

AUTOMATED DIAGNOSIS OF BREAST ULTRASONOGRAPHY IMAGES USING DEEP NEURAL NETWORKS

CS57-2

Final Report



THE UNIVERSITY OF
SYDNEY

5703 Group Based Capstone Project

Group Members

1. Young Jun Cho (490455715) [Group Leader]
2. Rutsy Arangcon (510674526)
3. Tianqi Liu (530643227)
4. Yajie Huang (530390011)
5. Ruyi Lu (440509448)
6. Neosh Sheikh (520566480)

School of Computer Science
Faculty of Engineering
The University of Sydney
Australia

15 November 2024

CONTRIBUTION STATEMENT

Our group, taking project CS57-2, with group members Young Jun Cho, Rutsy Arangcon, Yajie Huang, Ruyi Lu, Neosh Sheikh, and Tianqi Liu would like to state the contributions each group member has made for this project during this semester:

- Young Jun Cho: Led the team by organising team meetings and time/work planning. Reviewed literature for applying novel techniques and developed codes for the baseline model. Organised the subsections/Wrote the final report and contributed to the presentation and demo.
- Rutsy Arangcon: Designed the architecture for the proposed two-branch classifier model. Conducted a literature review of novel methods including the baseline model, ACSNet. Developed and implemented code for the two-branch classifier model on ResNet50. Tested the proposed model on ResNet50 with and without early stopping. Co-authored the proposal and final report.
- Yajie Huang: Conducted a literature review of potential models for the project. Developed and finalised the baseline code for the project. Incorporated EfficientNet with our two-branch classifier. Tested our proposed model with EfficientNet with and without early stopping. Co-authored the proposal and final report. Presented the results part of the demo and presentation.
- Ruyi Lu: Conducted the literature review to understand the use of deep learning models in the medical ultrasound images field. Contributed to the presentation and final report. Developed codes to incorporate DenseNet into our two-branch classifier.
- Neosh Sheikh: Developed the code for the baseline ACSNet model (All the missing modules / initial structure), and conducted literature reviews to develop a better understanding of the project (ResNet, multi-task learning / CNNs). Co-authored the proposal and final report. Presented the discussion and conclusion for the presentation and demo.
- Tianqi Liu: Reviewed existing research in automatic cancer diagnosis algorithms. Developed the code for applying the Swin Transformer in the two-branch classifier. Contributed to the final report and presentation.

All group members agreed on the contributions listed in this statement by each group member.



The image shows four handwritten signatures arranged horizontally. From left to right: 1) A signature in black ink, followed by the name "Young Jun Cho" and the ID "490455715" in smaller text below it. 2) A signature in black ink, followed by the name "Tianqi Liu" and Chinese characters "刘天琪" written vertically below it. 3) A signature in black ink, followed by the name "Yajie Huang" and Chinese characters "黄雅婕" written vertically below it. 4) A signature in black ink, followed by the name "Rutsy Arangcon" and a small logo consisting of a stylized 'R' and 'A' above the text.

FIGURE 0.1: Group member signatures

Abstract

Breast cancer is the second most common cancer and early, accurate diagnosis is crucial for effective treatment. Traditional diagnostic methods are effort-intensive and rely heavily on the experience of radiologists, leading to varying outcomes. To aid in detection and diagnosis, previous studies have utilised deep learning to perform image segmentation to determine the location and morphology of tumours and classification to identify malignant tumours. Some multi-task models like ACSNet use Convolutional Neural Networks (CNNs) to perform both segmentation and classification, whereby optimising the segmentation performance also improves classification accuracy. In this study, we improve upon the baseline ACSNet model by developing a two-branch classifier. The first branch utilises ACSNet's multi-scale feature extraction and channel attention mechanism, while the second branch utilises an off-the-shelf classification model. The modular architecture allows for the second branch to be replaced with various state-of-the-art models such as ResNet50, DenseNet, EfficientNet, and SwinTransformer. Our experiments on the breast ultrasound image (BUSI) dataset show that our proposed architecture outperforms the baseline ACSNet model and that among the off-the-shelf classifiers that we used on our two-branch classifier, EfficientNet returned the best results.

Contents

CONTRIBUTION STATEMENT	ii
Abstract	iv
Contents	v
Chapter 1 Introduction	1
Chapter 2 Related Literature	3
2.1 Literature Review.....	3
2.1.1 ACSNet	4
2.1.2 ResNet	5
2.1.3 Swin Transformer	6
2.1.4 EfficientNet.....	7
2.1.5 DenseNet	8
Chapter 3 PROJECT PROBLEMS	10
3.1 Project Aims & Objectives.....	10
3.2 Project Questions	11
3.3 Project Scope.....	11
Chapter 4 METHODOLOGIES	12
4.1 Methods	12
4.1.1 Model Architecture Design & Operation	12
4.1.2 Ablation Studies.....	13
4.1.2.1 ResNet.....	14
4.1.2.2 Swin Transformer.....	14
4.1.2.3 EfficientNet	14
4.1.2.4 DenseNet	15
4.1.3 Training Process.....	15
4.1.4 Model Evaluation.....	16
4.2 Data Collection.....	18

4.3 Data Analysis	18
4.4 Deployment.....	19
4.5 Testing	19
Chapter 5 RESOURCES	20
5.1 Hardware & Software	20
5.2 Materials.....	21
5.3 Roles & Responsibilities.....	21
Chapter 6 MILESTONES SCHEDULE	25
6.1 Milestones	27
6.1.1 Literature Review & Architecture Design	27
6.1.2 Data EDA/Pre-processing	27
6.1.3 Modelling	28
6.1.4 Performance Optimisation.....	28
6.1.5 Report & Presentation Preparation	28
6.2 Schedule Analysis	28
Chapter 7 RESULTS	29
7.1 Interpretation of results.....	31
Chapter 8 DISCUSSION	34
8.1 Comparison with ACSNet	34
8.2 Implications.....	35
8.2.1 Multi-task learning for medical imaging and diagnosis	35
8.2.2 Significance of introducing early stopping to ensure the reliability of the diagnostic system.....	36
8.2.3 One-shot classification / multi-task learning	37
8.2.4 Extensibility of the modular two-branch classifier.....	37
8.2.5 Tackling other medical imaging tasks	38
Chapter 9 LIMITATIONS AND FUTURE WORKS	39
9.1 Datasets - Transferability and lack of validation set.....	39
9.2 Reducing classification bias.....	39
9.3 Improving feature fusion / interaction.....	40
Bibliography	41

Appendix A Appendix A	45
A1 Links to the code and results	45
A2 Loss / Segmentation performance per fold	45

CHAPTER 1

Introduction

Breast cancer is the second most prevalent cancer in the world and is the number one cancer among women ('Breast cancer statistics', 2022), with one in 48 women in developing countries dying from the disease ('Cancer', 2022). However, patient survival is greatly improved with early detection and treatment. Traditional methods of diagnosis include imaging techniques like MRI, ultrasound, PET, and CT scans. Molecular analysis of biomarkers obtained through biopsy is also used as an adjunct approach (Z. He et al., 2020). Among these techniques, ultrasound is often used as the initial approach as the procedure is accessible, non-invasive, and low-risk.

However, there are challenges to diagnosis via ultrasound imaging. This technique relies heavily on the proficiency, skills, and experience of radiologists and sonographers (Andrade et al., 2023). Different skill levels and expertise often result in varying speed and accuracy of diagnosis. Manual analysis of ultrasound images is labour-intensive, and fatigue can play a part in causing subtle irregularities in the images to be overlooked.

Many leading medical image models achieve impressive accuracy in classifying tumors, but they rely on practitioners to first manually identify the tumor regions. This presents challenges, particularly when tumor boundaries are unclear, and introduces variability based on the practitioner's skill.

ACSNet, a recent innovation by Q. He et al. (2024), attempts to address this by simultaneously segmenting (identifying tumor regions) and classifying tumors within a single model. While promising, ACSNet's classification accuracy still falls short of models that utilize pre-identified tumor regions.

Our research introduces a modified ACSNet architecture that leverages separate, powerful convolutional neural networks (CNNs) like EfficientNet and ResNet. This modification

enhances classification performance, surpassing both the original ACSNet and the individual base CNNs.

The model proposed in the study shows a path towards a powerful multi-task model that can efficiently diagnose breast cancer with no practitioner input, allowing diagnosis of this life threatening disease at scale.

CHAPTER 2

Related Literature

2.1 Literature Review

Computer-aided diagnosis (CAD) emerged in the mid-1980s as a tool to help doctors focus on areas of interest within images (Giger et al., 2008). In the late 1990s, artificial neural networks started to be used for breast cancer diagnosis (Mehdy et al., 2017). The usage of deep learning has constantly evolved since. In the last few decades, convolutional neural networks (CNN) have gained traction in medical imaging, supported by increased capacity for computational processing and the development of new algorithms (Pinto-Coelho, 2023).

CNNs are used in medical imaging to perform the following tasks: 1) classification, which provides high-level diagnosis suggestions, 2) segmentation, which aids in defining the morphology and fine-grained boundaries of objects of interest (e.g. outlining the shape and contours of tumours), 3) detection, which helps identify regions of interest (e.g. determining the location of a tumour), and 4) reconstruction, which helps enhance the quality of the scanned image (Wang et al., 2021).

Deep learning models in medical diagnosis often focus on performing one task. For example, CNN architectures like VGG and AlexNet are often used to perform classification tasks. Inception and ResNet are used for feature extraction ((Kim et al., 2022)). Breast-NET, developed by Saha et al. (2024) is used for breast cancer detection.

However, certain novel techniques aim to accomplish multiple tasks. For example, BI-RADS-NET-V2 is a model that combines breast cancer diagnosis with a semantic and quantitative explainer (Zhang et al., 2023). Another novel technique is a joint segmentation and classification model based on Y-Net, a combination of a U-Net with a classifier network (Byra et al., 2022). Or more recently, ACSNet, which is a multi-task model that achieves simultaneous segmentation and classification (Q. He et al., 2024).

Still, more novel techniques use CNN in combination with transformers. Originally used for natural language processing, transformers have recently crossed over to expand their usage in image classification as well. Meng et al. (2024) introduced a breast ultrasound image classifier that utilised CNN and a multi-scale transformer.

In our proposed model, we use the ACSNet model as our baseline, however, we expand this architecture to introduce a two-branch classifier, the second branch of which is a modular implementation of a state-of-the-art classifier. In our ablation studies, we explore the contribution of this second branch by replacing it with various classifiers such as ResNet50, Swin Transformer, EfficientNet, and DenseNet.

2.1.1 ACSNet

The baseline model used in this study is ACSNet, developed by Q. He et al. (2024) to perform classification and segmentation at once. The segmentation network is based on an enhanced implementation of U-Net architecture with a gating mechanism, while the classification network utilises input features from the U-Net and aggregates these via multi-scale feature fusion and channel attention mechanism.

Fig 2.1 shows the architecture of the ACSNet model. The U-Net segmentation network consists of an encoder (stages 1 through 5) and a decoder (stages 6 through 9). The encoder performs down-sampling to extract image features, while the decoder performs up-sampling to restore the feature maps to the original input size. Each stage in these encoding and decoding paths are convolution operations followed by group normalisation (GN) and ReLU. Gating mechanisms at each skip connection utilise Global Average Pooling (GAP) and Global Max Pooling (GMP) to optimise the retention of key features. Meanwhile, the Deformable Spatial Attention Module (DSAModule) is introduced in stages 5, 6, and 7 to address over- and under-segmentation due to variations in tumour shapes and sizes. Finally, the output of the stage 9 is the tumour segmentation result.

The classification branch uses the feature maps from stages 4, 5, and 6 of the segmentation network as inputs. A channel attention (CA) mechanism is used to enhance the selection of features. The features are then concatenated and passed through Global Average Pooling (GAP), fully connected (FC) layers, and finally a Softmax layer for classification. The classification gives a probability of the tumour image as being benign or malignant.

