work and contribute to the research field within his lab team.

I also would like to express my express my deepest gratitude to Dr. Maria Begona Garcia Zapirain, my dissertation co-advisor within the collaboration of eVida group Research at Deusto University, Bilbao Spain, who continuously supported me and gave me all the assistance and help I needed for this work. Her thrilling ideas and precise suggestions have been an essential part of my accomplishments during the past years. I would also thank her team worker Cristiàn Castillo Olea who helped in providing the mammography dataset and its annotations from the National Institute of Cancerology (INCAN) in Mexico City.

A special thanks to Dr. Travis Ross, the director of Speed Technology Solutions, who gave me the chance to gain leadership and management skills as a graduate assistant. I cannot thank him enough for supporting my studies since the Master program. I perceive this opportunity in University of Louisville as a big milestone in my career and I will strive to use the gained skills and knowledge in the best possible way.

I would like to express my gratitude to my lab co-workers for sharing their research experience and ensuring a rich and pleasing environment. A special thanks to Lucas Camino for his assistance to deploy this work and make of this research work a possible product that could be tested online.

I express my gratitude to Dr. Daniel Sierra-Sosa, Dr. Monica Gentili, Dr. Juw Won Park, and Dr. Hui Zhang for accepting to be on my dissertation committee and taking the time to review my research work.

Last but not least, I would never forget the support and the encouragement that I had from my friends and the unconditional love of my family, especially my father, my mother, my sister, and my brother during my studies. Without them, I would have never taken these steps in my life to become the person I wanted to be. I will constantly strive to be the best version of myself and make you proud.

ABSTARCT

COMPUTER AIDED DIAGNOSIS SYSTEM FOR BREAST CANCER USING DEEP LEARNING

Asma Baccouche

July 25, 2022

The recent rise of big data technology surrounding the electronic systems and developed toolkits gave birth to new promises for Artificial Intelligence (AI). With the continuous use of data-centric systems and machines in our lives, such as social media, surveys, emails, reports, etc., there is no doubt that data has gained the center of attention by scientists and motivated them to provide more decision-making and operational support systems across multiple domains.

With the recent breakthroughs in artificial intelligence, the use of machine learning and deep learning models have achieved remarkable advances in computer vision, ecommerce, cybersecurity, and healthcare. Particularly, numerous applications provided efficient solutions to assist radiologists and doctors for medical imaging analysis, which has remained the essence of the visual representation that is used to construct the final observation and diagnosis.

Medical research in cancerology and oncology has been recently blended with the knowledge gained from computer engineering and data science experts. In this context, an automatic assistance or commonly known as Computer-aided Diagnosis (CAD) system has become a popular area of research and development in the last decades. As a result, the CAD systems have been developed using multidisciplinary knowledge and expertise and they have been used to analyze the patient information to assist clinicians and practitioners in their decision-making process. Treating and preventing cancer remains a crucial task that radiologists and oncologists face every day to detect and investigate abnormal tumors. Therefore, a CAD system could be developed to provide decision support for many applications in the cancer patient care processes, such as lesion detection, characterization, cancer staging, tumors assessment, recurrence, and prognosis prediction.

Breast cancer has been considered one of the common types of cancers in females across the world. It was also considered the leading cause of mortality among women, and it has been increased drastically every year. Early detection and diagnosis of abnormalities in screened breasts has been acknowledged as the optimal solution to examine the risk of developing breast cancer and thus reduce the increasing mortality rate.

Accordingly, this dissertation proposes a new state-of-the-art CAD system for breast cancer diagnosis that is based on deep learning technology and cutting-edge computer vision techniques. Mammography screening has been recognized as the most effective tool to early detect breast lesions for reducing the mortality rate. It helps reveal abnormalities in the breast such as Mass lesion, Architectural Distortion, Microcalcification. With the number of daily patients that were screened is continuously increasing, having a second reading tool or assistance system could leverage the process of breast cancer diagnosis. Mammograms could be obtained using different modalities such as X-ray scanner and Full-Field Digital mammography (FFDM) system. The quality of the mammograms, the characteristics of the breast (i.e., density, size) or/and the tumors (i.e., location, size, shape) could affect the final diagnosis. Therefore, radiologists could miss the lesions and consequently they could generate false detection and diagnosis. Therefore, this work was motivated to improve the reading of mammograms in order to increase the accuracy of the challenging tasks.

The efforts presented in this work consists of new design and implementation of neural network models for a fully integrated CAD system dedicated to breast cancer diagnosis. The approach is designed to automatically detect and identify breast lesions from the entire mammograms at a first step using fusion models' methodology. Then, the second step only focuses on the Mass lesions and thus the proposed system should segment the detected bounding boxes of the Mass lesions to mask their background. A new neural network architecture for mass segmentation was suggested that was integrated with a new data enhancement and augmentation technique. Finally, a third stage was conducted using a stacked ensemble of neural networks for classifying and diagnosing the pathology (i.e., malignant, or benign), the Breast Imaging Reporting and Data System (BI-RADS) assessment score (i.e., from 2 to 6), or/and the shape (i.e., round, oval, lobulated, irregular) of the segmented breast lesions.

Another contribution was achieved by applying the first stage of the CAD system for a retrospective analysis and comparison of the model on Prior mammograms of a private dataset. The work was conducted by joining the learning of the detection and classification model with the image-to-image mapping between Prior and Current screening views.

Each step presented in the CAD system was evaluated and tested on public and private datasets and consequently the results have been fairly compared with benchmark mammography datasets. The integrated framework for the CAD system was also tested for deployment and showcase. The performance of the CAD system for the detection and identification of breast masses reached an overall accuracy of 97%. The segmentation of breast masses was evaluated together with the previous stage and the approach achieved an overall performance of 92%. Finally, the classification and diagnosis step that defines the outcome of the CAD system reached an overall pathology classification accuracy of 96%, a BIRADS categorization accuracy of 93%, and a shape classification accuracy of 90%. Results given in this dissertation indicate that our suggested integrated framework might surpass the current deep learning approaches by using all the proposed automated steps. Limitations of the proposed work could occur on the long training time of the different methods which is due to the high computation of the developed neural networks that have a huge number of the trainable parameters.

Future works can include new orientations of the methodologies by combining different mammography datasets and improving the long training of deep learning models. Moreover, motivations could upgrade the CAD system by using annotated datasets to integrate more breast cancer lesions such as Calcification and Architectural distortion.

The proposed framework was first developed to help detect and identify suspicious breast lesions in X-ray mammograms. Next, the work focused only on Mass lesions and segment the detected ROIs to remove the tumor's background and highlight the contours, the texture, and the shape of the lesions. Finally, the diagnostic decision was predicted to classify the pathology of the lesions and investigate other characteristics such as the tumors' grading assessment and type of the shape.

The dissertation presented a CAD system to assist doctors and experts to identify the risk of breast cancer presence. Overall, the proposed CAD method incorporates the advances of image processing, deep learning, and image-to-image translation for a biomedical application.

TABLE OF CONTENTS

DEDICATION.....	III
ACKNOWLEDGMENTS	IV
ABSTARCT	V
LIST OF TABLES.....	XI
LIST OF FIGURES	XIII
CHAPTER I INTRODUCTION	1
A. BREAST CANCER	1
B. MAMMOGRAPHY	2
C. ARTIFICIAL INTELLIGENCE AND CAD SYSTEMS FOR BREAST CANCER	9
D. DISSERTATION ORGANIZATION.....	11
CHAPTER II LITITERATURE REVIEW MODELS AND METHODS FOR BREAST CANCER DIAGNOSIS	13
A. INTRODUCTION.....	13
B. BREAST CANCER DETECTION AND IDENTIFICATION.....	15
C. BREAST CANCER SEGMENTATION.....	18
D. IMAGE-TO-IMAGE TRANSLATION	21
E. BREAST CANCER PATHOLOGY & DIAGNOSIS CLASSIFICATION.....	22
F. CONCLUSION AND DISCUSSION	25
CHAPTER III PROPOSED METHODOLOGY.....	30
A. INTRODUCTION.....	30
B. YOLO-BASED FUSION MODEL FOR ABNORMALITIES DETECTION AND IDENTIFICATION.....	30

1. YOLO-Based Model	30
2. Fusion Models Approach	32
C. EARLY DETECTION AND CLASSIFICATION OF ABNORMALITIES IN PRIOR MAMMOGRAMS.....	33
D. NOVEL CONNECTED-UNETS ARCHITECTURE MODEL FOR BREAST MASS SEGMENTATION.....	35
E. IMAGE SYNTHESIS FOR DATA QUALITY ENHANCEMENT AND AUGMENTATION.....	39
F. STACKED ENSEMBLE OF NEURAL NETWORKS FOR TUMORS CLASSIFICATION AND DIAGNOSIS.....	42
1. ResNet base model: transfer learning and fine-tuning	42
2. Stacked Ensemble of ResNet models for Breast tumors classification.....	45
CHAPTER IV EXPERIMENTAL RESULTS & DISCUSSION.....	49
A. INTRODUCTION.....	49
B. MAMMOGRAPHY DATASETS.....	50
1. Public datasets.....	50
a. Curated Breast Imaging Subset of DDSM dataset	50
b. INbreast dataset	51
2. Private Datasets.....	52
a. INCAN private collection.....	52
b. UCHCDM private collection.....	53
3. Datasets annotation distribution.....	55
a. Detection and Identification task	56
b. Early detection of abnormalities in prior mammograms task.....	56
c. Mass segmentation task	57
d. Tumor Classification and Diagnosis task	58
4. Data preparation.....	59
C. EVALUATION METRICS & EXPERIMENTS SETUPS.....	63
1. Detection and Identification metrics	63
2. Retrospective early detection metrics	64
3. Segmentation metrics	64
4. Classification and Diagnosis metrics	65

5. Experimental setups	66
D. EVALUATION RESULTS.....	69
1. Detection and Identification.....	69
a. Single Models Evaluation.....	69
b. Fusion Models Evaluation.....	72
2. Early detection of abnormalities in prior mammograms.....	74
a. Evaluation of YOLO-based model on Current Mammograms.....	74
b. Evaluation of YOLO-based model on Prior Mammograms.....	78
c. Retrospective analysis for the early detection and identification on prior mammograms	82
3. Breast Mass Segmentation	88
a. Quantitative segmentation results.....	88
b. Qualitative segmentation results.....	94
4. Classification and Diagnosis.....	99
a. Quantitative classification results.....	99
i. Pathology classification.....	99
ii. BI-RADS category classification	101
iii. Shape classification.....	103
b. Qualitative classification results.....	104
E. TESTING & DEPLOYMENT.....	109
F. DISCUSSION	111
CHAPTER V CONCLUSION & FUTURE WORKS	114
REFERENCES	116
CURRICULUM VITAE.....	134

LIST OF TABLES

Table 1: BI-RADS assessment categories	9
Table 2: Related works	27
Table 3: ResNetV2 architecture layers	44
Table 4: Breast lesions data distribution.....	56
Table 5: Current and Prior data distribution	56
Table 6: Pathology class labels distribution for only Mass lesions	59
Table 7: BI-RADS category class labels distribution for only Mass lesions	59
Table 8: Shape class labels distribution for only Mass lesions	59
Table 9: Data distribution of the mammography datasets	67
Table 10: Cross Validation Folds: Data distribution across class labels	68
Table 11: Hyperparameters for the ResNetV2 base model	68
Table 12: Model performance for different configurations	70
Table 13: Model performance for different prediction classes.....	70
Table 14: Comparison performance using fusion models approach	73
Table 15: Comparison of mass detection with other works	74
Table 16: Comparison performance for different models across labeled classes on Test sets.....	75
Table 17: Performance results for Detection and Classification on Test sets	77
Table 18a: Inference results of YOLO Fusion model on Test sets of original Prior Mammograms	79
Table 18b: Inference results of YOLO Fusion model on Test sets of Prior Mammograms using CycleGAN method	80
Table 18c: Inference results of YOLO Fusion model on Test sets of Prior Mammograms using Pix2Pix method	80
Table 19: Comparison of Mass detection with other works	87
Table 20: Comparison of Early detection with other works	88

Table 21: Segmentation performance of our proposed networks on the test sets.....	88
Table 22: Comparison of the proposed architectures after adding synthetic CBIS-DDSM and INbreast	92
Table 23: Final Segmentation performance of our proposed networks on the test sets	93
Table 24: Comparison of the proposed architectures and state-of-the-art methods	93
Table 25a: Pathology classification results on the CBIS-DDSM dataset.....	99
Table 25b: Pathology classification results on the INbreast dataset.....	99
Table 25c: Pathology classification results on the INCAN private dataset.....	100
Table 26a: BI-RADS category classification results on the CBIS-DDSM dataset	101
Table 26b: BI-RADS category classification results on the INbreast dataset	102
Table 26c: BI-RADS category classification results on the INCAN private dataset	102
Table 27: Shape classification results on the CBIS-DDSM dataset.....	103
Table 28: Comparison of classification performance (accuracy %) using the proposed CAD system with and without mass segmentation.....	106
Table 29: Comparison of the proposed methodology and state-of-the-art methods	108

LIST OF FIGURES

FIGURE 1: Breast composition in case of metastatic breast cancer	2
FIGURE 2: Medical imaging for investigating and screening of breast lesions	3
FIGURE 3: Mammography screening process.....	4
FIGURE 4: Samples mammograms of Mediolateral oblique (MLO) views.....	4
FIGURE 5: Samples mammograms of craniocaudal (CC) views	5
FIGURE 6: Samples mammograms with Mass lesions (first row) and with Calcification lesions (second row)	6
FIGURE 7: Samples of region of interest (ROI) OF breast Mass lesions from public and private datasets ..	7
FIGURE 8: Samples of region of interest (ROI) OF breast Calcification lesions from public and private datasets.....	8
FIGURE 9: Samples of region of interest (ROI) OF breast Architectural Distortion lesions from public and private datasets.....	9
FIGURE 10: Integrated framework of CAD system for breast cancer (A mammogram case with a malignant mass lesion, BI-RADS score of 4 and irregular shape).	14
FIGURE 11: Proposed YOLO-based architecture	31
FIGURE 12: Flow chart of the fusion models' approach for the final prediction.....	33
FIGURE 13: YOLO-based Fusion model - Example of mammogram with a Mass lesion	34
FIGURE 15: The proposed Connect-UNets architecture	36
FIGURE 16: The Atrous Spatial Pyramid Pooling (ASPP) block	37
FIGURE 17: The proposed Connect-AUNets architecture	38
FIGURE 18: The attention block.....	38
FIGURE 19: The proposed Connect-ResUNets architecture	39
FIGURE 20: The residual block	39

FIGURE 21: Comparative scheme of standard GAN vs two variants for image-to-image translation: Pix2Pix and CycleGAN.....	41
FIGURE 22: CycleGAN architecture for data synthesis	42
FIGURE 23: Residual block of ResNetV2.....	43
FIGURE 24: Framework of the classification base model: a TF-ResNetV2 model which is a ResNetV2 model pre-trained on ImageNet data and modified and fine-tuned on mammography dataset.....	45
FIGURE 25: Framework of the classification Stacked Ensemble of ResNet models	46
FIGURE 26: The proposed integrated framework	47
FIGURE 27: The proposed integrated framework	47
FIGURE 28: The proposed integrated CAD framework.....	48
FIGURE 29: The sequence of different experiments using the public and private mammography datasets	49
FIGURE 30: Samples of mammograms from the CBIS-DDSM dataset. Cases with benign and malignant images of Mass and Calcification lesions.....	51
FIGURE 31: Samples of mammograms from the INbreast dataset. Cases with benign and malignant images of Mass and Calcification lesions	52
FIGURE 32: Samples of mammograms from the INCAN private dataset. Only cases with malignant images of Mass and Calcification lesions	53
FIGURE 33: Samples of Prior and Current mammograms screenings	54
FIGURE 34: Samples of Current mammograms screenings including cases with Mass, Calcification, and Architectural distortion.....	55
FIGURE 35: Samples from the public and private mammography datasets with zoomed-in ROI of mass ground truth (red). (a) Example from the CBIS-DDSM mammogram with MLO view, (b) Example from the INbreast mammogram with MLO view, and (c) Example from the INCAN mammogram with CC view ..	57
FIGURE 36: Samples of entire mammograms and ROI of detected and segmented Masses from different mammography datasets with their ground-truth of location and contour in red	58
FIGURE 37: Samples from Current exams of original (upper row) and preprocessed mammograms (bottom row).....	60

FIGURE 38: Samples of ROI masses and their ground-truth masks from each mammography dataset used for segmentation	62
FIGURE 39: Samples of the detected and segmented ROI masses for each class within different classification tasks from different mammography datasets	63
FIGURE 40: Hyperparameters tuning; (a) confidence score; (b) learning rate; (c) batch size and (d) number of epochs	67
FIGURE 41: Examples of different calcifications shape and localization (ground-truth of calcification is marked in green, ground-truth of mass is marked in red) for CBIS-DDSM, INbreast and INCAN private datasets (from left to right)	71
FIGURE 42: Examples of breast lesions detection and classification results and their confidence score toward different classes on CBIS-DDSM, INbreast and INCAN private datasets (from left to right): mass (green boxes) and calcification (yellow boxes)	72
FIGURE 43: FROC curve plots of the YOLO based proposed Detection and identification models per class label on Test sets.....	76
FIGURE 44: ROC curve plots of the proposed YOLO-based Fusion Model per class label on Test sets....	77
FIGURE 45: Confusion matrix of prediction results for Current Mammograms.....	78
FIGURE 46: Confusion matrix of prediction results for Prior Mammograms.....	82
FIGURE 47: Comparison of Mean follow-up exam time for prediction results across classes with and without image-to-image translation Pix2pix and CycleGAN on Prior mammograms	83
FIGURE 48: Comparison performance of YOLO Fusion model across classes with and without image-to-image translation.....	84
FIGURE 49: Comparison performance of YOLO Fusion model and Pix2Pix for image-to-image translation for different classes across follow-up exam time (years)	84
Figure 50: Example results on Prior mammograms vs Current mammograms using the YOLO-based Fusion model that was inferred on the translated Prior images from Pix2Pix method across classes: Mass (red bounding boxes, first row), Calcification (green bounding boxes, second row), and Architectural Distortion (yellow bounding boxes, third row). Red arrows point to the ground truth location. Last row belongs to the Normal class	86

FIGURE 51: Performance of mass segmentation using the different architectures in terms of ROC curves on the test sets of CBIS-DDSM, INbreast, and the INCAN private datasets	90
FIGURE 52: Samples of synthetic data from CBIS-DDSM and INbreast datasets generated by CycleGAN model using the INCAN private dataset.....	91
FIGURE 53 Examples of the segmentation results on the test set of the datasets. Subplot on the top shows two samples of mammograms from the CBIS-DDSM dataset.....	95
FIGURE 54: Examples of the segmentation results for the proposed architecture Connected-ResUNets comparing with and without using ASPP block.....	96
FIGURE 55: Examples of the segmented masses for the proposed architecture Connected-ResUNets comparing with the ResUNet.....	97
FIGURE 56: Examples of the segmentation results for the proposed architecture Connected-ResUNets with and without adding the synthetic data	98
FIGURE 57: Examples of the segmentation results for the proposed architecture models again two state-of-the-art methods FrCN and CRUNET	98
FIGURE 58: Performance of pathology classification using different models in terms of ROC curves and AUC score	101
FIGURE 59: Performance of BI-RADS category classification using different models in terms of ROC curves and AUC score.....	103
FIGURE 60: Performance of Shape classification using different models in terms of ROC curves and AUC score.....	104
FIGURE 61: Confusion matrix of the stacked ensemble of models for the pathology classification on the mammography datasets	105
FIGURE 62: Confusion matrix of the stacked ensemble of models for the BI-RADS category classification on the mammography datasets.....	105
FIGURE 63: Confusion matrix of the stacked ensemble of models for the Shape classification on the datasets	106
FIGURE 64: Framework of the deployed CAD system.....	110

CHAPTER I

INTRODUCTION

A. Breast Cancer

In recent decades, cancer diseases have occurred in many cases worldwide and the global number is predicted to grow to 27.5 million new cases by 2040 [1]. Cancer is described by the uncontrolled growth of abnormal cells within the body. There are quite 100 different types of cancer diseases which will occur anywhere within the body, but they are mainly described in five categories: carcinomas (i.e., skin and tissues affecting the inner organs), sarcomas (i.e., muscles, bones, or cartilage), leukemia (i.e., blood and bone marrow), lymphomas (i.e., immune system), and central nervous system (i.e., brain and spinal cord) [2]. Most commonly, lung cancer and prostate cancer occur in men, and breast cancer occurs in women.

Over years, breast cancer has remained the foremost frequently diagnosed non-skin cancer and therefore the leading explanation for death among females with a rate of 32% of total cancer cases [3]. According to the American Cancer Society, it is estimated that over 290,000 new cases will be reported, and 43,780 women will die from breast cancer in 2022 [4]. Early detection and diagnosis of breast cancer is the most effective way to treat this disease and reduce the mortality rate [5].

Breast cancer starts at the breast and replicates abnormal cells into lumps, often called tumors, as shown in Figure 1. Certain factors can increase the risk of developing the disease like aging and genetics [6]. The tumor can be classified as malignant or benign. While benign tumors are considered non-cancerous, the risk of malignant tumors must be studied before being spread into other health organs such as bones, lungs, liver, and adrenal gland [7]. Consequently, early detection and diagnosis of breast cancer has become required to reduce the high mortality rate among women; thus, diagnostic systems have been studied to help radiologists with more precise analysis [8-11].

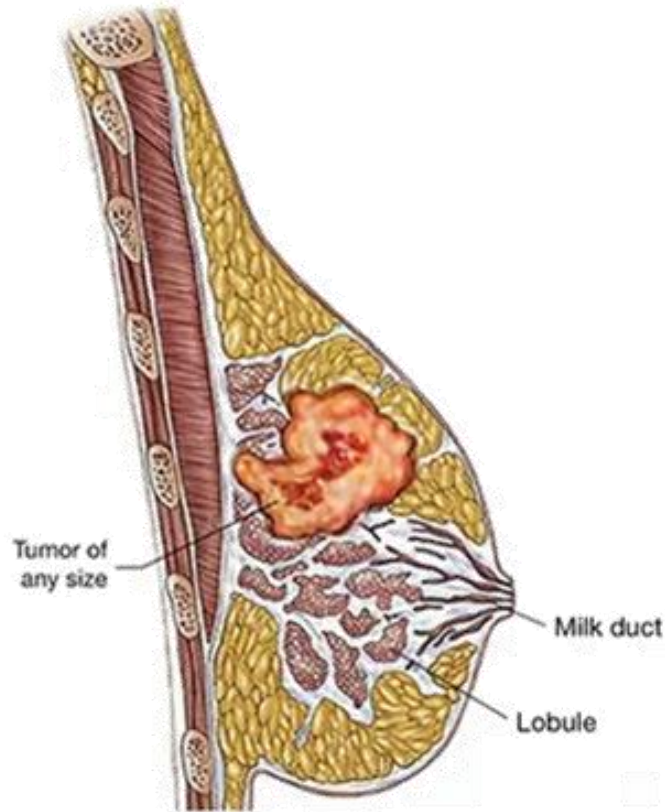


FIGURE 1: Breast composition in case of metastatic breast cancer

Source: <https://www.cancersupportcommunity.org/metastatic-breast-cancer>

B. Mammography

Since the invention of X-rays, medical imaging has significantly affected the cancerology. Over the last decades, a variety of imaging modalities and techniques have contributed to the screening process such as the computed tomography (CT), the ultrasonography, the magnetic resonance imaging (MRI), and the digital radiography [12]. Consequently, medical imaging has played a very important role within the diagnosis for acquiring medical images that are presented to radiology experts for interpretation and investigation as shown in Figure 2.



FIGURE 2: Medical imaging for investigating and screening of breast lesions

Source; <https://nypost.com/2019/09/03/breast-cancer-treatment-could-lead-to-incurable-tumors-years-later/>

Mammography has been proven the most reliable and preferred tool employed by radiologists to screen and investigate suspicious breast lesions [13]. The common clinical modalities that have been used since the 2000s are the real-time 3D ultrasound imaging, 2D and 3D cone-beam CT and full-field digital mammography (FFDM). The techniques are categorized into film-screen mammography, which creates a scanned photographic film, and digital mammography. Particularly, the FFDM method replaces the X-ray films by electronics and creates digital images. The general process of screening takes a few moments and uses a low-dose X-ray system to see inside the breasts and visualize any abnormal clusters and irregular nodules. The patient having the mammogram places their breast between two clear plates to be squeezed and scanned into digital images [14] as illustrated in Figure 3. The machine should take a picture from two angles for bilateral comparison into a mediolateral oblique (MLO) view and craniocaudal (CC) view. Correct positioning is critical to capture the whole breast and find out any abnormal regions.



FIGURE 3: Mammography screening process

Mammograms visualize the whole breast tissue on the MLO projection on a single image where the nipple should be well-positioned at the lower level of the pectoralis major. The CC view should show the medial tissue without the axillary tail of the breast and with a pectoralis major situated at the center of the view. Samples of mammograms from MLO and CC views are demonstrated in Figure 4 and Figure 5.

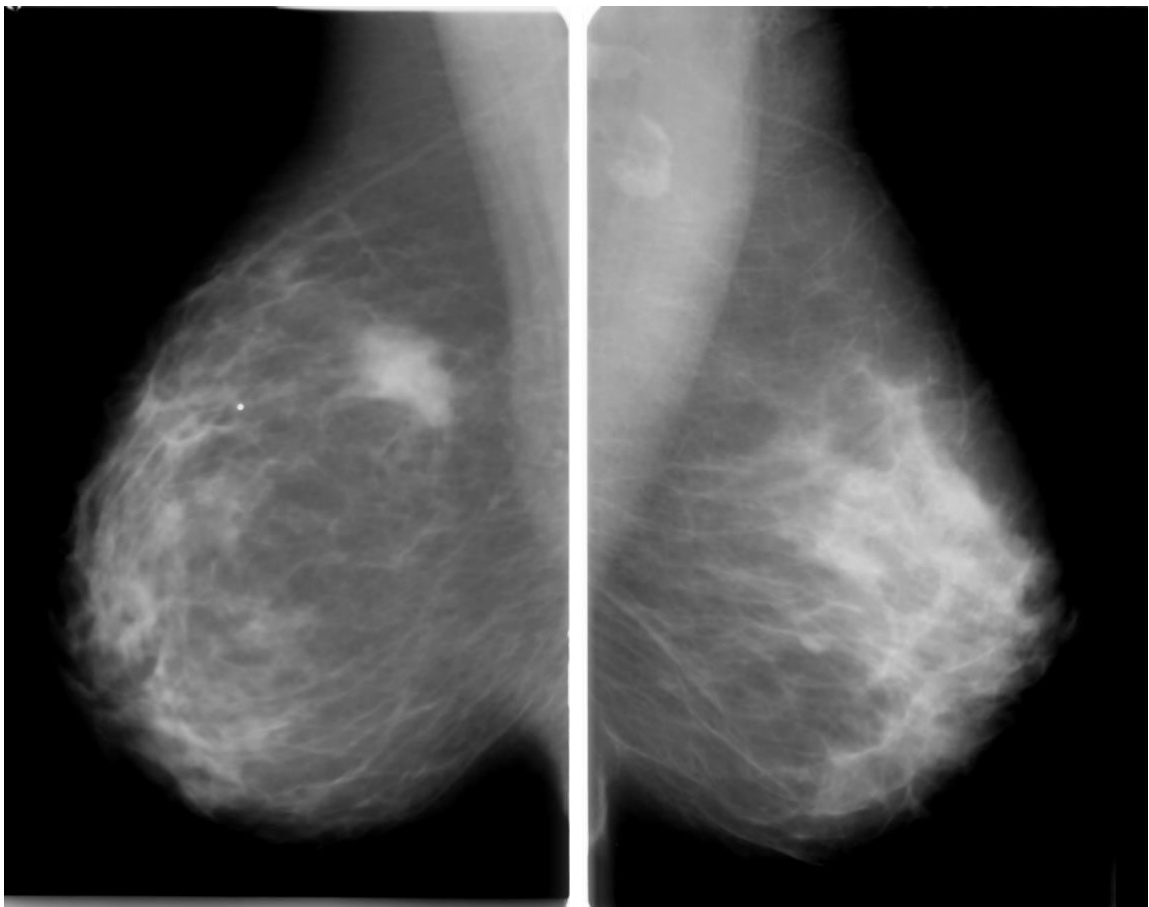


FIGURE 4: Samples mammograms of Mediolateral oblique (MLO) views

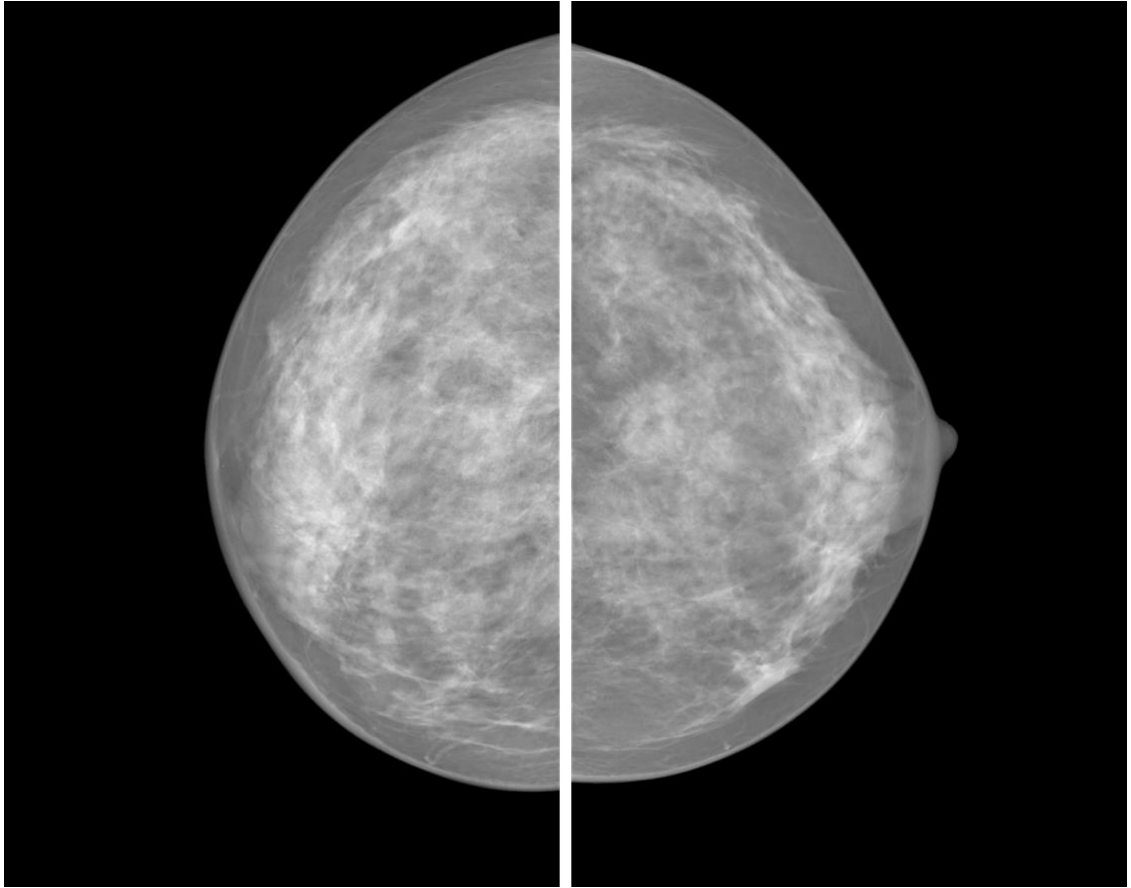


FIGURE 5: Samples mammograms of craniocaudal (CC) views

Radiologists should visually detect any area that does not look like normal tissue and may be a clear indicator of a cancer risk. During the diagnosis, the experts often search for areas of white, their position in the breast and note the density of the fatty tissue or/and the shape and size of the breast. A main concern taken from an observation occurs when a lump of a clear tumor appears as a condensed white area on a mammogram. A malignant tumor is likely to grow and change shape into a mass lesion. However, a benign tumor is usually non-cancerous and does not present a health risk and it is unlikely to grow. Other forms of irregular tissue can appear in the form of white spots and dots, and they are usually harmless. However, a continuous follow-up is required to track their shape and pattern that may lead to a sign of cancer. These abnormalities include calcification, asymmetry, architectural distortion, etc. Most of the non-mass lesions are benign and caused by summation of artifacts [15]. Samples below in Figure 6 present examples of mammograms that belong to public and private mammography datasets, and they were acquired using different modalities such as FFDM and digital mammographic screening.

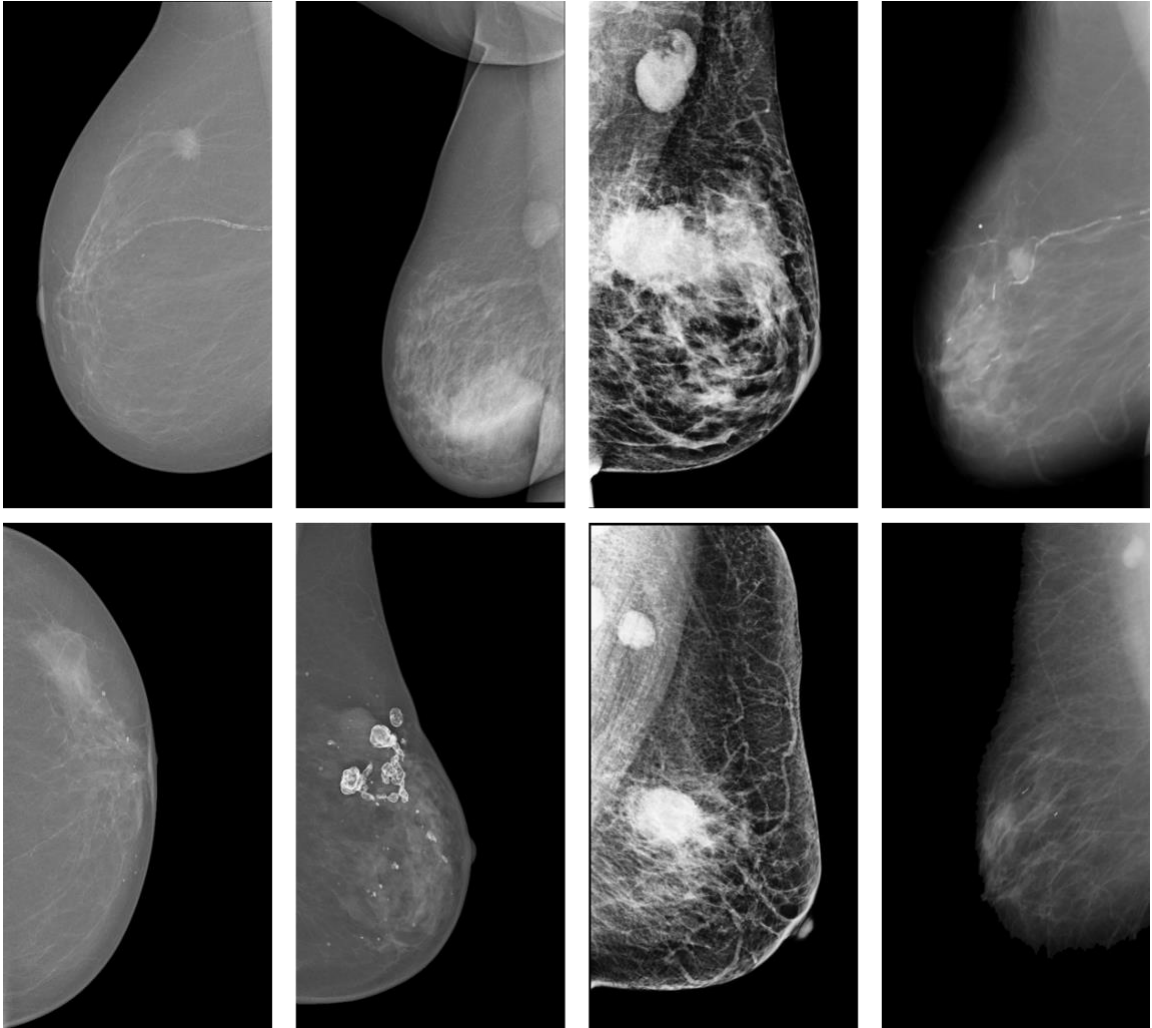


FIGURE 6: Samples mammograms with Mass lesions (first row) and with Calcification lesions (second row)

It is clear in Figure 6 that mammograms look different because they have different resolution and capture quality due to the different modality that was used to extract and save the images. Every mammogram may indicate one or more lesions of varying sizes and locations. Samples below present examples of different breast lesions that are commonly inspected in mammograms. Radiologists often investigate different screening views at different times and make a comparison to spot any unusual change or verify any suspected symptoms of breast cancer the patient has noticed.

Figure 7 shows samples of breast Mass lesions. Breast mass is a nodule or growth of tissue with defined borders. It is often characterized with a firm and hard area that causes a change in the tissue, the color, the size, or the density of the breast. A breast mass may be benign (non-cancerous) or malignant (harmful). A breast tumor can be confirmed with the feel (i.e., firm, margin, squishy, mobile), with

appearance in mammograms (i.e., fuzzy, lumpy, round, uniform), in an MRI (rapid or slow to light up), or with a biopsy (i.e., clusters of cells, irregular nuclei).

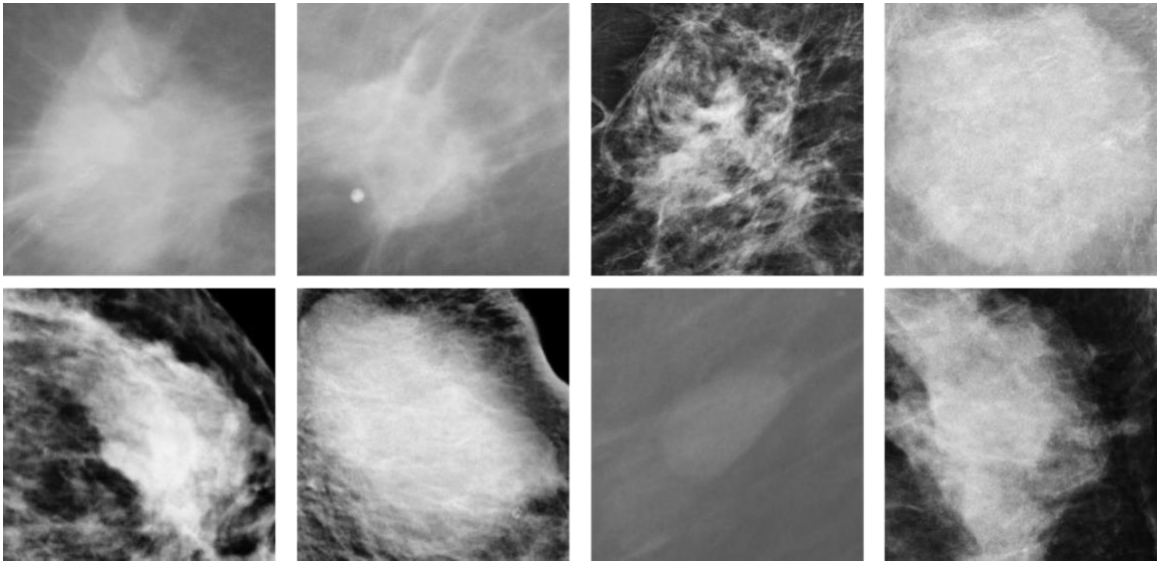


FIGURE 7: Samples of region of interest (ROI) OF breast Mass lesions from public and private datasets

Moreover, non-cancerous breast lesions like calcification and architectural distortion are shown in Figure 8 and Figure 9. Calcifications are often small deposits of calcium that are manifested in the form of white spots and dots. During the mammography, this built-up calcium absorbs the X-rays and shows up as a bright marker as shown in samples of Figure 8. Calcifications are not associated with breast cancer, and they are eventually benign cells, however they may be associated with presence of ductal carcinoma in situ [16]. Studies showed that a large calcification, known as macro calcification, is less specious. They should appear well-defined as dots and specks. On the other side, micro-calcifications are smaller and resemble mini flecks, but they require more follow-up.

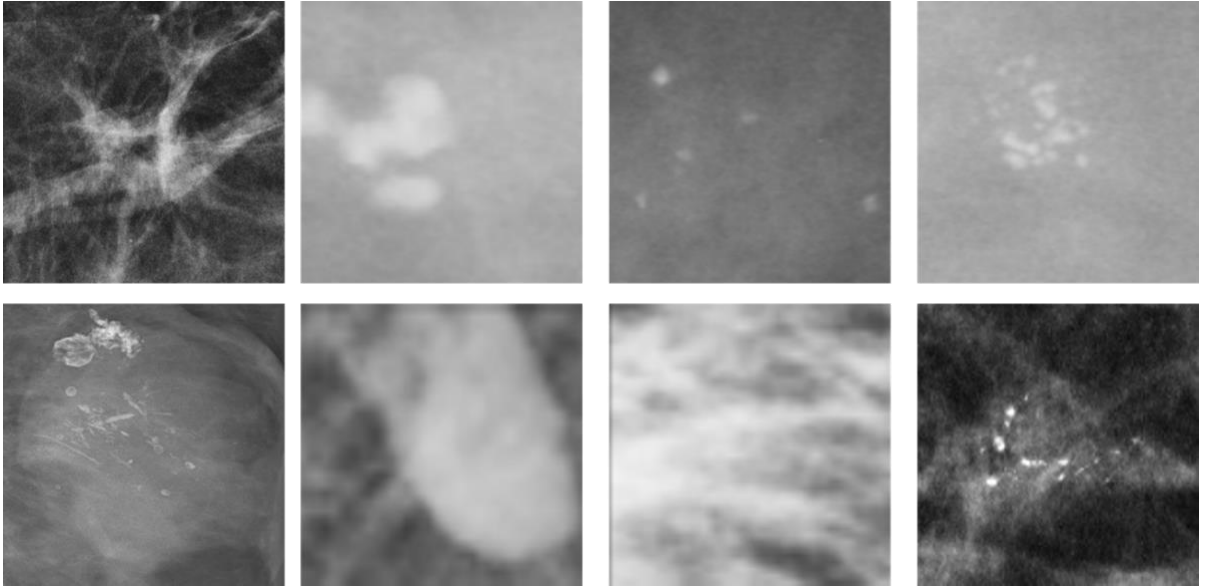


FIGURE 8: Samples of region of interest (ROI) OF breast Calcification lesions from public and private datasets

Other common harmless breast lesions are the architectural distortion that refers to clear distortion of the breast lobule without defining mass [17]. As shown in Figure 9, these non-cancerous lesions often appear as thin straight lines of spiculum in the form of dots, distortion, and focal retraction [18]. Architectural distortion is the third most common breast lesion, and it may be associated with calcifications or asymmetry, and it may represent an early indicator of breast cancer. Moreover, it is considered the hardest to diagnose, particularly in 2D mammography as it appears in variable presentation, size, and location.

