CHAPTER 1: INTRODUCTION

The Chapter briefly discusses about the layout of the model and various sub- structures involved underneath the abstract final model comprising of various CNN layers. The chapter, in depth provides insight to the application prospect of the technology along with the structure of the report.

1.1 BACKGROUND

The majority of cancer cases fall into the category of breast cancer and make the largest part of the total reported cancer cases worldwide overtaking lung cancer and is the second most prevalent cause of death among women [1] followed by heart diseases and has been on a rise with an estimation of 30% till 2025. Mammographic scans have long served as a foundation for early results in suspected cases of breast cancer. Mammography employs X-Rays to identify potential Regions of Interest (ROI), and the options make mammography a reasonable choice due to its cost effectiveness and ease of access in comparison to other techniques such as PET (Positron Emission Tomography), Ultrasound (specialisation required), MRI (Magnetic Resonance Imaging), and others, which are all expensive and less accessible resources.

In such a scenario using CESM images is a better alternative as it clearly demarcates the lesions in a scan and helps to allows for planning the proper surgical route for further investigation.

1.1.1 Challenges with Mammographs

Although mammograms can serve as a starting point for further investigations, they cannot be used to make an absolute diagnosis. Philips and Stinton in [2] discuss the

suggested 'double reading,' which means that the scans must be read independently by two radiologists to eliminate the possibility of probable recall cases and enhance sensitivity. The double reading however reduces the load on radiologists for only up to 50% instead to device a methodology that reduces the reduces the load further is desirable and to achieve this Lewin et. al. in [3] introduced Dual energy contrast enhanced spectral mammography (CESM) as a novel imaging modality for breast cancer screening which was then approved by FDA in [4] as a supplemental diagnostic imaging modality.

1.2 CADe/CADx

Yoon and Kim in [5] describe the primary goals of using artificial intelligence (AI) to assist image interpretation as follows:

1) Computer-aided detection (CADe), also known as automated lesion detection, focuses on identifying worrisome anomalies in an image.

2) Computer-aided diagnosis (CADx), the process of describing anomalies found by either the radiologist or the computer.

The interpreting radiologist decides the clinical significance of the discovered aberration and whether it requires additional study based on the CADe/CADx analysis.

Kim et. al. in [6] report that early investigations noted that computer-aided detection (CAD), which functions as an automated second reader by flagging potentially problematic regions for radiologists to analyse, may increase the sensitivity of initial screening.

1.3 CHOICE OF TECHNIQUE

Though image processing using DL (Deep Learning) is a comparatively newer domain but extensive research work has been already published. Earlier research

work like that of [7,8] report use of standard ML (Machine Learning) techniques like that of Decision Trees, SVM (Support Vector Machines) etc. and boosting algorithms like AdaBoost and XGBoost etc. The later discussed research part is extensively based on demographic, genetic etc, feature with some documents referring to mammographs as well but the overall use of DL has been limited due to high recall rates and low performance metrics. Constant effort is been made to check and test multiple probable models to increase performance metrics and recall rates. Latest techniques like YOLO are also been tested as in [9] and are yet on par with other DL methodologies.

1.4 TRANSFER LEARNING

Due to computational and hardware restrictions, running deep learning models requires operating several GPUs and a significant amount of RAM, which is out of the price range and reach for many people. So, with the ability to change the top and final layer, we can use models that have already been trained on large image datasets. RESNET50 [10], a residual network with around 150 layers that was trained by Microsoft, as well as VGG16 [11], VGG19 [12], which was also taught by Microsoft, AlexNet, etc. The final layer may be altered to accommodate the necessary number of classes, trained in real-time, or used as is with weights acquired from extensively large datasets like ImageNet, ChexNet by Stanford in [13].

1.5 COMPARISON OF MODELS

Performance metrics like confusion matrix, ROC (Receiver Operating Characteristic) curve, Specificity, Sensitivity, F1 Score etc. are used and are defined as follows:

1.5.1 Confusion matrix

A confusion matrix is a table that describes how well a categorization method performs.

- True Positive (TP): ROIs are those that are actually cancerous despite being labelled as such.
- False Positive (FP): A benign ROI that was mistakenly labelled as cancerous.
- True Negative (TN): A benign ROI is one that is actually benign.
- False Negative (FN): An ROI that is deemed benign but is actually cancerous.

1.5.2 Specificity and Sensitivity

The percentage of persons who test negative for a certain disease within a population of healthy individuals is shown by specificity, also known as the true negative rate.

Sensitivity, often known as the true positive rate, describes how successfully a test can find individuals who genuinely have a certain disease or condition.

1.5.3 Recall, Precision and F1 Score

Positive examples have an impact on a machine learning model's memory, whereas negative samples have no impact.

The precision of a machine learning model takes into account all positive samples that are rightly or wrongly identified as positive.

The F1 score assesses the model's precision. In cases of unequal class distribution, it is more beneficial than accurate.

1.5.4 ROC and AUC

The Receiver Operating Characteristic Curve (ROC curve) is a graph that displays

how well a classification model performs across all categorization levels. This curve has X and Y parameters as:

- True Positive Rate
- False Positive Rate

AUC stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve.

Table 1.1: Confusion Matrix

	PREDICTED CLASS		
ACTUAL CLASS		Class=No	Class=Yes
	Class= No	TN : True Negative	FP : False Positive
	Class=Yes	FN : False Negative	TP : True Positive

$$\text{Accuracy } \frac{TP+TN}{TP+TN+FP+FN} \quad (1.1)$$

$$\text{Precision - } \frac{TP}{TP+FP} \quad (1.2)$$

$$\text{Recall - } \frac{TP}{TP+FN} \quad (1.3)$$

$$\text{F1 Score - } \frac{2TP}{2TP+FN+FP} \quad (1.4)$$

$$\text{Specificity - } \frac{TN}{TN+FP} \quad (1.5)$$

$$\text{Sensitivity - } \frac{FN}{TP+FN} \quad (1.6)$$

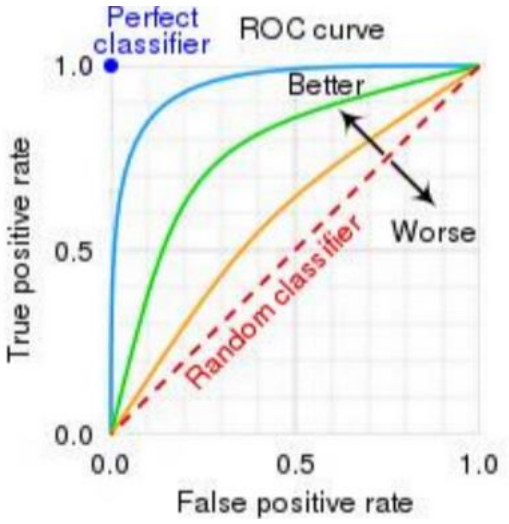


Figure 1.2: Sample ROC curve

1.6 TOOLS AND REQUIREMENTS

1. **ANACONDA**: The Anaconda platform facilitates the development and deployment of deep learning models predicated on neural networks.

2. **Google Colab**: This platform afforded us a complimentary 15 GB Graphics Processing Unit (GPU) infrastructure for the training of our architectural models.

3. Diverse Libraries such as Keras, TensorFlow, Matplotlib, and Keras Tuner were employed in the study.

Summary

The principal aim of our investigation was to formulate and assess distinct machine learning models, each specifically crafted to precisely predict the occurrence of breast cancer from potential lesions, and concurrently, to delineate the spatial localization of these lesions through the application of heatmaps. The integration of deep learning methodologies played a pivotal role in achieving a commendable level of accuracy. We employed pre-existing convolutional neural networks (CNNs) and multiple encoders to systematically analyze the intricate features inherent in Contrast-Enhanced Spectral Mammography (CESM) images.

Report Outline

Chapter 1 introduces the research with a focus on the challenges of mammographs. It explains the concepts of CADe/CADx, the chosen techniques, and why transfer learning is used. The chapter also compares different models using metrics like confusion matrix, specificity, sensitivity, recall, precision, F1 score, ROC, and AUC. Tools and requirements for the study are discussed, followed by a brief summary.

Moving on to Chapter 2, we explore "CESM Images." This chapter explains how CESM images work, the different views like craniocaudal and mediolateral oblique, and discusses image density. The advantages of CESM images over FFDM are also highlighted, wrapping up with a summary.

Chapter 3, titled "Literature Review," starts with an introduction and delves into CADe/CADx and decision support systems. A critical evaluation is presented based on the research's foundation and a review of relevant literature, concluding with a summary.

Chapter 4, "Methodology," goes into the nitty-gritty details. It covers pre-processing steps like data arrangement, resizing, pixel normalization, color space conversion, image equalization, and image augmentation. The deep neural network architecture is explained for models like ResNet50V2, ResNet101V2, InceptionV3, DenseNet169, DenseNet201, MobileNetV2, NASNetMobile, and NASNetLarge. The use of Keras Tuner, training parameters using Youden's Index, and localization techniques such as GradCAM and Saliency Heatmaps are discussed, followed by a summary.

Chapter 5, "Results and Discussion," is dedicated to analyzing classification and localization results. This includes a detailed examination of Grad-CAM results and Saliency Heat-maps, leading to a summarized discussion.

Chapter 6, "Conclusion and Future Scope," summarizes the study's findings and suggests potential future research directions.

The last section, "References," compiles all the sources cited throughout the report. Additionally, there's an appendix labeled "Appendix I" for any additional information.

CHAPTER 2: CESM IMAGES

In this chapter we explore the complex world of contrast-enhanced spectrometry (CESM) which is a key imaging modality for breast diagnostics. We look at the fundamental principles of CESM and its clinical applications, as well as the nuances that define the role of CESM in improving precision and effectiveness in the detection of breast cancer. In digital mammography, iodinated contrast material improves the visibility of hypertrophied tumours; however, there are still issues, especially when small enhancing tumours are hidden by overlapping high density breast tissue. CEM techniques have been developed to address this problem.

2.1 WORKING

The initial CEM method, known as temporal subtraction CEM, involves obtaining a non-contrast breast image followed by an intravenous contrast-enhanced image in the same position. These images are then subtracted, isolating the iodine signal. However, this technique faces significant limitations, including prolonged compression times, the ability to capture only one view of one breast, and susceptibility to misregistration artifacts due to any movement between image acquisitions. Additionally, the intravenous injection method precludes repeated administrations.[14-16]

Recognizing the drawbacks of temporal subtraction CEM, a dual-energy subtraction technique has emerged, capitalizing on the attenuation properties and k edge of iodine. This approach involves imaging the breast energies just above and below the k edge of iodine, allowing the acquisition of both traditional low-energy (LE) and iodine-enhanced images with a single, relatively brief compression.[16-18]

2.2 VIEWS

The following section briefly discusses various views and angles of capturing images

and scans.

2.2.1 Craniocaudal

The images are taken from the cranial to the caudal end of a structure. In general, the images are taken from a top perspective in the CC or the Craniocaudal view.

There are two types of CCs RCC and LCC ie. Right Craniocaudal and Left Craniocaudal respectively of each corresponding breast.

2.2.2 Mediolateral Oblique

The MLO or the Mediolateral Oblique view is taken from the centre of the chest outward and is the most important projection as it allows depiction of most breast tissue.

There are two types of MLOs RMLO and LMLO ie. Right Mediolateral Oblique and Left Mediolateral Oblique respectively of each corresponding breast.

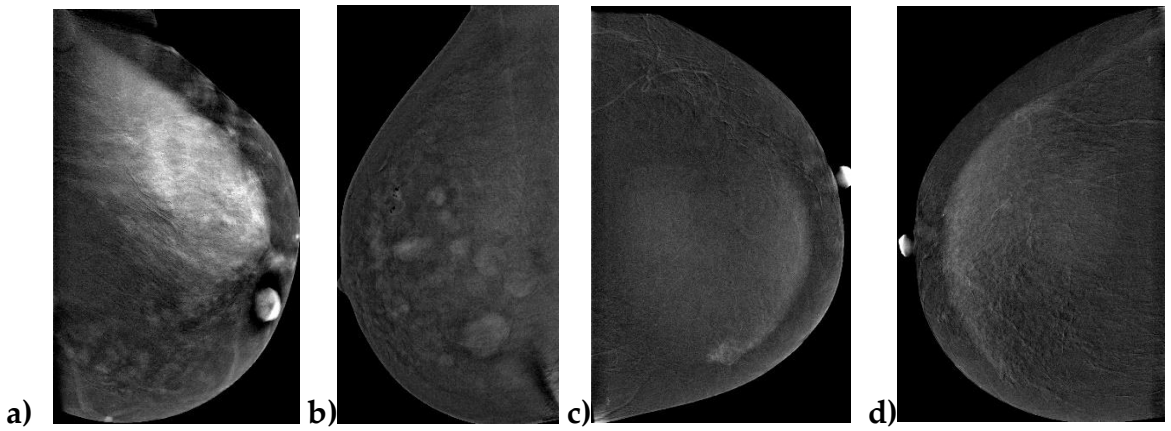


Figure 2. 1: a) RCC, b) LCC, c) RMLO, d) LMLO

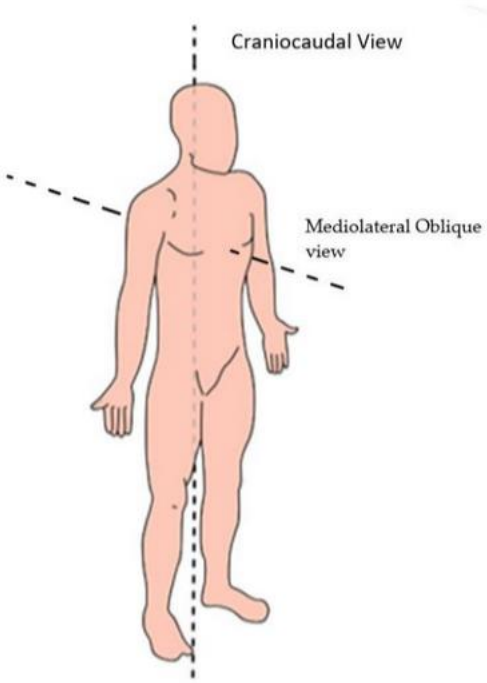


Figure 2. 2: Craniocaudal and Mediolateral Oblique View

2.3 DENSITY

Based on the density of the breasts the scans can be classified on basis of the BI-

RADS standards, which can be defined as follows:

- **BI-RADS 0**: Inadequate need for mammography and/or further imaging examination, recovering prior images that were inaccessible at the time of reading.
- **BI-RADS 1**: No masses, negative symmetry, architectural deformation, or aberrant calcifications.
- **BI-RADS 2**: Almost entirely fatty, benign, 0% likelihood of cancer
- **BI-RADS 3**: Scattered areas of fibro glandular density, short-interval follow-up is indicated, probably benign with a 2% probability of malignancy.
- **BI-RADS 4**: Heterogeneously dense, cancer is suspected. BI-RADS 4 may also be further categorized into the following categories, which mammography and ultrasound have a 2–94% probability of identifying:
    - **BI-RADS 4A**: Low risk of malignancy (2–9%).
    - **BI-RADS 4B**: High suspicion of malignancy (10–49% range)

- **BI-RADS 4C**: High risk of malignancy (50–94%)

  Consider doing a biopsy.