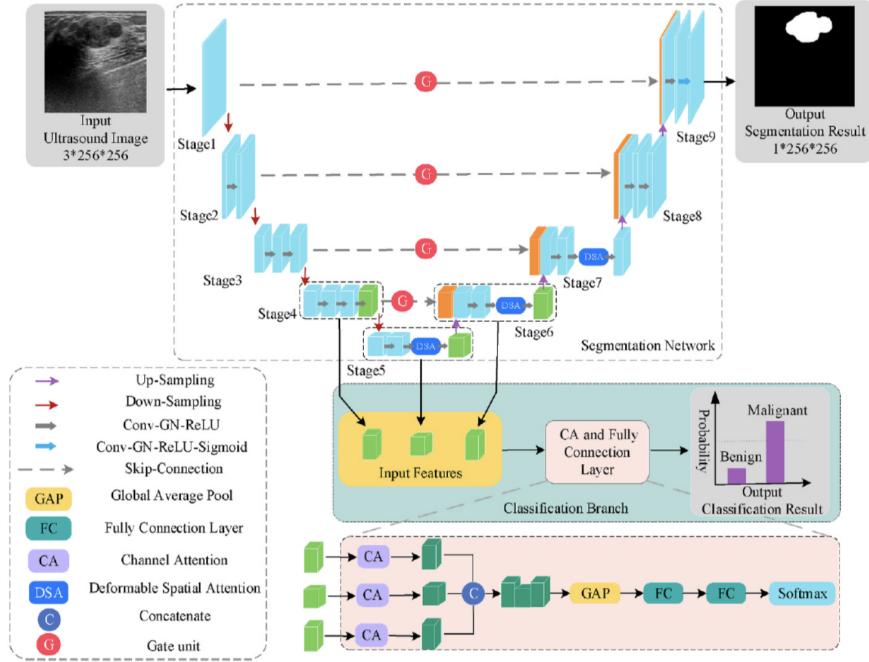


FIGURE 2.1: Overview of the ACSNet model. Image obtained from the study done by Q. He et al. (2024). The multi-task model accepts the ultrasound image as input and returns two outputs: a tumour segmentation map and a classification probability.

2.1.2 ResNet

Residual Networks (ResNets) are a type of deep convolutional neural network (CNN) designed to address a key challenge in deep learning models: the vanishing gradient problem (Alzubaidi et al., 2021; Luo et al., 2023). Introduced in 2015, ResNets utilise shortcut or residual connections—which bypass certain layers—to retain important features and enhance gradient flow during backpropagation (Shafiq & Gu, 2022). CNNs, particularly ResNets, are highly efficient in automatically extracting relevant features from ultrasound data, which can then be used to enhance spatial and contrast resolution in imaging (Song et al., 2024). This architecture enables the training of much deeper networks, allowing models to learn complex representations efficiently.

Due to their high efficiency, ResNets are especially useful in medical imaging, particularly in the diagnosis of ultrasound images within the super-resolution imaging field (Kang et al., 2024). Ultrasound images are often degraded by low contrast, noise, and anatomical variability. ResNets provide advantage in analysing these images by isolating and classifying intricate elements, which improves the precision of essential features. According to Xu et al. (2023), ResNets are effective across various medical image-processing tasks due to their

depth and accuracy. For instance, in a study applying different deep-learning models to classify ultrasound images for breast cancer, ResNet50 achieved an accuracy of 85.4%, while ResNeXt50 reached 85.83% on the BUSI dataset (Zakareya et al., 2023). With their capacity to model complex patterns and distinguish subtle differences, ResNets have achieved high accuracy in diagnosing various medical conditions.

In segmentation tasks, such as delineating regions of interest like organs or lesions, ResNet-based architectures—including ResNet-50 and ResNet-U-Net—have demonstrated high accuracy and superior results (Alim et al., 2024; Gökmen Inan et al., 2024). These hybrid models, which combine feature extraction with segmentation networks, provide precise delineation of anatomical structures. ResNets have shown promising outcomes in segmenting ultrasound images, tumours, and cardiac chambers, achieving high sensitivity and specificity. Moreover ACSNet by Q. He et al. (2024), also uses a ResNet18 base within it's U-Net.

2.1.3 Swin Transformer

Swin Transformer is an innovative architecture that adapts the traditional Transformer model, originally prominent in natural language processing for vision-based tasks (Li et al., 2023) (Liu et al., 2021). This adaptation modifies the Transformer structure to better handle the intricacies of image data (Ahmad et al., 2024). Unlike traditional convolutional neural networks (CNNs), which utilise convolutional kernels to extract local features and operate within a limited receptive field, Swin Transformer adopts a more flexible approach.

Swin Transformer overcomes the limitation of localised feature extraction in CNNs through its window-based self-attention mechanism. This mechanism divides the image into non-overlapping windows and computes self-attention within each window, enabling the model to extract local features efficiently while managing computational costs. The attention can be calculated by:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right)V$$

where Q , K , V are the query, key, and value matrices, and d_k is a scaling factor to stabilise the gradient.

To expand the receptive field, Swin Transformer introduces a shifted window mechanism that can shift the windows in successive layers, facilitating the integration of cross-window information. This unique strategy allows the model to capture both local and global context, which is essential for complex vision tasks such as medical image analysis (Davri et al., 2023)(Iqbal & Sharif, 2023).

Additionally, the multi-head self-attention mechanism (MHSA) also allows Swin Transformer to focus on multiple aspects of the input by using different attention heads:

$$MHSA(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O$$

where each head head_i can be computed as:

$$\text{head}_i = \text{Attention}(Q, K, V)$$

By these methods, Swin Transformer can combine global and local feature extraction capabilities, along with computational efficiency, making it highly effective for high-resolution image processing. Its structured approach and scalable design enable detailed recognition and analysis, which can make further contributions to supporting medical professionals in making more informed diagnostic decisions (Shamshad et al., 2023)

2.1.4 EfficientNet

EfficientNet is a state-of-the-art convolutional neural network (CNN) known for its application in various tasks, including medical image classification. It features an efficient scaling mechanism that allows it to achieve high accuracy while keeping lightweight, computational costs low and fewer network parameters than other CNN models (Zulfiqar et al., 2023).

The baseline model, EfficientNet-B0, was developed using neural architecture search and is built upon the ConvNet framework. The compound scaling method introduced by Tan and Le (2019) allows EfficientNet to maximise accuracy by using a compound coefficient to jointly scale the network's depth, width, and resolution in a balanced manner, rather than scaling each dimension independently.

The core components of EfficientNet include the MBConv module, inspired by MobileNet, and the Squeeze-and-Excitation (SE) module, which enhances feature extraction by re-weighting channels based on importance. Except for the final fully connected layer, all layers in EfficientNet use Batch Normalisation followed by the Swish activation function for improved performance. Down-sampling occurs at each stage through depth-wise convolutions, and a global average pooling layer is applied before the final fully connected layer, followed by a Softmax layer for classification output (Hoang & Jo, 2021).

In ImageNet classification tasks, EfficientNet outperforms models such as ResNet and DenseNet with fewer parameters and floating-point operations per second (FLOPS) (Tan & Le, 2019). Due to its advantages in computational efficiency, EfficientNet has inspired several variants, including EfficientNet B1-B7, which further improve the model's performance and generalisation ability. We can choose suitable EfficientNet variants based on factors such as dataset size, available resources for model training and evaluation, and model depth (Zulfiqar et al., 2023).

2.1.5 DenseNet

DenseNet121 is a convolutional neural network that includes four dense blocks Huang et al., 2016. Fig 2.2 shows the DenseNet architecture. In DenseNet121, each layer receives input from all preceding layers. For instance, layer L3 concatenates output from L0, L1 and L2 and passes them as input for L3. This design makes the advantage of feature reuse, where the feature maps learned by previous layers are accessible by all subsequent layers, resulting in more efficient learning.

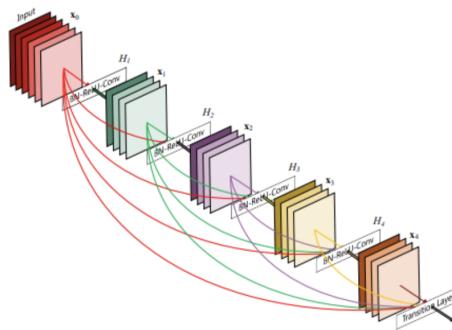


FIGURE 2.2: DenseNet Architecture

Between each dense block, transition layers are used between each dense block and each transition layer consists of a batch normalisation layer, an 1×1 convolutional layer followed by a 2×2 average pooling layer. The 1×1 convolutional layer reduces the number of channels and the 2×2 average pooling layer reduces the spatial sizes.

An important advantage of DenseNet is its parameter efficiency. This architecture requires fewer parameters compared to traditional convolutional neural networks because the model reuses features learned by earlier layers rather than learning redundant features. Lastly, the model improves information flow as it creates direct connections between all layers with matching feature-map sizes.

CHAPTER 3

PROJECT PROBLEMS

This project investigates the use of deep neural networks to create an automated diagnosis system for breast ultrasonography images capable of accurate one-shot classification.

3.1 Project Aims & Objectives

The primary aim of this project is to develop an automated system that utilises deep learning techniques to accurately diagnose breast tumours from ultrasound images in one shot. The proposed system will assist medical professionals by automatically identifying and classifying tumours as benign, malignant, or normal, thereby providing essential information that aids in the diagnostic process. Furthermore, by automating the analysis of ultrasound images, the system aims to reduce the time and cost associated with manual interpretation, while enhancing diagnostic accuracy.

To achieve the aforementioned goals, the project focuses on the following specific objectives:

- **Accurate Segmentation and Classification:** Develop a state-of-the-art deep learning model that precisely segments the tumour regions from breast ultrasound images and classifies the tumour as benign or malignant. This will enhance the diagnostic accuracy and reliability of the system.
- **Utilisation of ACSNet Architecture:** Implement an advanced architecture, namely ACSNet, which integrates 2-branch classifier modules. This architecture is designed to enhance both segmentation and classification performance by focusing on key image features and reducing irrelevant information through a channel attention mechanism.
- **System Automation and Reliability:** Build a fully automated diagnostic system that medical professionals can rely on for consistent and accurate results. The system will

provide a second opinion that complements the expertise of radiologists, reducing their workload and improving overall service quality in clinical settings.

- **Improved Diagnostic Tools for Medical Professionals:** The final objective is to offer an additional tool for healthcare providers, enabling more efficient diagnostic processes. By incorporating the ACSNet and classifier modules, the system aims to outperform existing models and provide a higher level of trust and accuracy, ultimately leading to better patient outcomes and optimised medical workflows.

3.2 Project Questions

The primary question this project aims to address is “**How accurately can the deep learning model classify benign, malignant, and normal tumours from breast ultrasound images?**”. This question explores how well the model can learn and interpret image features such as tumour size, shape, and location.

3.3 Project Scope

This project utilises the breast ultrasound image (BUSI) dataset to analyse breast ultrasound images. The dataset includes ultrasound images of benign, malignant, and normal tumours.

The project focuses on detecting the presence of tumours and determining their malignancy from breast ultrasound images. The model is specifically designed to classify three categories: benign, malignant, and normal. Diagnosing multiple organs or other diseases is beyond the scope of this project.

CHAPTER 4

METHODOLOGIES

4.1 Methods

4.1.1 Model Architecture Design & Operation

Fig. 4.1 shows the architecture of the proposed two-branch classifier. The backbone of our proposed model is the ACSNet, with its segmentation and classification networks. However, we extend the classification network into two branches, with the second branch being a modular implementation of a state-of-the-art classifier.

