Chapter 1: Introduction

1.1: Breast Cancer: A Global Challenge

Breast cancer is the most prevalent cancer affecting women worldwide. According to the World Health Organization (WHO), over 2.35 million new cases are diagnosed annually. Early detection significantly improves treatment efficacy, highlighting the crucial role of accurate and timely diagnosis. However, individual diagnostic methods often have limitations in capturing the entire spectrum of complexities associated with early-stage cancer detection. This necessitates exploring the potential of utilizing multiple medical imaging modalities for a more comprehensive diagnostic approach.

1.2: Limitations of Single-Modality Diagnosis

Conventional diagnostic methods for breast cancer, such as mammography and ultrasound, each possess limitations. Mammography, while being the most used screening method for breast cancer detection, can struggle with dense breast tissue, leading to missed diagnoses. Ultrasound, on the other hand, excels at differentiating between cystic and solid masses but might not always provide enough detail for definitive cancer classification.

1.3: Potential of Multimodal Medical Imaging

Through combining information from different modalities, such as mammography, ultrasound, and thermal imaging, a more comprehensive picture of breast tissue health can be obtained. This approach has the potential to overcome the limitations of individual modalities, leading to more accurate diagnoses, particularly in complex cases.

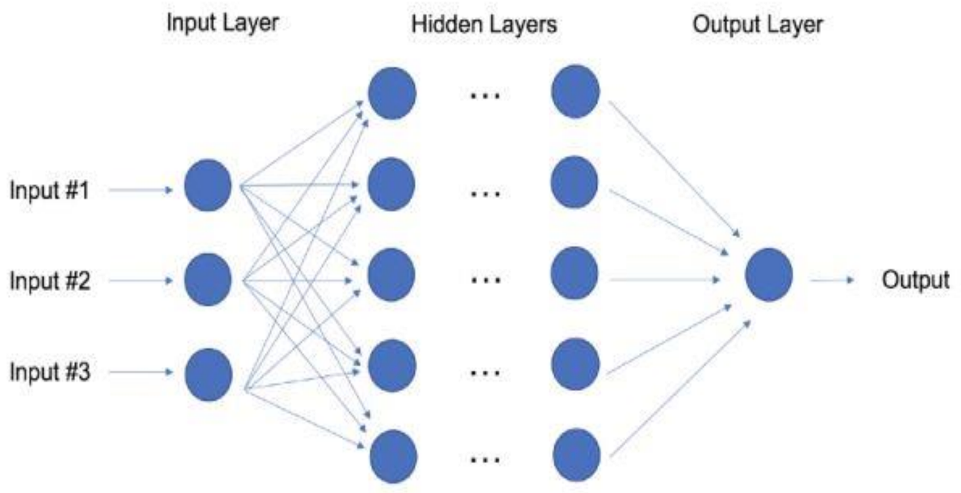
1.4: Background Information

1.4.1 Deep Learning

Deep learning is a subfield of machine learning that utilizes artificial neural networks with a significantly higher number of hidden layers compared to traditional shallow neural networks. Deep learning algorithms can automatically learn and extract features from data through a process called training. These features are then used to make predictions or classifications.

Deep learning models consist of numerous interconnected nodes arranged in multiple layers within a deep artificial neural network architecture (Figure 1.4.1). As new data enters the network, these nodes can automatically rearrange and adjust their connections, effectively learning and improving their performance over training

.



1.4.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a specific type of deep learning architecture particularly well-suited for image analysis tasks such as image classification, object detection, and image recognition.

A typical CNN architecture comprises several components:

<u>Convolutional Layer</u>**:** This layer applies filters to the input image, extracting feature maps that capture specific characteristics of the image. Different filter sizes can be used to detect features of varying scales.

<u>Pooling Layer</u>**:** This layer reduces the spatial dimensions of the data (width and height), thereby decreasing computational cost and mitigating overfitting. Common pooling techniques include max pooling, which selects the maximum value within a specific region, and average pooling, which computes the average value within a region.

<u>Fully Connected Layer</u>**:** This layer acts as the final classifier. It takes the flattened output from the previous layers and performs computations to determine the class probabilities for the input image.

