

Deep Learning-Based Breast Cancer Segmentation and Classification using CBIS-DDSM Dataset

1 Abstract

Breast cancer remains one of the leading causes of mortality among women globally. Early detection and accurate classification are essential for effective treatment and improved survival rates. In this project, we propose a comprehensive approach that leverages advanced deep learning techniques to automate the analysis of mammogram images. Using the CBIS-DDSM dataset, we develop models for three key tasks: classification of mammograms, detection of tumor regions from ROI masks, and precise segmentation of tumor boundaries. A wide range of deep learning architectures is employed, including convolutional neural networks (CNNs) such as VGG16, ResNet50, DenseNet121, MobileNet, and transformer-based models like Vision Transformer (ViT). For segmentation, models like U-Net, and DeepLabV3+ are utilized. Each model is rigorously evaluated using performance metrics relevant to its task. This work aims to identify the most effective strategy for integrating artificial intelligence into clinical breast cancer diagnosis.

2 Introduction

Breast cancer is a major global health issue, affecting millions of women each year. Mammography, a widely-used imaging technique, plays a critical role in the early detection of breast abnormalities. However, manual interpretation of mammograms can be challenging and prone to human error due to the complexity of the images and subtlety of early-stage symptoms. Artificial intelligence (AI), particularly deep learning, has shown great promise in medical imaging applications. By automating the analysis of mammograms, AI can help reduce diagnostic errors, enhance the speed of analysis, and assist radiologists in making accurate decisions. This project focuses on developing deep learning models capable of classifying mammograms, localizing tumors, and segmenting affected regions, with the ultimate goal of creating a robust diagnostic support system.

3 Dataset

The dataset utilized for this project is the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM), provided by The Cancer Imaging Archive (TCIA). It contains a large and diverse collection of annotated mammogram images

designed for research and development in medical imaging. Key attributes of the dataset include:

- Over 50,000 mammogram images from approximately 4,000 patients
- Detailed labels and annotations, including BI-RADS scores (a standard system for categorizing breast density and risk), tumor segmentation masks, and bounding box coordinates
- Categorization of cases into three main types: normal (no tumor), benign (non-cancerous tumor), and malignant (cancerous tumor)
- High-resolution grayscale images captured in standard clinical formats

4 Proposed Methodology

Our methodology is divided into several stages, from data preprocessing to model evaluation. The main components of the proposed system are detailed below.

4.1 Data Cleaning and Preprocessing

The raw mammogram images require several preprocessing steps before they can be used for model training:

- **Image Format Conversion:** Original DICOM images are converted to PNG or JPEG formats for ease of processing with deep learning libraries.
- **Resizing:** Images are resized to 224×224 pixels for classification tasks and 512×512 pixels for segmentation models to ensure consistent input size across models.
- **Normalization:** Pixel intensity values are scaled to a standard range, usually $[0,1]$, to stabilize and speed up training.
- **Contrast Enhancement:** CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied to enhance image quality and highlight important features.
- **Data Augmentation:** Various augmentation techniques such as rotation, flipping, zooming, and adding noise are applied to increase the diversity of training samples and prevent overfitting.

The CBIS-DDSM dataset provides three types of images: Original Mammograms (full breast scans), ROI Masks (binary tumor masks), and Cropped Mammograms. Among these, the ROI masks are manually annotated regions intended to highlight tumor locations, but they are not precisely aligned to the tumor boundaries at the pixel level. Instead, they offer only a rough or approximate indication of the tumor’s position. We extracted bounding boxes directly from these ROI masks. This was done by identifying the non-zero regions (tumor areas) in the mask and computing the tightest rectangle, defined by xmin, ymin, xmax,

and ymax coordinates, that encloses the tumor. These bounding boxes were then mapped onto the corresponding full mammogram images, giving the detection models access to both the global context of the breast and the localized tumor position. This method offered a practical balance between annotation precision and detection accuracy, making the dataset suitable for training object detection models.

Code Explanation: The preprocessing code includes loading DICOM files using pydicom, converting them to images using OpenCV, applying CLAHE to improve visibility, and augmenting the dataset using TensorFlow or PyTorch libraries. Functions are modular and reusable across training pipelines.

