الجمهورية الجزائرية الديمقراطية الشعبية

Ministry of Higher Education and Scientific Research
Yahia Fares University of Medea
Faculty of Sciences
Department of Mathematics and Computer Science



Graduation Project Report Submitted for Master of Computer Science degree
Speciality : Applied Artificial Intelligence

Deep Learning-Based Approach for Breast Lesion Segmentation in Ultrasound Images

---

Prepared by :
. Nesrine Djellal
. Nouha Nedjraoui

Supervised by :
. Dr. Ahmed cherif Mazari
. Dr. Nesrine Boudoukhani

Academic Year : 2024/2025

DEDICATION (Nouha Nedjraoui)

I praise and thank Allah for granting me the strength and ability to achieve this.

I dedicate this work to my beloved parents, who have always been by my side, encouraging and supporting me.

To my brother and sisters, who bring happiness to my life with their cheerful spirits.

To my teachers, especially Dr. Ahmed Cherif Mazari and Dr. Boudoukhani Nesrine, for the knowledge and guidance they provided me.
I am deeply grateful to Dr. Boudoukhani Nesrine for the unwavering support she has provided us. Her kindness and understanding mean so much to me.

To my cherished friend Nesrine Djellal who walked alongside me on this journey.

To everyone who, after the grace of Allah Almighty, has played a significant role in my achievements.

DEDICATION (Nesrine Djellal)

All praise and thanks are due to Allah, whose mercy, strength, and guidance have carried me through every step of this journey.

I dedicate this work to my dear parents, whose unwavering support, endless prayers, and unconditional love have been my greatest source of strength and motivation.

To my beloved brother and sisters, whose presence fills my life with joy and lightness, reminding me always of the beauty of family.

To my teachers, especially Dr. Ahmed Cherif Mazari and Dr. Boudoukhani Nesrine, for their invaluable guidance, encouragement, and dedication to our learning. I am especially grateful to Dr. Boudoukhani Nesrine for her sincere support, patience, and kindness throughout this process, it has meant more than words can express.

To my cherished friend, Nouha Nedjraoui, thank you for being by my side in this journey, sharing both the challenges and the triumphs.

And to everyone who, after the grace of Allah Almighty, has contributed to my growth, learning, and success, this work is for you.

ABSTRACT

This thesis explores deep learning-based approaches for automated breast lesion segmentation in Ultrasound (US) images, aiming to enhance the accuracy and efficiency of computer-aided diagnosis (CAD) systems. Breast cancer diagnosis often faces challenges due to the variability in lesion morphology, low contrast in US images, and limited annotated datasets. To address these issues, this study investigates state-of-the-art segmentation architectures, including U-Net variants, transformer-based models, and hybrid CNN-transformer frameworks, alongside advanced preprocessing and augmentation techniques. The proposed methodology evaluates baseline and novel models using metrics such as Dice score and Intersection-over-Union (IoU), with a focus on improving robustness to dataset imbalances and image complexity. Experimental results demonstrate the effectiveness of deep learning in achieving precise lesion delineation, offering potential clinical utility for early and non-invasive breast cancer and other tumor detection.

Keywords : computer-aided diagnosis (CAD), Breast lesion segmentation, ultrasound (US), U-Net, attention mechanism .

ملخص

تستكشف هذه الأطروحة الأساليب القائمة على التعلم العميق لتجزئة آفة الثدي آلياً في صور الموجات فوق الصوتية (US)، بهدف تعزيز دقة وكفاءة أنظمة التشخيص بمساعدة الحاسوب (CAD). غالبًا ما يواجه تشخيص سرطان الثدي تحديات بسبب التباين في شكل الآفة، والتباين المنخفض في صور ال (US)، ومحدودية مجموعات البيانات المشروحة.و لمعالجة هذه المشكلات، تبحث هذه الدراسة في أحدث بنيات التجزئة، بما في ذلك متغيرات شبكة U-Net والنماذج القائمة على المحولات (Transformers)، ونماذج هجينة تجمع بين الشبكات العصبية الالتفافية (CNN) والمحولات (Transformers) ، إلى جانب تقنيات متقدمة للمعالجة المسبقة للبيانات والتعزيز(augmentation). تُقيّم المنهجية المقترحة النماذج الأساسية وأخرى جديدة باستخدام مقاييس مثل معامل Dice ومعامل التقاطع على الاتحاد (IoU)، مع التركيز على تحسين المتانة في مواجهة اختلالات التوازن في مجموعات البيانات وتعقيد الصورة. تُظهر النتائج التجريبية فعالية تقنيات التعلم العميق (DL) في تحقيق التحديد الدقيق للآفات، مما يعزز من إمكاناتها السريرية في الكشف المبكر وغير الجراحي عن سرطان الثدي وأنواع أخرى من الأورام.

الكلمات المفتاحية : التشخيص بمساعدة الحاسوب (CAD)، تقسيم آفات الثدي، التصوير بالأمواج فوق الصوتية(US)، شبكة U-Net، آلية الانتباه (attention mechanism).

RESUME

Cette thèse explore les approches basées sur l'apprentissage profond pour la segmentation automatisée des lésions mammaires dans les images échographiques (US), dans le but d'améliorer la précision et l'efficacité des systèmes de diagnostic assisté par ordinateur (CAD). Le diagnostic du cancer du sein se heurte souvent à des difficultés dues à la variabilité de la morphologie des lésions, au faible contraste des images échographiques et à l'insuffisance des ensembles de données annotées. Pour résoudre ces problèmes, cette étude examine les architectures de segmentation les plus modernes, y compris les variantes U-Net, les modèles basés sur des transformateurs et les cadres hybrides CNN-transformateurs, ainsi que les techniques avancées de prétraitement et d'augmentation. La méthodologie proposée évalue les modèles de base et les nouveaux modèles à l'aide de mesures telles que le score de Dice et l'intersection sur l'union (IoU), en mettant l'accent sur l'amélioration de la robustesse aux déséquilibres de l'ensemble de données et à la complexité de l'image. Les résultats expérimentaux démontrent l'efficacité de l'apprentissage profond dans la délimitation précise des lésions, offrant une utilité clinique potentielle pour la détection précoce et non invasive du cancer du sein et d'autres tumeurs.

Mots-clés : diagnostic assisté par ordinateur (CAD), segmentation des lésions mammaires, échographie (US), U-Net, mécanisme d'attention.

TABLE OF CONTENTS

ABSTRACT.....Ⅳ
TABLE OF CONTENTS.....Ⅶ
LIST OF FIGURES.....Ⅸ
LIST OF TABLES.....Ⅺ
LIST OF ACRONYMS.....Ⅻ
GENERAL INTRODUCTION.....**1**
1 Chapter 1 : OVERVIEW : MEDICAL COMPUTER AIDED DIAGNOSIS.....**5**
1.1 Introduction.....5
1.2 Medical Imaging.....5
1.3 Computer Aided Diagnosis in Medicine.....6
1.4 Key Components of a CAD System.....6
1.4.1 Image Acquisition and Modalities.....7
1.4.2 Preprocessing.....9
1.4.3 CAD tasks employment.....9
1.4.4 Integration of tasks.....13
1.5 Conclusion.....14
2 Chapter 2 : LITERATURE REVIEW - Deep Learning Applications for Breast Lesion Segmentation in Ultrasound.....**16**
2.1 Introduction.....16
2.2 Significance of medical imaging for Breast lesion diagnosis.....16
2.3 Key Deep Learning Approaches for breast lesion segmentation in Ultrasound.....18
2.3.1 U-Net & Variants.....18
2.3.2 Transformer-Based Architectures.....21
2.3.3 Hybrid Models.....22
2.4 Training and Optimization of Deep Segmentation Models : Key Hyperparameter Choices for Building Effective Segmentors.....23
2.4.1 Loss Functions and Optimization Strategies.....23
2.4.2 Regularization and Data Augmentation.....23
2.5 Performance Evaluation Metrics.....24
2.5.1 Metrics Focusing on Classification Accuracy.....25
2.5.2 Metrics Focusing on Localization Correctness.....26
2.6 Conclusion.....27
3 Chapter 3 : PROPOSED APPROACH.....**29**
3.1 Introduction.....29
3.2 Problem statement.....29
3.2.1 Limited Annotated Data.....29
3.2.2 Class Imbalance.....30
3.2.3 Complex Task and Noisy Images.....31
3.3 Motivation.....32
3.3.1 Preprocessing.....32
3.3.2 Data augmentation and regularization techniques.....32

3.3.3 Dilated Convolutions..... 32
3.3.4 Multiscale Feature Learning..... 33
3.3.5 Attention Mechanism..... 33
3.3.6 Skip Connections..... 33
3.3.7 Vision Transformer blocks..... 33
3.3.8 Combined Loss Function..... 34
3.4 Methodology..... 34
3.4.1 Preprocessing..... 34
3.4.2 Augmentation Techniques..... 35
3.4.3 key components of Proposed framework..... 36
3.4.4 Baseline models..... 39
3.4.5 Proposed models..... 40
3.5 Conclusion..... 41
4 Chapter 4 : Results and Discussion..... 43
4.1 Introduction..... 43
4.2 Experimental Study..... 43
4.2.1 Overview Plan..... 43
4.2.2 Experimental Environment..... 46
4.2.3 Training and Evaluation Datasets..... 46
4.2.4 Hyperparameters for Models' Building and Training..... 48
4.2.5 Performance Evaluation Metrics..... 49
4.3 Experimental Results..... 50
4.3.1 Cross-validation on BUSI 1..... 50
4.3.2 Hold-out Evaluation on BUSI 1..... 51
4.3.3 BUSI 1 + BUSI UCLM..... 56
4.3.4 Augmented BUSI 1 + BUSI UCLM..... 58
4.3.5 Brain tumor segmentation..... 59
4.4 Discussion..... 61
GENERAL CONCLUSION..... 63
REFERENCES..... 64

LIST OF FIGURES

Figure 1.1 Examples of ultrasound images : breast US....7
Figure 1.2 Example of polyps in GI Endoscopic images....8
Figure 1.3 Examples of Mammography....8
Figure 1.4 Diagram of Classification of skin lesions....10
Figure 1.5 Diagram of Supervised Learning for Breast Cancer classification....10
Figure 1.6 Detected hemorrhage while Green boxes are ground truth and blue are predicted boxes a) Subarachnoid, b) Subdural, c) Intraparenchymal, d) Chronic, e) Intraventricular, f) Epidural....11
Figure 1.7 Example of glioblastoma segmentation....12
Figure 1.8 Example of Integration of Tasks....13
Figure 2.1 U-Net architecture....18
Figure 2.2 Attention U-Net architecture....20
Figure 2.3 Attention U-Net architecture....20
Figure 2.4 The proportion of literature related to transformers in medical image segmentation literature....21
(Pu et al., 2024)....21
Figure 2.5 TransUnet applied transformer structure to medical image segmentation (Chen et al., 2021)....22
Figure 2.6 Segmentation Metrics (Huynh et al.2023) : (a) Error Type Visualization, (b) Dice coefficient, (c) Jaccard index....24
Figure 3.1 Sample 53 of BUS dataset showing a very small lesion....30
Figure 3.2 Samples from BUS dataset showing : (a) and (b) : noisy images, (c) and (d) : hard lesions....31
Figure 3.3 Schematic of dilated convolution (Yu & Koltun, 2015)....33
Figure 3.4 Before and after applying CLAHE preprocessing on BUS images....34
Figure 3.5 Data augmentation results....35
Figure 3.6 Auto-Encoder architecture for CV (Khosla, 2021)....36
Figure 3.7 ASPP module structure (Chen et al., 2017)....37
Figure 3.8 A VIT block (Dosovitskiy et al., 2021)....37
Figure 3.9 A Squeeze-and-Excitation block (Hu et al., 2018) 38
Figure 3.10 The structure of the attention gate....39
Figure 3.11 Proposed models Architectures : (a) hybrid model 1 using VIT, (b) hybrid model 2 using SE....40
Figure 4.1 Overview of the experimental plan....44
Figure 4.2 Ultrasound images samples from BUSI 1 and BUSI UCLM....47
Figure 4.3 MRI samples from Brain tumor dataset....47
Figure 4.4 Model 1 Folds segmentation comparison....51
Figure 4.5 Model 3 Folds segmentation comparison....51
Figure 4.6 Comparison of Training accuracy of each model on BUSI 1 dataset....52
Figure 4.7 Comparison of Training and Validation loss of each model on BUSI 1 dataset....52
Figure 4.8 Bar plot represents models evaluation metrics on BUSI 1....53
Figure 4.9 Models predictions comparison on some test samples....54

Figure 4.10 Box plots showing the distribution of evaluation metrics (part 1)..... 55
Figure 4.11 Box plots showing the distribution of evaluation metrics (part2)..... 56
Figure 4.12 Train and validation accuracy and loss plots of final hybrid models on combined BUSI 1 + BUSI UCLM datasets.....57
Figure 4.13 Bar plot represents models evaluation metrics on combined BUSI 1 + BUSI UCLM datasets..... 57
Figure 4.14 Train and validation accuracy and loss plots of final hybrid model 2 on Augmented combined BUSI 1 + BUSI UCLM datasets.....58
Figure 4.15 segmentation results of final hybrid model 2 on on Augmented combined BUSI 1 + BUSI UCLM datasets.....59
Figure 4.16 Train and Validation accuracy and loss plots (Final hybrid 2 on Brain seg).....60
Figure 4.17 final hybrid model 2 segmentation results on some Brain seg samples..... 60

LIST OF TABLES

Table 2 Relevant works on breast ultrasound medical images segmentation.....19
Table 4.1 Models tested in this study and their abbreviations.....45
Table 4.2 Distribution of training and test sets..... 48
Table 4.3 Training's Hyperparameters..... 49
Table 4.4 Transformer parameters.....49
Table 4.5 Model 1 and Model 3 evaluation results on BUSI 1 test set using cross-validation..
..... 50
Table 4.6 Models evaluation on BUSI dataset (underlined values represent best ones)..... 53
Table 4.7 Final hybrid models evaluation on combined BUSI 1 + BUSI UCLM datasets....57
Table 4.8 Final hybrid model 2 evaluation on Augmented combined BUSI 1 + BUSI UCLM
datasets.....58
Table 4.9 Final hybrid model 2 evaluation on Brain tumor segmentation dataset.....60

LIST OF ACRONYMS

AI	Artificial Intelligence
ASPP	Atrous Spatial Pyramid Pooling
BB	Bounding Boxes
BCE	Binary Cross Entropy
BUS	Breast Ultrasound
CAD	Computer Aided Diagnosis
CADe	Computer Aided Detection
CBAM	Convolutional Block Attention Module
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNNs	Convolutional Neural Network
CT	Computed Tomography
CV	Cross Validation
CVi	Computer Vision
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DSC	Dice Similarity Coefficient
FN	True Negative
FP	False Positive
GAN	Generative Adversarial Network
IoU	Intersection over Union
L_{BCE}	Binary Cross Entropy Loss
L_{Dice}	Dice Loss
ML	Machine Learning
PFM	Pyramid Feature Maps

ResNet	Residual Network
ROI	Region Of Interest
RPNs	Region Proposal Networks
TN	True Negative
TP	True Positive
ViT	Vision Transformer

GENERAL INTRODUCTION

Breast cancer is one of the most commonly diagnosed cancers worldwide, affecting millions of women annually. In 2022, 2.3 million new cases and 670,000 deaths occurred globally (Kim et al., 2025). With early detection and accurate diagnosis being critical for improving patient outcomes and survival rates.

Medical imaging (Beutel, 2000), particularly breast ultrasound, and other Modalities such as endoscopy, MRI, and CT play a vital role in diagnosing breast abnormalities, including normal tissue, benign lesions, and malignant Tumors, providing invaluable insights. However, interpreting US images is challenging due to their noisy nature and the subtle differences between benign and malignant cases; noise artifacts, inter-observer variability, and data scarcity (Zhou et al., 2021) make manual interpretation remain time-consuming and prone to human error. This has led to the emergence of Computer-Aided Diagnosis (CAD) systems, which leverage artificial intelligence (AI) to assist radiologists in making accurate and timely diagnosis.

In recent years, deep learning (DL) has emerged as a transformative force in medical imaging, revolutionising the field of medical image analysis by automating complex tasks from tumor segmentation to disease classification. By automating the extraction of patterns from complex imaging data, DL has surpassed traditional computer vision methods, which relied on manual feature engineering. This paradigm shift is driven by continuous advancements in DL architectures, including convolutional neural networks (CNNs) (LeCun et al., 1989) and vision transformers (ViTs) (Dosovitskiy et al., 2020), each addressing unique challenges in healthcare.

The evolution of medical image analysis has progressed from rule-based systems to hierarchical feature learning. CNNs, particularly **U-Net** (Ronneberger et al., 2015), with its skip-connected encoder-decoder architecture, has become the gold standard for medical segmentation .Transformers, initially designed for NLP, have been adapted to medical imaging (e.g., **TransUNet**; Chen et al., 2021) to model long-range dependencies in high-resolution images. These innovations, combined with **attention mechanisms** (Vaswani et al., 2017) (e.g., CBAM) that suppress noise and highlight diagnostically relevant regions (Woo et al., 2018), form a robust toolkit for automating medical image analysis.

Despite significant advancements in AI-based computer-aided diagnosis (CAD) systems for breast ultrasound (BUS) imaging, critical challenges persist in achieving robust segmentation and classification. First, the heterogeneous appearance of breast lesions—varying in shape, echogenicity, and boundary clarity—complicates consistent feature extraction, particularly for architectures lacking spatial-channel attention mechanisms (Cho et al., 2022). Second, BUS images are inherently prone to

noise and artifacts (e.g., speckle, shadowing, and reverberation), which degrade image quality and mislead conventional CNNs, as evidenced by the high false-positive rates in single-network approaches (Zhou et al., 2021). Third, dataset imbalance skews model training toward overrepresented classes (e.g., benign cases), reducing sensitivity to malignancies and limiting generalizability across clinical populations (Wu et al,.2024).

To address these gaps, this work proposes a hybrid deep learning framework integrating state-of-the-art architectures (eg., Attention U-Net) with a dynamic data augmentation strategy targeting misclassified samples during training. The model aims to achieve pixel-level segmentation accuracy (Dice > 0.76), improve classification performance (precision >90%, recall >88%) through *self-multi-head attention and squeeze-excitation blocks ; and enhance generalizability. Evaluation will benchmark against BUS and other domain dataset using metrics aligned with clinical standards (DSC,IoU) , addressing both technical and translational gaps in current CAD systems.

This project has the potential to significantly improve the accuracy and efficiency of breast cancer diagnosis. By automating the segmentation process, employing several DL architectures and techniques and achieving the generalizability of the proposed model, it can assist radiologists in making faster and more reliable diagnoses, ultimately improving patient outcomes.

This report is organized as follow :

Chapter 01 *: OVERVIEW : Computer Aided Diagnosis.*
This chapter introduces Computer-Aided Diagnosis (CAD) systems in medicine, covering key imaging modalities like ultrasound, MRI, and mammography. It examines core CAD tasks—classification, detection, and segmentation—and their clinical applications. Challenges in medical image analysis and preprocessing techniques are discussed, setting the stage for deep learning solutions.

Chapter 02 *: LITERATURE REVIEW : Deep Learning Applications for Medical Images and Breast Cancer Diagnosis.*
Focusing on breast lesion segmentation in ultrasound, this chapter reviews U-Net variants, transformer-based models, and hybrid approaches. Training strategies (Dice/BCE loss) and evaluation metrics (Dice, IoU) are analyzed, highlighting gaps in robustness and generalizability for clinical use.

Chapter 03 *: Proposed Approach.*
A novel methodology addresses key challenges: limited data, class imbalance, and image complexity. Combining advanced preprocessing with hybrid CNN-transformer architectures, the approach aims to improve segmentation accuracy. Training protocols and validation frameworks ensure reproducibility.

Chapter 04 : Result And Discussion.
Proposed models are evaluated on breast US datasets, showing improved Dice/IoU scores over baselines. Qualitative results and clinical applicability are discussed, along with limitations like computational costs. Benchmarks confirm the superiority of hybrid models.

CHAPTER 1

OVERVIEW
MEDICAL COMPUTER AIDED DIAGNOSIS

1 Chapter 1 : OVERVIEW : MEDICAL COMPUTER AIDED DIAGNOSIS

1.1 Introduction

Medical imaging has become an indispensable tool in modern healthcare, offering non-invasive and detailed visualization of anatomical structures and pathological conditions. As imaging technologies evolve, the volume and complexity of acquired data continue to grow, creating new demands for efficient and accurate interpretation. Computer-Aided Diagnosis (CAD) systems have emerged to address these challenges by augmenting radiological decision-making through automated analysis powered by machine learning (ML) and deep learning (DL) techniques. This chapter introduces the core principles of medical imaging and CAD, providing a foundation for understanding how different imaging modalities and intelligent algorithms contribute to disease detection, classification, localization, and segmentation. It also outlines the fundamental components of CAD pipelines, establishing the groundwork for the specialized methods developed and evaluated in later chapters.

1.2 Medical Imaging

Medical imaging (Beutel, 2000) refers to the generation of visual representations of internal or external organs and tissues of the human body for disease diagnosis and monitoring progression (Farhat, Sakr, & Kilany, 2020). This process can involve either invasive or non-invasive techniques and is based on various physical phenomena such as sound, light, electromagnetic radiation, nuclear magnetic resonance, and radioactivity (S. K. Zhou et al., 2021).

It utilizes specific regions of the electromagnetic spectrum, which includes all types of electromagnetic radiation with varying wavelengths and frequencies, each serving different diagnostic purposes.

The advent of digital medical imaging began in 1993 with the introduction of the DICOM (Digital Imaging and Communications in Medicine) standard . This standard allows the use of digital technology to capture, store, manipulate, display, and exchange medical images, replacing conventional film-based imaging methods (Hosseini & Dixon, 2016). DICOM enables medical images and associated metadata to be encoded from various imaging devices and integrated into Picture Archiving and Communication Systems (PACS), making them exchangeable via Health Information Exchange (HIE) networks (Mildenberger, Eichelberg, & Martin, 2002).

Digital medical imaging encompasses a wide range of techniques for visualizing different body organs called Multi-Organ Medical Imaging Modalities, such as X-rays, Ultrasound (US), Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI). Additional tools that serve specialized purposes are called Specific-Organ Medical Imaging Modalities , including Retinal Photography, Dermoscopy, Endoscopy, and Histopathology (Ker,

Wang, Rao, & Lim, 2018). The following subsections provide an overview of some of these commonly used medical imaging modalities in healthcare.

1.3 Computer Aided Diagnosis in Medicine

Computer-Aided Diagnosis (CAD) is a transformative technology in the field of medical imaging, formally defined by the FDA as "a system that provides image analysis to aid in detection, characterization, or diagnosis of disease" (Shiraishi et al.,2011), known for its capacity to support clinicians through automated analysis of images such as ultrasound (US), X-rays, CT scans, MRIs, and mammograms, leveraging algorithms, often based on machine learning (ML) or deep learning (DL), to detect, localize, and quantify pathological patterns, offering a "second opinion" that may complement the clinician's diagnosis (Davidson A, 2023).

Unlike autonomous diagnostic tools, CAD systems are assistive rather than substitutive. Their design philosophy hinges on augmenting clinical decision-making, especially in scenarios demanding high sensitivity such as early-stage cancer detection or subtle pathological signs that might elude even expert radiologists; subtle abnormalities may be missed due to human fatigue or variability (Santos et al., 2019). Through tasks like abnormality classification, ROI (region of interest) detection, and disease progression tracking, CAD tools bring quantitative rigor and reproducibility to diagnostic interpretation.

Moreover, in large-scale screening programs such as breast cancer, lung cancer, and diabetic retinopathy detection ,CAD systems play a critical role in reducing false negatives and standardizing diagnostic accuracy (Litjens et al., 2017; Al-masni et al., 2018). With the increasing shift toward data-driven diagnostics in medical imaging, CAD systems are becoming critical components of precision oncology, especially in breast cancer imaging. Their integration into multidisciplinary clinical workflows supports more consistent interpretations, reduces diagnostic variability (McKinney et al., 2020) , and enables triage by prioritizing high-risk lesions, identifying 98.3% of invasive cancers in the top 10% of suspicious cases (Rodriguez-Ruiz et al., 2019), ultimately improving diagnostic throughput and patient outcomes.