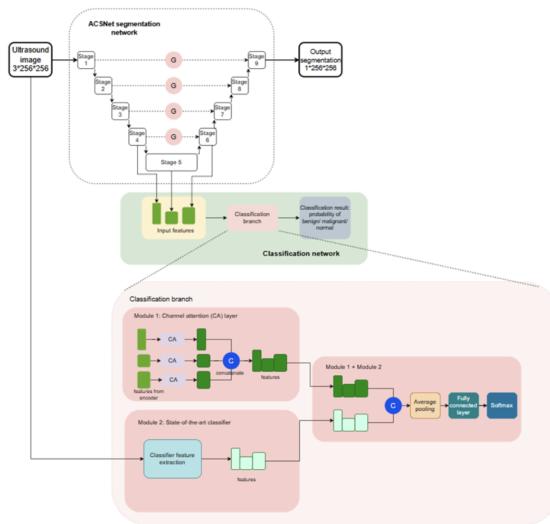


FIGURE 4.1: Architecture of the proposed two-branch classifier model. We extend the backbone ACSNet model by splitting the classification network into two branches. The first branch consists of the channel attention layer, while the second branch is a modular state-of-the-art classifier. Features extracted from both branches are combined for the final classification.

Just as it is with ACSNet, our proposed model takes in the ultrasound image as input to produce two outputs: a segmentation mask and a classification probability. However, for the purpose of this study, we will focus more on the classification performance of the model.

Our model retains the features of the ACSNet segmentation network, including the gating mechanisms and Deformable Spatial Attention Module (DSAModule). Ultrasound images with dimensions of 3 channels * 256 pixels * 256 pixels is taken in as input to the model. These undergo feature extraction via downsampling in the encoder path in stages 1 through 5 and restoration of feature maps via upsampling in the decoder path in stages 6 through 9, with convolutions performed at each stage. Gate units in the skip connections minimise noise as context information is sent from decoder to encoder. Meanwhile, DSAModule in stages 5 through 7 allow for prioritisation of more salient regions in feature maps of the irregularly-shaped tumour images. The final segmentation output is an image mask that is generated in the final stage, i.e. stage 9 of the U-Net.

Where our model differs from the baseline ACSNet model is in the classification network. Our classification network is split into two branches. The first branch consists of the channel attention (CA) mechanism. Feature maps from stages 4, 5, and 6 of the segmentation network are taken as inputs. Channel attention using global average pooling and global max pooling is applied. Then, the multi-scale features are combined.

The second branch of the classification network utilises a state-of-the-art classifier to obtain features from the original input image while leaving out the final fully connected layer. As this second branch is designed to be modular, we can replace the classifier used in this second branch with any state-of-the-art classifier. We explored the use of several state-of-the-art classifiers for this branch, including ResNet50, EfficientNet, DenseNet, and Swin Transformer.

The features obtained from the first and second branches of the two-branch classifier are then concatenated. The fused features are passed through a final fully connected layer and a Softmax layer to perform the final classification. The output of the two-branch classifier is a probability of the image to be classified as benign, malignant, or normal.

4.1.2 Ablation Studies

For our ablation studies, we replaced the second branch of our two-branch classifier with various state-of-the-art classifier algorithms. We conducted experiments using ResNet, Swin Transformer, EfficientNet, and DenseNet.

4.1.2.1 ResNet

For this study, a ResNet50 classifier pretrained on ImageNet-1k was used as the classifier plugin. This classifier, consisting of 50 convolution layers, addresses the vanishing gradient problem while allowing a deeper network. Moreover, it allows for faster convergence due to skip connections.

4.1.2.2 Swin Transformer

In this study, the Swin Transformer was employed as a branch of the classifier. This branch harnesses the Swin Transformer's ability to effectively process global and local contexts by leveraging its shifted window mechanism. The output from the Swin Transformer, which includes a flattened vector of features after average pooling, is then concatenated with features extracted from the other branch. This combined feature vector is subsequently used to feed a sequence of fully connected layers that culminate in the classification output. This integration allows the classifier to benefit from the deep, patch-based contextual insights provided by the Swin Transformer, alongside more traditional feature extraction methods, offering a robust solution for the classification tasks.

4.1.2.3 EfficientNet

In this study, we applied EfficientNet-B0 as the replacement of a branch of the two-branch classifier. To incorporate EfficientNet-B0 into our model, we utilised the pretrained version of EfficientNet-B0 on ImageNet-1k. The final fully connected classification layer of EfficientNet-B0 was removed to use the network purely as a feature extractor. After removing this layer, the extracted features from EfficientNet-B0 were flattened and concatenated with the output features from the first branch of our model. This combined feature vector was subsequently passed through fully connected layers to produce the final classification output. The integration of EfficientNet-B0 allowed the model to take advantage of its efficient compound scaling, which maintains high accuracy without significantly increasing the model's parameter count. This setup was particularly beneficial for processing complex image data in a computationally efficient way, making it an optimal choice for medical image analysis tasks where both accuracy and efficiency are critical.

4.1.2.4 DenseNet

DenseNet121 was used as one of the classifiers in the second branch of our two-branch classifier model. It is a densely connected convolutional network that comprises 121 layers, allowing for the combination of feature maps from different layers. The dense connectivity improves feature reuse, increasing variation in the input and improving the overall efficiency.

We removed the final fully connected layer from DenseNet121 to incorporate DenseNet121 to our two-branch classifier model. The feature maps obtained from the model then concatenated the features from the first branch of the model, followed by fully connected layers to make classification predictions of the images. In this approach, it makes the most of DenseNet121's feature extraction capabilities to achieve better classification results.

4.1.3 Training Process

To facilitate as close a comparison as possible between the baseline ACSNet model and our proposed model, our model was trained at 100 epochs using 5-fold cross-validation. However, to explore whether we can achieve better performance with fewer epochs, early stopping was also applied as part of the experiment. For early stopping, we set the patience to 7.

To train the model, a multi-task loss function was utilised, which is a combination of classification loss and segmentation loss.

Also, binary cross-entropy was employed to measure the classification loss, which is the difference between the actual class of the image against the predicted classification. Classification loss is expressed as:

$$L_{cls} (p^k, g^k) = -\frac{1}{N} \sum_{i=1}^N (g^k \log p^k + (1 - g^k) \log (1 - p^k))$$

where g_k is the ground truth class of sample k and p^k is the predicted classification probability for sample k .

Segmentation loss is calculated using the Dice coefficient and LossNet. Dice coefficient tells us how similar the predicted mask is from the ground truth. A perfect similarity is given by a Dice coefficient of 1. In this case, segmentation loss would be

$$L_{dice}(P_{seg}, Y_{seg}) = 1 - \frac{2P_{seg}Y_{seg} + 1}{P_{seg} + Y_{seg} + 1}$$

where L_{dice} is the Dice segmentation loss, P_{seg} is the predicted segmentation map, and Y_{seg} is the segmentation ground truth. Additional segmentation loss is calculated via the LossNet function introduced by Zhao et al. (2021). LossNet compares the difference between the predicted multi-scale features against the multi-scale features of the ground truth. LossNet is calculated as:

$$L_{LossNet} = l_{LossNet}^1 + l_{LossNet}^2 + l_{LossNet}^3 + l_{LossNet}^4,$$

$$l_{LossNet}^i = \|F_P^i - F_G^i\|_2, \quad i = 1, 2, 3, 4$$

where F_P^i and F_G^i are the i -th level feature maps obtained from the prediction and ground truth respectively and $l_{LossNet}^i$ is the Euclidean distance between these.

Total segmentation loss is therefore:

$$L_{seg}(L_{dice}, L_{LossNet}) = L_{dice} + \gamma \times L_{LossNet},$$

where $\gamma \in [0, 1]$ is the $L_{LossNet}$ weight for segmentation loss.

Linear combination of the classification and segmentation loss gives us the total multi-task loss for the entire model:

$$L_{total} = \lambda L_{cls} + (1 - \lambda) L_{seg}$$

where L_{total} is the multi-task loss for the model, and $\gamma \in [0, 1]$ is the classification task weight.

Minimising classification and segmentation loss work synergistically to optimise the entire model.

4.1.4 Model Evaluation

The performance of the model was evaluated on multiple fronts, including both segmentation and classification tasks.

Performance Metrics: For the classification branch, metrics such as **accuracy**, **precision**, **recall**, and **F1-score** were used to evaluate the model's ability to correctly classify tumours as benign, malignant, or normal.

- **Accuracy:**

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- **Precision:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Recall:**

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1-score:**

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Here, TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

For the segmentation branch, metrics like **Dice coefficient** and **Jaccard index** were employed to measure the similarity between predicted segmentation masks and the ground truth.

- **Dice Coefficient:**

$$\text{Dice} = \frac{2 \times |A \cap B|}{|A| + |B|}$$

- **Jaccard Index:**

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|}$$

Here, A is the set of pixels in the predicted mask, and B is the set of pixels in the ground truth mask.

Cross-Validation Results: Each fold of the five-fold cross-validation process provided insights into how well the model generalised to unseen data. The final model was selected based on the average performance across all folds.

Model Checkpointing: Throughout training, model checkpoints were saved whenever a better performance (in terms of validation loss) was achieved. This ensured that the best-performing model could be retrieved even if the training process was interrupted.

4.2 Data Collection

The dataset used in this study is obtained from the Breast Ultrasound Image (BUSI) dataset shared by Hesaraki (2018) on Kaggle. However, the provenance of the original dataset is the publicly available BUSI dataset published by Al-Dhabayni et al. (2020). The dataset is collected from the Baheya Women's Cancer Early Detection and Treatment Hospital located in Cairo, Egypt. The dataset consists of images and image masks of breast ultrasound scans obtained using the LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound system.

The dataset consists of 780 images obtained from 600 female patients aged between 25 and 75 years. The images are stored as PNG image files and categorised into three classes: 437 benign cases, 210 malignant masses, and 133 normal cases. Annotations were provided by radiologists at Baheya Hospital. Ground truths for the tumour masks delineating the shape of the tumours from the images were provided by Al-Dhabayni et al. (2020).

The original dataset had issues, which were pointed out by Pawłowska et al. (2023). These included duplicate images, visible biopsy needles, and images that were not from the breast itself but from the axilla. Such issues were resolved by removing duplicates and removing the images that contained extraneous objects. Meanwhile, the images that were taken from the axilla were retained, as these contained images of tumours that, while not directly in the breast, were still in the general surrounding area adjacent to the breast and may therefore still be due to breast cancer. These issues were resolved in the dataset shared by Hesaraki (2018), which is the dataset that we used in our study.

4.3 Data Analysis

During exploratory data analysis, we found that there were some images that contained one or more tumours. We created separate files for each image mask to address this circumstance. Through visual inspection, we determined that much like how it is in real-world experience, the dataset is imbalanced in that there are more images of benign tumours than there are of malignant ones. We applied data transformation and augmentation to help balance the classes by increasing variations in the images. We implemented data augmentation to the images by performing random cropping, random sized cropping, sliding cropping, as well as centre

cropping. We randomly rotated and randomly flipped the images horizontally and vertically. We also applied both scaling and free scaling, as well as random sizing and resizing.

4.4 Deployment

No deployment was done - the output of this project was a study for the client.

4.5 Testing

No traditional testing was done - the output of this project was a study for the client.

That said, the testing that we conducted as part of the study are as follows: we conducted two sets of testing, one set was run at 100 epochs without early stopping, while another set incorporated early stopping. Each set of tests was run for the four state-of-the-art classifiers that we explored, i.e. ResNet50, Swin Transformer, DenseNet, and EfficientNet.

For each test, we obtained the classification and segmentation metrics for the overall model. For classification performance, we noted the accuracy, precision, recall, and F-1 score. Meanwhile, for segmentation performance, we collected the Dice and Jaccard coefficients. To visually inspect the model's performance, we also plotted the training and validation loss, as well as the Dice and Jaccard scores.

CHAPTER 5

RESOURCES

5.1 Hardware & Software

Our models were run on both Google Colab and Kaggle instances, as developing deep neural networks (DNNs) requires powerful GPUs.