4.2 Classification

For the classification task, we prepared the dataset carefully to ensure that the images were standardized, balanced, and enhanced for robust model training. The original CBIS-DDSM mammogram images were first loaded and resized to a uniform size of 224x224 pixels, which is optimal for deep learning models like VGG, ResNet, and other convolutional neural networks. To improve contrast and enhance the visibility of features, especially tumor regions, CLAHE (Contrast Limited Adaptive Histogram Equalization) was applied to the images in LAB color space and converted back to RGB. This step helps the model to better capture subtle texture variations. Additionally, various data augmentations were introduced to increase dataset diversity and avoid overfitting. These included random rotations, horizontal and vertical flips, random brightness, contrast, and saturation adjustments, and the addition of Gaussian noise — all of which simulate real-world variations and improve generalization.

After augmentation, each image was saved with a unique filename to avoid duplicates and maintain dataset integrity. Furthermore, because the original dataset was imbalanced (benign vs malignant cases), a balancing step was performed where underrepresented classes were augmented further to ensure that both classes had an equal number of samples. This balanced and augmented dataset was finally saved into separate folders per class to be used for classification model training. Overall, this preprocessing pipeline was designed to standardize image size and quality, introduce variability through augmentation, and address class imbalance, thereby creating a clean and rich dataset for effective breast cancer classification. Classification involves assigning each mammogram to one of the predefined categories: normal, benign, or malignant. We use the following models:

4.2.1 VGG16 Model:

- Results comparison for VGG16 without pretraining and with pretraining on imagenet weights:

A VGG16-based classifier was developed for mammogram image classification. The architecture used VGG16 without pre-trained weights, freezing early layers while fine-tuning the last 8 layers. A custom classification head with global pooling, dense layers, and dropout was added. The model was trained with categorical cross-entropy loss and Adam optimizer, using callbacks for early stopping and best model saving. On evaluation, the model trained from scratch achieved 71.56% test accuracy, while the

pre-trained VGG16 (with ImageNet weights) performed better with 77.39% test accuracy, demonstrating the benefit of transfer learning for medical image classification tasks.

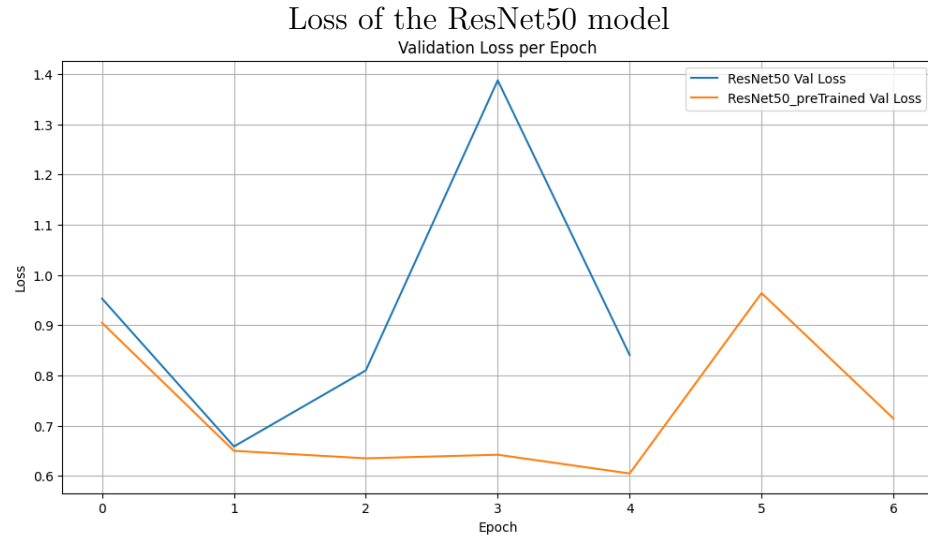
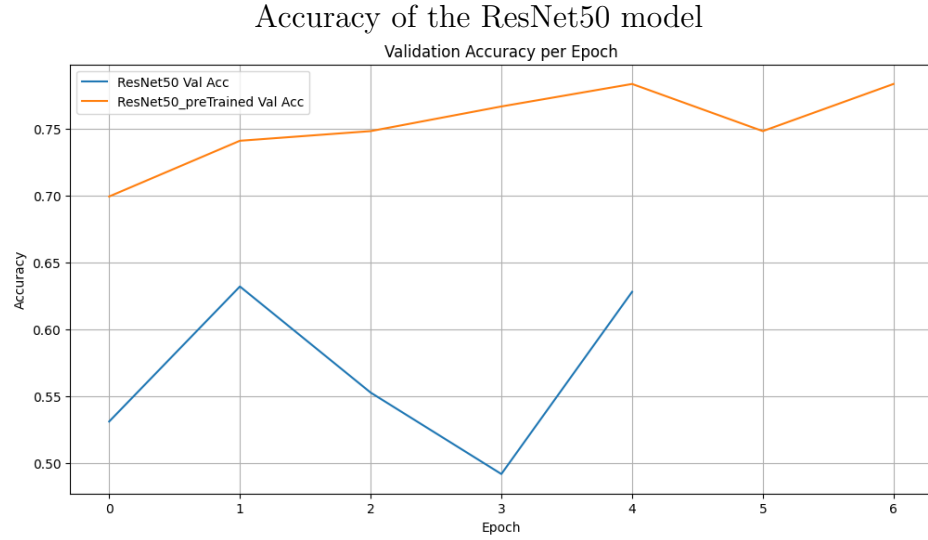