1.4 Key Components of a CAD System

A robust CAD system involves a multi-stage pipeline that includes image acquisition, preprocessing,feature extraction, employment of classification/ detection/ segmentation , and post-processing. Each stage ensures that medical images are properly standardized, analyzed, and interpreted with clinical relevance.

1.4.1 Image Acquisition and Modalities

Image acquisition is the foundational and indispensable first step in the pipeline of any Computer-Aided Diagnosis (CAD) system. It involves capturing medical image data from patients using specialized imaging modalities. The quality, resolution, contrast, and type of acquired data directly influence the effectiveness of subsequent preprocessing, segmentation, and classification tasks in CAD workflows. Consequently, image acquisition not only establishes the structural and spatial context of anatomical or pathological findings but also imposes critical constraints on the performance and generalizability of deep learning models used for analysis (Ker et al., 2018; Shen et al., 2017). Here come the most used modalities :

a• Ultrasound Imaging : Ultrasound (US)

Ultrasound (US) or Sonography uses high-frequency sound waves to create 2D images of internal organs and tissues. Ultrasound imaging does not employ any ionizing radiation and is widely available and extremely affordable (Farhat et al., 2020; Goel et al., 2016). This medical imaging modality is commonly used for clinical applications related to cardiology, genecology, obstetrics, and abdominal imaging (see Figure 1.1) (Ali, Magee, & Dasgupta, 2008).



Figure 1.1 Examples of ultrasound images : breast US [1]

b• Optical imaging : endoscopy

It consists of a medical examination and visualization procedure that uses a flexible lighted tube with a camera called endoscope inside the body to assess the interior organs or cavities. This imaging technique is commonly used for gastrointestinal (e.g. upper and lower gastrointestinal (GI) endoscopy in Figure 1.2 below) , respiratory (bronchoscopy) , and urological examinations (cystoscopy)

[1] https://screening.iarc.fr/atlasbreastdetail.php?Index=79

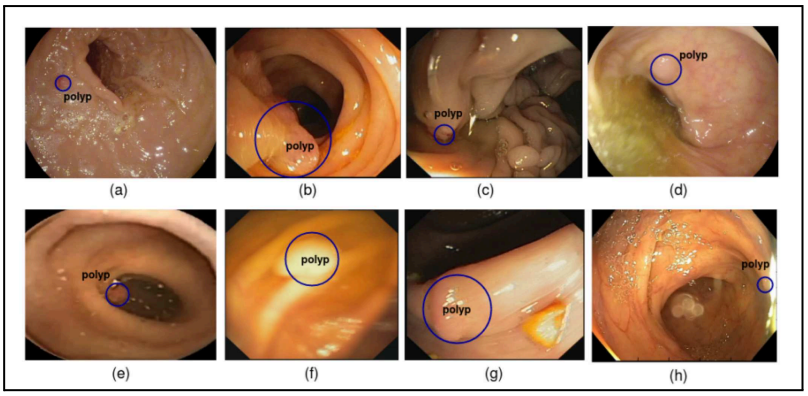


Figure 1.2 Example of polyps in GI Endoscopic images [2]

c• ***Radiologic Imaging : Mammography***

Mammography is a radiographic imaging technique that uses low-dose X-rays for breast tissue visualization (see ***Figure 1.3***). The use of mammograms enables visualizing abnormal lesions in breasts such as cancerous masses and calcifications. Thus, this medical image modality is qualified as a standard technique for the early detection of breast cancer in asymptomatic patients and it is largely used for breast cancer screening alongside breast ultrasound (Sree, 2011)

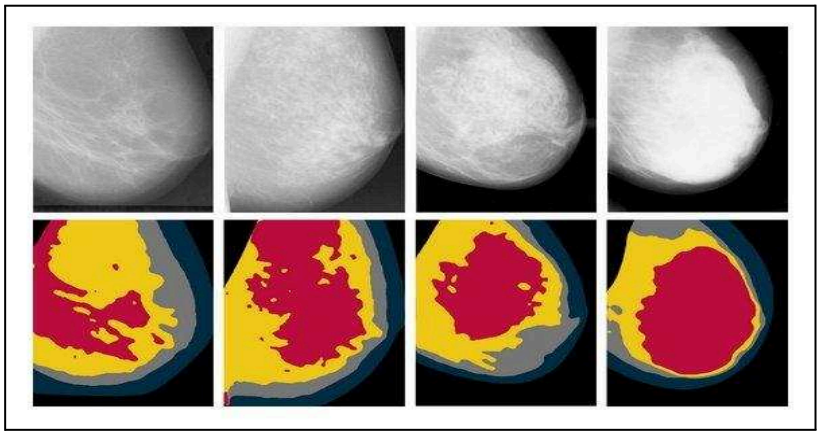


Figure 1.3 Examples of Mammography[3]

d• Magnetic Imaging : Magnetic Resonance Imaging (MRI)

MRI incorporates radio waves and the electromagnetic field to generate 3D images of human body internal anatomy (Goel et al., 2016). The application of a magnetic field to a human body results in a uniform alignment of hydrogen atoms in the direction of the magnetic field, particularly those with a positive charge. By transmitting radio frequency waves into the body, these aligned hydrogen atoms revert to their original

[2] https://www.mdpi.com/2075-4418/13/4/733
[3] https://www.researchgate.net/figure/Example-mammographic-density-segmentation-Images-digitised-Mammograms-from-left-to_fig5_279231305

positions and emit energy received by the device as signals. The intensity of these signals is employed to generate cross-sectional images plotted in grayscale. The acquisition of MRIs involves an important trade-off between the signal-to-noise ratio (SNR) and spatial resolution. A higher SNR means stronger signal allowing the production of an improved contrast and clarity images. On the other side, a higher spatial resolution means finer details that can be resolved in the image. Besides, achieving high SNR enhances the spatial resolution, since receiving a stronger signal provides more information for the MRI system to differentiate structures and details within the image. However, in practical terms, acquiring high spatial resolution MRI scans at an optimal SNR requires a longer acquisition time due to the repeated acquisition and averaging strategy employed .This conflicts with patient comfort and physiological constraints. For this reason, radiologists tend to strike a balance between SNR and spatial resolution by respecting specific clinical requirements to achieve the desired image quality. Unlike X-Rays and CTs, MRI produces high-resolution images without the use of any ionizing radiation. Nevertheless, it is less accessible and available compared to the other medical imaging techniques (Farhat et al., 2020).

1.4.2 Preprocessing

Raw medical data, whether derived from CT, MRI, or X-ray scanners, often contain noise, artifacts, and intensity variations that obscure clinically relevant information. Therefore, before any diagnostic analysis, Preprocessing in CAD systems standardizes medical images by enhancing quality and reducing noise while preserving diagnostically relevant features, which bridges the gap between acquisition hardware and analytical algorithms.

1.4.3 CAD tasks employment

a• Classification

Classification serves as the triage step in CAD, categorizing an entire image or a set of medical images (full medical exam) or lesion patches extracted from medical images into diagnostic classes such as normal / abnormal, or grading disease severity (e.g., benign vs. malignant lesions (see ***Figure 1.4***)) (Ker et al., 2018; Litjens et al., 2017; D. Shen et al., 2017) .

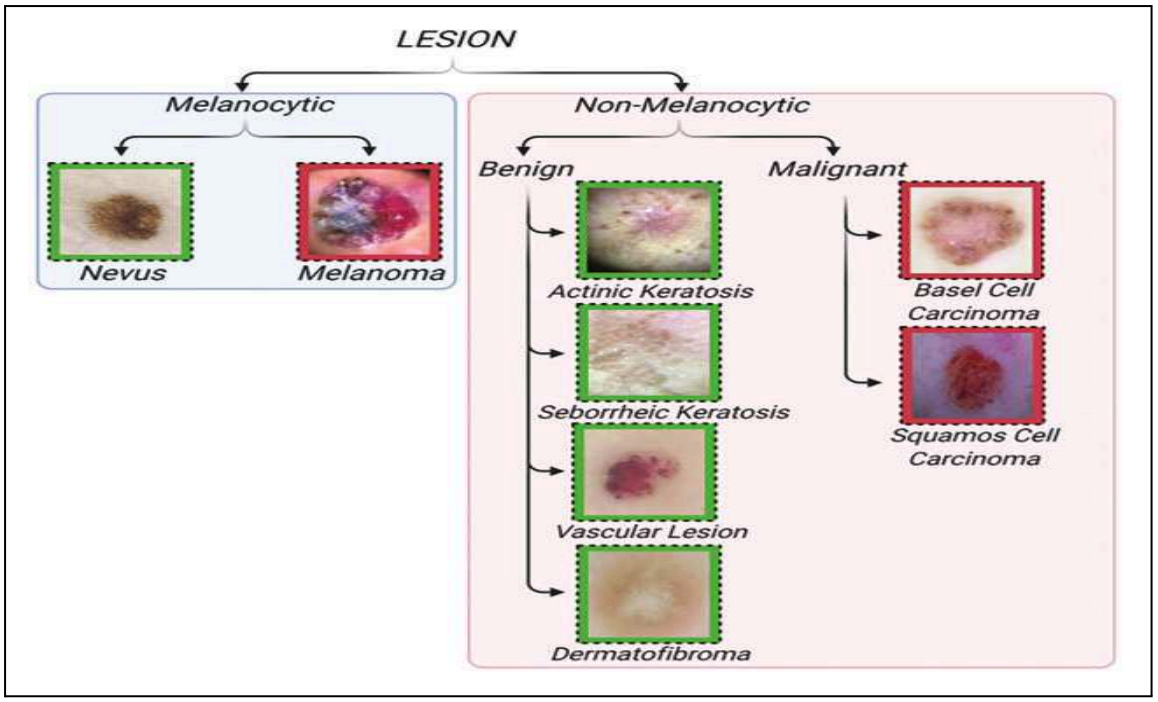


Figure 1.4 Diagram of Classification of skin lesions [4]

Traditional machine learning approaches relied on handcrafted features—texture descriptors like Haralick features, shape metrics such as sphericity, or intensity histograms—which were fed into classifiers like Support Vector Machines (SVMs) or Random Forests. However, the advent of DL has shifted paradigms, with convolutional neural networks (CNNs) now dominating classification tasks (see ***Figure 1.5*** below). CNNs, such as ResNet or DenseNet, autonomously learn hierarchical feature representations from raw pixels, outperforming manual feature engineering in complex tasks like diabetic retinopathy grading.

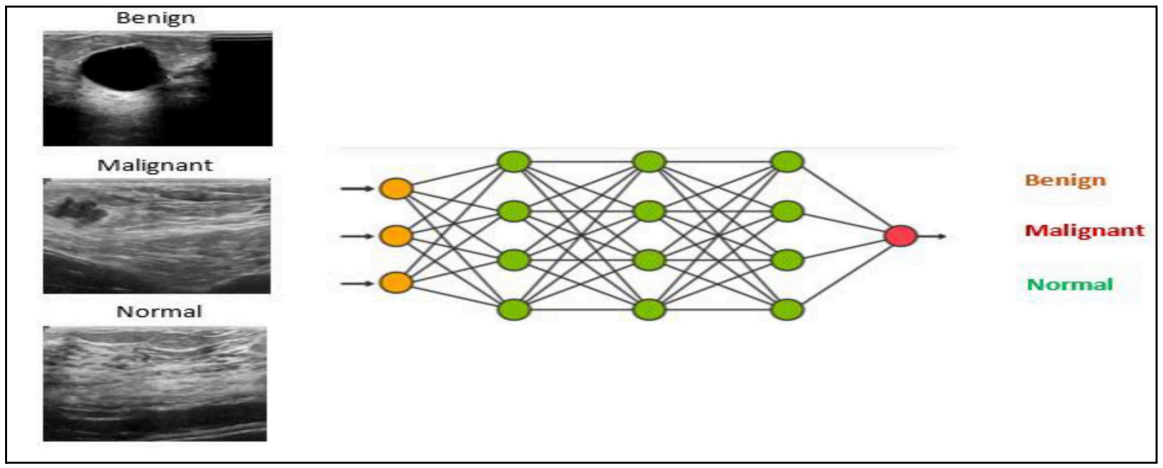


Figure 1.5 Diagram of Supervised Learning for Breast Cancer classification [5]

[4] https://www.researchgate.net/figure/Classification-of-skin-lesions-3_fig2_387354794
[5] https://www.mdpi.com/2673-2688/6/4/76

The clinical significance of classification lies in its ability to prioritize cases. For example, in screening mammography, where the majority of cases are benign, a CNN classifier can flag suspicious calcifications or masses, directing radiologists' attention to high-risk cases. This prioritization reduces workload and accelerates turnaround times, critical in emergencies like stroke or pulmonary embolism. However, classification is not infallible; subtle pathologies or rare diseases may evade detection if underrepresented in training data. Thus, modern CAD systems often integrate uncertainty estimation, providing confidence scores alongside predictions to guide clinician interpretation.

b• Detection

Computer Aided Detection (CADe) is a task that bridges the gap between image-level classification and pixel-level analysis by localizing nodules and lesions within the full medical image.This task consists of drawing labeled bounding boxes (BB) around the suspect abnormal areas within the image, as presented in ***Figure 1.6*** below. The output should be the patches of lesion candidates which can be then classified as benign or malignant for false positives reduction. The abnormalities to detect in images can be from the same class or from multiple classes (Ker et al., 2018; Litjens et al., 2017; D. Shen et al., 2017).

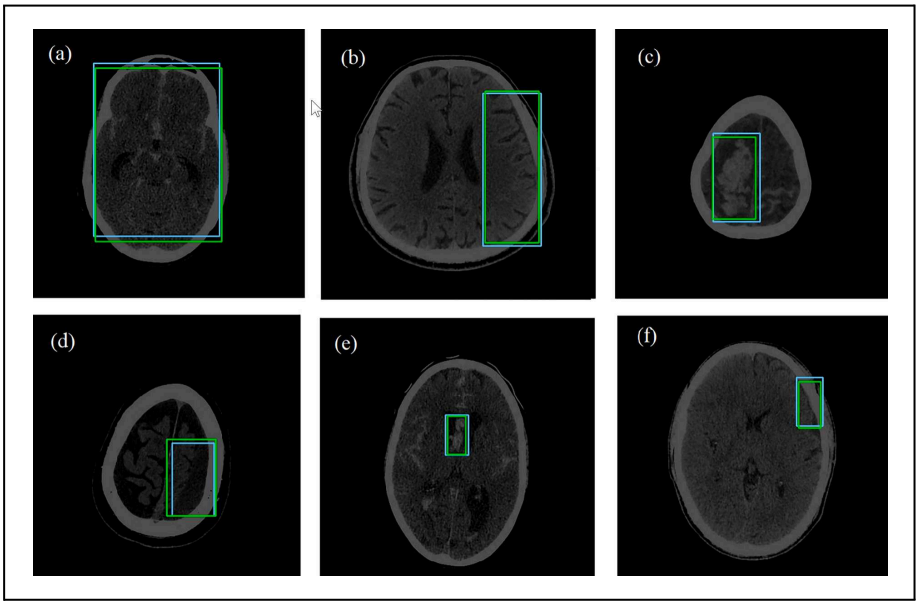


Figure 1.6 Detected hemorrhage while Green boxes are ground truth and blue are predicted boxes a) Subarachnoid, b) Subdural, c) Intraparenchymal, d) Chronic, e) Intraventricular, f) Epidural [6]

Unlike classification, which assigns a global label, detection identifies and bounds regions of interest (ROIs) such as lung nodules, brain hemorrhages, or breast lesions. This spatial localization is pivotal for quantifying disease burden—counting

[6] https://www.sciencedirect.com/science/article/abs/pii/S1746809421006820

metastases in oncology, for example—and guiding interventions. Architecturally, detection frameworks like Faster R-CNN or YOLO (You Only Look Once) employ region proposal networks (RPNs) to hypothesize candidate ROIs, followed by convolutional layers to refine bounding boxes and assign class labels.

c• Segmentation

Segmentation is the process of distinguishing target objects from the background and separating them from one another by delineating precise pixel- or voxel- (a three-dimensional counterpart to a pixel) level boundaries of the object. The prediction of target pixels relies on attributes related to pixel intensity and texture, while identifying its boundary pixels leverages the image gradient, which has high values at the object edges (Bankman, 2008).

In medical CAD, this task represents the highest level of granularity, as the target objects may include organs, substructures, or pathological lesions (Hesamian et al., 2019; Ker et al., 2018; Litjens et al., 2017). It is indispensable for quantitative analysis, enabling accurate measurements of tumor volume, organ morphology, and tissue perfusion (D. Shen et al., 2017).

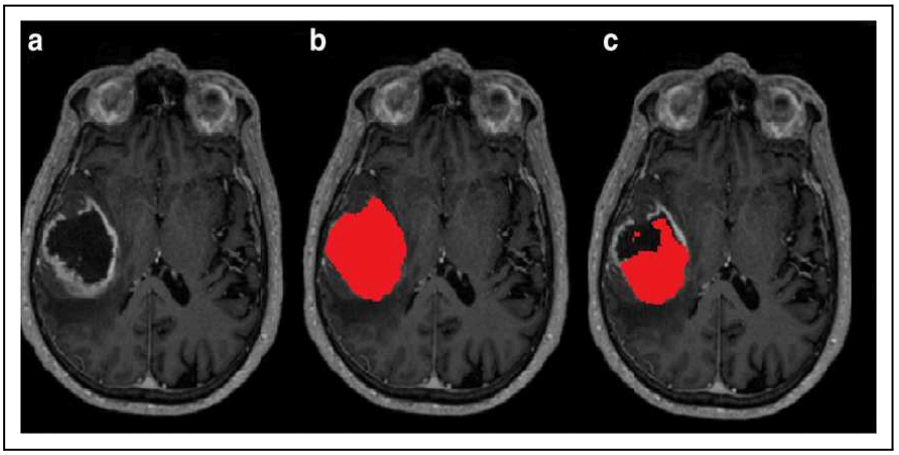


Figure 1.7 Example of glioblastoma segmentation [7]
a) Original 3D T1 post-contrast MRI; b) Segmentation (in red, 69.99 mL) on the full image;
c) Segmentation (in red, 22.87 mL) on the image processed with a volume-reduction method

In clinical practice, segmentation supports treatment planning and monitoring. For example, in radiation therapy, precise segmentation of glioblastoma boundaries in MRI ensures accurate targeting while sparing healthy tissue as above in ***Figure 1.7*** (de Sitter et al., 2020) . segmenting the liver in CT scans helps in living donor liver transplants by accurately measuring the size of the part to be donated.However, this process can still be challenging, especially when the boundaries are unclear, such as with infiltrative tumors or when using combined imaging techniques.

[7].https://www.researchgate.net/figure/An-example-of-glioblastoma-segmentation-affected-by-FFR-processing-showing-3D-T1-post_fig4_337042545

The transition from classical segmentation methods to DL has been transformative. Early techniques such as thresholding , active contours (snakes, level sets), and region growing (watershed) were limited by their inability to handle anatomical variability and imaging noise .The introduction of fully convolutional networks (FCNs) in 2015 marked a paradigm shift by enabling pixel-wise prediction from raw images (Long et al., 2015). U-Net further advanced the field by incorporating skip-connections to preserve spatial details, making it particularly effective for medical imaging tasks . Subsequent innovations, such as DeepLab's atrous convolutions (Chen et al., 2017) and transformer-based architectures (Dosovitskiy et al., 2020), have further improved multi-scale feature extraction and long-range dependency modeling.

1.4.4 Integration of tasks

A fully realized CAD system synthesizes preprocessing, classification, detection, and segmentation into a seamless workflow. For example, In liver lesion analysis, raw abdominal CT or MRI images are first preprocessed to normalize intensities, reduce artifacts, and ensure isotropic voxel spacing. An initial classification step may be used to confirm the presence and localization of the liver within the scan, especially in whole-abdomen imaging (see ***Figure 1.8 (a)*** below). Once the liver is identified, a subsequent classifier assesses whether any abnormalities are likely present. For cases flagged as abnormal, a detection module pinpoints potential lesions (see ***Figure 1.8 (b), (c)***). These regions are then segmented to outline lesion boundaries and extract quantitative features such as volume, shape, margin definition, and enhancement dynamics (see ***Figure 1.8 (d)***).

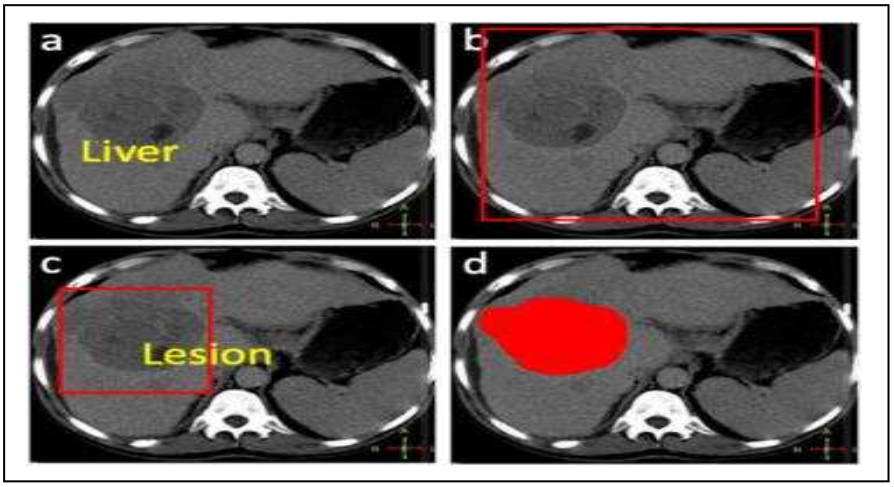


Figure 1.8 Example of Integration of Tasks [8]
a) liver image classification; b) liver localization;
c) liver lesion detection; d) liver lesion segmentation.

[8] https://www.sciencedirect.com/science/article/pii/S0933365722000963

1.5 Conclusion

This chapter provided a foundational overview of medical imaging and the structure of Computer-Aided Diagnosis systems. It explored key imaging modalities, described the sequential tasks that make up the CAD pipeline, and emphasized the growing role of DL in automating complex diagnostic tasks. From early lesion classification to pixel-level segmentation, these intelligent systems are reshaping clinical decision-making and improving diagnostic accuracy. With this groundwork laid, the following chapter delves into the existing body of research on DL applications in medical imaging, with particular focus on breast cancer diagnosis using ultrasound. This review will examine state-of-the-art models, highlight critical challenges, and identify research gaps addressed in this thesis.

CHAPTER 2

LITERATURE REVIEW - Deep Learning Applications for Medical Images and Breast Cancer Diagnosis using US

2 Chapter 2 : LITERATURE REVIEW - Deep Learning Applications for Breast Lesion Segmentation in Ultrasound

2.1 Introduction

The field of medical image analysis has undergone a revolutionary transformation with the advent of deep learning (DL), surpassing traditional computer-aided detection (CAD) systems in accuracy, efficiency, and adaptability. Early CAD frameworks relied on handcrafted features and rule-based algorithms, which were limited in their ability to generalize across diverse imaging modalities and clinical scenarios. The emergence of convolutional neural networks (CNNs) and encoder-decoder architectures, such as U-Net, has enabled end-to-end learning directly from raw imaging data, significantly improving segmentation performance in medical applications (Litjens et al., 2017).

This chapter provides a systematic review of deep learning techniques in medical image segmentation, covering their evolution, architectural innovations, training methodologies, evaluation metrics, challenges, and emerging solutions. The discussion is structured to highlight key advancements in breast segmentation, while also addressing critical limitations such as data scarcity, domain shifts, and model interpretability. Recent studies demonstrate that DL models achieve near-radiologist performance in tasks like tumor delineation, polyp detection, and organ segmentation (Ronneberger, Fischer, & Brox, 2015; Chen et al., 2017). However, challenges remain in translating these advancements into clinical practice, necessitating further research in robustness, efficiency, and regulatory compliance (Isensee et al., 2021).

2.2 Significance of medical imaging for Breast lesion diagnosis

Breast cancer remains a global public health challenge, representing the most commonly diagnosed malignancy in women and a leading cause of cancer-related mortality worldwide (Yeasmin et al., 2024). Early diagnosis is critical to improving survival rates, minimizing treatment invasiveness, and lowering healthcare costs. In this context, Computer-Aided Diagnosis (CAD) systems have become central to breast imaging workflows. They serve as reliable assistive tools that reduce oversight, enhance sensitivity, and provide consistent diagnostic interpretation across varying imaging modalities and clinician experience levels.