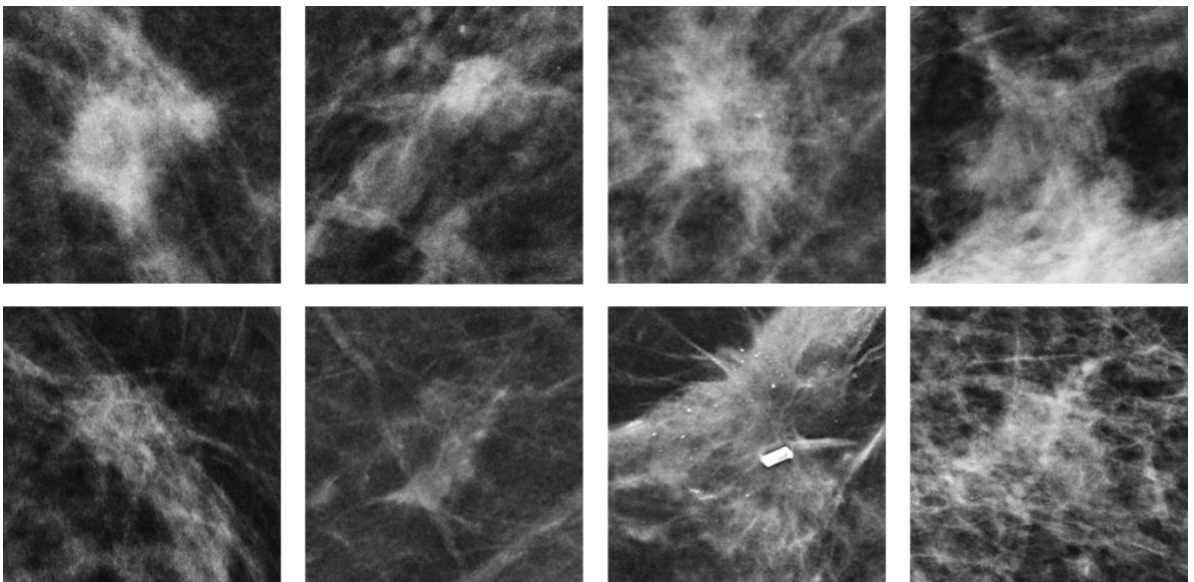


FIGURE 9: Samples of region of interest (ROI) OF breast Architectural Distortion lesions from public and private datasets

During the last decades, breast cancer reporting has been standardized using the Breast Imaging Reporting and Data System (BI-RADS) that is universally used by radiologists and experts for risk assessment and tumor grading. BI-RADS was developed by the American College of Radiology [19] that scaled the diagnosis assessment with a score from 0 to 6. The score is commonly used in all screening modalities to help determine the severity of the breast lesions and quantify the finding. It also contributes to the tracking of change in breast lesions and classifies the growth of the lesions. Table 1 summarizes the assessment categories and description.

Table 1: BI-RADS assessment categories

Category	Diagnosis assessment
0	Incomplete: Not enough information
1	Negative: No assessment
2	Benign
3	Probably benign: High probability of benign
4	Probably malignant: Reasonable probability of malignancy
5	High suggestive of malignancy
6	Known biopsy proven malignancy

Mammograms are used for detection and segmentation of abnormalities in the breast to remove the tumors background and provide affordable diagnosis like the tumor’s pathology (i.e., malignant, or benign) and classifying its shape and grading (i.e., BI-RADS score). Because most of these clinical tasks are usually conducted manually, their precision and efficiency are usually associated with the radiologist’s expertise. In fact, automated analysis of mammograms could enhance the reading and reduce the false detection and decision towards the breast cancer diagnosis.

C. Artificial Intelligence and CAD systems for Breast Cancer

Studies emphasized the importance of frequent mammography screening to reduce the mortality rate by early detecting the breast tumors before being spread to normal tissues and other healthy organs [20]. Therefore, mammograms are inspected every day by radiology experts to search for abnormal lesions and detect the location, shape, and type of any suspicious regions in the breast [21]. Although this process is

considered crucial and requires more precision and accuracy, it remains expensive and exposed to error, due to the increasing number of daily-screened mammograms [22].

During the screening process, radiologists should detect the abnormal lesions; determine their location on the breast mammogram, and identify their type to distinguish between mass, calcification, and other common abnormalities. Another task the medical experts conduct is the image segmentation, which helps doctors to extract detailed information of the suspicious regions of tumors for further diagnosis and pathology findings. Finally, the doctors should make a final decision regarding the tumors and determine their BI-RADS grading assessment and pathology diagnosis.

With the increase in the number of daily-screened mammograms, an efficient diagnostic methodology is necessary to assist doctors in the timely procedure of breast cancer. Thus, computer-aided diagnosis (CAD) systems could provide a second suggestion and reading to the final examination of the experts regarding the presence of breast cancer [23]. The automated systems are designed to perform computational image analysis and benefit from the high number of mammograms to handle the screening process automatically [24].

In the last years, the advance of Artificial Intelligence (AI) in computer vision applications has presented algorithms that showed remarkable results in developing tools to assist doctors. The systems could handle detecting, segmenting, and classifying tumors with the lowest possible error in many medical image applications and particularly in mammography [25, 26, 27].

Conventional systems relied on extracting hand-crafted and low-level features to localize and classify potential regions using simple image processing and machine learning techniques [28, 29, 30]. So far, these solutions have become inaccurate and resulted in a high false positive rate, and thus they have been substituted with the novel deep learning approaches [31, 32, 33, 34]. Accordingly, traditional techniques for tumor segmentation, such as region-growing, active contour, and watershed, relied on extracting handcrafted features that only represent gray-level, texture, and morphology to label the pixels and indicate the contour surrounding the mass tumors, while excluding the background tissue [35, 36]. Moreover, different algorithms have been widely implemented in CAD systems for tumor classification, and the most used algorithms are conventional machine learning classifiers and threshold-based methods that are based on handcrafted features [35, 36, 37].

With the continuous enhancement of computational capacity of computers, different deep learning models have been widely implemented to offer a better alternative. They aim to automatically extract deep and high-level features directly from raw images without knowledge requirement [38]. It has helped to improve results of automated systems and maintain a good tradeoff between precision of lesions detection and accuracy of distinguishing between different types of lesions from a single mammogram [39, 40, 41]. Deep learning algorithms can extract deep and multiple-scaled information and integrate them to help experts in making final decisions. Accordingly, their strength to adapt to different cases has been proved for objects detection and classification tasks in many applications [42, 43, 44, 45].

The development of computer-aided diagnosis (CAD) for breast cancer imaging has lately been revised to deal with the rapid rise of deep learning algorithms and artificial intelligence, and it emphasizes innovative systems that may have significant potential to improve clinical treatment [46, 47, 48].

With the practical challenges that breast tumors offer due to their variation in size, shape, location, and texture, there has been a significant need to improve the overall performance of CAD systems and reduce false positive and negative cases. Consequently, deep learning methodology has been broadly suggested in biomedical applications and particularly in CAD systems for mammography [49, 50, 51].

In the last two decades, deep learning has shown a growing success in many computers vision tasks and has proven a capability to overcome complex problems in the medical imaging domain. As a result, several works have been suggested and applied particularly in mammography, such as for tumors detection [52, 53], breast lesions segmentation [54, 55], and classification [56, 57].

D. Dissertation organization

The dissertation consists of five main chapters as detailed in the following paragraph:

- Chapter I presents the scope of the research, and it introduces the motivations taken from applying medical imaging for breast cancer. It also presents the mammography process and the role of AI in CAD systems.
- Chapter II surveys the related works and the state-of-the-art methodologies of computer vision for CAD applications. It also reveals the limitations and challenges of AI techniques and deep learning technology for breast cancer diagnosis and highlights the potential solutions for improving the CAD system performance.

- Chapter III focuses on the proposed methods and techniques presented towards an integrated framework for a CAD system. It details the novel implementations and deep learning architectures that were proposed for each step defining the entire CAD system for breast cancer diagnosis.
- Chapter IV details the used mammography datasets and the setting and metrics employed for the experimental evaluation. Before presenting the general discussion and the comparative results, it illustrates a testing and deployment section for showcase of the integrated CAD system.
- Chapter V summarizes the entire work and the proposed methodologies and lays out ten future works.

CHAPTER II

LITTERATURE REVIEW MODELS AND METHODS FOR BREAST CANCER DIAGNOSIS

The present chapter highlights the recent success of several research studies on developing automated solutions using cutting-edge technology for many medical applications and particularly breast cancer. Computer vision and AI technology have been merged to produce feasible Computer-aided diagnosis (CAD) systems to assist doctors during the diagnosis task. As a medical imaging application, an automated CAD system could be beneficial to save the time on the screening procedure and provide a second reading that helps reveal any hidden pattern and indication of cancer that could be overlooked. The CAD system could also contribute to lowering the mortality rate and decreasing the false positive and negative diagnosis.

All the surveyed works in this chapter are based on image analysis and processing methods, pattern recognition techniques, and machine learning and deep learning technologies. The chapter is organized according to the main steps that form the integrated framework of a CAD system: tumor detection and identification, tumor segmentation and background masking, and tumor classification and diagnosis that presents the final decision. The tasks were conducted in the order we mentioned so that the segmentation step was applied on the detected region of interest (ROI) of tumors, and similarly the classification and diagnosis step was applied on the detected and segmented ROI of tumors.

In this chapter, the state-of-the-art methods and architecture models are presented with their applications and performance results that will be later compared to our proposed methodologies.

A. Introduction

Breast cancer is the most common type of cancer that is leading to death among women, where 43,250 death cases were reported in the United States in 2022 and it represents a rate of 15% of estimated deaths against the other types of cancer. However, the overall death rate decreased by 1% per year from 2013

to 2020 because of treatment advancements and early diagnosis through screening. The median age of women to have cancer is 60 but the risk may occur from the age of 30 to 50. Early detection and diagnosis of breast cancer is the most effective way to treat the disease to postpone its spread to other healthy organs. The process of screening may reveal small indicators of abnormalities that may lead later to risky factors of cancer. The most common screening procedure is the digital mammography that takes snapshot images of the breast composition and visualizes any irregular pattern for breast cancer.

Due to the increasing number of daily screening mammograms, automated machine could benefit from the huge amount of the stored medical images and develop a CAD system that performs computational image analysis and advance features extraction to automatically decide toward the location of the abnormal tumors, their type, pathology, shape, and assessment.

Accordingly, as shown in Figure 10, a completely integrated CAD system would start its first stage, with the detection and localization of suspicious lesions and distinguishing between their types, i.e., mass, calcification, architectural distortion, etc. Then, at a second stage, the CAD system should perform a segmentation of the obtained region of interest (ROI) surrounding the breast lesion to recognize its anatomical contour and remove its tissue background without losing its shape precision. Finally, diagnostic information could be extracted regarding the pathology to classify the decided lesion as either malignant or benign and identify its characteristics such as a tumor grading using Breast Imaging Reporting and Data System (BI-RADS) score, or/and a shape categorization. As the automated procedure relies on connected stages, each output information must be generated precisely to generate a fast and accurate final decision.

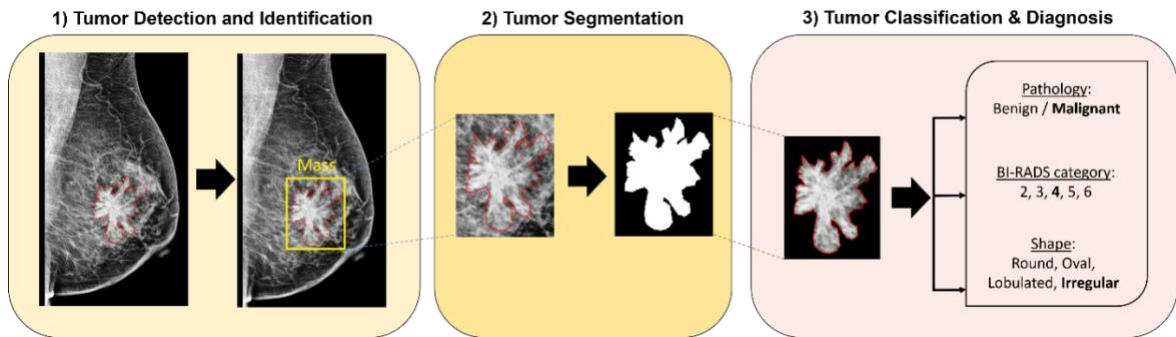


FIGURE 10: Integrated framework of CAD system for breast cancer (A mammogram case with a malignant mass lesion, BI-RADS score of 4 and irregular shape).

The rest of the chapter details recent works that were proposed for each step of the CAD system showing recent techniques and models for implementing every task accordingly. Each subsection compares the performance results for the conducted task and presents their advantages and limitations.

B. Breast Cancer Detection and Identification

Since its discovery in 1913, mammography has been considered an essential key for early detection and diagnosis of specious lesions. Mammography screening has helped radiologists identify breast cancer and several studies showed its impact for a significant reduction in mortality rate [58]. With the remarkable advances in computer vision and artificial intelligence to assist doctors for medical imaging analysis, many studies showed the effectiveness of CAD systems to automatically detect suspicious lesions from raw screened mammograms [59]. The introduction of neural network models changed the CAD's approach and substituted the use of hand-crafted features extraction with deep learning architectures that are capable of learning complex features at different scales [60].

With the development of machine learning technology, many applications have given more attention in adopting deep learning to solve complex problems, particularly in the fields of computer vision, image recognition, object detection [32, 38, 39] and segmentation [40, 41]. Due to the limitations of hand-crafted features derived from raw images, many studies have shown that traditional methodologies have tried to establish highly accurate models. Indeed, traditional CAD systems that were proposed for breast lesions detection and classification could not overcome the huge variations in lesions size and texture, compared to deep learning methods [51, 52]. As a result, several CAD systems based on deep learning architectures have been successfully developed to improve the detection and classification of organ lesions such as liver lesions, lung nodules, and, in particular, breast lesions [57, 58].

Researchers have demonstrated the feasibility of regional-based models to build an end-to-end system for detecting and classifying malignant and benign tumors in the INbreast mammograms and achieved a detection rate of 89.4% [52]. Peng et al. provided an alternative approach in a recent paper. [61] that introduced an automated mass detection approach. The method integrated the Faster R-CNN model and multiscale-feature pyramid network and yielded a true positive rate of 0.93 on CBIS-DDSM and 0.95 on INbreast dataset.

Accordingly, Al-Antari et al. [62] employed the YOLO model for breast masses detection that reported a detection accuracy of 98.96%. The output served after that for mass segmentation and recognition to provide a fully integrated CAD system for digital X-ray mammograms. Another work by Al-Antari et al. [63] in 2020 improved the results of the breast lesions detection and classification by adopting first the YOLO model for detection and then compared feedforward CNN, ResNet-50, and InceptionResNet-V2 for classification. Similarly, Al-masni et al. [49] proposed a CAD system framework that first detected breast masses using YOLO model with an overall accuracy of 99.7%, and then classified them into malignant and benign using FC-NNs (Fully Connected Neural Networks) that achieved with an accuracy of 97%.

Deep convolutional neural networks (DCNN) were also suggested for mammographic mass detection by using transfer-learning strategy from natural images [64]. Ribli et al. presented a paper in 2018 [65] and proposed a CAD system based on Faster R-CNN framework to detect and classify malignant and benign lesions and obtained an AUC score of 0.95 on INbreast dataset. Another work employed a fully convolutional network (FCN) with adversarial learning in an unsupervised fashion to align different domains while conducting mass detection in mammograms [66].

Indeed, the breast tumor detection task is a crucial step that remains a challenge for CAD systems where many reliable models were used to support the automatic diagnosis. For example, Singh et al. relied on the Single Shot Detector (SSD) model to localize tumors in mammograms, and then extracted output boxes to apply segmentation and classification tasks [67]. It yielded a sufficient true positive rate of 0.97 on the INbreast dataset. Other recent studies proposed using the YOLO model to achieve a better performance in detecting bounding boxes surrounding breast tumors. For example, Al-masni et al. [68] presented a YOLO-based CAD system that achieved an overall accuracy of 85.52% on DDSM dataset.

Moreover, a tumor localization task was conducted in a detection framework for cancer metastasis using a patch-based classification stage and a heatmap-based post-processing stage [69]. It achieved a score of 0.7051 and served for whole slide image classification. Breast tumor detection was also addressed in 2016 by Akselrod-Ballin et al. [70] where images were divided into overlapped patches and fed into a cascaded R-CNN model to first detect masses and then classify them into malignant or benign. Dhungel et al. presented research in 2015 [71] and relied on a multi-scale Deep Belief Network (DBN) to first extract all suspicious regions from entire mammograms and then filter out the best regions using Random Forest (RF). The

technique achieved a true positive rate of 96%. Akselrod-Ballin et al. presented a paper in 2017 [72] and developed a three-stage cascade of Faster-RCNN models to detect and classify abnormal regions in mammograms. The work was evaluated on the INbreast dataset and achieved an overall detection and classification accuracy of 72% and 77%.

In another work by Li et al. [73], a bilateral mass detection method was introduced using two networks: a registration network between left and right breasts and a Siamese-Faster-RCNN network to detect masses from pairs of registered mammograms. They reported results of a true positive rate of 0.88 on the INbreast dataset and 0.85 on a private dataset. In addition, Hamed et al. [74] presented a YOLOV4-based CAD system with 2-path detection of masses in full and cropped mammograms and then classified them into benign and malignant.

Moreover, early detection and diagnosis of breast cancer in mammography using the deep learning-based CAD systems can help prevent development of tumors by marking lesions, and thus it can effectively decrease death rate [75]. A retrospective study by Watanabe et al. [76] showed a potential area of improvement for radiologists' interpretation of screening mammograms for early detection using Artificial Intelligence. The studied CAD system succeeded to mark 30 (86%) of 35 missed micro-calcifications and 58 (73%) of 80 missed masses. In addition, missed malignant lesions were flagged as early as 70 months Prior to recall or diagnostic follow-up. In consequence, CAD systems could benefit from the change that occurred between Prior and Current mammographic exams. A recent work by Timp et al. [77] tried to improve the characterization of mass lesions by adding information about the tumor behavior over time. The authors presented a CAD program to detect temporal changes between two consecutive screening images using a regional registration method to localize lesions detected on the current views and their corresponding on the Prior views. After that, a Support Vector Machines (SVM) classifier was applied to show the effectiveness of temporal features. In a different study, Timp et al. [78] attempted to improve detection methods by including temporal information in the CAD system. A regional registration technique along feature space was used to map suspicious locations on the Current mammograms with a corresponding location on the Prior mammograms with 72% accuracy. Accordingly, Loizidou et al. [79] tried to increase the microcalcification detection accuracy to 99.2% by adding temporal subtraction between mammogram pairs before applying SVMs classifier. In the same context, a recent study by Loizidou et al. [80] extended their

previous work of breast micro-classification detection and classification by adding an image registration step of Prior mammograms before applying temporal subtraction of pairs. In a different work by Zheng et al. [81], follow-up digital mammography images were integrated together to develop a CAD method for breast cancer detection. All regional images were detected using the Haar features, local binary pattern, and histogram of oriented gradient via the AdaBoost approach and then fed into a CNN to filter out the false positive cases.

Most of the reviewed works and their diagnosis results showed how artificial intelligence has successfully contributed to solve the challenge of breast cancer detection. In fact, practical implementation, and system evaluation, as well as the tremendous complexity of memory and time, remain unsolved issues. Most of the works have tackled the problem of detecting only mass tumors in the entire breast and then classifying them into malignant and benign. Our approach was developed differently to address the task of detection and identification of at least two types of breast lesions (i.e., mass and calcification). Our methodology was also expanded by presenting a fusion-models approach that combines predictions of different models to improve the results.

C. Breast Cancer Segmentation

With the significant advance in computer vision applications, several studies have focused on developing deep learning methods for the detection and segmentation of breast mass tumors [82, 83, 84]. Convolutional Neural Networks (CNNs) were among the first architectures that attempted to automatically learn features from raw images to label pixels surrounding objects at different scales and shapes [85].

Regarding medical image segmentation, one of the well-known architectures is the Fully Convolutional Network (FCN), which introduced the first encoder-decoder network design. In fact, the encoder part applies convolutional, max-pooling and down-sampling layers and finally fully connected layers of VGG-16 to end up with low-level features. The decoder part is like the encoder part by using a deconvolutional and up-sampling layer [86]. The network was introduced by Hai et al. [87] that added multiscale image information for automated breast tumor segmentation, and it achieved a Dice score of 76.97% and an IoU score of 60.41% on a private dataset. Al-antari et al. introduced the Full resolution Convolutional Network (FrCN) [62], which is a version of the FCN, to segment the detected breast masses, and it produced a Dice score of 92.69% and a Jaccard similarity coefficient of 86.37% on INbreast dataset. Accordingly, Zhu et al. [88] employed a multi-scale FCN model followed by a conditional random field

(CRF) for mammographic mass segmentation, and they achieved Dice score of 90.97% on the INbreast dataset and 91.30% on the DDSM-BCRP dataset. Another work proposed by Singh et al. [89] was inspired by the FCN architecture and developed a conditional Generative Adversarial Network (cGAN) for breast tumor segmentation. The work achieved a Dice score of 92.11% and an IoU score of 84.55% on the INbreast dataset and a Dice score of 88.12% and an IoU score of 79.87% on a private dataset.

Medical image segmentation usually presents challenging cases; and consequently, this type of network suffers from low segmentation accuracy due to the loss of spatial resolution in case of small objects and irregular shapes. As a result, Ronneberger et al. developed a new model called UNet [90] to overcome the limitation of FCN models. UNet proposed to integrate the high-level features from the decoder, which preserve the semantic information, with the low-level features from the encoder, which represent the spatial information. The fusion was maintained with the skip connections that made the UNet architecture adequate in several medical applications and particularly in mammography. A work presented by Sathyan et al. [91] employed the UNet model to segment mass and calcification in whole mammograms and achieved a Dice score of 67.3% on the CBIS-DDSM dataset. Another work by Soulami et al. [92] relied on an end-to-end UNet model for the detection, segmentation, and classification of breast masses in one stage, where the segmentation evaluation showed a Dice score of 90.5% for both DDSM and INbreast datasets.

Many potential scopes of improvement were recently suggested on the UNet architecture to improve its performance and enhance the quality of the segmentation [93]. Inspired by the residual mechanism, Abdelhafiz et al. [94] proposed the residual UNet, called RUNet or ResUNet, by adding residual blocks to the standard convolutional layer in the encoder pathway to add a deeper effect to the network. The work was applied for mass segmentation and then detected binary maps were fed to a ResNet model for classification into benign or malignant. Segmentation results yielded a Dice score of 90.5% and a mean IoU score of 89.1% on the INbreast dataset. Similarly, Ibtehaz et al. [95] developed a novel architecture, called MultiResUNet, which showed a remarkable gain in performance for biomedical image datasets. The work replaced the standard blocks of the encoder-decoder by multi-residual blocks and altered the plain skip connection with residual paths. Accordingly, a Conditional Residual UNet, called CRUNet, was also suggested by Li et al. [96] to improve the performance of the standard UNet for breast mass segmentation, and it achieved a Dice score of 92.72% on the INbreast dataset.

Another variation of the UNet was suggested using the attention mechanism that showed remarkable success in medical image segmentation [97]. Consequently, Oktay et al. [98] integrated the attention gate into the standard UNet to propose a new Attention UNet, called AUNet. It improved the prediction performance across CT pancreas segmentation and yielded a Dice score of 83.1%. Similarly, Li et al. [99] developed an Attention dense UNet for breast mass segmentation that achieved better results than UNet, AUNet and DenseNet with a Dice score of 82.24% on the DDSM database. In another work suggested by Sun et al. [100], a novel attention-guided dense-upsampling network was developed for breast mass segmentation in whole mammograms. The architecture achieved a Dice score of 81.8% on the CBIS-DDSM dataset and 79.1% on the INbreast dataset. Similarly, Ravitha Rajalakshmi et al. [101] introduced a deeply supervised U-Net model (DS U-Net) associated with dense CRFs to segment suspicious regions on mammograms. The model was tested and gave a Dice score of 82.9% and 79% respectively on CBIS-DDSM and INbreast datasets.

Aligned with the improvement made in encoder-decoder architecture to deal with the limitations encountered in medical images segmentation, the Atrous Spatial Pyramid Pooling (ASPP) module was successfully integrated in many networks [102]. It showed effectiveness in breast mass segmentation in a work presented by Wang et al. [103] that achieved a Dice score of 91.10% and 91.69%, respectively on the INbreast and DDSM-BCRP datasets.

The depth of the UNet architecture was exposed through studying it, as well as the limited design of skip connections. Therefore, a novel architecture, named UNet++, was introduced by Zhou et al. [104] to alleviate the network depth and redesign the standard skip connections. The research was tested on six medical image datasets with diverse modalities, and it showed similar results for semantic and instance segmentation tasks. A similar variation model, called U-Net+, was employed by Tsochatzidis et al. [105] to segment ROI mass before integrating it with the classification stage by a CNN model. On the CBIS-DDSM and DDSM-400 datasets, the segmentation performance had a Dice score of 0.722 and 0.738, and a Jaccard index of 0.565 and 0.585, respectively.

Furthermore, to cope with difficult medical imagery, Jha et al. [106] presented a DoubleUNets architecture that uses two encoders and two decoders in sequence and an ASPP module. On four medical segmentation datasets, the network outperformed both the baselines and UNet Das et al. [107] proposed a

Contour-Aware Residual W-Net, termed WRC-Net, which consists of double UNets. The first UNet was designed to predict objects boundaries and the second UNet generated the segmentation map. The work was evaluated on stained cell images, and it outperformed the state-of-the-art nuclei segmentation techniques.

Additionally, a new variation of the UNet was presented by Tran et al. [108], named TMD-UNet, that modified the interconnection of the network node, replaced the standard convolutions with dilated convolutions layers, and developed new dense skip connections. For liver, polyp, skin lesion, spleen, nuclei, and left atrium segmentation, the network outperformed popular models.

D. Image-to-image Translation

With the significant attention given on improving the performance of neural networks algorithms, many studies have focused on enhancing the quality of medical images that are acquired using multiple imaging modalities. Lately, image-to-image translation has been employed to solve many computer vision applications in medical imaging, such as image synthesis and reconstruction.

In fact, deep learning models require a large amount of annotated data to achieve a robust behavior. However, it is often difficult for medical applications to collect enough instances. Therefore, synthetic data was recently adopted to increase the size of dataset, within either the same image modality, or using cross-modality translation. In a work presented by Senaras et al. [109], a conditional Generative Adversarial Network was applied to generate synthetic images that were indistinguishable from real images of a breast cancer database. Accordingly, Alyafi et al. [110] employed Deep Convolutional GAN (DCGAN) to generate synthetic mammograms with mass lesions to enhance the classification performance in imbalanced datasets. Another recent technique that was widely used for unpaired image-to-image translation is the cycle-consistent Generative Adversarial Network (CycleGAN), and it was developed by Zhu et al. [111]. The technique learns two mappings by transforming images between two different domains using two GANS and maintains their reconstruction by a cycle-consistency loss and hence the name. Becker et al. [112] presented CycleGAN model to artificially inject or remove suspicious features and thus increase the size of the BCDR and INbreast datasets. Moreover, a cross-modality synthesis approach was introduced by Cai et al. [113]. It was motivated by CycleGAN between CT and MRI, and it was tested on 2D/3D images for segmentation.

Another similar work by Hiasa et al. [114] extended the CycleGAN approach by adding gradient consistency loss and aimed for MRI-to-CT synthesis. The work yielded a better segmentation precision on

musculoskeletal images. Upon such an idea, Huo et al. [115] proposed an end-to-end synthesis and segmentation network (EssNet) to conduct the unpaired MRI to CT images synthesis and CT splenomegaly segmentation without using manually annotated CT. It achieved a higher Dice score of 91.88% than the state-of-the-art performance.

A similar architecture called Pix2Pix was also suggested for image-to-image translation but using paired datasets [116]. A recent application by Shen et al. [117] employed the Pix2Pix network for image-to-mask segmentation in mammography. Pix2pix was also employed by Liao et al. [118] to artificially remove artifacts in CT scans and the method showed improvement for clinical image reconstruction.

E. Breast Cancer Pathology & Diagnosis Classification

Computer-aided diagnosis (CAD) systems have recently used advanced computer vision techniques, artificial intelligence, and machine learning approaches to automatically extract information from digital medical images to assist experts in their final diagnostic decisions. Several research studies have attempted to suggest methods to assist doctors in their diagnosis and have focused on improving results of breast mass classification in digital mammography. In this context, Dhahri et al. [119] used a Tabu search to select the most significant features and then fed them into a K-Nearest Neighbors (KNN) algorithm to classify breast lesions into malignant or benign. This methodology required four iterative stages and achieved an accuracy of 98.24% on the WDBC dataset.

Since their development, many studies have given more attention to incorporating deep learning methods in CAD systems as they showed better efficiency than traditional CAD systems that require extensive features extraction. For instance, an end-to-end approach was developed by Shen et al. [38] to classify whole-digital mammograms into cancer or normal. The work demonstrated a new CNN structure using VGG network and the residual network (ResNet) and accomplished an AUC of 0.91 on the CBIS-DDSM dataset and an AUC of 0.98 on the INbreast dataset. Another end-to-end model, called DiaGRAM, was built by Shams et al. [120] that combined CNN and Generative Adversarial Networks (GAN). The work was conducted to classify mammograms as benign or cancerous. Experiments showed an accuracy of 89% on the DDSM dataset and 93.5% on the INbreast dataset. An improved deep learning method, called the new DenseNet-II model, was invented in a work by Li et al. [121] for classification of benign and malignant mammography images. The model was analyzed and applied on a private collection of mammography images

and reached an accuracy of 94.55%. Accordingly, a novel model for mass classification was proposed by Zhang et al. [122] that fused texture features from local binary pattern extraction, with deep CNN features. The model was tested on a CBIS-DDSM dataset and achieved an accuracy of 94.30%. In another work by Muramatsu et al. [123], a CNN model's performance was improved by adding synthetic data generated from lung nodules in computed tomography (CT) using cycle generative adversarial networks. The classification performance was tested on a DDSM dataset and achieved an accuracy of 81.4%. Recently, Chakravarthy et al. [124] proposed a novel method that coordinates the deep learning with an extreme learning machine (ELM) optimized by employing a basic crow-search algorithm for classifying ROI images with abnormality into malignant or benign. The proposed work achieved a maximum overall accuracy of 97.19% on DDSM, 98.13% on the Mammographic Image Analysis Society (MIAS) dataset and 98.26% on INbreast datasets. In a recent work by Khan et al. [125], multi-view feature fusion (MVFF) based-CAD system was implemented to increase the performance of CNN by combining information of four views of mammograms to classify them first into abnormal or normal, then into mass or calcification, and finally into malignant or benign. The final pathology classification achieved an AUC of 0.84 on the CBIS-DDSM and mini-MIAS databases.

Moreover, Kumar et al. [126] suggested a classification framework for breast density using an ensemble of 4-class neural network classifiers. The work showed an overall classification accuracy of 90.8% on the DDSM dataset. Besides ensemble learning methodology, transfer learning was also adapted with deep learning techniques to develop an approach for differentiation between benign and malignant breast cancer cases in X-ray images [127]. Hence, in a work by Alkhaleefah et al. [128], double-shot transfer learning (DSTL) was used by fine-tuning various pre-trained networks once on an ImageNet dataset, and another time on a larger dataset like the target dataset. The method was trained on CBIS-DDSM and showed a better performance than single-shot transfer learning with an average AUC of 0.99 on the MIAS dataset and 0.94% on the BCDR dataset. Similarly, Falconí et al. [129] used transfer learning on a NasNet Mobile model and fine tune on VGG models to classify mammogram images according to the BI-RADS scale. The work achieved an accuracy of 90.9% on the INbreast dataset. Recently, a work by Medeiro et al [130] combined DenseNet201 and multi-perceptron layer (MLP) models to classify the pathology within BI-RADS levels 3 and 4 for malignancy of breast masses. The model achieved an accuracy of 63% surpassing the performance of a human expert by 9.0%.

To accomplish an efficient mass classification and diagnosis procedure, researchers have shown that capturing texture and morphological characteristics could help doctors understand the nature of the breast tumor and assess its malignancy scale. For instance, research by Bi et al. [131] showed that the probability of malignancy is highly correlated with the shape and morphology of a breast lesion. Therefore, several works have incorporated the segmentation stage to provide a complete, significant diagnosis [132]. In a previous work by Tsochatzidis et al. [105] modified convolutional layers of a CNN to integrate both input images and their corresponding segmentation maps to improve the diagnosis of breast cancer in mammograms. The method was applied on DDSM-400 and CBIS-DDSM datasets and achieved a diagnosis performance AUC of 0.89 and 0.86. Similarly, a dual convolutional neural network was suggested by Li et al. [133], which computed the mass segmentation and simultaneously predicted the diagnosis results. The model contributed an improvement to the mass segmentation and cancer classification problem at the same time and achieved an AUC of 0.85 on the DDSM dataset and 0.93 on the INbreast dataset.

Recently, most of the developed CAD systems have automated the breast cancer diagnosis procedure that gets an entire mammogram image and returns a final diagnosis. Thus, many studies have integrated the first stage of identifying the suspicious region of breast lesions and based on its automated output, performed the segmentation and classification tasks. For instance, Sarkar et al. [134] proposed an automated CAD system that detects suspicious regions of potential lesions using a deep hierarchical prediction network and then classifies them into mass or non-mass, and finally into malignant or benign using a CNN structure. The work was tested and achieved an accuracy of 98.05% on the DDSM dataset and 98.14% on the INbreast dataset. Another fully automated system by Dhungel et al. [135] for breast mass classification integrated mass detection and segmentation in a complete CAD system. The methodology used a multi-scale deep belief network (m-DBN) classifier followed by a cascade of CNNs and random forest classifiers for false positive reduction for mass detection, a conditional random field (CRF) for mass segmentation, and a multi-view deep residual neural network (mResNet) for mass classification. The proposed work achieved an AUC of 0.8 on the INbreast dataset. Another recent work by Singh et al. [136] presented automatic workflow that detects breast tumor regions from mammograms using the Single Shot Detector (SSD), and then outlines its segmented mask using conditional Generative Adversarial Network (cGAN) that was finally used for shape classification using a CNN. The framework achieved an overall accuracy of 80% for the shape classification.

Similarly, Al-Antari et al. [62] proposed a fully integrated CAD system for digital mammograms via deep learning techniques. It started with a mass detection using the You-Only Look Once (YOLO) architecture model, then performed a mass segmentation on the detected regions using Full resolution convolutional network (FrCN), and finally classified the detected and segmented masses into benign or malignant using a CNN model. The entire framework had an overall classification accuracy of 95.64% on the INbreast dataset. The mass classification step was differently solved in a recent work of Al-Antari et al. [137] that separately adopted three conventional deep learning models including regular feedforward CNN, ResNet-50, and InceptionResNet-V2. The work achieved a maximum accuracy of 95.32% on the INbreast dataset.

F. Conclusion and discussion

The reviewed state-of-the-art works highlighted models that have proved outstanding success on medical images. These models were evolved from a simple Convolutional Neural Networks (CNNs) model to present other variations such as R-CNNs, Fast CNNs and Faster R-CNNs Models. These popular models have overcome many limitations of deep learning such as computational time, redundancy, overfitting, and parameter size. However, most of these models take a long time to train and implement and require a lot of computing resources. Therefore, another variation called You-Only-Look-Once (YOLO), which is characterized with a low-memory dependence, has been recently recognized as a fast object detection model and suitable for CAD systems.

In this dissertation, we propose an end-to-end system that is based on the YOLO-based model to simultaneously detect and identify breast lesions into two most common abnormalities, Mass tumors or Calcification. Our approach contributes with a new feature, which is an end-to-end system that can recognize both types of suspicious lesions whether only one type exists in an image, or both simultaneously appear in the same image. As the choice of YOLO model was stated earlier, the proposed implementation will also serve as a basic step for the integrated framework for CAD system to present a complete breast cancer diagnosis.

The performance of the first step was demonstrated on different mammography datasets using deep learning methodologies (i.e., data augmentation, early stopping, hyperparameters tuning and transfer learning). An additional contribution was presented in this work to boost the lesions detection and identification performance as follows. Because the performance varies according to the input data of the

model, single evaluation results were first reported over the variations of images, then different fusion models were developed to increase the final detection accuracy rate and join models with different configurations. The suggested methodology will ensure the best-detected bounding boxes and remove the bad predictions that can mislead the future diagnostic tasks. Experiments were conducted on two most widely used datasets: CBIS-DDSM and INbreast, and on an independent private dataset. The outcome of this work will justify the performance of the YOLO-based model for deep learning lesion detection and identification on mammography. It will also serve as a comparison study of YOLO-based model performance on various mammograms.

Inspired by the reviewed works and their diagnosis results, we expanded the study on another mammography dataset that includes Current and Prior screening views for conducting an early detection task retrospectively. We first apply the YOLO-based fusion models on the most recent screening mammograms to detect and identify three types of breast lesions (i.e., Mass, Calcification, Architectural distortion). Second, we suggest replicating the early-screened mammograms with cancer diagnosis and maintaining prior shape and appearance into synthetic Prior mammograms that resemble the Current mammograms. The model that was trained on the Current views will be inferred retrospectively on the new Prior views to predict the suspicious findings that were missed at initial screening.

We evaluated two state-of-the-art techniques for image-to-image translation, CycleGAN and Pix2Pix and compared their performance on predicting location and type of lesions on Prior mammograms at early screening.

Motivated by the surveyed works on breast tumors segmentation, we suggest a novel architecture that incorporates all the recent approaches that were suggested to improve the breast masses segmentation task. We elaborate a new segmentation architecture that is based on the standard UNet and its two variations. Our approach also aims to evoke the CycleGAN method for a synthetic data augmentation between different mammography datasets to make benefit from the similar domains.

The second step of the CAD system was applied only on the detected Mass lesions after filtering their type and location from the mammograms. Other abnormalities were discarded for the lack of their ground truth diagnostic information.

Inspired by the continuous success of the CNN model and its variations for breast mass classification, and the high performance of the ensemble model technique on machine learning, we propose a stacked ensemble of residual network (ResNet) models to classify and diagnose previously detected and segmented mass lesions. The proposed model uses three different architectures of the ResNet model, ResNet50V2, ResNet101V2, and ResNet152V2 that are transferred and fine-tuned on the public and private mammography datasets. The models' layers are stacked together and reconfigured into a new entire model for an overall classification and diagnosis of 1) the pathology as malignant or benign; 2) the BI-RADS category as assessment score from 2 to 6; and 3) the shape as round, oval, lobulated, or irregular. All surveyed works are grouped and categorized into the Table 2.

Table 2: Related works

Reference	Year	Methods	Dataset
1) Breast Lesions Detection & Identification			
Kozegar et al. [31]	2013	Adaptive threshold machine learning	INbreast
Agarwal et al. [32]	2019	CNN patch classifier and mass probability	CBIS-DDSM
		map (MPM)	INbreast
Al-masni et al. [49]	2018	YOLO	DDSM
Aly, G. et al. [52]	2020	YOLO	INbreast
Peng et al. [61]	2020	Faster R-CNN	CBIS-DDSM
			INbreast
Al-Antari et al. [62]	2018	YOLO	INbreast
Al-Antari et al. [63]	2020	YOLO	INbreast
Singh et al. [89]	2020	Single Shot Detector (SSD)	INbreast
Ribli et al. [65]	2018	Faster R-CNN + FCN	INbreast
Al-masni et al. [68]	2017	YOLO	DDSM
Akselrod-Ballin et al. [70]	2016	cascaded R-CNN	INbreast
Dhungel et al. [71]	2015	Deep Belief Network (DBN)	DDSM-BCRP
			INbreast
Akselrod-Ballin et al. [72]	2017	three-stage cascade of Faster-RCNN models	INbreast
Li et al. [73]	2020	Siamese-Faster-RCNN	INbreast
Hamed et al. [74]	2021	YOLOV4	INbreast
Watanabe et al. [76]	2019	cmAssist – Custom deep learning networks	Private
Timp et al. [77]	2007	temporal change features + SVM	Private
Timp et al. [78]	2006	regional registration technique	Private
Loizidou et al. [79]	2019	Temporal subtraction	Custom

Loizidou et al. [80]	2020	Image registration	Custom
Zheng et al [81]	2018	Detection: 3 cascading detectors (Haar, LBP, and HOG) Classification: VGG-19	UCHCDM
Dhungel et al. [167]	2017	Cascade Deep Learning and Random Forest	INbreast
Dhungel et al. [83]	2015	Deep structured output learning + refinement	INbreast
Dhungel et al. [135]	2017	conditional GAN (cGAN)	INbreast
Isfahani et al. [186]	2021	Growth regional method	MIAS
Silalahi et al. [187]	2021	CNN (VGG + ResNet)	MIAS
Yassir et al. [188]	2022	YOLOv5 + Mask RCNN	INbreast
2) Biomedical Lesions Segmentation			
Al-Antari et al. [62]	2018	AUNet	INbreast
Hai et al. [87]	2019	Fully connected VGG-16	Private
Zhu et al. [88]	2018	Full resolution convolutional network (Front)	INbreast
Singh et al. [89]	2020	CNN + CRF	INbreast
Ronneberger et al. [90]	2015	UNet	Biomedical Images
Sathyan et al. [91]	2020	UNet	CBIS-DDSM
Soulami et al. [92]	2021	UNet	INbreast DDSM
Abdelhafiz et al. [94]	2019	UNet-128	INbreast
Ibtehaz et al. [95]	2020	MultiResUNet	Biomedical Images
Li et al. [96]	2018	Conditional Residual Unet	INbreast
Oktay et al. [98]	2019	Attention dense Unet	DDSM
Sun et al. [101]	2020	ASPP-FC-DenseNet	private
		R-Unet	INbreast
Ravitha Rajalakshmi et al. [101]	2020	ResNet34 + ASPP	INbreast
		CRUNet	INbreast
Wang et al. [103]	2019	Multi-scale FCN-CRF	INbreast
Zhou et al. [104]	2019	Unet++	Biomedical Images
Tsochatzidis et al. [105]	2021	Unet+	CBIS-DDSM
Jha et al. [106]	2020	DoubleUNets	Biomedical Images
Das et al [107]	2019	Contour-Aware Residual W-Net (WRC-Net)	Biomedical Images
Tran et al. [108]	2021	TMD-Unet	Biomedical Images
Irfan et al. [182]	2021	Di-CNN + morphological erosion operation.	Ultrasonic
Li et al. [183]	2022	Class activation mapping and deep level set (CAM-DLS)	Breast Ultrasound
Sharma et al. [184]	2022	Xception model + SVM classifier	Histology Images

3) Breast Cancer Diagnosis & Classification			
Shen et al. [38]	2017	CNN + VGG	CBIS-DDSM INbreast
Singh et al. [89]	2020	CNN	DDSM
Al-Antari et al. [62]	2018	Ensemble of AlexNet-based CNN	INbreast
Shams et al. [120]	2018	CNN + GAN	DDSM INbreast
Li et al. [121]	2019	DenseNet-II	Private
Zhang et al. [122]	2020	Local binary patter and Deep CNN features	CBIS-DDSM
Chakravarthy et al. [124]	2020	Improved Crow-Search Optimized Extreme Learning Machine (ICS-ELM) algorithm	INbreast
Khan et al. [125]	2019	multi-view feature fusion (MVFF)	CBIS-DDSM
Alkhaleefah et al. [128]	2020	AlexNet, VGG, GoogLeNet, ResNet + fine tuning	CBIS-DDSM
Falconí et al. [127]	2019	MobileNet and NasNet + fine tuning	CBIS-DDSM
Medeiro et al [130]	2020	DenseNet201 + MLP	CBIS-DDSM
Li et al. [133]	2020	DualCoreNet: Texture and shape features fusion	INbreast
Tsochatzidis et al. [105]	2021	CNN	CBIS-DDSM
Dhungel et al. [135]	2017	multi-view deep residual neural network (mResNet)	INbreast
Falconí et al. [129]	2020	NasNet + fine tuning on VGG16 and VGG19	INbreast
Falconí et al. [178]	2020	VGG16 + fine tuning	CBIS-DDSM
Irfan et al. [182]	2021	Fusion of CNN-activated feature vectors and DenseNet201-activated feature vectors combined with SVM	Ultrasonic
Ayana et al. [184]	2022	esNet50-Adagrad-based multistage transfer learning (MTL)	Ultrasonic
Zhou et al. [189]	2022	Resolution adaptive network (RANet)	Histopathological

CHAPTER III PROPOSED METHODOLOGY

A. Introduction

This chapter presents technical details about the architecture models and approaches that were adapted to develop a complete CAD system for breast cancer diagnosis. The models are based on deep learning technology and motivated by the latest cutting-edge computer vision techniques. Each step of the integrated CAD system is solved using a novel architecture model that was trained on mammograms to extract deep features for the detection of breast abnormalities and localize them in the entire mammogram; the segmentation of the breast tumors and masking their background; and the classification of the tumors for a final diagnostic decision. The proposed detection and identification step was conducted with a fusion-models approach that is based on the YOLO architecture model. The same idea was tested on an external dataset that acquires current mammograms and their corresponding prior screening views. The model was inferred on synthetic prior images that were generated using an image-to-image translation technique Pix2Pix and retrospective analysis was used to evaluate the early detection and classification of abnormalities in prior mammograms. Next, the segmentation step was solved with a novel segmentation architecture model, called Connected-UNets. The performance was enhanced using additional synthetic images using image-to-image translation methods. Finally, the classification step was developed using a stacked ensemble of Residual Neural Networks (ResNet) models.