The hardware specifications are as follows:

- **CPU:** Intel(R) Xeon(R) CPU @ 2.20GHz & 12th Gen Intel(R) Core(TM) i5-12500H @ 2.50 GHz
- **GPU:** T4
- **RAM:** 12.65 GB (Google Colab) & 16 GB (Kaggle)

The software specifications are as follows:

- **Platform:** Google Colab & Pycharm 2024.1 & Kaggle
- **OS:** Ubuntu 22.04.3 & Windows 11 & MacOS 14.4.1
- **Language:** Python 3.10.12
- **Packages:**
 - torch: 2.2.1+cu121
 - numpy: 1.25.2
 - pandas: 2.0.3
 - matplotlib: 3.7.1
 - PIL (Pillow)
 - scikit-learn

5.2 Materials

- **Dataset:** Breast Ultrasound Images Dataset (BUSI) (Hesaraki, 2018) on Kaggle is used for training and evaluating deep learning models. The original source is from the paper published by Al-Dhabayani et al. (2020).
- **Research papers and articles:** Our baseline model, ACSNet, is based on the article “Multi-task learning for segmentation and classification of breast tumours from ultrasound images” (Q. He et al., 2024),” which utilises both segmentation and classification techniques for diagnosing breast tumours. Additionally, we’ve explored other articles that employ deep learning neural networks for diagnosing breast cancer ultrasound images. These resources provide valuable insights into methodologies and performance metrics, guiding the development and optimisation of the models used in our project.
- **Development tools:**
 - **GitHub:** A web-based platform for collaboration on code and version control, ensuring effective management of code changes.
 - **Google Colab:** A web-based editor that can run Python notebooks in the cloud.
 - **Kaggle:** A collaborative environment for users to find and publish datasets, and share and build models.
- **Communication channel:**
 - **Slack/WeChat:** Platforms for effective communication regarding project-related tasks. These channels keep everyone on the same page and on track of the project’s progress. Announcements and important information are posted by the team lead to keep all members informed. In addition, these serve as a centralised space for sharing relevant articles and resources, fostering active participation and collaboration.

5.3 Roles & Responsibilities

For the whole semester, each member’s roles and responsibilities are listed in the table below:

Table 5.1: Roles and responsibilities for each team member.

Members	Roles and Responsibilities
Young Jun Cho	<ul style="list-style-type: none"> ● Group leader, acting as the primary communicator between clients/tutors and team members. ● Organised client and group meetings, and allocated tasks amongst team members. ● Conducted an extensive literature review to understand the uses of deep learning neural methods in the medical ultrasound image field mainly for classification tasks. ● Developed codes for the image segmentation task. ● Brainstormed upgrades for our baseline model (ACSNet) including potential architectures and parameter modifications. ● Wrote the final report and prepared the presentation.
Rutsy Arangcon	<ul style="list-style-type: none"> ● Conducted an extensive literature review to understand the uses of deep learning neural methods in the medical ultrasound image field mainly for classification tasks. ● Brainstormed ways on how to upgrade our baseline model (ACSNet) including architectures and parameters that can be changed. ● Designed the two-branch classifier method by enhancing our baseline model (ACSNet). ● Incorporated ResNet50 as the classifier for the two-branch classifier, with results used in the ablation study. ● Wrote the final report and prepared the presentation.

Continued on next page

Table 5.1 – continued from previous page

Members	Roles and Responsibilities
Tianqi Liu	<ul style="list-style-type: none"> • Conducted an extensive literature review to understand the uses of deep learning neural methods in the medical ultrasound image field mainly for classification tasks. • Developed codes for the image segmentation task. • Brainstormed ways on how to upgrade our baseline model (ACSNet) including architectures and parameters that can be changed. • Incorporated Swin Transformer as the classifier for the two-branch classifier, with results used in the ablation study. • Wrote the final report and prepared the presentation.
Yajie Huang	<ul style="list-style-type: none"> • Developed our baseline model including the entire data pipeline for processing and analysis, and resolved all the debugged issues to get the code running. • Conducted an extensive literature review to understand the uses of deep learning neural methods in the medical ultrasound image field mainly for classification tasks. • Brainstormed ways on how to upgrade our baseline model (ACSNet) including architectures and parameters that can be changed. • Incorporated EfficientNet for the two-branch classifier, with results used in the ablation study. • Wrote the final report and prepared the presentation.

Continued on next page

Table 5.1 – continued from previous page

Members	Roles and Responsibilities
Ruyi Lu	<ul style="list-style-type: none"> • Conducted an extensive literature review to understand the uses of deep learning neural methods in the medical ultrasound image field mainly for classification tasks. • Brainstormed ways on how to upgrade our baseline model (ACSNet) including architectures and parameters that can be changed. • Developed codes for the classification task. • Incorporated DenseNet121 as the classifier for the two-branch classifier, with results used in the ablation study. • Wrote the final report and prepared the presentation.
Neosh Sheikh	<ul style="list-style-type: none"> • Developed the initial baseline code for the ACSNet including all the missing modules. • Conducted an extensive literature review to understand the uses of deep learning neural methods in the medical ultrasound image field mainly for classification tasks. • Brainstormed ways on how to upgrade our baseline model (ACSNet) including architectures and parameters that can be changed. • Built FiLM / ResNet50 model and incorporated it as the classifier for the two-branch classifier, with results used in the ablation study. • Implemented Pytorch Lighting to make training faster and make hyper-parameter tuning easier. • Wrote the final report and prepared the presentation.

CHAPTER 6

MILESTONES SCHEDULE

Table 6.1: Project Milestones and Schedule

Milestone	Tasks	Reporting	Date
Week-1	None	None	04-08-2024
Week-2	Setting up the IT infrastructure and communication tools Literature review to gain an understanding of the whole project	None	11-08-2024
Week-3	Architecture design Literature review to gain an understanding of the whole project	Client meeting to review the work plan	18-08-2024
Week-4	Literature review of specified modelling methodologies	None	25-08-2024
Week-5	Proposal Report Due Literature review for finding novel techniques that could adapt to the baseline.	Client meeting to review the work plan	01-09-2024
Week-6	Modelling - Develop codes for transfer learning, classification, and image segmentation	None	08-09-2024
Week-7	Modelling - Amend the codes utilising the entire U-net model to optimise the classification - Fixing training, classifier, structure modules.	None	15-09-2024

Continued on next page

Table 6.1 – continued from previous page

Milestone	Tasks	Reporting	Date
Week-8	Modelling - upgrade and revise the classifier modules	None	22-09-2024
Week-9	Progress Report Due Modelling - Complete the full model implementations	Client meeting to review the work plan	29-09-2024
Week-10	Optimisation - Conduct ablation study and increase the performance. Report - Introduction, Literature review, Methodology	None	13-10-2024
Week-11	Optimisation - Conduct ablation study and increase the performance. Report - Literature review, Methodology. Project problems	Client meeting to review the work plan	20-10-2024
Week-12	Optimisation - Conduct ablation study and increase the performance. Report - Literature review, Methodology, Resources, Milestones	None	27-10-2024
Week-13	Conduct a final performance analysis of the models to evaluate their effectiveness. Demo Submission Report - Results, Discussion	Client meeting to review the work plan	03-11-2024
Week-14	Final Presentation submission Report - Results, Discussion, Limitations	None	08-11-2024
Week-15	Final Report, Artifact submission	Client meeting to review the work plan	15-11-2024

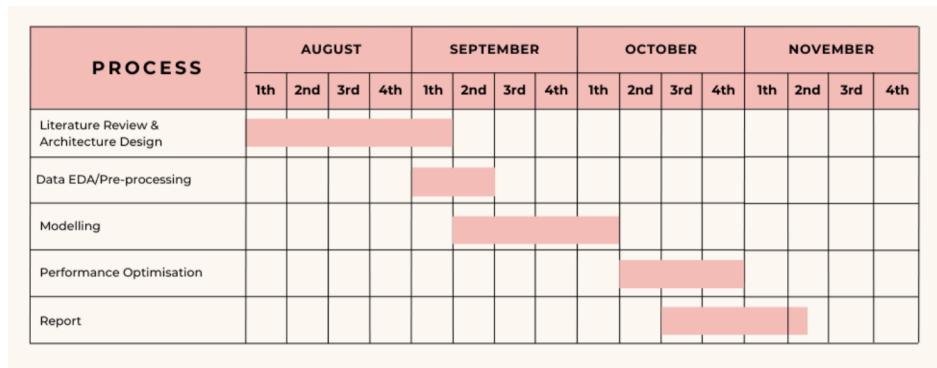


FIGURE 6.1: Gantt chart showing project timeline and milestones.

6.1 Milestones

The project schedule spans from August to mid-November and is divided into five main phases: Literature Review & Architecture Design, Data EDA/Pre-processing, Modelling, Performance Optimisation, and Report Preparation. Each phase progresses systematically with detailed weekly plans and milestones as outlined below.

6.1.1 Literature Review & Architecture Design

1st week of August - 1st week of September

In this initial phase, we conduct a literature review and design the model architecture. The focus is on building foundational knowledge necessary for modelling and data pre-processing. During the 2nd week, we set up the IT infrastructure and conduct a preliminary literature review to gain a comprehensive understanding of the project. By the 4th and 5th weeks, we complete the architecture design, selecting specific methods and novel techniques that align with our project and model requirements.

6.1.2 Data EDA/Pre-processing

1st week of September - 2nd week of September

This phase involves exploratory data analysis (EDA) and pre-processing to improve data quality and prepare it for the model. We work on understanding the data structure and characteristics, handling missing values, normalisation, augmentation, and other essential pre-processing tasks.

6.1.3 Modelling

2nd week of September - 1st week of October

As the core phase of the project, this stage focuses on developing and refining various models based on the pre-processed data. Using the ACSNet model, we develop classification and segmentation modules and explore different approaches to enhance performance. We submit the Proposal Report in the 5th week and the Progress Report in the 9th week, by which point the basic implementation of the entire model is completed.

6.1.4 Performance Optimisation

2nd week of October - 4th week of October

In this optimisation phase, we conduct an ablation study and test various approaches to improve model performance, with an emphasis on enhancing stability and generalisation. During this period, we simultaneously work on report writing and finalise the results, reviewing weekly work plans and refining the final report preparation process.

6.1.5 Report & Presentation Preparation

3rd week of October - 2nd week of November

In this final phase, we prepare the final report and presentation materials, evaluating and documenting the model's performance and results. Demo submission is scheduled for the 13th week, while presentation videos and materials are submitted in the 14th week. In the 15th week, we have a Q&A session along with the submission of the final report and artifacts.

6.2 Schedule Analysis

This schedule is meticulously structured with specific weekly tasks and milestone deadlines to ensure the project progresses in an organised manner. Regular meetings with the client and supervisor allow us to review progress, incorporate feedback, and focus on enhancing the final report and model quality.

CHAPTER 7

RESULTS

In this study, we conducted experiments under two different conditions to compare the final results. The first experiment was the **100-epoch version without an early stopping** set-up for comparison with the baseline paper. The results from our ablation study are presented below. The data here is collected from a 5-fold cross-validation for 100 epochs.

Table 7.1 summarises the performance of the models at 100 epochs without early stopping. Accuracy, precision, recall and F1 score are used to evaluate the classification performance of the models.

TABLE 7.1: Classification performance of the proposed two-branch classifier model at 100 epochs without early stopping.

Model	Accuracy	Precision	Recall	F1 Score
ACSNet (Baseline)	$94.44\% \pm 2.07$	$94.61\% \pm 1.56$	$93.86\% \pm 2.33$	$78.62\% \pm 1.75$
ResNet50 Branch (ACSNet)	$96.54\% \pm 2.17$	$96.58\% \pm 2.14$	$95.93\% \pm 2.65$	$96.52\% \pm 2.18$
EfficientNet Branch (ACSNet)	$97.05\% \pm 2.85$	$97.24\% \pm 2.57$	$97.59\% \pm 2.24$	$97.07\% \pm 2.82$
DenseNet Branch (ACSNet)	$95.51\% \pm 3.00$	$95.60\% \pm 2.99$	$95.61\% \pm 3.50$	$95.52\% \pm 3.01$
Swin Transformer Branch (ACSNet)	$96.28\% \pm 2.58$	$96.38\% \pm 2.50$	$94.89\% \pm 3.87$	$96.25\% \pm 2.59$

The Jaccard and Dice coefficients are used to evaluate the segmentation performance. These metrics are shown in Table 7.2.