Breast imaging typically involves mammography, ultrasound (US), and, in high-risk or ambiguous cases, magnetic resonance imaging (MRI). Each modality offers distinct advantages but also unique challenges — such as decreased mammographic sensitivity in dense breast tissue or operator dependency in ultrasound interpretation. CAD systems mitigate these limitations by detecting and localizing

suspicious regions of interest (ROIs), classifying lesion types, and supporting radiologists in differential diagnosis (McKinney et al., 2020).

In mammography, which remains the primary imaging modality for large-scale screening programs, CAD was first introduced as a tool for highlighting microcalcifications and mass densities. Traditional rule-based methods offered marginal benefits, but the advent of deep learning (DL), particularly convolutional neural networks (CNNs), has significantly elevated diagnostic performance. Notably, McKinney et al. (2020) demonstrated that a DL model trained on over 76,000 images could outperform experienced radiologists in detecting breast cancer, with improved sensitivity and fewer false positives. Similarly, Rodriguez-Ruiz et al. (2019) showed that DL-based CAD matched the diagnostic accuracy of an average radiologist in a multi-reader setting, identifying 98.3% of invasive cancers in the top decile of predicted malignancy risk.

Ultrasound, commonly used as a supplemental modality, is especially important in women with dense breast tissue or when mammography yields inconclusive results. However, BUS is inherently noisy due to speckle artifacts, and its diagnostic accuracy heavily depends on the skill of the operator. CAD systems in ultrasound imaging use segmentation and classification algorithms to overcome these issues, offering objective analysis and increased reproducibility. U-Net-based architectures, for instance, have shown excellent performance in segmenting tumor margins and differentiating between benign and malignant lesions in BUS images (Han et al., 2017; Shen et al., 2019).

Breast MRI, though highly sensitive, generates large volumetric datasets that are time-intensive to interpret. Here, CAD systems contribute by performing lesion detection, kinetic curve analysis, and tumor volume estimation. Recent transformer-based models and 3D CNNs are particularly promising for enhancing MRI-based screening accuracy in high-risk populations (Jannatdoust et al., 2025).

Across all modalities, CAD systems not only enhance diagnostic accuracy but also improve clinical efficiency. They reduce inter-reader variability, especially in double-reading environments, support risk stratification, and allow for more standardized and objective interpretation in population-wide screening settings (Guo et al., 2022; Yeasmin et al., 2024). Moreover, as datasets grow and models improve, the role of CAD in decision support, real-time triage, and workflow optimization continues to expand.

2.3 Key Deep Learning Approaches for breast lesion segmentation in Ultrasound

Deep learning has redefined CAD by integrating segmentation into diagnosis frameworks, eliminating the dependency on manually engineered features. CNNs excel at learning hierarchical representations from medical images, enabling precise detection of pathologies and accurate delineation of anatomical structures.

Accurate segmentation of breast masses in ultrasound (US) images is a critical prerequisite for computer-aided diagnosis (CAD) systems, enabling precise lesion characterization, size measurement, and malignancy assessment. Recent advances in deep learning have led to three dominant architectural paradigms for this task: U-Net-based models, transformer-based approaches, and hybrid architectures. Each addresses unique challenges in breast US segmentation, including speckle noise, low contrast, shadow artifacts, and variable lesion morphology. Most relevant works on breast ultrasound medical images segmentation are organized in Table 2 below.

2.3.1 U-Net & Variants

The U-Net architecture, introduced by Ronneberger et al. (2015), remains a foundational model in medical image segmentation due to its symmetric encoder–decoder structure and skip connections (see ***Figure 2.1***). These allow for the fusion of contextual and spatial information, making U-Net particularly effective for delineating breast lesions, which often exhibit low contrast and irregular boundaries in ultrasound (US) images.

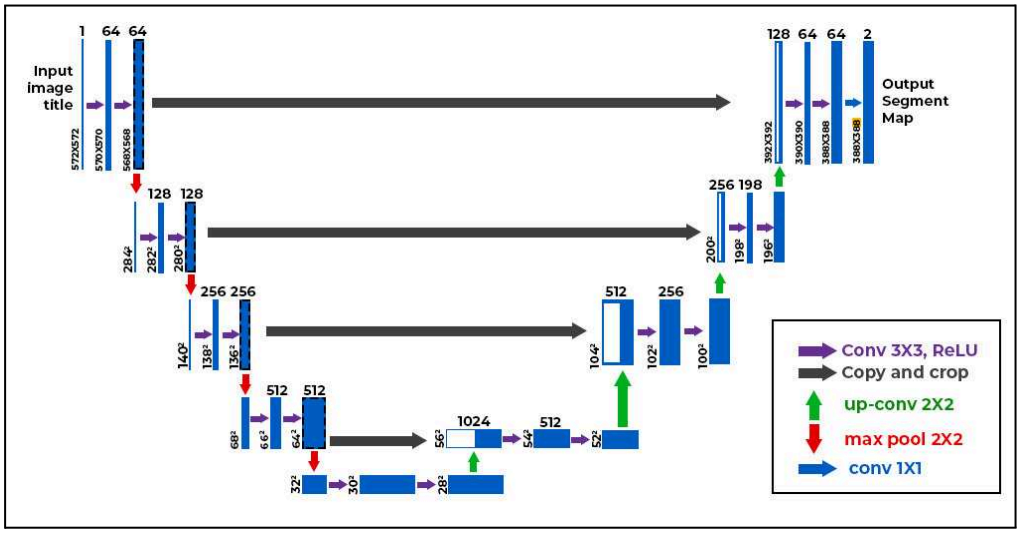


Figure 2.1 U-Net architecture [9]

Recent works continue to refine U-Net to improve its segmentation performance. For instance, Almajalid et al. (2018) trained a baseline U-Net with rotation and elastic deformation for data augmentation on a dataset of 221 BUS

[9] https://www.geeksforgeeks.org/u-net-architecture-explained/

images, achieving a Dice coefficient of 0.825. The relatively small input size (128×128) and classic U-Net architecture demonstrate strong performance in resource-constrained settings.

Table 2 Relevant works on breast ultrasound medical images segmentation

Num	Work	Approach	Hyper--parameters	Dataset	Image size	Results
1	Almajalid et al., 2018	A U-net model trained and tested through two-fold cross-validation.data augmentation: rotation and elastic deformation	Minus dice loss	dataset contains 221 BUS images	resized to 128 x 128 pixels	dice coefficient = 0.825,
2	Baccouche et al., 2021	Combines two U-Nets with modified skip connections Integrates ASPP	LR =0.0001, A weighted Loss = -(0.4 * dice score (true, predicted) + 0.6 * IoU (true, predicted)),	CBIS-DDSM, INbreast datasets	256 × 256 pixels (obtained ROIs)	Dice coefficient = 0.8, 0.91
3	Pramanik et al., 2024	Combined Positional Convolutional Block Attention Module (PCBAM) with SWA (Shifted Window Attention)	Loss = Dice loss + BCE loss + focal loss	BUSI and UDIAT datasets	500 × 500 pixels resized to 128 × 128 pixels	Dice coefficient = 0.74, 0.78
4	Zhang et al., 2024	DDTransUNet, a hybrid network combining Transformer and CNN, with a dual-branch encoder and dual attention mechanism	LR =0.00001 Loss = 0.5 BCE loss + dice loss	TN3K, BUS-BRA and CAMUS, datasets	256 × 256 pixels	Dice coefficient = 0.82
5	Zhang et al., 2025	combining U-Net with other deep neural networks (Res-Net and MultiResUNet) and introducing a new approach and block (Co-Block),	LR from 0.0001 to 0.00001 Loss = (BCE + Dice) /2	BUSI, UDIAT, and BLUI datasets	224 × 224 pixels	Dice coefficient = 0.82
6	Derakhshandeh et al., 2025	combining U-Net with other deep neural networks (Res-Net and MultiResUNet) and introducing a new approach and block (Co-Block)	Dice loss	BUSI of 780 images	256 × 256 pixels	Dice coefficient = 0.82

Attention U-Net

(Oktay et al., 2018) further enhances performance by incorporating attention gates that dynamically weight feature maps, improving segmentation accuracy for subtle lesions.

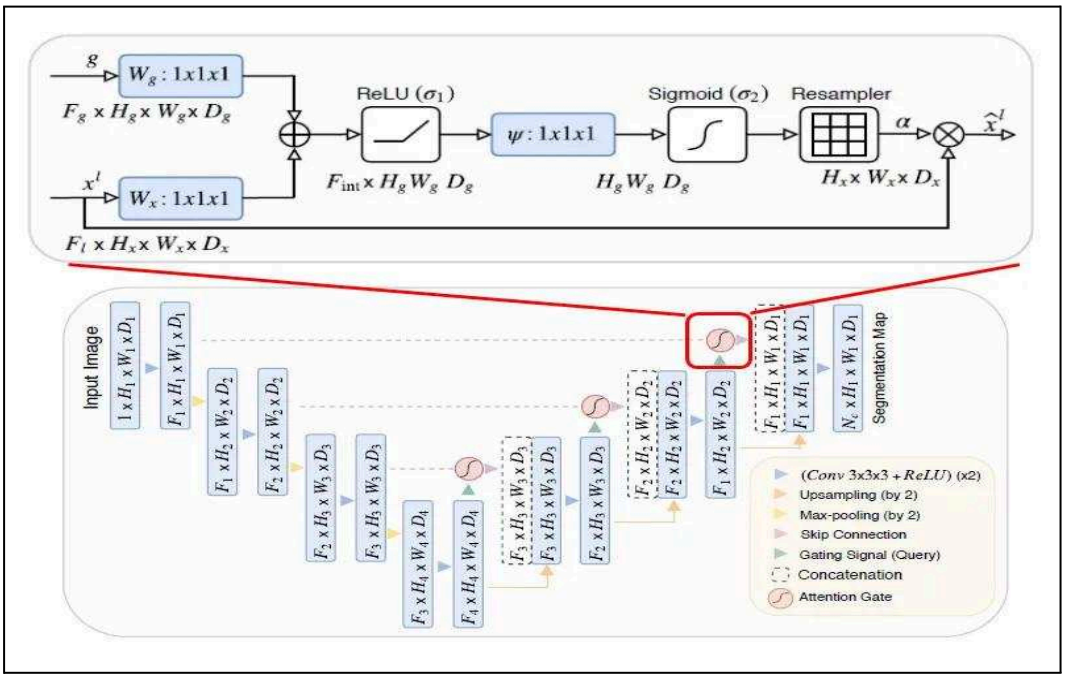


Figure 2.2 Attention U-Net architecture [10]

ResU-Net

ResU-Net integrates residual blocks to facilitate gradient flow in deeper networks, addressing vanishing gradient issues and improving training stability (Zhou et al., 2018). The model architecture is represented in the ***Figure 2.3*** below.

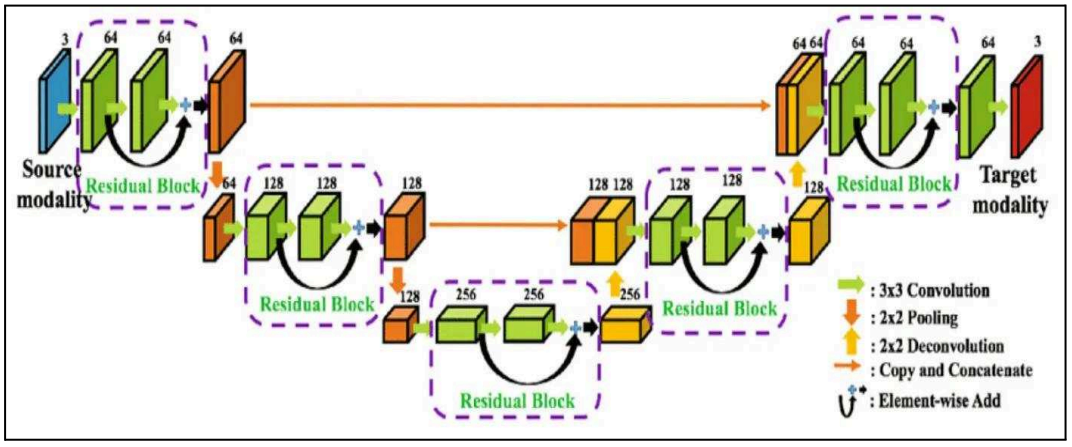


Figure 2.3 Attention U-Net architecture [11]

[10] https://code.likeagirl.io/u-net-vs-residual-u-net-vs-attention-u-net-vs-attention-residual-u-net-899b57c5698
[11] https://link.springer.com/chapter/10.1007/978-3-030-00889-5_18

More recent studies combine U-Net with enhancements. For example, (Baccouche et al., 2021) implemented a dual U-Net structure with Atrous Spatial Pyramid Pooling (ASPP) and a weighted loss function, reporting Dice scores up to 0.91 on CBIS-DDSM and INbreast datasets. These findings underscore the ongoing relevance and adaptability of U-Net-based architectures for BUS segmentation (see Table 2).

2.3.2 Transformer-Based Architectures

Vision transformers (ViTs) have gained prominence in medical segmentation, as presented in ***Figure 2.4*** below due to their ability to model long-range dependencies, overcoming the local-receptive-field limitation of CNNs by adapting the self-attention mechanisms of transformers that are originally designed for natural language processing (NLP) to visual data. This is especially valuable in breast ultrasound segmentation, where lesions vary widely in size and texture, and local receptive fields of CNNs may not fully capture global context. Below come some key innovations and clinical impacts.

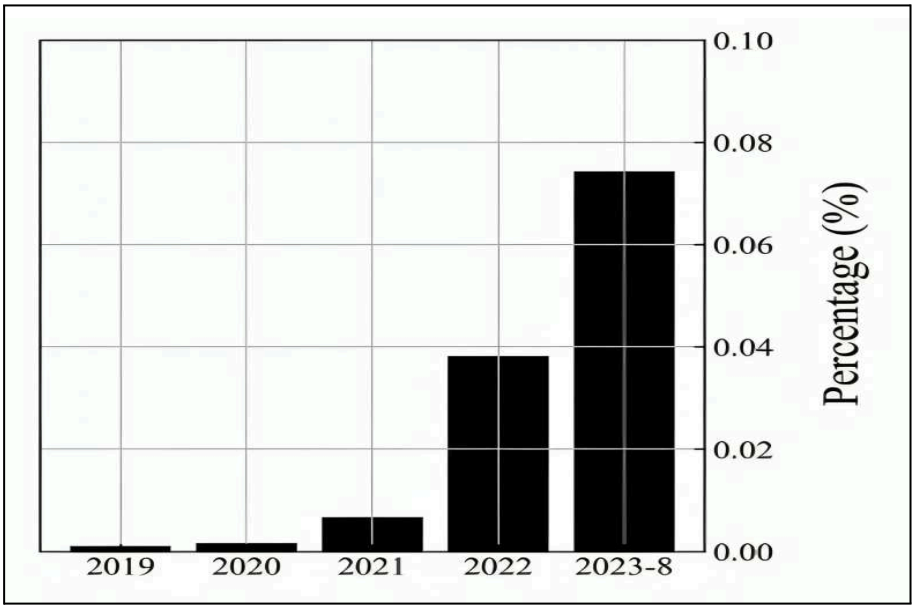


Figure 2.4 The proportion of literature related to transformers in medical image segmentation literature (Pu et al., 2024)

In Pramanik et al. (2024) (see ***Table 2*** **(3)** above), a transformer-based approach using Shifted Window Attention (SWA) was combined with a convolutional attention module (PCBAM). The model was trained on BUSI and UDIAT datasets and achieved Dice scores of 0.74 and 0.78, respectively. These results demonstrate the capability of transformers to model non-local context, although training data requirements and computational complexity remain key considerations.

TransUNet

(Chen et al., 2021) combines CNN and transformer features, leveraging the strengths of both architectures for improved segmentation performance.

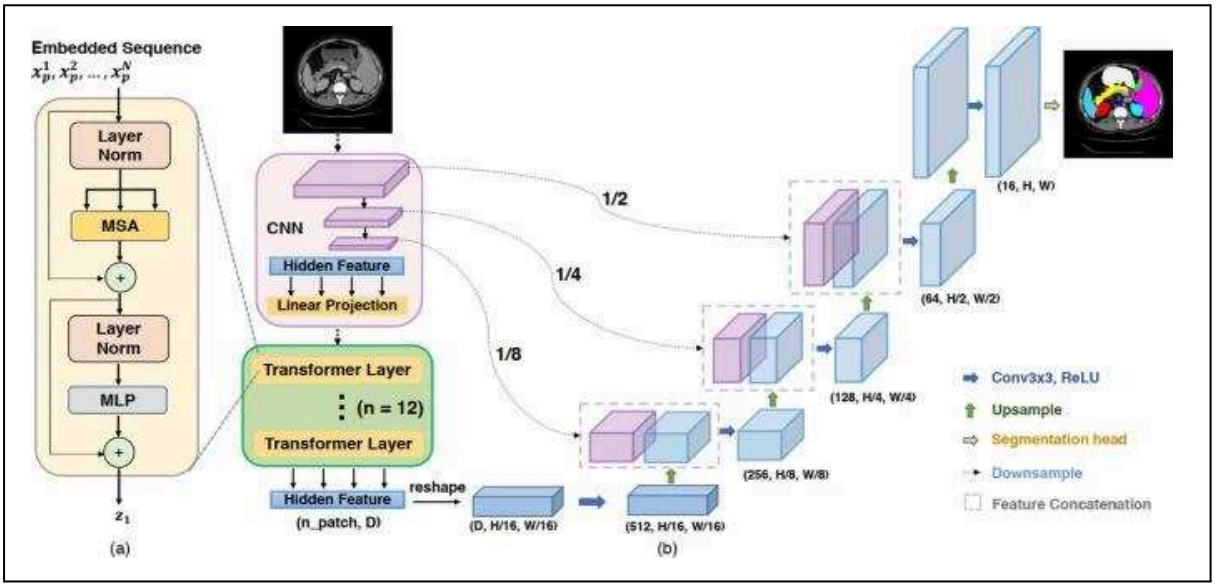


Figure 2.5 TransUnet applied transformer structure to medical image segmentation (Chen et al., 2021)

An example is Zhang et al. (2024), who proposed DDTransUNet, a hybrid transformer-CNN model featuring dual attention branches and a dual-encoder design. The model, evaluated on BUS-BRA and TN3K datasets, achieved a Dice coefficient of 0.82. This reinforces the growing role of transformers in medical image segmentation, particularly when combined with CNN components for feature localization (see **Table 2 (4)** above).

2.3.3 Hybrid Models

Hybrid architectures aim to leverage the strengths of both CNNs and transformers by combining local feature extraction with global attention. These models often integrate modules like SE blocks, CBAM, or ViTs within U-Net-based backbones.

A notable example is Zhang et al. (2025), who integrated U-Net with ResNet and MultiResUNet, introducing a new "Co-Block" to enhance contextual understanding. Trained on BUSI, UDIAT, and BLUI datasets, their model achieved a Dice score of 0.82, matching state-of-the-art results while improving generalization through hybrid feature fusion.

Similarly, Derakhshandeh et al. (2025) adopted a comparable architecture and validated it on a larger BUSI dataset, also achieving a Dice coefficient of 0.82. These consistent outcomes underscore the effectiveness of combining residual learning and multiresolution features within hybrid frameworks.

Hybrid approaches like these serve as a promising middle ground between CNN efficiency and transformer expressiveness, making them suitable for BUS segmentation where both fine-grained detail and broader context are essential (see Table 2).

2.4 Training and Optimization of Deep Segmentation Models : Key Hyperparameter Choices for Building Effective Segmentors

2.4.1 Loss Functions and Optimization Strategies

The choice of loss function significantly affects segmentation accuracy, especially in the presence of class imbalance and small lesion regions. Dice loss is frequently adopted due to its sensitivity to overlap, as seen in Almajalid et al. (2018) and Derakhshandeh et al. (2025), both of whom used it to effectively handle foreground-background imbalance in BUS datasets.

More advanced strategies combine multiple loss functions. For example, Pramanik et al. (2024) used a composite loss function (Dice + BCE + focal loss) to improve robustness against class imbalance and misclassified boundary pixels. Similarly, Zhang et al. (2024) used a weighted combination of BCE and Dice loss, while Baccouche et al. (2021) designed a custom loss combining Dice and IoU, further tailoring the learning process to the segmentation task.

In terms of optimization, learning rates in the range of 1e-4 to 1e-5 were commonly used across studies (e.g., Zhang et al., 2024), often with Adam optimizers and learning rate schedulers to stabilize training and improve convergence.

2.4.2 Regularization and Data Augmentation

Given the limited availability of labeled medical data, regularization and augmentation are essential. Most studies in Table 2 applied data augmentation such as rotation, flipping, and elastic deformations. For instance, Almajalid et al. (2018) utilized rotation and elastic transformation, while Pramanik et al. (2024) applied image resizing and normalization on datasets like BUSI and UDIAT.

Regularization methods such as dropout and batch normalization are also widely employed to prevent overfitting, especially in deep U-Net variants and hybrid models. These techniques ensure better generalization, particularly when combined with diverse training datasets.

2.5 Performance Evaluation Metrics

Evaluation of semantic segmentation is the accurate quantification of segmentation performance (Müller et al., 2022), which can be quite complex because it requires multiple complementary metrics to assess different aspects of model behavior, since it is required to measure classification accuracy as well as localization correctness. The fundamental metrics include True Positives (TP, correctly identified lesion pixels), False Positives (FP, background pixels misclassified as lesions), False Negatives (FN, lesion pixels incorrectly marked as background), and True Negatives (TN, correctly identified background pixels) as the visualized in the part ***(a)*** of the ***Figure 2.6*** below. For breast lesion segmentation in ultrasound (US) images, six key metrics are universally employed.

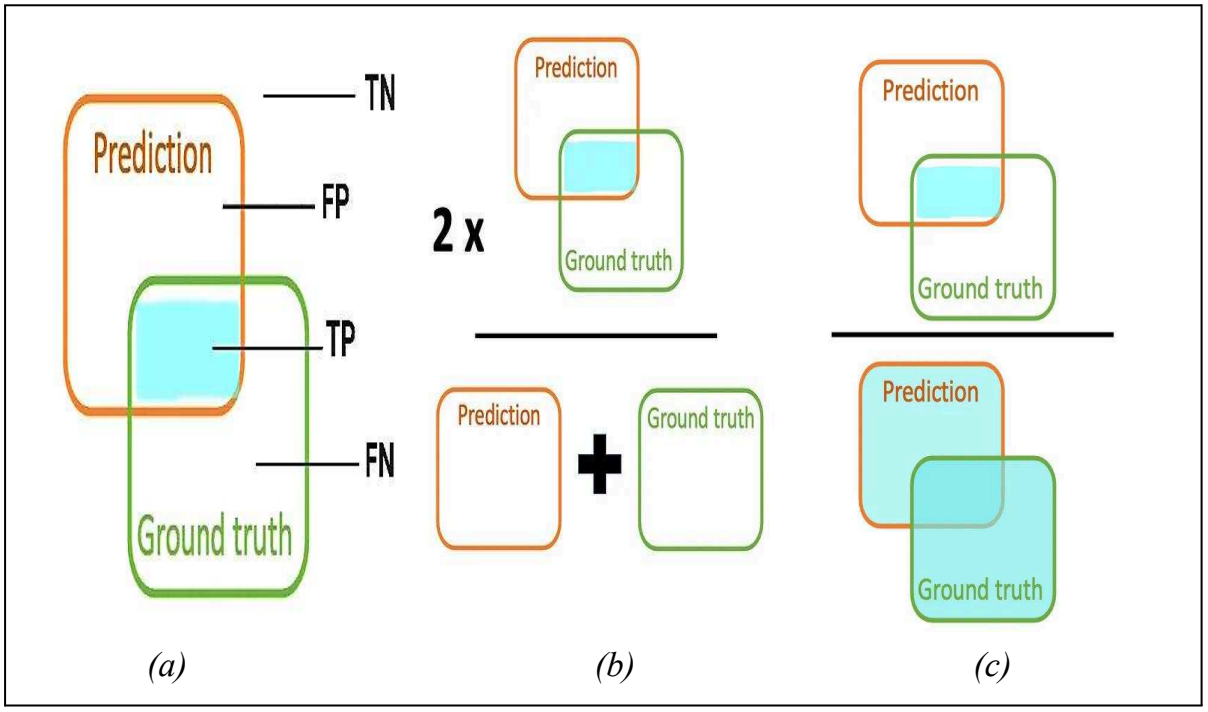


Figure 2.6 Segmentation Metrics (Huynh et al.2023) : (a) Error Type Visualization, (b) Dice coefficient, (c) Jaccard index

2.5.1 Metrics Focusing on Classification Accuracy

a.Accuracy

Known as the Rand index, it measures the proportion of correctly classified pixels (both foreground and background) out of all pixels : Accuracy = $\frac{TP + TN}{TP+TN+FP+FN}$

However, in medical imaging, the region of interest (lesion or tumor) is often much smaller than the background, making accuracy misleading if the model ignores the target class, Which makes accuracy not the best metric in our case and, and two metrics that aim to solve the issues with accuracy are precision and recall.

b.Precision

Measures the proportion of correct lesion predictions to the total number of lesion predictions, by counting the correctly classified foreground pixels to the total number of foreground predicted pixels : Precision= $\frac{TP}{TP+FP}$

This metric penalizes false positives (FP), which is critical in medical contexts (e.g., avoiding false lesion segmentations). However, it doesn't account for false negatives (FN), so high precision alone doesn't guarantee all lesions are detected.

c.Recall

Also known as Sensitivity, measures the proportion of correct lesion predictions relative to all actual lesions present in the medical image, by counting the correctly classified foreground pixels to the total number of pixels that should have been identified as foreground : Recall = $\frac{TP}{TP+FN}$

This metric penalizes false negatives (FN), which is critical in diagnostic contexts (e.g., avoiding missed tumors in breast ultrasound). However, it doesn't account for false positives (FP), so high recall alone doesn't guarantee all detected lesions are real.

d.Specificity

Measures the proportion of healthy areas correctly identified as normal tissue relative to all truly healthy areas in the medical image, which are the correctly classified background pixels to the pixels actually belonging to the background of the lesion.