- **BI-RADS 5**: Extremely dense, strongly indicative of cancer (>95% chance of malignancy)
- **BI-RADS 6**: Known cancer with a positive biopsy

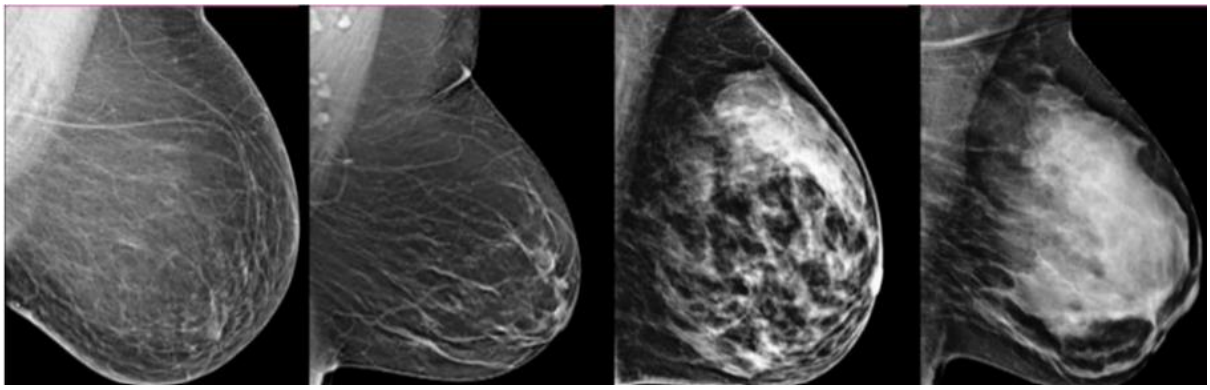


Figure 2. 3: Standard 4 class BI-RADS classification

2.4 ADVANTAGES OF CESM IMAGES OVER FFDM

1. Enhanced Lesion Detection:

- Need: CESM addresses the limitations of standard mammography by introducing a contrast agent that highlights blood vessels and abnormalities, significantly enhancing the detection of lesions.

- Technology: The injection of a contrast agent, typically iodine-based, coupled with dual-energy imaging, allows for the creation of subtracted images that emphasize areas with increased vascularity, improving lesion visibility.

2. Improved Sensitivity in Dense Breast Tissue:

- Need: Standard mammography faces challenges in detecting lesions in dense breast tissue. Enhanced sensitivity is crucial for early detection in such cases.

- Technology: CESM's dynamic contrast enhancement and dual-energy acquisition overcome the limitations of dense breast tissue, providing superior sensitivity through detailed vascular characterization.

3. Precise Differentiation of Lesion Characteristics:

- Need: Accurate differentiation between benign and malignant lesions is vital for effective clinical decision-making.

- Technology: CESM's emphasis on vascular characteristics allows for precise differentiation between benign and malignant lesions, contributing to more accurate diagnoses.

4. Advanced Imaging for Comprehensive Assessment:

- Need: Comprehensive assessment of breast abnormalities requires advanced imaging modalities to reveal intricate details.

- Technology: The integration of CESM involves obtaining a series of low- and high-energy images, followed by subtraction techniques, providing a more detailed and comprehensive view of breast tissue abnormalities.

5. Contributions to Personalized Medicine:

- Need: As medicine moves towards personalized approaches, there is a growing demand for imaging technologies that offer nuanced insights tailored to individual patient profiles.

- Technology: CESM's ability to capture detailed vascular information aligns with the trend towards personalized medicine, enabling tailored diagnostic and treatment strategies based on the specific characteristics of lesions.

Summary

In summary, the chapter discusses challenges in detecting small enhancing tumours obscured by high-density breast tissue in digital mammography, despite the use of iodinated contrast material. To overcome this, subtraction contrast-enhanced mammography (CEM) techniques were introduced. The initial method, temporal subtraction CEM, faced limitations such as prolonged compression times, capturing only one view of one breast, and susceptibility to misregistration artifacts. In response, a dual-energy subtraction technique utilizing the attenuation properties and k edge of iodine was developed, enabling the acquisition of both low-energy and iodine-enhanced images in a single, shorter compression.

CESM addresses the need for improved lesion detection, especially in dense breast tissue, and its technology, involving dynamic contrast enhancement and dual-energy imaging, empowers clinicians with enhanced sensitivity and detailed vascular characterization for accurate and personalized breast cancer diagnostics.

CHAPTER 3: LITERATURE REVIEW

3.1 INTRODUCTION

Cancer is the second leading cause of mortality among women, after only heart disease. Breast cancer will kill around 685,000 individuals worldwide by 2020, according to the World Health Organisation. Breast cancer accounts for the majority of all recorded cancer cases worldwide, overtaking lung cancer as the second greatest cause of death among women [1], followed by heart disease, and is anticipated to climb by 30% by 2025. Mammograms have long been used to assist early detection of breast cancer. Mammography use X-Rays to find potential ROI, and the possibilities make mammography an appealing solution due to its low cost and ease of accessibility in compared to other techniques like as PET, Ultrasound, MRI, and others, which are all expensive and difficult to get.

3.2 CADe/CADx AND DECSION SUPPORT SYSTEMS

Yoon and Kim in [5] describe the primary goals of using artificial intelligence (AI) to assist image interpretation as follows: 1) Computer-aided detection (CADe), also known as automated lesion detection, focuses on identifying worrisome anomalies in an image. 2) Computer-aided diagnosis (CADx), the process of describing anomalies found by either the radiologist or the computer. The interpreting radiologist decides the clinical significance of the discovered aberration and whether it requires additional study based on the CADe/CADx analysis. Kim et. al. in [6] report that early investigations noted that computer-aided detection (CAD), which functions as an automated second reader by flagging potentially problematic regions for radiologists to analyse, may increase the sensitivity of mammography.

Though image processing using DL (Deep Learning) is a comparatively newer domain but extensive research work has been already published. Earlier research work like that

of [7,8] report use of standard ML (Machine Learning) techniques like that of Decision Trees, SVM (Support Vector Machines) etc. and boosting algorithms like AdaBoost and XGBoost etc.

Contrast-enhanced mammography (CEM) uses iodinated contrast medium to enhance breast lesion visualization and assess tumour neovascularity. By combining high- and low-energy x-ray images, areas of contrast accumulation are highlighted. However, CEM introduces distinct workflows, artifacts, and risks related to contrast medium and radiation dosage. It's essential to understand the differences between CEM and conventional mammography, emphasizing the mechanisms enhancing image contrast in CEM. [19]

3.3 CRITICAL EVALUATION

3.3.1 Basis Of The Research

Khalid R. *et. al*. in [20] lay foundation for CESM data set that can be optimised for further use in diagnostic purposes using Artificial Intelligence being the basis . The data set comprises of three major parts namely low energy, high energy and the associated subtracted images derived using pixel wise subtraction and Contrast Enhancement of the two formulated in (1). The data set is one of the largest and only of the few open-sourced data sets available for public exploration of the same. The data set consists of total 2006 numbers of images collected over 326 female patients with normal images standing at 757 the benign and malignant at 587 and 662 respectively. The data set in 20 covers wide array of cancers. Some of the few amongst which are Invasive ductal carcinoma 445 (67.5%), Invasive lobular carcinoma 42 (6.3%), Mixed invasive ductal carcinoma and invasive lobular carcinoma 28 (4.2%), Ductal carcinoma insitu purely 17 (2.5%), Inflammatory breast cancer and others. The data used unlike the other public data sets encompasses over multiple age groups ranging from less than 40 years to greater than 70 years. The paper reports a manual annotation comprising of age, ACR, BIRADS,

pathology, reports tags image preview etc. in a separate folder long with manual annotation.

$$CE(b) = -\sum_{c=1}^{C}\sum_{i=1}^{N} y_{i_c}.\log \hat{y}_{i_c} \quad (1)$$

The research enclosed herewith tries classify the benign and malignant images using highly deep learning models using fine tuning of largest models (transfer learning) and localising the result in the breast images using GradCam [22]. This helps the radiologist / oncologist reading the reports to differentiate the reports between the calcification which are non-Mass lesions in the mammographs and can significantly reduce the pain and the associated cost and labour required to perform biopsies.

3.3.2 Critical Review

Hegazy R. *et. al.* in [23] work over a smaller dataset comprising of patients who were suspected to have intraductal papilloma on full-field digital mammography (FFDM) and ultrasound (US) examinations. The dataset suffers the two major issues major issues; 1) The dataset is quite small to jump to any significant conclusions, 2) the dataset provided in [23] covers only one type of cancer i.e. intraductal papilloma. Other significant works include that of Sorin V. *et. al.* in [24] where the author uses a total of 15 cancers present in 14 different women the patients in the study included intraductal papilloma but again the study has a serious concern of a significantly small dataset the paper also reports that the sensitivity of CESM when compared to FFDM jumped significantly from 50% to a maximum of 87.5%.

Hassan A. S. *et. al.* in [25] uses deep convolutional neural networks and transfer learning to classify breast cancer masses the DCNNs were trained for standard FFDMs instead of CESM images the AUC reported at the experimentation was 88% and consecutively were also tested. Yala A. *et. al.* in [26] also introduces the use of deep learning for use in classification of standard mammograph images and results in an AUC of 0.7 for image only DL where an AUC of 0.68 for hybrid-DL. Shen L. *et. al. in* [27] published in reputed

Nature also uses DL to improvise cancer detection on FFDM images reporting a sensitivity of 86.1% and a specificity of 80.1% while being trained on CBIS-DDSM dataset. Out of all the papers reviewed the best standing paper bubbled out to be that by Chi X. *et. al.* in [28] where the study aimed to improve the diagnostic accuracy of breast diseases by combining breast imaging–reporting and data system (BI–RADS) with the enhancement intensity and pattern of contrast-enhanced spectral mammography (CESM). The paper uses the Receiver Operating Characteristic (ROC) curve as the performance metric and reports an AUC of 0.858 in case of BaC and only 0.790 in case of CESM category and explores multiple type of cancers. The documents also published an extensive comparison of using only BI-RADS, CESM and BaC (combination of BI-RADS, CESM) using sensitivity, specificity, Youden's Index etc. However, the only issue with the experimentation is that of the number of images which stand at 312 with benign lesions at 109 and malignant lesions at 203.

ElSaid *et. al.* in [29][30] explains the CESM as the technique relies on the assessment of the vascularity of the tumor where the non-ionic iodine based contrast material injected during mammography is only taken by the viable cancer cells.

Summary

The chapter discusses breast cancer as a significant cause of mortality globally, highlighting the role of mammography and contrast-enhanced mammography (CEM) in early detection. It introduces the use of artificial intelligence (AI) for computer-aided detection (CADe) and diagnosis (CADx). A critical evaluation of research studies using deep learning for breast cancer classification is provided, emphasizing the importance of large datasets. The text concludes by summarizing a study combining BI-RADS with contrast-enhanced spectral mammography for improved diagnostic accuracy, acknowledging limitations in sample size. Overall, it offers a comprehensive overview of breast cancer diagnosis challenges, AI applications, and critical assessments of research findings.

CHAPTER 4: METHODOLOGY

The following chapter gives in brief the methodologies adapted in order to receive better than average results.

4.1 PRE-PROCESSING

4.1.1 Data Arrangement

The initial stage is of data arrangement where the root folder of the data set which comprised of Low-Energy images, High-Energy images and Subtracted images folders is split into two sub folders titled Normal, Benign and Malignant, the frequency of each of which is displayed in figure 4.1. Each of the three folders is consecutively split into training, testing and validation folders. The standard split of training and testing is set at 4:1 or 80% to 20%. The validation data is further carved out of training data, the composition here as well remains 80% to 20%.

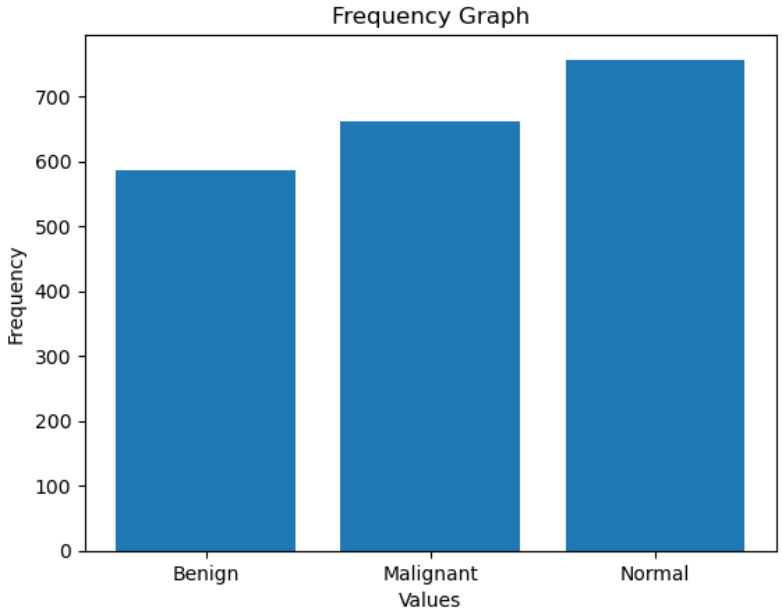


Figure 4. 1: Frequency Graph of Images

4.1.2 Resizing

The images are resized to a standard of 224x224 and the transfer learning model are tuned accordingly in order to align to the input size. Resizing is a key preprocessing

method employed in the fields of image processing and computer vision. The process entails modifying the proportions of a picture to conform to a defined size or aspect ratio. The process of resizing is of utmost importance in order to guarantee that all photos within a dataset possess same dimensions, hence rendering them suitable for integration with deep learning models that need consistent input sizes. Various techniques are commonly employed to resize images, such as scaling, cropping, or padding, with the aim of attaining the necessary dimensions while maintaining the image's aspect ratio.