The following sections detail each methodology and approach that was proposed in this work.

B. YOLO-based fusion model for Abnormalities Detection and Identification

1. YOLO-Based Model

Object detection refers to a regression problem that maps the right coordinates of image pixels to a bounding box that surrounds a specific object. After being fed into a CNN that builds a convolutional feature

map, popular regional-based neural networks models predict multiple bounding boxes and use regions to locate objects within images. This approach applies a selective search that extracts most adequate regions from images and then predicts the offset values for the final bounding boxes. Typically, this technique is experimentally slow and memory consuming, therefore YOLO (You-Only-Look-Once) deep learning network was proposed where a single CNN predicts at the same time bounding boxes allocation and their class label probabilities from entire images. The fact that YOLO does not involve extracting features from sliding windows leads to its low computing cost. In fact, it only uses features from the entire image to directly detect each bounding box and predict its class label probability.

YOLO architecture, as explained in Figure 11, is simply based on the fully convolutional neural network (FCNN) design. Particularly, it splits each entire image into $m \times m$ grids and for each grid, B bounding boxes are returned with a confidence score and C class probabilities.

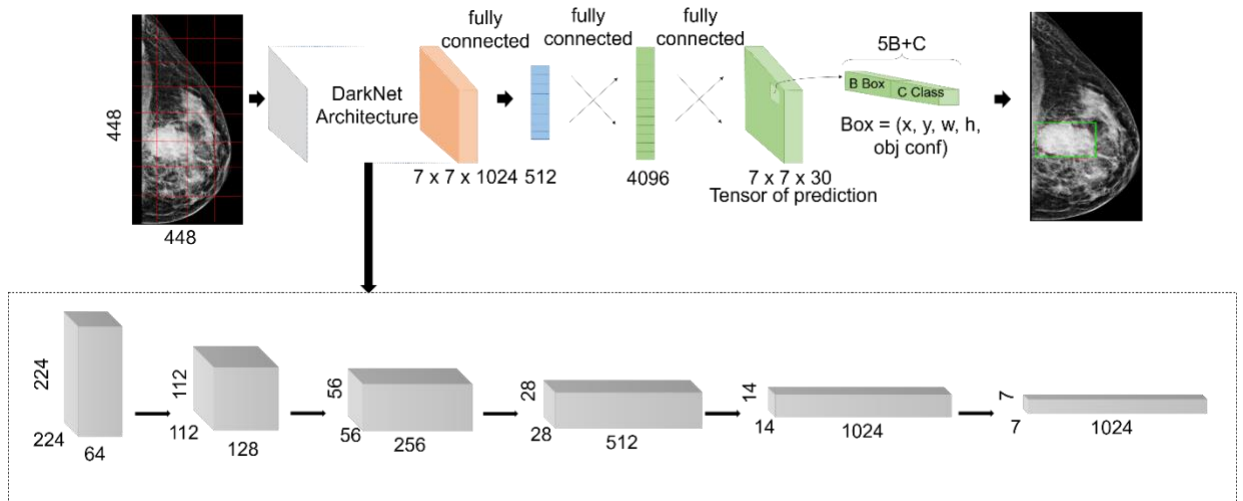


FIGURE 11: Proposed YOLO-based architecture

Confidence score is computed by multiplying the probability of an existing class object with the intersection over the union (IoU) score as detailed in Eq. (1).

$$\text{Confidence} = \text{Prob}(\text{object}) \times \text{IoU score} \quad (1)$$

In addition, the detected object is classified as mass or calcification according to its class probability and its confidence score for that specific class label as explained below in Eq. (2).

$$\text{C class probability} = \text{Prob}(\text{Class object}) \times \text{IoU score} \quad (2)$$

In this dissertation, at the time our recent work was published, we adopted YOLO-V3, which is the third improved version of YOLO networks, to detect more different scaled objects, and it uses multi-scale features extraction and detection. As shown in Figure 11, the architecture first employs an extraction step that is based on the DarkNet backbone framework [138]. It was inspired by the ResNet architecture and VGG-16, and it presents a new design of 53 layers, as illustrated in the lowest block in Figure 11. The model uses skip connections to prevent gradients from diminishing and vanishing while propagating through deep layers. After that, the extracted features at different scales are fed into the detection part that presents three fully connected layers. After that, it applies the concept of anchor boxes that is borrowed from Faster-RCNNs model. In fact, the prior boxes were pre-determined by training a K-means algorithm on the entire images. After that, the output matrixes of multi-scale features were defined as grid cells with anchor boxes. This helps to determine the IoU percentage between the defined ground-truth and anchor boxes. It also ensures selecting the boxes with the best scores compared to a certain threshold. At the end, four offset values of bounding boxes against each anchor box were predicted with a confidence score and a class label probability. Hence, the detection considered correct bounding boxes that had both scores exceeding a certain threshold [139].

2. Fusion Models Approach

According to the generalized YOLO-based model we presented earlier in Figure 11, the bounding boxes that surround suspicious breast lesions are detected with certain confidence scores as explained in the previous subsection. This score is affected by model settings, input data provided to the model, and YOLO's internal classification process to calculate the class label probability score (i.e., Mass or Calcification). Based on this hypothesis, evaluation of such a model can be expanded to improve the final predictions results.

We proposed in this dissertation that the best predicted bounding boxes be chosen first among all augmented images (i.e., rotated, morphed, translated, etc.) based on their IoU score. The idea helped to determine the best representative mammograms to correctly localize and classify breast lesions. Second, we suggested joining different predictions of the model's implementation to lower the error rate and combine performance of differently configured models. These models were trained and set differently, resulting in a fusion-based model devoted to achieving the best results.

In fact, we note that Model1, referred to as M1, is trained and configured differently for one class targeting either Mass or Calcification. Therefore, the two developed models from M1 are now referenced as

M1(Mass) for Mass class and M1(Calcification) for Calcification class. Model2, referred to as M2, is configured for multi-class training and identification and used for fusion to improve the performance of single-class models. The model M2 will now be identified as M2(Mass and Calcification) since it targets multiple classes.

After developing and testing each model M_i , our proposed fusion approach is to create a fusion model for Mass class using M1(Mass) and for Calcification class using M1(Calcification), while benefiting from the M2(Mass and Calcification) to improve the performance of the M1 models.

We first report the Mass predictions1 using M1(Mass) that have IoU score more than threshold1. Next, we select only images with Mass lesions and report their predictions using M2(Mass and Calcification) and another threshold2. After that, we filter out predicted images that are not within the Mass predictions1 and save them as Mass predictions 2. We finally combine the two predictions into final Mass predictions as shown in Figure 12. We repeat the same logic for Calcification predictions according to the flow in Figure 12. In all our fusion models, we used a threshold1 to be 0.5 and threshold2 to be 0.35 that yielded satisfying results.

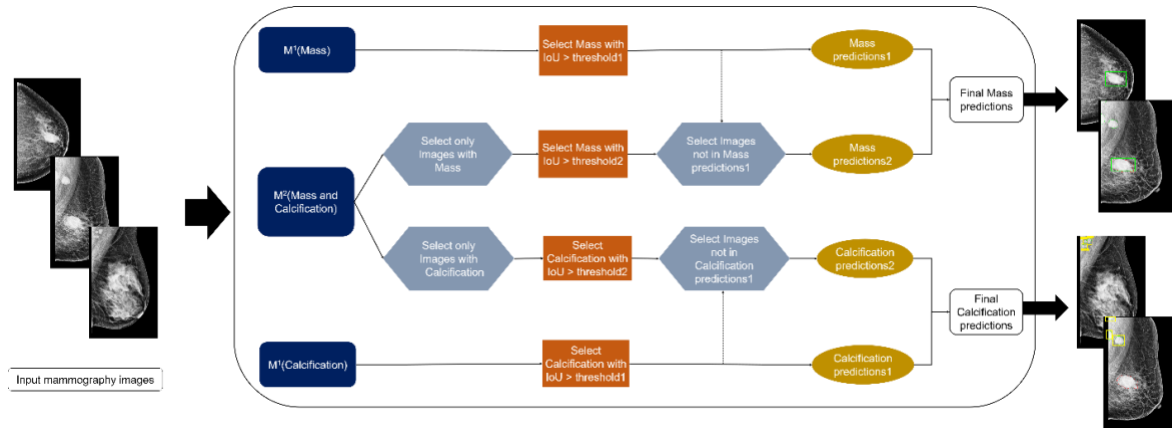


FIGURE 12: Flow chart of the fusion models' approach for the final prediction (examples of Input mammography images that includes single lesions and different lesions cases from the CBIS-DDSM dataset)

C. Early Detection and Classification of Abnormalities in Prior Mammograms

As the previous work detailed, we employed a YOLO-based model in a different evaluation fashion. The basic model was initially trained using different configurations (i.e., target class labels). Then, each experiment was evaluated by selecting the best predicted bounding boxes within all augmented images (i.e., original, and rotated images) having the highest confidence score. The technique proved an effective way to

determine the best representative images to precisely detect and classify breast lesions in each mammogram. After that, as shown in Figure 13, the idea of YOLO-based fusion models was implemented to improve the final prediction results. Different predictions were joined to lower the final error rate and to combine models that were differently configured. We used the same notation by referring Model1 to the YOLO-based model that was trained and configured for one class either Mass, Calcification, or Architectural Distortion. Therefore, Model2 refers to the YOLO-based model that was configured for multiple classes training (i.e., all three classes together). Finally, the Fusion Model refers to the combined evaluation of Model1 and Model2 that was used to improve the overall detection performance. The final model should select predictions that were not within the single class predictions according to a threshold of 0.5, which showed satisfying results. All models were developed and tested on the Current mammograms from the most recent screening, with either Mass, Calcification or Architectural Distortion lesions. Different from our previous work, we added a class label, ‘Normal’ for the current mammograms that were not diagnosed with abnormal findings during the follow-up screening. Our trained YOLO-based model on abnormal mammograms was applied on Normal mammograms to ensure that no bounding boxes were predicted, and consequently, classify the mammograms as Normal.

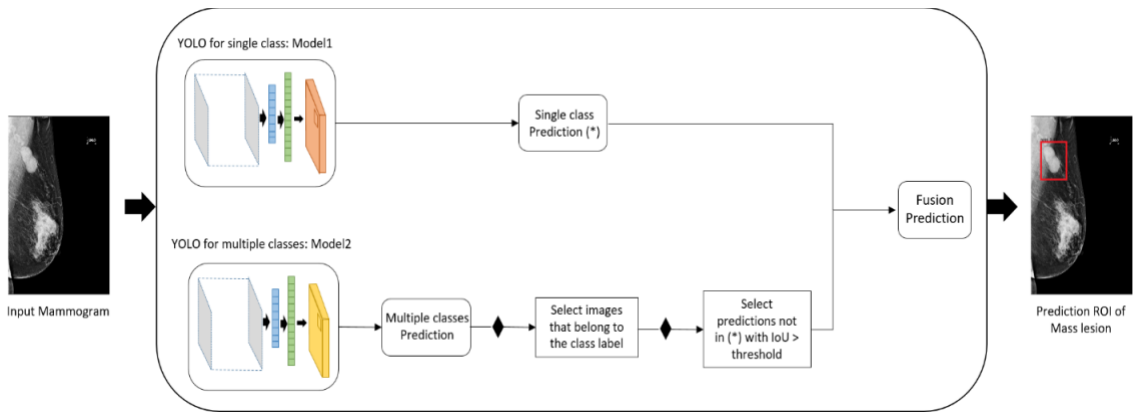


FIGURE 13: YOLO-based Fusion model - Example of mammogram with a Mass lesion

In this task, we first apply and evaluate the YOLO technique on the Current mammograms to detect different breast lesions and classify them into Mass, Calcification, or Architectural Distortion, and the rest to Normal. Second, we consider two image-to-image techniques, Pix2Pix and CycleGAN, to learn mapping between Current mammograms and their corresponding Prior mammograms. As shown in Figure 14, new synthetic Prior mammograms are generated to overcome the misalignment between the screenings due to

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

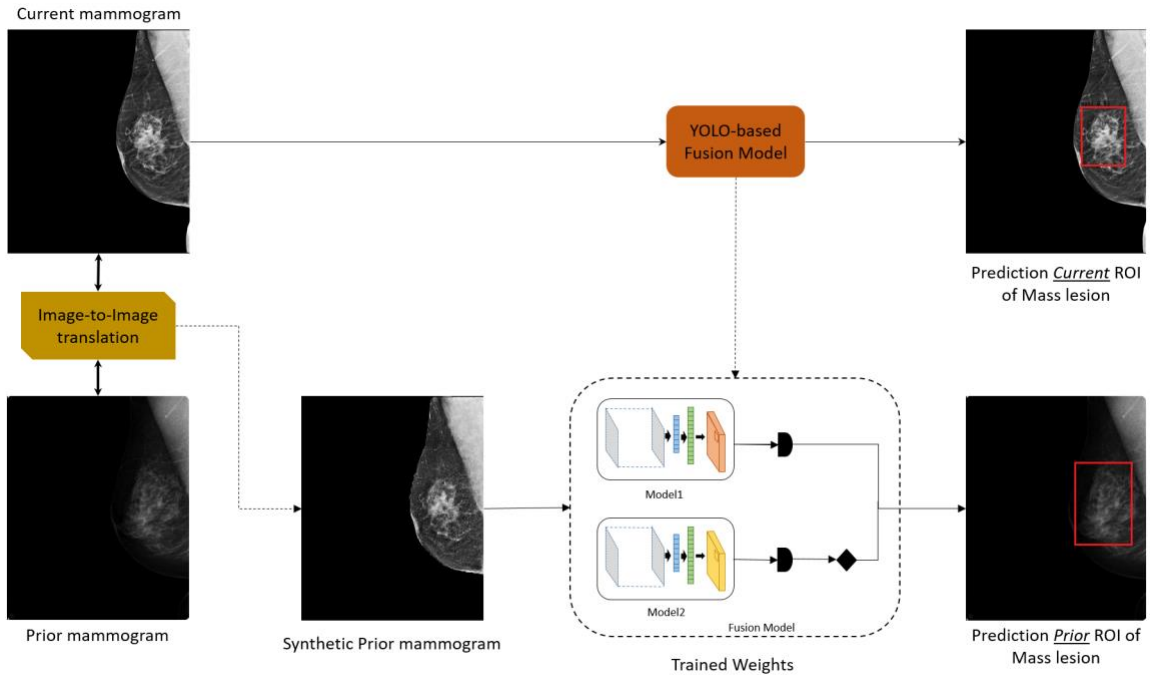


FIGURE 14: Framework for early detection and classification on Prior mammograms (Example of Prior mammogram with normal diagnosis and Current mammogram with Mass lesions (red bounding boxes))

D. Novel Connected-UNets architecture model for Breast Mass Segmentation

UNet is considered as a state-of-the-art model for medical image segmentation. Inspired by the Fully Convolutional Network (FCN), it is based on the encoder-decoder structure by removing the fully connected layers. The network has a symmetric architecture with a u-shape, as the name suggests. It has two paths: one for down-sampling and one for up-sampling. The remarkable contribution of UNet architecture was the introduction of the skip connections path that added an advantage to the standard encoder-decoder architecture. This helps recover the spatial information that gets lost during the down-sampling path due to the pooling operations.

Inspired by the efficiency of the skip connections, we propose a new architecture, called Connected-UNets, which connects two UNets using additional skip connections. Figure 15 shows an overview of the proposed architecture, where it consists of two standard encoder and decoder blocks and two Atrous Spatial Pyramid Pooling (ASPP) blocks for the transition between the two pathways. We suggest connecting the first decoder and the second encoder blocks with new modified skip connections to reconstruct the decoded information in the first UNet before being encoded again in the second UNet. Each encoder block includes two convolution units, which consist of 3 x 3 convolutions followed by an activation ReLU (Rectified Linear Unit) and a batch normalization (BN) layer. Before sending the information to the next encoder, each encoder block's output is subjected to a maximum pooling function. Each decoder block consists of a 2 x 2 transposed convolution unit (i.e., deconvolution layer) that is concatenated with the previous encoder output, and then the result is fed into two convolution blocks, which consist of 3 x 3 convolutions followed by an activation ReLU and a BN layer.

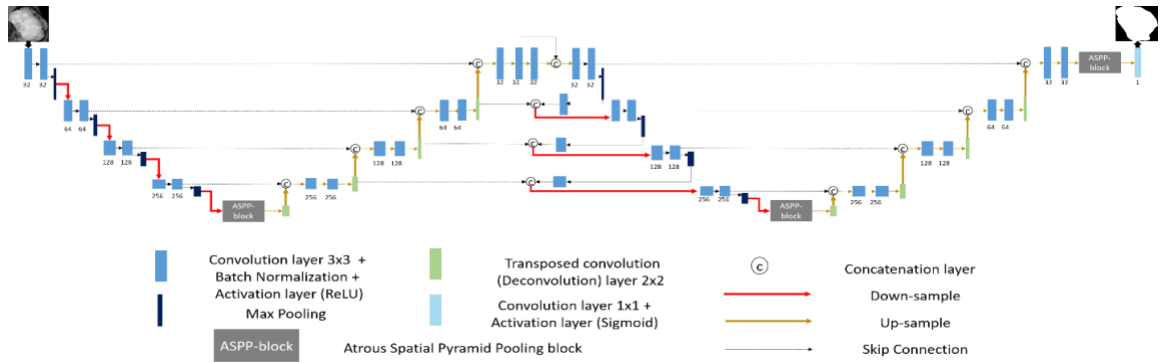


FIGURE 15: The proposed Connect-UNets architecture

An ASPP block is used to transition between the down-sample and up-sample routes. As the name indicates, this technique uses “Atrous” (which means “holes” in French) convolution to allow having a larger receptive field in the transition path without losing resolution. As shown in Figure 16, it is composed of four dilated 3 x 3 convolutions layers with dilation rate $r = \{0, 6, 12, 18\}$ and followed by BN layers. The four sub-blocks are next added to create a multi-scaled features block that is finally fed into a 1 x 1 convolutions layer.

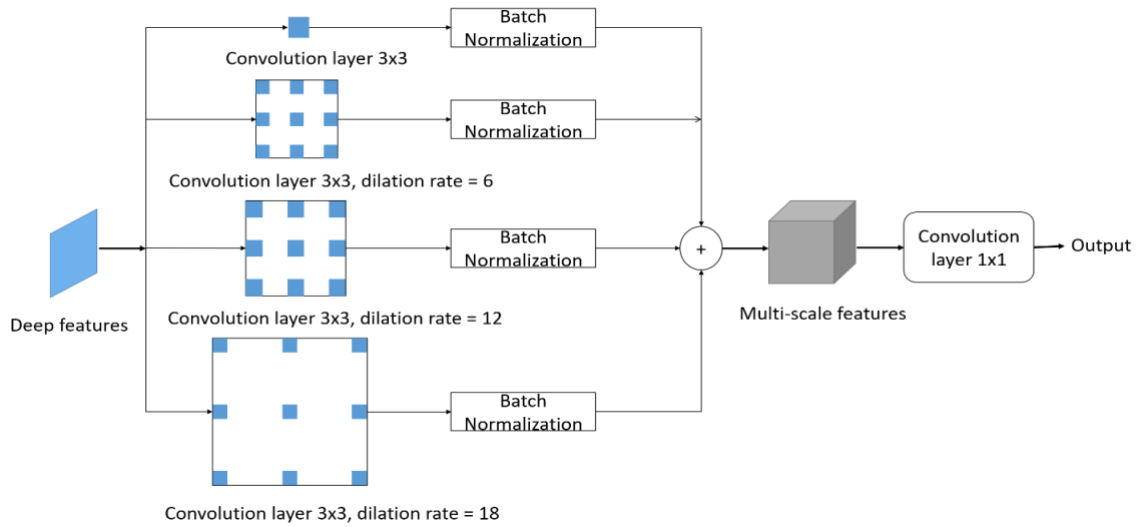


FIGURE 16: The Atrous Spatial Pyramid Pooling (ASPP) block

After going through the first UNet, a second UNet is attached through new skip connections that use information from the first up-sampling pathway. First, the result of the last decoder block is concatenated with the same result after being fed into a 3 x 3 convolutions layer followed by an activation ReLU and a BN layer. This is used as the input for the second UNet's initial encoder block. After that, the output of the maximum pooling operations of each three encoder blocks are fed into 3 x 3 convolutions layers and then concatenated with the output of the last previous decoder block. The result is next down sampled to the next encoder block. The second UNet's last encoder block is transferred to the ASPP block, and the rest is the same as the first UNet. Finally, the last output is given to a 1 x 1 convolutions layer that is followed by a sigmoid activation layer to generate the predicted mask.

In addition to the proposed design, which is based on the standard UNet, we present the AUNet model, which includes an attention block throughout the up-sampling process. This combines the attention technique with the encoder and decoder blocks' skip connections. Indeed, the additional attention block should allow the network to weight the low-level features (i.e., information that has been down-sampled) before being concatenated with the high-level features (i.e., up-sampled information) during the skip connections. Thus, a new Connected-AUNets architecture is introduced as illustrated in Figure 17.

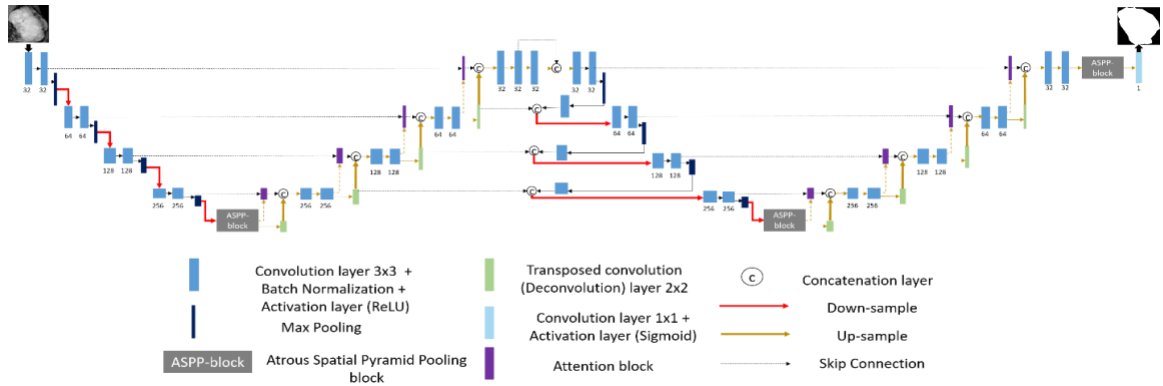


FIGURE 17: The proposed Connect-AUNets architecture

In fact, the attention block, as shown in Figure 18, consists of a 2×2 transposed convolutions layer with strides equal to (2,2) and takes low-level features as input. Next, the output is concatenated with the high-level features and the result is fed into a ReLU activation layer followed by a 2×2 transposed convolutions layer with strides equal to (1,1) and a sigmoid activation layer. This generates the attention map that is next multiplied by the skip connect input to produce the final output of the attention block, which serves as a new input of the decoder block.

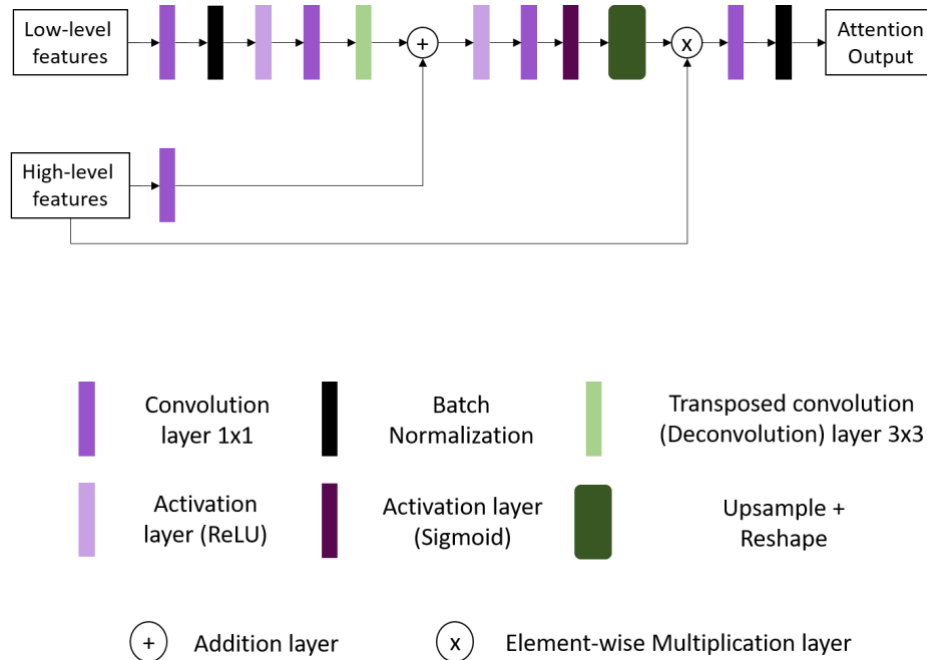


FIGURE 18: The attention block

We replace the usual convolution blocks with residual convolution blocks, motivated by the improvement made to the UNet architecture to make it strong enough for segmenting medical images with

diverse sizes. Therefore, the proposed architecture has become the Residual UNet (ResUNet), and consequently we proposed a new Connected-ResUNets architecture as detailed in Figure 19.

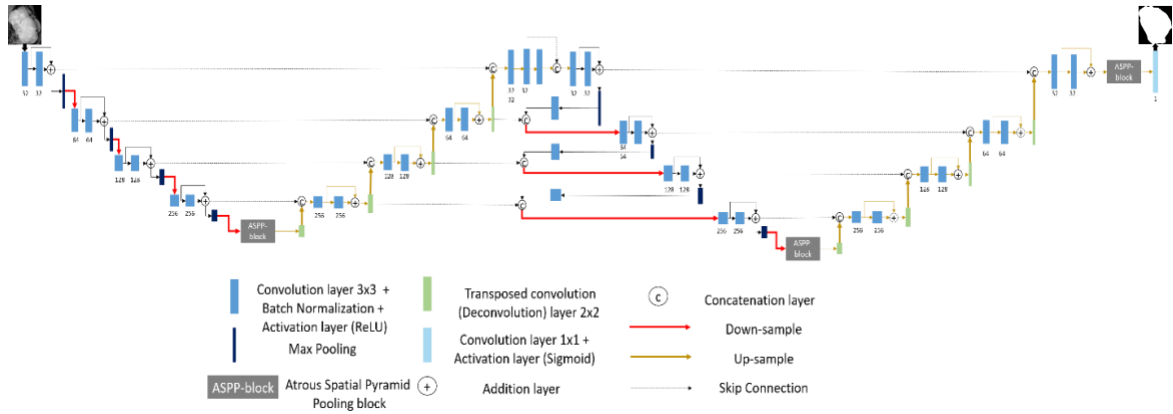


FIGURE 19: The proposed Connect-ResUNets architecture

Consequently, adding the residual convolution blocks should enhance the UNet architecture to reconcile the features learnt at each scale of the down-sampling pathway and take full advantage of the information propagated, which may result in the deep network's degeneration. As shown in Figure 20, a residual convolution block consists of two 3 x 3 convolution layers, followed by BN and ReLU activation layers. The output of the residual block is then yielded by adding the input features to the outputs of the stacked layers.

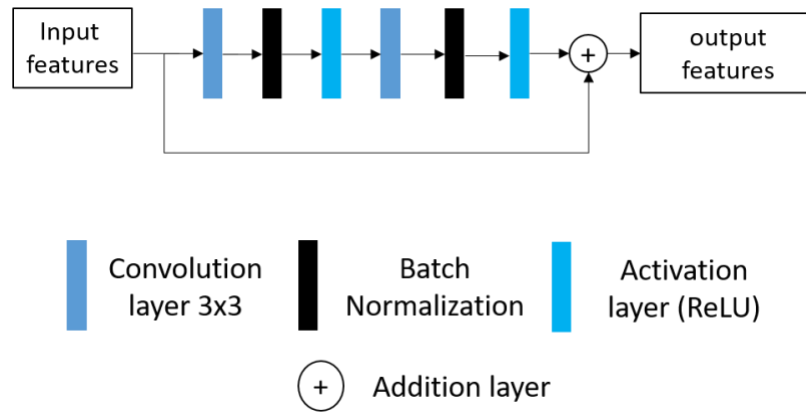


FIGURE 20: The residual block

E. Image Synthesis for Data Quality Enhancement and Augmentation

Deep convolutional networks have been enormously improved to provide cutting-edge solutions to computer vision and they have given the ability to manipulate images for complex image-related tasks such

as image synthesis, image reconstruction, image translation, etc. Recently, these tasks were significantly treated thanks to the discovery of Generative Adversarial Networks (GANs) architecture. A standard GAN comprises two models, a generator, and a discriminator. These models compete against each other to produce fake data that is realistic enough to fool the discriminator. The architecture has known success in medical imaging applications [140] and many variants were introduced such as conditional GAN (cGAN), Wasserstein Generative Adversarial Network (WGAN), etc. Further work extended the idea to create multiple GANs that can serve for synthetic data augmentation, domain adaptation, and style transfer. This allowed using a pair of generators to learn mappings of images and a pair of discriminators to distinguish two different types of images. The idea emphasized the image-to-image translation task that leverages external labeled dataset to effectively reconstruct the source domain images with additional characteristics of a target domain such as pixels, color distribution, shape, and texture. In this context, Pix2Pix and CycleGAN are two common models that were developed to apply image-to-image translation techniques. As shown in Figure 21, like the standard GAN, the two models have the target of translating images between two domains. However, the difference is that the Pix2Pix model works with paired data sets but only accepts one image from source domain (A) and indeed it corrects and updates the training using its corresponding image from a target domain (B). Differently, CycleGAN model works with unpaired datasets, accepts two images, and performs a cyclic translation across domains to return new synthetic images.

In fact, Pix2Pix [141] is based on conditional GAN (cGAN) architecture to learn a mapping between images where the network is composed from a generator $G_{A \rightarrow B}$ and a discriminator D . The generator has an encoder-decoder structure, and it tries to transfer special characteristics of an input image x_r^A to get an output image x_g^B . The discriminator uses PatchGAN architecture, and it compares the input image to the generated image on one time and the input image to the corresponding image from the external dataset x_r^B another time to update the generator learning.

Moreover, the Cycle Generative Adversarial Network, called CycleGAN [142], was designed to learn mapping between images without the need to have correlations and one-to-one matches. The idea was built on the top of Pix2Pix architecture but with the use of two generators $G_{A \rightarrow B}$ and $G_{B \rightarrow A}$ for cycled images mapping and two discriminators D_A and D_B to distinguish between real and synthetic images. Additionally, the CycleGAN technique employs a cycle consistency for the generators to ensure a good reconstruction of

the new image back to their original look. Consequently, the technique helps to capture both domains' features and style without mismatch.

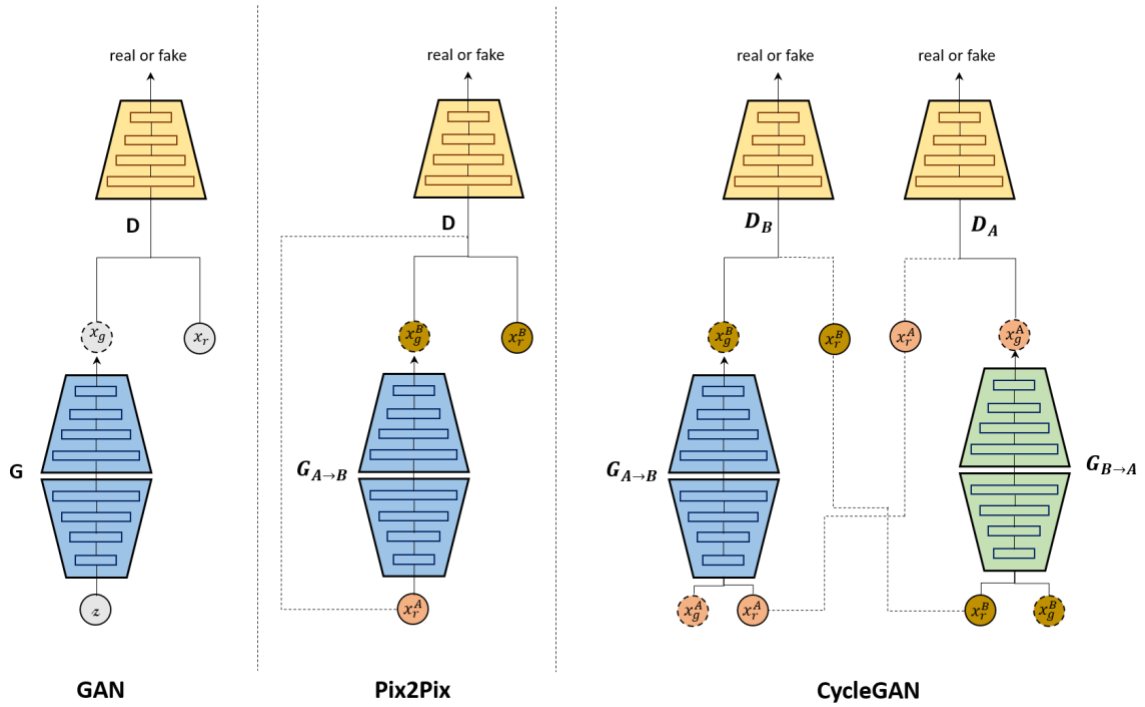


FIGURE 21: Comparative scheme of standard GAN vs two variants for image-to-image translation: Pix2Pix and CycleGAN

Given the limited size of annotated datasets and the difference of their resolutions, we propose to apply image synthesis on our mammography datasets to improve the results of the segmentation. In a CycleGAN architecture, a deep learning model learns the mapping pixel, color distribution, shape, and texture between two datasets [143]. It has been considered as an effective deep learning technique for style transfer, domain adaptation and data synthesis [144, 145, 146, 147]. Figure 22 details the CycleGAN architecture that additionally evokes a cycle consistency for the generators G and F to ensure the reconstruction of the images back to their original domains, where $F(G(X)) = X$ and $G(F(Y)) = Y$. This helps the final model to capture the characteristic features of the two domains and transfer the style without requiring any paired dataset. Thus, the network uses the standard adversarial losses and a cycle consistency loss, defined as $\|F(G(X)) - X\|_1 + \|G(F(Y)) - Y\|_1$.

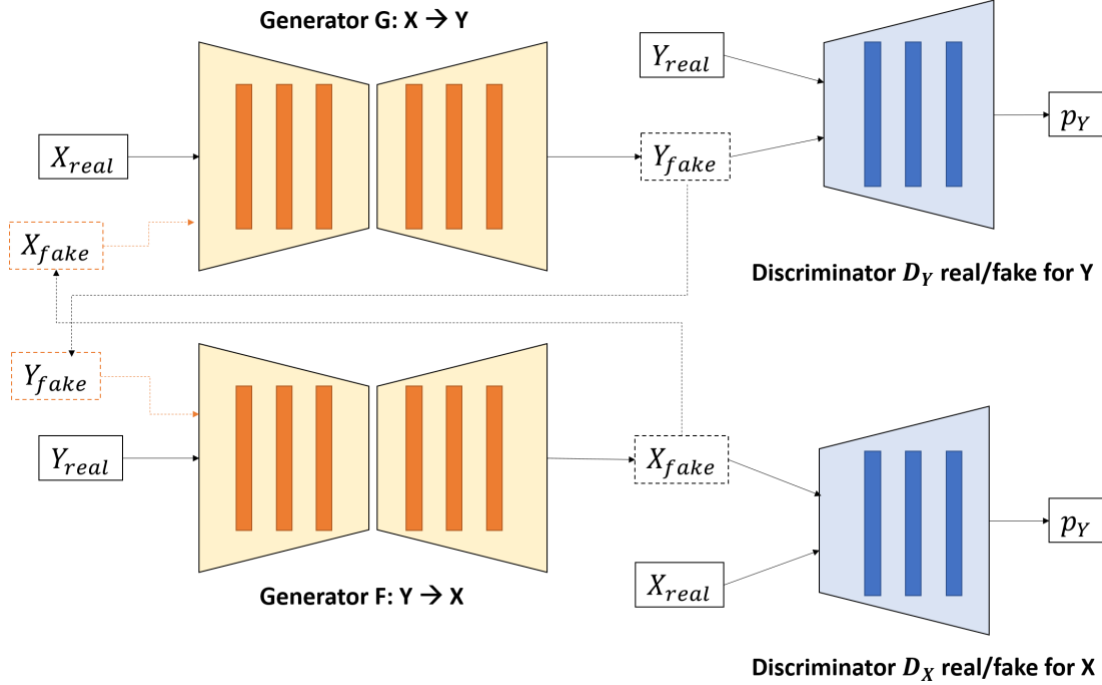


FIGURE 22: CycleGAN architecture for data synthesis

In this work, we developed the CycleGAN model using the tutorial available in Keras webpage (<https://keras.io/examples/generative/cyclegan>). The architecture model has two generators and two discriminators' networks. The generator network consists of two down-sampling blocks with filter sizes {128, 256}, nine residual blocks with filter size 256, and two up-sampling blocks with filter sizes {128, 64}. The discriminator network is based on four down-sampling blocks with filter sizes {64, 128, 256, 512}. For the Pix2Pix model, we similarly used two generators and two discriminators' networks. The generator network contains seven down-sampling and up-sampling blocks with filter sizes {64, 128, 256, 512, 512, 512, 512} and we used the same discriminator network from the CycleGAN architecture model.

F. Stacked Ensemble of Neural Networks for Tumors Classification and Diagnosis

In this step, we propose a stacked ensemble of models to classify and diagnose detected and segmented breast masses. The base model comes from the ResNet architecture and its variations. Our methodology employs different strategies: transfer learning, stacked ensemble learning, and image data augmentation.

1. ResNet base model: transfer learning and fine-tuning

Since its introduction, ResNet has been one of the recent architectures that has known common success in medical imaging applications [148, 149]. ResNet is a deep CNN architecture suggested by He et al. [150]

that uses residual blocks with skip connection between layers to bypass a few convolution layers at a time. This new architecture accelerated the convergence of a larger number of deep layers, and consequently it has been found efficient to provide a compact representation of input images and improve the classification task performance [151]. The ResNet has some common architectures such as ResNet-50, 101, and 152, which indicate the number of deep layers, but all use the same residual blocks containing three convolutional layers with stride 1 and 3 followed by a Batch Normalization (BN) layer and a non-linearity layer Rectified Linear Unit (ReLU) on its shortcut path [152]. Alternatively, ResNet architecture presented an improved version ResNetV2 by He et al. [153], where the last ReLU was removed to clear the shortcut path using a simple identity connection, as shown in Figure 23.

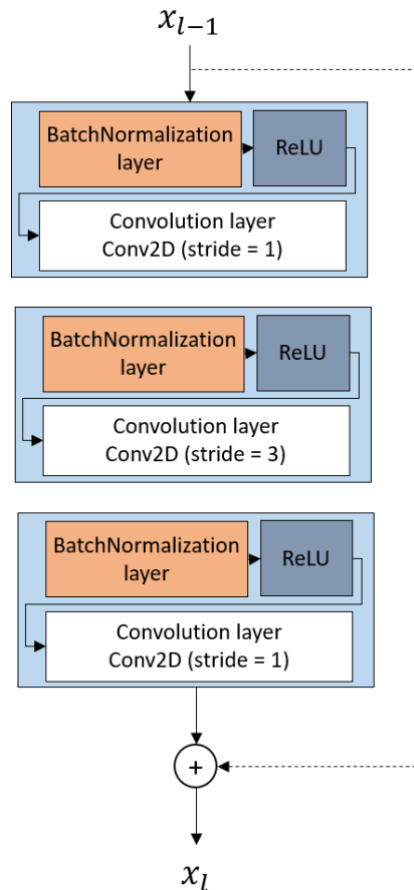


FIGURE 23: Residual block of ResNetV2

Our methodology employs three pre-trained ResNetV2 architectures, detailed below in Table 3. The architectures accept RGB input images of size 224 x 224, and they are composed of an initial convolutional layer with a filter size 7 x 7 and stride 2, followed by a pooling layer of size 3 x 3 and stride 2. The base

architecture should then use four residual blocks with different kernel sizes of 64, 128, 256, and 512. Each architecture has a different number of residual blocks that define the depth of layers. Finally, a global average pooling layer with size 7 x 7 is used with a final fully connected layer (FC).