4.2.2 ResNet50 Model:

- Results comparison for ResNet50 without pretraining and with pretraining on imagenet weights:

A ResNet50-based model was built for mammogram classification, leveraging transfer learning. The backbone used ResNet50 pre-trained on ImageNet, with the last 30 layers unfrozen for fine-tuning while freezing earlier layers to retain learned generic features. A custom classification head with global average pooling, dense layers, batch normalization, and dropout was added. The model was trained using categorical cross-entropy loss and Adam optimizer with early stopping and best model saving strategies. Evaluation results showed that the pre-trained ResNet50 achieved 76.46% test accuracy,

outperforming the version trained from scratch, which reached 61.54% accuracy, thus validating the positive impact of pre-trained knowledge in medical image classification tasks.



4.2.3 DenseNet Model:

- Results comparison for DenseNet without pretraining and with pretraining on imagenet weights:

A DenseNet121-based deep learning model was designed for breast cancer classification on mammogram images. Initially, the model used non-pretrained (random) weights, with the top classifier excluded. The last 100 layers of the DenseNet backbone were unfrozen to enable fine-tuning while earlier layers remained frozen to preserve generalizable features. The classification head was added with global average pooling, dense layers, and dropout for regularization. The model was compiled with categorical cross-entropy loss and Adam optimizer. On evaluation, the model trained from scratch

achieved only 61.31% test accuracy, whereas using pre-trained DenseNet121 weights improved test accuracy to 76.92%, demonstrating that transfer learning significantly boosts performance on medical imaging tasks.



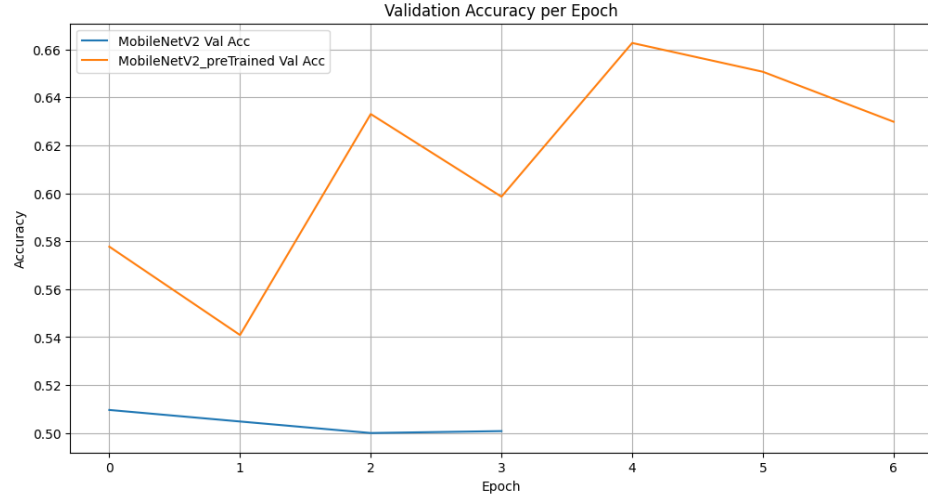
4.2.4 MobileNet Model:

- Results comparison for MobileNet without pretraining and with pretraining on imagenet weights:

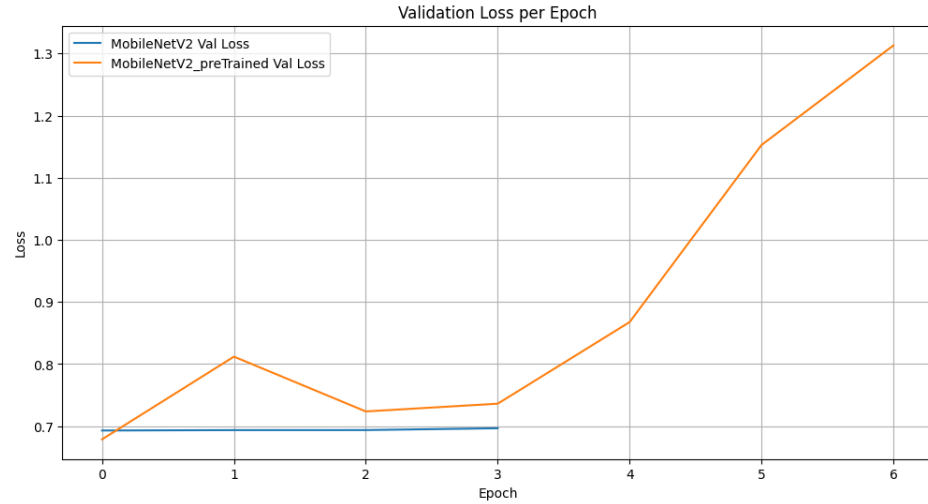
A MobileNetV2 architecture was implemented to classify mammogram images for breast cancer detection. Initially, the model was constructed without pre-trained weights and the last 120 layers were unfrozen for fine-tuning while the rest remained frozen to retain general features. The classification head included global average pooling, dense layers, and dropout for better generalization. After compilation using categorical cross-entropy and Adam optimizer, the non-pretrained version yielded only 51.05% test accuracy, showing limited capability. However, using pre-trained ImageNet

weights significantly improved the test accuracy to 62.94%, indicating that transfer learning offers essential advantages even for lightweight models like MobileNetV2 in medical image classification.

Accuracy of the MobileNet model



Loss of the MobileNet model



4.3 Vision Transformer:

- A Vision Transformer (ViT) model was implemented for mammogram classification using the google/vit-base-patch16-224 pretrained model as the backbone. The backbone was frozen to leverage learned representations from large-scale natural image datasets. A custom Keras layer (ViTLayer) was created to integrate the transformer, extracting the CLS token for downstream classification. This token was passed through a lightweight head comprising dense layers with batch normalization and dropout for robust classification. Finally, a softmax layer predicted the binary class. The model achieved a test accuracy of 72.73%, demonstrating that transformer-based architectures, when fine-tuned properly, can generalize effectively even in specialized medical

imaging tasks like breast cancer detection.

All models are pretrained on the ImageNet dataset and fine-tuned on the CBIS-DDSM dataset. We use regularization techniques such as dropout and batch normalization, and apply learning rate scheduling for optimal convergence.

Code Explanation: The classification notebook includes steps for loading the processed dataset, splitting it into training and testing sets, and defining CNN or ViT models using Keras or PyTorch. Training routines include callbacks for early stopping and learning rate reduction. Model performance is monitored through accuracy and loss graphs.

4.4 Segmentation

The dataset provides full mammograms and ROI masks, where the ROI masks indicate tumor regions but are not cropped or aligned perfectly with the tumors. To prepare this data for effective segmentation model training, several steps were taken. Initially, paired datasets were created by extracting the full mammogram images and their corresponding ROI masks. Both images and masks were resized to a consistent size of 512x512 pixels to ensure uniformity and compatibility with neural network input requirements. However, since using full mammograms along with full ROI masks resulted in poor DICE coefficients due to imprecise localization and excessive background, a refined approach was implemented.