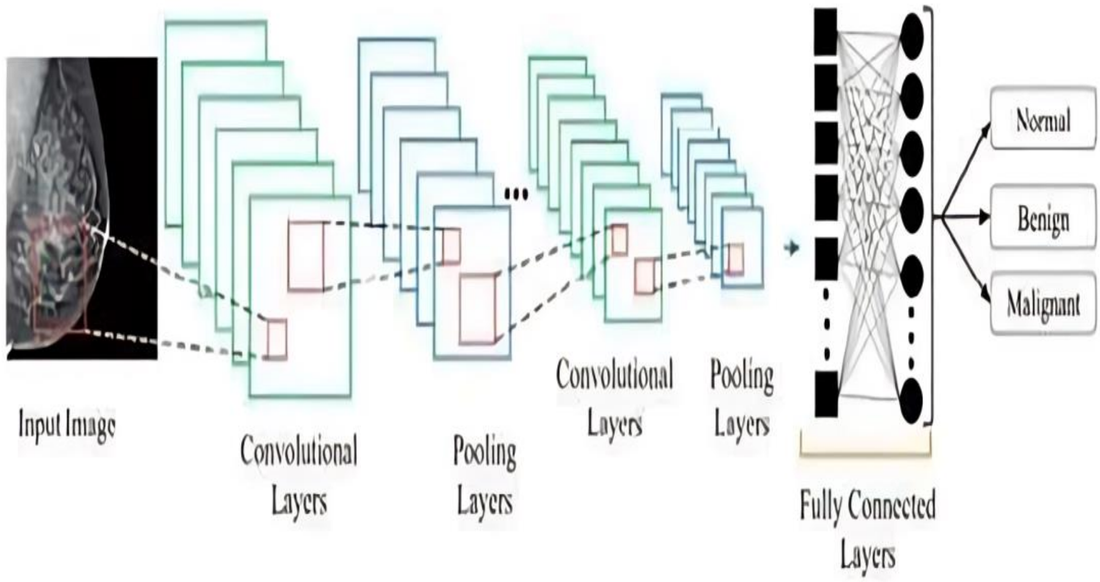


Figure 1.4.2: CNN Model Training on Mammography Image

1.5: AIM & OBJECTIVE

The aim of this project is to explore the potential of medical based pre-trained CNN models for multimodal breast cancer classification.

The specific objectives are as follows:

Creating Unimodal Models with Medical Pre-trained CNN Models**:**
Apply pre-trained CNN models of RadImageNet, ChexNet that had been trained on medical images on individual medical image modalities (mammograms, ultrasounds, and thermal images) with its specific architectural requirements.

Training & Optimizing the Models**:**
Fine-tune the parameters of these pre-trained models and employ data augmentation technique to enhance their classification accuracy for breast cancer detection.

Develop Multimodal Models:
Create multimodal CNN classification models with the best performing unimodal models for each image modality by extracting and combining complex features from multimodal images at feature level fusion.

Performance Evaluation:

Evaluate the performance of the developed Uni & multimodal CNN models with evaluation metrics.

Chapter 2: Literature Review

This chapter focuses on the literature survey relevant to the main topic of this dissertation, i.e. role of medical based pre-trained CNN models in classification of medical images.

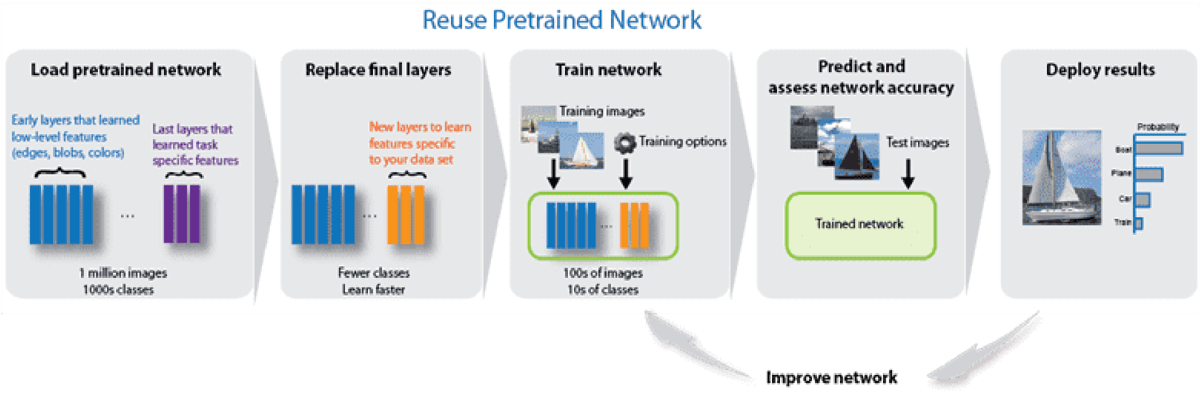
2.1: Overcoming The Traditional CNN Challenges For Medical Diagnosis

In the early 2000s, CNNs were computationally expensive and very prone to over fitting due to lack of available medical labeled data, class imbalance, and inherent interpretability issues with medical datasets. Additionally, the black-box nature of deep learning models was hindering their adoption in clinical settings where interpretability and trustworthiness are crucial.