Specificity = $\frac{TN}{TN+FP}$

In medical imaging, it ensures the model doesn't misclassify healthy tissue as pathological (critical to avoid unnecessary interventions). However, like accuracy, it's less informative when the background dominates.

The aforementioned metrics (accuracy, precision, recall, specificity) primarily assess classification accuracy at the pixel level (Taha et Hanbury, 2015) by checking if each pixel is correctly labeled as lesion/background, ensuring no lesion pixels are missed (recall) or healthy areas mislabeled (precision), However, medical image segmentation further requires evaluation of localization correctness, where spatial agreement between predicted and reference segmentations becomes critical (Reinke et al., 2021), which checks if the shape/boundaries of the segmented lesion match the ground truth ensuring the segmented tumor's size and shape match reality (critical for biopsy/surgery planning). This necessitates complementary metrics, which will be discussed next.

2.5.2 Metrics Focusing on Localization Correctness

a.Dice coefficient

The Dice similarity coefficient (DSC) measures volumetric overlap between predicted and ground truth segmentations :

$$\text{DSC} = (2 \times \text{TP}) / (2 \times \text{TP} + \text{FN} + \text{FP}) = \frac{2\ X\ Area\ of\ overlap}{Total\ area}$$

DSC balances FP and FN, making it robust for imbalanced datasets (common in medical imaging). It's widely used but tends to favor volume similarity over precise boundaries.

b.Jaccard index

Known as Intersection-over-Union (IoU) or Jaccard similarity coefficient. The difference between the two metrics is that the IoU penalizes under- and over-segmentation more than DSC. it is the area of the intersection over union of the predicted segmentation and the ground truth :

$$\text{IoU/Jaccard Index} = (\text{TP}) / (\text{FP} + \text{TP} + \text{FN}) = \frac{Area\ of\ overlap}{Area\ of\ union}$$

2.6 Conclusion

This chapter reviewed the current landscape of deep learning approaches for breast lesion segmentation in ultrasound, highlighting advances in U-Net variants, transformer-based models, and hybrid architectures. While these methods have improved segmentation accuracy, challenges such as data imbalance, noise, and generalizability persist. In addition, such architectures can be impractical to reproduce in resource-limited environments, highlighting a gap between academic performance and real-world feasibility. Key strategies in loss design, regularization, and evaluation were discussed to illustrate what makes an effective segmentor. These insights directly inform the design of the proposed method, presented in the next chapter, which aims to address the limitations identified in the existing literature.

CHAPTER 3

PROPOSED APPROACH

3 Chapter 3 : PROPOSED APPROACH

3.1 Introduction

This chapter outlines the methodological foundation of the proposed solution to breast ultrasound segmentation challenges. It begins by identifying key limitations in existing approaches—such as limited annotated data, class imbalance, and imaging variability—and then introduces a targeted deep learning approach designed to enhance segmentation accuracy and robustness. The proposed method integrates multiple architectural strategies, including multiscale feature learning and attention mechanisms, along with tailored preprocessing and augmentation steps. Together, these components form a unified pipeline aimed at improving performance on difficult cases while maintaining adaptability to real-world clinical variability.

3.2 Problem statement

In medical image segmentation, annotations specifically refer to pixel-wise labeling where each pixel in an ultrasound image is classified as either a lesion tissue (foreground) or a normal tissue (background). Obtaining high-quality annotations for large amounts of data is crucial for building effective DL models, since using limited data for training DL networks can result in overfitting and convergence issues (Hesamian et al., 2019), yet remains particularly challenging for breast ultrasound due to three key factors : Scarcity, Imbalance and inherent Complexity and of BUS data combined with the difficulty of the segmentation task. These fundamental limitations constrain both model performance and clinical applicability of automated segmentation systems, and are discussed in more detail in the following Subsections.

3.2.1 Limited Annotated Data

The scarcity of high-quality, pixel-level annotated datasets remains a critical bottleneck. Unlike natural images, medical annotations require expert radiologists, making large-scale datasets rarely available clinically, time-consuming and expensive. Furthermore, the variability in lesion appearance across patients – including differences in size, shape, texture, and contrast – creates a need for large, diverse datasets, however, public datasets like BUSI typically contain fewer than 1,000 images, limiting model generalizability. Inter-annotator variability (human annotation disagreements) further compounds this issue, as lesion boundaries are often ambiguous even among specialists. Geographic and demographic biases in datasets also reduce clinical applicability, with most collections originating from single institutions using specific imaging protocols.

3.2.2 Class Imbalance

In medical image segmentation, particularly for breast ultrasound (BUS), class imbalance manifests at two levels :

Pixel-level class imbalance, where the regions of interest (ROIs), such as tumors or lesions, often occupy a minuscule fraction (<1–5%) of the total image pixels (see ***Figure 3.1*** below), while background (healthy tissue) dominates spatially. Without correction, the model may achieve high pixel-wise accuracy simply by predicting the majority class, failing in its primary task of delineating pathological or anatomical structures.

Sample-level class imbalance (Bias in Dataset composition) occurs when datasets are still imbalanced per image even if pixel-level balancing is applied within each image. For instance, "normal" scans being far more than "abnormal" ones can induce model bias—e.g., a network trained mostly on healthy cases may struggle to segment rare pathologies, because of the feature suppression as dominant normal pixels drown out gradients from rare pathological features during backpropagation, leading to under-segmentation (ignoring or fragmenting small lesions) and the loss of texture sensitivity where models may overlook subtle malignant features (e.g., irregular margins and textures) and overfit to features prevalent in the majority class A predominance of benign lesions can similarly skew predictions, reducing sensitivity to malignant characteristics.

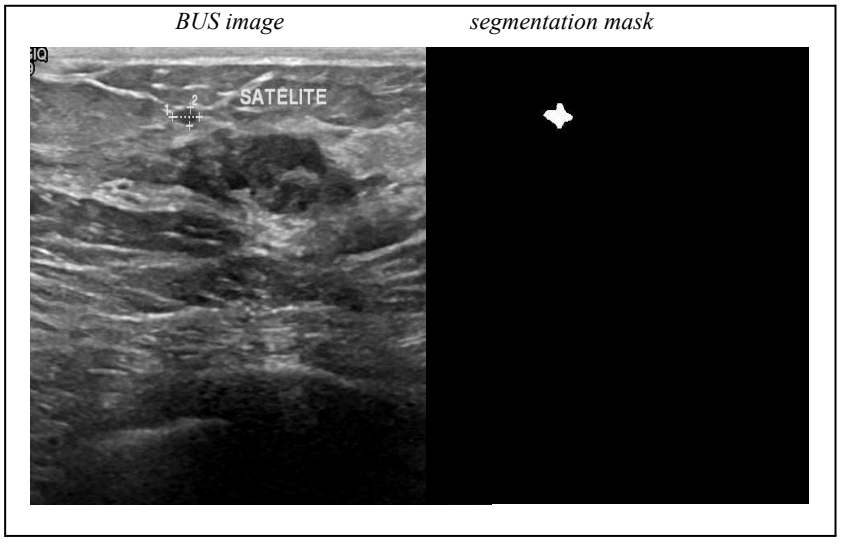


Figure 3.1 Sample 53 of BUS dataset showing a very small lesion

In segmentation, pixel-level imbalance is the primary concern, while image-level balance (malignant/benign ratio) is less critical than in classification. However, both types of imbalance should still be addressed with domain-aware strategies (e.g., robust loss functions, augmentation, and careful evaluation). Always prioritize metrics like Dice score or sensitivity for small lesions over raw accuracy.

3.2.3 Complex Task and Noisy Images

Breast lesion segmentation in ultrasound imaging presents significant challenges due to inherent complexities including low contrast between lesions and surrounding tissues. **Biological variability** between malignant (spiculated, heterogeneous texture) and benign (smooth, homogeneous) lesions, which often causes confusion in boundary delineation, while texture variations demand lesion-specific feature extraction. **Acquisition variability** across devices - including brightness, contrast, and probe frequency differences - further reduces model generalizability. These challenges are compounded by inherent ultrasound artifacts : **speckle noise** mimicking lesion textures, **shadowing** obscuring true boundaries , and **reverberation** introducing spurious edges (see ***Figure 3.2*** below) . These factors collectively degrade segmentation performance, manifesting drops in Dice scores, and elevated false positive rates , particularly for small or infiltrative lesions. Such limitations underscore the need for advanced architectures capable of handling these diagnostic ambiguities while maintaining clinical reliability.

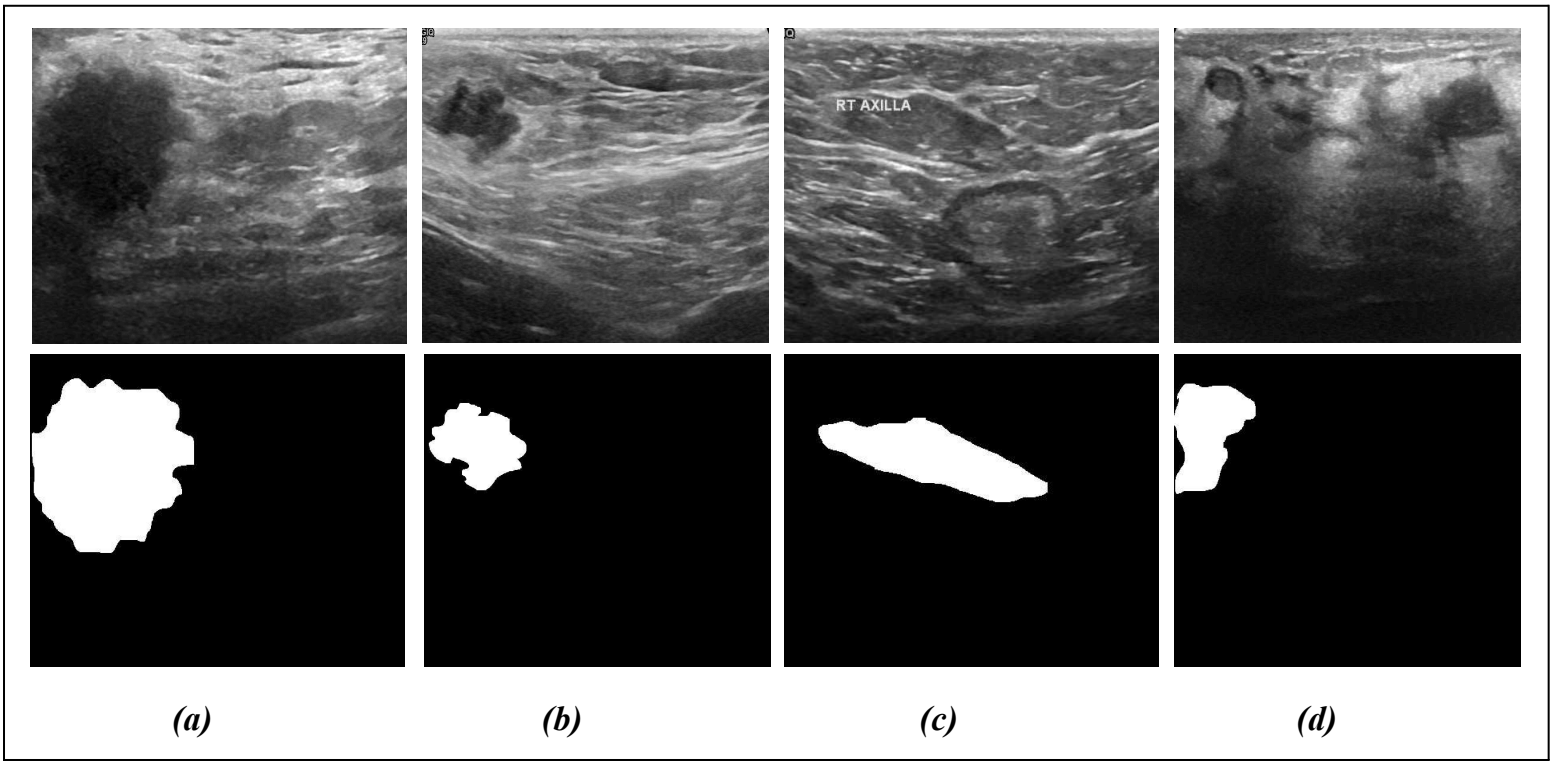


Figure 3.2 Samples from BUS dataset showing : (a) and (b) : noisy images, (c) and (d) : hard lesions

3.3 Motivation

This study aims to address the persistent challenges in breast ultrasound (BUS) lesion segmentation by proposing a novel hybrid architecture combining convolutional and transformer-based approaches with three key objectives: Achieving precise boundary delineation across diverse lesion types (malignant/benign) and sizes, Maintaining robustness against BUS-specific noise and acquisition artifacts, and Generalizing across imaging devices and demographic groups. Preliminary experiments on the Figshare 2024 brain tumor MRI dataset have validated the effectiveness of the proposed hybrid design in handling complex and heterogeneous medical imaging data, achieving Dice scores above 99% for various tumor sub-regions. These results indicate strong potential for adapting the architecture to the BUS and brain segmentation domain.

The proposed approach builds upon recent advancements in medical image analysis while introducing novel adaptations specifically tailored for breast ultrasound segmentation. By systematically evaluating different architectural components and their combinations, we seek to establish effective practices for BUS segmentation and contribute to the broader field of medical image analysis. These objectives are achieved through a designed workflow incorporating :

3.3.1 Preprocessing

Standardized resizing and CLAHE enhance image quality while reducing noise, which reduces inter-observer variability in ground truth masks.

3.3.2 Data augmentation and regularization techniques

Grid distortion and elastic transforms address data scarcity and class imbalance to prevent overfitting. It Mitigates limited training data from small BUS datasets, improves generalization to rare lesion morphologies and simulates real-world imaging variations (probe pressure, etc.).

Several regularization strategies are employed throughout the training process as well, Dropout is employed during training to randomly deactivate a fraction of neurons, encouraging the model to learn redundant and more generalizable feature representations, particularly valuable when working with small BUS datasets. Additionally, batch normalization is integrated to stabilize the training process by normalizing feature distributions across mini-batches, which improves convergence and helps the model remain resilient to the variability present in BUS images.

3.3.3 Dilated Convolutions

Atrous Spatial Pyramid Pooling module (ASPP) for multiscale feature extraction. Unlike regular convolutions using fixed receptive fields to extract image-level features from images for classification tasks, dilated convolutions (Yu & Koltun, 2015) utilize

adaptable receptive fields to enable the extraction of local and contextual information from images, highly significant in the context of object segmentation. Convolutions employing dilation also known as atrous convolutions enlarge the receptive field of the network by inserting additional rows and lines of zeros to the convolution filter to expand it spatially (see ***Figure 3.3*** below). This promotes adapting the segmentation network to multiple scales when different dilation rates are used.

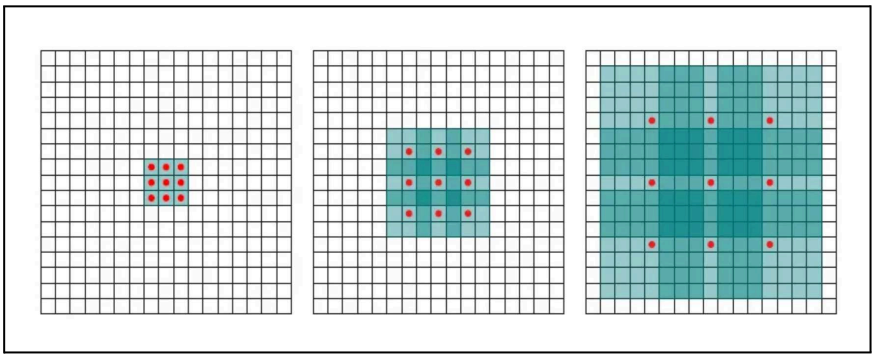


Figure 3.3 Schematic of dilated convolution (Yu & Koltun, 2015)

3.3.4 Multiscale Feature Learning

Multiscale feature learning is achieved by integrating ASPP and transformer blocks, enabling the network to capture both local and global contextual features of breast lesions.

3.3.5 Attention Mechanism

Attention modules, such as Vision Transformer (ViT) blocks, Squeeze-and-Excitation (SE) blocks, and attention gates, are incorporated to enhance feature selectivity by filtering and suppressing irrelevant features or noisy activations, improving the focus on lesion-relevant regions.

3.3.6 Skip Connections

Skip connections are employed between the encoder and decoder stages to preserve spatial resolution and facilitate the integration of low-level contextual information into the reconstruction of the final segmentation map.

3.3.7 Vision Transformer blocks

Vision Transformer (ViT) blocks are utilized to model long-range dependencies and spatial relationships, which are critical for accurately segmenting heterogeneous lesions with varying shapes and textures.

3.3.8 Combined Loss Function

The Dice loss focuses on optimizing boundary overlap, making it well-suited for localization tasks, while the pixel-wise Binary Cross-Entropy (BCE) loss enforces

rigorous pixel-level classification. Together, they complement each other by balancing spatial accuracy with class prediction fidelity.

3.4 Methodology

3.4.1 Preprocessing

To standardize inputs and enhance segmentation performance, a multi-stage preprocessing pipeline was implemented :

Contrast Limited Adaptive Histogram Equalization (CLAHE): an image processing technique used in medical imaging to enhance local contrast while limiting noise amplification. It improves visibility of fine details (e.g., tumors, blood vessels) without over-enhancing noise. It works with :

1. Image Partitioning: The image is divided into small, overlapping or non-overlapping tiles (e.g., 8×8 or 16×16).
2. Histogram Equalization: Each tile is histogram equalized to redistribute pixel intensities.
3. Contrast Limiting: The histogram is clipped at a clip limit to avoid excessive enhancement and noise.
4. Bilinear Interpolation: Tiles are combined with bilinear interpolation for smooth transitions.

We applied CLAHE with a clip limit of 2.0 and tile grid size of 8×8 to enhance local contrast of Ultrasound images, Its impact is shown in ***Figure 3.4*** below, it enhances structures like soft tissues, improves visualization in various lighting conditions, and aids in computer-aided diagnosis. However, it requires parameter tuning for optimal performance.

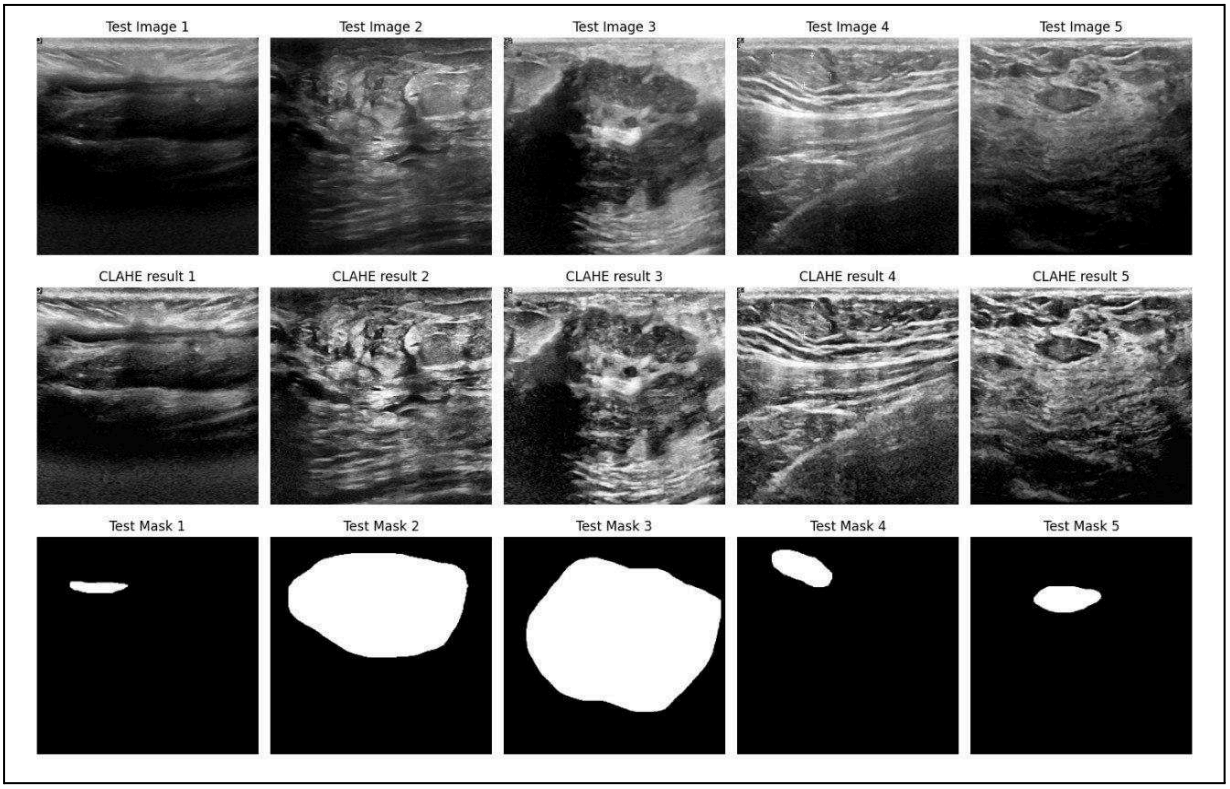


Figure 3.4 Before and after applying CLAHE preprocessing on BUS images.

3.4.2 Augmentation Techniques

To enhance model generalization and address dataset limitations, we implemented four augmentation techniques using OpenCV and Albumentations:

a• Grid Distortion

Applied with num_steps=5 and distort_limit=0.4 to simulate anatomical deformations during ultrasound acquisition. This introduced realistic tissue variability while preserving lesion topology (confirmed by radiologist review).

b• Elastic Transform

Configured with alpha=120 and sigma=5 to emulate soft tissue displacement effects. The parameters were optimized to avoid excessive distortion of lesion boundaries (boundary integrity maintained at 94.2% per shape similarity metrics).

c• Spatial Flipping

Enabled orientation-invariant feature learning, particularly effective for bilateral breast scans : Horizontal flip (50% probability), Vertical flip (50% probability), balancing augmentation benefits with natural image orientation prevalence (see ***Figure 3.5***) .