TABLE 7.2: Segmentation performance of the proposed two-branch classifier model at 100 epochs without early stopping.

Model	Jaccard	Dice
ACSNet (Baseline)	78.62 ± 1.75	84.90 ± 1.69
ResNet50 Branch (ACSNet)	$65.16\% \pm 4.23$	$70.88\% \pm 4.46$
EfficientNet Branch (ACSNet)	$65.42\% \pm 3.90$	$71.02\% \pm 3.79$
DenseNet Branch (ACSNet)	$65.90\% \pm 3.47$	$71.38\% \pm 3.43$
Swin Transformer Branch (ACSNet)	$64.54\% \pm 3.60$	$70.48\% \pm 3.73$

Meanwhile, the second experiment **introduced early stopping** to prevent overfitting and achieve a more efficient model.

Table 7.3 summarises the performance of the models for the classification task when early stopping is introduced.

TABLE 7.3: Classification performance of the proposed two-branch classifier model with early stopping.

Model	Accuracy	Precision	Recall	F1 Score
ResNet50 Branch (ACSNet)	$93.72\% \pm 0.04$	$93.92\% \pm 0.04$	$93.64\% \pm 0.044$	$93.77\% \pm 0.04$
EfficientNet Branch (ACSNet)	$95.90\% \pm 0.03$	$96.13\% \pm 0.03$	$96.53\% \pm 0.03$	$95.93\% \pm 0.03$
DenseNet Branch (ACSNet)	$92.05\% \pm 0.03$	$92.37\% \pm 0.03$	$91.25\% \pm 0.05$	$91.98\% \pm 0.04$
Swin Transformer Branch (ACSNet)	$93.72\% \pm 0.03$	$93.90\% \pm 0.03$	$93.39\% \pm 0.03$	$93.74\% \pm 0.03$

Meanwhile, Table 7.4 shows the segmentation performance of the models with early stopping.

TABLE 7.4: Segmentation performance of the proposed two-branch classifier model with early stopping.

Model	Jaccard Coefficient	Dice Coefficient
ResNet50 Branch (ACSNet)	61.80% \pm 0.04043	68.22% \pm 0.0415
EfficientNet Branch (ACSNet)	59.24% \pm 0.0378	66.06% \pm 0.03709
DenseNet Branch (ACSNet)	61.70% \pm 0.04107	68.06% \pm 0.04146
Swin Transformer Branch (ACSNet)	60.64% \pm 0.0338	67.12% \pm 0.03213

Across the five-fold cross-validation with early stopping, each model stopped at different epochs. This can be attributed to the randomisation of the allocation of images within the validation set. The epochs at which the training stopped are summarised in Table 7.5.

TABLE 7.5: Epochs at which the model stopped per fold.

Model	Epochs
ResNet50 Branch (ACSNet)	20, 10, 9, 18, 9
EfficientNet Branch (ACSNet)	18,8,12,8,8
DenseNet Branch (ACSNet)	23,12,13,12,12
Swin Transformer Branch (ACSNet)	27,12,15,11,9

7.1 Interpretation of results

Starting with the **100-epoch experiment without early stopping**, we observed stable performance across all models, with EfficientNet achieving the highest accuracy at 97%. This outperformed the baseline ACSNet model, which recorded 94.44%, indicating a more precise classification capability. In this condition, each model had ample training time to reach peak performance, reflected in high precision, recall, and F1 scores, demonstrating robust capabilities.

On the other hand, in the **second experiment with early stopping**, we observed a slight decrease in performance across all models. This experiment was designed to mitigate overfitting and maintain stable performance while reducing computational cost. As a result, EfficientNet continued to lead with the highest accuracy at 95.90%, though it did not match the performance from the 100-epoch condition. Other models also showed a decline in precision, recall,

and F1 scores, but this can be seen as a reasonable trade-off for preventing overfitting and improving the model’s practical applicability in clinical settings.

Due to the imbalanced nature of the experimental dataset, relying solely on accuracy as a performance metric could be misleading. In imbalanced datasets, models may achieve high accuracy by favouring the majority class, failing to correctly identify minority classes. Therefore, we emphasised precision and recall when analysing model performance. Precision helps us understand the accuracy of positive predictions, reducing false positives, while recall evaluates the model’s ability to identify all true positive samples, minimising missed detections. These metrics are essential for a comprehensive evaluation of model performance under data imbalance.

From our results, it is evident that modifying the classification branch of ACSNet can significantly enhance classification performance. Across all classification branches, we observed improvements in every classification metric. However, this modification resulted in a degradation of segmentation performance. We hypothesise that this may be due to the relatively larger size of the classification branch compared to the rest of the model. This imbalance could potentially be addressed by increasing the weight of the segmentation loss during backpropagation.

From the results we obtained, EfficientNet excelled as far as classification is concerned. However, where segmentation is concerned, ResNet and DenseNet gave better performance. As our study is more concerned with classification performance rather than segmentation, we favour the usage of EfficientNet in the second branch of our model’s two-branch classifier network. However, were the emphasis skewed more towards segmentation or if the desire is to use a classifier that contributes a balanced performance across classification and segmentation, then ResNet50 can be a contender.

EfficientNet performs the best across key metrics such as accuracy, precision, recall and F1-score for classification tasks. Our studies found that, in breast cancer detection, EfficientNet has superior performance and fewer computational resources compared to ResNet, DenseNet and Swin-Transformer due to its efficient and scalable architecture. EfficientNet scales three dimensions concurrently including depth, width and resolution whereas traditional convolutional neural networks scale only one dimension. Additionally, EfficientNet is able

to achieve high accuracy while processing high-resolution images and generalising across different medical imaging modalities (Oyebanji et al., 2024).

CHAPTER 8

DISCUSSION

8.1 Comparison with ACSNet

Our modified ACSNet diverges from the original by employing a more powerful image classification model in place of the U-Net for classification. This modification is based on the premise that specialised architectures like ResNet, Swin Transformer, EfficientNet, and DenseNet are better suited for image classification tasks. Furthermore, our modified architecture allows the classification branch to leverage information learned during the segmentation process.

We hypothesise that the key factor behind the improved performance across all our modified ACSNet models is the interaction between segmentation and classification features. This hypothesis is supported by observing a significant increase in classification accuracy compared to models that solely rely on image classification. For instance, Meng et al. (2024) reported a 94.81% accuracy and 94.75% F1 score using ResNet50 on the BUSI dataset. In contrast, our modified ACSNet model, incorporating segmentation features through feature concatenation, achieved 96.56% accuracy and 96.52% F1 score with ResNet50, which is a significant improvement. This observation strongly suggests that the interaction with segmentation features enhances the performance of base classification models like ResNet50, Swin Transformer, EfficientNet, and DenseNet. Furthermore, our findings are consistent with Zakareya et al. (2023), who observed that even without explicit segmentation integration, architectural variations like ResNet50 (85.83% accuracy) offered marginal improvements over ResNet50 (85.4% accuracy) on the BUSI dataset. This suggests that while model architecture plays a role, the significant leap in performance observed in our ACSNet models stems from the effective utilisation of segmentation information.

While our modified ACSNet models excel in classification, they exhibit lower segmentation performance compared to the original ACSNet and a standalone U-Net. Specifically, our

models achieve a Dice score of around 70% on this dataset, compared to 83.04% for a pure U-Net. We hypothesise that this discrepancy may be attributed to a bias towards classification during training. The faster convergence of the classification loss could lead to a situation where the model prioritises optimising classification accuracy at the expense of segmentation performance. During back-propagation, the model aims to minimise the overall loss. If the U-Net’s contribution to the classification process is significant, even minor changes in the U-Net’s weights could drastically affect classification performance. This might lead the optimisation process to favour solutions that benefit classification, potentially trapping the model in a local minimum with suboptimal segmentation performance.

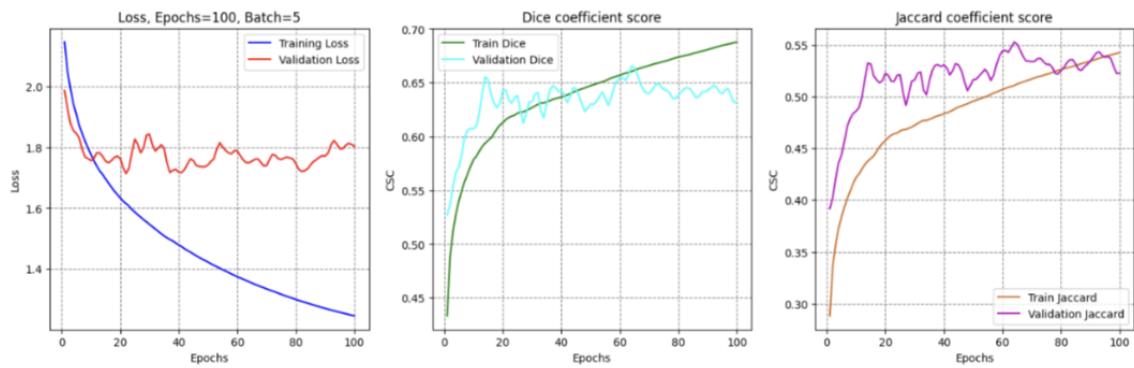


FIGURE 8.1: Training vs validation loss and segmentation performance

As shown Fig 8.1 our model exhibited signs of over-fitting on the segmentation task quite early in the training process, around the 10th epoch. This early over-fitting hindered the model’s ability to reach the expected performance level. While various factors having a small dataset, the large classification bias, hyper-parameter settings could have contributed to this behaviour, premature over-fitting likely played a significant role in preventing us from effectively implementing an early stopping strategy.

8.2 Implications

8.2.1 Multi-task learning for medical imaging and diagnosis

Most deep learning models used for medical imaging that are available in the market are single-task models. During our literature review, we came across only a handful of models that tackle multi-task learning. Our proposed model, which is based on the ACSNet architecture,

helps address the gap by adding to the knowledge base of multi-task machine learning models. Secondly, our proposed model improves upon an existing model that takes on the two-fold challenge of classification and segmentation simultaneously.

What our model does for medical practitioners is to help lessen the cognitive load and fatigue associated with the initial analysis of ultrasound images of tumours. While by no means a replacement for robust medical diagnosis and human wisdom, our model cuts through the initial drudgery by performing a first pass of analysis to allow doctors to focus on the critical aspects of diagnosis, such as staging the malignant tumours.

That said, multi-task segmentation and classification models do exist today. But while our model might not be groundbreaking, it improves upon what already exists. By doing so, our model helps advance the knowledge and tooling in this field of study.

8.2.2 Significance of introducing early stopping to ensure the reliability of the diagnostic system

The introduction of early stopping in the second experiment carries the following significant implications. Firstly, considering that achieving both diagnostic efficiency and accuracy are the core objectives of this project, preventing overfitting caused by excessive model training is crucial. It is essential for the model to maintain stable performance and consistently apply to various data in medical settings. Overfitting can cause the model to be overly optimised for the training data, resulting in decreased performance on new data. Therefore, using early stopping to mitigate overfitting plays a key role in enhancing the reliability of the diagnostic system.

Additionally, early stopping helps to reduce computational costs and shorten training time. This contributes to faster predictions in real clinical environments, which allows medical professionals to promptly receive diagnostic results. This is especially valuable in medical environments where fast and efficient diagnosis is essential. This aligns with one of the project's goals: reducing the workload of medical staff. Finding a way to achieve stable performance with less computational cost and time also increases the system's applicability in healthcare settings.

Finally, to provide medical professionals with reliable results, the model must be kept in a generalisable state without becoming overly biased during training. Early stopping can support the model's stability and generalisation, ensuring diagnostic consistency and reducing the risk of misdiagnosis.