Table 3: ResNetV2 architecture layers

Layer Name	Output Size	ResNet50V2	ResNet101V2	ResNet152V2
Input Layer		224 x 224 x 3		
		Number of filters = 64		
		Filter size = 7 x 7		
		Strides = 2		
		Padding = 3		
conv1	112 x 112 x 64			
		Strides = 2		
		Pooling size = 3 x 3		
pool1	56 x 56 x 64			
conv2_x	56 x 56 x 64	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$
conv3_x	28 x 28 x 128	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 8$
conv4_x	14 x 14 x 256	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 36$
conv5_x	7 x 7 x 512	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$
Global average pooling	2048	Pooling size = 7 x 7		
		Strides = 7		
Fully connected	1000	1 x 1 x 2048		

Training a deep learning model often requires a large amount of annotated data that helps optimize the high number of parameters and computations needed in the architecture. However, a limited size of medical imaging datasets are usually available that suffer from either missing labels or imbalanced data distribution. To overcome these challenges, transfer learning has been a common solution used in many recent medical image application [154, 155] by training a model on a large and diverse dataset (i.e., ImageNet, MSCOCO, etc.) to capture universal features like curves, edges, and boundaries in its early layers that are relevant for image classification. After that, the pre-trained model should be alerted and fine-tuned on a custom and specific dataset to reflect the final classification. This procedure provides a fast and generalizable training of small datasets and avoids the overfitting problem that deep learning commonly suffers from.

As Figure 24 indicates, we apply transfer learning to the base architecture ResNetV2 for our proposed methodology to become a TF-ResNetV2. In fact, the model was initially pre-trained on ImageNet, and then the first four residual blocks of layers were frozen with exception to the BN layers that needed to be retrained to improve the training convergence. After that, the entire architecture was modified by adding another FC layer with size of 1024, followed by a dropout regularization layer to maintain a generalization aspect for the training. A new final FC layer was placed according to the number of classes for each classification task and the entire TF-ResNetV2 is re-trained.

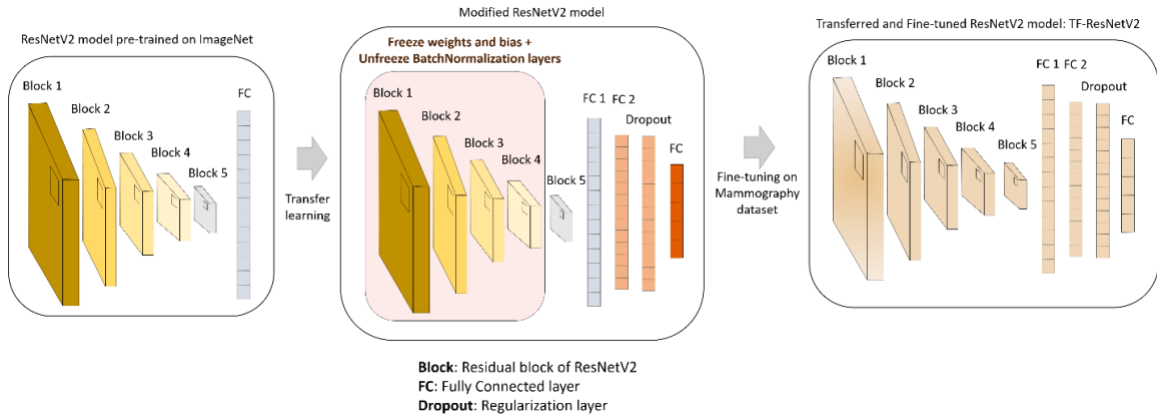


FIGURE 24: Framework of the classification base model: a TF-ResNetV2 model which is a ResNetV2 model pre-trained on ImageNet data and modified and fine-tuned on mammography dataset

2. Stacked Ensemble of ResNet models for Breast tumors classification

Ensemble learning has been considered efficient to improve the classification task results. Combining weaker classifiers to create a better final classification prediction has been adopted by either bagging, boosting, or stacking models [156]. While bagging is achieved by learning independently from different models and then averaging the predictions, boosting happens by sequentially learning from homogenous learners and iteratively combining them into a final model. On the other hand, stacking has been considered a way to learn different weak learners parallelly and combine them into a meta-model that is later trained to achieve the classification prediction [157].

We propose a stacked ensemble of three different ResNet models to conduct our classification tasks. After removing the last FC layer of each ResNetV2 architecture, a two-layer network is considered as a meta-classifier model that concatenates the three models' layers, and stacks three different FC layers of size 1000, 100 and 10, coupled with activation functions Sigmoid and ReLU. As shown in Figure 25, pre-trained weights

of each model were extracted as images features of size 1024 based on previous layer predictions and considered as new input of the entire stacked ensemble of ResNet models for the final class prediction after training independently ResNet50V2, ResNet101V2, and ResNet152V2.

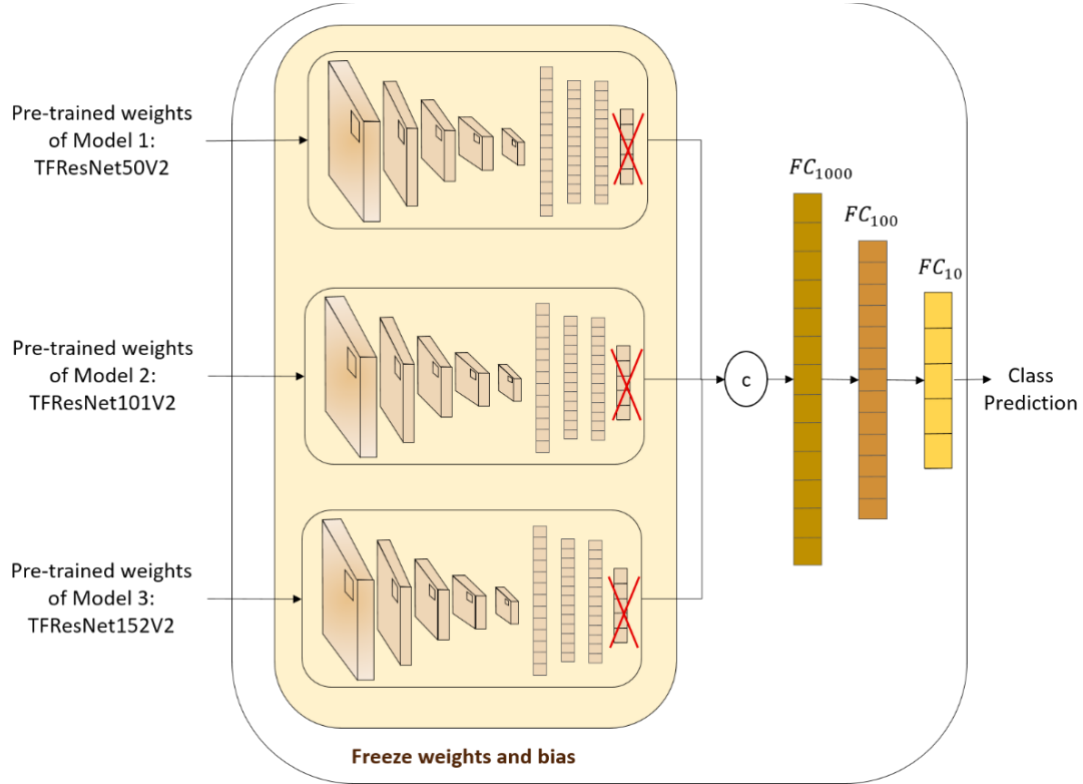


FIGURE 25: Framework of the classification Stacked Ensemble of ResNet models

G. Integrated Framework of CAD system for Breast Cancer

Our final framework detects and localizes breast masses in a first step, and then segments them in a second step. As a preparatory step before implementing the mass segmentation, an advanced data-enhancement method is used. This approach will not only improve the quality of low-resolution mammograms, but it will also increase the size of mammography datasets.

In fact, the new architecture is used to apply to the ROIs of breast masses discovered in the previous stage. Our framework applied the You- Only-Look-Once (YOLO) model in the previous step to locate suspicious breast lesions and distinguish between mass and calcification lesions. As a result, the full mammograms were used to forecast bounding boxes around worrisome items. At this stage, the architecture is applied only on the ROIs of breast masses that were detected from the previous stage. Due to the lack of ground truth annotation of the calcification lesions, only mass lesions segmentation is conducted in this study.

Our framework applied the YOLO-based fusion model from the first step to locate suspicious breast lesions and distinguish between mass and calcification lesions. Hence, the detected ROI masses are fed directly into the second segmentation step using our new proposed architecture as shown in Figure 26. Given the different scale of breast masses, our methodology expands some bounding boxes coordinates by adding extra space around the small tumors. Thus, we obtain the ROI images, and we scale them into 256 x 256 pixels, which is the optimal input size found experimentally for segmentation networks.

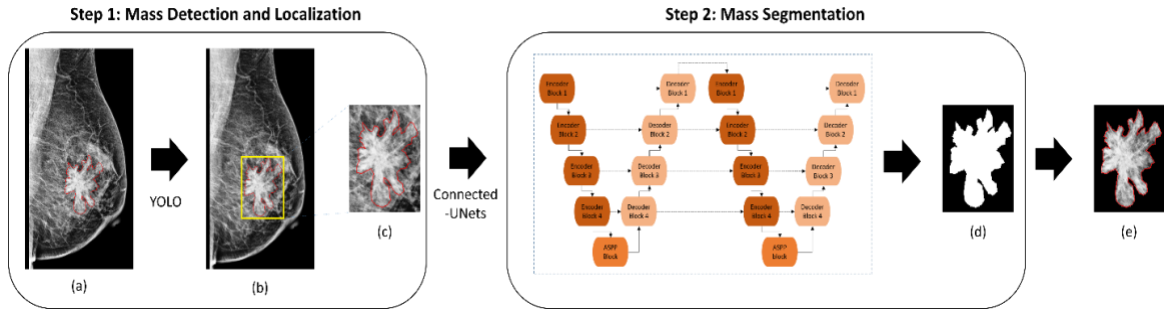


FIGURE 26: The proposed integrated framework

(a) Original mammogram with ground truth of mass (red), (b) Detected ROI of mass (yellow) superimposed on the original mammogram, (c) Detected ROI (i.e., input mass) obtained with ground truth (red), (d) Output segmented binary mask of input mass, and (e) Segmented output mass where tissue is masked

After that, we enhanced the performance by integrating synthetic images as shown in Figure 27. During the segmentation, the original detected ROI masses' images and their generated synthetic images are fed directly into the segmentation stage using our proposed architecture.

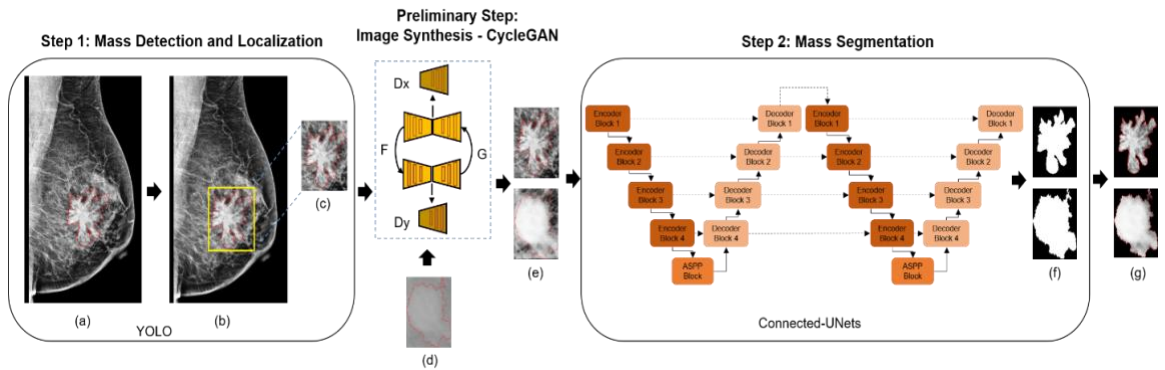


FIGURE 27: The proposed integrated framework

(a) Original mammogram with ground truth of mass (red). (b) Detected ROI of mass (yellow) superimposed on the original mammogram. (c) Detected ROI (i.e., input mass) obtained with ground truth (red) (Domain X). (d) Detected ROI obtained from a different mammography dataset (Domain Y). (e) Original ROI (Domain X) and Synthetic ROI

(transferred from Domain Y to Domain X). (f) Output segmented binary mask of input mass. (g) Segmented output mass where tissue is masked.

Finally, the segmented and detected ROI of breast masses generated with masked tissue are used for the third and final classification step. The stacked ensemble of ResNet models is trained independently on the input ROI masses for each classification task to finally predict the pathology as either malignant or benign, the BI-RADS category with an assessment score between 2 to 6, or/and the shape as either round, oval, lobulated or irregular. Therefore, the final framework is complete with all automated steps for breast cancer analysis and diagnosis as shown in Figure 28.

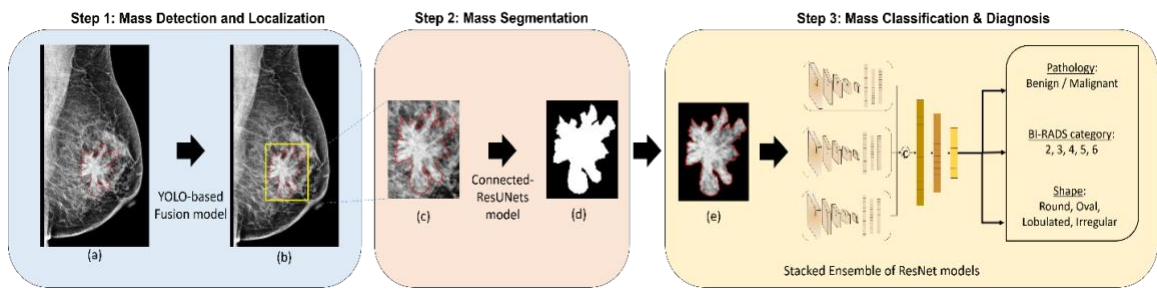


FIGURE 28: The proposed integrated CAD framework

(a) Original mammogram with ground truth of mass (red), (b) Detected ROI of mass (yellow) superimposed on the original mammogram, (c) Detected ROI mass obtained with ground truth (red), (d) Output segmented binary mask of ROI mass, and (e) Segmented ROI mass with marked tissue.

CHAPTER IV EXPERIMENTAL RESULTS & DISCUSSION

A. Introduction

In this chapter, different experiments will be detailed to explain the evaluation of all proposed methods and models. As mentioned before, the integrated framework of the CAD system was designed to present a complete diagnosis of breast cancer as shown in the Figure 29. However, every step forming the CAD system should return an output that was used somehow as an input of the consecutive step. Therefore, all steps should be evaluated individually and together to ensure the final performance. The used mammography datasets are annotated to help validate the learning and evaluation metrics were used accordingly for each specific task. This chapter presents the datasets we used for this dissertation, details the evaluation metrics, and finally reports the results of each proposed methodology.

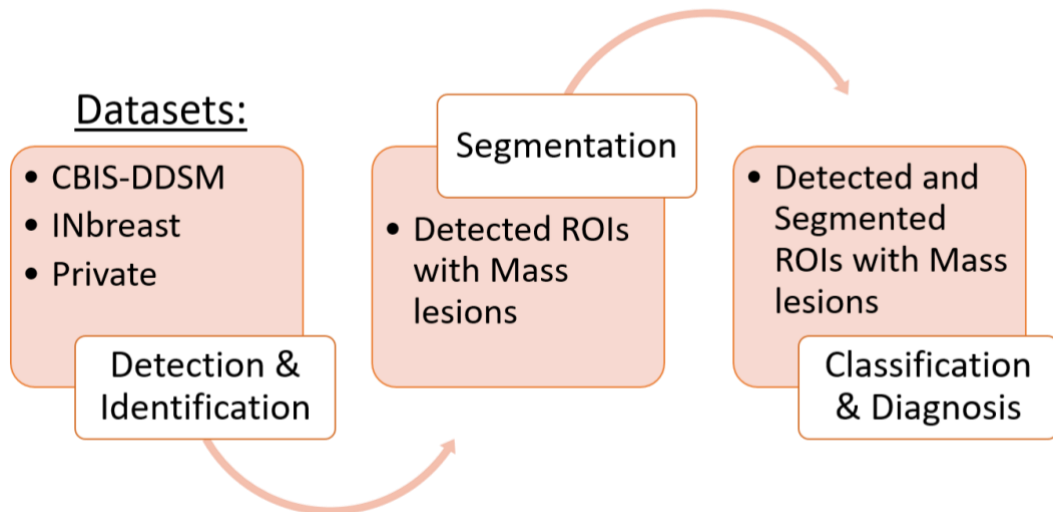


FIGURE 29: The sequence of different experiments using the public and private mammography datasets

B. Mammography Datasets

In this dissertation, we evaluated the proposed architectures models and methods on two public datasets, the CBIS-DDSM and INbreast datasets and two private datasets that will be detailed in the subsections.

1. Public datasets

a. Curated Breast Imaging Subset of DDSM dataset

The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) dataset [158] is an updated and standardized version of the DDSM dataset, in which images were transformed from Lossless Joint Photographic Experts Group (LJPEG) to Digital Imaging and Communications in Medicine (DICOM) format. After removing incorrect samples, it was evaluated by radiologists and validated with histopathological classification. It was reviewed by radiologists after eliminating inaccurate cases and confirmed with the histopathology classification. It contains 2907 mammograms from 1555 patients, and it is organized in two categories of pathology: Mass images (50.5%) and Calcification images (49.5%). Mammograms were collected with two different views for each breast (i.e., MLO and CC) and samples are presented in Figure 30. Original images have average size of 3000×4800 pixels and are associated with their pixel-level ground-truth for suspicious regions location and type, and class labels annotation (i.e., Pathology, BI-RADS category and Shape).

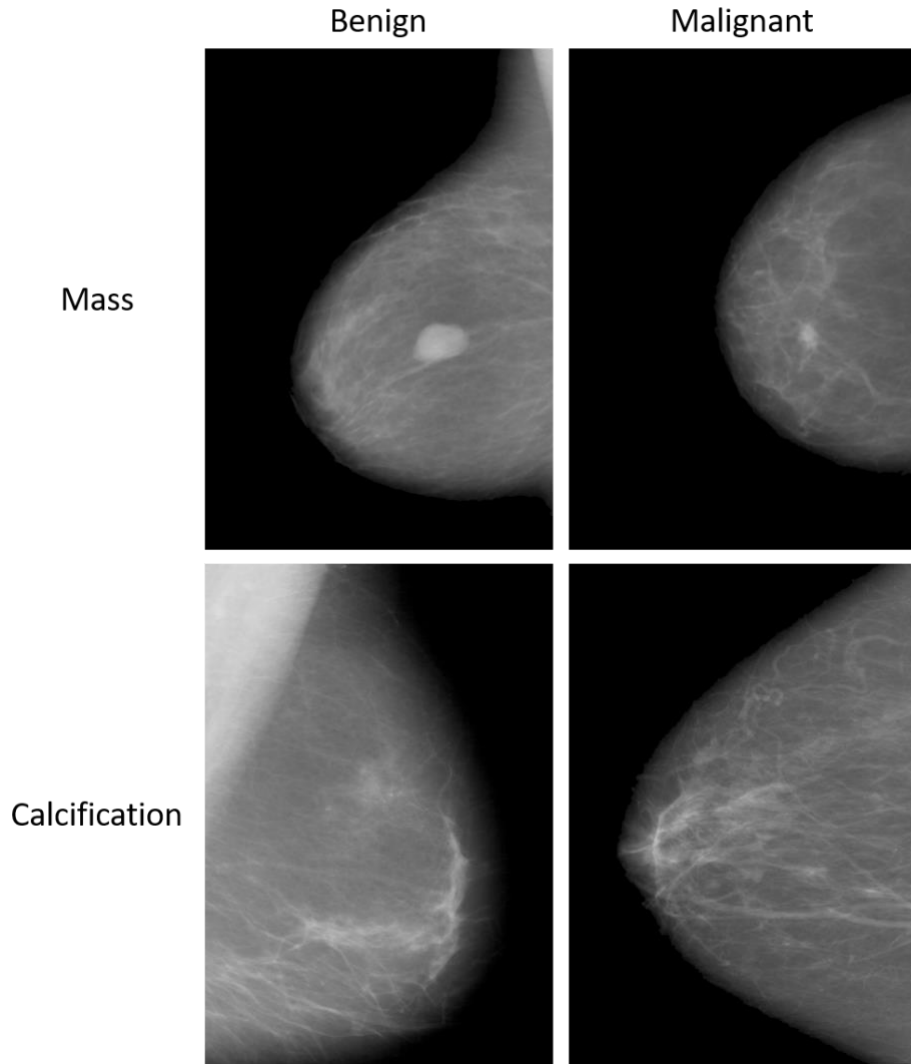


FIGURE 30: Samples of mammograms from the CBIS-DDSM dataset. Cases with benign and malignant images of Mass and Calcification lesions

b. INbreast dataset

INbreast [159] is a public dataset of images acquired using the *MammoNovation Siemens* Full-Field Digital Mammography (FFDM) that are stored in DICOM format. Normal mammograms were eliminated from the database, which comprises 410 mammograms, 235 of which had anomalies in both MLO and CC views from 115 patents. Images are also represented with their annotated ground-truth and have average size of 3328×4084 pixels. There are 45.5% of images that include Mass lesions and 54.5% of images that include Calcifications lesions, and samples are shown in Figure 31.

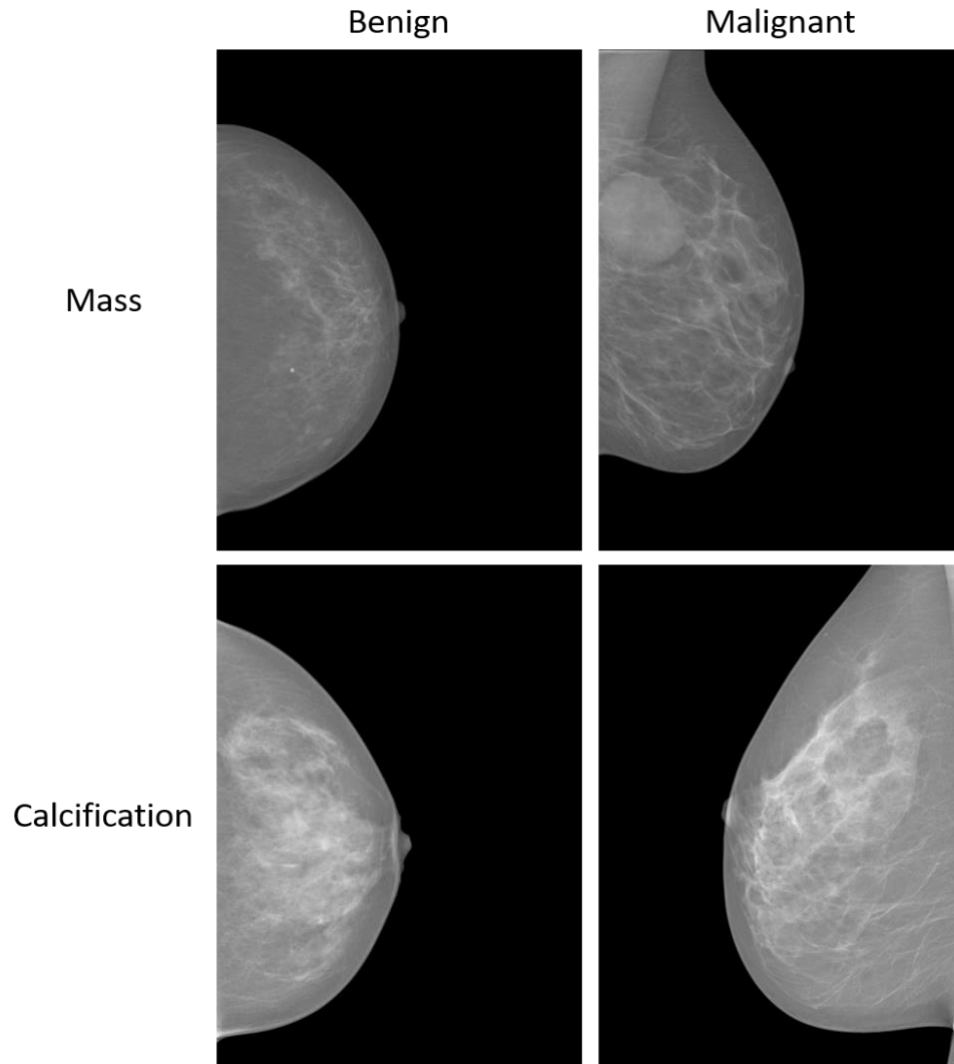


FIGURE 31: Samples of mammograms from the INbreast dataset. Cases with benign and malignant images of Mass and Calcification lesions

2. Private Datasets

a. INCAN private collection

A private dataset was acquired from the National Institute of Cancerology (INCAN) in Mexico City. It contains 489 mammograms with only stage 3 and 4 breast cancer (i.e., malignant) where 487 cases include abnormal lesions from 208 patients, where 80% of images include Mass lesions and the rest includes Calcifications. Images have an average of 300×700 pixels collected from CC, and MLO views. All mammograms may have one or multiple lesions with different sizes and locations and are presented in Figure 32.

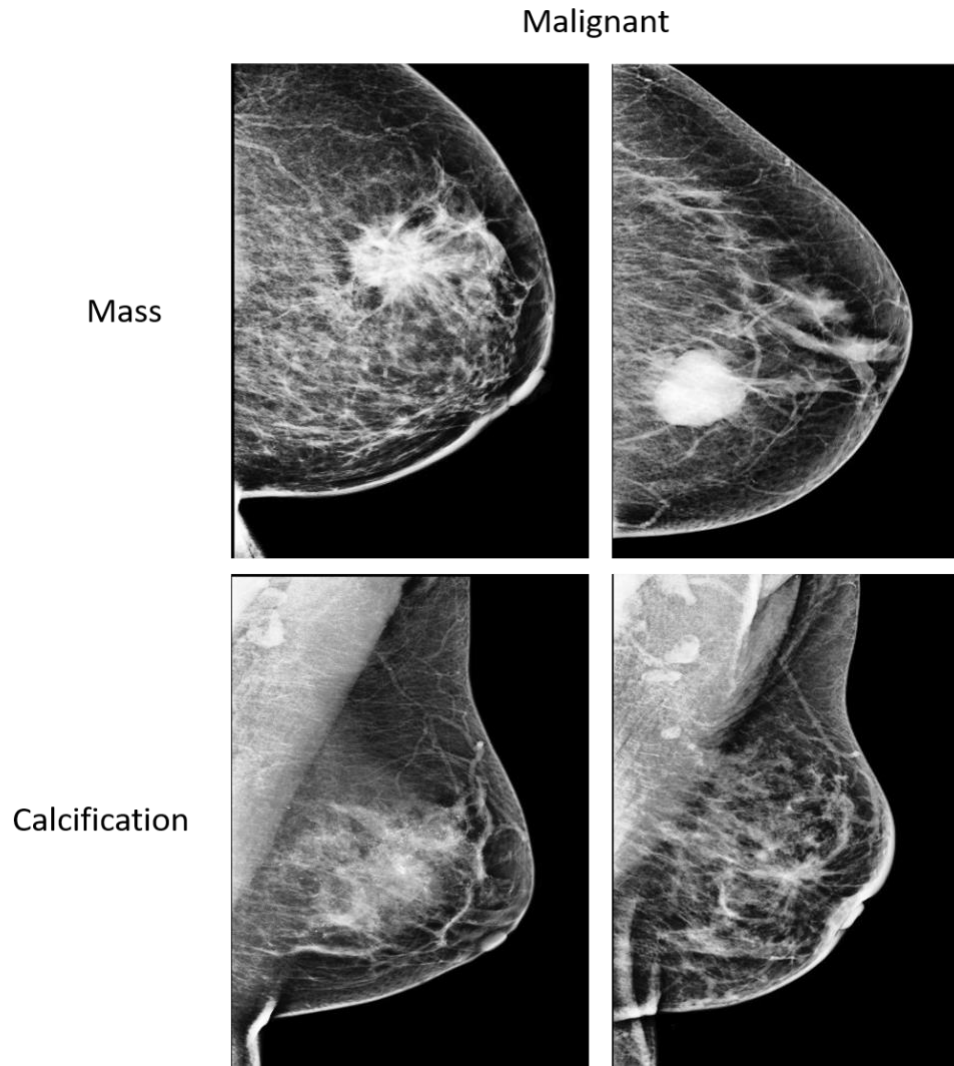


FIGURE 32: Samples of mammograms from the INCAN private dataset. Only cases with malignant images of Mass and Calcification lesions

b. UCHCDM private collection

Another collection of private datasets was acquired from the University of Connecticut Center (UHC), named UHC DigiMammo (UCHCDM) database [160]. The dataset contains screening mammograms of 230 patients, where each case had an initial screening, called Prior exam, and a second follow-up screening between 1 to 6 years, called the Current exam, and samples are displayed in Figure 33.

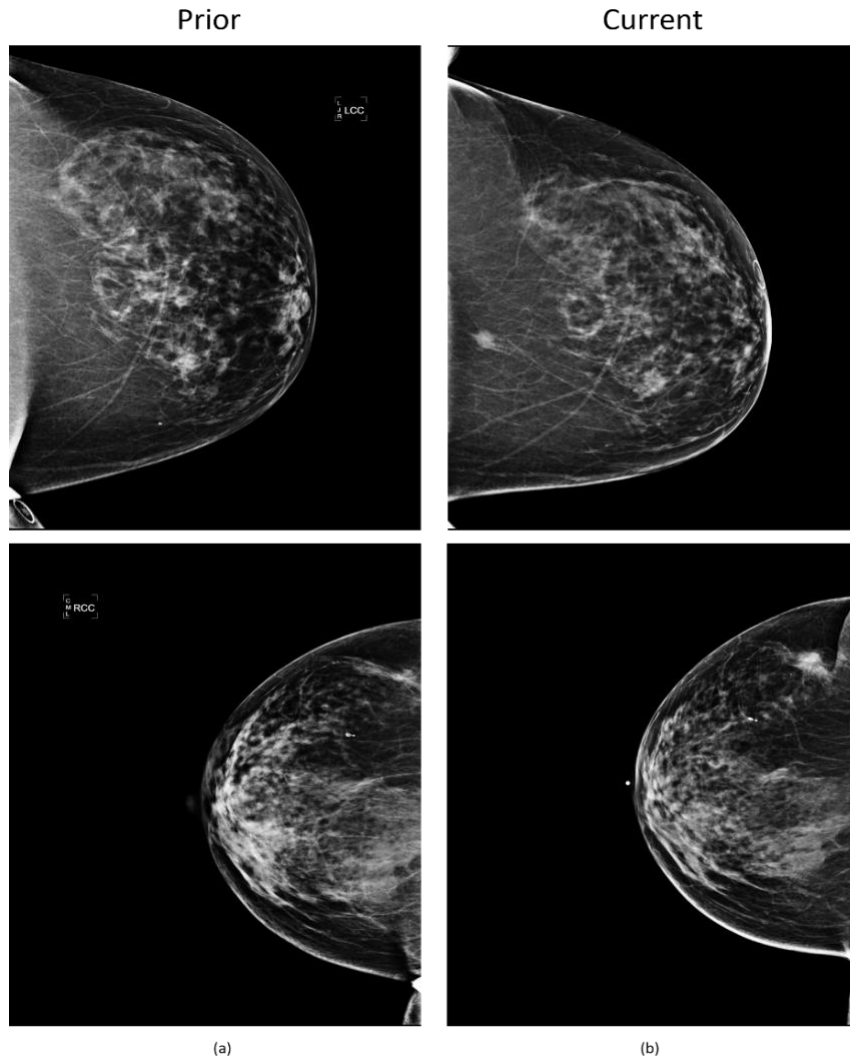


FIGURE 33: Samples of Prior and Current mammograms screenings

Row 1: Case# 31 for Right CC View and Prior exam 2.5 years. Row2: Case#14 for Left CC View and Prior exam 3 years.
 (a) Prior exam with Normal mammogram (i.e., No diagnosis). (b) Current exam with Mass present.

This dataset was specially used for the early detection in prior screening views using a retrospective analysis. Each screening in the dataset acquires two different views, CC, and MLO. All images were saved with the Digital Imaging and Communications in Medicine (DICOM) format and were annotated by expert radiologists in a description text file with corresponding pathology of a mammographic finding (i.e., Mass, Calcification, Architectural Distortion, Normal), as shown in Figure 34. Pixel-level ground-truth images were also provided separately where suspicious locations were circulated. A total of 413 mammograms are

considered separately for Current (cancerous and healthy) and Prior exams (i.e., non-cancerous), and they have an average size of 2950 x 3650 pixels.

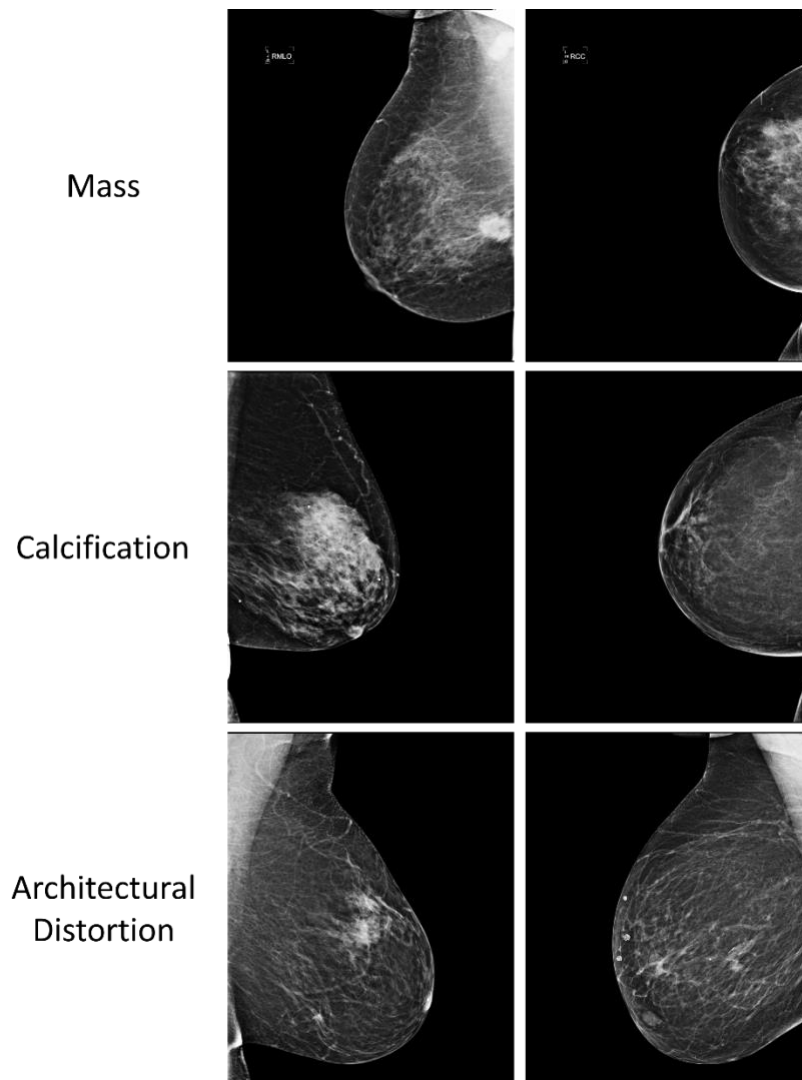


FIGURE 34: Samples of Current mammograms screenings including cases with Mass, Calcification, and Architectural distortion

3. Datasets annotation distribution

All the tasks that were conducted during the CAD system require ground-truth annotation and class labels to train the machine learning techniques. The proposed methods and models should use labeled datasets to classify the images according to each task and predict the outcomes precisely and accurately. All the deep learning models we suggested in this dissertation are based on a supervised machine learning technique that adjusts the model's weights through iterations until the model is fitted appropriately. Each subsection below

describes how the data is distributed according to the class labels that were truly annotated by experts and radiologists.

a. Detection and Identification task

During the simultaneous detection and identification of abnormal lesions in mammograms, the proposed YOLO-based fusion model should predict the class label probability of the detected bounding boxes. The data that was fed to the model should have either single class labels, either Mass or Calcification, or multiple labels. In this first step of the CAD system, we used the two public datasets, and the private dataset and details are shown in Table 4. It is to mention that some mammograms may have one or multiple lesions with different sizes and locations.

Table 4: Breast lesions data distribution

Dataset	Mass	Calcification	Total	Unique Patients
CBIS-DDSM	1,467	1,440	2,907	1,555
INbreast	107	128	235	410
INCAN	390	97	487	208

* We only consider images with existing Ground truth (pixel-level) for the segmentation task

b. Early detection of abnormalities in prior mammograms task

In this step, the first task of detection and identification was re-evaluated on an external private dataset to detect different breast lesions and classify them into Mass, Calcification, or Architectural Distortion and the rest to Normal as detailed in Table 5. After that, the proposed framework was furthermore developed to integrate the Prior mammograms corresponding to all used follow-up screenings and provide an early detection and classification on initial screened mammograms. The work evaluated the retrospective prediction on Prior mammograms that were diagnosed as Normal but at a later stage, they were reported with a clear presence and progress of abnormal findings.

Table 5: Current and Prior data distribution

Dataset	Normal	Mass	Calcification	Architectural distortion	Total	Unique Patients
Prior views	413	0	0	0	413	230
Current views	42	181	116	74		

c. Mass segmentation task

After detecting the location of abnormal lesions and distinguishing between their types, the outcome of the first step is coordinates of the ROI images that surround each abnormality and indicate their class label probability. In this dissertation, we only focused on ROI masses due to the lack of pixel-level annotation of other lesions in some datasets. Hence, the detected ROIs of only mass cases were retained for the second step of segmentation and samples of entire mammograms and their ROI masses are illustrated in Figure 35. It can be visually observed that the images have different resolution, and this is due to the different modality and tools configuration that were used to acquire and store the mammograms.

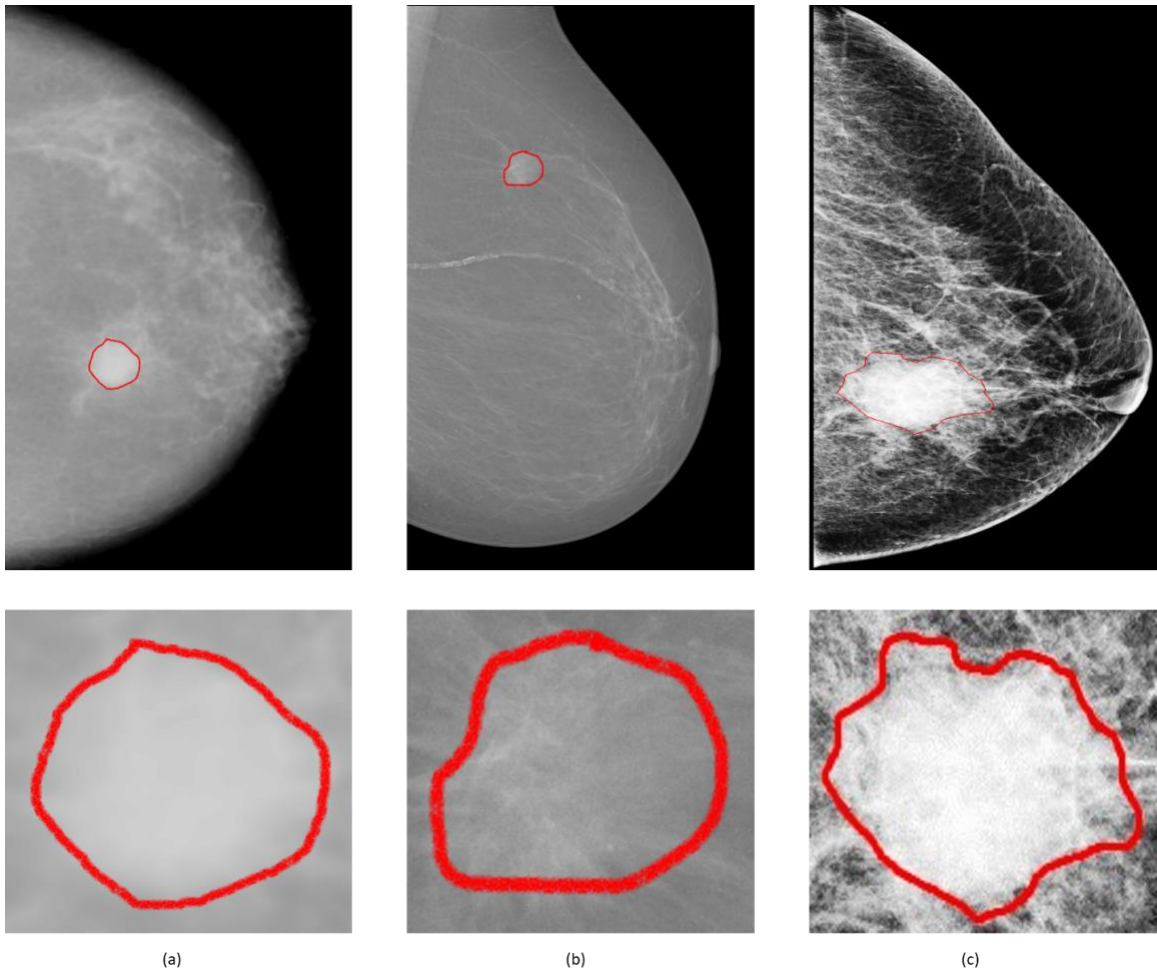


FIGURE 35: Samples from the public and private mammography datasets with zoomed-in ROI of mass ground truth (red). (a) Example from the CBIS-DDSM mammogram with MLO view, (b) Example from the INbreast mammogram with MLO view, and (c) Example from the INCAN mammogram with CC view

d. Tumor Classification and Diagnosis task

In the last step, after segmenting the detected ROI of masses, the predicted binary masks showing only the tumors' contours were applied on the ROIs images. The outcome of the second step should predict the pixel-level labels to mask the background of the breast mass and emphasize the texture and the morphology of the tumors. Figure 36 illustrates samples of original mammograms and their ROI masses compared to the detected and segmented ROI masses from different datasets.