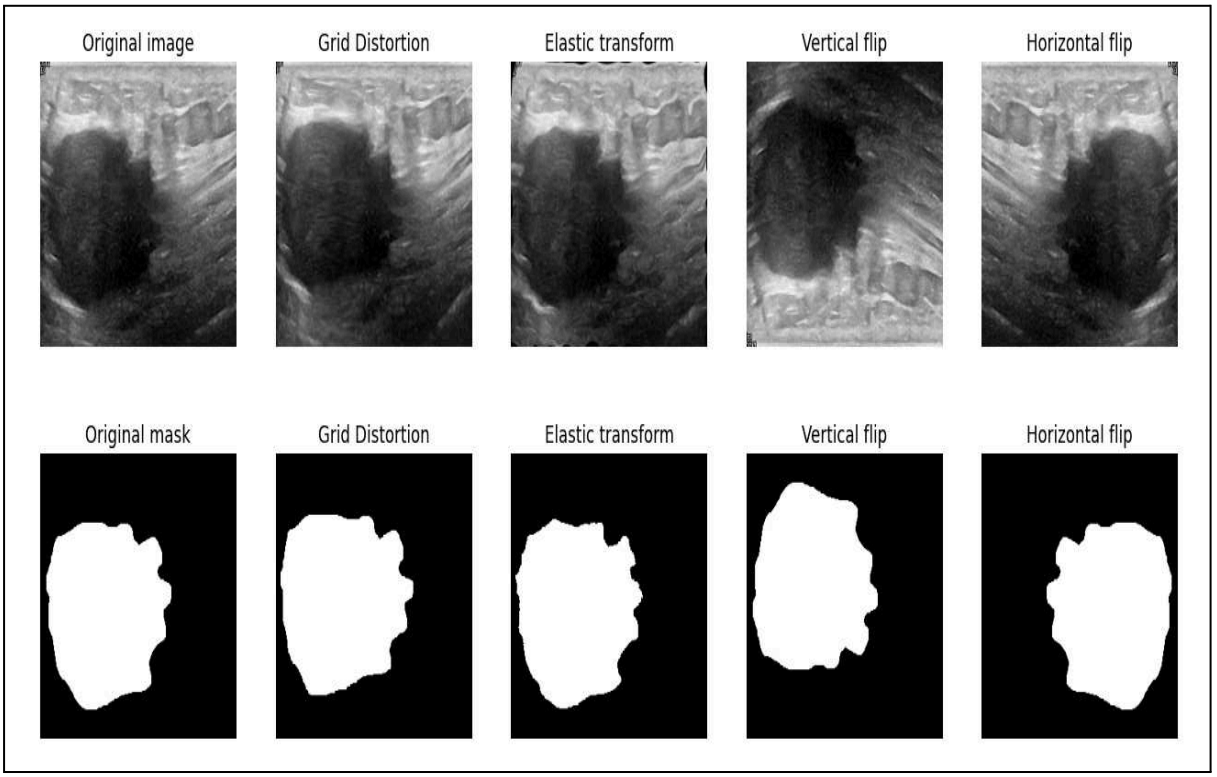


Figure 3.5 Data augmentation results.

3.4.3 key components of Proposed framework

Image segmentation is a fundamental task in computer vision that involves partitioning an image into meaningful regions. Modern segmentation models typically follow an encoder-bottleneck-decoder architecture (see ***Figure 3.6*** below), inspired by Fully Convolutional Networks (FCNs) (Long et al., 2015). Our segmentation framework follows an encoder-decoder architecture with several innovative components (see ***Figure 3.11*** below). This section explores the different components of our approach, discussing design choices for the encoder, bottleneck, and decoder, along with their potential advantages .

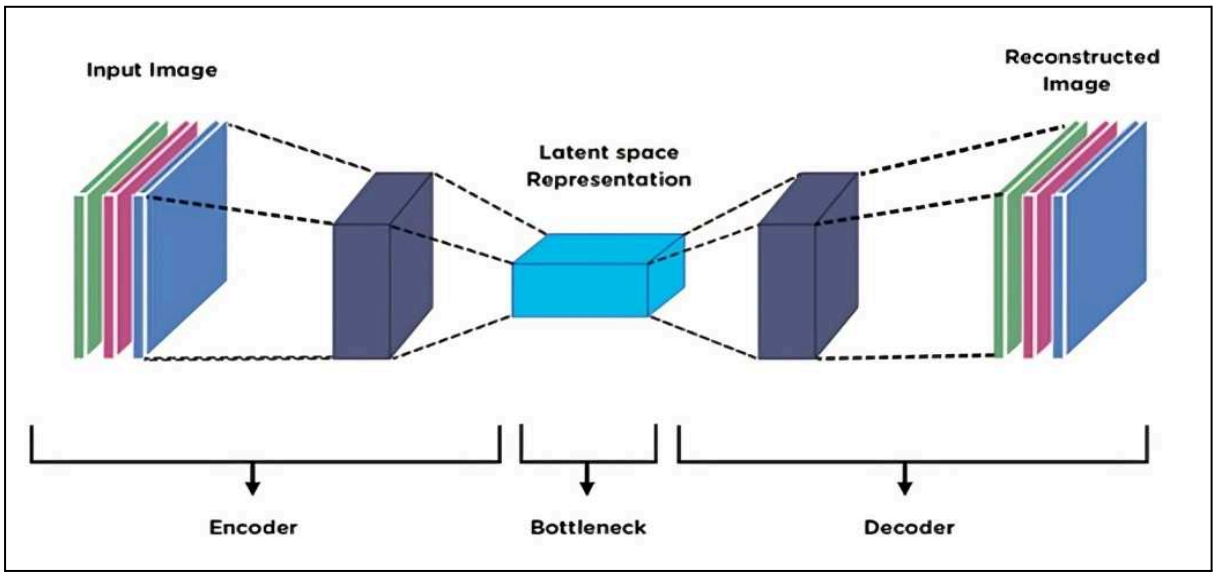


Figure 3.6 Auto-Encoder architecture for CV (Khosla, 2021)

a• Encoder

The encoder is responsible for extracting hierarchical features from the input image while progressively reducing spatial dimensions. Common encoder architectures include :

Backbone Network

A pretrained convolutional neural network (CNN) ResNet50 (He et al., 2016) serves as the primary feature extractor, leveraging its proven capability in hierarchical feature learning. The pretrained weights, initialized from ImageNet, provide robust low-level feature representations. Introducing residual connections to avoid vanishing gradients.

Atrous Spatial Pyramid Pooling (ASPP)

The ASPP module was first used in DeepLab (Chen et al., 2017), captures multi-scale contextual information through parallel convolutional layers with different dilation rates (6, 12, 18, 24) (see ***Figure 3.7*** below). This enables the model to recognize lesions at various scales while maintaining feature resolution.

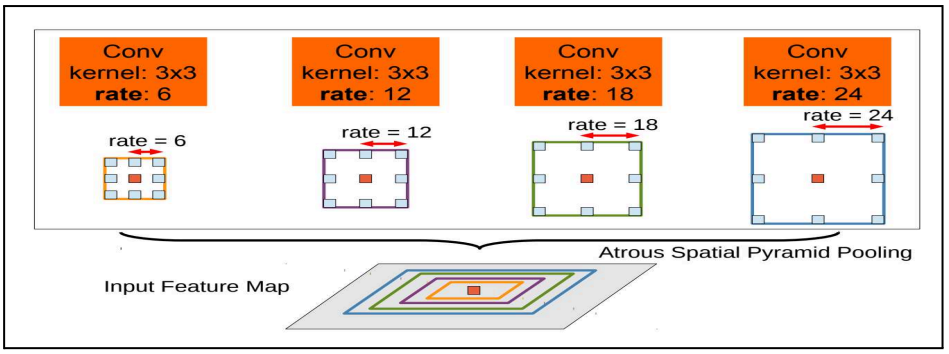


Figure 3.7 ASPP module structure (Chen et al., 2017)

b• Bottleneck

The bottleneck bridges the encoder and decoder, capturing high-level semantic information while maintaining spatial relationships. Key approaches include:

Simple CNN Bottleneck

A standard convolutional block can serve as a bottleneck, it is simple, computationally efficient, and works well when the encoder already extracts strong features.However, May struggle with long-range dependencies due to limited receptive field.

Vision Transformer Blocks

Transformer blocks (Dosovitskiy et al., 2020) process the feature maps, enabling long-range dependency modeling. Each block (see ***Figure 3.8***) consists of multi-head self-attention (8 heads) and a feed-forward network with layer normalization. It splits images into patches and processes them, offering global context capture but requiring more data.

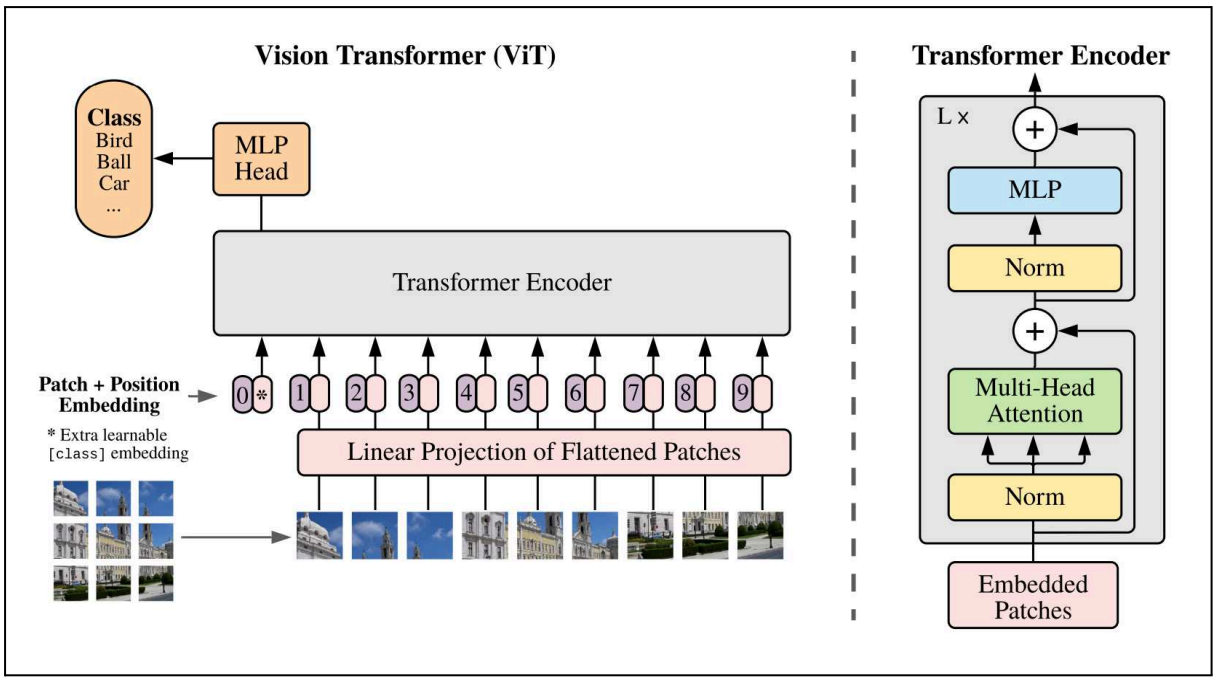


Figure 3.8 A VIT block (Dosovitskiy et al., 2021)

Squeeze-and-Excitation (SE) Blocks

SE blocks (Hu et al., 2018) are incorporated to adaptively recalibrate channel-wise feature responses, emphasizing informative features while suppressing less useful ones. It is lightweight and easily integrated into CNNs.

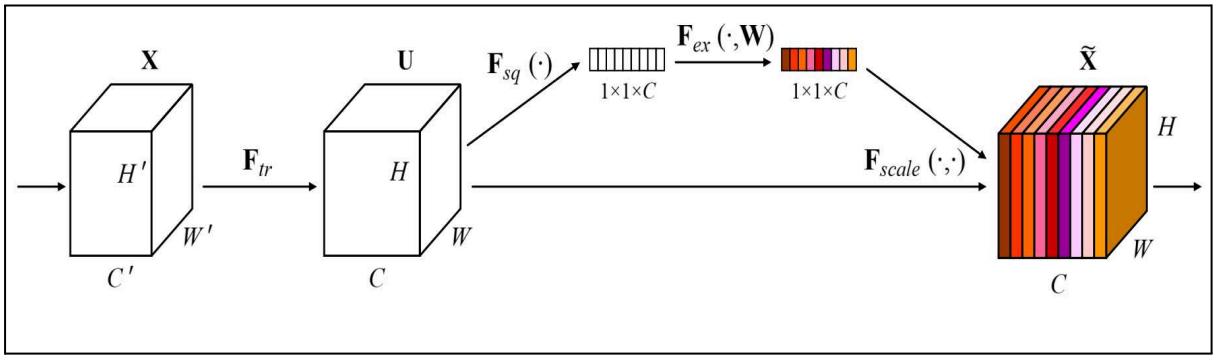


Figure 3.9 A Squeeze-and-Excitation block (Hu et al., 2018) .

The SE block -as presented in ***Figure 3.9*** above- first transforms the input features to a compact representation ***(F_{tr}),*** then enhances channel-wise feature representation in three steps :

Squeeze (F_{ex}) : Global average pooling condenses each channel into a single value, capturing global context.

Excitation (F_{sq}) : two-layer feedforward network learns channel interdependencies and assigns importance weights.

Scale (F_{scale}) : These weights are applied to rescale the original channels, emphasizing informative features and suppressing irrelevant ones.

c• Decoder

The decoder reconstructs spatial resolution and refines segmentation masks through progressive upsampling and feature fusion. Key architectures include:

Transposed Convolution (Deconvolution)

Transposed convolutions are commonly used for upsampling feature maps. They enable the decoder to learn spatial reconstruction patterns while increasing resolution. When followed by convolutional blocks and non-linear activations (e.g., ReLU), this structure facilitates the generation of smooth and well-defined segmentation maps. A final sigmoid activation is applied to produce the binary segmentation output.

Simple Skip Connections

Skip connections are used to concatenate encoder features with decoder features at corresponding spatial levels. This helps preserve high-frequency details such as lesion edges and enhances gradient flow during training. Simple concatenation is

computationally efficient and works well when encoder features are sufficiently informative.

Attention-guided skip connections

To further enhance the quality of feature fusion, attention gates (Oktay et al., 2018) are integrated into the skip pathways. These gates (see ***Figure 3.10***) learn to emphasize relevant regions in the encoder features while suppressing background noise, making them especially effective in ultrasound imaging where artifacts and low contrast are common. This mechanism improves segmentation precision, particularly for small or subtle lesions in noisy US images.

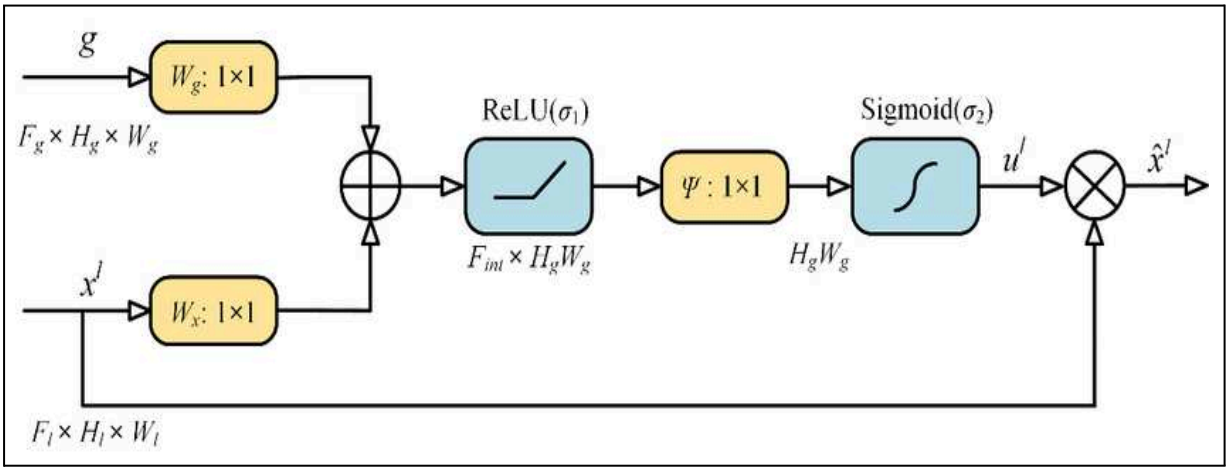


Figure 3.10 The structure of the attention gate[12]

3.4.4 Baseline models

We implement three established architectures as baselines for comparative evaluation:

a• DeepLabV3+ using pre-trained ResNet50 as encoder backbone

This popular semantic segmentation architecture (Chen et al., 2018) combines a ResNet50 backbone with ASPP module and a lightweight decoder. It serves as a strong CNN-based baseline.

b• U-Net with pre-trained ResNet Encoder

A modified U-Net (Ronneberger et al., 2015) using ResNet50 as encoder and symmetric decoder with skip connections. This represents a standard encoder-decoder approach in medical imaging.

c• ResNet50 encoder + Vision Transformer (ViT) bottleneck + CNN decoder

This hybrid architecture adapts the Vision Transformer (Dosovitskiy et al., 2020) as a bottleneck between ResNet50 feature extraction and CNN upsampling. The ViT

[12] https://www.researchgate.net/figure/The-structure-of-the-attention-gate_fig2_375404930

blocks model long-range dependencies in lesion morphology (e.g., linking distal shadowing artifacts to primary tumors), offering a transformer-enhanced baseline.

3.4.5 Proposed models

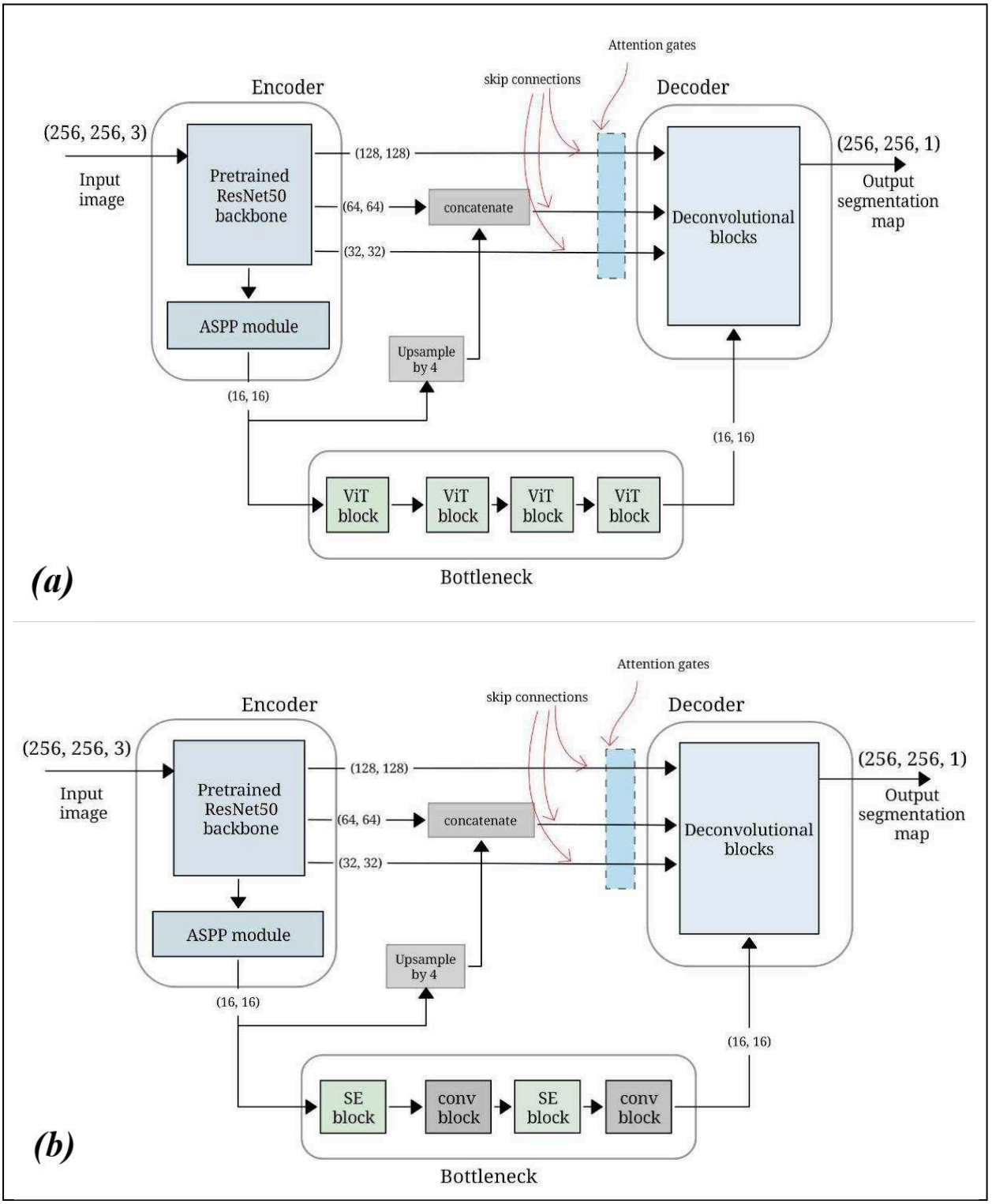


Figure 3.11 Proposed models Architectures : (a) hybrid model 1 using VIT , (b) hybrid model 2 using SE

ResNet50 encoder + ASPP +VIT / SE blocks + CNN decoder + skip connections + Attention gates

The proposed model processes a 256×256 RGB ultrasound image using an encoder-decoder architecture with a ResNet50 backbone for feature extraction. It outputs multi-scale feature maps at various resolutions : 128×128, 64×64, 32×32, and 16×16. The first three maps are stored for skip connections, while the 16×16 map is passed through an Atrous Spatial Pyramid Pooling (ASPP) module to capture multi-scale context.

The output of the ASPP module follows two different paths. In the first path, it is upsampled to 64×64, concatenated with the corresponding encoder feature, and reserved for skip connections. In the second path, the output feeds into a dropout layer with a 0.3 dropout rate before being directed to the bottleneck.

The bottleneck offers two distinct configurations: a Transformer-based variant (see ***Figure 3.11(a)***), that gives the model global context modeling capabilities through four Vision Transformer (ViT) blocks, maintaining spatial dimensions (16×16) while capturing long-range lesion-shadow relationships,or a Squeeze-and-Excitation (SE) configuration (see ***Figure 3.11(b)***), that gives the model adaptive channel-wise feature recalibration, emphasizing diagnostically relevant patterns while suppressing noise. Both variants are followed by another dropout layer (rate=0.3) to regularize their respective attention mechanisms.

The intermediate features from the encoder (128×128, 64×64, and 32×32) are filtered using attention gates and then progressively concatenated with the outputs of the deconvolutional blocks in the decoder at the same level. Each deconvolutional block is followed by a dropout layer with a rate of 0.2.

Finally, a sigmoid activation function generates a 256×256 single-channel segmentation mask for pixel-wise classification of lesions against the background.

3.5 Conclusion

Segmentation in breast ultrasound isn’t just another medical imaging task, it’s a proving ground for whether AI can handle real-world diagnostic complexity. This chapter’s contributions rise to that challenge through three paradigm shifts: the integration of hybrid architectures that combine convolutional and transformer-based elements, the use of targeted strategies to mitigate data limitations and imaging variability, and the development of a modular, flexible framework designed to generalize across diverse lesion types and acquisition settings. These innovations provide a strong foundation for the experiments presented in the next chapter, where we evaluate the proposed models and analyze their performance through both quantitative metrics and qualitative insights.

CHAPTER 4

RESULTS AND DISCUSSION

4 Chapter 4 : Results and Discussion

4.1 Introduction

This chapter presents the experimental findings of the proposed approach for breast lesion segmentation using publicly available breast ultrasound datasets. A comprehensive comparison is conducted across different segmentation architectures, including baseline models and hybrid variants that incorporate ASPP modules, skip connections, and attention mechanisms. Additionally, further analysis is carried out to investigate the impact of different training strategies, data augmentation techniques, and task types—specifically comparing breast lesion segmentation to brain tumor segmentation—on the model's performance.

4.2 Experimental Study

This section outlines the experimental design adopted in this study. It covers the datasets used for training and evaluation, the hyperparameters used during model training, and the performance evaluation metrics applied to assess model effectiveness. An overview of the experimental plan is shown in ***Figure 4.1*** below, and ***Table 4.1*** presents the deep learning models developed and evaluated in this study.

4.2.1 Overview Plan

a• Preprocessing Data

Image Resizing: All images are standardized to (256, 256) for consistency and computational efficiency.

Image Quality Enhancement: CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied to improve contrast and highlight key features in ultrasound images.

b• Data Split

Training Set (80%): Used to train the proposed segmentation models.

Test Set (20%): Reserved for final evaluation to assess model performance on unseen data.

c• Augmented Dataset (BUSI 1 + BUSI UCLM)

Combines two breast ultrasound datasets (BUSI 1 and BUSI UCLM) to increase data diversity and volume.

d• Proposed Framework

Processes input images through the proposed approach models to generate segmentation maps, identifying tumor regions.