4.1.3 Pixel Normalisation

Each of the images from the data set after being resized into a standard size is normalized using the normal values of mean and deviation from the large ImageNet which are imported using the OpenCV library. The normalization each image is carried after conversion to a standard data format like .jpg or .png. Each of the three channels i.e. R-Red, G-Green, B-Blue is normalized and standardized separately. A normalized image has been displayed in figure 4.2.

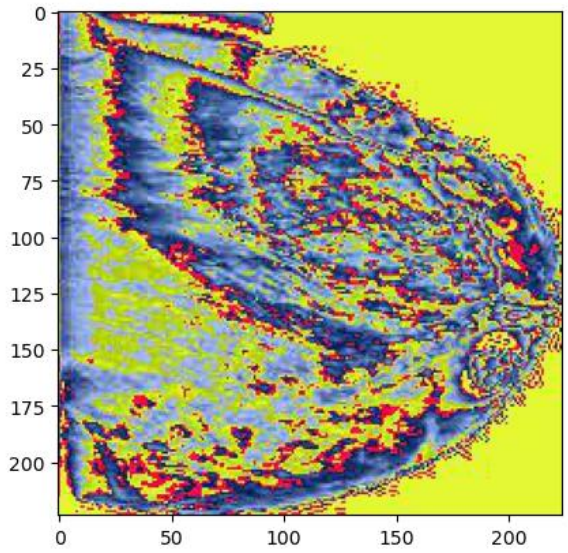


Figure 4. 2: Normalised CESM image

4.1.4 Colour Space Conversion

The images were initially transformed into grayscale. The conversion of pictures to grayscale, a technique that alters full-color images to black and white, presents several benefits in the fields of image processing and computer vision. To begin with, grayscale images provide computational simplicity, hence decreasing memory and processing demands in contrast to their full-color counterparts. This characteristic renders them more appropriate for applications with limited resources or for processing in real-time scenarios. Furthermore, the utilization of grayscale pictures has the potential to improve the resilience of computer vision algorithms by mitigating the adverse effects caused by fluctuations in colour, shadows, and lighting circumstances. The process of simplification frequently leads to enhanced performance in object identification, edge detection, and feature extraction. Furthermore, grayscale pictures possess enhanced interpretability, hence facilitating a more lucid visualization and study of texture, patterns, and structural intricacies within the image.

4.1.5 Image Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a highly effective method employed in the field of image processing for the purpose of enhancing images. In contrast to conventional histogram equalization, the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique demonstrates the ability to adjust to the specific attributes of a given picture, hence restricting the amplification of noise in regions exhibiting less contrast. The adaptable characteristic of CLAHE enables it to effectively increase details in regions with both high and low lighting conditions. This quality makes it particularly advantageous for enhancing the visibility of intricate structures and features in many domains such as medical imaging, remote sensing, and other related industries. In addition, the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique offers the ability to regulate the degree of contrast

augmentation, so mitigating the risk of excessive amplification of localized characteristics and guaranteeing the maintenance of the image's overall integrity.

4.1.6 Image Augmentation

Image augmentation is a crucial technique in deep learning, enhancing the robustness and generalization of models by artificially expanding the diversity of the training dataset. For Contrast-Enhanced Spectral Mammography (CESM) images, the ImageDataGenerator class is often employed with specific parameters to introduce variations while maintaining the integrity of the medical images.

In this particular setup, the ImageDataGenerator is configured with rescale=1/255 to normalize pixel values, ensuring compatibility with neural network training. The validation_split parameter is set to 0.2, allowing for the automatic creation of a validation dataset during training, aiding in model evaluation.

The augmentation parameters include rotation_range, width_shift_range, height_shift_range, shear_range, zoom_range, and horizontal_flip. Rotation_range introduces random rotations up to 10 degrees, while width_shift_range and height_shift_range allow horizontal and vertical shifts of up to 10% of the image dimensions. Shear_range introduces shearing transformations, zoom_range enables random zooming up to 10%, and horizontal_flip randomly flips images horizontally.

By applying these augmentation techniques, the ImageDataGenerator diversifies the training dataset, exposing the model to various perspectives, orientations, and scales of CESM images. This not only aids in preventing overfitting but also ensures the model's adaptability to variations in patient positioning, scanner settings, and other factors encountered in real-world scenarios.

4.2 DEEP NEURAL NETWORK ARCHITECTURE

The "V2" versions typically refer to updated or improved versions of the original models or frameworks. For example, a pre-trained model like ResNet might have a ResNetV2

version that includes enhancements, architectural changes, or improvements in training methodology. These updates aim to address limitations or challenges observed in the original versions, providing better performance, efficiency, or generalization capabilities. The "V2" versions denote updated or improved versions of these models, incorporating refinements and advancements to enhance their capabilities.

External Head – The custom output head of the transfer learning models is designed as shown in figure 4.3.

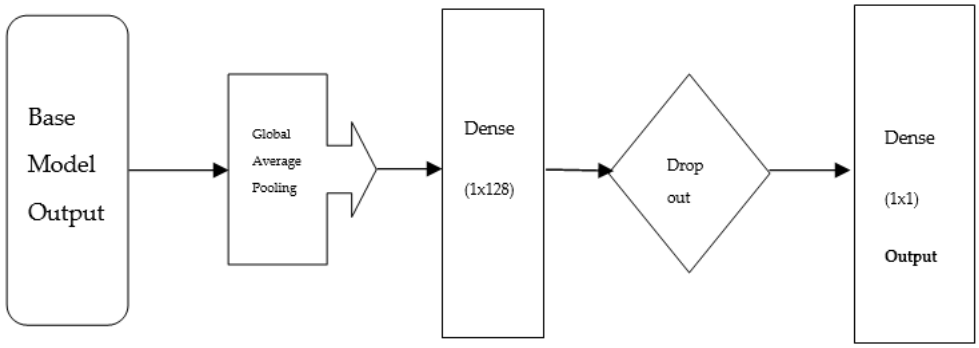


Figure 4. 3: Output head of the model

4.2.1 ResNet50V2

ResNet50V2 [10] is an enhanced version of the ResNet50 architecture, renowned for its impact on deep learning, especially in computer vision. Developed by Microsoft Research, ResNet50V2 introduces a pre-activation mechanism, optimizing skip connections for better information flow and training efficiency. The architecture maintains a modular design, making it adaptable to various tasks. With its improved structure, ResNet50V2 continues to be influential in tasks such as image classification, object detection, and image segmentation. The key innovation in ResNet50V2 is the introduction of a new bottleneck design. This design optimizes the use of skip connections, or residual connections, by incorporating a 'pre-activation' mechanism. Unlike the original ResNet where the activation was applied after the convolutional layer, ResNet50V2 applies activation functions before the convolutional layers. This alteration facilitates better information flow, reduces the likelihood of vanishing

gradients, and enhances the overall training efficiency of the network. The last 3 layers in figure 4.4 are replaced by architecture defined in figure 4.3. and consecutively the stages are repeated for 50 times.

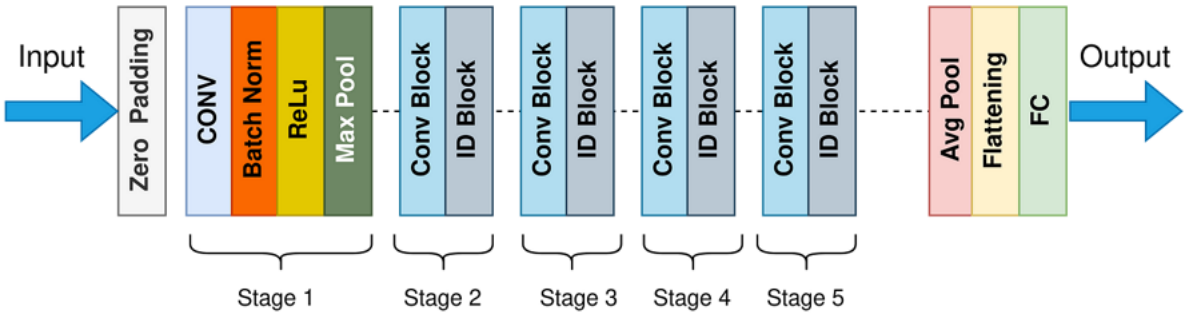


Figure 4. 4: Architecture of ResNet50V2

4.2.2 ResNet101V2

ResNet101V2 [10] and ResNet50V2 are both iterations of the ResNet architecture, sharing fundamental features such as the pre-activation mechanism, modularity, and the utilization of skip connections. Both models are well-suited for computer vision tasks, including image classification, object detection, and image segmentation. They demonstrate effectiveness in various benchmarks and competitions, providing reliable solutions for diverse applications. The key point of differentiation lies in their depth; ResNet101V2 boasts a more extensive architecture with 101 layers, offering a higher model capacity, while ResNet50V2 is comparatively shallower with 50 layers. This depth discrepancy influences factors such as computational requirements and the ability to capture intricate features. The choice between ResNet101V2 and ResNet50V2 depends on the specific demands of the task, with ResNet50V2 often favoured for scenarios requiring a balance between performance and computational efficiency, and ResNet101V2 chosen for more complex tasks that benefit from its deeper architecture.

4.2.3 InceptionV3

The InceptionV3 [31] architecture is a well-known convolutional neural network that has made significant progress in the field of computer vision. Inception V3, a model

developed by Google Research, belongs to the Inception model lineage. Inception V3 is widely recognised for its exceptional efficiency and efficacy, particularly in the domains of picture categorization and related tasks. The objective of this investigation is to investigate the operating mechanisms and essential properties of Inception V3. The Inception V3 model is characterized by its integration of "Inception modules," which are comprised of a series of convolutional layers featuring many branching, each employing different kernel sizes. These modules facilitate the network's ability to simultaneously capture features at various scales. A typical Inception module is composed of convolutional filters of sizes 1x1, 3x3, and 5x5, along with max-pooling processes. The network demonstrates efficacy in gathering information across various scales through the utilization of filters with varying sizes. The Inception V3 model employs 1x1 convolutions, known as "bottleneck" layers, within its Inception modules to efficiently reduce the dimensionality of the feature maps. The bottleneck layers serve two main purposes: reducing computational load and improving the network's capacity to acquire concise and informative representations. As a result, they facilitate faster and more efficient training processes. Parallel processing is a basic characteristic of the Inception modules, as they are specifically engineered to do concurrent computations on several receptive field widths. This enhances the network's capacity to effectively capture a wide range of features at various scales. The final result of each Inception module is achieved by combining the outputs of its several branches along the channel dimension through concatenation. The Inception V3 model integrates auxiliary classifiers at two specific depths inside the network architecture. The use of auxiliary classifiers is employed to address the challenge of the vanishing gradient problem and offer additional assistance during the training phase. The primary objective of their design is to optimize the process of generating intermediate predictions and gradients, hence improving the overall efficiency of error backpropagation. Batch Normalization is a technique implemented in Inception V3, a current deep learning architecture, that serves to normalize the inputs to each layer during the training process. This specific methodology

facilitates the stabilization and acceleration of the training process by addressing the issue of internal covariate shift.

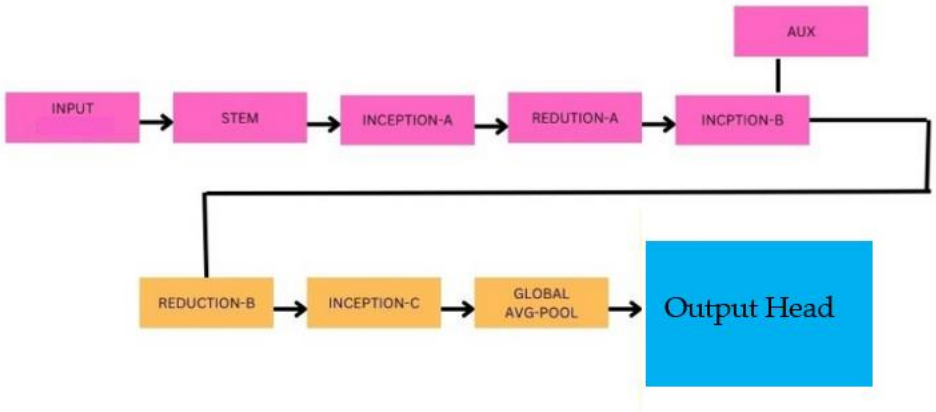


Figure 4. 5 : InceptionV3 architecture

4.2.4 DenseNet169

DenseNet169 [32], an extension of the DenseNet architecture, stands out as a powerful convolutional neural network with a distinctive approach to feature connectivity. The "169" in its name signifies its depth, encompassing a total of 169 layers.