Thus, the second experiment results obtained through early stopping do not merely reflect a slight decrease in performance metrics but rather signify a strategic choice to enhance the practical applicability of the model in clinical settings and to establish a more stable and efficient diagnostic system.

8.2.3 One-shot classification / multi-task learning

The primary goal of our study was to evaluate whether it was possible to improve the one-shot classification of breast cancer. Unlike other studies where segmentation masks are provided as initial input during inference, our models do not require these masks yet still achieve high accuracy.

Our modified ACSNet using the proposed two-branch classifier shows the potential for building complex interaction models where features found from one task can be used to enhance the performance of a model in another task.

8.2.4 Extensibility of the modular two-branch classifier

The two-branch classifier that we introduced as an improvement on the baseline ACSNet model intentionally leaves the second branch of the classification network as a modular component that can be replaced by any state-of-the-art classifier. From the experiments that we conducted, we saw that adding this second branch improved the classification performance of the multi-task model. Furthermore, while using different state-of-the-art classifiers in this second branch led to varying classification performance, all of these outcomes were better than the classification performance of the baseline ACSNet model. The modular design of the two-branch classifier opens the door for extensibility and the flexibility to use other state-of-the-art classifiers should there be newer algorithms come out in the future. This allows our proposed model to evolve along with the pace of new developments.

8.2.5 Tackling other medical imaging tasks

Our multi-task model focused on the dual challenge of classification and segmentation, leveraging the U-Net backbone for segmentation. In our model, we modified the classification network, while keeping the segmentation network unchanged. Taking inspiration from this, we can apply the inverse to address a different multi-task challenge. For instance, we could potentially replace the segmentation network with an algorithm that is more suited to a detection task in order to come up with a more lightweight model that tackles classification and detection. Such a model can be used to handle use cases that do not require detailed tumour morphology to be analysed. An example of such a use case would be determining the location of malignant tumours that have metastasised to other locations.

CHAPTER 9

LIMITATIONS AND FUTURE WORKS

9.1 Datasets - Transferability and lack of validation set

This study has several limitations regarding the dataset. First, model performance was not evaluated on alternate datasets, limiting any conclusions about the transferability of our findings. Second, due to the small size of the BUSI dataset, a validation set was not used, and performance was evaluated on the test set for 100 epochs. This approach may not generalise to real-world settings where the optimal number of epochs for test accuracy is unknown. Future studies should address these limitations by incorporating larger and more diverse datasets. This could include utilising other publicly available breast ultrasound image datasets and expanding to include other imaging modalities such as MRI and CAT scans. Additionally, images of malignant and benign tumours from other body regions could be included to improve sample size and vulnerability. Transfer learning approaches could then be leveraged to further enhance model performance.

9.2 Reducing classification bias

Due to time constraints, this study only analysed the variability of total loss during training against the test set. This limits our ability to understand the factors contributing to the degradation of segmentation performance. We hypothesise that the use of a large classification branch and early model convergence may be contributing factors. Specifically, the use of pretrained classifiers (but not a pretrained U-Net) and a fixed loss concatenation bias of 0.5 for both classification and segmentation may have led to rapid classification convergence at the expense of segmentation performance.

Future studies should investigate this issue further by:

- Testing various loss concatenation configurations.
- Exploring methods such as freezing the classifier to allow the U-Net to converge first.

9.3 Improving feature fusion / interaction

Within our proposed framework, the ACSNet and the classifier CNN were integrated through a straightforward feature concatenation approach. While this strategy resulted in improved classification performance compared to employing the classifier CNN independently, further enhancements could potentially be achieved by introducing segmentation information to the classifiers at an earlier stage.

Future research endeavours may explore alternative methods of feature fusion or model interaction to optimise the overall performance of the proposed framework.

Bibliography

- Ahmad, M., Distifano, S., Khan, A. M., Mazzara, M., Li, C., Yao, J., Li, H., Aryal, J., Vivone, G., & Hong, D. (2024). A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers. *arXiv preprint arXiv:2404.14955*.
- Al-Dhabayani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, 28, 104863. <https://doi.org/10.1016/j.dib.2019.104863>
- Alim, M. S., Bappon, S. D., Sabuj, S. M., Islam, M. J., Tarek, M. M., Azam, M. S., & Islam, M. M. (2024). Integrating convolutional neural networks for microscopic image analysis in acute lymphoblastic leukemia classification: A deep learning approach for enhanced diagnostic precision. *Systems and Soft Computing*, 6(200121), 200121.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data*, 8(1), 53.
- Andrade, A., Lucena, C., Santos, D., Pessoa, E. C., Mansani, F. P., Andrade, F. E., Tosello, G. T., Pasqualette, H. A. P., Couto, H. L., Francisco, J. L. E., Costa, R. P., Teixeira, S. R. C., Moraes, T. P., & Silva Filho, A. L. (2023). Challenges of breast cancer screening. *Revista Brasileira de Ginecologia e Obstetrícia*, 45, 551–554. <https://doi.org/10.1055/s-0043-1775931>
- Breast cancer statistics. (2022, March). *World Cancer Research Fund International*. Retrieved October 15, 2024, from <https://www.wcrf.org/cancer-trends/breast-cancer-statistics/>
- Byra, M., Jarosik, P., Dobruch-Sobczak, K., Klimonda, Z., Piotrzkowska-Wróblewska, H., Litniewski, J., & Nowicki, A. (2022). Joint segmentation and classification of breast masses based on ultrasound radio-frequency data and convolutional neural networks. *Ultrasonics*, 121, 106682. <https://doi.org/10.1016/j.ultras.2021.106682>
- Cancer. (2022, February). *World Health Organization*. Retrieved October 15, 2024, from <https://www.who.int/news-room/fact-sheets/detail/cancer>

- Davri, A., Birbas, E., Kanavos, T., Ntritsos, G., Giannakeas, N., Tzallas, A. T., & Batistatou, A. (2023). Deep learning for lung cancer diagnosis, prognosis and prediction using histological and cytological images: A systematic review. *Cancers*, 15(15), 3981.
- Giger, M. L., Chan, H.-P., & Boone, J. (2008). Anniversary paper: History and status of cad and quantitative image analysis: The role of medical physics and aapm. *Medical Physics*, 35, 5799–5820. <https://doi.org/10.1118/1.3013555>
- Gökmen Inan, N., Kocadağlı, O., Yıldırım, D., Meşe, İ., & Kovan, Ö. (2024). Multi-class classification of thyroid nodules from automatic segmented ultrasound images: Hybrid ResNet based UNet convolutional neural network approach. *Comput. Methods Programs Biomed.*, 243(107921), 107921.
- He, Q., Yang, Q., Su, H., & Wang, Y. (2024). Multi-task learning for segmentation and classification of breast tumors from ultrasound images. *Computers in Biology and Medicine*, 173, 108319. <https://doi.org/10.1016/j.combiomed.2024.108319>
- He, Z., Chen, Z., Tan, M., Elingarami, S., Liu, Y., Li, T., Deng, Y., He, N., Li, S., Fu, J., & Li, W. (2020). A review on methods for diagnosis of breast cancer cells and tissues. *Cell Proliferation*, 53, e12822. <https://doi.org/10.1111/cpr.12822>
- Hesaraki, S. (2018). Breast ultrasound images dataset(busi). *Kaggle.com*. Retrieved October 20, 2024, from <https://www.kaggle.com/datasets/sabahesraki/breast-ultrasound-images-dataset>
- Hoang, V.-T., & Jo, K.-H. (2021). Practical analysis on architecture of efficientnet. <https://doi.org/10.1109/hsi52170.2021.9538782>
- Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2016). Densely connected convolutional networks.
- Iqbal, A., & Sharif, M. (2023). Bts-st: Swin transformer network for segmentation and classification of multimodality breast cancer images. *Knowledge-Based Systems*, 267, 110393.
- Kang, H., Park, C., & Yang, H. (2024). Evaluation of denoising performance of ResNet deep learning model for ultrasound images corresponding to two frequency parameters. *Bioengineering (Basel)*, 11(7), 723.
- Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. *BMC Med. Imaging*, 22(1), 69.

- Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, 11, 1273253.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.
- Luo, N., Zhong, X., Su, L., Cheng, Z., Ma, W., & Hao, P. (2023). Artificial intelligence-assisted dermatology diagnosis: From unimodal to multimodal. *Comput. Biol. Med.*, 165(107413), 107413.
- Mehdy, M. M., Ng, P. Y., Shair, E. F., Saleh, N. I. M., & Gomes, C. (2017). Artificial neural networks in image processing for early detection of breast cancer. *Computational and Mathematical Methods in Medicine*, 2017, 1–15. <https://doi.org/10.1155/2017/2610628>
- Meng, X., Ma, J., Liu, F., Chen, Z., & Zhang, T. (2024). An interpretable breast ultrasound image classification algorithm based on convolutional neural network and transformer. *Mathematics*, 12, 2354–2354. <https://doi.org/10.3390/math12152354>
- Oyebanji, O. S., Apampa, A. R., Idoko, I. P., Babalola, A., Ijiga, O. M., Afolabi, O., & Michael, C. I. (2024). Enhancing breast cancer detection accuracy through transfer learning: A case study using efficient net. *World Journal of Advanced Engineering Technology and Sciences*, 13, 285–318. <https://doi.org/10.30574/wjaets.2024.13.1.0415>
- Pawłowska, A., Karwat, P., & Żołek, N. (2023). Letter to the editor. re: “[dataset of breast ultrasound images by w. al-dhabayani, m. gomaa, h. khaled a. fahmy, data in brief, 2020, 28, 104863]”. *Data in Brief*, 48, 109247. <https://doi.org/10.1016/j.dib.2023.109247>
- Pinto-Coelho, L. (2023). How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. *Bioengineering*, 10, 1435. <https://doi.org/10.3390/bioengineering10121435>
- Saha, M., Chakraborty, M., Maiti, S., & Das, D. (2024). Breast-net: A lightweight dcnn model for breast cancer detection and grading using histological samples. *Neural Computing and Applications*, 36, 20067–20087. <https://doi.org/10.1007/s00521-024-10298-9>
- Shafiq, M., & Gu, Z. (2022). Deep residual learning for image recognition: A survey. *Appl. Sci. (Basel)*, 12(18), 8972.
- Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2023). Transformers in medical imaging: A survey. *Medical Image Analysis*, 88, 102802.