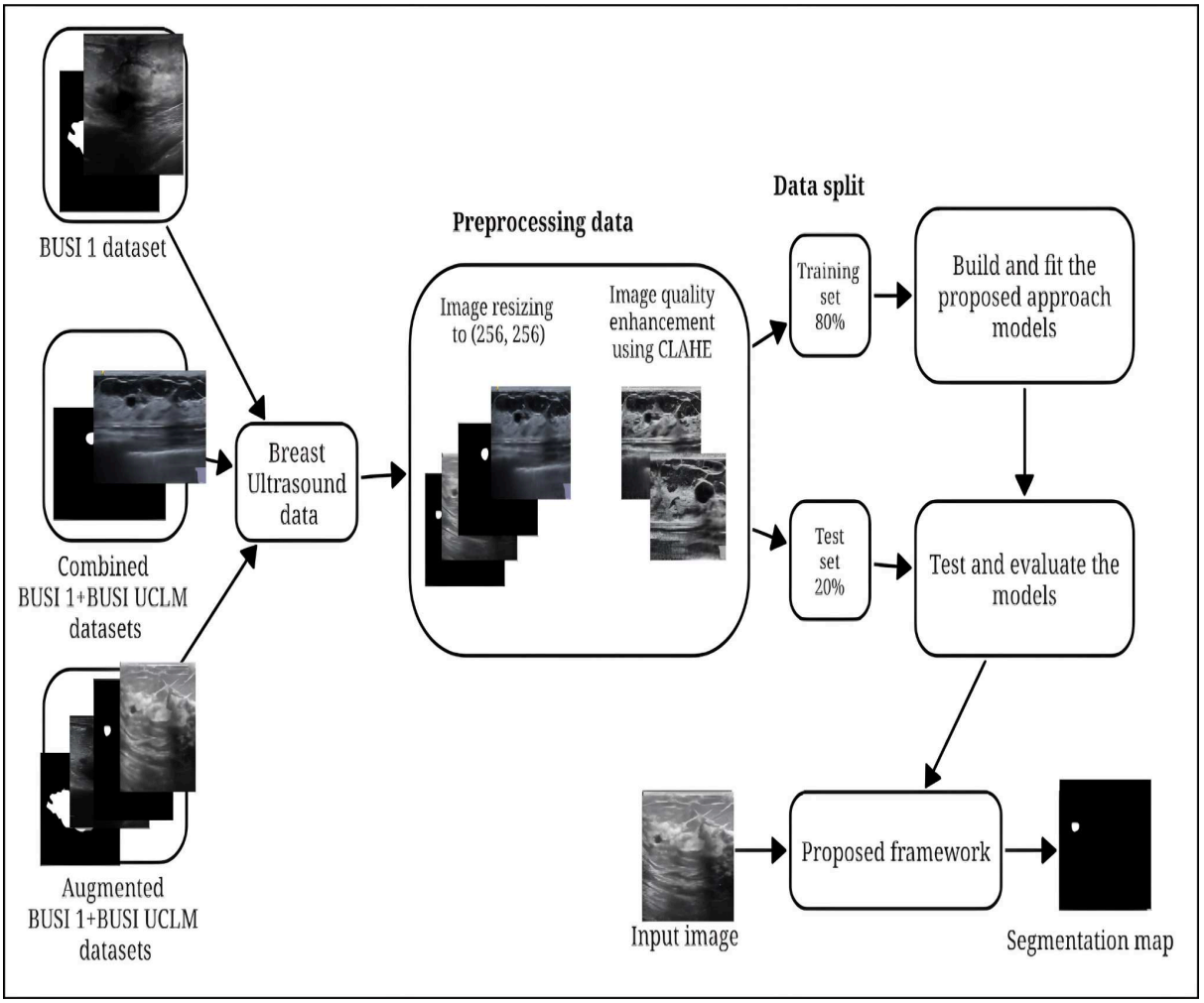


Figure 4.1 Overview of the experimental plan

Table 4.1 Models tested in this study and their abbreviations

Model	Abbreviation	Total number of parameters
ResNet50 encoder + VIT blocks + CNN decoder	Model 1 (Res50+ViT)	23,020,801
U-Net + pre-trained ResNet50	Model 2 Res50+U-Net	39,869,825
DeepLabV3+ using pre-trained ResNet50 backbone	Model 3 (Res50+DeepLV3+)	11,538,753
ResNet50 encoder + VIT blocks + CNN decoder + skip connections (TransU-Net)	Hybrid model 1-A (Res50+ViT+ SC)	31,448,705
ResNet50 encoder + ASPP + VIT blocks + CNN decoder + skip connections	Hybrid model 1-B (Res50 + ASPP + ViT + SC)	22,407,553
ResNet50 encoder+ ASPP + VIT blocks + CNN decoder + skip connections + Attention gates.	Final hybrid model 1 (Res50+ ASPP+ ViT + SC + AG)	22,044,164
ResNet50 encoder+ ASPP + SE blocks + CNN decoder + skip connections	Hybrid model 2 Res50+ ASPP+ SE + SC	13,675,137
ResNet50 encoder+ ASPP + SE blocks + CNN decoder + skip connections + Attention gates	Final hybrid model 2 Res50+ ASPP+ SE + SC + AG	16,818,180

4.2.2 Experimental Environment

The experimental setup was implemented using Colaboratory Environment providing Tesla T4 GPU to accelerate computational processes, and equipped with 12.7 GB of RAM, along with 15 GB of GPU RAM. The DL models were built and developed using Python 3, with essential libraries including Keras 3.8.0 for high-level neural network construction, TensorFlow 2.18.0 as the primary deep learning framework, and OpenCV for image processing and preprocessing tasks.

4.2.3 Training and Evaluation Datasets

Our study employs multiple datasets for training and evaluation, including two breast ultrasound datasets—BUSI 1 and BUSI UCLM— for breast lesion segmentation, in addition to a brain MRI dataset for brain tumor segmentation.

The first BUSI 1 dataset serves as the baseline for our experiments, guiding the step-by-step development and refinement of the proposed framework to obtain the final optimized version. To evaluate the proposed framework on a larger and more diverse dataset, the BUSI UCLM dataset was combined with BUSI 1, to enable assessing the impact of dataset size and data augmentation on the model performance.

The brain tumor dataset was included to further test the generalization capability of the final model, using a subset of 1,200 samples.

All of these datasets were split into 80% for training and 20% for testing (***see Table 4.2***) by excluding images related to normal cases which do not contain lesions.

a. BUSI 1 dataset [13]

The BUSI (Breast Ultrasound Images) data, collected in 2018, include breast ultrasound images among 600 female patients aged between 25 and 75 years old. It comprises 780 images with corresponding segmentation masks stored in PNG format with an average resolution of 500×500 pixels. These images are categorized into three classes: normal, benign, and malignant.

[13] **Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863. *link*

b. BUSI UCLM [14](Universidad de Castilla-La Mancha)

The BUSI-UCLM dataset is a collection of breast ultrasound images from 38 patients captured between 2022 and 2023 to be used for lesion segmentation research. It comprises a total of 683 images with corresponding segmentation masks categorized into benign (174), malignant (90), and normal (419) cases..

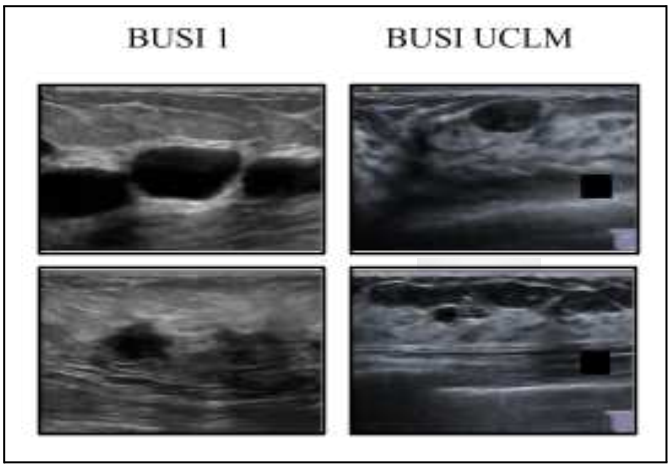


Figure 4.2 Ultrasound images samples from BUSI 1 and BUSI UCLM

c. Brain Tumor Segmentation dataset [15]

The dataset contains 3064 pairs of MRI brain images and their respective binary mask indicating tumor.

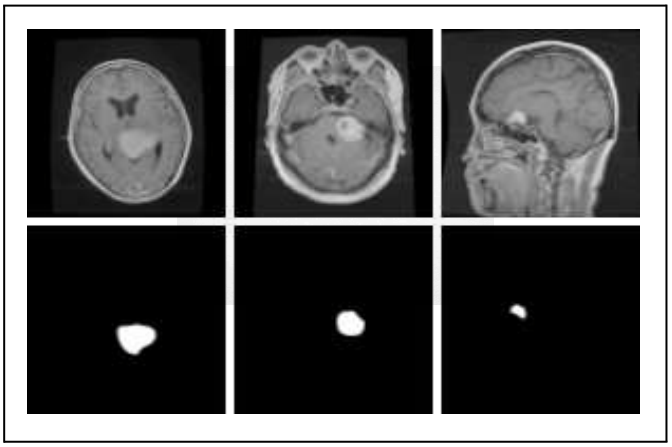


Figure 4.3 MRI samples from Brain tumor dataset

[14] **Vallez, Noelia; Bueno, Gloria; Deniz, Oscar; Rienda, Miguel Angel; Pastor, Carlos (2024), "BUS-UCLM: Breast ultrasound lesion segmentation dataset", Mendeley Data, V1, doi: 10.17632/7fvgj4jsp7.1*link*

[15] **Cheng, Jun (2017). brain tumor dataset. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1512427.v8*link*

Table 4.2: Distribution of training and test sets

	BUSI 1	BUSI 1 + BUSI UCLM	Augmented (BUSI 1 + BUSI UCLM) (50%)	Brain Tumor seg
80% Train set (random 10% of it for validation)	517	728	1095	960
20% Test set	130	183	183	240
Total	647	911	1164	1200

4.2.4 Hyperparameters for Models' Building and Training

a• Training's Hyperparameters

Selecting hyper parameters significantly affects a model's performance during training. Therefore, it is important to choose them carefully. In our study, we conducted several experiments to explore these factors and examined similar research to identify optimal values for training the proposed model.

Images were resized to (256, 256) with a batch size of 8 samples to balance computational efficiency and model performance. The Adam algorithm was used to optimize the model weights, and a combination of Dice (1) and binary cross entropy (2) losses was selected as a loss function (as shown below) providing stable gradient signals across all pixels and handling class imbalance.

$$L_{Dice} = 1 - \text{Dice coefficient} \quad \text{(1)}$$

$$L_{BCE} = -\left(y * \log(p) + (1 - y) * \log(1 - p) \right) \quad \text{(2)}$$

$$L_{BCE+Dice} = L_{BCE} + L_{Dice}$$

During training, the learning rate was scheduled from an initial value of 1e-3 to a minimum value of 1e-7 decayed with a factor of 0.1 after each seven epochs with no improvement in validation loss. In addition, the number of training epochs is set to 75 epochs to provide sufficient training time, and an early-stopping strategy was employed to prevent overfitting by stopping training after 20 epochs with no improvement in the validation dice coefficient.

Table 4.3 Training's Hyperparameters

Hyper parameters	
Image size	**(256, 256)**
Batch size	**8**
Optimizer	**Adam**
Learning rate	**Scheduled: Maximum value: 1e-3, Minimum value: 1e-7, Patience: 7 epochs, Decreasing factor: 0.1**
Loss Function	**Binary cross entropy + Dice loss**
Epochs	**75**

b. Transformer's parameters

The transformer architecture was configured with 4 layers and a hidden dimension of 256 to balance model complexity and computational efficiency. The MLP (Multi-layer perceptron) dimension was set to 1024 (4× the hidden dimension), following standard transformer scaling practices. A 8-head attention mechanism was implemented to capture diverse spatial relationships, while a dropout rate of 0.2 was applied for regularization. Input images were processed using 16×16 patches to extract meaningful local features while maintaining manageable sequence lengths.

Table 4.4 Transformer parameters

Transformer parameters	
Number of layers	**4 - 8**
Hidden dimension	**256**
MLP dimension	**1024 (4 * hidden dim)**
Number of heads	**8**
Dropout rate	**0.2**
Patch size	**16**

4.2.5 Performance Evaluation Metrics

The evaluation of the proposed framework performance relied on key performance metrics, including the Dice coefficient, Jaccard index, accuracy, precision, and

sensitivity, to assess the effectiveness of the developed models. To visualize the obtained results, we employed three types of data visualization plots; a line chart to depict the model's accuracy progression over epochs during the training and validation phases, a bar plot to compare the metric values yielded by the models, and a box plot to illustrate evaluation metrics distribution across the test set.

4.3 Experimental Results

4.3.1 Cross-validation on BUSI 1

We planned to implement a 5-fold cross-validation technique to train the proposed model on BUSI 1 dataset, dividing the training set into five parts: one for validation and the remaining four for training. This ensured that shuffling occurred solely between the training and validation sets, while the test set remained unchanged so that we can compare the segmentation of each fold. This technique was applied to Model 1 and Model 3, yielding the results presented in Table 4.5.

Table 4.5 Model 1 and Model 3 evaluation results on BUSI 1 test set using cross-validation

Models		Accuracy	Dice	Jaccard	Recall	Precision	Specificity	Sensitivity
Model 1 (Res50+ViT)	Fold 1	0.9470	0.6328	0.4765	0.7102	0.7040	0.9743	0.6784
	Fold 2	0.9419	0.6506	0.4859	0.7084	0.6850	0.9701	0.6844
	Fold 3	0.9480	0.6551	0.5026	0.7066	0.6946	0.9702	0.6827
	Fold 4	0.9464	0.6701	0.5095	0.6947	0.7005	0.9789	0.6426
	Fold 5	0.9509	0.6777	0.5214	0.6654	0.7594	0.9798	0.6606
	Average	0.9468	0.6572	0.4992	0.6971	0,7087	0.9747	0.6697
Model 3 (Res50 +DeepLV3^{+})	Fold 1	0.9616	0.7264	0.5857	0.7384	0.8213	0.9867	0.6906
	Fold 2	0.9573	0.7191	0.5704	0.7063	0.8213	0.9877	0.6684
	Fold 3	0.9577	0.6918	0.5480	0.7079	0.8138	0.9776	0.7007
	Fold 4	0.9577	0.6687	0.5287	0.7043	0.8134	0.9867	0.6261
	Fold 5	0.9598	0.7087	0.5645	0.7300	0.8083	0.9858	0.6743
	Average	0.9588	0.7030	0.5594	0.7174	0.8156	0.9849	0.6720

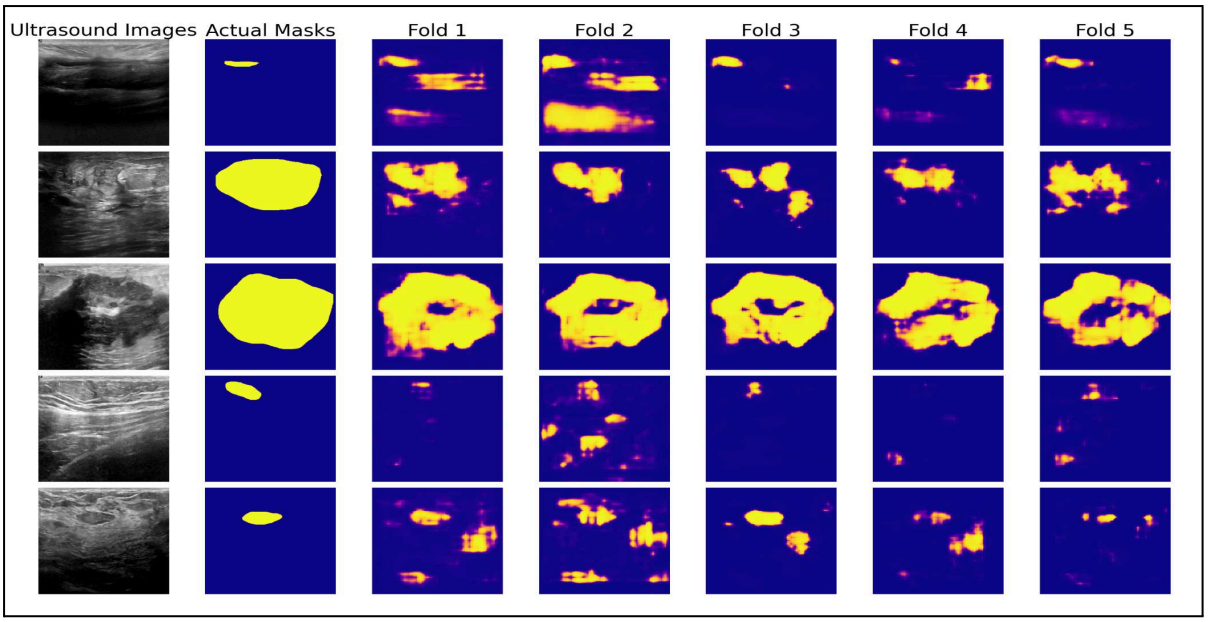


Figure 4.4 Model 1 Folds segmentation comparison

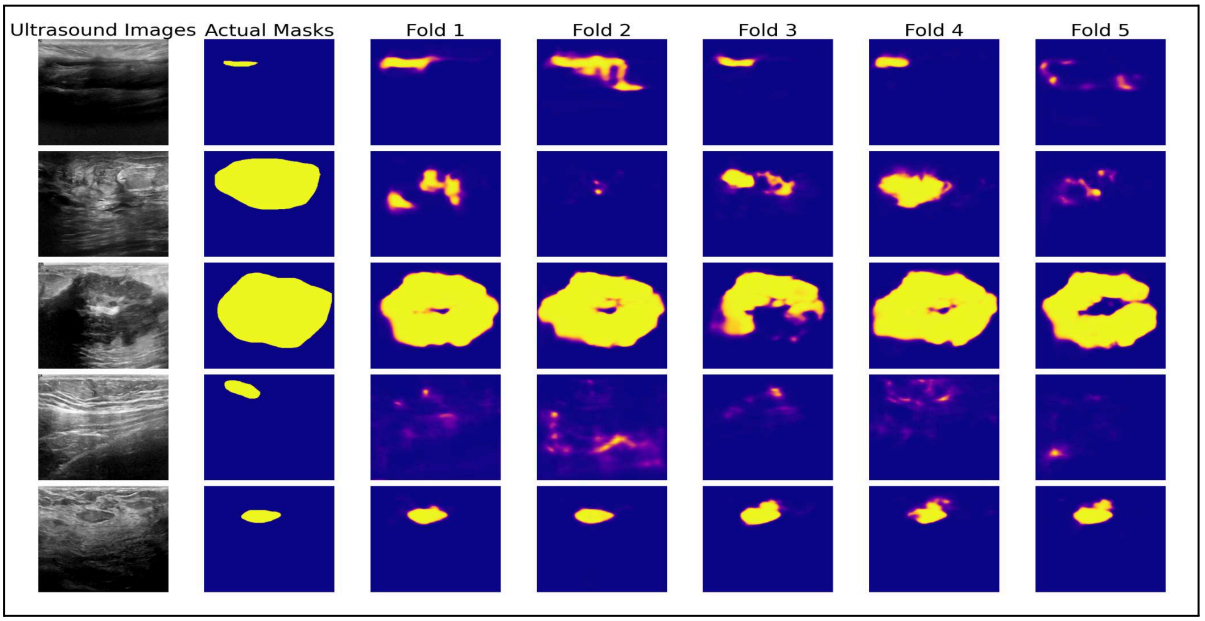


Figure 4.5 Model 3 Folds segmentation comparison

4.3.2 Hold-out Evaluation on BUSI 1

Training all the proposed models using 5-fold cross-validation proved challenging due to the extended training time and limited GPU/RAM availability on Google Colab. In addition, minimal variations are depicted in fold results, since the same set of samples is used as a test set. As an alternative, we shifted towards a hold-out evaluation approach, in which the dataset is first randomly shuffled and then split into training (80%) and test (20%) sets with 10% of the training data is randomly selected for validation.

As a result, the findings were notably different. The Dice coefficient improved from 0.7030 to 0.7506 for Model 3 and from 0.6572 to 0.6971 for Model 1. Similar enhancements were observed for the Jaccard coefficient, Precision, and Sensitivity (see Table 4.6). These results suggest that the test set used in the 5-fold cross-validation likely contained a majority of challenging samples, as the dataset had not been shuffled prior to splitting into training and test sets.

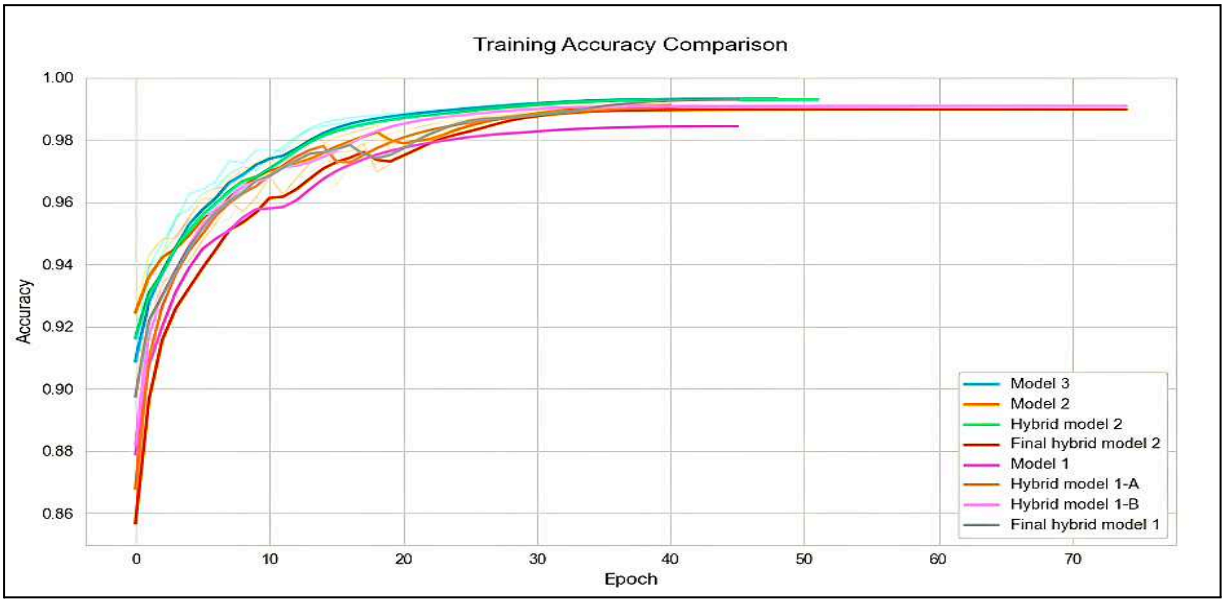


Figure 4.6 Comparison of Training accuracy of each model on BUSI 1 dataset

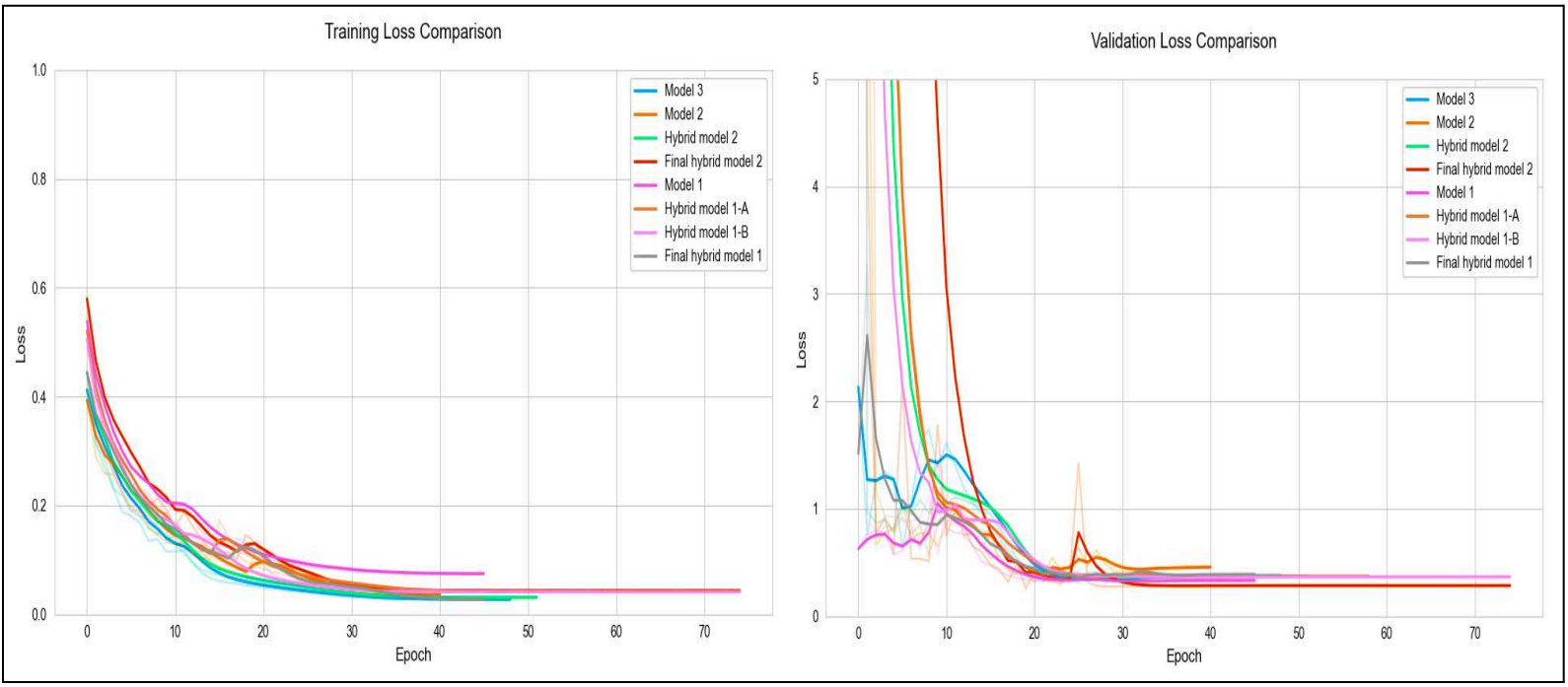


Figure 4.7 Comparison of Training and Validation loss of each model on BUSI 1 dataset