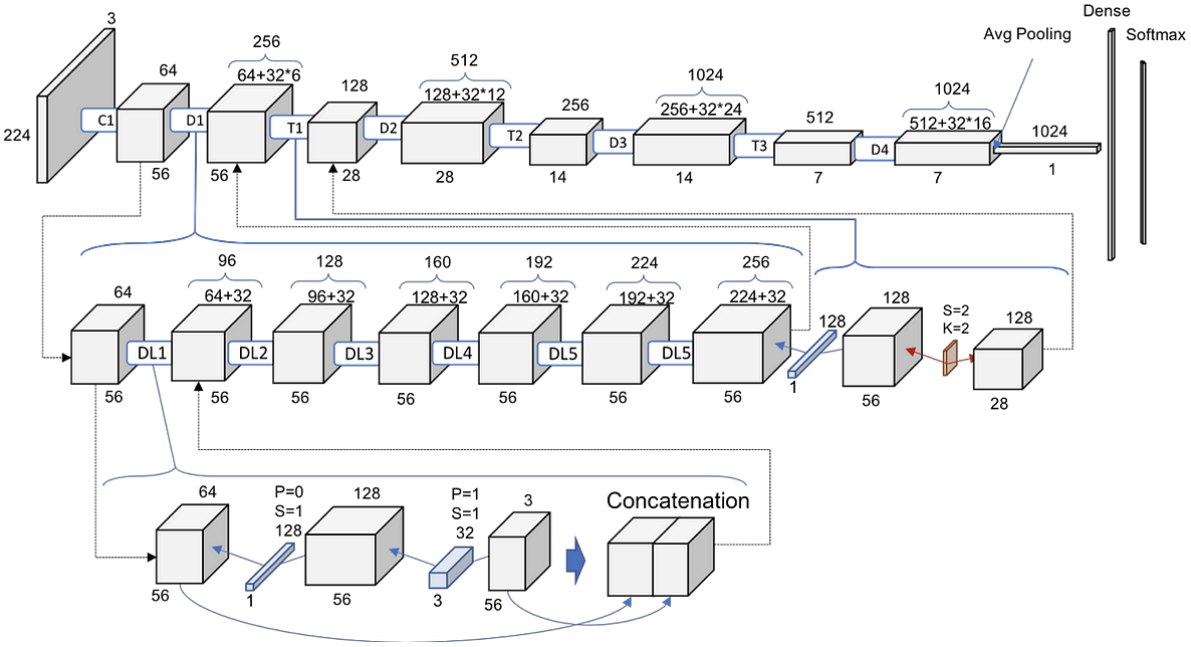


Figure 4. 6: DenseNet architecture

DenseNet architectures are characterized by dense connectivity, a unique feature where each layer receives direct input from all preceding layers. This dense connectivity fosters a highly efficient flow of information and gradients throughout the network.

Designed primarily for image-related tasks, DenseNet169 excels in capturing intricate patterns and features within data. The dense connectivity encourages feature reuse, mitigates vanishing gradient issues, and contributes to the model's overall efficiency. The architecture is composed of dense blocks, transition layers, and a global average pooling layer, ensuring a compact yet powerful representation.

4.2.5 DenseNet201

DenseNet201 [32] and DenseNet169 both belong to the DenseNet architecture, renowned for its dense connectivity and efficient parameter sharing across layers. Sharing common features such as dense blocks, transition layers, and global average pooling, these architectures are designed for computer vision tasks, including image classification and object detection, where their dense connectivity proves beneficial for nuanced feature learning. However, the key distinction lies in their depth, as indicated by the numerical suffixes. DenseNet201, with 201 layers, boasts a higher model capacity, potentially capturing more complex features. On the other hand, DenseNet169, with 169 layers, maintains a slightly shallower architecture, offering a balance between performance and computational efficiency. The choice between DenseNet201 and DenseNet169 depends on the specific demands of the task, with the former favored for intricate tasks requiring deeper networks and the latter considered for scenarios where computational resources are a constraint. Both models are versatile and often employed in transfer learning, adapting pre-trained networks to specific tasks with relative ease.

4.2.6 MobileNetV2

MobileNetV2 [33] is a lightweight and efficient convolutional neural network architecture designed for mobile and edge devices. It serves as an evolution of the original MobileNetV1, introducing improvements in terms of both performance and

computational efficiency. MobileNetV2 is characterized by a streamlined architecture that employs depth wise separable convolutions, which significantly reduce the number of parameters and computations compared to traditional convolutions. The network is built upon inverted residuals and linear bottlenecks, allowing for better information flow within the network while minimizing computational costs. With a focus on achieving high accuracy with reduced latency, MobileNetV2 is well-suited for real-time applications on devices with limited computational resources. Its compact design makes it particularly valuable for tasks such as image classification, object detection, and semantic segmentation in scenarios where computational efficiency is a priority. MobileNetV2's success lies in its ability to strike a balance between model accuracy and the demand for lightweight, deployable models on resource-constrained devices. The figure 4.7 shows architecture of the MobileNetV2 and the output head once again is replaced by the output architecture from the figure 4.3.

4.2.7 NASNetMobile

NASNetMobile [34], short for Neural Architecture Search Network Mobile, is a neural network architecture designed through neural architecture search (NAS), a process that automates the exploration of model architectures. Developed by Google,

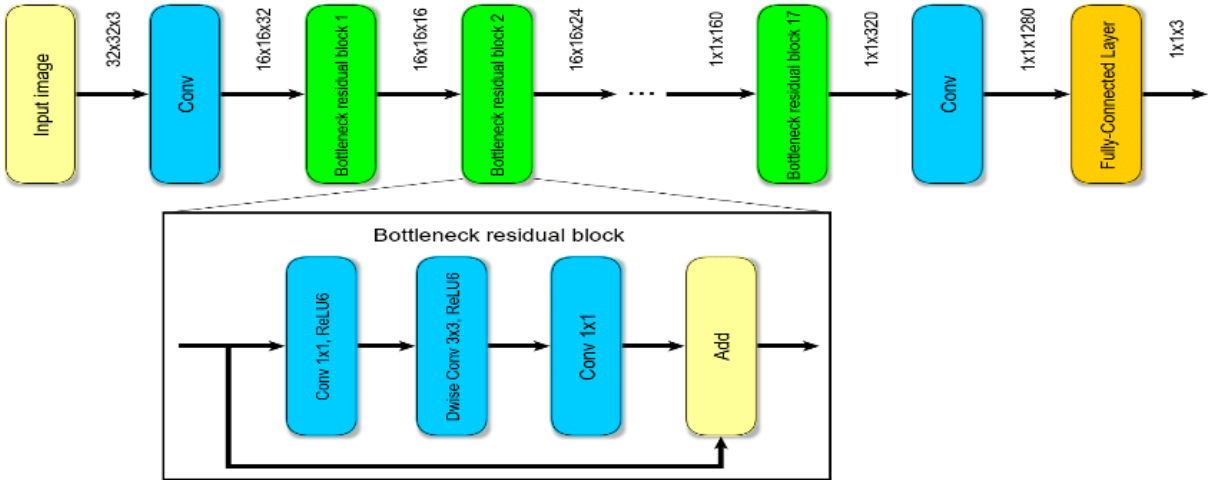


Figure 4. 7: MobileNetV2 architecture

NASNetMobile is part of the NASNet family of models and is specifically designed for mobile and edge devices. Notable for its efficiency and adaptability, NASNetMobile incorporates architectural innovations discovered through an automated search process.

Key characteristics of NASNetMobile include the use of hierarchical blocks, where the network learns to repeat certain blocks at different scales, and the integration of learned cell structures that optimize the trade-off between accuracy and computational efficiency. This neural architecture search enables NASNetMobile to adapt its structure to different tasks and datasets, making it versatile for various computer vision applications.

Due to its automated design process, NASNetMobile is capable of achieving competitive performance while maintaining a compact architecture. This makes it well-suited for deployment on devices with limited computational resources, where its efficiency is crucial. NASNetMobile has found applications in tasks such as image classification, object detection, and feature extraction, showcasing its effectiveness in scenarios where lightweight models are essential for real-time and resource-constrained environments. Figure 4.8 shows NASNetMobile architecture where the output head is replaced by the figure 4.3 architecture.

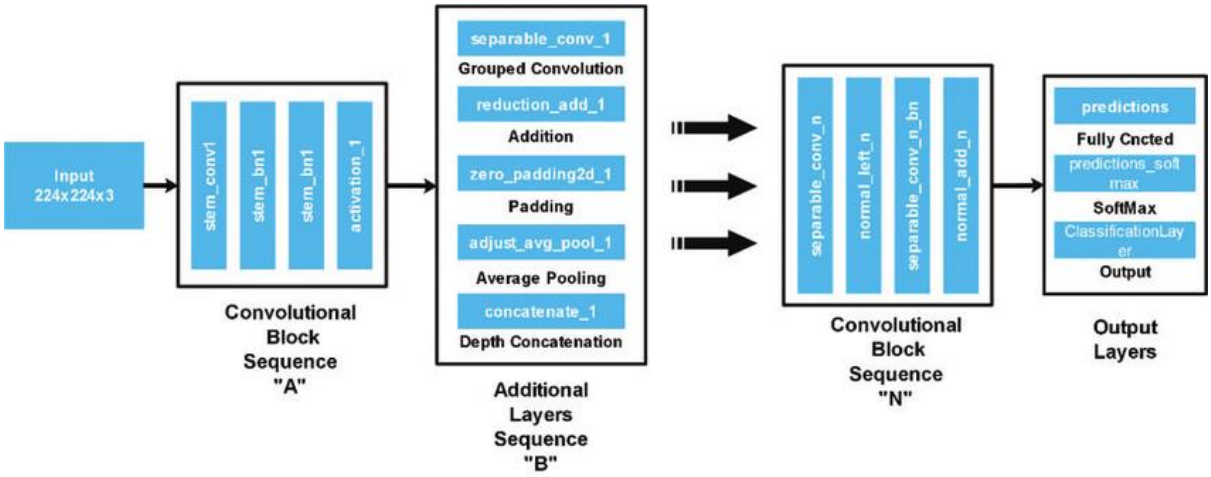


Figure 4. 8: NASNetMobile Architecture

4.2.8 NASNetLarge

NASNetLarge represents a more substantial and complex architecture optimized for high performance in computer vision tasks. With an emphasis on accuracy, NASNetLarge is equipped with a larger number of parameters and intricate connections, enabling it to capture intricate features and representations in data. This variant is suitable for applications with more abundant computational resources, where achieving top-tier accuracy in tasks such as image classification and semantic segmentation is paramount. NASNetMobile prioritizes efficiency, making it well-suited for mobile and edge deployments. In contrast, NASNetLarge sacrifices some efficiency for enhanced accuracy, making it more suitable for scenarios where ample computational resources are available. The choice between the two depends on the specific requirements of the task, the available hardware, and the importance placed on achieving the highest accuracy versus operational efficiency.

4.3 KERAS TUNER

Keras Tuner is a powerful tool for automating the optimization of hyperparameters in deep learning models. It streamlines the process by defining a search space for hyperparameters, selecting an appropriate tuner, and specifying a model-building function and objective metric to optimize, such as accuracy for classification tasks. During the tuning process, the tuner iteratively trains and evaluates various model configurations, sampling hyperparameters from the defined space. The goal is to discover the combination that maximizes the chosen metric which was accuracy in our case. Keras Tuner reduces the manual effort of fine-tuning and helps find the best-performing model architecture and hyperparameter settings for your specific task.

4.4 TRAINING PARAMETER – YOUDEN'S INDEX

Youden's Index, also known as the Youden's J statistic or J index, is a performance metric commonly used in binary classification tasks. It provides a comprehensive measure of a diagnostic test's ability to balance sensitivity and specificity. In the context of medical

imaging, particularly Contrast-Enhanced Spectral Mammography (CESM) images, Youden's Index plays a crucial role in evaluating the performance of classification models. Formula for Youden's Index (J)

$$J = \text{Sensitivity} + \text{Specificity} - 1 \quad (4.1)$$

Where sensitivity and specificity are from equation number 1.5 and 1.6.

Interpretation

- Youden's Index ranges from -1 to 1.
- A value of 0 indicates that the test has no discriminatory power.
- The closer the index is to 1, the better the balance between sensitivity and specificity.

Application in CESM Images

In the realm of medical imaging, such as CESM, Youden's Index is often utilized as a training parameter for machine learning models, particularly those involved in the detection or classification of breast lesions. CESM, being a contrast-enhanced imaging modality, aims to enhance the visualization of breast lesions and neovascularity associated with tumors.

When utilizing Youden's Index in the training of machine learning models for CESM images, the objective is to find a classification threshold that maximizes both sensitivity (the ability to correctly identify positive cases) and specificity (the ability to correctly identify negative cases). By optimizing Youden's Index during the training process, the model aims to strike an optimal balance, minimizing false positives and false negatives.

The training parameter derived from Youden's Index helps in determining a classification threshold that is well-suited for the specific characteristics of CESM images. This is essential in clinical applications where accurate detection of abnormalities, such as tumors or lesions, is critical. Fine-tuning machine learning models with Youden's

Index as a guiding metric contributes to enhancing the diagnostic accuracy of CESM-based systems, making them more reliable tools for early detection and assessment of breast-related pathologies.

4.5 LOCALIZATION

4.5.1 GradCAM [22]

The localisation of the region affecting the output of the Deep Convolutional Neural Network (DCNN) can be traced out using tools like GradCam. Grad-CAM, short for Gradient-weighted Class Activation Mapping, is an interpretability tool used in the field of deep learning to visualize the regions of an input image that contribute most significantly to the decision made by a convolutional neural network (CNN). This technique has proven to be valuable in elucidating the features and patterns within an image that are crucial for the model's classification or decision-making process.

The utilization of Grad-CAM involves several key steps. Firstly, a pre-trained CNN, typically employed for image classification tasks, is selected. Following this, the last convolutional layer of the chosen network is identified. Grad-CAM operates by deriving the gradients of the target class with respect to the final convolutional layer's feature maps. These gradients essentially signify the importance of each feature map in the decision process.

The subsequent step involves computing the global average pooling of the gradients obtained. This pooling step produces a set of weights that reflect the importance of each feature map. These weights are then used to compute a weighted sum of the feature maps, resulting in a coarse localization map. This map highlights the regions in the input image that are influential in the network's decision for a specific class.

The final Grad-CAM output is a heatmap overlaid on the original input image, visually indicating the regions that had the most impact on the network's classification. This

interpretability tool facilitates a more transparent understanding of the CNN's decision-making process, contributing to enhanced model transparency and trust.

In the realm of breast cancer diagnosis, particularly in Contrast-Enhanced Spectral Mammography (CESM) images, Grad-CAM stands as a valuable tool for localization, offering insights into regions of interest (ROIs) that influence the model's decision-making process. This involves selecting a pre-trained convolutional neural network (CNN) model that has been adeptly trained on relevant datasets for breast cancer detection. Grad-CAM is then integrated into the model to allow for the visualization of significant regions within CESM images during the evaluation phase.