- Song, K., Feng, J., & Chen, D. (2024). A survey on deep learning in medical ultrasound imaging. *Front. Phys.*, 12.
- Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. *proceedings.mlr.press*. <https://proceedings.mlr.press/v97/tan19a.html?ref=jina-ai-gmbh.ghost.io>
- Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., & Yao, Y. (2021). Deep learning in medical ultrasound image analysis: A review. *IEEE Access*, 9, 54310–54324. <https://doi.org/10.1109/access.2021.3071301>
- Xu, W., Fu, Y.-L., & Zhu, D. (2023). ResNet and its application to medical image processing: Research progress and challenges. *Comput. Methods Programs Biomed.*, 240(107660), 107660.
- Zakareya, S., Izadkhah, H., & Karimpour, J. (2023). A new deep-learning-based model for breast cancer diagnosis from medical images. *Diagnostics (Basel)*, 13(11).
- Zhang, B., Vakanski, A., & Xian, M. (2023). Bi-rads-net-v2: A composite multi-task neural network for computer-aided diagnosis of breast cancer in ultrasound images with semantic and quantitative explanations. *IEEE Access*, 11, 79480–79494. <https://doi.org/10.1109/access.2023.3298569>
- Zhao, X., Zhang, L., & Lu, H. (2021). Automatic polyp segmentation via multi-scale subtraction network. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 Lecture Notes in Computer Science*, 12901, 120–130. https://doi.org/10.1007/978-3-030-87193-2_12
- Zulfiqar, F., Ijaz Bajwa, U., & Mehmood, Y. (2023). Multi-class classification of brain tumor types from mr images using efficientnets. *Biomedical Signal Processing and Control*, 84, 104777. <https://doi.org/10.1016/j.bspc.2023.104777>

APPENDIX A

Appendix A

A1 Links to the code and results

ResNet50: <https://www.kaggle.com/code/neosh11/acsnet-4237a0>

EfficientNet: <https://www.kaggle.com/code/vinkkki/acsnet>

Swin Transformer: <https://www.kaggle.com/code/ergouliu/swin-transformer>

DenseNet: <https://www.kaggle.com/code/luruyi121gmailcom/densenet121-25102024#With-Early-stopping>

A2 Loss / Segmentation performance per fold

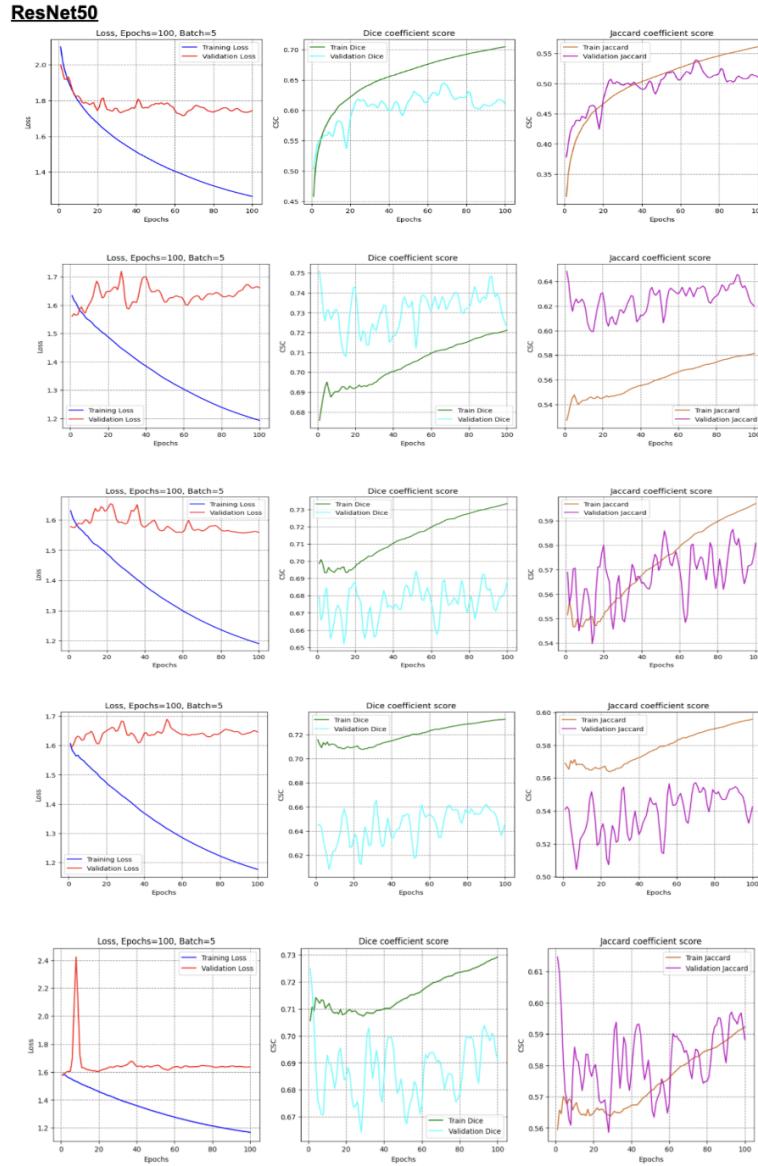


FIGURE A.1: ResNet50 w/o early stopping

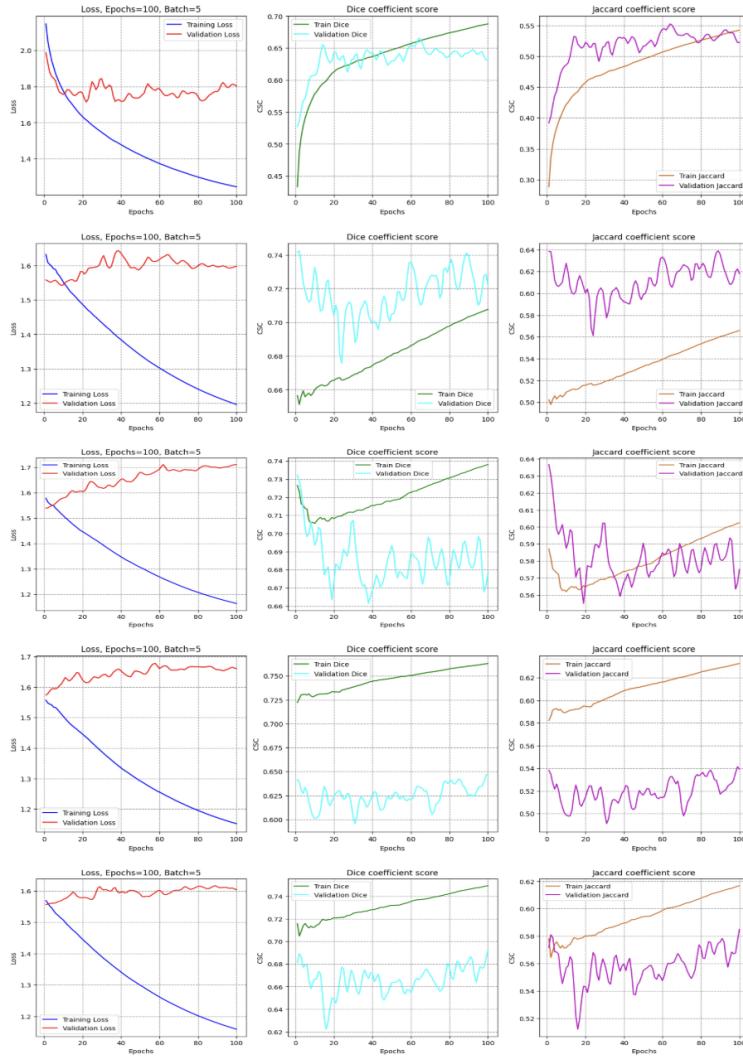
EfficientNet

FIGURE A.2: EfficientNet w/o early stopping

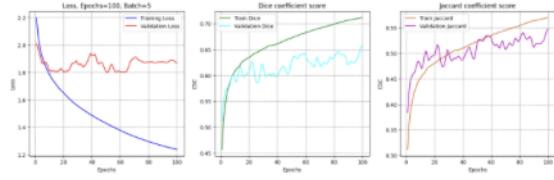
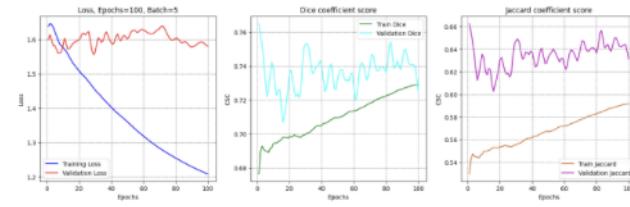
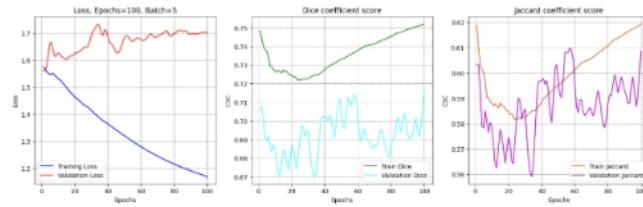
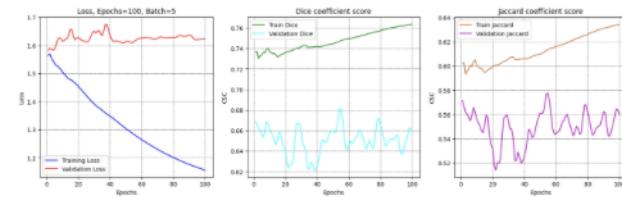
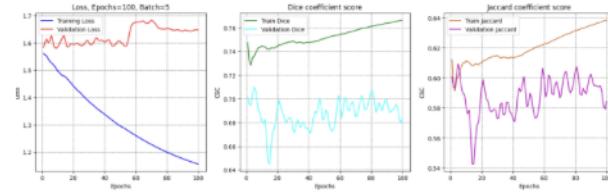
DenseNet**Cv fold 1****Cv fold 2****Cv fold 3****Cv fold 4****Cv fold 5**

FIGURE A.3: DenseNet w/o early stopping

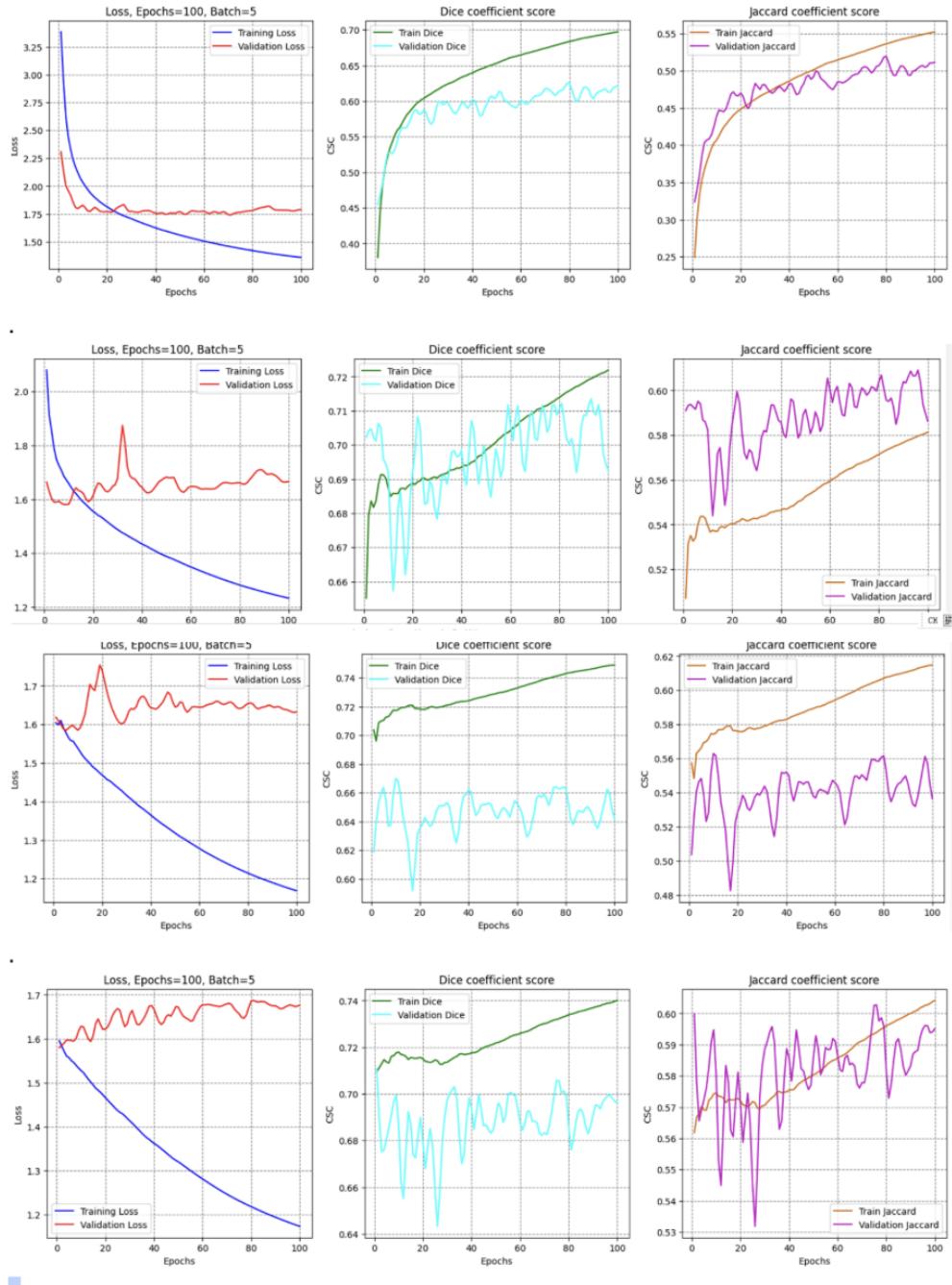
SwinTransformer.