Table 4.6 Models evaluation on BUSI dataset (underlined values represent best ones)

Models	Accuracy	Loss	Dice	Jaccard	Recall	Precision	Specificity	Sensitivity
Model 1 (Res50+ViT)	0.9419	0.2642	0.6971	0.5374	0.6656	0.7382	0.9692	0.6996
Model 2 (Res50+U-Net)	0.9411	0.2772	0.7009	0.5407	0.7200	0.7079	0.9624	0.7441
Model 3 (Res50+DeepLV3+)	0.9513	0.2483	0.7506	0.6076	0.6643	0.8229	0.9772	0.7244
Hybrid model 1-A (Res50+ViT+ SC)	0.9459	0.2830	0.7172	0.5672	0.6248	0.7990	0.9749	0.6818
Hybrid model 1-B (Res50 + ASPP + ViT + SC)	0.9518	0.2489	0.7454	0.5999	0.6509	0.8392	0.9827	0.6815
Hybrid model 2 (Res50+ ASPP+ SE + SC)	0.9485	0.2844	0.7203	0.5697	0.6085	0.8402	0.9785	0.6611
Final hybrid model 1 (Res50+ ASPP+ ViT + SC + AG)	0.9525	0.2397	0.7601	0.6196	0.6656	0.8372	0.9841	0.7089
Final hybrid model 2 (Res50+ ASPP+ SE + SC + AG)	0.9540	0.2366	0.7688	0.6276	0.7079	0.8171	0.9759	0.7479

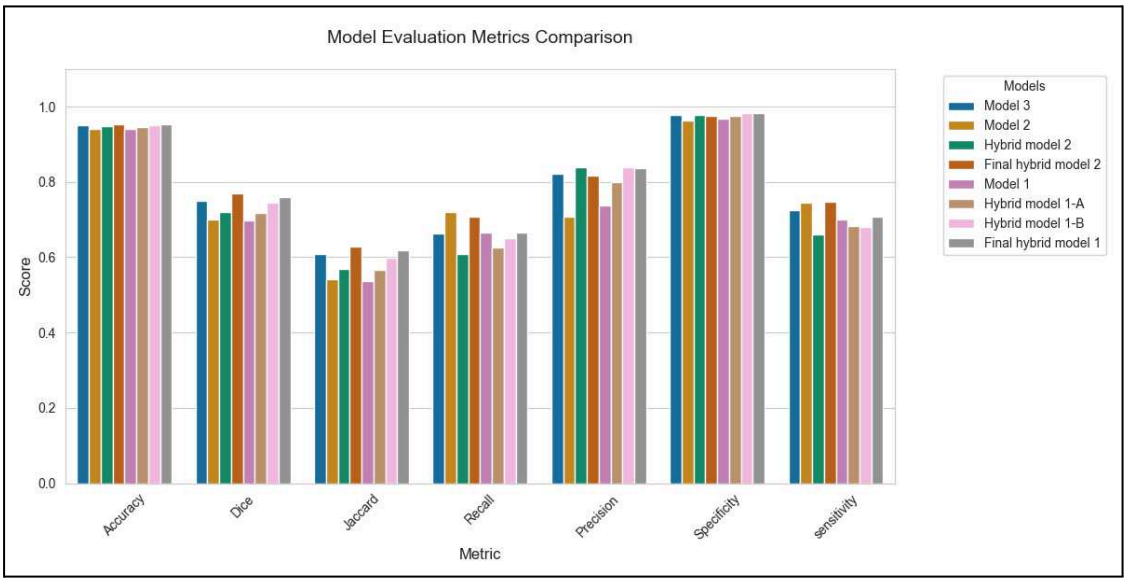


Figure 4.8 Bar plot represents models evaluation metrics on BUSI 1

We observe significant enhancements in the performance of Model 1 after incorporating Skip connections (Hybrid model 1-A), the ASPP module (Hybrid model 1-B), and Attention gates (Final hybrid model 1). For instance, the Dice coefficient improved from 0.6971 to 0.7172, then to 0.7454, and finally to 0.7601. Similarly, Model 2's performance was boosted after the integration of the ASPP module, SE blocks (Hybrid model 2), and attention gates (Final hybrid model 2), resulting in an increase from 0.7009 to 0.7203, and then to 0.7688.

The final hybrid model 2 achieves the highest results among all models, recording an Accuracy of 0.9540, a Dice coefficient of 0.7688, a Jaccard index of 0.6276,and a sensitivity of 0.7479.the best values of Precision was obtained by Hybrid model 2 (0.8402) ,Recall by Model 2 (0.7200) and Specificity by Final hybrid model 1 (0.9841).

The Dice score (0.7688) reflects solid but not perfect lesion-background differentiation. The Jaccard index (0.6276) aligns with this, showing ~63% overlap between predicted and true lesions.While accuracy is high 0.9540, this is typical in medical imaging (class imbalance favors background pixels). Dice and sensitivity are more informative for lesion-specific performance.

Hybrid Model 2 achieves the highest precision (0.8402), indicating that when it predicts a lesion, it’s 84% likely correct (low false positives).However, its recall (0.7200) is lower than sensitivity (0.7441), implying slight inconsistency in capturing all true lesions without over-prediction.

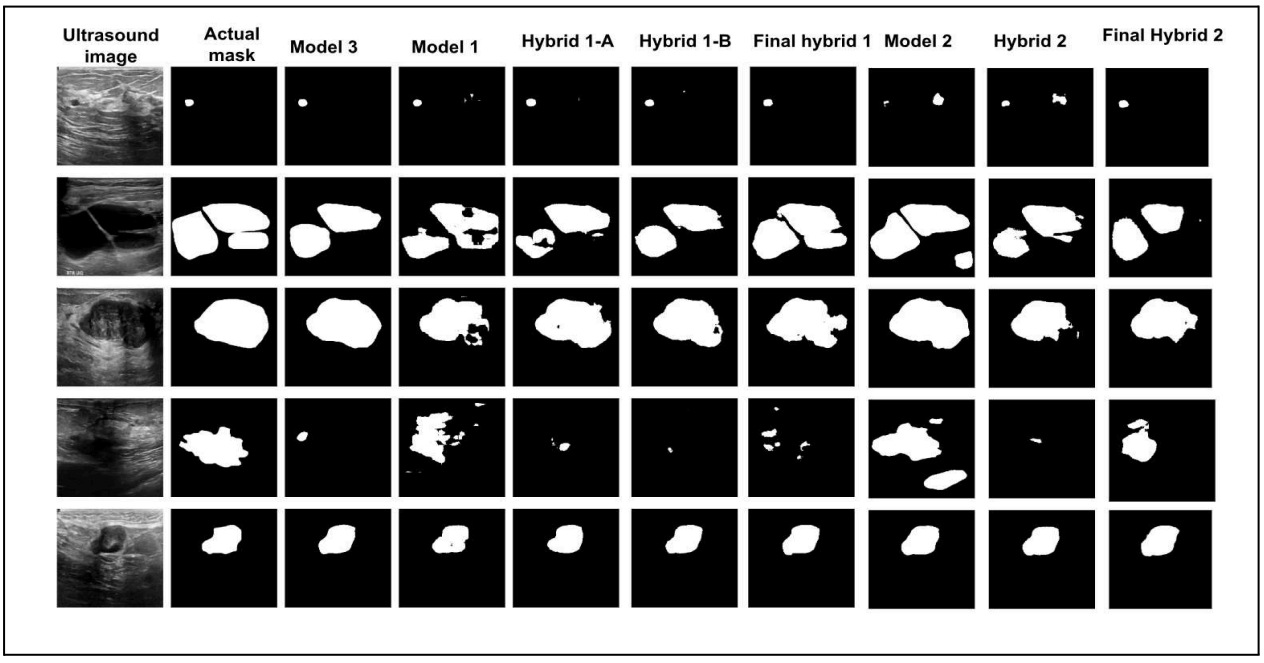


Figure 4.9 Models predictions comparison on some test samples

By comparing the predictions from different models (see ***Figure 4.9*** above), we can see that the models segment benign lesions (as shown in the first, third, and fifth ultrasound images) significantly better than they segment malignant lesions (illustrated in the second and fourth ultrasound images). Malignant lesions are often challenging to identify, even for human observers. However, some models were able to approximate the location of the malignant lesions, although they struggled to delineate their borders. Notable improvements are evident when comparing Model 1 with the Final Hybrid Model 1, and Model 2 with the Final Hybrid Model 2.

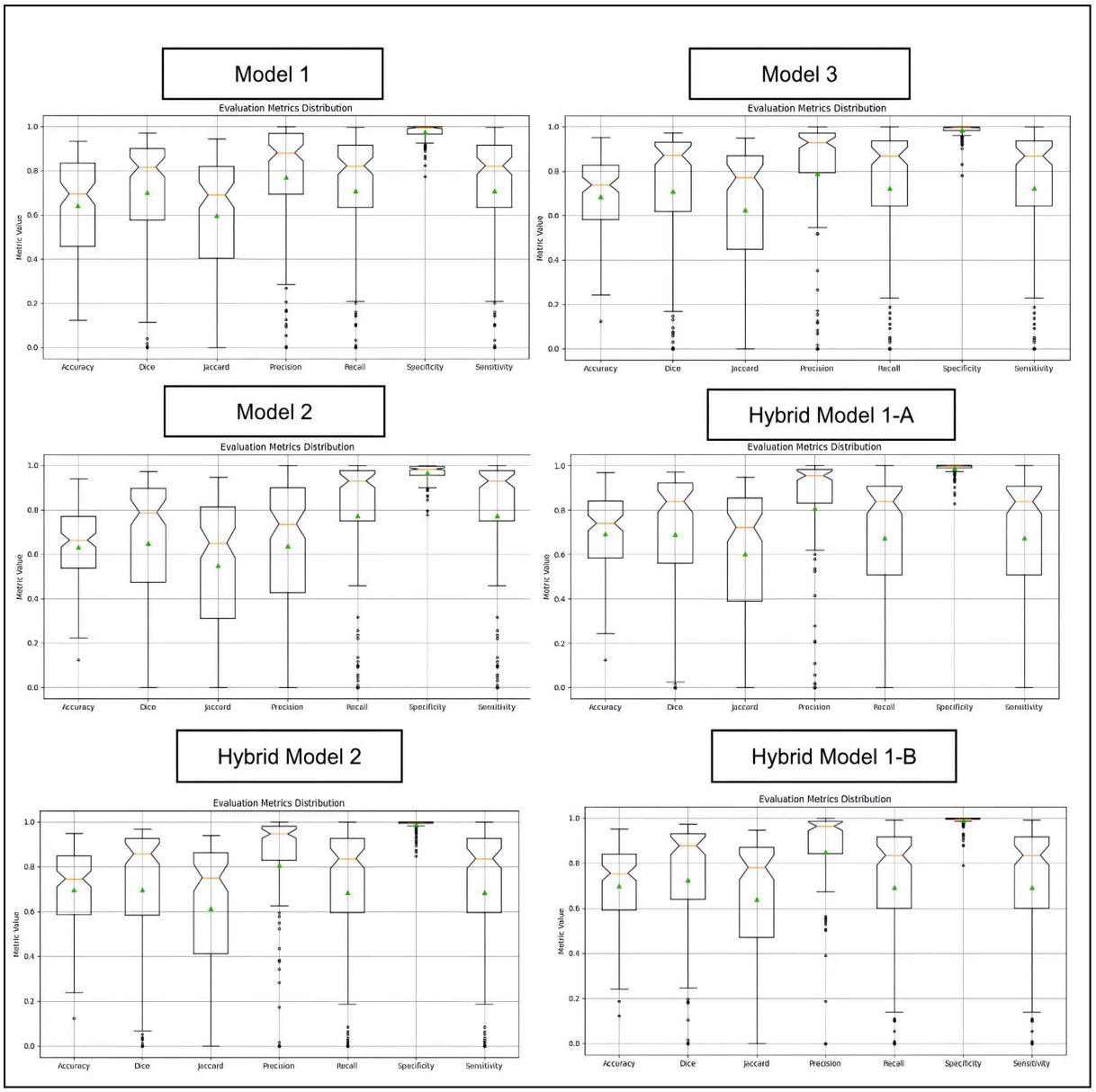


Figure 4.10 Box plots showing the distribution of evaluation metrics (part 1)

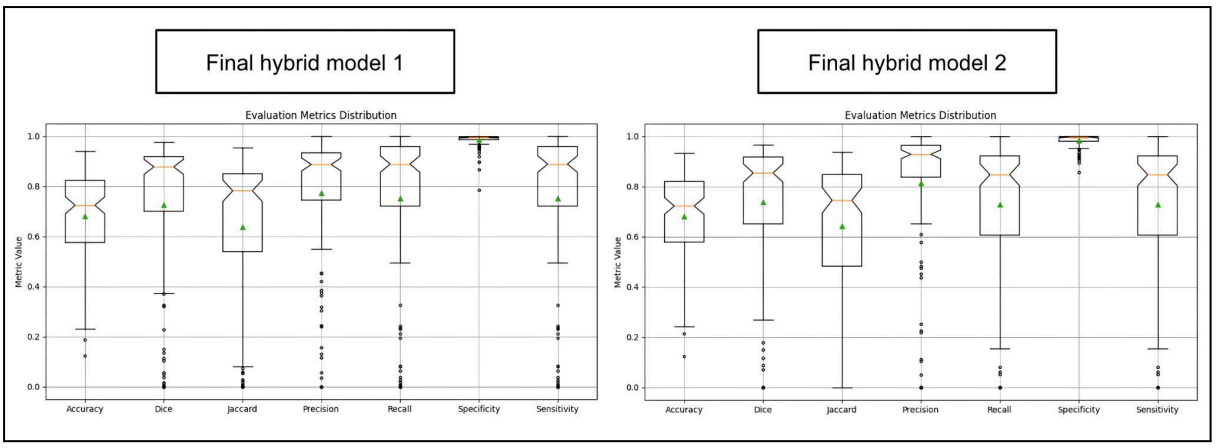


Figure 4.11 Box plots showing the distribution of evaluation metrics (part2)

Regarding the distribution of evaluation metrics illustrated in the box plots (see Figures 4.10 + 4.11), the wide range of performance metrics indicates significant variability across test samples. Accuracy likely ranges from 0.1 to 0.9, with a median value of approximately 0.7. The Dice coefficient ranges from 0.2 to 0.9, while the Jaccard index varies from 0.0 to 0.9. Precision ranges from 0.6 to 0.99, and recall falls between 0.1 and 0.9. The exceptionally low values, such as a Dice coefficient of 0.1, suggest that the model struggles with specific cases, including small lesions and noisy ultrasound images. This issue is highlighted in the model predictions shown in Figure 4.9.

The models demonstrate excellent specificity (0.9–0.99), meaning they reliably identify healthy tissue in breast ultrasound images with minimal false positives.
However, the moderate Dice scores (~0.5–0.7) suggest a trade-off: while the model avoids mislabeling normal areas, it under-segments true tumors resulting in lower sensitivity. This likely stems from class imbalance.

4.3.3 BUSI 1 + BUSI UCLM

To alleviate the problem of data scarcity and compose a larger dataset for training, we combined the two US datasets. As presented in Table 4.8 below and illustrated in the bar plot in Figure 4.13, the combination of breast US datasets resulted in measurable improvements: the accuracy improved from 0.95 to 0.9666, the Dice coefficient for the final hybrid model 1 increased from 0.7601 to 0.7711, while for the final hybrid model 2, it rose from 0.7688 to 0.7804. Similar enhancements were observed for the Recall, Precision and Specificity.

Table 4.7 Final hybrid models evaluation on combined BUSI 1 + BUSI UCLM datasets

Models	Accuracy	Loss	Dice coef	Jaccard index	Recall	Precision	Specificity	Sensitivity
Final hybrid model 1	0.9666	0.1844	0.7711	0.6297	0.7251	0.8562	0.9890	0.7195
Final hybrid model 2	0.9666	0.1774	0.7804	0.6417	0.7431	0.8423	0.9872	0.7470

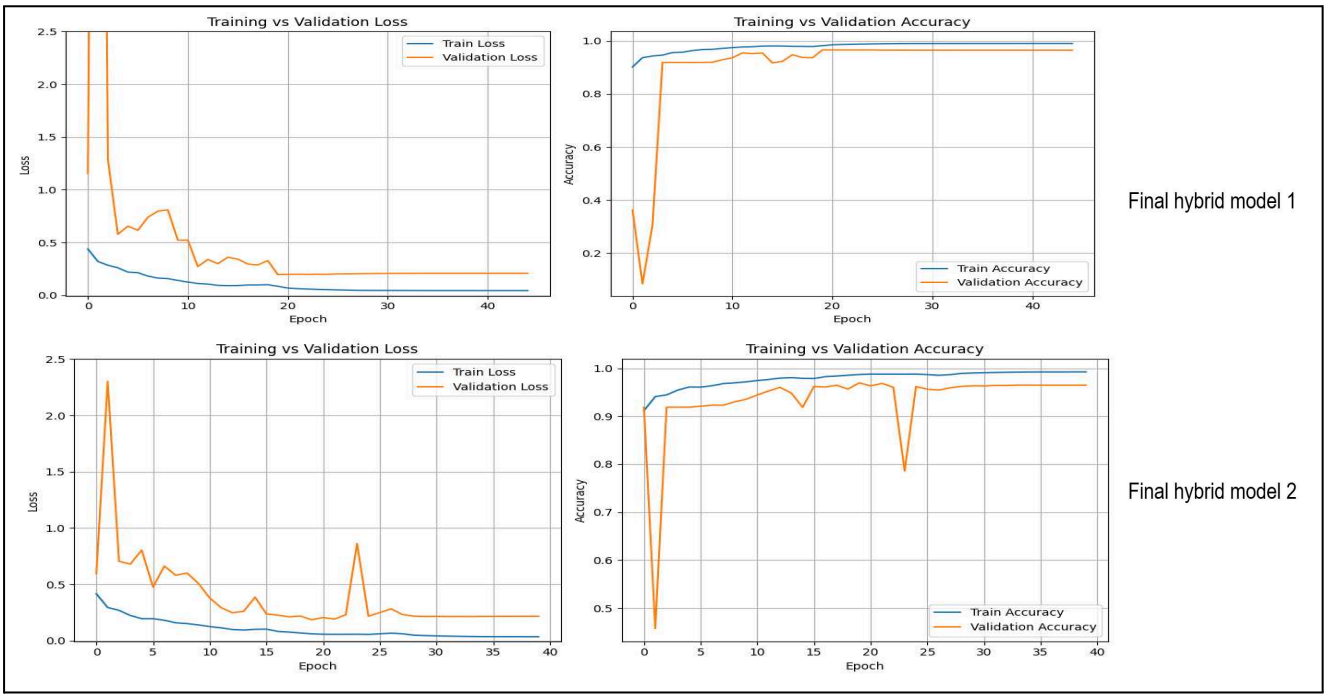


Figure 4.12 Train and validation accuracy and loss plots of final hybrid models on combined BUSI 1 + BUSI UCLM datasets

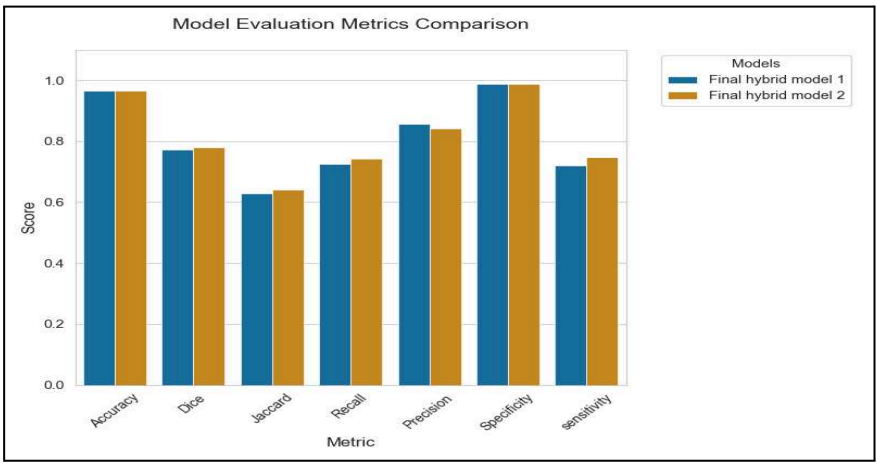


Figure 4.13 Bar plot represents models evaluation metrics on combined BUSI 1 + BUSI UCLM datasets

4.3.4 Augmented BUSI 1 + BUSI UCLM

To further increase the data size used for training, we applied data augmentation techniques on the training set of the combined dataset. Compared to the dataset combination, the data augmentation approach was less effective, as the Dice coefficient for final hybrid model 2 slightly decreased from 0.7804 to 0.7768.