The integration process requires fine-tuning or training the model on a CESM-specific dataset, complete with annotations indicating the presence and location of abnormalities, such as tumours or lesions. Upon successful training, Grad-CAM is employed to generate heatmaps that visually highlight areas in CESM images that play a pivotal role in the model's classification decisions. These heatmaps essentially provide a spatial representation of the region’s most crucial for the model's predictions.

We can leverage these Grad-CAM heatmaps by overlaying them onto the original CESM images. This visual interpretation aids in identifying potential areas of concern and understanding the specific features or patterns considered significant by the model. Grad-CAM becomes instrumental in providing transparency into the decision-making process, enhancing the interpretability of the AI model's outputs.

The assistance offered by Grad-CAM extends beyond visualization; it plays a critical role in the decision-making process for biopsies. The localized areas highlighted by Grad-CAM serve as a guide for clinicians in determining whether further diagnostic procedures, particularly biopsies, are warranted. Consistent patterns of attention identified by Grad-CAM may indicate the presence of abnormalities that merit closer inspection and validation through additional medical procedures.

Clinical validation studies are paramount in assessing the reliability and accuracy of Grad-CAM-assisted localization in CESM images. Collaboration with medical professionals ensures that Grad-CAM predictions align with actual biopsy results and clinical findings. Through iterative improvement, the model, and its integration with Grad-CAM, can be refined based on feedback from clinicians, ultimately enhancing its accuracy and reliability over time. In essence, Grad-CAM emerges as a supportive tool, aiding in the transparent localization of potential abnormalities in CESM images and contributing to informed decision-making in the context of breast cancer diagnosis.

4.5.2 Saliency Heatmaps [35,36]

Saliency heatmaps, similar to Grad-CAM, are instrumental in providing interpretability and transparency in deep learning models, particularly in the context of breast cancer diagnosis using Contrast-Enhanced Spectral Mammography (CESM) images. Saliency maps highlight regions of importance within an image, offering insights into the features that significantly influence a model's decision.

The utilization of saliency heat maps in the context of CESM images involves similar initial steps. A pre-trained convolutional neural network (CNN) designed for breast cancer detection is selected and integrated with the saliency map generation process. The model is trained or fine-tuned using a CESM-specific dataset, incorporating annotations that indicate the presence and location of abnormalities.

Saliency maps are generated by computing the gradient of the predicted class score with respect to the input image pixels. In simpler terms, this process identifies which pixels contribute the most to the model's decision. The resulting saliency heat map is then overlaid onto the original CESM image, visually emphasizing regions that have a substantial impact on the model's classification.

For clinicians and radiologists, these saliency heat maps serve as a powerful tool for understanding the focal points that guide the model's decision-making process. By visually highlighting areas of significance, the heat maps facilitate a more intuitive

interpretation of why a particular classification was made. This transparency is crucial in the medical field, where understanding the reasoning behind a model's decision is paramount.

In the context of breast cancer diagnosis, saliency heat maps can aid in localizing potential abnormalities within CESM images. Similar to Grad-CAM, these maps contribute to the decision-making process regarding the necessity of biopsies. Consistent patterns of attention in saliency heat maps may indicate regions that require further investigation, assisting clinicians in making informed decisions about additional diagnostic procedures.

As with any interpretability tool, the effectiveness of saliency heat maps is contingent on rigorous validation against clinical outcomes. Collaborative efforts between data scientists and medical professionals ensure that the highlighted regions align with actual pathological findings. Continuous refinement of the model based on this feedback enhances its reliability, making saliency heat maps a valuable asset in the arsenal of tools for transparent and interpretable breast cancer diagnosis using CESM images.

Summary

This chapter serves as a detailed guide through the intricacies of model training, offering insights into the methodologies and techniques applied to enhance overall performance. The narrative begins with a comprehensive exploration of data preprocessing, emphasizing the critical role of techniques such as data augmentation and normalization in bolstering the model's resilience.

The strategic implementation of transfer learning and fine-tuning is elaborated upon, showcasing their significance in leveraging pre-trained models for expedited training and improved generalization.

In addition to model training, the chapter delves into sophisticated regularization methods, including dropout and weight decay, employed to curb overfitting and elevate

the model's capacity for generalization. The intricate process of hyperparameter tuning is thoroughly examined, emphasizing the delicate equilibrium sought between model complexity and performance.

Notably, the chapter extends its purview to include localization strategies, shedding light on techniques employed to identify and highlight specific regions of interest within the data.

In summary, this chapter stands as a comprehensive guide to the model training journey, encompassing techniques that refine and optimize the model. The inclusion of localization strategies broadens its scope, highlighting the model's ability to discern and emphasize specific regions of interest within the dataset.

CHAPTER 5: RESULTS AND DISCUSSION

This chapter serves as the canvas upon which the outcomes of our endeavors are painted, providing a comprehensive portrayal of the empirical findings and their subsequent interpretation.

5.1 CLASSIFICATION

The initial model tested and trained over the dataset was that of ResNet50V2 and yielded a result of AUC reaching approximately 0.73 and Sensitivity (True Positive Rate or TPR) at 0.889, the Specificity (True Negative Rate or TNR) at 0.615. On using the higher version with 101 residual cross connections ResNet101V2 the final AUC score of 0.72 was achieved consecutively the Sensitivity or TPR was reported at 0.90 with Specificity or TNR was achieved with a score reaching approximately 0.62. While using the ResNet152V2 version the AUC score was as high as 0.79 while the Sensitivity remained stable at 0.89 the Specificity dropped to score of 0.54.

While using the InceptionV3 the AUC score of the corresponding ROC curve soared to a high of 0.83 the Sensitivity also soared drastically to 0.95 while the Specificity was considerably shrunk to the score of 0.31.

The DenseNet169 had a normal AUC score of 0.71 the Sensitivity and Specificity both reported at 0.72 and 0.62 respectively. Thus, the TPR to TNR ratio was nearest to 1 while using DenseNet. The DesnseNet201 gave an AUC score of 0.66 the Sensitivity was reported to be 0.72, the Specificity also dropped though not very significantly to a score of 0.62

The MobileNetV2 and MobileNet performed quite similarly though poorly in terms of comparison with other models as the AOC remained at approx. 0.72 with Sensitivity stationary at nearly 0.89 and the Specificity approximately dipping to a low of 0.46

The NASNetMobile resulted in an AUC score of 0.78, the TPR and TNR while using the NASNetMobile came up to be 0.78 and 0.70 respectively. Although the model did not yield any significant score both in terms of AUC, Sensitivity and Specificity but the ratio of TPR to TNR stood quite close to 1. While using the NASNetLarge the AUC jumped although not very significantly to an AUC score of 0.80. Also, the Sensitivity stood at 0.889 and the Specificity was reported to be 0.692.

From the analysis above it can be drawn that certain models show and report quite similar output results, the reason for which can be concluded due to the fact that the models were quite similar in terms of the output head attached over these pre-trained models but working of different models is different. The other reason include that the outputs are significantly rounded off and the models reported are only those of which are considered apt for the same.

Table 5. 1: Performance Metrics of various Models

Model	AUC	Sensitivity	Specificity	TPR/TNR Ratio
ResNet50V2	0.73	0.889	0.615	1.44
ResNet101V2	0.72	0.90	0.62	1.45
ResNet152V2	0.79	0.89	0.54	1.65
InceptionV3	0.83	0.95	0.31	3.06
DenseNet169	0.71	0.72	0.62	1.16
DenseNet201	0.66	0.72	0.62	1.16
MobileNetV2	0.72	0.89	0.46	1.93
NASNetMobile	0.78	0.78	0.70	1.11
NASNetLarge	0.80	0.889	0.692	1.28

The graphs for the training accuracy (Youden's Index), the loss and the same for the validation set along with the ROC curves can be found attached as Appendix-I.

5.2 LOCALIZATION

5.2.1 Grad-CAM results

- GradCAM visualization for Malignant case in MLO view

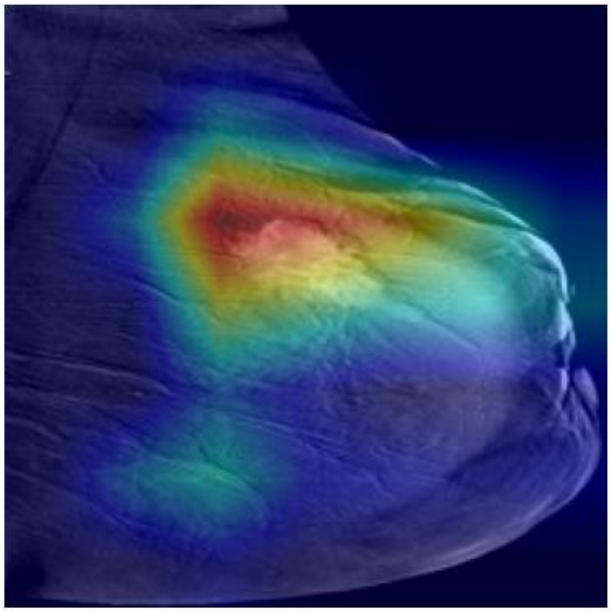


Figure 5. 1: GradCAM visualization for malignant case (MLO)

- GradCAM visualization for Malignant case in CC view with multiple lesions

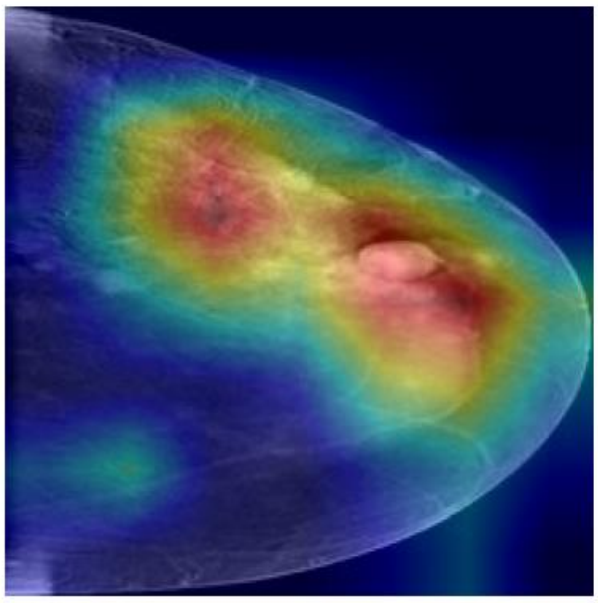


Figure 5. 2: GradCAM visualization for malignant case (CC)

- GradCAM visualization for benign case in MLO view.