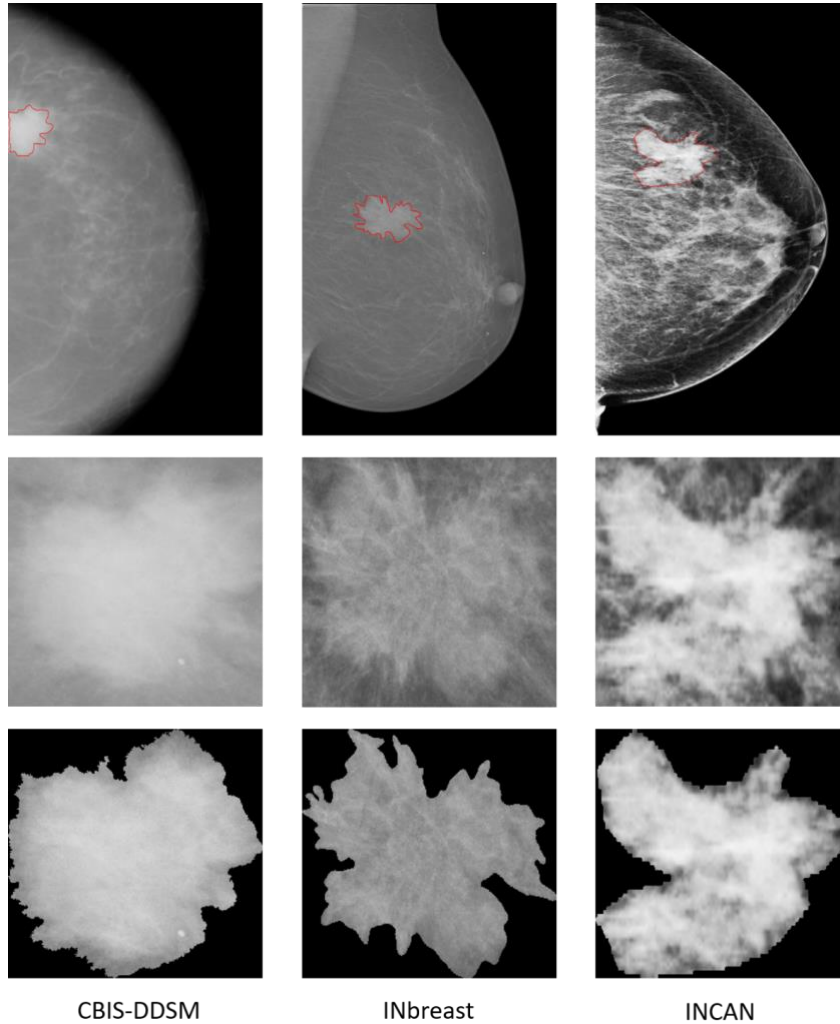


FIGURE 36: Samples of entire mammograms and ROI of detected and segmented Masses from different mammography datasets with their ground-truth of location and contour in red

Each mammography dataset has a different quality of images in terms of pixel quality, existing annotated labels, and class distribution, as detailed in Table 6, 7 and 8. In fact, only the CBIS-DDSM dataset includes true class labels for lesions' shape. Accordingly, the INbreast dataset indicates cases with BI-RADS

score from 2 to 6, however the CBIS-DDSM dataset presents cases in the BI-RADS category 2 to 5, and both INCAN and UCHCDM private datasets have only malignant cases as they indicate breast cancer cases from only stage 3 and 4. Consequently, all mammograms from the private datasets fall into BI-RADS categories 4 and 5.

Table 6: Pathology class labels distribution for only Mass lesions

Dataset	Pathology	
	Benign	Malignant
CBIS-DDSM*	717	750
INbreast**	44	63
INCAN***	0	390
UCHCDM***	0	181

* Cases with Benign_without_callback are considered Benign

** Cases with BI-RADS score > 3, are considered Malignant otherwise Benign

*** Only Malignant Cases with BI-RADS score = 4 and 5

Table 7: BI-RADS category class labels distribution for only Mass lesions

Dataset	BI-RADS				
	Category 2	Category 3	Category 4	Category 5	Category 6
CBIS-DDSM	189	323	388	567	0
INbreast	22	13	21	43	8
INCAN	0	0	211	287	0
UCHCDM	0	0	105	76	0

Table 8: Shape class labels distribution for only Mass lesions

Dataset	Shape			
	Round	Oval	Lobulated	Irregular
CBIS-DDSM	132	341	352	642

4. Data preparation

All mammogram images were collected using the scanning technique of digital X-Ray mammography that usually compresses the breast. The technique may generate deformable breast regions and degrade the quality of mammography images [161, 162]. Therefore, some preprocessing steps could be applied to correct the data and remove additional noise [47, 63]. In this work, we only applied the histogram equalization

technique on the public and private datasets to enhance any compressed region and create a smooth pixel-equalization that could help distinguish suspicious regions from the normal regions.

Furthermore, our suggested YOLO-based model requires input mammograms and their corresponding ground-truth coordinates of regions of interest (ROI) that surround the breast lesions. According to the existing annotation that represents the experts' annotations, we extracted the lesion coordinates represented in x, y, width and height and the class (Mass or Calcification) accordingly. Next, all mammograms were resized using a bi-cubic interpolation over a 4 x 4 neighborhood. For experimental reasons, we used images sizes of 448 x 448 because the input size should be divisible by 32 according to DarkNet backbone architecture of YOLO-V3, and this size should also fit on the GPU memory. Samples of original and preprocessed images are illustrated below in Figure 37.

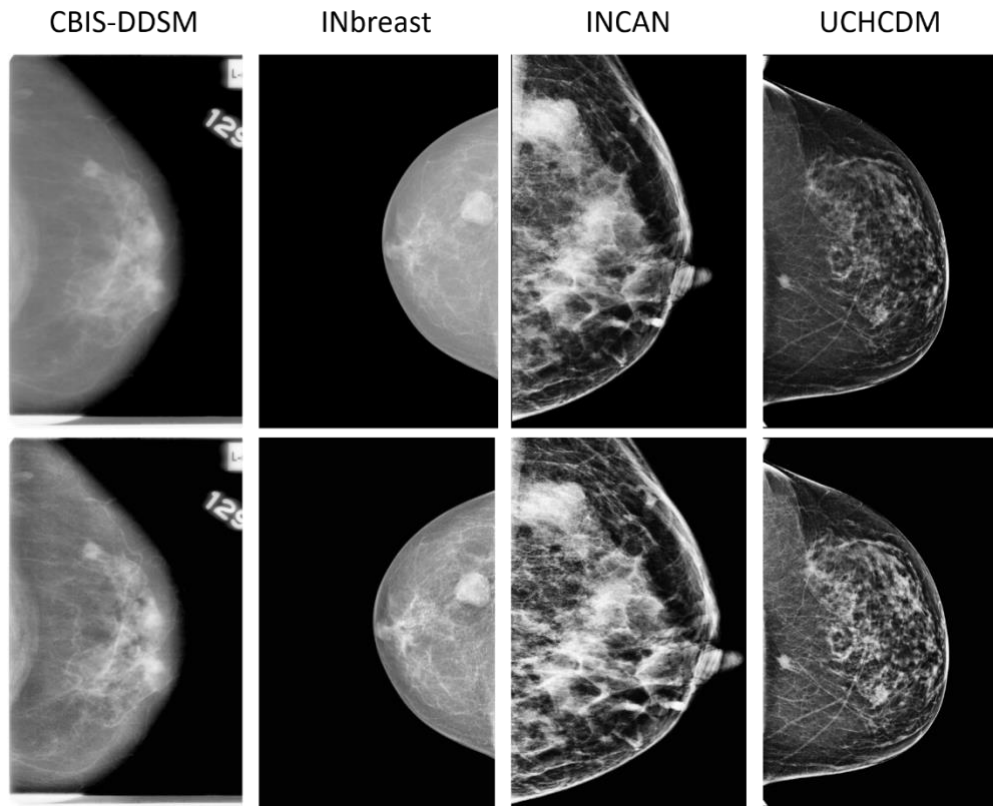


FIGURE 37: Samples from Current exams of original (upper row) and preprocessed mammograms (bottom row)

Deep learning models require a huge quantity of labeled data to be trained, which improves in their generalization. For medical applications, most of the collected datasets have small numbers of instances and often suffer from an imbalanced distribution, which remains a challenge for training deep learning models

[163]. Many researchers have lately used two techniques to solve this problem: data augmentation and transfer learning. Data augmentation is a technique for extending the quantity of a dataset in an experimental setting. In this work and for the detection task, we augmented the original mammograms six times. First, we rotated original images with the angles $\Delta\theta = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ and we transformed them using Contrast Limited Adaptive Histogram Equalization (CLAHE) method [164] with two variations {tile grid size of (4, 4) and a contrast threshold of 40, tile grid size of (8, 8) and a contrast threshold of 30}. Thus, a total of 18,909; 1,410; 2,922 and 1,652 mammograms were respectively collected for the CBIS-DDSM and INbreast public datasets, and the INCAN and UCHCDM private datasets to train and test the proposed model.

The initialization of the trainable parameters is the first step in deep learning models (i.e., weights, biases). To do that, there are two commonly adopted methods: random initialization and transfer learning [165, 166]. We exclusively used the transfer learning approach in our work, which involved applying the weights of a pre-trained model on a larger annotated dataset (i.e., ImageNet, MSCOCO, etc.) and then re-training and fine-tuning the new weights on our unique task and augmented dataset. This helped to accelerate the convergence and avoid overfitting problems. As a result, we employed the weights that were learned on the Microsoft COCO dataset using the DarkNet backbone architecture. The pre-trained model architecture was originally based on the VGG-16 model.

In the segmentation stage, only detected and classified masses from the two public dataset and the INCAN private dataset there were correctly predicted by the YOLO model were considered and the false predictions were discarded as similarly highlighted in previous works [167, 168]. The UCHCDM private dataset was not used because it does not acquire pixel-level annotation.

Some cases of mammograms have more than one detected mass lesion, therefore, a total of 1,467, 112, and 638 masses were respectively considered from the CBIS-DDSM, INbreast, and the INCAN dataset. Because our network only works on single detected ROIs, we wanted to examine mammograms with multiple lesions at the detection stage and treat them as single mass lesions ROIs. The estimated ROI masses were then downsized into 256 x 256 using either a bi-cubic interpolation or an inter-area resampling interpolation, depending on whether the original size was small or big. All images were preprocessed to remove additional noise and degradation caused by the scanning technique of digital X-ray mammography [49, 80]. As a result, we used histogram equalization to improve the compressed regions and smooth the pixel distribution, which

also improves the pixel segmentation. All images were normalized to a range of [0, 1]. Samples of ROI images of masses and their pixel-level mapping are shown in Figure 38.

To generalize the learning curve of the proposed segmentation deep learning models, many annotated samples need to be provided. Because each dataset has a restricted number of ROI masses, we have augmented the original ROIs four times by rotating them at angles of 0° , 90° , 180° , and 270° . We also used the Contrast Limited Adaptive Histogram Equalization (CLAHE) approach to alter them twice. To train and evaluate the suggested architectures, raw data of single ROI images were augmented six times into a total of 8,802; 672, and 3,828 ROI masses were prepared, respectively, from the CBIS-DDSM, INbreast, and INCAN private datasets.

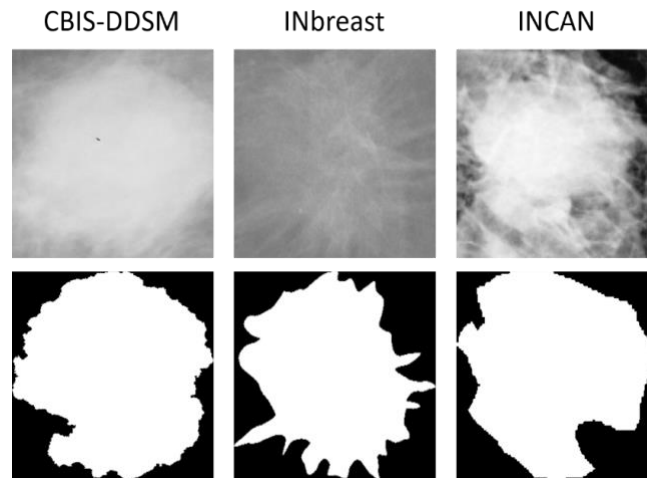


FIGURE 38: Samples of ROI masses and their ground-truth masks from each mammography dataset used for segmentation

Finally, the detected and segmented ROIs from the previous steps are used for the last classification and diagnosis step. The outcome of the last third step defines the diagnostic decision of the CAD system that provides a class label probability of the pathology, the BI-RADS category or/and the shape. Furthermore, the pre-trained ResNet models, which were used for the last step, require input image size of 224×224 ; therefore, the detected and segmented ROIs were down sampled from 256×256 using an inter-area resampling interpolation. Finally, all images were normalized to a range of [0, 1]. Samples of input data for each classification class are illustrated in Figure 39 where ROIs are distributed according to different class labels from the mammography datasets.

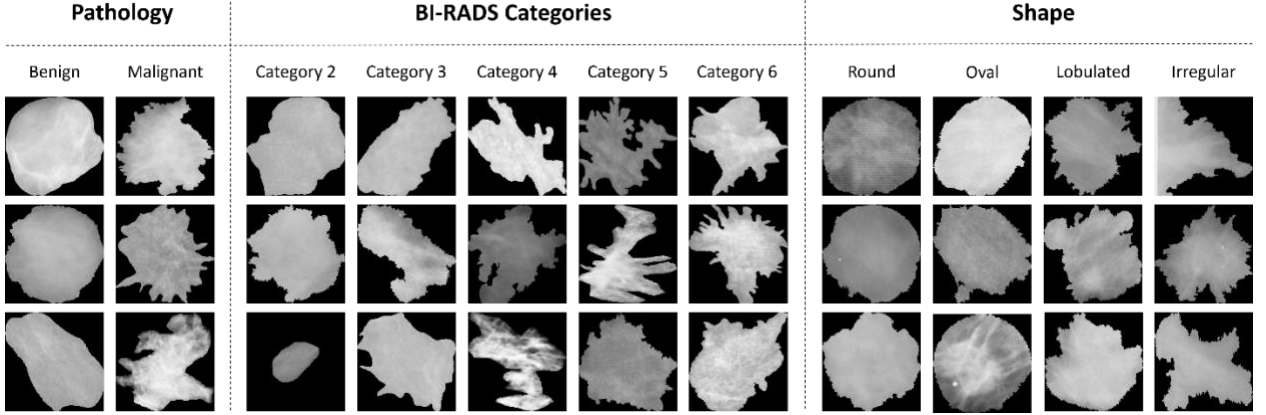


FIGURE 39: Samples of the detected and segmented ROI masses for each class within different classification tasks from different mammography datasets

C. Evaluation Metrics & Experiments Setups

1. Detection and Identification metrics

In this dissertation, we used only object detection and identification measures to evaluate the performance of our proposed YOLO-based model. To ensure the true detection of breast lesions in the mammograms, we first measured the intersection over union (IoU) score between each detected box and its ground-truth (i.e. (x, y, h, w) coordinates and class label), and then we verified if it exceeded a particular confidence score threshold of 0.35 that will be discussed later. Eq. (1) details the IoU score formula.

$$IoU\ score = \frac{Area\ of\ Intersection}{Area\ of\ Union} \quad (1)$$

We also used an objective metric that considered the genuine detected boxes' estimated class probability. Inspired by the work [139], we computed the number of true detected masses and calcifications over the total number of mammograms as defined in Eq. (2).

$$Detection\ accuracy\ rate = \frac{True\ detected\ cases}{Total\ number\ of\ cases} \quad (2)$$

This means that before estimating the final detection accuracy rate, we removed cases with a lower IoU score. Predicted boxes with confidence probability scores equal to or greater than the confidence score threshold were only considered when calculating the final detection accuracy rate. We measured the detection accuracy rate globally and for each independent class to evaluate the performance of the simultaneous detection and classification.

2. Retrospective early detection metrics

In this part, the YOLO-based fusion model was trained on the Current views of the UCHCDM private dataset. The same evaluation metrics for detection and identification were used as explained above. Additionally, we particularly reported the Current mammograms prediction results using the area under curve (AUC) that reflects the performance of the model and the trade-off between the true positive rate and false positive rate for each target class label. We used three additional metrics called precision, recall, and sensitivity that are computed using the TP, FP, and FN that are defined per predicted class to represent the number of true positive, false positive, and false negative predictions, respectively as shown in Eq. (3), Eq. (4), and Eq. (5).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Sensitivity = 1 - FNR = 1 - \frac{FN}{FN + TP} \quad (5)$$

The early detection of abnormalities in Prior mammograms was evaluated using two metrics, a true prediction rate and false prediction rate. We considered a true prediction where the location and type of breast lesions were correctly captured using the inference model, retrospectively on non-cancerous screening views at $t=0$ years. The inference evaluation was concluded using the ground-truth labels of the Current views that were generated by experts later at $t= [1, 6]$ years. Consequently, all predictions should be fairly compared to 0 predictions (i.e., all were missed) by experts at $t=0$. Both metrics were compared to the “Experts prediction” that missed detecting the risk of future breast cancer.

3. Segmentation metrics

Segmentation stage is evaluated using the Dice similarity score, also called F1-score, which represents a coupled average of the intersection between areas and the total areas as indicated in Eq. (6). Accordingly, we use another evaluation metric, the Intersection Over Union (IoU) score, also called the Jaccard score, which is detailed in Eq. (7). The pixels surrounding all the masses are correctly segmented, resulting in a binary mask being formed from the segmented contour of the mass lesions with a high Dice score and IoU score.

$$Dice\ score(A, B) = \frac{2 \times Area\ of\ Intersection(A, B)}{Area\ of\ (A) + Area\ of\ (B)} = \frac{2 \times (A \cap B)}{A + B} \quad (6)$$

$$IoU\ score(A, B) = \frac{Area\ of\ Intersection(A, B)}{Area\ of\ Union(A, B)} = \frac{A \cap B}{A \cup B} \quad (7)$$

To train the proposed architecture models, a learning rate of 0.0001 with Adam optimizer is employed. A weighted sum of Dice and IoU losses is used as a segmentation loss function using the Dice score and IoU score between true and predicted samples, as detailed in Eq. (8).

$$\begin{aligned} Segmentation\ loss(true, predicted) & \quad (8) \\ & = -(0.4 \times Dice\ score(true, predicted) + 0.6 \times IoU(true, predicted)) \end{aligned}$$

To evaluate our integrated framework system, we first define a segmentation accuracy measure to be the mean IoU score for correctly identified ROIs based on a 90% overlap threshold and we refer to it as IoU_{90} score as shown in Eq. (9). Then, a final segmentation accuracy is introduced as an end-to-end accuracy for the two stages, explained in Eq. (10).

$$\begin{aligned} IoU_{90}\ score & = \{mean(IoU\ scores\ \forall\ ROIs) & \quad if\ IoU\ score(A, B) & \quad (9) \\ & \geq 90\ Not\ applicable & \quad if\ IoU\ score < 90 \end{aligned}$$

$$Final\ segmentation\ accuracy = Detection\ accuracy\ rate \times IoU_{90}\ score \quad (10)$$

4. Classification and Diagnosis metrics

All classification tasks are evaluated overall using the accuracy, and area under curve (AUC) that reflect the performance of the model while considering the unbalanced mammography datasets. Particularly, for pathology classification, which presents a binary-class case, we use three additional metrics called sensitivity, specificity scores, and F1-score, as shown in Eq. (11), Eq. (12), and Eq. (13). The F1-score is a coefficient that represents a harmonic average between the specificity and sensitivity, where its maximum score of 1 indicates perfect specificity and sensitivity and of 0 the worst performance. Moreover, the accuracy score is a rate of correct predictions over all cases as detailed in Eq. (14) where TP, TN, FP, and FN are defined per predicted class to represent the number of true positive, true negative, false positive, and false negative predictions, respectively.

$$Sensitivity = \frac{TP}{TP + FN} \quad (11)$$

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

$$F1 - score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (13)$$

$$Accuracy = \frac{True\ prediction}{Total\ cases} = \frac{TP + TN}{TP + FP + TN + FN} \quad (14)$$

In the pathology classification, positive refers to the malignant class and negative refers to the benign class. In the BI-RADS category and shape classification, macro-averaging is used to compute the accuracy and the AUC scores. Consequently, a confusion matrix can be driven from these measurements to show the tradeoff between the true and predicted class labels.

5. Experimental setups

The proposed YOLO-based model presents a list of hyperparameters that includes learning rate, number of epochs, dropout rate, batch size, number of hidden units, confidence score threshold and so on. Only three hyperparameters were chosen for tuning because of their impact on the model's performance. For all datasets, we randomly split all mammograms for each class into groups of 70%, 20%, and 10% respectively for training, testing, and validation sets.

Trainable parameters were kept constant while hyperparameters were modified in each experiment. Adam was used to optimize all experimental datasets, and the detection accuracy rate was utilized to report all experiments. First, we set the learning rate to 0.001, number of epochs to 100 and the batch size to 64 according to the work [47], and then we trained the model with different confidence score thresholds until we report the value that provided satisfying detected objects for further tasks (i.e., segmentation and shape classification). As shown in Figure 40a, the best confidence score value for all datasets is 0.35 to accept all detected objects the model confident from them by more than 35%. Next, we repeated the experiments, but we varied learning rate values to report the best detection accuracy rate for all datasets as shown in Figure 40b. In addition, the early stopping strategy for the second half of iterations was used to reduce the learning rate by 10% if the loss function did not decrease every 10 epochs.

Next, we selected the best learning rate, which is 0.001, and we varied the batch size to report the best results for the three datasets as illustrated in Figure 40c. Finally, we modified the number of epochs until all datasets indicated the greatest performance for 100 epochs, as shown in Figure 40d.

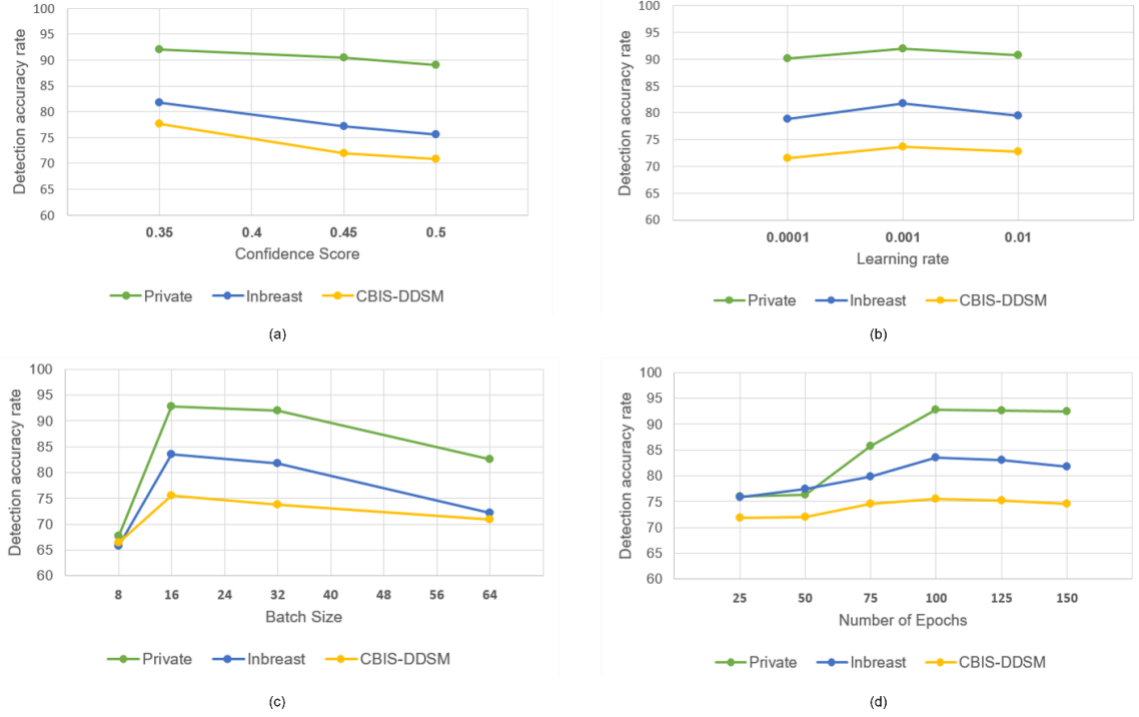


FIGURE 40: Hyperparameters tuning; (a) confidence score; (b) learning rate; (c) batch size and (d) number of epochs

Similarly, for the segmentation step, each mammography dataset is randomly split into groups of 70%, 20%, and 10%, respectively, for training, testing, and validation sets as detailed in Table 9. It is important to highlight that some of the raw MGs have multiple ROIs. Accordingly, 100 epochs and 8 mini batches were used to optimize the network parameters with the training and validation sets.

Table 9: Data distribution of the mammography datasets

Dataset	Raw MGs Data	Raw ROIs Data	Augmented Data (ROIs*6)	Training Data (70%)	Testing Data (20%)	Validation Data (10%)
CBIS-DDSM	1,467	1,467	8,802	6,161	1,760	881
INbreast	107	112	672	470	134	68
INCAN	389	638	3,828	2,679	766	383

Furthermore, for the two image-to-image translation models, they were trained and evaluated on 100 epochs and optimized using the Adam technique with learning rate of 0.0002 and beta score of 0.5.

During the early detection on prior mammograms, the YOLO-based fusion model was differently evaluated using a 5-fold cross validation by training and testing the model using different test sets of random

mammograms to ensure the model robustness. Consequently, the entire UCHCDM dataset was randomly divided into equal 5 folds of 1,324 training images (80%) and 328 testing images (20%) with respect to the imbalanced classes as detailed in Table 10. Finally, we reported the average of results over all the folds.

Table 10: Cross Validation Folds: Data distribution across class labels

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

For all experiments using the YOLO-based model, we set the learning rate to be 0.001, the batch size to be 8, and the number of epochs to be 100. The loss function combined the bounding box regression loss, the class label loss, and the confidence loss. All functions were based on cross-entropy, and they were scaled to handle the imbalance of class labels on each batch. In addition, the early stopping method was used for the second half of iterations to dynamically reduce the learning rate by 10% every 10 epochs in case of constant loss function value. To prevent overfitting, all models were initialized by weights from a pre-trained model on a large public dataset, MSCOCO. Then, the models were re-trained and new layers were fine-tuned on our mammography dataset.

Finally, during the classification and diagnosis step, extensive experiments with different variations in hyperparameters were conducted to select the best parameters for the base ResNetV2 model. Considering their effect on the classification performance, only hyperparameters detailed in Table 11 were tuned to select the best configured network that outperforms the evaluated networks on all mammography datasets.

Table 11: Hyperparameters for the ResNetV2 base model

Hyperparameters	Values explored	Description
Batch size	32, 64	mini-batch training size
Epochs	20, 30, 50	Number of training epochs
Dropout	0%, 20%, 30%	% of neuron of hidden layers “dropped” for regularization
LR	10^{-1} , 10^{-2} , 10^{-3}	Learning rate for the Adam optimizer
Smoothing	0% 20%, 25%	% of label smoothing for the loss function

For all datasets, we randomly split images for each class into groups of 80% for training, 20% divided equally between testing and validation sets. In each experiment, the same trainable parameters were used and each hyperparameter was varied accordingly. For all datasets and classification tasks, we used Adam optimized and evaluation was reported with a weighted accuracy score to reflect the class imbalance during the training and testing. The loss function was employed according to the classification task, a Binary Cross-entropy function for binary classes and Categorical Cross-entropy for multiple classes. In both cases, a label smoothing technique for regularization to help overcome overfitting and provide a generalized model. The technique works by explicitly updating the labels during the loss function and decreasing the model's confidence when it starts diverging [169]. In addition, training was monitored using a method that reduces the learning rate if the accuracy stops improving. Thus, we applied the stated strategy with a factor of 0.5 when the accuracy did not improve after two iterations. Conclusively, the best evaluation was reported with a batch size of 32, 30 epochs, a dropout rate of 30%, a learning rate of 10^{-2} , and a smoothing label of 25%.

D. Evaluation Results

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

1. Detection and Identification

During the first step of the CAD system, different experiments were conducted to assess the effect of varying input images data and target classes (i.e., Mass, Calcification) of our suggested YOLO-based model. To improve the outcomes, more tests were undertaken with the fusion model's technique.

a. Single Models Evaluation

The breast lesions detection and classification model was trained differently over the mammography datasets. We varied the input data fed to the model and configured the classification to be with multiple classes using M^2 . Performance of the model is reported in Table 12.

Table 12: Model performance for different configurations

Dataset	Configuration	Overall Detection accuracy	Inference time per
		rate (%)	image (s)
CBIS- DDSM	Original images + Original size	63.7	1.23
	Augmented images + Resize 448 x 448	73.7	0.57
INbreast	Original images + Original size	75.2	0.98
	Augmented images + Resize 448 x 448	81.7	0.57
INCAN	Original images + Original size	62.4	0.79
	Augmented images + Resize 448 x 448	92	0.57

Results show the advantage of data augmentation and resize over the original mammography datasets. In fact, the performance increased with 10% for CBIS-DDSM dataset with almost half of inference time. Similarly, the model achieved a better detection accuracy rate with more than 6.5% and 40% less inference time. The same improvement with 29.6% is noticed on the INCAN private dataset with a 28% drop-in inference time. Accordingly, using the augmented and resized datasets, we varied the prediction classes by training M^1 independently on Mass and Calcification, and M^2 on both, and results are reported in Table 13 below

Table 13: Model performance for different prediction classes

Dataset	Detection accuracy rate (%)					Inference time per image (s)
	Model1		Model2		Overall	
	Mass	Calcification	Mass	Calcification		
CBIS-DDSM	85.1	62.2	84.7	60.2	73.7	0.55
INbreast	93.4	71.8	93.1	68.7	81.7	0.58
INCAN	97.9	68.5	96.2	67.8	92	0.52

Results show that INCAN private dataset had the highest performance compared with the public datasets and this can be explained with the good resolutions and the easy localization of most of the lesions in those mammograms. Furthermore, the public datasets had more degraded lesions, which are more difficult to identify and classify at the same time. Accordingly, results in Table 13 show the clear ability of the YOLO-based model to better detect and classify the mass lesions from the entire mammograms than the calcification

lesions. This corresponds to the differences in shape, size, and texture of the two types of lesions. In fact, calcifications are often small and randomly distributed in challenging positions within the breast [170]. As shown in Figure 41, calcifications do not have standard shape and they can be bilateral, thick linear, clustered, pleomorphic, and vascular, etc. These varied shapes can limit the detection and classification for this type of lesions and yield more failed cases than for the other lesions. Below in Figure 41, it shows a case of a coarse-like calcification that has crossed thick lines with irregular size (image on the left, taken from the CBIS-DDSM dataset). Pleomorphic calcifications with a random distribution are seen in another example (image on the middle, taken from the INbreast dataset). In addition, an example of clustered calcifications located on the pectoral muscle that presents a challenging case in mammography (image on the right, taken from the INCAN private dataset).

Moreover, we notice that both models have the best results toward mass lesions using the INCAN private dataset, and toward calcification lesions using the INbreast dataset. This can be explained with the degraded quality presented in the digitized X-rays mammograms of CBIS-DDSM dataset. Consequently, performance is affected by the image quality and our study proved that detection and classification highly require full-field digital mammography images which involve direct conversion and preserve the shape and textures of breast lesions [171].

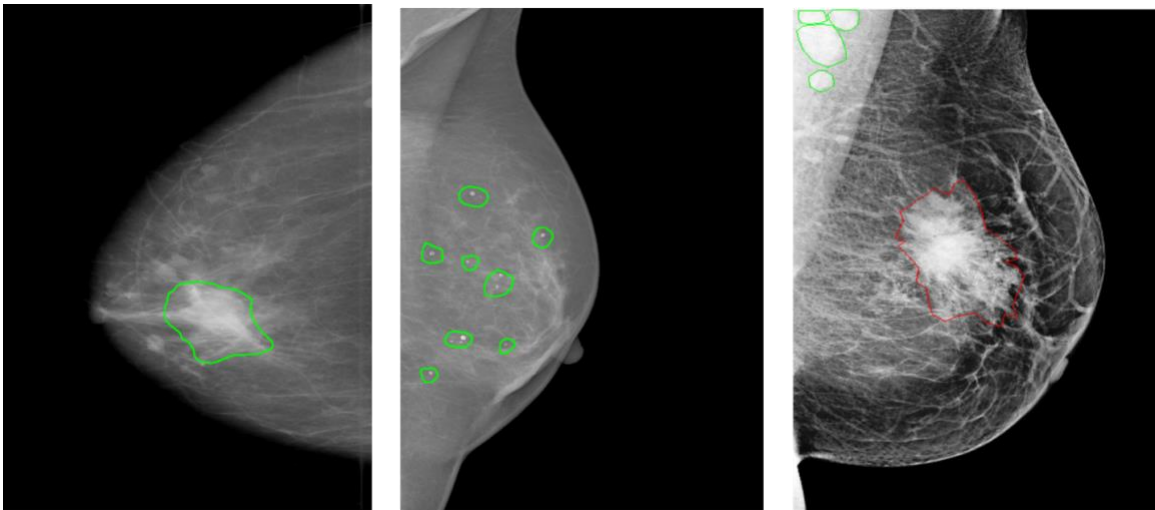


FIGURE 41: Examples of different calcifications shape and localization (ground-truth of calcification is marked in green, ground-truth of mass is marked in red) for CBIS-DDSM, INbreast and INCAN private datasets (from left to right)

Moreover, Table 13 demonstrates that training the model on both prediction classes slightly decreased the performance and this can be explained by the inability of the YOLO-based model to detect and distinguish some different types of lesions having similar shapes. However, we proved the robustness of our suggested model toward mass detection with a maximum detection accuracy rate of 96.2 using the private dataset. Inference time was similar in all studies, with a maximum of 0.58 seconds. Examples from each dataset are illustrated in Figure 42, and each lesion's breast has its confidence score. Multiple lesions were accurately discovered in the same mammogram, as we can see.

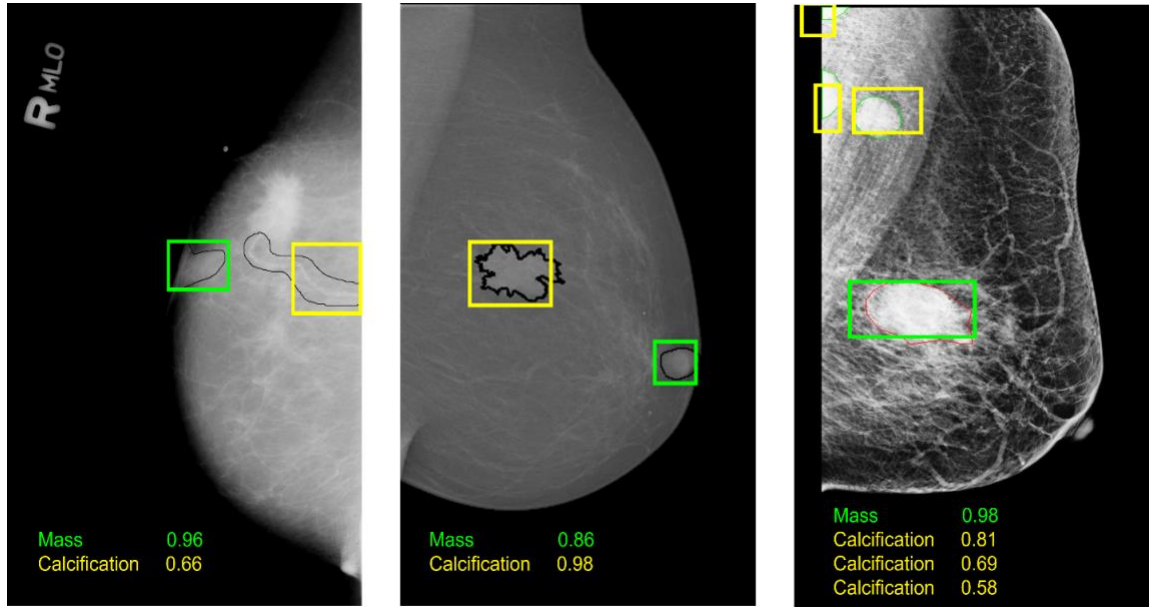


FIGURE 42: Examples of breast lesions detection and classification results and their confidence score toward different classes on CBIS-DDSM, INbreast and INCAN private datasets (from left to right): mass (green boxes) and calcification (yellow boxes)

b. Fusion Models Evaluation

To test the simultaneous detection and classification model, this study introduced an additional step. It presents an expanded evaluation which fuses models that are trained with different settings as detailed in the previous chapter. In fact, before presenting the results, single models M^1 and M^2 were first reported over best-selected mammograms from the augmented datasets. This means for every set of predicted mammograms including the original and their five augmented images (i.e., rotated, transformed), we selected the image having the highest IoU score. Next, different models were fused into a new Fusion model, as detailed in Table 14, and we measured the detection accuracy rate toward every prediction class.

Table 14: Comparison performance using fusion models' approach

Dataset	Detection accuracy rate (%)					
	Model1		Model2		Fusion model	
	Mass	Calcification	Mass	Calcification	Mass	Calcification
CBIS-DDSM	85.1	62.2	84.7	60.2	95.7	74.4
INbreast	93.4	71.8	93.1	68.7	98.1	72
INCAN	97.9	68.5	96.2	67.8	98	73.2

Indeed, the performance of detection and classification using the fusion model was increased for each type of breast lesion compared to the single models. For the CBIS-DDSM dataset, mass lesions had a detection accuracy rate of 95.7%, which is higher than 85.1%. Besides, we boosted the performance with 12.2% for calcification lesions. For the INbreast dataset, we achieved a final detection accuracy rate of 98.1% for mass lesions and 72% for calcification lesions, which are better than results reported for single experiments in Table 13. Similarly, performance was improved for the private dataset with 98% detection accuracy rate for mass lesions and 73.2% for calcification lesions.

Our proposed fusion models significantly increased the detection and classification outcomes on mammography pictures, as can be seen. Indeed, fusion strategies were reviewed in the past for medical image segmentation [172, 173, 174], and our approach is a new decision-level fusion strategy for object detection and classification that proved the advantage of fusing results of multiple models.

Finally, a comparison of mass detection results of the latest studies and similar methods are listed in Table 15. Our method, which is based on fusion models, is sufficiently fast and accurate. Comparing both detection accuracy rate and inference time with the other works shows that we achieved a better overall performance on the public datasets: CBIS-DDSM with a detection accuracy rate of 95.7% and INbreast with a detection accuracy rate of 98.1%.

It's worth noting that comparing results with state-of-the-art approaches were based on both detection accuracy rate and testing inference time, thus while Al-Antari et al. [62] exceeded the INbreast detection results, it was more expensive in terms of inference time than our implementation. Furthermore, the trials in each paper used distinct preprocessing strategies that performed differently on both standard datasets. The framework of the YOLO-based fusion model for the detection and identification of breast cancer

abnormalities was published in the Computers, Materials & Continua Journal at the Recent Advances in Deep Learning for Medical Image Analysis Special Issue [175].

Table 15: Comparison of mass detection with other works

Reference	Year	Method	Dataset	Detection accuracy rate (%)	Inference time per image (s)
Dhungel et al. [167]	2017	Cascade Deep Learning and Random Forest	INbreast	96	39
Kozegar et al. [31]	2013	Adaptive threshold machine learning	INbreast	87	108
Agarwal et al. [32]	2019	CNN patch classifier and mass probability map (MPM)	CBIS-DDSM	82	NA
			INbreast	98	
Aly, G. et al. [52]	2021	YOLO	INbreast	89.5	0.009
Peng et al. [61]	2020	Faster R-CNN	CBIS-DDSM	93.45	0.134
			INbreast	95.54	
Al-Antari et al. [62]	2018	YOLO	INbreast	98.96	3
Al-Antari et al. [63]	2020	YOLO	INbreast	97.27	0.025
Singh et al. [89]	2020	Single Shot Detector (SSD)	INbreast	97	NA
Proposed Methodology	2021	YOLO-based Fusion Models	CBIS-DDSM	95.7	0.55
			INbreast	98.1	0.58
			INCAN	98	0.52

2. Early detection of abnormalities in prior mammograms

a. Evaluation of YOLO-based model on Current Mammograms

First part of the study considered only Current mammograms from the most recent screening exams. The YOLO-based models were trained differently over the Current views of the UCHCDM private dataset. We varied the models according to the input dataset and the target class. Hence, Model1 was configured for

single classes and Model2 was configured for mixed classes. Finally, the Fusion Model was designated to combine Model1 and Model2 for each target class according to the approach previously. Table 16 shows quantitative comparison of the detection accuracy rate and count that were reported using the 5-fold cross validation as $\mu \pm \sigma$, where μ and σ refer to the mean and standard deviation, respectively.

Table 16: Comparison performance for different models across labeled classes on Test sets

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

Results show the advantage of the adapted Fusion Model and confirm its highest results overall and for each class label. Fusion Model had the highest score of 95% for Architectural Distortion lesions and a score of 92% overall. Moreover, results in Table 16 show the ability of YOLO architecture to detect and classify the breast lesions with a maximum accuracy rate of 93% for mammograms with Mass lesions, 88% for mammograms with Calcification lesions, and 95% for mammograms with Architectural Distortion lesions. Appropriately, Normal mammograms were also correctly classified with a maximum accuracy score of 94% where no bounding boxes were detected. All experiments had similar inference time with a maximum value of 0.62 seconds per image.

Additionally, to get a better understanding of the models' performance, we generated the free-response receiver operating characteristic (FROC) curves to illustrate the number of false positives per image (FPI) for each target class label. Plots of the FROCs between Average sensitivity and the average number of false positives are shown in Figure 43 that specifically compares Model1, Model2, and the Fusion Model.

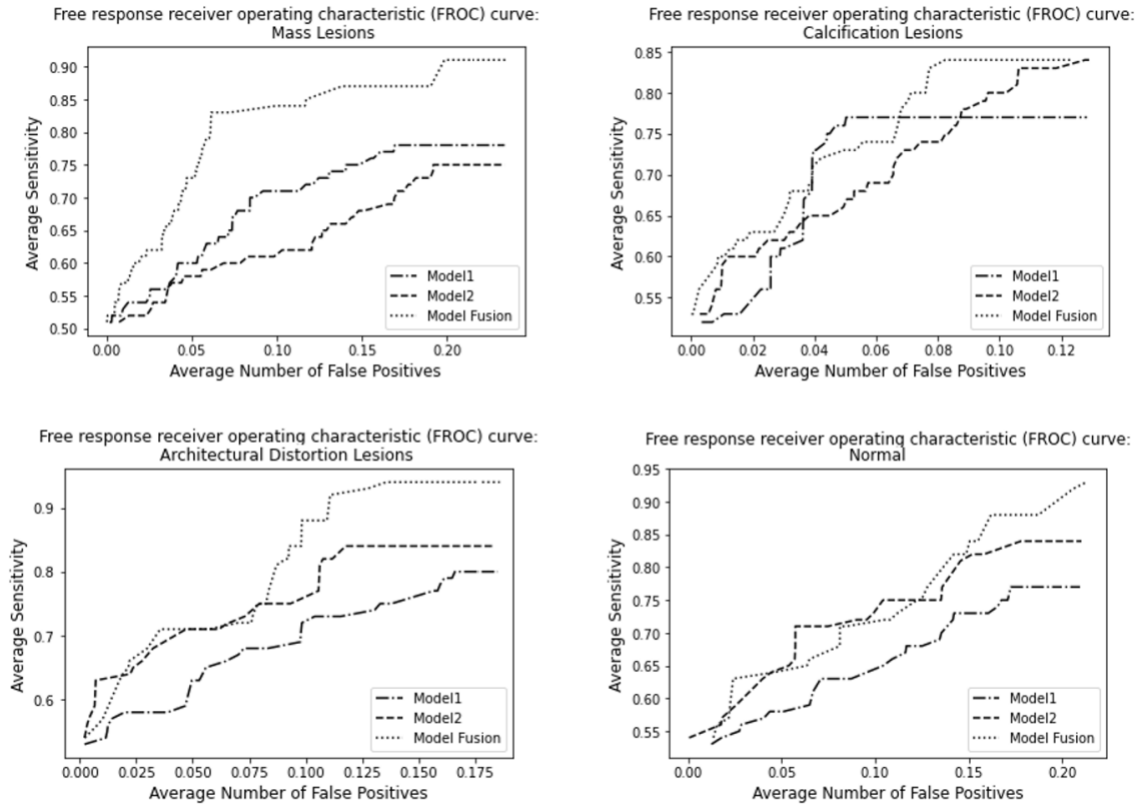


FIGURE 43: FROC curve plots of the YOLO based proposed Detection and identification models per class label on Test sets

By varying the threshold and the range of false positives between 0.05 and 0.20 overall, we could achieve an average sensitivity between 0.7 and 0.95 for all cases. Figure 43 clearly shows that the Fusion model had the highest performance compared to the other evaluated models. It is observed that the proposed model could obtain an average sensitivity of more than 0.90 with an average FPI of 0.20 for Mass lesions, an average sensitivity of more than 0.85 with an average FPI of 0.12 for Calcification lesions, and an average sensitivity of more than 0.90 with an average FPI of 0.175 for Architectural Distortion lesions. Accordingly, Normal cases in Current views were evaluated using the FROC analysis and a false positive was considered when no detection should have occurred in a non-cancerous case, but it was missed by the model. It is to notice that we could obtain an average sensitivity of around 0.95 with an average FPI of 0.20.