This involved extracting precise tumor-focused crops from the full mammograms using the ROI masks. The masks were thresholded to binary and processed to find the largest contour, representing the main tumor region. Bounding boxes were computed around these contours, with an added margin to retain some surrounding tissue context. These bounding boxes were then used to crop both the full mammograms and the corresponding masks, focusing specifically on the abnormality. The cropped images and masks were then resized back to 512x512 to maintain a standardized input shape. This method ensured that the model would focus only on the tumor regions during segmentation, reducing background noise and enhancing the accuracy of the tumor boundary predictions. Finally, the cropped and paired datasets were saved and zipped for efficient loading during model training. Overall, these preprocessing steps addressed the challenges posed by the loosely aligned ROI masks and optimized the dataset for precise and efficient tumor segmentation. The models used for segmentation include:

- **U-Net:** For the segmentation task, an advanced UNet architecture was designed to effectively learn the spatial and contextual patterns of tumors in mammogram images. The network followed the classical encoder-decoder pattern with skip connections, but also included modern enhancements such as Batch Normalization for stable learning and Dropout layers to prevent overfitting during training. The decoder used transposed convolutions to progressively upsample the feature maps and refine the segmentation boundaries. The final layer used a sigmoid activation to predict pixel-wise probabilities for binary segmentation (tumor vs. background). The model achieved impressive performance, reaching a pixel test accuracy of 95.66%, with a Dice coefficient of 0.8402 and an IoU score of 0.7246, reflecting its strong ability to delineate tumor regions

accurately. The use of Batch Normalization, dropout regularization, and careful up-sampling helped in achieving robust and generalizable segmentation performance on unseen mammograms.

- **DeepLabV3+:** For the segmentation task, an advanced DeepLabV3+ architecture with ResNet50 backbone was adopted to improve the ability to capture multi-scale tumor regions in mammogram images. Since the mammograms were grayscale, they were repeated into 3 channels to align with the expected ResNet50 input format. The model leveraged the Atrous Spatial Pyramid Pooling (ASPP) module to effectively capture contextual information at multiple receptive fields and scales, helping to detect both small and large tumors. Additionally, skip connections with low-level features from earlier ResNet layers ensured that fine-grained spatial details were preserved. The decoder path further refined the feature maps using convolutional blocks before upsampling to the original image size for pixel-wise predictions. The final activation layer used a sigmoid activation for binary segmentation. This architecture resulted in robust segmentation performance with a pixel Test Accuracy of 95.27%, a Dice Coefficient of 0.8294, and an IoU Score of 0.7087, showing its strong capability in localizing and delineating tumors with precision.

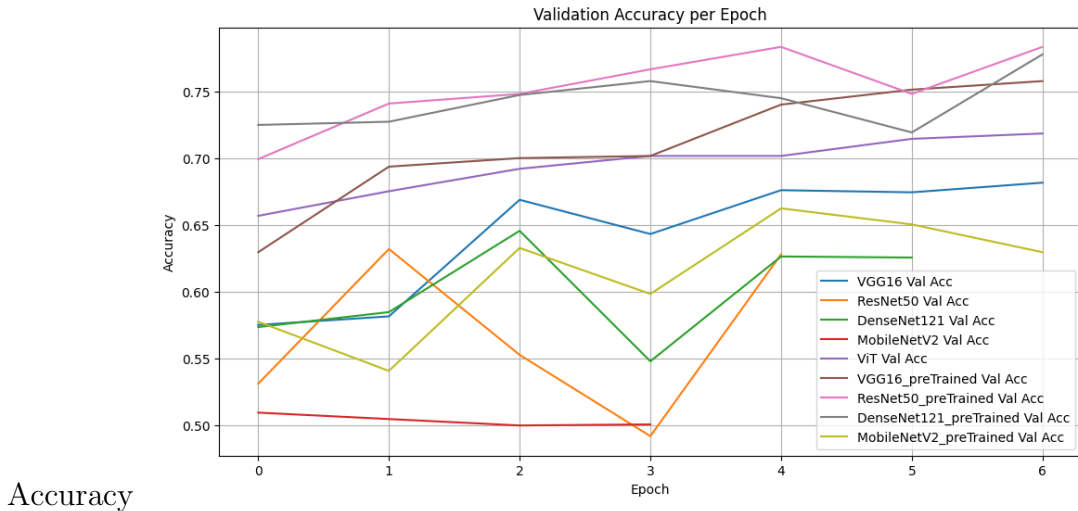
Segmentation performance is evaluated using:

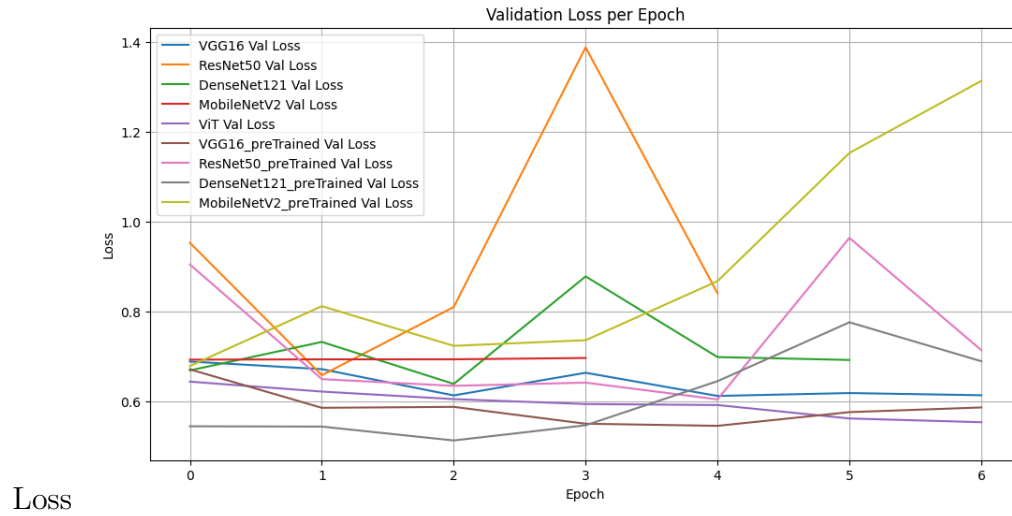
- **Dice Coefficient:** Measures the overlap between the predicted and actual tumor regions.
- **Jaccard Index (IoU):** Another overlap-based metric that compares intersection over union of predicted vs. actual regions.

5 Results

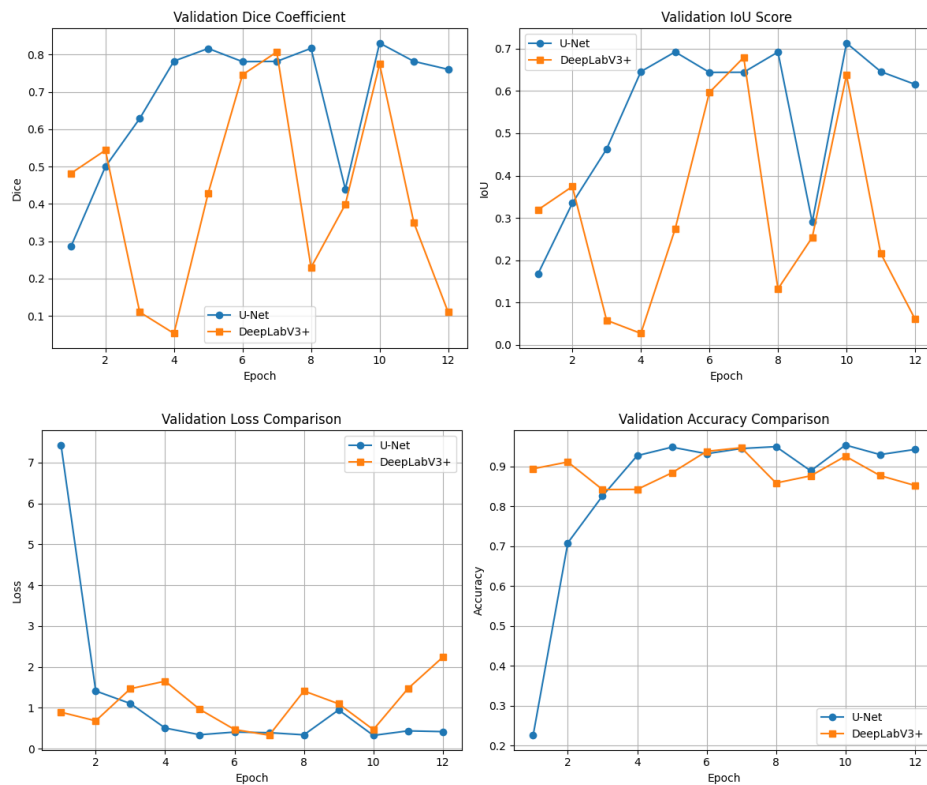
Model performance is assessed using different sets of metrics depending on the task:

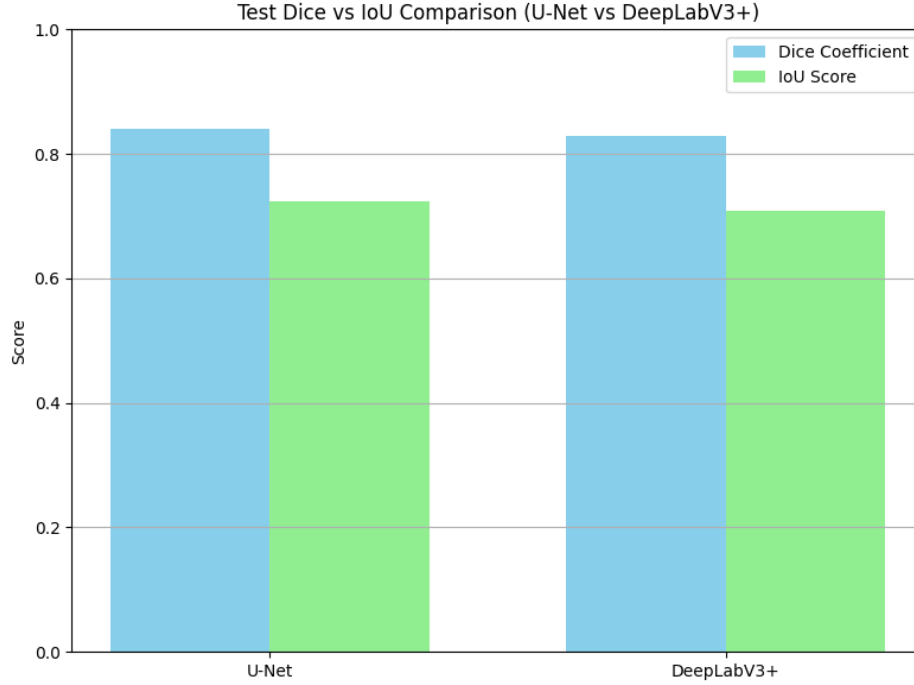
- **Classification:** The accuracy, loss are used to evaluate how well the models can distinguish between classes.





- **Segmentation:** Dice and Jaccard scores assess the spatial accuracy of tumor segmentation.





6 Conclusion

classification, segmentation, and detection tasks using advanced deep learning models. For classification, various state-of-the-art architectures such as VGG16, ResNet50, DenseNet121, MobileNetV2, and Vision Transformer (ViT) were explored, both with and without pre-trained weights, revealing the clear advantage of transfer learning in medical imaging applications. For segmentation, robust models like UNet and DeepLabV3+ were implemented to accurately localize tumor regions, achieving high Dice and IoU scores that reflect precise boundary detection. Furthermore, for detection tasks, bounding boxes were derived from ROI masks. Throughout the project, careful data preprocessing, augmentation, and class balancing ensured data quality and improved model generalization. The outcomes of this study not only highlight the effectiveness of deep learning in breast cancer image analysis but also lay the groundwork for potential real-world clinical applications that can assist radiologists in faster and more accurate diagnosis.

References

- [1] Sawyer-Lee, R., Gimenez, F., Hoogi, A., & Rubin, D. (2016). Curated Breast Imaging Subset of DDSM (CBIS-DDSM). The Cancer Imaging Archive. Available: <https://www.cancerimagingarchive.net/collection/cbis-ddsm/>
- [2] Oyelade, O. N., Ezugwu, A. E., & Omogbadegun, Z. O. (2022). Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Literature Review. *Applied Sciences*, 12(3).

- [3] Gupta, A., & Sinha, A. (2021). Enhancing breast cancer segmentation and classification using deep learning techniques. *Biomedical Signal Processing and Control*, 68, 102741.
- [4] Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Scientific Reports*, 6, 27327.
- [5] Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, 7, e6201.
- [6] Al-antari, M. A., Al-masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics*, 117, 44-54.
- [7] Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical Image Analysis*, 37, 114-128.
- [8] Ribli, D., Horváth, A., Unger, Z., Pollner, P., & Csabai, I. (2018). Detecting and classifying lesions in mammograms with deep learning. *Scientific Reports*, 8(1), 4165.
- [9] Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., ... & Jiang, T. (2020). A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks. *IEEE Access*, 8, 90931-90956.
- [10] Pun, N. S., & Agarwal, S. (2021). RCA-IU-net: A residual cross-spatial attention guided inception U-Net model for tumor segmentation in breast ultrasound imaging. *Computers in Biology and Medicine*, 135, 104541.