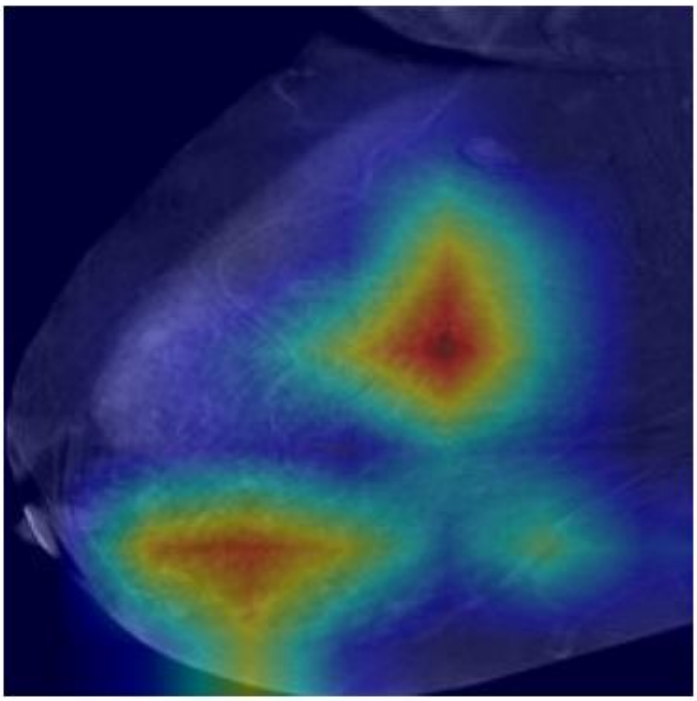


Figure 5. 3: GradCAM visualization for benign case

5.2.2 Saliency Heat-maps

- Saliency heatmap for Malignant case

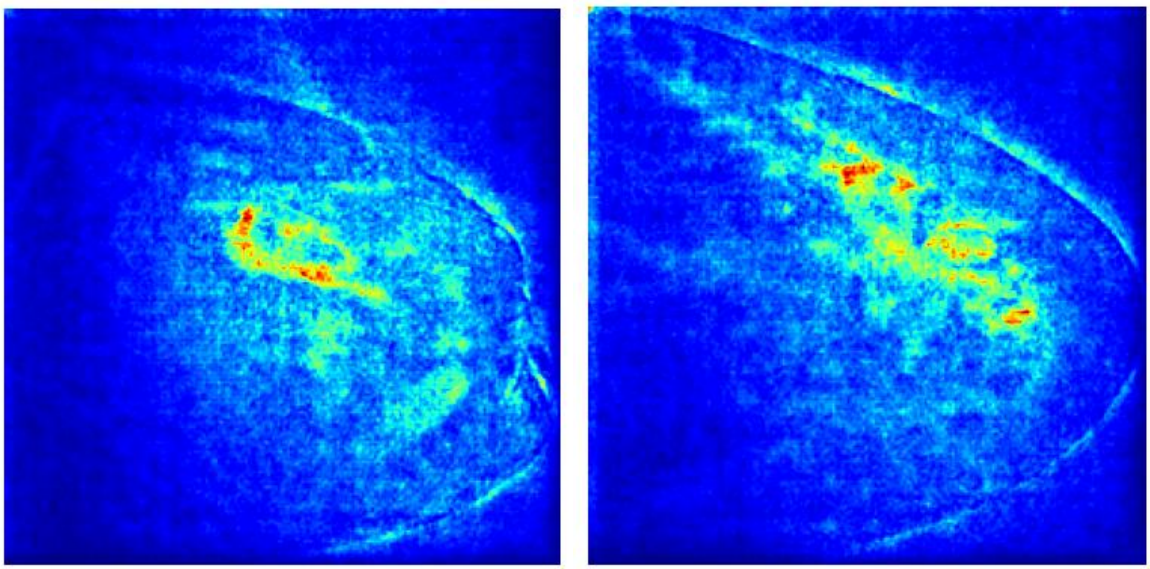


Figure 5. 4: Saliency Heatmap for Malignant cases

- Saliency heatmap for Benign case

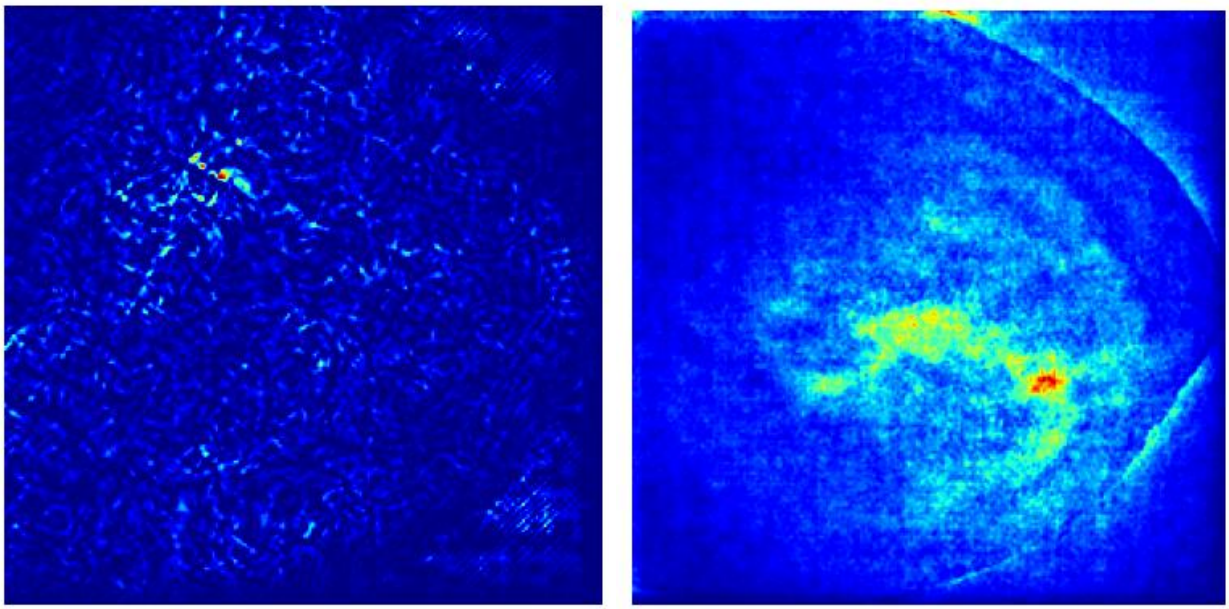


Figure 5. 5: Saliency Heatmap for benign cases

Summary

The section provides a detailed examination of various deep learning models applied to the dataset. Notable findings include ResNet50V2 with an AUC of 0.73, ResNet101V2 exhibiting increased Sensitivity at 0.90, and ResNet152V2 with a higher AUC of 0.79. InceptionV3 stands out with an impressive AUC of 0.83 but sacrifices Specificity. DenseNet169 maintains consistency with an AUC of 0.71, while DenseNet201 shows similar patterns. MobileNetV2 and MobileNet perform less favourably with lower Specificity. NASNetMobile achieves a balanced AUC of 0.78, and NASNetLarge improves with an AUC of 0.80. The section emphasizes nuanced model behavior, encouraging a deeper understanding of their effectiveness. Appendices provide graphical representations for further insights.

CHAPTER 6: CONCLUSION AND FUTURE SCOPE

As we wrap up our study, let's take stock of what we've learned from exploring different deep learning models. In this final chapter, we summarize we figure out what it all means.

6.1 CONCLUSION

In this report we cover various nuances of transfer learning and deep learning to be used as a backbone on order to generate a Decision Support System or DSS which can assist radiologists and Oncologists to detect and localize lesions in various CESM images which are high energy images and are better when converted to FFDM as the FFDMs can convey only structural information that is evident by the differences in x-ray attenuation [37].

CESM images were here chosen as a standard choice because the tumours in breast are highly vascular and develop their own blood supply hence when we apply neuro-reflective iodine injection it can be clearly seen that the tumours are of various shapes and types. Subsequently the low energy images (without iodine) and high energy images (with iodine injected) are subtracted in order to expose the tumorous regions. But while exposing the tumorous regions the calcifications can also be exposed as the fibro glandular tissue has attenuation similar to that of breast lesions and thus similar image gray levels. The similar attenuation between fibro glandular tissue and a lesion reduces the visibility of (i.e., masks) a lesion because the lesion gets lost in the background which can raise the False Positive results hence increasing the overall cost and labour involved in biopsies. The task is significantly difficult in case of highly dense breast tissues can be vastly helpful. Figure 6.1 shows how a DSS system can be used in stand-alone fashion to classify the images and the validation is performed none-other than the model itself as another model classifies the same image, if the two results are similar the answer is

accepted, although the method is highly dubious but still provides some level of resilience. In order, to overcome this latency and redundancy of results the model is converted to a DSS as shown in figure 6.2 and a check of radiologist is introduced who verifies the output. The radiologist can also classify the images on the BI-RADS scale prior to feeding to the system. This later setup is called as Fully Automated Pipeline System (FAPS) as described in [38].

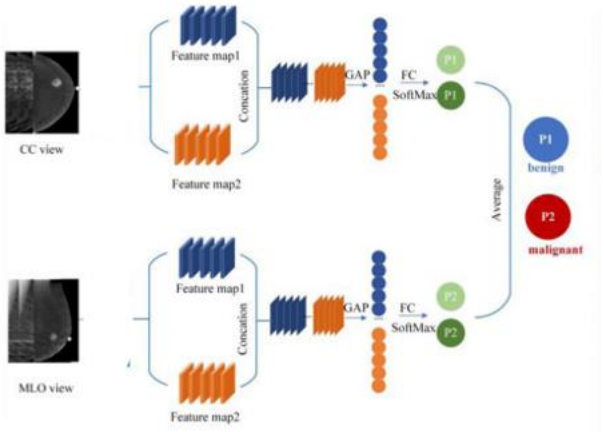


Figure 6. 1: Classification in case of autonomous model

In figure 6.2 the blue symbols represent patients whose BI-RADS categories were upgraded, and the red ones were downgraded.

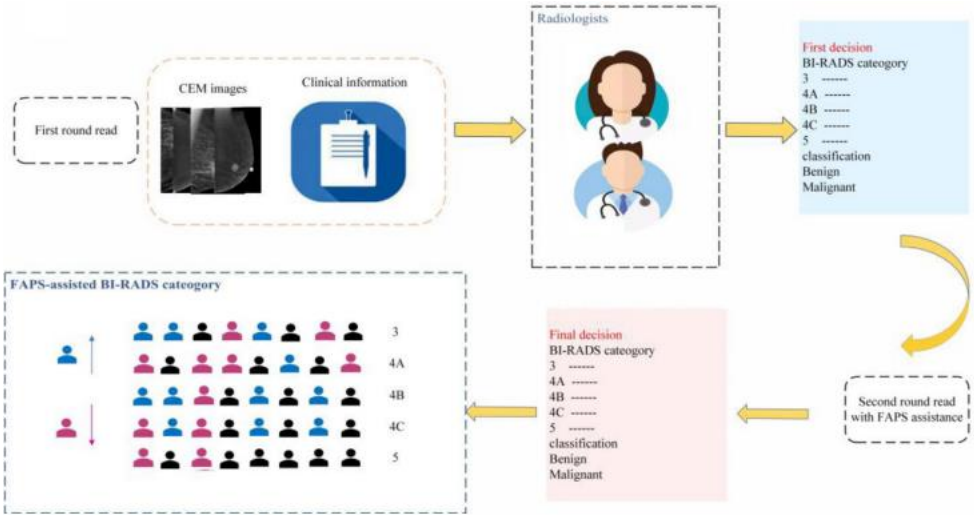


Figure 6. 2: Classification of CESM images in BIRADs categories

The model can also be used to verify as the double reading standards required by the doctors can be highly laborious and subsequently can increase the cost and labour of required by significant amount. The arrangement for the double reading can be seen in the set up depicted in figure 6.3.

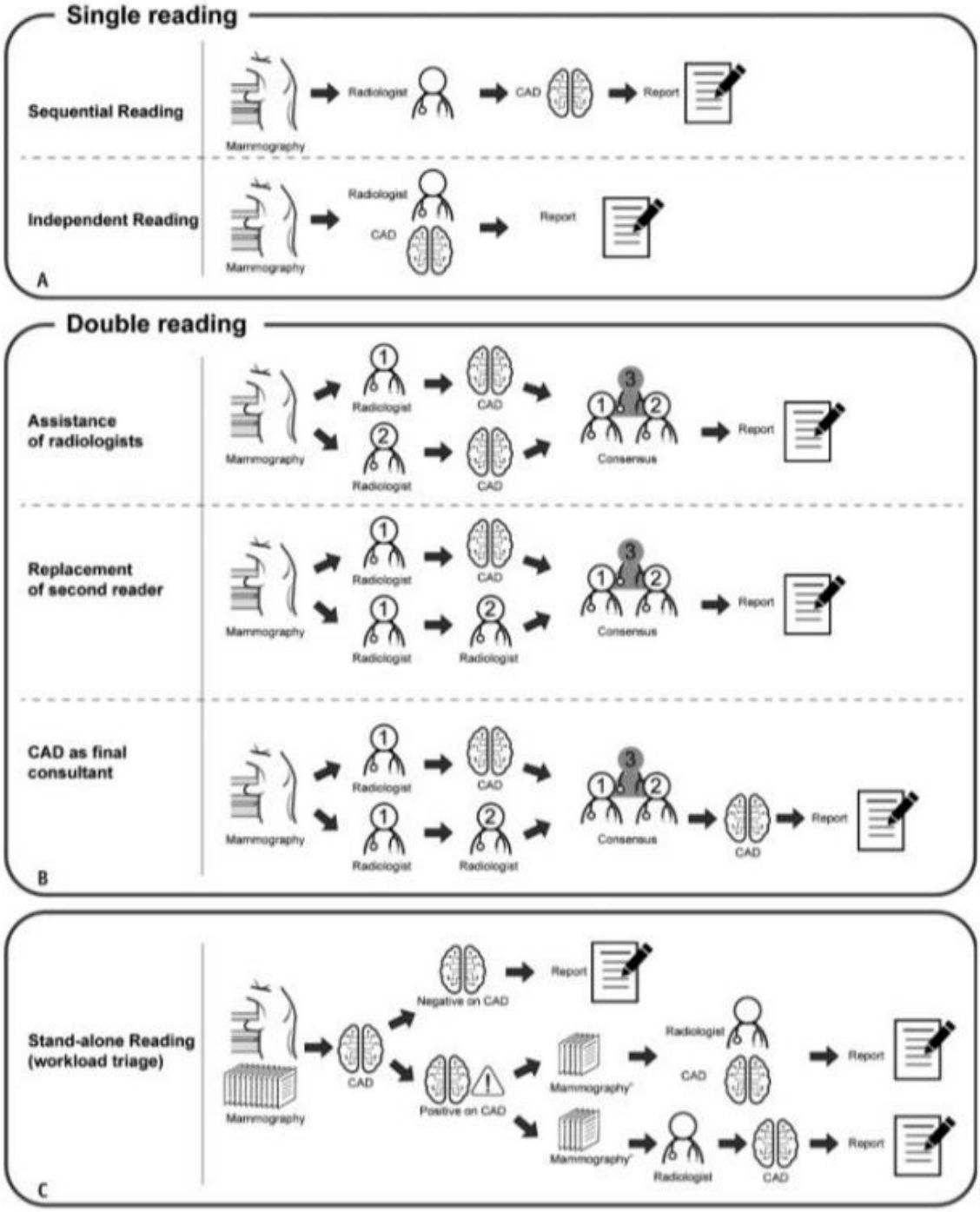


Figure 6. 3: Various setups for using CAD as DSS

6.2 FUTURE SCOPE

The CESM technology is newer for diagnosticians and isn't still widely available to public. The datasets in this respect are very limited and various uses are still unexplored for the technology. The introduction of more public datasets will definitely increase the quality of research.

Introduction of newer deep models in conditions persistent to medical images will be highly useful like that of ChexNet designed specifically for pneumonia detection by Stanford [13]. Datasets with manual annotations in format of YOLO or You Only Look Once will be a boon as the same will increase the computation cost drastically for the classification task. Other latest technologies like Visual Transformers or ViT are also still unexplored.

Due to computational limitations segmentation models like U-Net could not be implemented, custom models could also not be implemented to the same. Models like EfficientNet, VGG etc still need to be optimized heavily as these models overfitted the results very easily. Sophisticated models like XceptionNet are also novel for this kind of research and hence can be explored in depth.

Summary

The chapter draws conclusions and defines outlines for future scope of work.

REFERENCES

[1] M.-L. Huang and T.-Y. Lin, "Dataset of breast mammography images with Masses," Data in Brief, vol. 31, p. 105928, 2020.

[2] S. Taylor-Phillips and C. Stinton, "Double reading in breast cancer screening:Considerations for policy-making," The British Journal of Radiology, vol. 93, no. 1106, p. 20190610, 2020.

[3] Lewin JM, Isaacs PK, Vance V, Larke FJ. Dual-Energy Contrast-enhanced Digital Subtraction Mammography: Feasibility. Radiology 2003;229:261-8. 10.1148/radiol.2291021276

[4] FDA 510(k) clearance for GE Contrast Enhanced Spectral Mammography (CESM). Available online: https://www.accessdata.fda.gov/cdrh_docs/pdf10/K103485.pdf

[5] J. H. Yoon and E.-K. Kim, "Deep learning-based Artificial Intelligence for Mammography," Korean Journal of Radiology, vol. 22, no. 8, p. 1225, 2021.

[6] E.-K. Kim, H.-E. Kim, K. Han, B. J. Kang, Y.-M. Sohn, O. H. Woo, and C. W. Lee, "Applying data-driven imaging biomarker in mammography for breast cancer

screening: Preliminary study," Scientific Reports, vol. 8, no. 1, 2018.