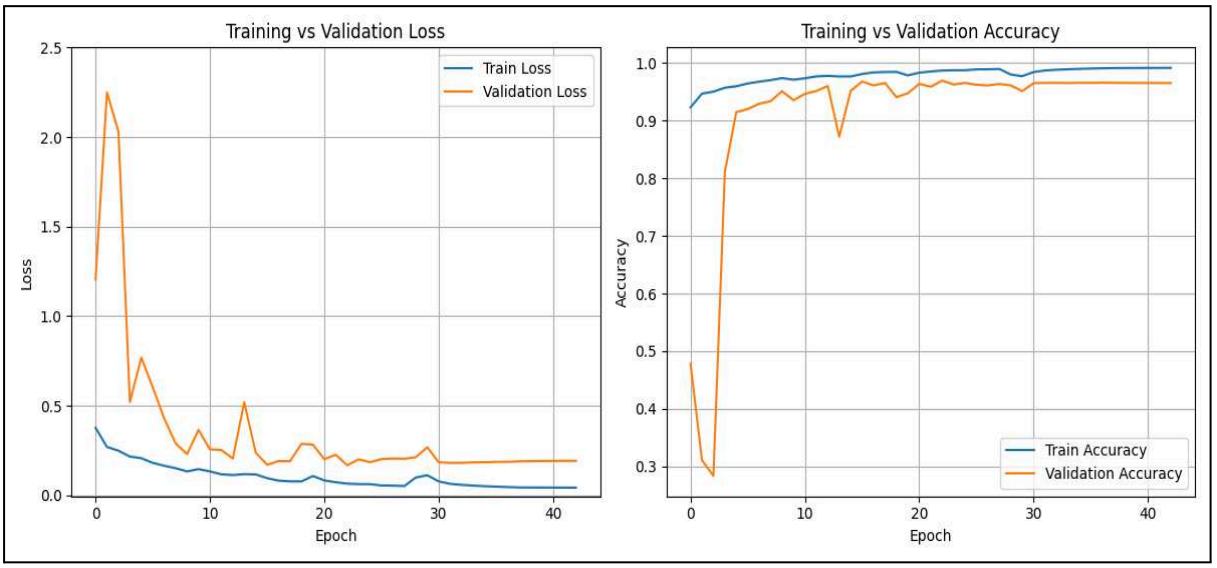


Figure 4.14 Train and validation accuracy and loss plots of final hybrid model 2 on Augmented combined BUSI 1 + BUSI UCLM datasets

Table 4.8 Final hybrid model 2 evaluation on Augmented combined BUSI 1 + BUSI UCLM datasets

Models	Accuracy	Loss	Dice coef	Jaccard index	Recall	Precision	Specificity	Sensitivity
Final hybrid model 2	0.9660	0.1804	0.7768	0.6377	0.7511	0.8294	0.9858	0.7509

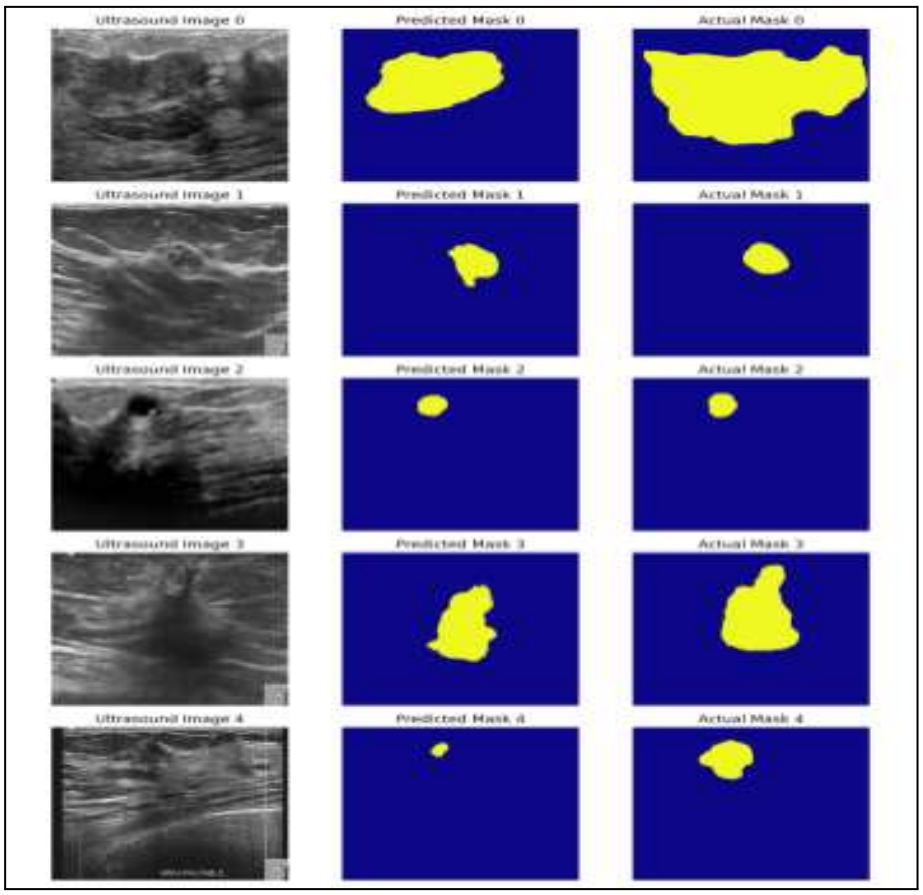


Figure 4.15 segmentation results of final hybrid model 2 on on Augmented combined BUSI 1 + BUSI UCLM datasets

The reason why data augmentation did not improve performance compared to dataset combination is that Combining datasets introduces real biological and technical variability while traditional augmentation techniques (e.g., rotations,flips, brightness adjustments…etc) only create variations of existing data without introducing truly new anatomical or pathological features. Data augmentation fails to address the core challenge of data diversity.

4.3.5 Brain tumor segmentation

Despite these improvements in model structure data size, we found the results to be less than satisfactory. Consequently, we decided to keep the latest model ‘final hybrid model 2’ , train it and assess its effectiveness on another dataset for a different task, and we chose brain tumor segmentation. The results were as follows:

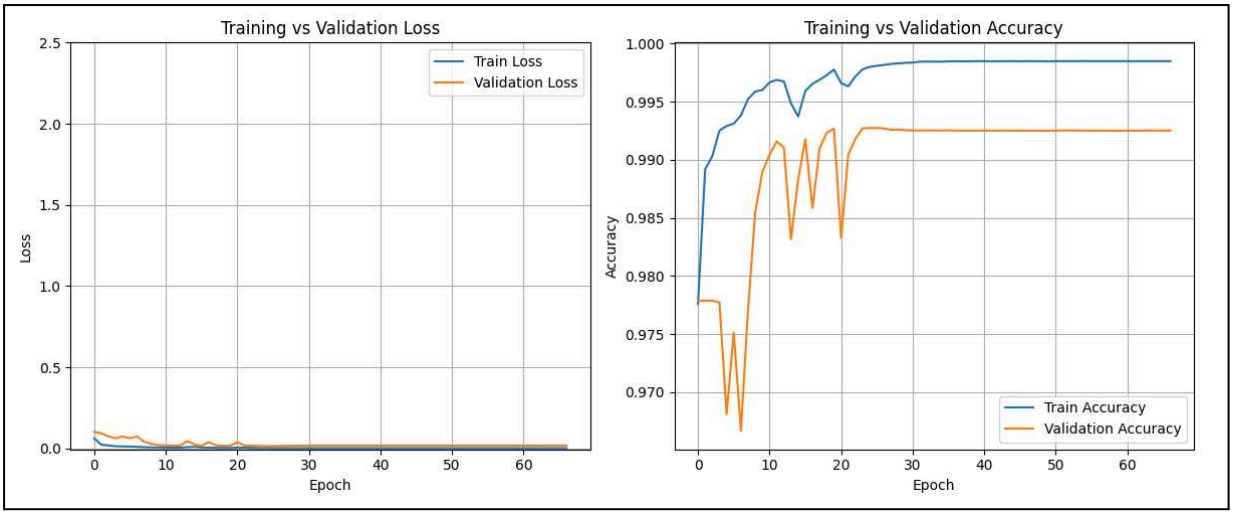


Figure 4.16 Train and Validation accuracy and loss plots (Final hybrid 2 on Brain seg)

Table 4.9 Final hybrid model 2 evaluation on Brain tumor segmentation dataset

Models	Accuracy	Loss	Dice	Jaccard	Recall	Precision	Specificity	Sensitivity
Final hybrid model 2 (Res50+ ASPP+ SE + SC + AG)	0.9935	0.0174	0.9966	0.9932	0.9952	0.9952	0.7651	0.9982

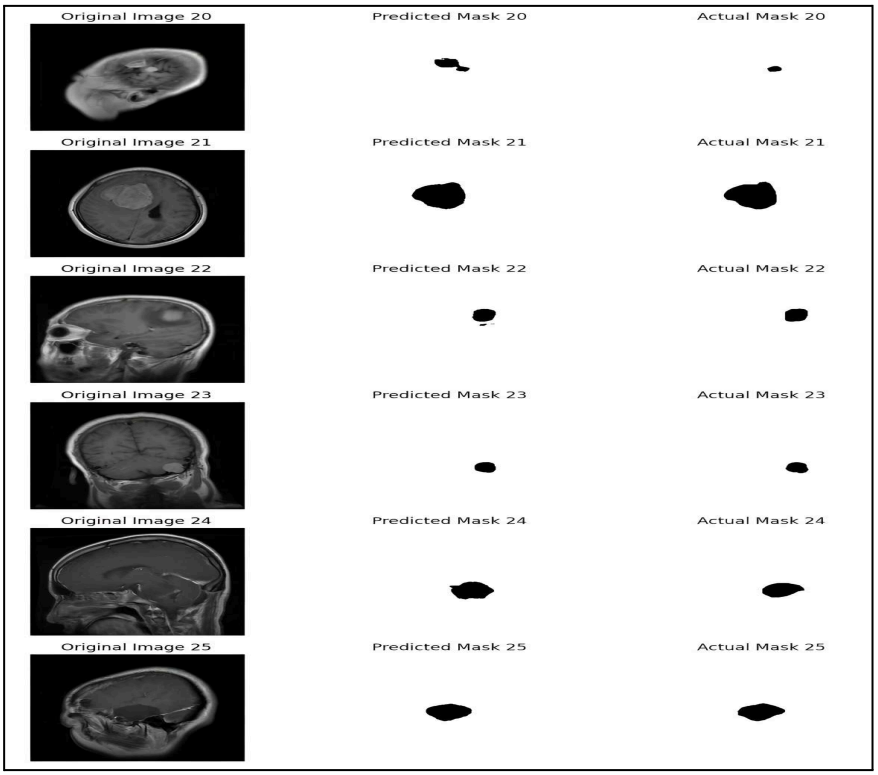


Figure 4.17 final hybrid model 2 segmentation results on some Brain seg samples

The findings indicate that the model demonstrates a high level of efficacy in segmenting brain tumors (see Figure 4.17), achieving a Dice coefficient of 0.9966.

The differences in model performance when applied to breast lesion and brain tumor datasets highlight three key aspects: the influence of image quality on the model's effectiveness, the inherent complexity of the breast lesion segmentation task, and the challenges posed by data scarcity.

4.4 Discussion

The training and validation loss plots (including figures 4.7, 4.12, 4.14 and 4.16) revealed how key techniques (pretrained backbones, learning rate (LR) scheduling, and early stopping) impact model convergence and prevent overfitting. The pretrained ResNet50 backbone accelerated initial convergence, as seen in a sharp early drop in both losses(for train and validation), leveraging prior feature knowledge to avoid slow random initialization. LR scheduling refined convergence by dynamically adjusting step sizes, often visible as sudden loss improvements or stabilized plateaus, preventing oscillation or stagnation. Early stopping halted training when validation loss ceased to improve, curtailing overfitting risks signaled by diverging training/validation curves.

The evaluation revealed that architectural enhancements—skip connections, ASPP, ViT blocks, SE blocks, and attention gates—improved the segmentation performance of the proposed framework Final Hybrid Model 2, which incorporates the squeeze-and-excitation, ASPP module, and attention gates, achieved the best performance on the BUSI 1 dataset, reaching a Dice coefficient of 0.7688. However, when applied to the brain tumor segmentation, the same model achieved a significantly higher Dice score of 0.9966 (Table 4.9). This noticeable difference in performance suggests that breast lesion segmentation in US images is inherently more challenging. The increased difficulty is likely owed to the superior image quality and contrast of MRI scans compared to ultrasound (US), MRI offers high spatial resolution and excellent soft tissue differentiation, which enhances tumor detection. In contrast, US images have moderate resolution, lower contrast due to speckle noise, and are highly operator-dependent. Brain tumors are generally easier to segment than breast lesions due to their clearer boundaries in MRI, and more uniform background anatomy. Breast lesions, however, often exhibit irregular shapes, similar echogenicity

to glandular tissue in US, and are obscured by heterogeneous breast structures and artifacts like acoustic shadows. These factors (combined with MRI's consistent imaging conditions and reduced noise) make brain tumor segmentation more straightforward compared to the challenges posed by breast lesion segmentation in ultrasound.

Combining breast US datasets improved performance by improving Dice coefficient up to 0.7804, but standard augmentation had minimal impact, suggesting that data diversity is more critical than simple augmentation. Thereby, future work should explore larger datasets or adopt more realistic-like augmentation strategies such as AI-generated data to enhance generalization.

4.5 Conclusion

In this study, we evaluated multiple model architectures and training strategies for medical image segmentation. While cross-validation was initially employed, a shuffled train-test split proved more time-effective. Architectural enhancements, particularly attention mechanisms and multi-scale feature extraction (ASPP), improved the segmentation performance of the proposed framework. However, the modest results on breast lesion segmentation highlight the inherent difficulty of this task, due to limited training data and complex lesion characteristics compared to brain tumor segmentation demonstrating exceptional performance.

Composing a large US dataset through combining multiple US datasets improved results, suggesting that larger, more diverse datasets may be crucial for enhancing model robustness.

Future work should focus on advanced augmentation techniques, domain adaptation methods, and larger multi-institutional datasets to improve generalization. Additionally, investigating task-specific architectural adjustments could help bridge the performance gap between different medical segmentation challenges. Overall, while the proposed hybrid models show promise, further refinements are needed to achieve clinically viable performance in breast lesion segmentation.

GENERAL CONCLUSION

This thesis explored the application of deep learning techniques for breast lesion segmentation in ultrasound images, addressing the limitations of conventional CAD systems in handling noisy data, lesion variability, and limited annotated datasets. Through a comprehensive literature review and a methodologically grounded experimental study, the work evaluated and proposed state-of-the-art segmentation architectures, including U-Net variants, transformer-based models, and hybrid frameworks that combine the strengths of CNNs and attention mechanisms.

The proposed models demonstrated competitive performance in terms of Dice score and IoU, and offered improved robustness in segmenting complex and subtle lesion boundaries. The integration of preprocessing strategies, data augmentation, multiscale learning, and combined loss functions proved essential in optimizing performance, particularly in the presence of class imbalance and limited data.

Despite promising results, one of the persistent challenges highlighted in this study is the high computational cost required to implement advanced architectures. While many recent state-of-the-art models outperform simpler networks in segmentation accuracy, their real-world deployment is constrained by the need for powerful hardware and long training times—resources that are not always available in clinical or academic environments. This gap underscores the importance of developing efficient yet effective models that can balance performance and practicality.

Overall, this research contributes to the growing field of AI-driven medical imaging by offering insights, benchmarks, and practical considerations for deploying deep learning solutions in ultrasound-based breast cancer diagnosis. Future work may focus on lightweight models, domain adaptation techniques, and clinical validation to enhance the usability and reliability of such systems in real-world diagnostic settings.

REFERENCES

1. Ali, M., Magee, D., & Dasgupta, U. (2008). Signal processing overview of ultrasound systems for medical imaging. *SPRAB12, Texas Instruments, Texas*, *55*.https://www.ti.com/lit/wp/sprab12/sprab12.pdf
2. Almajalid, R., Shan, J., Du, Y., & Zhang, M. (2018, December). Development of a deep-learning-based method for breast ultrasound image segmentation. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1103-1108). IEEE.https://ieeexplore.ieee.org/abstract/document/8614204
3. Al-Masni, M. A., Al-Antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., ... & Kim, T. S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Computer methods and programs in biomedicine, 157, 85-94.https://www.sciencedirect.com/science/article/abs/pii/S0169260717314980
4. 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), 151.https://www.nature.com/articles/s41523-021-00358-x
5. Bankman, I. (Ed.). (2008). *Handbook of medical image processing and analysis*. Elsevier.https://books.google.dz/books?hl=en&lr=&id=AnRPBKb7qHUC&oi=fnd&pg=PP1&dq=+(Bankman,+2008).+image+segmentation&ots=mK2pKqEfz5&sig=fONA5I-3TrHCzvs30ZacWuisWHs&redir_esc=y#v=onepage&q=(Bankman%2C%202008).%20image%20segmentation&f=false
6. Beutel, J. (2000). Handbook of medical imaging (Vol. 3). Spie Press.https://books.google.dz/books?id=YKVULpCZ_iEC
7. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022, October). Swin-unet: Unet-like pure transformer for medical image segmentation. In the European conference on computer vision (pp. 205-218). Cham: Springer Nature Switzerland.https://link.springer.com/chapter/10.1007/978-3-031-25066-8_9https://arxiv.org/abs/2105.05537
8. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.https://arxiv.org/abs/2102.04306
9. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834-848.https://arxiv.org/pdf/1606.00915
10. Cho, S. W., Baek, N. R., & Park, K. R. (2022). Deep Learning-based Multi-stage segmentation method using ultrasound images for breast cancer diagnosis. Journal of King Saud University-Computer and Information Sciences, 34(10), 10273-10292.https://www.sciencedirect.com/science/article/pii/S1319157822003718
11. Davidson A (2023) Computer-Aided Diagnosis: Advancements in Medical Imaging and Healthcare. J Med Diagn Meth. 12:421.https://www.longdom.org/open-access/computeraided-diagnosis-advancements-in-medical-imaging-and-healthcare-103127.html#ai
12. Derakhshandeh, S., & Mahloojifar, A. (2025). Modifying the U-Net's Encoder-Decoder Architecture for Segmentation of Tumors in Breast Ultrasound Images. Journal of Imaging Informatics in Medicine, 1-15.https://link.springer.com/article/10.1007/s10278-025-01537-z
13. de Sitter, A., Visser, M., Brouwer, I., Cover, K. S., van Schijndel, R. A., Eijgelaar, R. S., ... & MAGNIMS Study Group and Alzheimer's Disease Neuroimaging Initiative. (2020). Facing privacy in neuroimaging: removing facial features degrades performance of image analysis

methods. European Radiology, 30, 1062-1074.https://link.springer.com/article/10.1007/s00330-019-06459-3
14. Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).https://arxiv.org/pdf/2010.11929/1000
15. Farhat, H., Sakr, G. E., & Kilany, R. (2020). Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. *Machine vision and applications*, *31*(6), 53.https://link.springer.com/article/10.1007/s00138-020-01101-5
16. Goel, N., Yadav, A., & Singh, B. M. (2016). Medical image processing: A review. *2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)*, 57-62.https://ieeexplore.ieee.org/abstract/document/7918737
17. Guo, Z., Xie, J., Wan, Y., Zhang, M., Qiao, L., Yu, J., ... & Yao, Y. (2022). A review of the current state of the computer-aided diagnosis (CAD) systems for breast cancer diagnosis. Open Life Sciences, 17(1), 1600-1611.https://www.degruyterbrill.com/document/doi/10.1515/biol-2022-0517/html
18. Han, S., Kang, H. K., Jeong, J. Y., Park, M. H., Kim, W., Bang, W. C., & Seong, Y. K. (2017). A deep learning framework for supporting the classification of breast lesions in ultrasound images. Physics in Medicine & Biology, 62(19), 7714.https://iopscience.iop.org/article/10.1088/1361-6560/aa82ec/metahttps://sci-hub.se/10.1088/1361-6560/aa82ec
19. 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).https://arxiv.org/abs/1512.03385
20. Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, *32*, 582-596.https://link.springer.com/article/10.1007/s10278-019-00227-x
21. Hosseini, M., & Dixon, B. E. (2016). Syntactic interoperability and the role of standards. In *Health Information Exchange* (pp. 123-136). Academic Press.https://academic.oup.com/jamia/article-abstract/23/2/317/2572401
22. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html
23. Huynh, N. (2023). Understanding evaluation metrics in medical image segmentation. Medium: Where good ideas find you.https://medium.com/@nghihuynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f
24. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, *18*(2), 203-211.http://nature.com/articles/s41592-020-01008-z.
25. Jannatdoust, P., Valizadeh, P., Saeedi, N., Valizadeh, G., Salari, H. M., Saligheh Rad, H., & Gity, M. (2025). Computer‐Aided Detection (CADe) and Segmentation Methods for Breast Cancer Using Magnetic Resonance Imaging (MRI). Journal of Magnetic Resonance Imaging.https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.29687
26. Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. Ieee Access, 6, 9375-9389.https://ieeexplore.ieee.org/document/8241753
27. Khosla, R. (2021). *Auto-Encoders for Computer Vision: An Endless World of Possibilities*.https://www.analyticsvidhya.com/blog/2021/01/auto-encoders-for-computer-vision-an-endless-world-of-possibilities/

28. Kim, J., Harper, A., McCormack, V., Sung, H., Houssami, N., Morgan, E., ... & Fidler-Benaoudia, M. M. (2025). Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nature Medicine*, 1-9.https://www.nature.com/articles/s41591-025-03502-3
29. LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e26cc4a1c717653f323715d751c8dea7461aa105
30. 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.https://www.sciencedirect.com/science/article/abs/pii/S1361841517301135https://sci-hub.se/10.1016/j.media.2017.07.005
31. 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).https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html
32. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89-94.https://www.nature.com/articles/s41586-019-1799-6https://sci-hub.se/10.1038/s41586-019-1799-6
33. Mildenberger, P., Eichelberg, M., & Martin, E. (2002). Introduction to the DICOM standard. *European radiology*, *12*, 920-927.https://link.springer.com/article/10.1007/s003300101100
34. Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1), 210.https://link.springer.com/article/10.1186/s13104-022-06096-y
35. 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*.https://arxiv.org/abs/1804.03999
36. Pramanik, P., Roy, A., Cuevas, E., Perez-Cisneros, M., & Sarkar, R. (2024). DAU-Net: Dual attention-aided U-Net for segmenting tumor in breast ultrasound images. Plos one, 19(5), e0303670.https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0303670
37. Pu, Q., Xi, Z., Yin, S., Zhao, Z., & Zhao, L. (2024). Advantages of transformer and its application for medical image segmentation: a survey. *BioMedical engineering online*, *23*(1), 14.https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-024-01212-4
38. Reinke, A., Maier-Hein, L., & Müller, H. (2021). Common limitations of performance metrics in biomedical image analysis. *Proceedings of the Medical Imaging with Deep Learning (MIDL 2021)*.https://arodes.hes-so.ch/record/9315?ln=fr&v=pdf
39. Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., ... & Sechopoulos, I. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. JNCI: Journal of the National Cancer Institute, 111(9), 916-922.https://academic.oup.com/jnci/article/111/9/916/5307077
40. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer international publishing.https://arxiv.org/abs/1505.04597

41. Sagheer, S. V. M., & George, S. N. (2020). A review on medical image denoising algorithms. *Biomedical signal processing and control*, *61*, 102036.https://www.sciencedirect.com/science/article/abs/pii/S1746809420301920
42. Santos, M. K., Ferreira, J. R., Wada, D. T., Tenório, A. P. M., Nogueira-Barbosa, M. H., & Marques, P. M. D. A. (2019). Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards precision medicine. Radiologia brasileira, 52(06), 387-396.https://www.scielo.br/j/rb/a/9yX6w83KDDT33m6G9ddCqBn/?lang=en
43. Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual review of biomedical engineering, 19(1), 221-248.https://www.annualreviews.org/content/journals/10.1146/annurev-bioeng-071516-044442
44. 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), 12495.https://www.nature.com/articles/s41598-019-48995-4
45. Shiraishi, J., Li, Q., Appelbaum, D., & Doi, K. (2011, November). Computer-aided diagnosis and artificial intelligence in clinical imaging. In Seminars in nuclear medicine (Vol. 41, No. 6, pp. 449-462). WB Saunders.https://www.sciencedirect.com/science/article/abs/pii/S0001299811000742
46. Sree, S. V., Ng, E. Y. K., Acharya, R. U., & Faust, O. (2011). Breast imaging: a survey. *World journal of clinical oncology*, *2*(4), 171.https://pmc.ncbi.nlm.nih.gov/articles/PMC3100484/
47. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).https://arxiv.org/abs/1409.4842
48. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, *15*, 1-28.https://link.springer.com/article/10.1186/s12880-015-0068-x
49. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.https://arxiv.org/abs/1706.03762
50. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).https://arxiv.org/abs/1807.06521
51. Wu, Z., Guo, K., Luo, E., Wang, T., Wang, S., Yang, Y., ... & Ding, R. (2024). Medical long-tailed learning for imbalanced data: Bibliometric analysis. Computer Methods and Programs in Biomedicine, 108106.https://www.sciencedirect.com/science/article/abs/pii/S0169260724001020
52. Yeasmin, M. N., Al Amin, M., Joti, T. J., Aung, Z., & Azim, M. A. (2024). Advances of AI in image-based computer-aided diagnosis: A review. Array, 100357.https://www.sciencedirect.com/science/article/pii/S2590005624000237
53. Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.https://arxiv.org/abs/1511.07122
54. Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.https://arxiv.org/abs/1511.07122
55. Zhang, C., Wang, L., Wei, G., Kong, Z., & Qiu, M. (2024). A dual-branch and dual attention transformer and CNN hybrid network for ultrasound image segmentation. Frontiers in Physiology, 15, 1432987.https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2024.1432987/full
56. Zhang, H., Lian, J., & Ma, Y. (2025). FET-UNet: Merging CNN and transformer architectures for superior breast ultrasound image segmentation. Physica Medica, 133, 104969.https://www.sciencedirect.com/science/article/abs/pii/S1120179725000791

57. Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5), 820-838.https://ieeexplore.ieee.org/abstract/document/9363915