Finally, we analyzed the performance results with a particular focus on the classification task that was conducted by the YOLO-based Fusion model. Table 17 explores the calculated classification metrics by

each class label, where we achieved the highest sensitivity of 94.11% on the cancer cases with Architectural Distortion and a sensitivity of 92.09% on the non-cancerous cases.

Table 17: Performance results for Detection and Classification on Test sets

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

Additionally, Figure 44 illustrates a visual comparison of the trade-off between the false positive rate (FPR) and the true positive rate (TPR) according to the ROC curve plot between the different cases. We observed a highest AUC score of 0.95 for the Mass and the Architectural Distortion cases, and an AUC score of 0.96 for the Normal cases. The low results for the Calcification lesions could be explained with the fact that this type of breast lesions do not have standard shape and location and they are often small and randomly distributed which can limit the automatic detection.

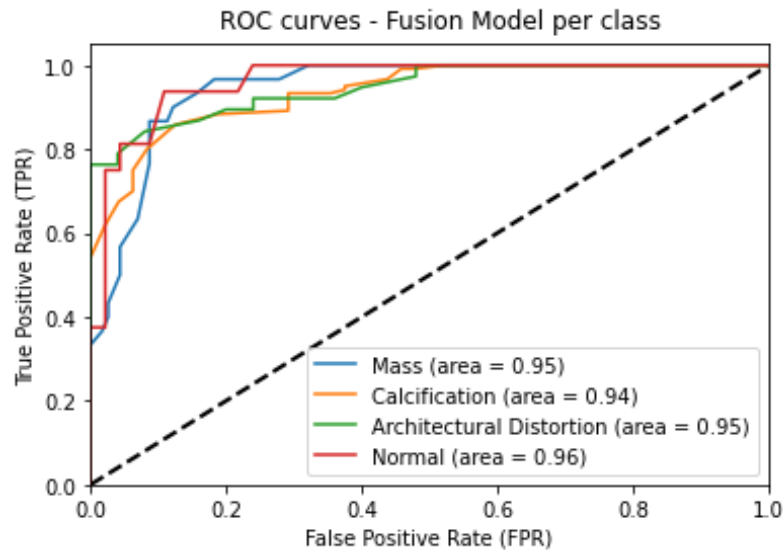


FIGURE 44: ROC curve plots of the proposed YOLO-based Fusion Model per class label on Test sets

Moreover, Figure 45 illustrates the confusion matrix for the classification of the true detected bounding boxes applied on the Current mammograms, where three types of lesions are presented with the Normal cases (i.e., correct prediction without detected lesions). Clearly, the prediction error for different classes is low with a high rate of 6.2% corresponding to the Normal class label and 7.6% corresponding to

the Calcification cases. The distribution of error within classes could be explained by the inability of the YOLO-based model to detect and distinguish some different types of lesions having similar shapes such as Calcification and Architectural Distortion that often have irregular shapes in challenging positions within the breast.

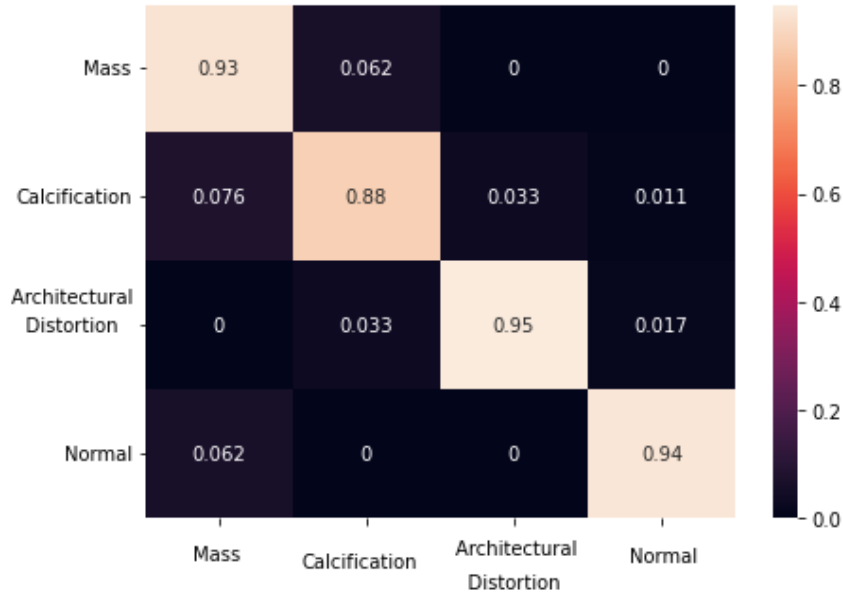


FIGURE 45: Confusion matrix of prediction results for Current Mammograms

b. Evaluation of YOLO-based model on Prior Mammograms

Second part of the study focused on using the pairs of mammograms, including current views and their Prior screening exams to provide an early detection and classification of lesions on the Prior screening exams. All Prior mammograms were not annotated with diagnosis and thus were considered Normal (i.e., non-cancerous, corresponding to 0s in “Experts prediction” row in Tables 18a, 18b, and 18c). In this part, we introduce a retrospective approach to look back at the Prior mammograms and try to explore any patterns of breast lesions before waiting on a follow-up screening.

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

exams were generated to resemble the Current mammograms and preserve the general texture of the Prior mammograms.

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

Table 18a, 18b and 18c present results of early prediction on Prior mammograms that are reported using the 5-fold cross validation as $\mu \pm \sigma$, where μ and σ refer to the mean and standard deviation, respectively. We considered a true prediction where the location and type of breast lesions were correctly captured using the inference model, retrospectively on non-cancerous screening views at $t=0$ years. The inference evaluation was concluded using the ground-truth labels of the Current views that were generated by experts later at $t= [1, 6]$ years. Consequently, all predictions should be fairly compared to 0 predictions (i.e., all were missed) by experts at $t=0$.

Table 18a: Inference results of YOLO Fusion model on Test sets of original Prior Mammograms

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

Table 18b: Inference results of YOLO Fusion model on Test sets of Prior Mammograms using CycleGAN method

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

Table 18c: Inference results of YOLO Fusion model on Test sets of Prior Mammograms using Pix2Pix method

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

Results in Table 18a, 18b and 18c show the count and percentage of mammograms for each class and overall, that were correctly predicted at Prior views and considered for an early detection and classification. All true predictions presented two scenarios; one for all correct predictions on both Current mammograms and their corresponding Prior views from the first exams (i.e., $t=0$), and another scenario for only correct prediction on Prior mammograms even though their corresponding Current views were not correctly predicted.

It is observed that the highest results were reported by the YOLO-based model that was inferred on synthetic Prior mammograms by Pix2Pix technique, where a total number of 52 mammograms (36% \pm 0.01)

were accurately anticipated. We also noticed a high percentage of $36\% \pm 0.01$ was shown for Mass lesions, $14\% \pm 0.01$ for Calcification lesions, and $50\% \pm 0.02$ for Architectural Distortion lesions. In addition, $90\% \pm 0.06$ of Normal mammograms were accordingly classified on Prior exam screenings. The inference time per each configuration was reported with a maximum value of 0.63 seconds per image.

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

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

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

Additionally, Figure 46 illustrates the confusion matrix for the classification of the true detected bounding boxes on the Prior mammograms, where three types of lesions are presented with the Normal cases (i.e., correct prediction without detected lesions). Prediction error for different classes is low with a maximum rate of 3.5%.

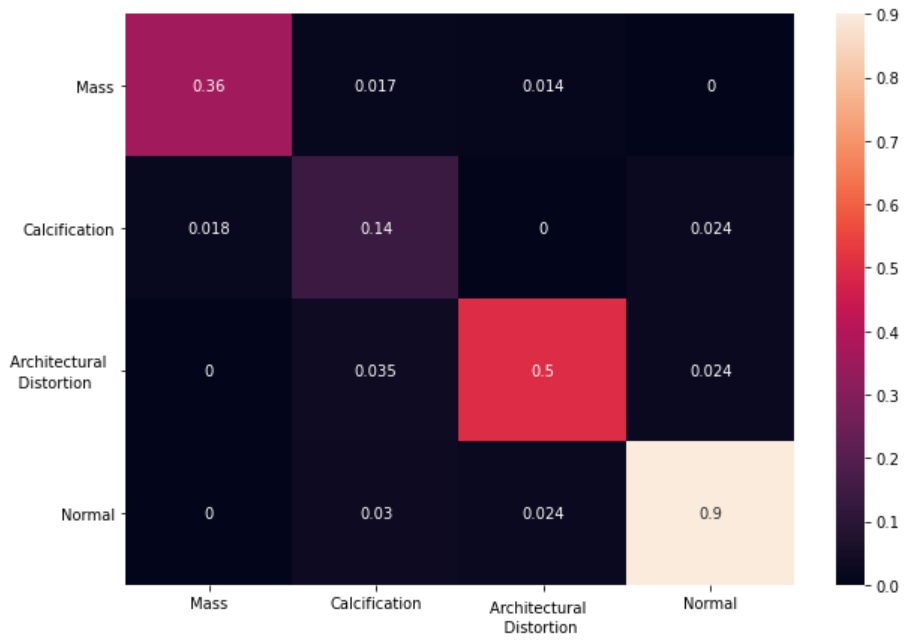


FIGURE 46: Confusion matrix of prediction results for Prior Mammograms

c. Retrospective analysis for the early detection and identification on prior mammograms

In this part, we investigate the follow-up exam time (i.e., originally between 1 to 6 years) of the true early prediction results for each class label. Figure 47 illustrates retrospectively that our suggested methodology can anticipate the presence of breast lesions that were originally diagnosed at a later exam (i.e., $t > 0$). Mass lesions were predicted beforehand but later detected by experts and radiologists within 2 to 3.5 years. The comparative figure also represents the latest follow-up exam time of results from using the image-to-image translation techniques versus the original mammograms.

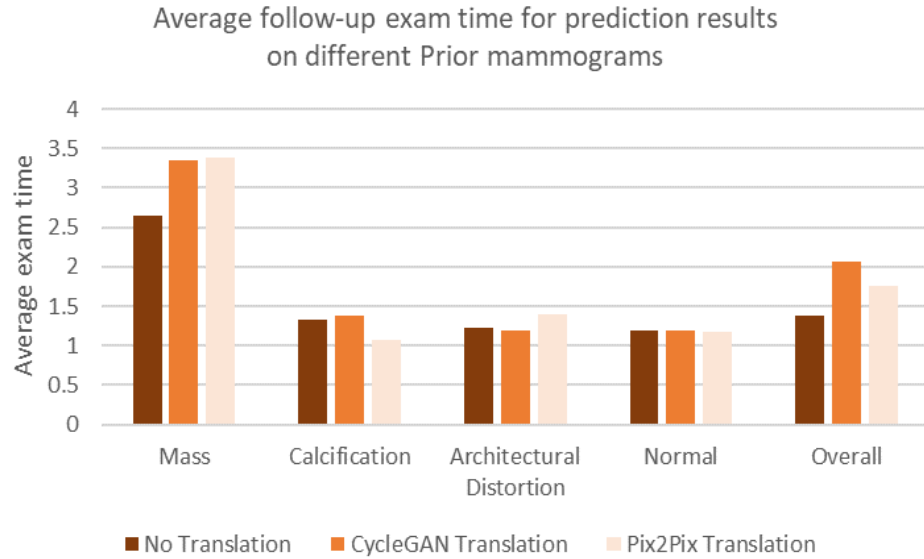


FIGURE 47: Comparison of Mean follow-up exam time for prediction results across classes with and without image-to-image translation Pix2pix and CycleGAN on Prior mammograms

Moreover, Figure 48 illustrates a comparison of inference results of the YOLO-based model with and without image-to-image translation across the different classes and overall. We can visually conclude that the Pix2Pix translation method had the best performance, which could also be explained by the fact the Pix2Pix model was trained between paired images, compared to the CycleGAN model that used unpaired images. Hence, the Pix2Pix model is more efficient than the CycleGAN for the task of image-to-image translation considering the advantage of image alignment it presents between the paired datasets.

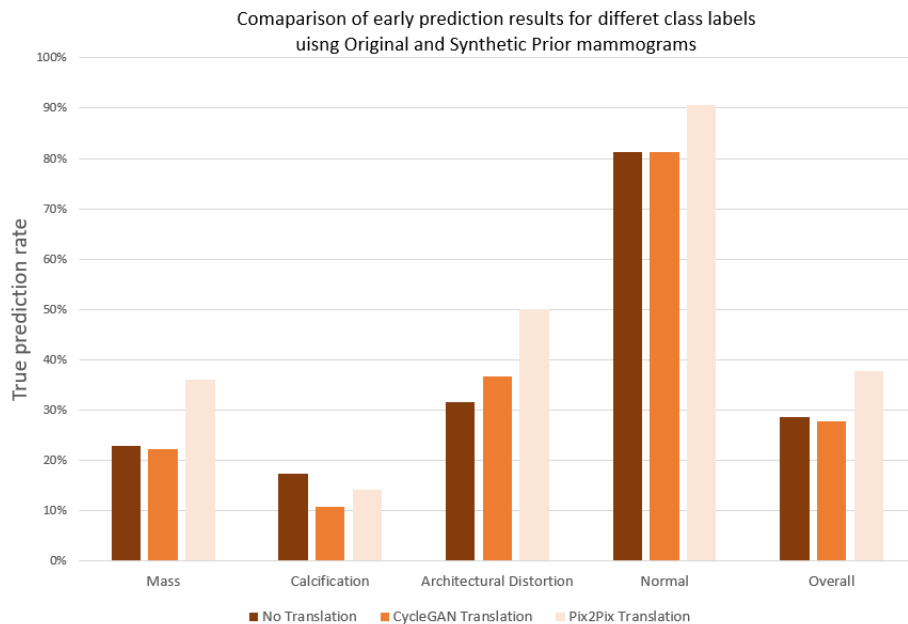


FIGURE 48: Comparison performance of YOLO Fusion model across classes with and without image-to-image translation

Furthermore, we compared results of the early detection and classification across the Prior exam’s time that varied between 1 to 6 years. Figure 49 provides a visual observation of the percentage of correctly predicted Prior mammograms for each class label using the best-reported experiment (i.e., using the Pix2Pix translation). The follow-up exam time of 1 year had the highest rate of predicted images. This emphasizes the success of our methodology to early localizing and identifying lesions that are often considered the hardest to diagnose. Another observation is that our methodology captured the Mass lesions that had follow-up requests of later than 3 years, which might be too late to diagnose patients with Mass breast lesions.

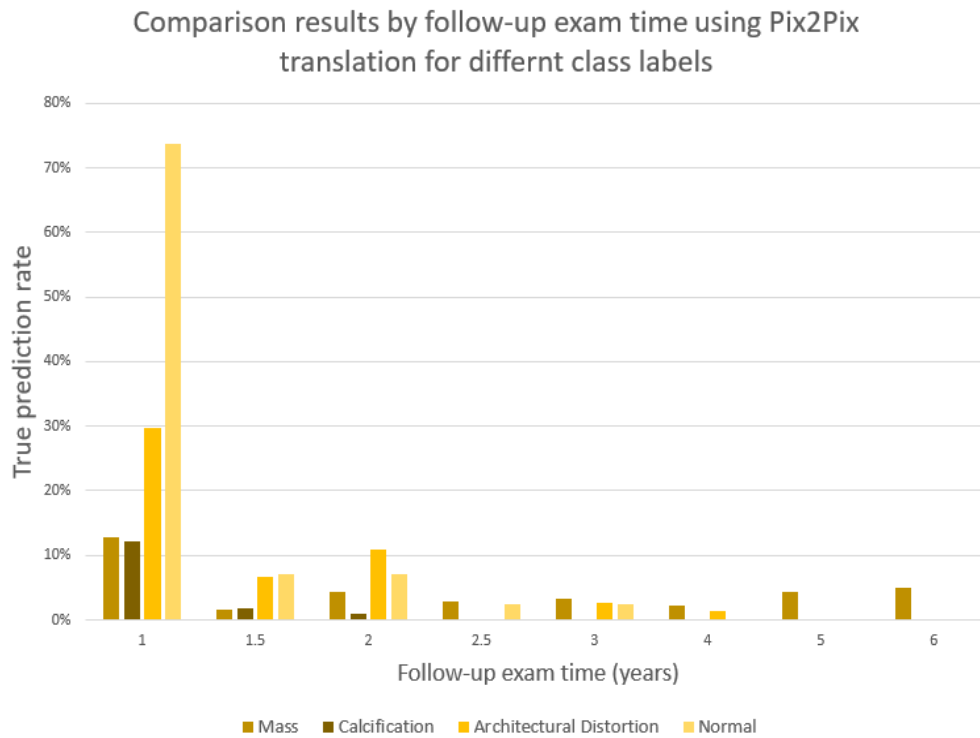


FIGURE 49: Comparison performance of YOLO Fusion model and Pix2Pix for image-to-image translation for different classes across follow-up exam time (years)

Finally, two samples of mammograms that were taken from different patients, including Prior exam views and their corresponding Current exam views are shown below in Figure 50. Two cases of results are demonstrated: 1) When both Current and Prior mammograms were correctly predicted, and 2) When only the Prior mammograms were correctly predicted. It seems that when the model failed to predict lesions in

some Current mammograms, their corresponding Prior mammograms were successfully predicted using the inference model. Predicted bounding boxes were slightly different between views but they exceeded the threshold score, and this could be explained with the different quality of the acquired images and the type of the detected lesions. Moreover, Normal mammograms were shown in the last row accordingly where correct predictions were demonstrated for both screening and for only Predicted mammograms.

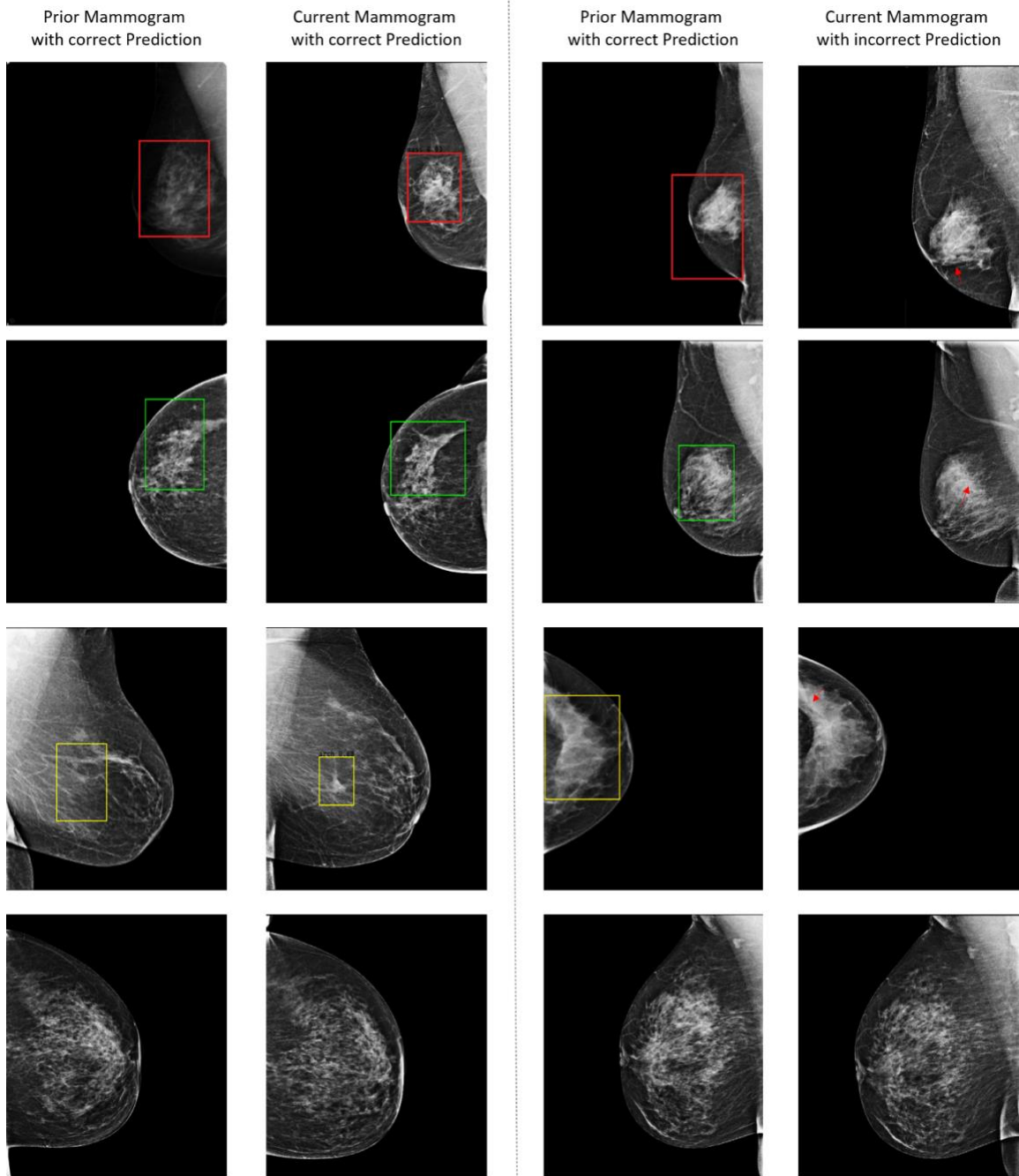


Figure 50: Example results on Prior mammograms vs Current mammograms using the YOLO-based Fusion model that was inferred on the translated Prior images from Pix2Pix method across classes: Mass (red bounding boxes, first row), Calcification (green bounding boxes, second row), and Architectural Distortion (yellow bounding boxes, third row). Red arrows point to the ground truth location. Last row belongs to the Normal class

Finally, a comparison of latest studies and similar methods were reported against our proposed methodology. For a complete and fair comparison, only works that were applied for Mass lesions detection were reported and compared in Table 19. Comparing both detection accuracy rate and inference time with the other works that were evaluated on the two public datasets CBIS-DDSM and INbreast, our YOLO-based fusion models achieved overall better than previous works. Our recent work was considered having the best trade-off between the detection accuracy and testing time comparing to the work by Peng et al. [61] that had a better inference time of 0.134 second per image but it only had a detection accuracy rate of 93.45% on the CBIS-DDSM dataset. Accordingly, the work by Al-Antari et al. [63] had faster inference time of 0.025 seconds per image on Mass detection for the INbreast dataset, but our results exceeded their detection accuracy rate of only 97.27%. Additionally, it is fair to mention that all experiments in the related works were conducted using different configurations and preprocessing techniques, which may show different performance on public datasets. We also compared the work of Zheng et al. [152] that similarly conducted the detection and classification tasks on the UCHCDM dataset. Although the surveyed work achieved a better overall performance of 92.8% than our reported results, the tasks were not simultaneous and required separate inference time of 0.62 seconds per image for the detection method and 0.88 seconds per image on the classification.

The framework of the retrospective analysis of YOLO-based fusion model prediction on prior mammograms was recently published in the Computer Methods and Programs in Biomedicine Journal [190].

Table 19: Comparison of Mass detection with other works

Reference	Year	Method	Dataset	Detection accuracy rate (%)	Inference time per image (sec)
Al-masni et al. [68]	2017	YOLO	DDSM	85.25	NA
Dhungel et al. [167]	2017	Cascade Deep Learning and Random Forest	INbreast	96	39
Al-masni et al. [49]	2018	YOLO	DDSM	99.7	NA
Zheng et al [81]	2018	Detection: 3 cascading detectors (Haar, LBP, and HOG)	UCHCDM	92.8	0.62
		Classification: VGG-19			0.88
Agarwal et al. [32]	2019	CNN patch classifier and mass probability map (MPM)	CBIS-DDSM	82	NA
			INbreast	98	
Aly, G. et al. [52]	2020	YOLO	INbreast	89.5	0.009
Peng et al. [61]	2020	Faster R-CNN	CBIS-DDSM	93.45	0.134
			INbreast	95.54	
Al-Antari et al. [63]	2020	YOLO	INbreast	97.27	0.025
Singh et al. [89]	2020	Single Shot Detector (SSD)	INbreast	97	NA
Proposed Methodology	2021	YOLO-based Fusion Models	CBIS-DDSM	95.7	0.55
			INbreast	98.1	0.58
			INCAN	98	0.52
	2022	YOLO-based Fusion Models (Current mammograms)	UCHCDM	92.09	0.62

Furthermore, we compared the effort of similar works on conducting an early detection of breast lesions against our work contribution. Table 20 shows two recent works that had the closest similarity on

integrating Prior mammograms views to predict the location and type of abnormal lesions. Our work surpassed the work of Watanabe et al. [168] that was able to accurately detect and distinguish Mass lesions with an early detection accuracy rate of 27%. However, our proposed methodology had a lower early detection accuracy rate on Calcification lesions where they had 20% on a custom dataset that was generated using the temporal subtraction technique. Genuinely, all the reviewed works were assessed on private datasets and the reported results could be distinctive compared to our study’s outcome. All comparable works did not measure the testing time, but our proposed method achieved an inference time of only 0.63 seconds per image.

Table 20: Comparison of Early detection with other works

Reference	Year	Method	Dataset	Class Label	Early Detection accuracy rate (%)	Inference time per image (sec)
Watanabe et al. [76]	2019	cmAssist – Custom deep learning networks	Private	Mass	27	NA
Loizidou et al. [79]	2019	Temporal subtraction	Custom	Calcification	20	NA
Proposed Methodology	2022	YOLO-based	UCHCDM + Synthetic dataset	Mass	36 ± 0.01	0.63
		Fusion Models + Pix2Pix translation (Prior mammograms)		Calcification	14 ± 0.01	
		Architectural Distortion		50 ± 0.02		
		Normal		90 ± 0.06		
		Overall		37 ± 0.10		

3. Breast Mass Segmentation

a. Quantitative segmentation results

As shown in Table 21, the results are measured for each testing set where we computed the two-evaluation metrics for the segmented maps per pixel and compared them to the original ground truth.

Table 21: Segmentation performance of our proposed networks on the test sets

Proposed Architectures	Dice score	IoU score	Dice score	IoU score	Dice score	IoU score
	(%)	(%)	(%)	(%)	(%)	(%)
	CBIS-DDSM		INbreast		INCAN	
Standard UNet	78.62	64.87	89.21	79.5	89.87	86.43
Connected-UNets	82.22	69.82	93.36	85.75	95.72	91.95
Standard AUNet	80.39	67.29	91.35	82.59	90.25	88.02
Connected-AUNets	83.84	72.19	93.52	86.01	95.82	92.17
Standard ResUNet	80.94	68.05	92.71	84.58	93.58	89.79
Connected-ResUNets	85.01	73.95	94.13	87.63	95.88	92.27

The results reveal that the proposed Connected-UNets architecture outperforms the regular UNet in terms of Dice and IoU scores across all experimental datasets. We also enhanced the segmentation performance of the standard AUNet and ResUNet using the architecture. As a result, the results reveal a comparison of the conventional architectures, with ResUNet outperforming the AUNet and the later architecture outperforming the UNet. The findings highlight the benefits of the attention mechanism as well as the residual blocks added to the simple UNet. We clearly notice an improvement of Dice score by 3.6% using the Connected-UNets, 3.4% using the Connected-AUNets, and 4% using the Connected-ResUNets on the CBIS-DDSM dataset. With the Connected-UNets, we enhanced the Dice score by 4.15 %, 2.17 % with the Connected-AUNets, and 1.42 % with the Connected-ResUNets for the INbreast dataset. Similarly, we had an improvement of Dice score on the INCAN private dataset by 5.85% using the Connected-UNets, 5.57% using the Connected-AUNets, and 2.3% using the Connected-ResUNets.

Furthermore, the area under curve (AUC) over test sets of all datasets was used to evaluate the segmentation performance of our proposed Connected-UNets and its modifications against the conventional UNet, AUNet, and ResUNet. Each model was used to create segmented images with pixels estimated to be between 0 and 255. The anticipated images were then normalized using scores between 0 and 1. Ground truth images were similarly adjusted to have values of 0 or 1. As a result, the problem was turned into a binary pixel classification task, and the receiver operating characteristic (ROC) between the predicted pixels and their true values was calculated. Figure 51 shows a comparison of ROC curves where we clearly notice that the proposed architectures outperform all standard models with an average AUC of 0.79 for the CBIS-DDSM, 0.94 for the INbreast, and 0.95 for the INCAN private dataset.

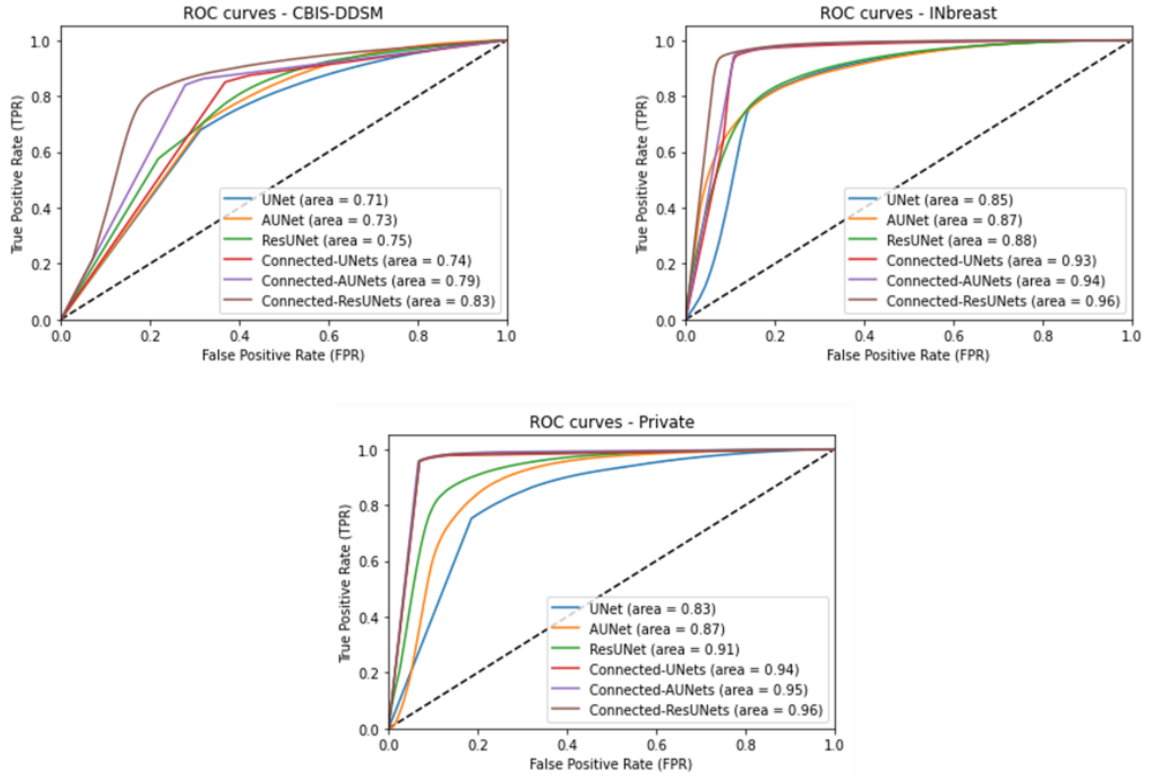


FIGURE 51: Performance of mass segmentation using the different architectures in terms of ROC curves on the test sets of CBIS-DDSM, INbreast, and the INCAN private datasets

ROC curve plots with True positive Rate (TPR) against the False Positive Rate (FPR) and area under curve for pixel-wise evaluation of the standard models (UNet, AUNet, and ResUNet) and for the proposed architecture models (Connected-UNets, Connected-AUNets, and Connected-ResUNets).

According to Table 21, the INCAN private dataset had the best segmentation performance along with the proposed architectures as it represents the best image resolution among the used mammography datasets. Therefore, we applied the CycleGAN model to translate images from CBIS-DDSM and INbreast datasets (i.e., weak domains) into the INCAN private dataset (i.e., strong domain). After training the CycleGAN model between the unpaired datasets and producing new ROI masses for the CBIS-DDSM and the INbreast, synthetic pictures were constructed, as demonstrated in the examples below in Figure 52, where we clearly see the enhanced quality of the new ROI masses that benefit from each dataset's texture.

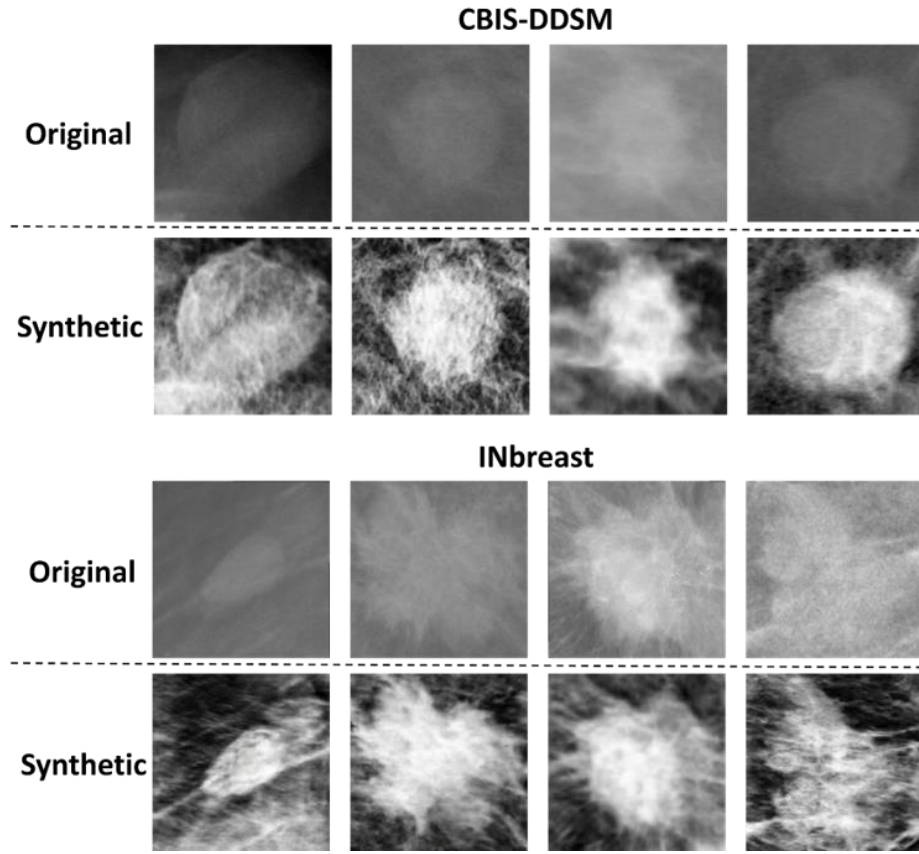


FIGURE 52: Samples of synthetic data from CBIS-DDSM and INbreast datasets generated by CycleGAN model using the INCAN private dataset.

Furthermore, we trained the proposed architectures on the original and synthetic images to predict the segmentation mappings. Table 22 shows the improvement of segmentation's performance of all the standard and architectures using the joined dataset of original and synthetic images. In fact, we notice an increase of Dice score on the CBIS-DDSM by 3.76% using the standard UNet, 3.97% using the standard AUNet, and 4.17% using the ResUNet. Similarly, we have an improved Dice score of 4.8% using the Connected-UNets, 4.11% using the Connected-AUNets, and 4.51% using the Connected-ResUNets.

Table 22: Comparison of the proposed architectures after adding synthetic CBIS-DDSM and INbreast

Proposed Architectures	Dice score (%)	IoU score (%)	Dice score (%)	IoU score (%)	Dice score (%)	IoU score (%)	Dice score (%)	IoU score (%)
	CBIS-DDSM without Synthetic data		CBIS-DDSM with Synthetic data (CycleGAN)		INbreast without Synthetic data		INbreast with Synthetic data (CycleGAN)	
	Standard UNet	78.62	64.87	82.38	72.59	89.21	79.5	93.45
Connected-UNets	82.22	69.82	87.02	77.07	93.36	85.75	95.16	90.77
Standard AUNet	80.39	67.29	84.36	74.02	91.35	82.59	94.73	89.99
Connected-AUNets	83.84	72.19	87.95	78.89	93.52	86.01	94.89	90.28
Standard ResUNet	80.94	68.05	85.11	76.13	92.71	84.58	94.48	89.59
Connected-ResUNets	85.01	73.95	89.52	80.02	94.13	87.63	95.28	91.03

Finally, all the suggested models are used to evaluate the integrated framework. As the end-to-end performance depends on the first detection and localization step which used the YOLO model, the segmentation step was first reported using the segmentation accuracy measure IoU_{90} score that was later multiplied by the detection accuracy rate to form a final segmentation accuracy. Table 23 shows a comparison of final segmentation results of the different models after using the detection accuracy rate of 95.7%, 98.1%, and 98%, respectively, for CBIS-DDSM, INbreast, and the INCAN private dataset. Consequently, we reported a final segmentation performance with a maximum accuracy of 86.91%, 93.03%, and 95.39% using the Connected-ResUNets architecture model, respectively, for CBIS-DDSM, INbreast, and the INCAN private dataset.

Table 23: Final Segmentation performance of our proposed networks on the test sets

Proposed Architectures	IoU ₉₀	Final	IoU ₉₀	Final	IoU ₉₀	Final
	score	Segmentation	score	Segmentation	score	Segmentation
	(%)	accuracy (%)	(%)	accuracy (%)	(%)	accuracy (%)
	CBIS-DDSM		INbreast		INCAN	
Connected-UNets	90.05	86.18	94.06	92.27	96.99	95.05
Connected-AUNets	90.24	86.36	94.63	92.83	97.22	95.27
Connected-ResUNets	90.82	86.91	94.83	93.03	97.34	95.39

Finally, a comparison of the results of the latest state-of-the-art methods and models to segment the breast masses is listed in Table 24. Our proposed architectures outperformed the UNet model and its current variations. On the two public datasets, we achieved the best segmentation performance using the designs when comparing the Dice and IoU scores to the other methods: CBIS-DDSM with a Dice score of 89.52% and an IoU score of 80.02%, and INbreast with a Dice score of 95.28% and an IoU score of 91.03% using the Connected-ResUNets. We outperformed Ravitha Rajalakshmi et al. [53] by 6.62% Dice score on the CBIS-DDSM dataset, and the work of Li et al. [32] by 2.56% Dice score on the INbreast dataset.

The integrated framework of the new Connected-UNets architecture model and the previous YOLO-based fusion model was published in the Nature Partner Journal – Breast Cancer [176].

Table 24: Comparison of the proposed architectures and state-of-the-art methods

Reference	Year	Method	Dataset	Dice score (%)	IoU score (%)
Dhungel et al. [76]	2015	Deep structured output learning + refinement	INbreast	85	--
Dhungel et al. [135]	2017	CNN + CRF	INbreast	90.06	--
Zhu et al. [88]	2018	Multi-scale FCN-CRF	INbreast	90.97	--
Fully resolution					
Al-Antari et al. [62]	2018	Convolutional Network (FrCN)	INbreast	92.63	86.37
Li et al. [96]	2018	Conditional Residual UNet	INbreast	92.72	--
Wang et al. [103]	2019	ResNet34 + ASPP	INbreast	91.1	--
Abdelhafiz et al. [94]	2019	R-UNet	INbreast	90.5	89.1
Sun et al. [100]	2020	Attention UNet	INbreast	79.1	--

			CBIS-DDSM	81.8	--
Ravitha Rajalakshmi et al. [101]	2020	Deeply supervised U-Net (DS U-Net)	INbreast	79	83.2
			CBIS-DDSM	82.9	--
Singh et al. [89]	2020	conditional GAN (cGAN)	INbreast	91.47	83.58
Tsochatzidis et al. [105]	2021	UNet+	CBIS-DDSM	72.2	56.5
		Connected-UNets	INbreast	95.16	90.77
			CBIS-DDSM	87.02	77.07
Proposed architectures	2021	Connected-AUNets	INbreast	94.89	90.28
			CBIS-DDSM	87.95	78.89
		Connected-ResUNets	INbreast	95.28	91.03
			CBIS-DDSM	89.52	80.02

b. Qualitative segmentation results

We performed a post-processing step on all segmented ROI masses by simply eliminating any outlier points that were too far away from the lesions' primary contour. As a result, we took all the possible contours from the binary masks and only chose the one with the greatest area. The output of the conventional UNet models, the Connected-UNets model, and their variants were subjected to this.

Figure 53 shows examples of the segmented ROI masses generated by the experimental models against their ground truth images. We can plainly see that the Connected-UNets segmentation maps and outcomes are of varying quality, and that their variations always contain less error and capture more precision than the ground truth. We can see from the segmentation results that Connected-ResUNets is better than the other designs at predicting the smallest aspects of the tumor's boundaries.