[7] R. Rabiei, "Prediction of breast cancer using machine learning approaches," Journal of Biomedical Physics and Engineering, vol. 12, no. 3, 2022.

[8]M. A. Naji, S. E. Filali, K. Aarika, E. L. H. Benlahmar, R. A. Abdelouhahid, and O.Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," Procedia Computer Science, vol. 191, pp. 487–492, 2021.

[9]Y. Su, Q. Liu, W. Xie, and P. Hu, "Yolo-logo: A Transformer-based Yolo segmentation model for breast mass detection and segmentation in digital mammograms," Computer Methods and Programs in Biomedicine, vol. 221, p. 106903, 2022.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv.org, 10-Apr-2015. [Online]. Available:https://arxiv.org/abs/1409.1556.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv.org, 10-Apr-2015. [Online]. Available:https://arxiv.org/abs/1409.1556v6.

[13] Rajpurkar, P., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning", arXiv e-prints, 2017. doi:10.48550/arXiv.1711.05225.

[14] Lobbes MB, Smidt ML, Houwers J, Tjan-Heijnen VC, Wildberger JE. Contrast enhanced mammography: techniques, current results, and potential indications. Clin Radiol 2013;68(9):935–944. Crossref, Medline, Google Scholar

[15] Castillo-López JP, Cruz-Rodríguez JC, Galván-Espinoza HA, Berumen F, Villaseñor-Navarro Y, Brandan ME. Optimization of acquisition parameters for the detection of secondary breast lesions applying temporal contrast enhanced digital mammography. In: Krupinski EA, ed. 14th International Workshop on Breast Imaging (IWBI 2018). Vol 10718. Bellingham, Wash: International Society for Optics and Photonics, 2018; 107181Y. Crossref, Google Scholar

[16] Lewin JM, Isaacs PK, Vance V, Larke FJ. Dual-energy contrast-enhanced digital subtraction mammography: feasibility. Radiology 2003;229(1):261–268. Link, Google Scholar

[17] Kelcz F, Mistretta CA, Riederer SJ. Spectral considerations for absorption-edge fluoroscopy. Med Phys 1977;4(1):26–35. Crossref, Medline, Google Scholar

[18] Kruger RA, Mistretta CA, Crummy AB, et al. Digital k-edge subtraction radiography. Radiology 1977;125(1):243–245. Link, Google Scholar

[19] [1] W. F. Sensakovic et al., "Contrast-Enhanced Mammography: How Does It Work?," RadioGraphics, vol. 41, no. 3, pp. 829–839, 2021. doi:10.1148/rg.2021200167

[20] Khaled, R., Helal, M., Alfarghaly, O. et al. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. Sci Data 9, 122 (2022). https://doi.org/10.1038/s41597-022-01238-0

[21] Mohamed, S.A.S., Moftah, S.G., Chalabi, N.A.E.M. et al. Added value of contrast-enhanced spectral mammography in symptomatic patients with dense breasts. Egypt J Radiol Nucl Med 52, 8 (2021). https://doi.org/10.1186/s43055-020-00372-2

[22] R. R. Selvaraju et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization," International Journal of Computer Vision, vol. 128, no. 2, pp. 336–359, 2019. doi:10.1007/s11263-019-01228-7

[23] Hegazy, R., Adel, L. & Yasin, R. The value of CESM in the evaluation of intraductal breast papilloma: a comparative study with DCE-MRI. Egypt J Radiol Nucl Med 51, 27 (2020). https://doi.org/10.1186/s43055-019-0122-8

[24] Sorin V, Sklair-Levy M. Dual-energy contrast-enhanced spectral mammography (CESM) for breast cancer screening. Quant Imaging Med Surg 2019;9(11):1914-1917. doi: 10.21037/qims.2019.10.13

[25] Hassan, S.A., Sayed, M.S., Abdalla, M.I. et al. Breast cancer masses classification using deep convolutional neural networks and transfer learning. Multimed Tools Appl 79, 30735–30768 (2020).https://doi.org/10.1007/s11042-020-09518-w

[26] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for Improved Breast Cancer   Risk Prediction," Radiology, vol. 292, no. 1, pp. 60–66, 2019. doi:10.1148/radiol.2019182716

[27] Shen, L., Margolies, L.R., Rothstein, J.H. et al. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Sci Rep 9, 12495 (2019). https://doi.org/10.1038/s41598-019-48995-4

[28] Chi, Xiaoxiao MSca; Zhang, Lei BSb; Xing, Dong BSa; Gong, Peiyou PhDa; Chen, Qianqian PhDc; Lv, Yongbin MSca,∗. Diagnostic value of the enhancement intensity and enhancement pattern of CESM to benign and malignant breast lesions. Medicine 99(37):p e22097, September 11, 2020. | DOI: 10.1097/MD.0000000000022097

[29] [1] N. A. ElSaid, H. G. M. Mahmoud, A. Salama, M. Nabil, and E. D. ElDesouky, "Role of contrast enhanced spectral mammography in predicting pathological response of locally

Advanced Breast Cancer Post neo-adjuvant chemotherapy," The Egyptian Journal of Radiology and Nuclear Medicine, vol. 48, no. 2, pp. 519–527, 2017. doi:10.1016/j.ejrnm.2017.03.022 ,

[30] Travieso Aja MM, Rodríguez Rodríguez M, Alayón Hernández S, Vega Benítez V, Luzardo OP. Dual-energy Contrast-enhanced mammography. Radiología (English Edition) 2014;56(5):390–9.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826.

[32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700-4708.

[33] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," arXiv preprint arXiv:1707.07012, 2018.

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520.

[35] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. International Conference on

Learning Representations Workshop, 2014.

[36] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). "Learning Deep Features for Discriminative Localization." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921-2929.

[37] William F. Sensakovic, Molly B. Carnahan, Christopher D. Czaplicki, Samuel Fahrenholtz, Anshuman Panda, Yuxiang Zhou, William Pavlicek, and Bhavika Patel RadioGraphics 2021 41:3, 829-839

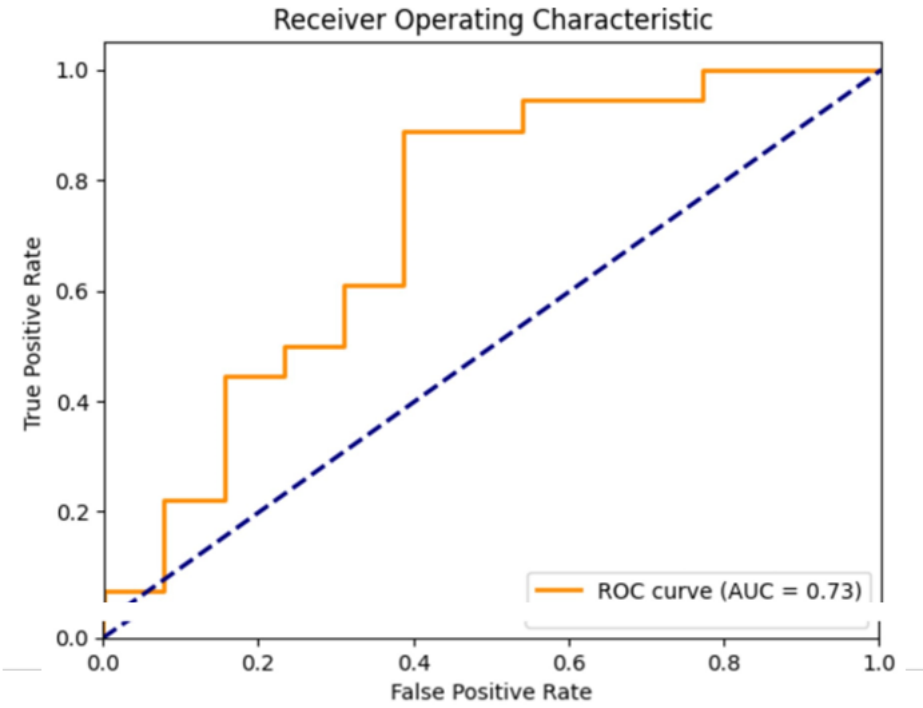
[38] T. Zheng et al., "Deep learning-enabled fully automated pipeline system for segmentation and classification of single-mass breast lesions using contrast-enhanced mammography: A

prospective, multicentre study," eClinicalMedicine, vol. 58, p. 101913, 2023. doi:10.1016/j.eclinm.2023.101913

APPENDIX I

ROC curves for various models.

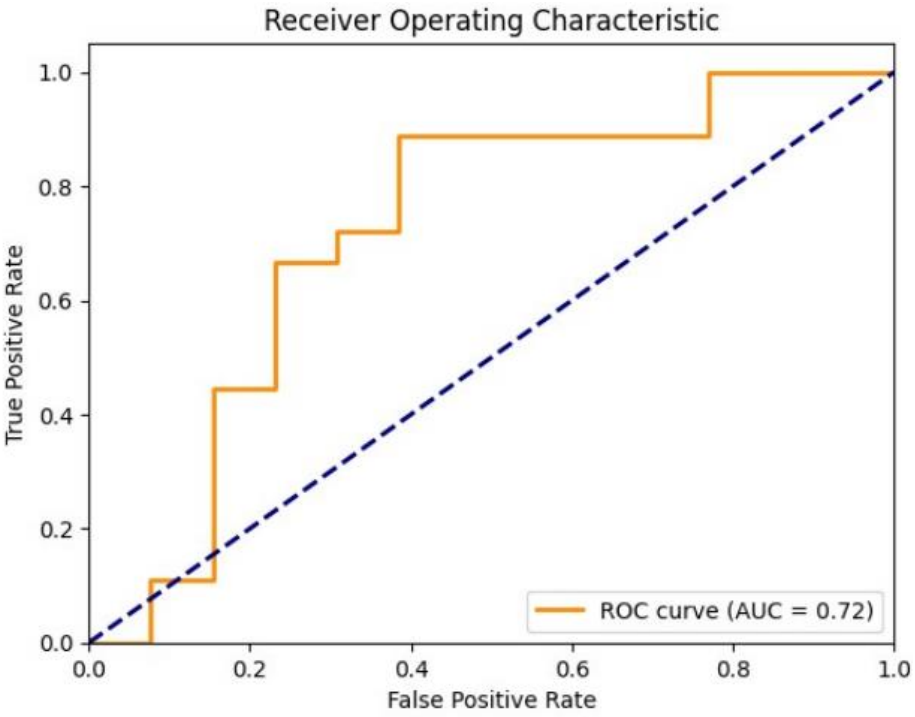
- **ResNet50V2**



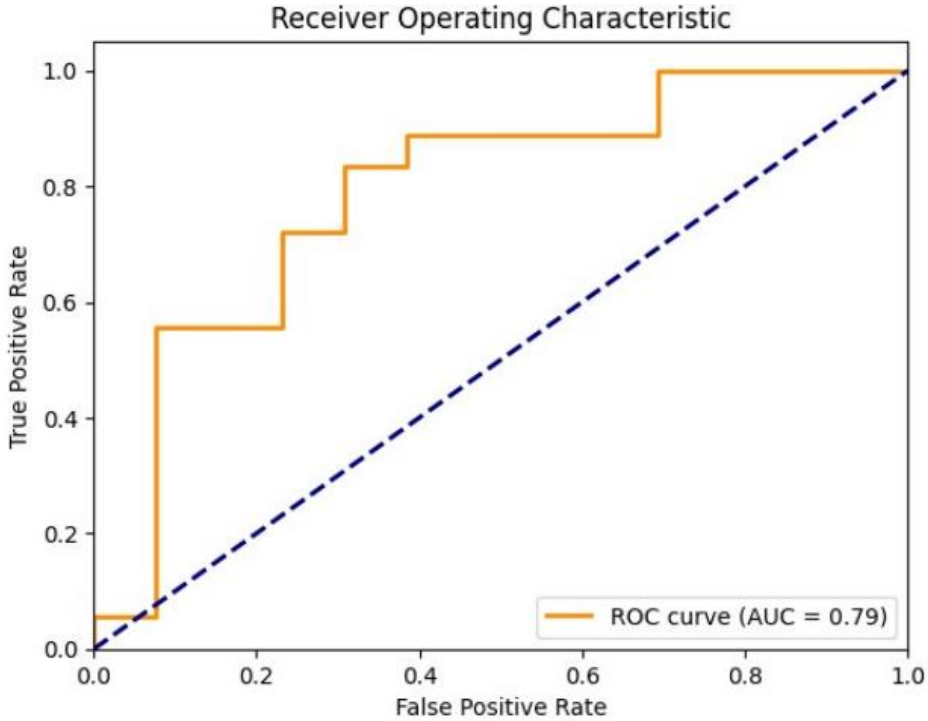

```
Confusion Matrix:
[[ 8  5]
 [ 2 16]]
```