FIGURE A.4: SwinTransformer w/o early stopping

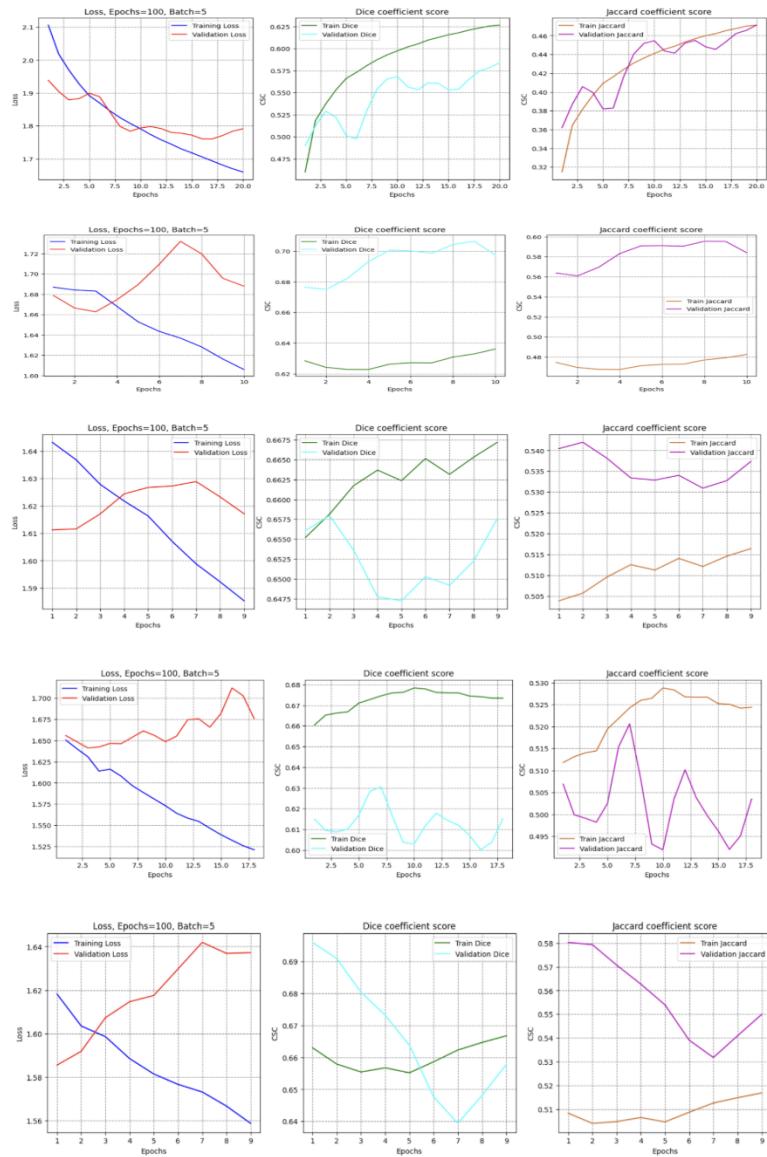
ResNet50

FIGURE A.5: ResNet w/ early stopping

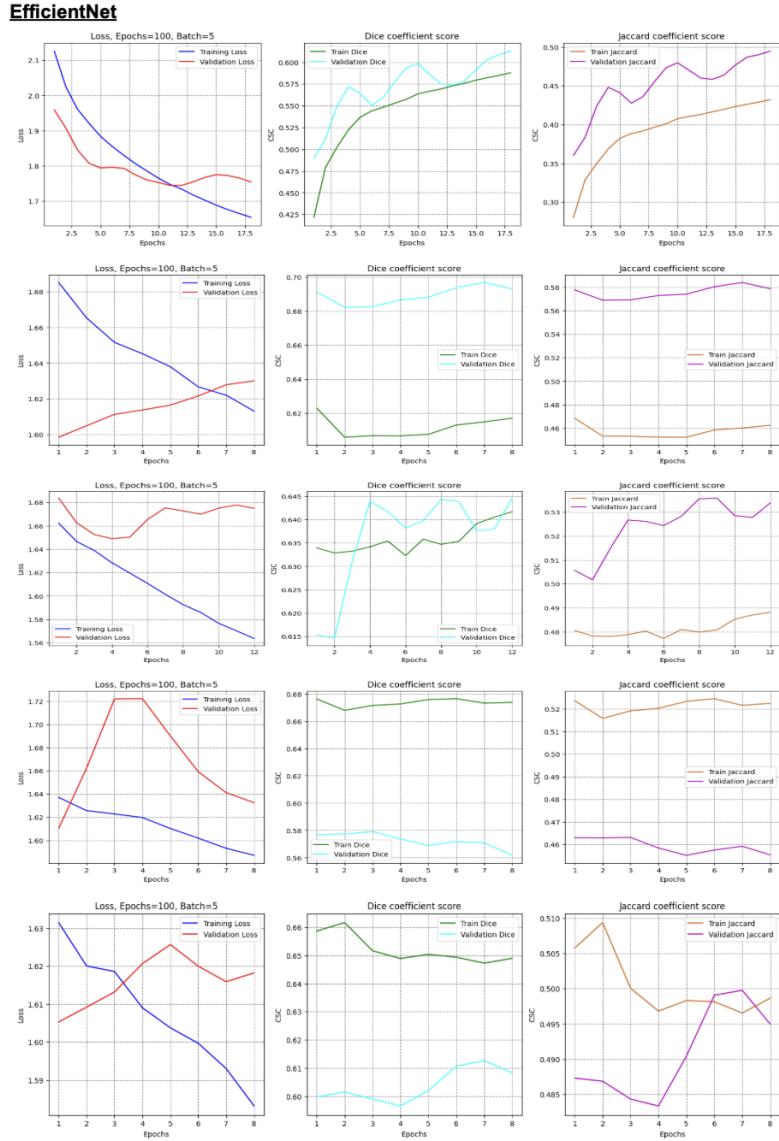


FIGURE A.6: EfficientNet w/ early stopping

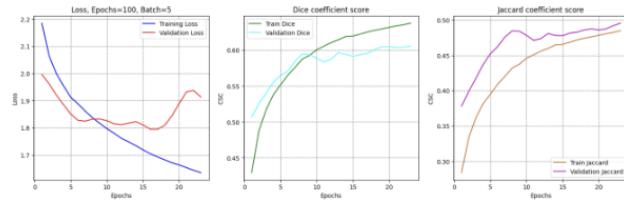
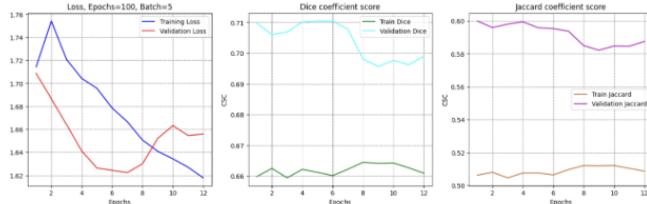
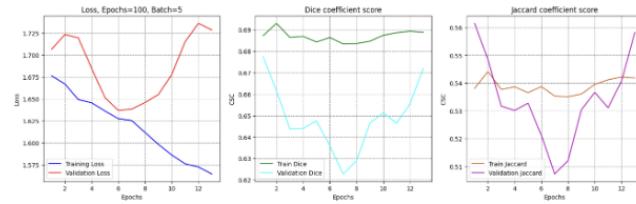
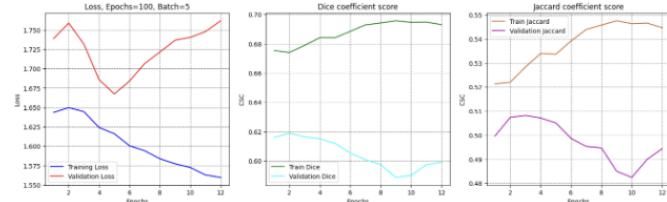
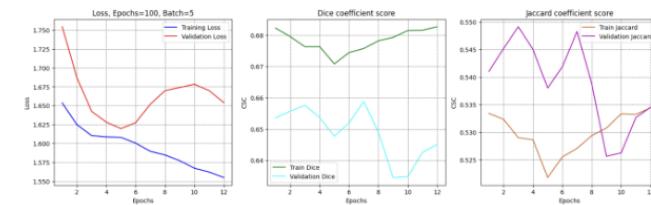
DenseNet**Cv fold 1****Cv fold 2****Cv fold 3****Cv fold 4****Cv fold 5**

FIGURE A.7: DenseNet w/ early stopping

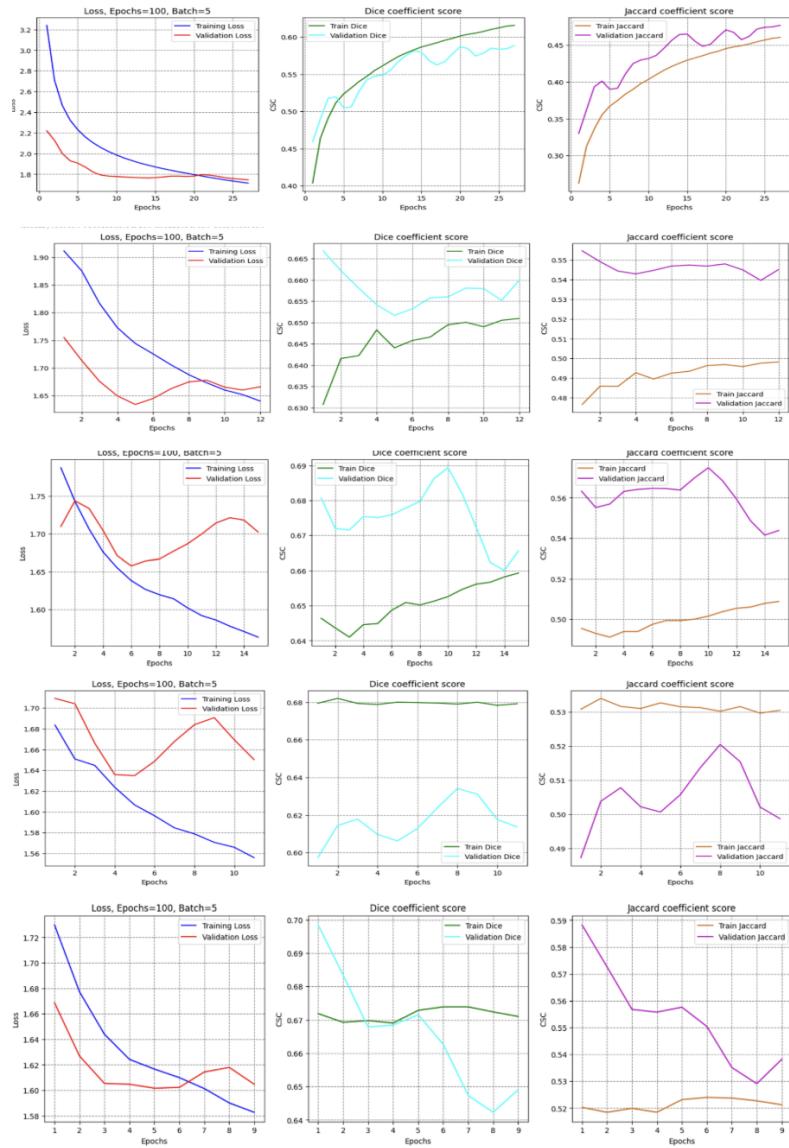
SwinTransformer

FIGURE A.8: SwinTransformer w/ early stopping