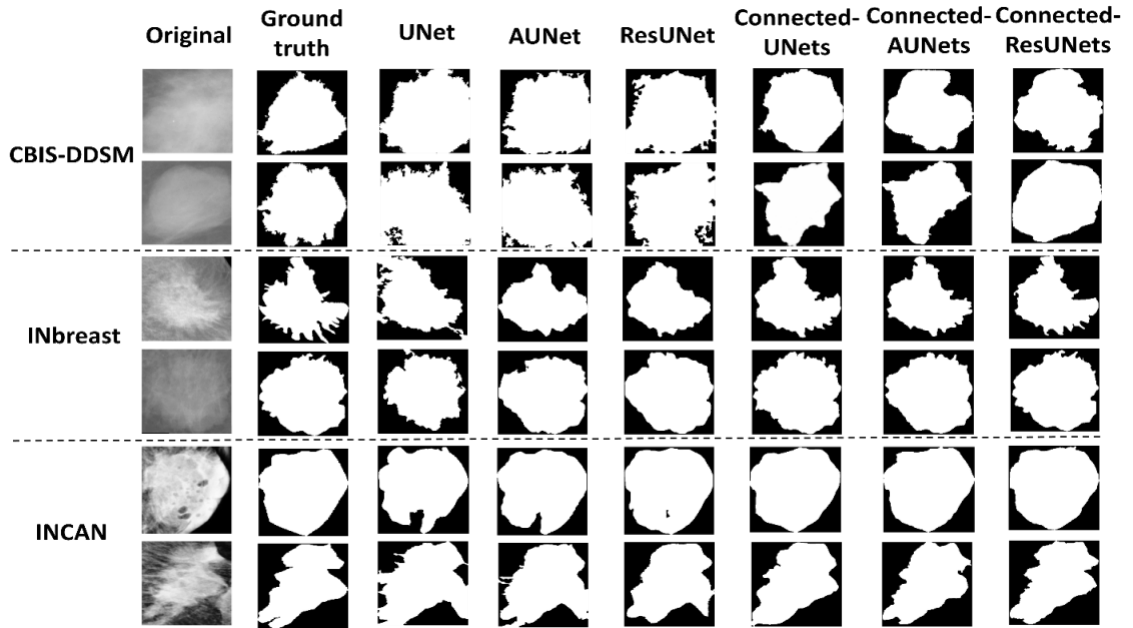


FIGURE 53 Examples of the segmentation results on the test set of the datasets. Subplot on the top shows two samples of mammograms from the CBIS-DDSM dataset.

Overall, the proposed architectures outperform the typical architectures, demonstrating their ability to learn complicated features thanks to the connections made between the two UNets in the proposed Connected-UNets, which uses decoded features as an additional input in the encoder route.

As a result, Figure 54 shows a visual comparison of the Connected-ResUNets model, which uses the recommended ASPP block to connect each encoder and decoder, against the same model without the ASPP block. We can conclude that the ASPP block improved the segmentation results' precision.

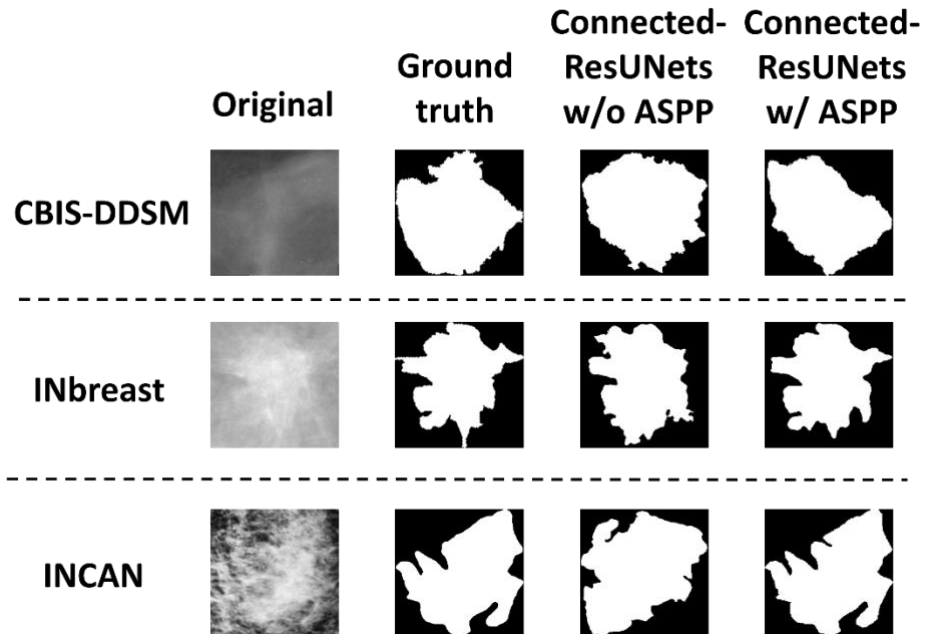


FIGURE 54: Examples of the segmentation results for the proposed architecture Connected-ResUNets comparing with and without using ASPP block.

After that, a qualitative segmentation comparison of the proposed Connected-ResUNets against the basic ResUNet architecture is presented in Figure 55. In addition, for each associated ROI mass, a comparison of Dice and IoU scores is made. We found that the proposed architecture model can capture the tiniest characteristics of tumors of all forms and sizes across all datasets. As a result, it is obvious that the Connected-ResUNets predicted contours are the most like the ground truth contours, as seen by the greatest Dice and IoU score values. After adding the synthetic images created by the CycleGAN model to the training data, we compared the segmentation results of one of the recommended architecture models, Connected-ResUNets.

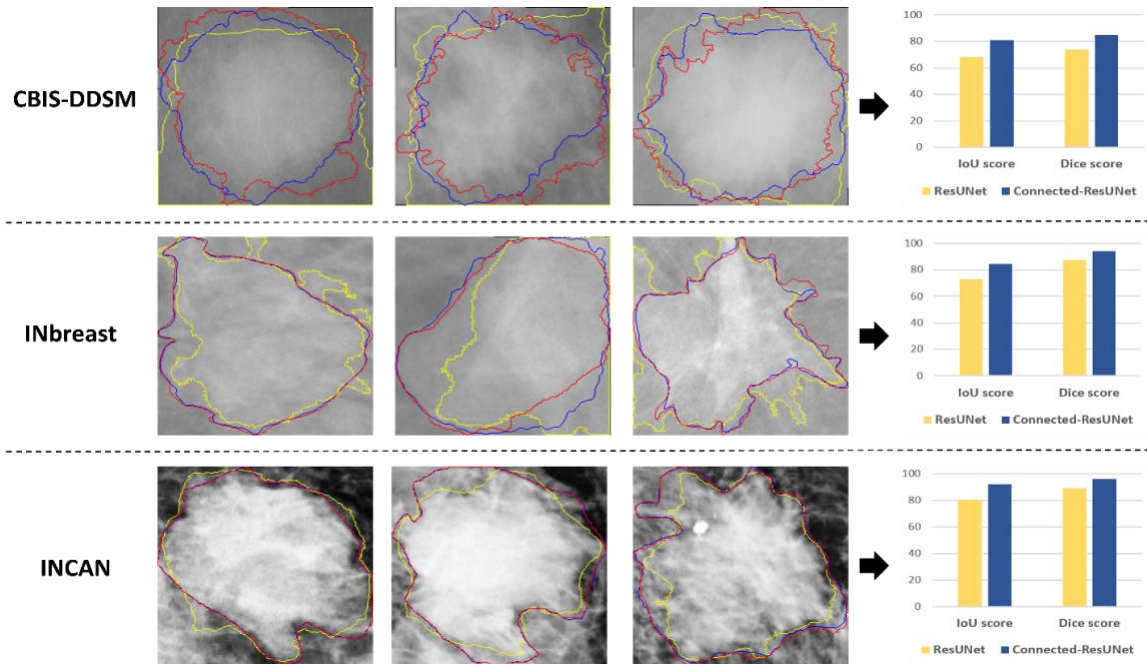


FIGURE 55: Examples of the segmented masses for the proposed architecture Connected-ResUNets comparing with the ResUNet

Figure 56 shows a better segmented contour of the mass tumor using the additional synthetic images. The new training data produces more exact pixel segmentation that closely resembles the ground truth images. As a result, the quality of the segmentation findings demonstrates the benefit of using synthetic images to improve segmentation quality, as well as the capacity of cross-modality synthesis to increase data quantity and improve quality by incorporating additional related domains. Finally, we applied two state-of-the-art methods that we discussed, by Al-Antari et al. [22] and Li et al. [32], to segment ROIs from all the mammography datasets, and visual comparison shows that predictions of two suggested models FrCN and CRUNET are slightly close to the ground truth images but they do not represent the contours exactly.

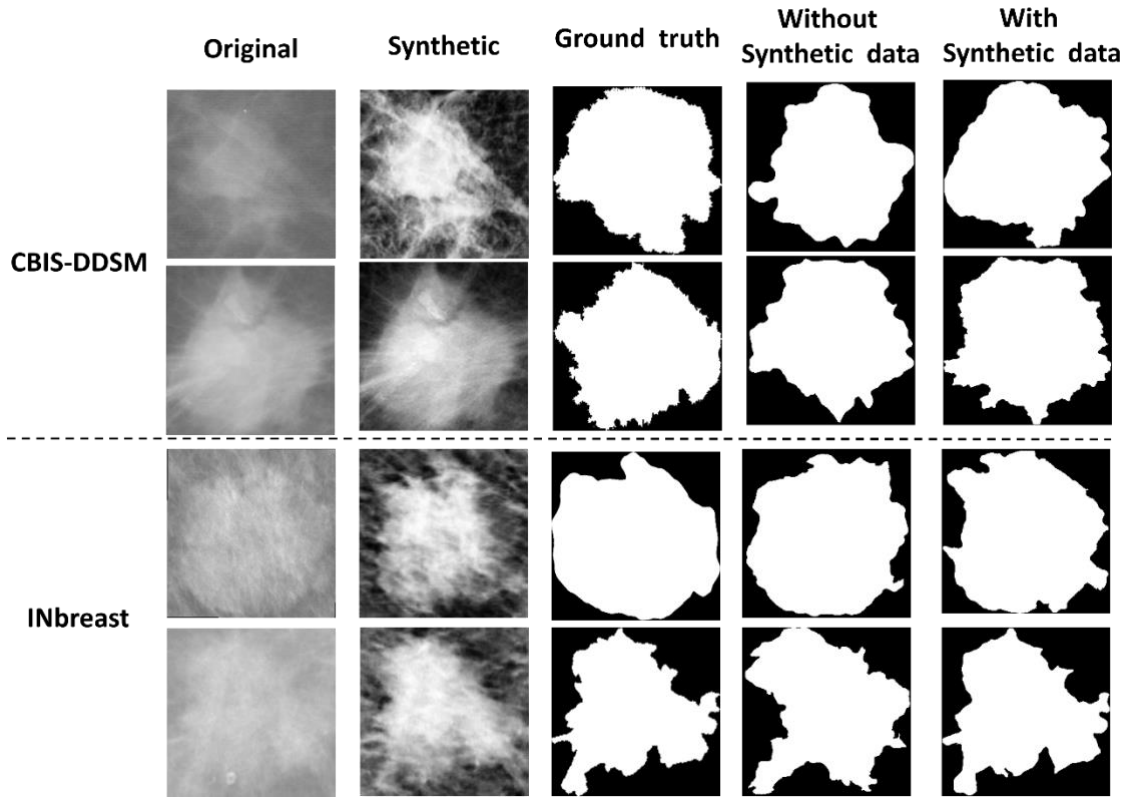


FIGURE 56: Examples of the segmentation results for the proposed architecture Connected-ResUNets with and without adding the synthetic data

Examples shown in Figure 57 are selected to be challenging for segmentation and our proposed architecture models showed better visual results to segment the mass lesions.

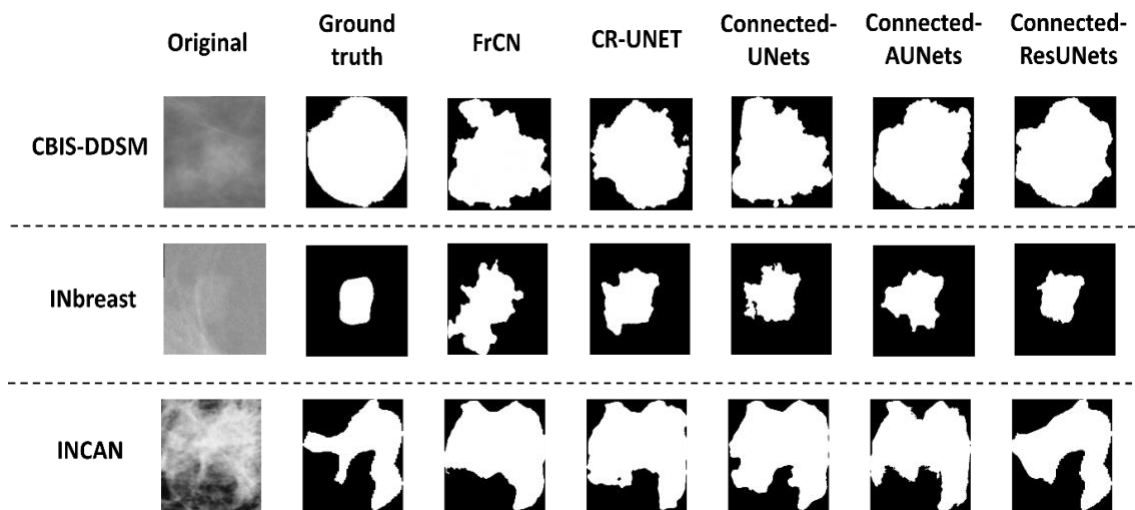


FIGURE 57: Examples of the segmentation results for the proposed architecture models against two state-of-the-art methods FrCN and CRUNET

4. Classification and Diagnosis

a. Quantitative classification results

The proposed breast mass classification model was trained and compared to single base models for each presented task on the different mammography datasets. We also compared the stacked ensemble of models to a conventional average of different models' weights with an XGBoost classifier.

i. Pathology classification

As shown in Table 25a, Table 25b, and Table 25c, the pathology classification results are compared between different models respectively for CBIS-DDSM, INbreast, and INCAN private datasets. It is reasonable to mention that because the INCAN private dataset includes only malignant cases, we trained and tested the model on a combination of all datasets.

Table 25a: Pathology classification results on the CBIS-DDSM dataset

Model	Accuracy	Sensitivity	Specificity	F1-score	AUC
Model1: ResNet50V2	89.97	0.89	0.91	0.9	0.9
Model2: ResNet101V2	93.57	0.92	0.95	0.94	0.93
Model3: ResNet152V2	92.11	0.92	0.92	0.92	0.92
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	91.04	0.85	0.98	0.91	0.91
Stacked Ensemble of models	95.13	0.93	0.97	0.95	0.95

Table 25b: Pathology classification results on the INbreast dataset

Model	Accuracy	Sensitivity	Specificity	F1-score	AUC
Model1: ResNet50V2	98.52	1.0	0.93	0.97	0.96
Model2: ResNet101V2	95.58	1.0	0.80	0.93	0.9
Model3: ResNet152V2	96.6	1.0	0.90	0.91	0.94
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	97.9	1.0	0.96	0.98	0.97
Stacked Ensemble of models	99.2	1.0	0.98	0.99	0.99

Table 25c: Pathology classification results on the INCAN private dataset

Model	Accuracy	Sensitivity	Specificity	F1-score	AUC
Model1: ResNet50V2	92.83	0.94	0.90	0.91	0.92
Model2: ResNet101V2	94.18	0.94	0.95	0.93	0.94
Model3: ResNet152V2	94.60	0.94	0.94	0.94	0.94
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	94.89	0.94	0.97	0.93	0.94
Stacked Ensemble of models	95.88	0.93	0.97	0.95	0.96

The comparative results show that the proposed stacked ensemble of models performs better than the base ResNet models having different numbers of deep layers (i.e., ResNet50V2, ResNet101V2 and ResNet152V2). Accordingly, our proposed methodology outperformed the average ensemble of models with an XGBoost classifier that performed slightly better than individual models. We clearly notice a high accuracy of 95.13% on the CBIS-DDSM dataset, 99.2% on the INbreast dataset, and 95.88% on the INCAN private dataset. Besides, our proposed model achieved a high sensitivity rate of 0.93 on the CBIS-DDSM dataset, 1.0 on the INbreast dataset, and 0.93 on the INCAN private dataset. Consequently, the results emphasize generally the advantage of ensemble learning technique on improving the classification performance, and particularly the improvement achieved by the stacking method using deep learning models.

Moreover, the pathology classification performance was compared against the different models using the AUC over the test sets of all datasets. Figure 58 shows plots of the Receiver Operating Characteristic (ROC) curves of the True positive Rate (TPR) against the False Positive Rate (FPR), and we clearly notice that the proposed model outperformed all experimental techniques with an AUC of 0.95 for the CBIS-DDSM dataset, 0.99 for the INbreast dataset, and 0.96 for the INCAN private dataset.

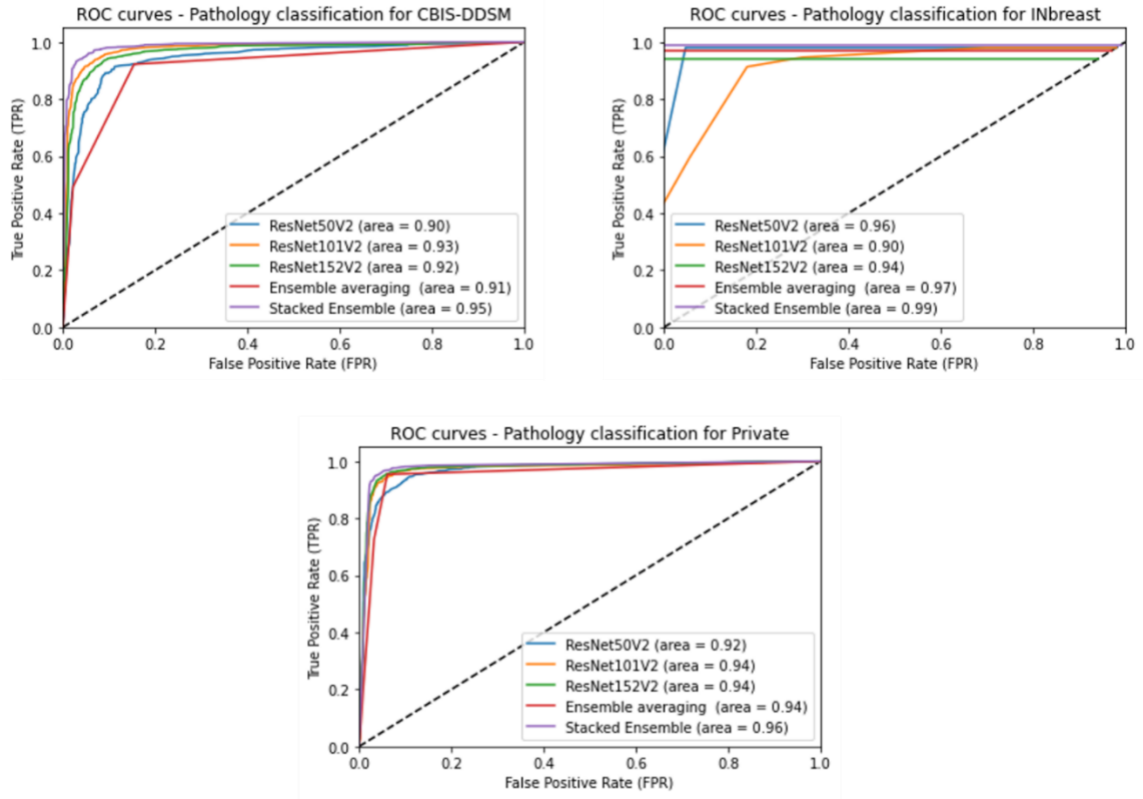


FIGURE 58: Performance of pathology classification using different models in terms of ROC curves and AUC score

ii. BI-RADS category classification

Results shown in Table 26a, Table 26b and Table 26c for BI-RADS category classification illustrate the comparison between different models for all mammography datasets. As mentioned in the section on Datasets description, each dataset has different class labels that vary from category 2 to category 6.

Table 26a: BI-RADS category classification results on the CBIS-DDSM dataset

Model	Accuracy	AUC
Model1: ResNet50V2	80.07	0.93
Model2: ResNet101V2	77.12	0.92
Model3: ResNet152V2	80.08	0.93
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	79.48	0.85
Stacked Ensemble of models	83.84	0.94

Table 26b: BI-RADS category classification results on the INbreast dataset

Model	Accuracy	AUC
Model1: ResNet50V2	98.0	0.94
Model2: ResNet101V2	98.0	0.97
Model3: ResNet152V2	96.1	0.92
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	97.1	0.99
Stacked Ensemble of models	99.0	1.00

Table 26c: BI-RADS category classification results on the INCAN private dataset

Model	Accuracy	AUC
Model1: ResNet50V2	91.91	0.91
Model2: ResNet101V2	92.43	0.92
Model3: ResNet152V2	94.25	0.94
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	92.95	0.93
Stacked Ensemble of models	96.08	0.95

The classification results presented above demonstrate a clear improvement of the performance using our proposed stacked ensemble of models comparing to the basic models with an accuracy of at least 3.78% on the CBIS-DDSM dataset, 1% on the INbreast dataset, and 1.83% on the INCAN private dataset. Moreover, our methodology achieved a better AUC score than the average ensemble model with an XGBoost classifier where we notice a high AUC of 0.94 for the CBIS-DDSM dataset, 1.00 on the INbreast dataset, and 0.95% on the INCAN private dataset. This can be confirmed with a visual comparison of ROC curve plots between employed models as illustrated in Figure 59.

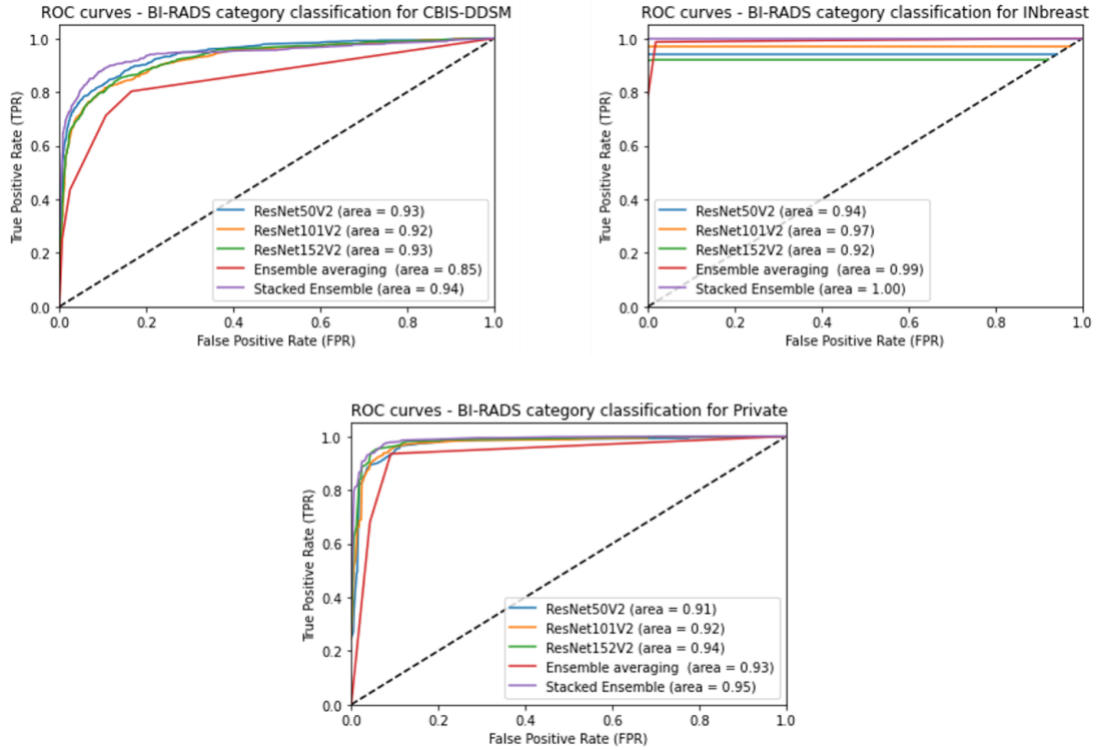


FIGURE 59: Performance of BI-RADS category classification using different models in terms of ROC curves and AUC score

iii. Shape classification

Lastly, the proposed model was trained on the CBIS-DDSM dataset for classifying the shape of breast masses, as it is the only dataset that possesses shape annotation by experts. Equivalently, all trained models were tested, and comparison is shown in Table 27. Undoubtedly, our suggested stacked ensemble of models had the highest accuracy score of 90.02% among the employed models, which improved the performance of separate models notably with 1.7% and remarkably with 10.66% compared to the average ensemble of models with an XGBoost classifier.

Table 27: Shape classification results on the CBIS-DDSM dataset

Model	Accuracy	AUC
Model1: ResNet50V2	75.51	0.90
Model2: ResNet101V2	89.90	0.95
Model3: ResNet152V2	88.32	0.97
Average Weights of Model1, Model2 and Model3 + XGBoost Classifier	79.36	0.84
Stacked Ensemble of models	90.02	0.98

Furthermore, Figure 60 presents a comparison of ROC curve plots for the different employed models and represents the AUC score accordingly. We clearly notice that our proposed model had the highest AUC of 0.98 among the presented models, which was close to the ResNet152V2 performance but with a slightly better accuracy rate.

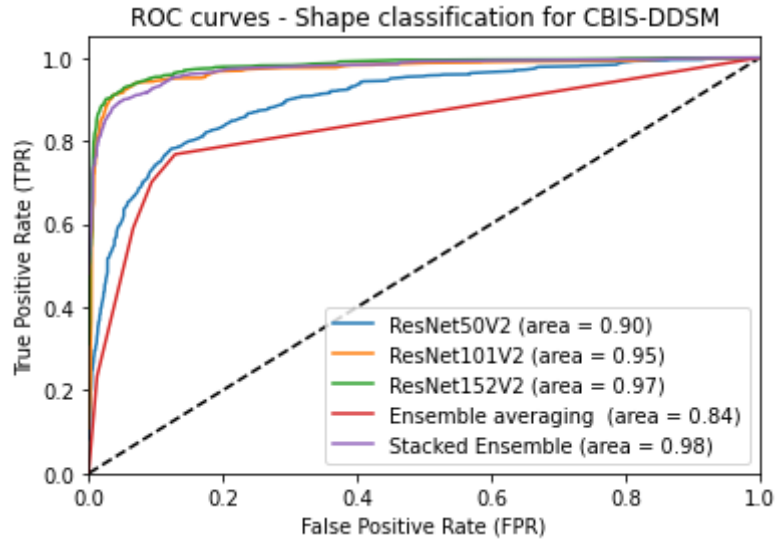


FIGURE 60: Performance of Shape classification using different models in terms of ROC curves and AUC score

b. Qualitative classification results

Previous comparison results highlighted that our proposed methodology yielded the best results for the different classification tasks. Consequently, we analyzed the classification prediction between different datasets using the confusion matrix that summarizes the results across the class labels.

Figure 61, Figure 62, and Figure 63 respectively present the normalized confusion matrix plots for each classification problem. As indicated below, the INbreast dataset had the best pathology classification tradeoff between malignant and benign classes, and this can be explained with the high-quality resolution of the mammograms collected in FFDM format that helps distinguish between the two class labels. The INCAN private dataset had also a remarkable confusion matrix with close recall and precision scores, which were like the CBIS-DDSM dataset's performance.

Moreover, the INbreast dataset had the best BI-RADS categorization tradeoff with a notable prediction per class from 0.92 to 1.0. Concerning the INCAN private dataset, it has only two BI-RADS categories 4 and 5, and we notice a similar satisfying confusion matrix with prediction scores of 0.93 and 0.96. The CBIS-DDSM dataset has a slightly worse tradeoff for the BI-RADS category classification, and

this is due to the low resolution of deteriorated ROI images from the digitized X-rays mammograms. In fact, the confusion matrix shows values from 0.80 to 0.89 and we clearly notice that the four categories have a close prediction score due to the similarity of the pixel distribution caused by the quality presented in the public dataset.

Finally, the confusion matrix for the shape classification showed an overall sufficient tradeoff between the class labels. We clearly observe similar predicted results for the irregular and lobulated cases with a maximum value of 0.93, and this can be interpreted by the close appearance of the two lesions' shape. However, oval, and round cases had worse results, and in particular, the round class label had a performance score of 0.73.

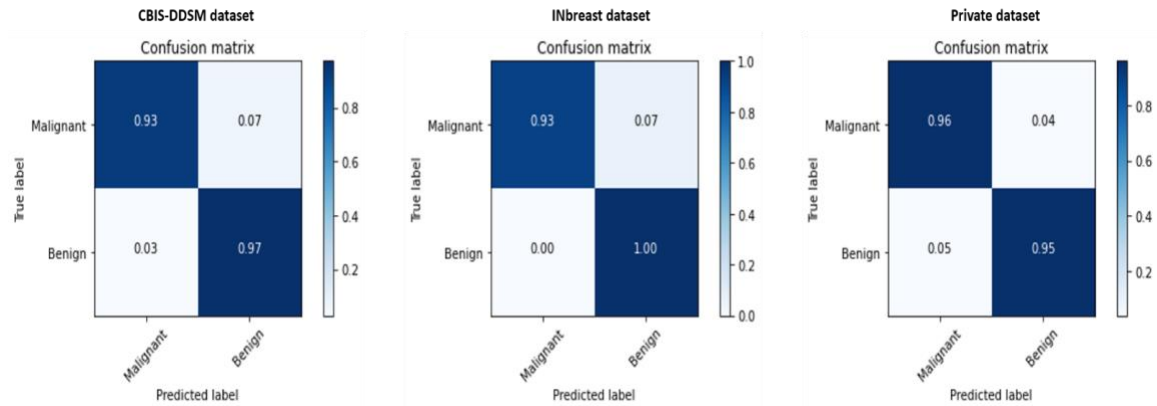


FIGURE 61: Confusion matrix of the stacked ensemble of models for the pathology classification on the mammography datasets

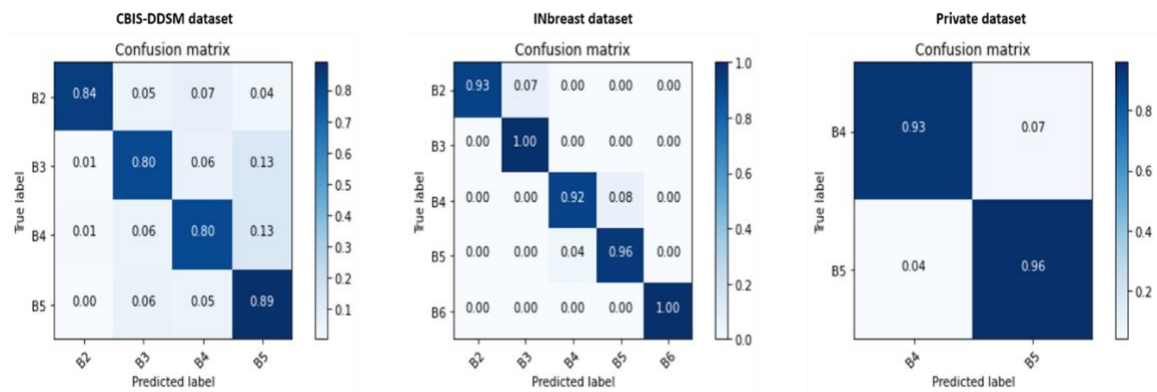


FIGURE 62: Confusion matrix of the stacked ensemble of models for the BI-RADS category classification on the mammography datasets

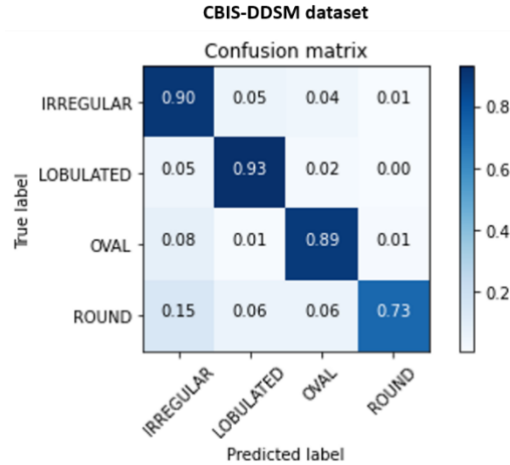


FIGURE 63: Confusion matrix of the stacked ensemble of models for the Shape classification on the datasets

Additionally, we reported the classification results of the final integrated CAD system using the segmentation step and compared it to results of the proposed method without using the segmented ROIs. Table 28 shows a better classification performance for each task on all mammography datasets, where the pathology classification demonstrated improvement of 4.26% on the CBIS-DDSM dataset, 4% on the INbreast dataset, and 5.5% on the INCAN private dataset. Accordingly, the BI-RADS category classification presented enhanced performance using the detected and segmented images with accuracy difference of 3.69% on the CBIS-DDSM dataset, 1.5% on the INbreast dataset, and 0.38% on the INCAN private dataset. Finally, the CBIS-DDSM dataset had improved the shape classification results with 4.36% accuracy. Consequently, it is clearly observed that the CAD system with the integrated detection and segmentation stages achieved much better results for the classification and diagnosis of breast masses.

Table 28: Comparison of classification performance (accuracy %) using the proposed CAD system with and without mass segmentation

Dataset	Pathology		BI-RADS Category		Shape	
	CAD system without mass segmentation	CAD system with mass segmentation	CAD system without mass segmentation	CAD system with mass segmentation	CAD system without mass segmentation	CAD system with mass segmentation
CBIS-DDSM	90.87	95.13	81.69	85.38	85.66	90.02
INbreast	95.20	99.20	97.50	99	NA	NA
INCAN	90.38	95.88	95.70	96.08	NA	NA

Finally, a comparison of results of the latest state-of-the-art methods and similar models to classify the breast masses is listed in Table 29. Our proposed CAD system that integrates the previous detection and segmentation steps and the current proposed classification framework outperformed the previous deep learning models applied for pathology, BI-RADS category, and shape classification. Compared with other techniques that used segmented ROIs, we exceeded the performance of the work by Falconí et al. [127] that only achieved an accuracy of 78.4% using the MobileNet model on the CBIS-DDSM dataset. On the other hand, we also outperformed the work of Alkhaleefah et al. [128] even though they did not use segmented input images and reported an accuracy of 93.47%. Moreover, recent works on the INbreast dataset were all surpassed where the highest accuracy of 98.26% was reported by Chakravarthy et al. [124] using the ICS-ELM algorithm on original ROI masses. We also reported better accuracy score for the pathology classification applied on segmented ROIs from the INbreast dataset, where Al-Antari et al. [62] only achieved an accuracy of 95.64% and an AUC score of 0.94. The results of the BI-RADS categorization also outperformed the previous works on CBIS-DDSM dataset with the work suggested by Medeiros et al. [130] that applied DenseNet201 on original ROI masses and only achieved an accuracy of 63.4%. No previous work applied the BI-RADS category classification on segmented images and therefore we could not compare with our proposed work. Accordingly, our method surpassed the performance on the INbreast dataset by the work of Falconí et al. [129] that only reported an accuracy of 90.9% using NasNet and VGG models. Lastly, our methodology gained the best shape classification performance compared to a recent work of Singh et al. [89] that applied a CNN model on a similar dataset, and it is reasonable to say that this reviewed work is the only comparable work that applied shape classification on detected and segmented ROIs but only achieved an accuracy of 80%.

The entire integrated framework of the Stacked Ensemble of Residual neural networks model and the previous Connected-UNets architecture model and the YOLO-based fusion model was submitted to the Scientific Reports Journal, and it is currently under revision [177].

Table 29: Comparison of the proposed methodology and state-of-the-art methods

Reference	Year	Method	Classification	Dataset	Segmented Images	Accuracy	AUC
Dhungel et al. [135]	2017	multi-view deep residual neural network (mResNet)	Pathology	INbreast	Yes	--	0.80
Al-Antari et al. [62]	2018	Ensemble of AlexNet-based CNN	Pathology	INbreast	Yes	95.64	0.94
Falconí et al. [127]	2019	MobileNet and NasNet + fine tuning	Pathology	CBIS-DDSM	Yes	78.4	--
Falconí et al. [178]	2020	VGG16 + fine tuning	Pathology	CBIS-DDSM	No	84.4	0.84
Chakravarthy et al. [124]	2020	Improved Crow-Search Optimized Extreme Learning Machine (ICS-ELM) algorithm	Pathology	INbreast	No	98.26	--
Alkhaleefah et al. [128]	2020	AlexNet, VGG, GoogLeNet, ResNet + fine tuning	Pathology	CBIS-DDSM	No	93.47	0.97
Medeiros et al [130]	2020	DenseNet201 + MLP	BI-RADS	CBIS-DDSM	No	63.4	--
Falconí et al. [129]	2020	NasNet + fine tuning on VGG16 and VGG19	BI-RADS	INbreast	No	90.9	0.99
Singh et al. [89]	2020	CNN	Shape	DDSM	Yes	80	0.80
Li et al. [133]	2021	DualCoreNet: Texture and	Pathology	INbreast	Yes	--	0.93

		shape features fusion					
Proposed Methodology	2022	Stacked Ensemble of ResNet models	Pathology	CBIS-DDSM	Yes	95.13	0.95
				INbreast		99.20	0.99
				INCAN		95.88	0.95
		BI-RADS	CBIS-DDSM	Yes	85.38	0.94	
			INbreast		99	1.00	
			INCAN		96.08	0.95	
Shape	CBIS-DDSM	Yes	90.02	0.98			

E. Testing & Deployment

The entire connected framework we suggested for the CAD system of breast cancer diagnosis was deployed on a server to showcase the performance and results of the experimental tool. To deploy the developed tool that is based on several deep learning architecture models, all the trained models' weights and biases were saved for inference. Because of the large number of parameters and weights matrix in each model, we tried first to lightweight the saved models before being used in the final implementation. To do that, we applied a post-training quantization technique to the models we saved in Keras framework [179, 180]. Therefore, we optimized the deployment of our CAD tool by reducing the size of the entire model while maintaining the performance and accuracy.

During the training and validation stages, we relied on a GPU and large storage to learn the large models where each model was worth at least 3 GB. However, during the deployment, the server may accept the size of the inferred models, but it would not respond fast. Consequently, the tool will perform expensive computations and generate late results.

We applied the post-training optimization technique that reduces the CPU and hardware accelerator latency and processing and optimized the model size with quite a drop in performance [181]. The technique applies to convert the already-trained Tensorflow models into Tensorflow Lite models. As a result, the models were saved into TFLite format files that will be placed within the backend package for the deployment. In fact, the files' sizes have been reduced to 75% using the quantized models.

We initially tested the integrated CAD system offline to evaluate the *TFLite* conversion by building a script that calls and executes the three main steps in the order of the suggested logic. As we discussed in the previous chapters, the output of one-step should serve as an input of the next step, and the integrated framework should be starting from an entire mammogram to yield the final diagnosis prediction results that include the pathology, the BI-RADS assessment score, and the shape.

Before applying the model quantization to the deployment, the offline testing took around 6 minutes per image to execute the whole system, which is unfortunately slow. However, after using the post-training dynamic range optimization, the testing time changed to less than 1 minute per image thanks to the lightweight models.

After preparing the backend package, we implemented the design of web service that was developed using JavaScript and React library. As shown in in Figure 64, the web server should process users' requests by just allowing them uploading mammography images, in either DICOM format or image format (i.e. .png, .jpg, etc.). After that, the system accepts the images and assigns a corresponding transaction code to the user to start the processing.

Next, the backend script should sequentially execute an image, first return the detected ROI, and then identify if it is Mass lesion where the detected ROI box is also highlighted on the entire mammogram. Second, the system should generate the segmented image of the detected ROI, and finally return the three classification and diagnosis. As soon as the system is ready to return the results, the code assigned to the user should be used to check the results and send an e-mail to notify the user with the results.

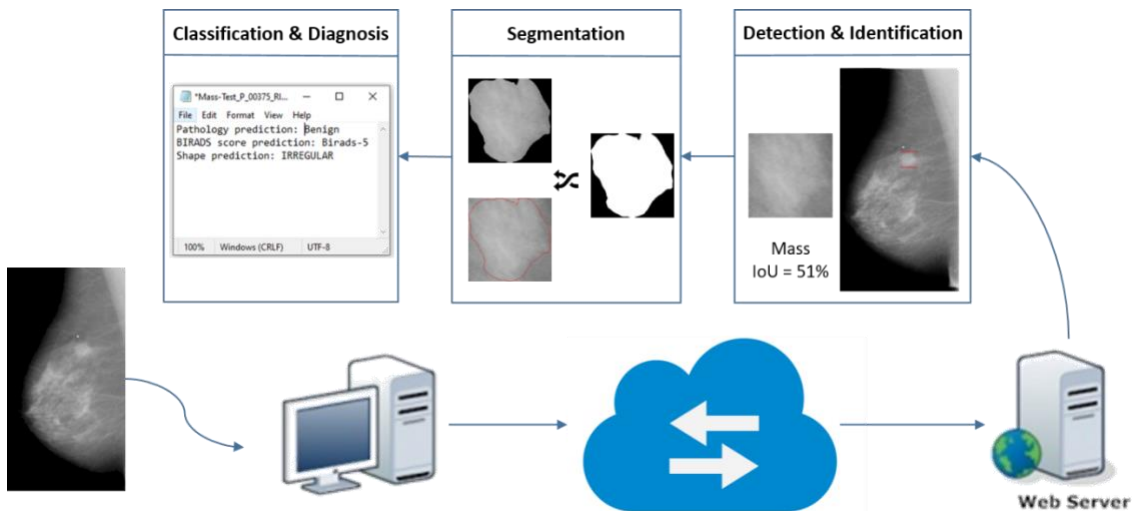


FIGURE 64: Framework of the deployed CAD system

F. Discussion

In this dissertation, we have designed implemented an integrated CAD system using deep learning models for detection and identifying abnormal lesions in mammograms, segment and mask the background of masses, and classifying their pathology and final diagnosis.

In the first step, we have implemented a deep learning YOLO model to simultaneously detect and classify the suspicious lesions in the breast. Similar studies focused solely on the identification of mass lesions and the extraction of areas of interest for further diagnosis. In contrast, our study expands the ability of YOLO-based model to conduct simultaneous detection and classification on mammograms [49].

Furthermore, the presented approach revealed the advantage of YOLO model as detector and classifier toward different clinical mammographic images (i.e., digitized X-rays, full-field digital mammography, etc.). The resilience of YOLO to correctly identify breast lesions above pectoral muscle, near to breast nipples, or over thick tissues is also demonstrated by the quality of anticipated pictures, as illustrated in Figure 41. Experimental results showed that training YOLO-based deep learning model is overall fast and accurate, where our results outperform the SSD method [89], the Faster R-CNN model [63], the CNN model [32] and other machine learning techniques [31, 167] that had a maximum detection accuracy rate of 98% on INbreast dataset but a significantly high inference time. The comparison revealed that YOLO model is the right choice for mass detection in mammography as presented in other existing YOLO implementations [52, 632, 63] with a maximum detection accuracy rate of 97.27% on INbreast dataset, and our study enhanced the state-of-the-art results to be 98.1%. However, the suggested YOLO model has constraints in the training configuration, which is dependent on creating the correct format of input data. Thus, input images should be accompanied by the true locations and class labels of the lesions during the training. As a result, the YOLO model has an input dependency since the locations of lesions must be extracted from the ground truth.

In the second step, we introduced a novel architecture, called Connected-UNets, which fully connects two single UNets using additional skip connection paths. The ASPP mechanism is also used by the network as a transition block to address the problem of resolution loss, which is particularly problematic in the case of small tumors. The proposed mass segmentation architecture expands the ability of skip connections to reconstruct the details lost in the encoding pathway by revoking the first decoded features and connect them with the additional encoded inputs. The design was implemented on two UNet variants: the

Attention UNet (AUNet) and the Residual UNet (ResUNet). The results of the proposed architectures showed the segmentation improvement compared to the basic architectures as shown in Table 21 with a maximum Dice score of 89.52% on the CBISDDSM dataset and 95.28% on the INbreast dataset. Furthermore, the quantitative examination revealed that ResUNet and AUNet had an edge in segmenting bulk tumors. Hence, the improved architectures' Connected-AUNets and Connected-ResUNets outperformed the Connected-UNets in all the used mammography datasets. Comparison of the segmentation map results of each model approve the enhancement made to the standard models to provide a precise segmentation of the mass boundaries as shown in Figure 52. Limitations of the proposed architectures can occur on the long training time of an average of 0.638 s per epoch, which is due to the high computation of the neural networks that have more trainable parameters than the standard architecture models.