- **ResNet101V2**

```
Confusion Matrix:
[[ 8  5]
 [ 2 16]]
```

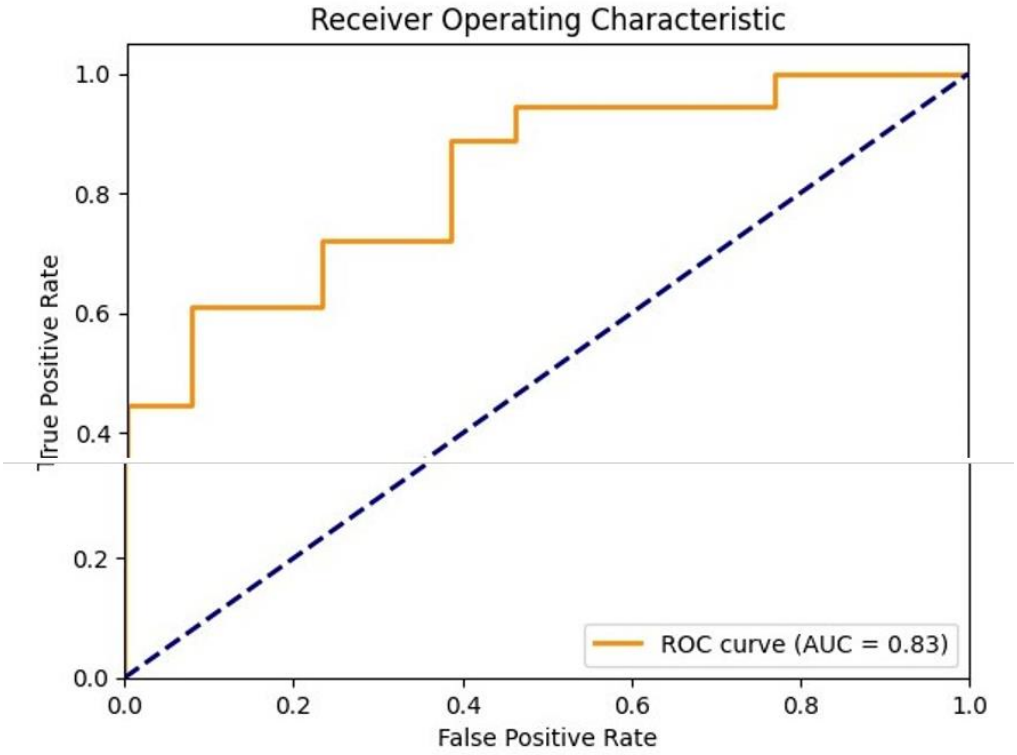


- **ResNet152V2**



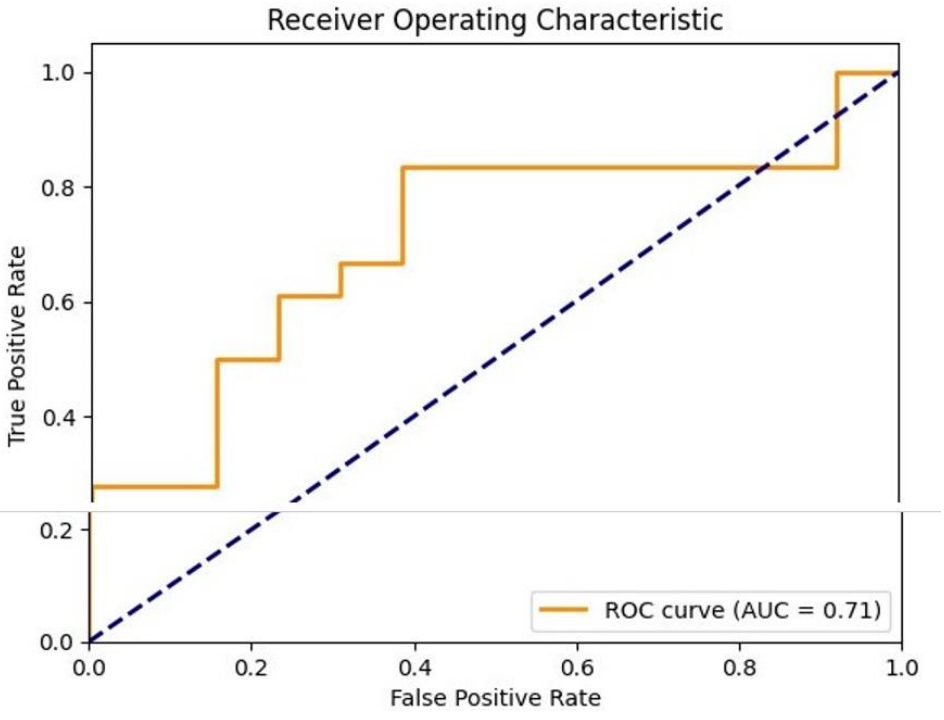

```
Confusion Matrix:
[[ 7  6]
 [ 2 16]]
```

- **InceptionV3**



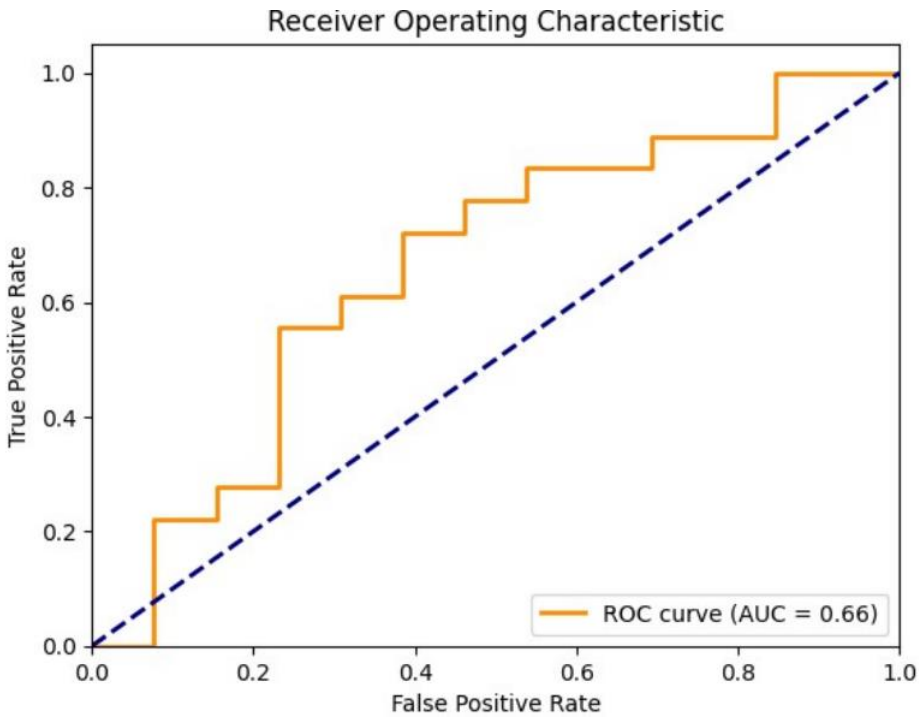

```
Confusion Matrix
[[ 4  9]
 [ 1 17]]
```

- **DenseNet169**



```
Confusion Matrix:
[[ 8  5]
 [ 5 13]]
```

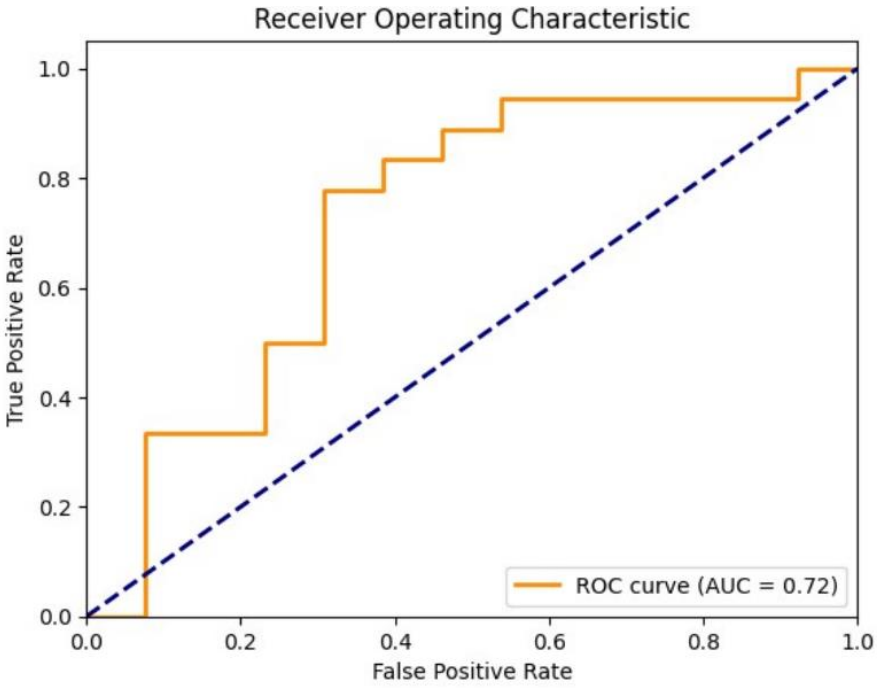
- **DenseNet201**



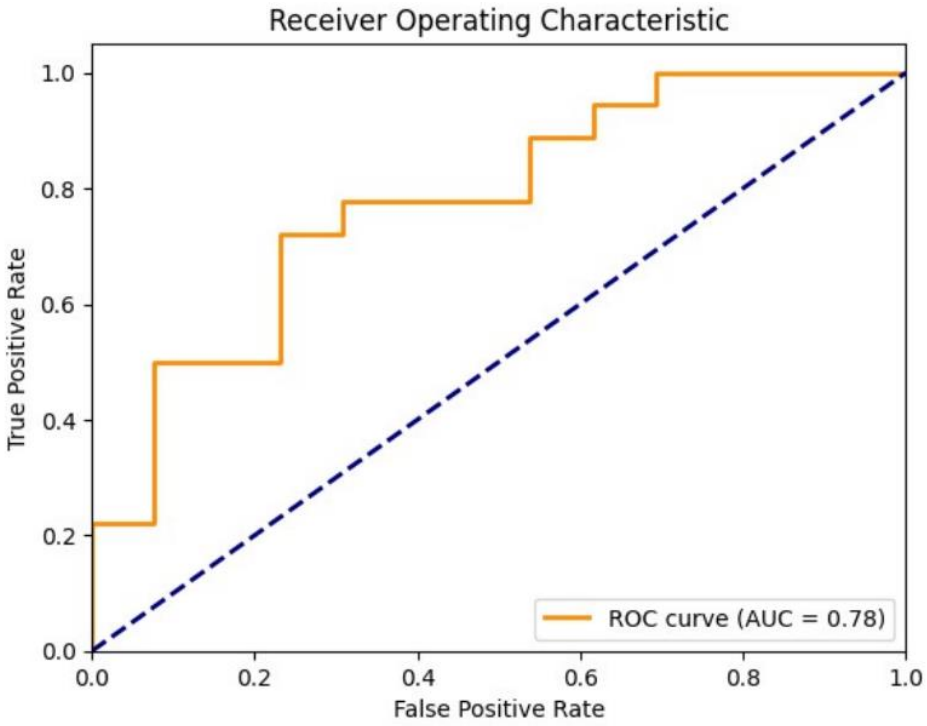

```
Confusion Matrix:
[[ 8  5]
 [ 5 13]]
```

- **MobileNetV2**

```
Confusion Matrix:
[[ 6  7]
 [ 2 16]]
```

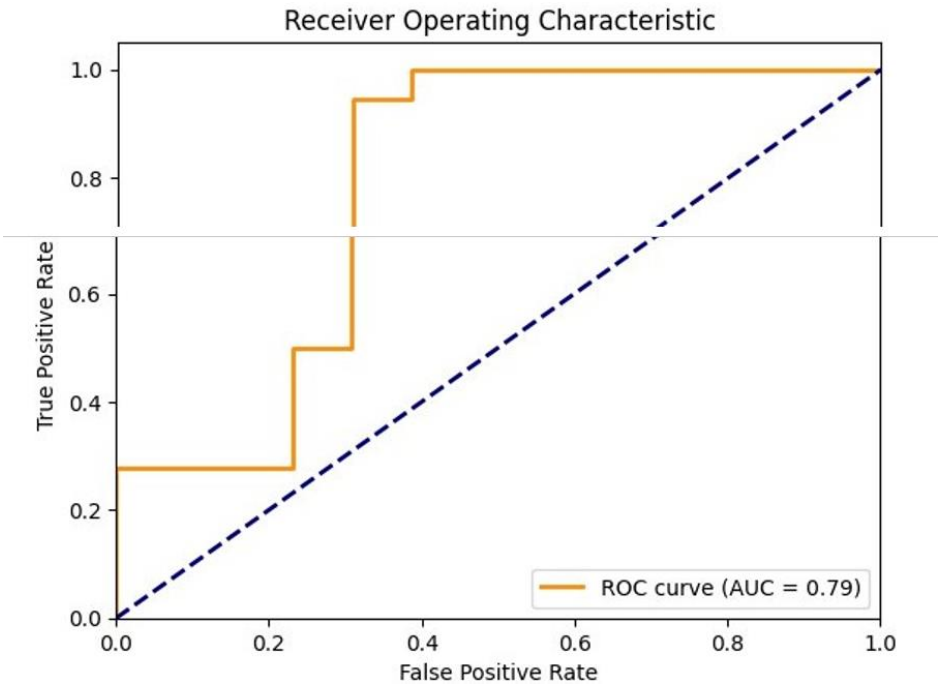


- **NASNetMobile**



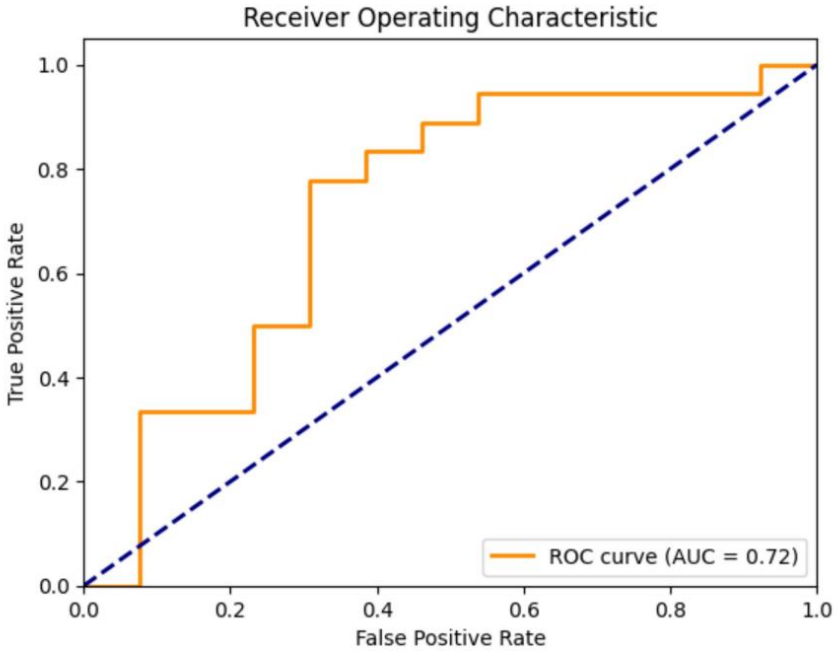

```
Confusion Matrix:
[[ 9  4]
 [ 4 14]]
```

- **NASNetLarge**




```
Confusion Matrix:
[[ 9  4]
 [ 4 14]]
```

- **MobileNet**




```
Confusion Matrix:
[[ 6  7]
 [ 2 16]]
```