Finally, in the last step, we have implemented a stacked ensemble of ResNet models to classify breast masses as malignant or benign and diagnose their BI-RADS category assessment with a score from 2 to 6 and their shape as oval, round, lobulated or irregular. The results of the proposed methodology showed the classification performance's improvement compared to the individual architectures and to the other methods applied on the existing benchmark datasets. Tab.10 shows that we achieved the highest pathology classification performance on the two public datasets: CBIS-DDSM with an accuracy of 95.13% and an AUC score of 0.95, and INbreast with an accuracy of 99.20% and an AUC score of 0.99. Furthermore, we surpassed the results of other models for the BI-RADS categorization on the CBIS-DDSM dataset with an accuracy of 85.38% and an AUC score of 0.94, and on the INbreast dataset with an accuracy of 99% and an AUC score of 1.0. We also reported the highest results on the shape classification for the CBIS-DDSM dataset with an accuracy of 90.02% and an AUC score of 0.98.

Comparing with the similar frameworks that applied the presented classification tasks on segmented ROI masses, our model outperformed the MobileNet and NasNet models [129] for the pathology classification on the CBIS-DDSM dataset and the Ensemble of AlexNet-based CNN model [150] on the INbreast dataset. Moreover, the shape classification achieved better results on a similar dataset DDSM that was evaluated with an individual CNN model [149]. As a result, the stacking model technique provided an efficient way to learn from various depths of neural networks and combine them in another neural network classifier model to benefit from the different weights that were trained individually. Limitations of the

proposed methodology can occur on the long training time of 0.74s per epoch, which is due to the high number of trainable parameters and computations of the ResNetV2 model.

CHAPTER V

CONCLUSION & FUTURE WORKS

In this dissertation, we proposed an integrated framework for Computer-aided Diagnosis (CAD) system that is based on deep learning techniques. The system will serve as a second-reading tool to help doctors investigating the risk of breast cancer and provide an advanced analysis for the suspicious lesions. The presented work proposed three main steps to process a mammogram to first detect the location of abnormalities and distinguish their types to either Mass, Calcification, or Architectural distortion, etc. Next, the system focused on Mass lesions in mammograms and generated a precise contour of the Mass tumors to help determining their malignancy and grading score. Consequently, the ROI of masses were segmented to mask the tissue background and highlight the tumor. Finally, the integrated system predicted the pathology of the detected tumors and classified their BI-RADS score and shape.

The first step was conducted with a YOLO-based fusion model and presented a method that overcomes the problem of predicting location and type of two common abnormal findings in the entire mammograms: Mass and Calcification. Results showed the capability of our proposed methodology to accurately achieve state-of-the-art performance. In addition, the dissertation provided feasible and promising results using the proposed fusion models approach that was considered to join different models and lower the miss-prediction error.

Moreover, as the breast lesions detection plays a critical role in the CAD systems and fully integrated breast cancer diagnosis our methodology provided an improved detection performance compared with the recent deep learning models. This helps to prevent making additional mistakes when diagnosing the lesions that have been found. Furthermore, the second step provided a novel architecture to segment the breast masses in mammograms. The proposed architecture incorporated the recent modifications that were suggested to overcome the challenges of pixel-to-pixel segmentation in medical images, such as attention

mechanism, residual blocks, ASPP concept, etc. The improved segmentation performance was made after benefiting from the information decoded using one UNet and propagated again to a second UNet. In addition, synthetic data was developed to supplement the training data using the CycleGAN model. This uses a quality translation across domains to account for the various quality levels of current mammography datasets (such as X-ray filter and full-field digital mammography (FFDM)).

Finally, the third step integrated our recent works of the YOLO-based fusion models [175] and the Connected-UNets model [176] that generated the detected and segmented ROIs of breast masses. The final diagnosis step was suggested using a stacked ensemble of Neural network models (ResNets) to predict the pathology (malignant or benign), the BI-RADS grading score (2 to 6), and the pathology (round, oval, lobulated, irregular) to provide an entire framework for breast cancer diagnosis that may also include clinical reports analysis. Indeed, an increase of performance using the segmented ROIs has indicated the advantage of masking the background tissues from the tumors' boundaries to help improve the overall classification and diagnosis and decrease the false positive and negative rates.

In summary, this dissertation presents the integrated framework for a breast cancer CAD system via deep learning models. The three stages of detection, segmentation and classification aim to provide a complete clinical tool that can assist radiologists with a second suggestion for an automated mass tumor diagnosis.

Future works can include combining different mammography datasets and improving the long training of deep learning models for the three tasks. Potential ideas may include integrating more abnormalities to the entire system such as Calcification and Architectural distortion and upgrade the segmentation and classification with more annotated datasets.

More orientations to the future could include the radiology reports and medical text data and merge them within the images data for more precision and affordable explainability to the AI and machine learning for medical applications.

This work aimed to focus on mammography images for breast cancer analysis and diagnosis, but we believe it can be transferred to other medical imaging application such Lung cancer, Pulmonary diseases, Brain cancer, etc. Moreover, the proposed architecture models were designed for 2D medical images such mammograms, but we may be reconstructed for 3D medical images such as CT and MRI.

REFERENCES

- [1] American Cancer Society. Global Cancer Facts & Figures. <https://www.cancer.org/research/cancer-facts-statistics/global.html>. Accessed March 27, 2022.
- [2] Cancer Treatment Center of America. What is cancer? <https://www.cancercenter.com/what-is-cancer>. Accessed March 27, 2022.
- [3] Ferlay, J. et al. Cancer statistics for the year 2020: An overview. *International Journal of Cancer* 149, 778-789 (2021).
- [4] American Cancer Society. Cancer Statistics Center. <http://cancerstatisticscenter.cancer.org>. Accessed January 17, 2022.
- [5] Duffy, S. W. et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. *Cancer* 126, 2971-2979 (2020).
- [6] Kaushal, C., Bhat, S., Koundal, D., & Singla, A. (2019). Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images. *Irbm*, 40(4), 211-227.
- [7] Cancer Support Community. Metastatic Breast Cancer <https://www.cancersupportcommunity.org/metastatic-breast-cancer>. Accessed March 27, 2022.
- [8] Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: advances, challenges, and applications. *BMC bioinformatics*, 20(11), 1-20.
- [9] Al-Antari, M. A., Al-Masni, M. A., Park, S. U., Park, J., Metwally, M. K., Kadah, Y. M., ... & Kim, T. S. (2018). An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network. *Journal of Medical and Biological Engineering*, 38(3), 443-456.
- [10] Xi, P., Shu, C., & Goubran, R. (2018, June). Abnormality detection in mammography using deep convolutional neural networks. In *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 1-6). IEEE.

- [11] Samala, R. K., Chan, H. P., Hadjiiski, L., Helvie, M. A., Wei, J., & Cha, K. (2016). Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Medical physics*, 43(12), 6654-6666.
- [12] Doi, K. (2006). Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Physics in Medicine & Biology*, 51(13), R5.
- [13] Dibden, A., Offman, J., Duffy, S. W., & Gabe, R. (2020). Worldwide review and meta-analysis of cohort studies measuring the effect of mammography screening programmes on incidence-based breast cancer mortality. *Cancers*, 12(4), 976.
- [14] Fiorica, J. V. (2016). Breast cancer screening, mammography, and other modalities. *Clinical obstetrics and gynecology*, 59(4), 688-709.
- [15] Johnson, B. (2021). Asymmetries in Mammography. *Radiologic Technology*, 92(3), 281M-298M.
- [16] van Seijen, M., Lips, E. H., Thompson, A. M., Nik-Zainal, S., Futreal, A., Hwang, E. S., ... & Wesseling, J. (2019). Ductal carcinoma in situ: to treat or not to treat, that is the question. *British journal of cancer*, 121(4), 285-292.
- [17] Lamb, L. R., Mohallem Fonseca, M., Verma, R., & Seely, J. M. (2020). Missed breast cancer: effects of subconscious bias and lesion characteristics. *Radiographics*, 40(4), 941-960.
- [18] American Journal of Roentgenology. 2013;201: W662-W670. 10.2214/AJR.12.10153.
- [19] Burnside, E. S., Sickles, E. A., Bassett, L. W., Rubin, D. L., Lee, C. H., Ikeda, D. M., ... & D'Orsi, C. J. (2009). The ACR BI-RADS® experience: learning from history. *Journal of the American College of Radiology*, 6(12), 851-860.
- [20] Lauby-Secretan, B. et al. Breast-cancer screening—viewpoint of the IARC Working Group. *N. Engl. J. Med.* 372, 2353–2358 (2015).
- [21] Couture, H. D., Williams, L. A., Geradts, J., Nyante, S. J., Butler, E. N., Marron, J. S., ... & Niethammer, M. (2018). Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ breast cancer*, 4(1), 1-8.
- [22] Celik, Y. et al. Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognit. Lett.* 133, 232–239 (2020).

- [23] Rahman, M. M. et al. Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features. *Irbm* 42, 215-226 (2021).
- [24] Ramadan, S. Z. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *Journal of healthcare engineering* (2020).
- [25] Taghanaki, S. A. et al. Deep semantic segmentation of natural and medical images: a review. *Artif. Intell. Rev.* 54, 137–178 (2021).
- [26] Wang, G., Li, W., Ourselin, S. & Vercauteren, T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI Brainlesion Workshop*, 178–190 (Springer, Cham, 2017).
- [27] Giacomello, E., Loiacono, D. & Mainardi, L. Brain MRI tumor segmentation with adversarial networks. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (IEEE, 2020).
- [28] Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. *Computational and structural biotechnology journal*, 15, 75-85.
- [29] Li, H., Meng, X., Wang, T., Tang, Y., & Yin, Y. Breast masses in mammography classification with local contour features. *Biomedical engineering online* 16, 1-12 (2017).
- [30] Virmani, J., Dey, N., & Kumar, V. (2016). PCA-PNN and PCA-SVM based CAD systems for breast density classification. In *Applications of intelligent optimization in biology and medicine* (pp. 159-180). Springer, Cham.
- [31] Kozegar, E., Soryani, M., Minaei, B., & Domingues, I. (2013). Assessment of a novel mass detection algorithm in mammograms. *Journal of cancer research and therapeutics*, 9(4), 592.
- [32] Agarwal, R., Diaz, O., Lladó, X., Yap, M. H., & Martí, R. (2019). Automatic mass detection in mammograms using deep convolutional neural networks. *Journal of Medical Imaging*, 6(3), 031409.
- [33] Baccouche, A., Ahmed, S., Sierra-Sosa, D., & Elmaghraby, A. (2020). Malicious text identification: deep learning from public comments and emails. *Information*, 11(6), 312.
- [34] Baccouche, A., Garcia-Zapirain, B., & Elmaghraby, A. (2018, December). Annotation technique for health-related tweets sentiment analysis. In *2018 IEEE international symposium on signal processing and information technology (ISSPIT)* (pp. 382-387). IEEE.

- [35] Yassin, N. I., Omran, S., El Houby, E. M., & Allam, H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer methods and programs in biomedicine* 156, 25-45 (2018).
- [36] Paramkusham, S., Thotempuddi, J., & Rayudu, M. S. Breast masses classification using contour shape descriptors based on Beam Angle Statistics. In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1340-1345. (IEEE, 2021).
- [37] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, 9(1), 1-12.
- [38] Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3), 257-273.
- [39] Qiu, Y., Yan, S., Gundreddy, R. R., Wang, Y., Cheng, S., Liu, H., & Zheng, B. (2017). A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. *Journal of X-ray Science and Technology*, 25(5), 751-763.
- [40] Yap, M. H., Pons, G., Martí, J., Ganau, S., Sentis, M., Zwiggelaar, R., ... & Martí, R. (2017). Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 22(4), 1218-1226.
- [41] Chakraborty, J., Midya, A., & Rabidas, R. (2018). Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. *Expert Systems with Applications*, 99, 168-179.
- [42] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 764-773).
- [43] Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., ... & Wang, J. (2019). High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv:1904.04514*.
- [44] Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2012). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1915-1929.
- [45] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904-1916.

- [46] Gao, Y., Geras, K. J., Lewin, A. A. & Moy, L. New frontiers: an update on computeraided diagnosis for breast imaging in the age of artificial intelligence. *Am. J. Roentgenol.* 212, 300–307 (2019).
- [47] Henriksen, E. L., Carlsen, J. F., Vejborg, I. M., Nielsen, M. B. & Lauridsen, C. A. The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review. *Acta Radiol.* 60, 13–18 (2019).
- [48] Mullooly, M. et al. Application of convolutional neural networks to breast biopsies to delineate tissue correlates of mammographic breast density. *npj Breast Cancer* 5, 1–11 (2019).
- [49] Al-Masni, M. A. et al. 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 (2018).
- [50] Eltrass, A. S., & Salama, M. S. Fully automated scheme for computer-aided detection and breast cancer diagnosis using digitised mammograms. *IET Image Processing* 14, 495-505 (2020).
- [51] Siddiqui, S. Y. et al. Intelligent breast cancer prediction empowered with fusion and deep learning. *Computer, Materials and Continua* 67, 1033–1049 (2021).
- [52] Aly, G. H., Marey, M., El-Sayed, S. A., & Tolba, M. F. YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms. *Computer Methods and Programs in Biomedicine* 200, 105823 (2021).
- [53] Abdelhafiz, D., Bi, J., Ammar, R., Yang, C. & Nabavi, S. Convolutional neural network for automated mass segmentation in mammography. *BMC Bioinformatics* 21, 1–19 (2020).
- [54] Taghanaki, S. A. et al. Deep semantic segmentation of natural and medical images: a review. *Artif. Intell. Rev.* 54, 137–178 (2021).
- [55] Jiao, Z., Gao, X., Wang, Y., & Li, J. A parasitic metric learning net for breast mass classification based on mammography. *Pattern Recognition* 75, 292-301 (2018).
- [56] Rasti, R., Teshnehlal, M., & Phung, S. L. Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks. *Pattern Recognition* 72, 381-390 (2017).
- [57] Antropova, N. O., Abe, H., & Giger, M. L. (2018). Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *Journal of Medical Imaging*, 5(1), 014503.

- [58] Vorontsov, E., Cerny, M., Régnier, P., Di Jorio, L., Pal, C. J., Lapointe, R., ... & Tang, A. (2019). Deep learning for automated segmentation of liver lesions at CT in patients with colorectal cancer liver metastases. *Radiology: Artificial Intelligence*, 1(2), 180014.
- [59] Gardezi, S. J. S., Elazab, A., Lei, B., & Wang, T. (2019). Breast cancer detection and diagnosis using mammographic data: Systematic review. *Journal of medical Internet research*, 21(7), e14464.
- [60] Houssami, N., Kirkpatrick-Jones, G., Noguchi, N., & Lee, C. I. (2019). 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*, 16(5), 351-362.
- [61] Peng, J., Bao, C., Hu, C., Wang, X., Jian, W., & Liu, W. (2020). Automated mammographic mass detection using deformable convolution and multiscale features. *Medical & biological engineering & computing*, 58(7), 1405-1417.
- [62] Al-Antari, M. A., Al-Masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International journal of medical informatics*, 117, 44-54.
- [63] Al-Antari, M. A., Han, S. M., & Kim, T. S. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. *Computer methods and programs in biomedicine* 196, 105584 (2020).
- [64] Suzuki, S., Zhang, X., Homma, N., Ichiji, K., Sugita, N., Kawasumi, Y., ... & Yoshizawa, M. (2016, September). Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis. In 2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE) (pp. 1382-1386). IEEE.
- [65] Ribli, D., Horváth, A., Unger, Z., Pollner, P., & Csabai, I. (2018). Detecting and classifying lesions in mammograms with deep learning. *Scientific reports*, 8(1), 1-7.
- [66] Shen, R., Yao, J., Yan, K., Tian, K., Jiang, C., & Zhou, K. (2020). Unsupervised domain adaptation with adversarial learning for mass detection in mammogram. *Neurocomputing*, 393, 27-37.
- [67] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

- [68] Al-masni, M. A., Al-antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., ... & Kim, T. S. (2017, July). Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1230-1233). IEEE.
- [69] Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718.
- [70] Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., & Barkan, E. (2016). A region based convolutional network for tumor detection and classification in breast mammography. In Deep learning and data labeling for medical applications (pp. 197-205). Springer, Cham.
- [71] Dhungel, N., Carneiro, G., & Bradley, A. P. (2015, November). Automated mass detection in mammograms using cascaded deep learning and random forests. In 2015 international conference on digital image computing: techniques and applications (DICTA) (pp. 1-8). IEEE.
- [72] Akselrod-Ballin, A., Karlinsky, L., Hazan, A., Bakalo, R., Horesh, A. B., Shoshan, Y., & Barkan, E. (2017). Deep learning for automatic detection of abnormal findings in breast mammography. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 321-329). Springer, Cham.
- [73] Li, Y., Zhang, L., Chen, H., & Cheng, L. (2020). Mass detection in mammograms by bilateral analysis using convolution neural network. *Computer methods and programs in biomedicine*, 195, 105518.
- [74] Hamed, G., Marey, M., Amin, S. E., & Tolba, M. F. (2021). Automated Breast Cancer Detection and Classification in Full Field Digital Mammograms Using Two Full and Cropped Detection Paths Approach. *IEEE Access*, 9, 116898-116913.
- [75] Kumar, A., Mukherjee, S., & Luhach, A. K. (2019). Deep learning with perspective modeling for early detection of malignancy in mammograms. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(4), 627-643.
- [76] Watanabe, A. T., Lim, V., Vu, H. X., Chim, R., Weise, E., Liu, J., ... & Comstock, C. E. (2019). Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography. *Journal of digital imaging*, 32(4), 625-637.

- [77] Timp, S., Varela, C., & Karssemeijer, N. (2007). Temporal change analysis for characterization of mass lesions in mammography. *IEEE Transactions on Medical imaging*, 26(7), 945-953.
- [78] Timp, S., & Karssemeijer, N. (2006). Interval change analysis to improve computer aided detection in mammography. *Medical Image Analysis*, 10(1), 82-95.
- [79] Loizidou, K., Skouroumouni, G., Nikolaou, C., & Pitris, C. (2019, May). A new method for breast microcalcification detection and characterization using digital temporal subtraction of mammogram pairs. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (pp. 1-4). IEEE.
- [80] Loizidou, K., Skouroumouni, G., Nikolaou, C., & Pitris, C. (2020). An automated breast microcalcification detection and classification technique using temporal subtraction of mammograms. *IEEE Access*, 8, 52785-52795.
- [81] Zheng, Y., Yang, C., & Merkulov, A. (2018, May). Breast cancer screening using convolutional neural network and follow-up digital mammography. In *Computational Imaging III* (Vol. 10669, p. 1066905). International Society for Optics and Photonics.
- [82] Domínguez, A. R., & Nandi, A. K. (2009). Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognition*, 42(6), 1138-1148.
- [83] Dhungel, N., Carneiro, G., & Bradley, A. P. (2015, October). Deep learning and structured prediction for the segmentation of mass in mammograms. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 605-612). Springer, Cham.
- [84] Cardenas, C. E., Yang, J., Anderson, B. M., Court, L. E., & Brock, K. B. (2019, July). Advances in auto-segmentation. In *Seminars in radiation oncology* (Vol. 29, No. 3, pp. 185-197). WB Saunders.
- [85] Xu, X., Fu, L., Chen, Y., Larsson, R., Zhang, D., Suo, S., ... & Zhao, J. (2018, July). Breast region segmentation using convolutional neural network in dynamic contrast enhanced MRI. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 750-753). IEEE.
- [86] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- [87] Hai, J., Qiao, K., Chen, J., Tan, H., Xu, J., Zeng, L., ... & Yan, B. (2019). Fully convolutional densenet with multiscale context for automated breast tumor segmentation. *Journal of healthcare engineering*, 2019.

- [88] Zhu, W., Xiang, X., Tran, T. D., Hager, G. D., & Xie, X. (2018, April). Adversarial deep structured nets for mass segmentation from mammograms. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 847-850). IEEE.
- [89] Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, M. M. K., ... & Torrents-Barrena, J. (2020). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Systems with Applications*, 139, 112855.
- [90] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
- [91] Sathyan, A., Martis, D., & Cohen, K. (2020, November). Mass and Calcification Detection from Digital Mammograms Using UNets. In *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCFMI)* (pp. 229-232). IEEE.
- [92] Soulam, K. B., Kaabouch, N., Saidi, M. N., & Tamtaoui, A. (2021). Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. *Biomedical Signal Processing and Control*, 66, 102481.
- [93] Liu, L., Cheng, J., Quan, Q., Wu, F. X., Wang, Y. P., & Wang, J. (2020). A survey on U-shaped networks in medical image segmentations. *Neurocomputing*, 409, 244-258.
- [94] Abdelhafiz, D., Nabavi, S., Ammar, R., Yang, C., & Bi, J. (2019, September). Residual Deep Learning System for Mass Segmentation and Classification in Mammography. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 475-484).
- [95] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74-87.
- [96] Li, H., Chen, D., Nailon, W. H., Davies, M. E., & Laurenson, D. (2018). Improved breast mass segmentation in mammograms with conditional residual u-net. In *Image Analysis for Moving Organ, Breast, and Thoracic Images* (pp. 81-89). Springer, Cham.
- [97] Sinha, A., & Dolz, J. (2020). Multi-scale self-guided attention for medical image segmentation. *IEEE journal of biomedical and health informatics*.

- [98] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
- [99] Li, S., Dong, M., Du, G., & Mu, X. (2019). Attention dense-u-net for automatic breast mass segmentation in digital mammogram. *IEEE Access*, 7, 59037-59047.
- [100] Sun, H., Li, C., Liu, B., Liu, Z., Wang, M., Zheng, H., ... & Wang, S. (2020). AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. *Physics in Medicine & Biology*, 65(5), 055005.
- [101] Ravitha Rajalakshmi, N., Vidhyapriya, R., Elango, N., & Ramesh, N. (2020). Deeply supervised U-Net for mass segmentation in digital mammograms. *International Journal of Imaging Systems and Technology*.
- [102] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.
- [103] Wang, R., Ma, Y., Sun, W., Guo, Y., Wang, W., Qi, Y., & Gong, X. (2019). Multi-level nested pyramid network for mass segmentation in mammograms. *Neurocomputing*, 363, 313-320.
- [104] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6), 1856-1867.
- [105] Tsochatzidis, L., Koutla, P., Costaridou, L., & Pratikakis, I. (2021). Integrating segmentation information into CNN for breast cancer diagnosis of mammographic masses. *Computer Methods and Programs in Biomedicine*, 200, 105913.
- [106] Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., & Johansen, H. D. (2020, July). Doubleu-net: A deep convolutional neural network for medical image segmentation. In 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 558-564). IEEE.
- [107] Das, S., Deka, A., Iwahori, Y., Bhuyan, M. K., Iwamoto, T., & Ueda, J. (2019). Contour-Aware Residual W-Net for Nuclei Segmentation. *Procedia Computer Science*, 159, 1479-1488.
- [108] Tran, S. T., Cheng, C. H., Nguyen, T. T., Le, M. H., & Liu, D. G. (2021, January). TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation. In *Healthcare* (Vol. 9, No. 1, p. 54). Multidisciplinary Digital Publishing Institute.

- [109] Senaras, C., Niazi, M. K. K., Sahiner, B., Pennell, M. P., Tozbikian, G., Lozanski, G., & Gurcan, M. N. (2018). Optimized generation of high-resolution phantom images using cGAN: Application to quantification of Ki67 breast cancer images. *PloS one*, 13(5), e0196846.
- [110] Alyafi, B., Diaz, O., Elangovan, P., Vilanova, J. C., del Riego, J., & Marti, R. (2020, May). Quality analysis of DCGAN-generated mammography lesions. In *15th International Workshop on Breast Imaging (IWBI2020)* (Vol. 11513, p. 115130B). International Society for Optics and Photonics.
- [111] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).
- [112] Becker, A. S., Jendele, L., Skopek, O., Berger, N., Ghafoor, S., Marcon, M., & Konukoglu, E. (2019). Injecting and removing suspicious features in breast imaging with CycleGAN: A pilot study of automated adversarial attacks using neural networks on small images. *European journal of radiology*, 120, 108649.
- [113] Cai, J., Zhang, Z., Cui, L., Zheng, Y., & Yang, L. (2019). Towards cross-modal organ translation and segmentation: a cycle-and shape-consistent generative adversarial network. *Medical image analysis*, 52, 174-184.
- [114] Hiasa, Y., Otake, Y., Takao, M., Matsuoka, T., Takashima, K., Carass, A., ... & Sato, Y. (2018, September). Cross-modality image synthesis from unpaired data using CycleGAN. In *International workshop on simulation and synthesis in medical imaging* (pp. 31-41). Springer, Cham.
- [115] Huo, Y., Xu, Z., Bao, S., Assad, A., Abramson, R. G., & Landman, B. A. (2018, April). Adversarial synthesis learning enables segmentation without target modality ground truth. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)* (pp. 1217-1220). IEEE.
- [116] Kaji, S., & Kida, S. (2019). Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological physics and technology*, 12(3), 235-248.
- [117] Shen, T., Gou, C., Wang, J., & Wang, F. Y. (2020, November). Collaborative Adversarial Networks for Joint Synthesis and Segmentation of X-ray Breast Mass Images. In *2020 Chinese Automation Congress (CAC)* (pp. 1743-1747). IEEE.

- [118] Liao, H., Huo, Z., Sehnert, W. J., Zhou, S. K., & Luo, J. (2018, September). Adversarial sparse-view CBCT artifact reduction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 154-162). Springer, Cham.
- [119] Dhahri, H., Rahmany, I., Mahmood, A., Al Maghayreh, E., & Elkilani, W. (2020). Tabu search and machine-learning classification of benign and malignant proliferative breast lesions. *BioMed research international*, 2020.
- [120] Shams, S., Platania, R., Zhang, J., Kim, J., Lee, K., & Park, S. J. (2018, September). Deep generative breast cancer screening and diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 859-867). Springer, Cham.
- [121] Li, H., Zhuang, S., Li, D. A., Zhao, J., & Ma, Y. (2019). Benign and malignant classification of mammogram images based on deep learning. *Biomedical Signal Processing and Control*, 51, 347-354.
- [122] Zhang, Q., Li, Y., Zhao, G., Man, P., Lin, Y., & Wang, M. (2020). A novel algorithm for breast mass classification in digital mammography based on feature fusion. *Journal of Healthcare Engineering*, 2020.
- [123] Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., ... & Fujita, H. (2020). Improving breast mass classification by shared data with domain transformation using a generative adversarial network. *Computers in biology and medicine*, 119, 103698.
- [124] Chakravarthy, S. S., & Rajaguru, H. (2021). Automatic Detection and Classification of Mammograms Using Improved Extreme Learning Machine with Deep Learning. *IRBM*.
- [125] Khan, H. N., Shahid, A. R., Raza, B., Dar, A. H., & Alquhayz, H. (2019). Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. *IEEE Access*, 7, 165724-165733.
- [126] Kumar, I., Bhadauria, H. S., Virmani, J., & Thakur, S. (2017). A classification framework for prediction of breast density using an ensemble of neural network classifiers. *Biocybernetics and Biomedical Engineering*, 37(1), 217-228.
- [127] Falconí, L. G., Pérez, M., & Aguilar, W. G. Transfer learning in breast mammogram abnormalities classification with mobilenet and nasnet. In *2019 International Conference on Systems, Signals, and Image Processing (IWSSIP)*, 109-114 (IEEE, 2019).

- [128] Alkhaleefah, M., Ma, S. C., Chang, Y. L., Huang, B., Chittem, P. K., & Achhannagari, V. P. (2020). Double-shot transfer learning for breast cancer classification from X-ray images. *Applied Sciences*, 10(11), 3999.
- [129] Falconí, L., Pérez, M., Aguilar, W., & Conci, A. (2020, July). Transfer Learning and Fine Tuning in Mammogram BI-RADS Classification. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 475-480). IEEE.
- [130] Medeiros, A., Ohata, E. F., Silva, F. H., Rego, P. A., & Reboucas Filho, P. P. (2020, July). An approach to BI-RADS uncertainty levels classification via deep learning with transfer learning technique. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 603-608). IEEE.
- [131] Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., ... & Aerts, H. J. (2019). Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: a cancer journal for clinicians*, 69(2), 127-157.
- [132] Tariq, M., Iqbal, S., Ayesha, H., Abbas, I., Ahmad, K. T., & Niazi, M. F. K. (2021). Medical image based breast cancer diagnosis: State of the art and future directions. *Expert Systems with Applications*, 167, 114095.
- [133] Li, H., Chen, D., Nailon, W. H., Davies, M. E., & Laurenson, D. (2020). Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography. *arXiv preprint arXiv:2008.02957*.
- [134] Sarkar, P. R., Prabhakar, P., Mishra, D., & Subrahmanyam, G. (2019, October). Towards automated breast mass classification using deep learning framework. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 453-462). IEEE.
- [135] Dhungel, N., Carneiro, G., & Bradley, A. P. (2017, April). Fully automated classification of mammograms using deep residual neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 310-314). IEEE.
- [136] Singh, V. K., Rashwan, H. A., Abdel-Nasser, M., Sarker, M., Kamal, M., Akram, F., ... & Puig, D. (2019). An efficient solution for breast tumor segmentation and classification in ultrasound images using deep adversarial learning. *arXiv preprint arXiv:1907.00887*.
- [137] Al-Antari, M. A., Al-Masni, M. A., & Kim, T. S. (2020). Deep learning computer-aided diagnosis for breast lesion in digital mammogram. *Deep Learning in Medical Image Analysis*, 59-72.

- [138] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
- [139] Wang, D., Li, C., Wen, S., Han, Q. L., Nepal, S., Zhang, X., & Xiang, Y. (2021). Daedalus: Breaking nonmaximum suppression in object detection via adversarial examples. *IEEE Transactions on Cybernetics*.
- [140] Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58, 101552.
- [141] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).
- [142] Hiasa, Y. et al. Cross-modality image synthesis from unpaired data using CycleGAN. In *International workshop on simulation and synthesis in medical imaging*, 31-41 (Springer Cham 2018).
- [143] Yoo, T. K., Choi, J. Y. & Kim, H. K. CycleGAN-based deep learning technique for artifact reduction in fundus photography. *Graefe's Archive for Clinical and Experimental Ophthalmology* 258, 1631-1637 (2020).
- [144] Deng, W. et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 994-1003 (2018).
- [145] Russ, T. et al. Synthesis of CT images from digital body phantoms using CycleGAN. *International journal of computer assisted radiology and surgery* 14, 1741-1750 (2019).
- [146] Sandfort, V., Yan, K., Pickhardt, P. J. & Summers, R. M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific reports* 9, 1-9 (2019).
- [147] Bao, F., Neumann, M. & Vu, N. T. CycleGAN-Based Emotion Style Transfer as Data Augmentation for Speech Emotion Recognition. In *INTERSPEECH*, 2828-2832 (2019).
- [148] Yu, X., Kang, C., Guttery, D. S., Kadry, S., Chen, Y., & Zhang, Y. D. (2020). ResNet-SCDA-50 for breast abnormality classification. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(1), 94-102.

- [149] Garcia-Gasulla, D., Parés, F., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., ... & Suzumura, T. (2018). On the behavior of convolutional nets for feature extraction. *Journal of Artificial Intelligence Research*, 61, 563-592.
- [150] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [151] Ragab, D. A., Attallah, O., Sharkas, M., Ren, J., & Marshall, S. (2021). A framework for breast cancer classification using multi-DCNNs. *Computers in Biology and Medicine*, 131, 104245.
- [152] Ferreira, C. A., Melo, T., Sousa, P., Meyer, M. I., Shakibapour, E., Costa, P., & Campilho, A. (2018, June). Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In *International conference image analysis and recognition* (pp. 763-770). Springer, Cham.
- [153] He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In *European conference on computer vision* (pp. 630-645). Springer, Cham.
- [154] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.
- [155] Chen, Y., Zhang, Q., Wu, Y., Liu, B., Wang, M., & Lin, Y. (2018, September). Fine-tuning ResNet for breast cancer classification from mammography. In *The International Conference on Healthcare Science and Engineering* (pp. 83-96). Springer, Singapore.
- [156] Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., & Elmaghraby, A. (2020). Ensemble deep learning models for heart disease classification: a case study from Mexico. *Information*, 11(4), 207.
- [157] Bellmann, P., Thiam, P., & Schwenker, F. (2018). Multi-classifier-systems: architectures, algorithms, and applications. In *Computational Intelligence for Pattern Recognition* (pp. 83-113). Springer, Cham.
- [158] Lee, R. S. et al. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data* 4, 1–9 (2017).
- [159] Moreira, I. C. et al. INbreast: toward a full-field digital mammographic database. *Acad. Radiol.* 19, 236–248 (2012).
- [160] Zheng, Y., Yang, C., Merkulov, A., & Bandari, M. (2016, May). Early breast cancer detection with digital mammograms using Haar-like features and AdaBoost algorithm. In *Sensing and Analysis*

Technologies for Biomedical and Cognitive Applications 2016 (Vol. 9871, p. 98710D). International Society for Optics and Photonics.

[161] Kallenberg, M., & Karssemeijer, N. (2010, June). Comparison of tilt correction methods in full field digital mammograms. In *International Workshop on Digital Mammography* (pp. 191-196). Springer, Berlin, Heidelberg.

[162] Mehmood, M., Ayub, E., Ahmad, F., Alruwaili, M., Alrowaili, Z. A., Alanazi, S., ... & Alyas, T. (2021). Machine learning enabled early detection of breast cancer by structural analysis of mammograms. *Computers, Materials and Continua*, 67(1), 641-657.

[163] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.

[164] Yadav, G., Maheshwari, S., & Agarwal, A. (2014, September). Contrast limited adaptive histogram equalization based enhancement for real time video system. In *2014 International Conference on Advances in Computing, Communications, and Informatics (ICACCI)* (pp. 2392-2397). IEEE.

[165] Wang, Y., Tao, D., Gao, X., Li, X., & Wang, B. (2011). Mammographic mass segmentation: embedding multiple features in vector-valued level set in ambiguous regions. *Pattern Recognition*, 44(9), 1903-1915.

[166] Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P. A. (2016). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging*, 36(4), 994-1004.

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

[168] Zhang, Y., Chan, S., Park, V. Y., Chang, K. T., Mehta, S., Kim, M. J., ... & Su, M. Y. (2020). Automatic detection and segmentation of breast cancer on MRI using mask R-CNN trained on non-fat-sat images and tested on fat-sat images. *Academic Radiology*.

[169] Müller, R., Kornblith, S., & Hinton, G. (2019). When does label smoothing help? *arXiv preprint arXiv:1906.02629*.

[170] Hernández, P. A., Estrada, T. T., Pizarro, A. L., Cisternas, M. D., & Tapia, C. S. (2016). Breast calcifications: description and classification according to bi-rads 5th edition. *Rev. Chil. Radiol*, 22, 80-91.

- [171] Rajendran, P. T., Krishnapillai, V., Tamanang, S., & Chelliah, K. K. (2012). Comparison of image quality criteria between digital storage phosphor plate in mammography and full-field digital mammography in the detection of breast cancer. *The Malaysian journal of medical sciences: MJMS*, 19(1), 52.
- [172] Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3, 100004.
- [173] Antropova, N., Huynh, B. Q., & Giger, M. L. (2017). A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical physics*, 44(10), 5162-5171.
- [174] Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., & Xin, J. (2019). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*, 7, 105146-105158.
- [175] Baccouche, A., Garcia-Zapirain, B., Olea, C. C., & Elmaghraby, A. S. (2021). Breast lesions detection and classification via yolo-based fusion models. *Comput. Mater. Contin.*, 69, 1407-1425.
- [176] Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., & Elmaghraby, A. S. (2021). Connected-UNets: a deep learning architecture for breast mass segmentation. *NPJ Breast Cancer*, 7(1), 1-12.
- [177] Baccouche, A., Garcia-Zapirain, B., & Elmaghraby, A. (2022). An integrated Framework for Breast Mass Classification and Diagnosis using Stacked Ensemble of Residual Neural Networks. Preprint.
- [178] Falconi, L. G., Perez, M., Aguilar, W. G., & Conci, A. Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. *Advances in Science, Technology and Engineering Systems* 5, 154-165 (2020).
- [179] Coelho, C. N., Kuusela, A., Li, S., Zhuang, H., Ngadiuba, J., Aarrestad, T. K., ... & Summers, S. (2021). Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors. *Nature Machine Intelligence*, 3(8), 675-686.
- [180] Chang, D., Lee, J., & Heo, J. (2021). Lightweight of ONNX using Quantization-based Model Compression. *The Journal of the Institute of Internet, Broadcasting and Communication*, 21(1), 93-98.
- [181] Ishan Dutta, Optimizing Models with Post-Training Quantization in Keras - Part I. Weights & Biases. <https://wandb.ai/ishandutta/Post%20Training%20Quantization/reports/Optimizing-Models-with-Post-Training-Quantization-in-Keras-Part-I--VmlldzoxNDI3ODQz>. Accessed April 10, 2022.

- [182] Irfan, R., Almazroi, A. A., Rauf, H. T., Damaševičius, R., Nasr, E. A., & Abdelgawad, A. E. (2021). Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion. *Diagnostics*, *11*(7), 1212.
- [183] Li, Y., Liu, Y., Huang, L., Wang, Z., & Luo, J. (2022). Deep weakly-supervised breast tumor segmentation in ultrasound images with explicit anatomical constraints. *Medical Image Analysis*, *76*, 102315.
- [184] Sharma, S., & Kumar, S. (2022). The Xception model: A potential feature extractor in breast cancer histology images classification. *ICT Express*, *8*(1), 101-108.
- [185] Ayana, G., Park, J., Jeong, J. W., & Choe, S. W. (2022). A Novel Multistage Transfer Learning for Ultrasound Breast Cancer Image Classification. *Diagnostics*, *12*(1), 135.
- [186] Isfahani, Z. N., Jannat-Dastjerdi, I., Eskandari, F., Ghouschi, S. J., & Pourasad, Y. (2021). Presentation of novel hybrid algorithm for detection and classification of breast cancer using growth region method and probabilistic neural network. *Computational Intelligence and Neuroscience*, 2021.
- [187] Silalahi, A. R. J. (2021, March). Breast Cancer Lesion Detection and Classification in mammograms using Deep Neural. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1115, No. 1, p. 012018). IOP Publishing.
- [188] Yasir, N., Anwar, S., & Khan, M. T. (2022). Machine Vision based Intelligent Breast Cancer Detection. *Pakistan Journal of Engineering and Technology*, *5*(1), 1-10.
- [189] Zhou, Y., Zhang, C., & Gao, S. (2022). Breast Cancer Classification From Histopathological Images Using Resolution Adaptive Network. *IEEE Access*, *10*, 35977-35991.
- [190] Baccouche, A., Garcia-Zapirain, B., 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*, 106884.

CURRICULUM VITAE

Asma Baccouche

Northwestern Memorial Healthcare

Chicago, IL, USA

E-mail: asma.baccouche@louisville.edu,

asmabaccouche@live.fr

Tel: (502) 907-8970

Education

- 2019-Present Ph.D. Student, Computer Science and Engineering
PhD Candidate since Fall 2020, GPA = 3.86
Department of Computer Engineering and Computer Science,
University of Louisville, Louisville, KY 40292, USA
- 2018-2019 M.Sc., Computer Science, GPA = 3.91
Department of Computer Engineering and Computer Science,
University of Louisville, Louisville, KY 40292, USA
M.Sc. Project: A Qualitative Study of Speed School Publications Using Text
Clustering Analysis
- 2014-2017 BSc., Engineering, Distinction: Excellent
Tunisia Polytechnic School, La Marsa, 2078, Tunisia
Sr. Project: Enhancing Sentiment Analysis of Health-Related Tweets Using Deep
Learning

Experience

2021-2022	Data Engineering Intern at Kindred Healthcare – Hive at UofL, Louisville, KY
2018-2022	Graduate Assistant at Speed Technology Solutions - Speed School of Engineering, Louisville, KY
2017	Natural Language Processing Researcher at Innovative and Emerging Technologies Lab, Department of Computer Science and Engineering, University of Louisville, Louisville, KY
2016	Engineering Intern at Axe Finance – R&D Team, Tunis, Tunisia

Awards and Recognition

- *CSE Doctoral Award*, Speed School of Engineering, University of Louisville, Louisville, KY
- *1st Winner for Master Category of E-Expo*, Student Research Exposition Engineering Exposition E-Expo, University of Louisville, Louisville, KY
- *Top 10* in national competition for admission to Engineering Schools, Tunis, Tunisia

Publications

- **Journal Articles:**
 - ◇ **Baccouche, A.**, Garcia-Zapirain, B., 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*, 106884.
 - ◇ **Baccouche, A.**, Garcia-Zapirain, B., & Elmaghraby, A. (2022). An integrated Framework for Breast Mass Classification and Diagnosis using Stacked Ensemble of Residual Neural Networks. *Scientific Reports – Preprint 2022*
 - ◇ **Baccouche, A.**, Garcia-Zapirain, B., Castillo Olea, C., & Elmaghraby, A. S. (2021). Connected-UNets: a deep learning architecture for breast mass segmentation. *NPJ Breast Cancer*, 7(1), 1-12.

- ◇ **Baccouche, A.**, Garcia-Zapirain, B., Olea, C. C., & Elmaghraby, A. S. (2021). Breast lesions detection and classification via yolo-based fusion models. *Comput. Mater. Contin.*, 69, 1407-1425.
- ◇ **Baccouche, A.**, Ahmed, S., Sierra-Sosa, D., & Elmaghraby, A. (2020). Malicious text identification: deep learning from public comments and emails. *Information*, 11(6), 312.
- ◇ **Baccouche, A.**, Garcia-Zapirain, B., Castillo Olea, C., & Elmaghraby, A. (2020). Ensemble deep learning models for heart disease classification: a case study from Mexico. *Information*, 11(4), 207.
- **Conference Proceedings:**
 - ◇ Baccouche, A., Garcia-Zapirain, B., & Elmaghraby, A. (2018, December). Annotation technique for health-related tweets sentiment analysis. In *2018 IEEE international symposium on signal processing and information technology (ISSPIT)* (pp. 382-387). IEEE.

Graduate Advisors

Dr. Adel Elmaghraby, Dissertation Co-advisor, Department of Computer Science and Engineering, University of Louisville; **Dr. Maria Begona Garcia Zapirain**, Dissertation Co-advisor, eVida Research Group, Department of Computer Science, Electronics and Telecommunications, Deusto University, Bilbao Spain; and **Dr. Daniel Sierra-Sosa**, Dissertation Co-advisor, Department of Computer Science and Information Technology, Hood College.