However, in the last decade, due to the exponential increament in computational power by availability of GPUs and development of techniques like transfer learning that leverages pre-trained models on large-scale datasets such as ImageNet and fine-tunes them on medical imaging data, has emerged as a powerful strategy to mitigate the challenges of limited annotated medical datasets. Additionally, ensemble learning techniques combining multiple deep learning models have been explored to enhance classification accuracy and robustness, thereby improving diagnostic outcomes in clinical settings.

2.2: Transfer Learning in Deep Learning

Transfer learning allows a pre-trained model on a large dataset to be adapted for a new task (target domain) with a smaller dataset. This approach leverages the knowledge learned from the source domain to improve performance on the target task. In CNNs, the trainable parameters, which include filters in convolutional layers, weights in dense layers, and other components adjusted during backpropagation, store this knowledge. By transferring these parameters and fine-tuning the final layers, the model can adapt to the new classification problem in the target domain. Popular deep learning frameworks like Keras provide access to pre-trained models on datasets like ImageNet. These models excel at recognizing fundamental visual features like edges, shapes, and textures – building blocks for extracting visual information from images.



<u>Benefits of Transfer Learning:</u>

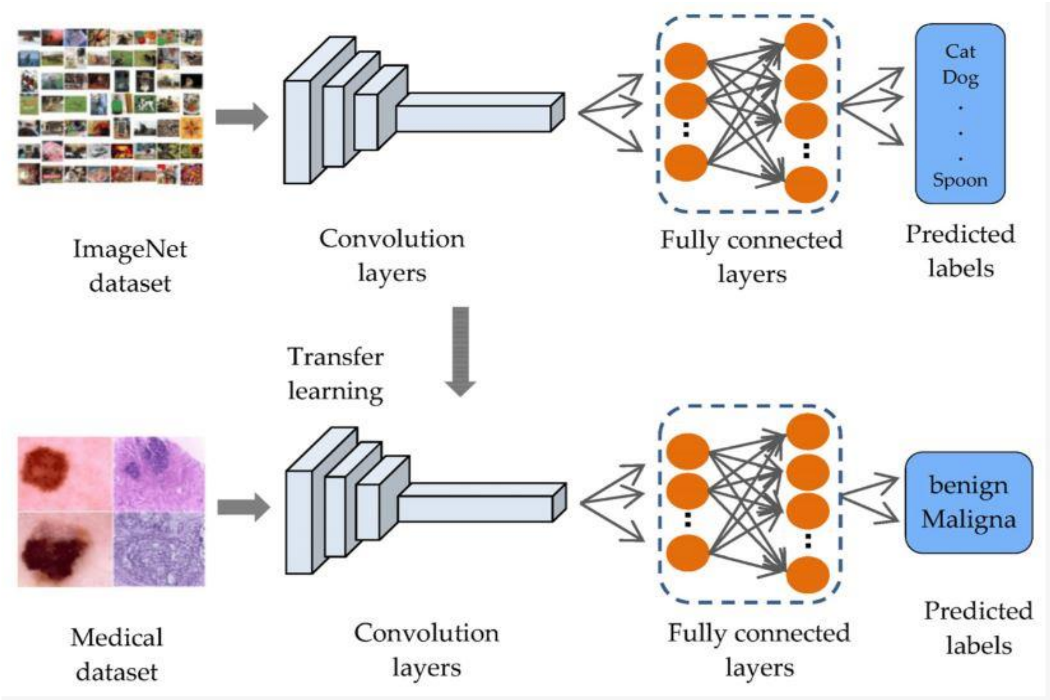
- *Reduced Training Time*: Pre-trained models contain features learned from vast amounts of data, allowing faster convergence on the target task with a smaller medical dataset.
- *Improved Performance:* Transfer learning enables leveraging pre-trained features for medical image classification, even with limited data, leading to better accuracy compared to training from scratch.
- *Reduced Computational Resources:* By leveraging pre-trained models, the need for extensive training on powerful computing systems is reduced.

2.2.1 Transfer Learning with ImageNet

ImageNet, a large dataset of everyday objects, has been a popular source for pre-training CNNs for medical image classification. These pre-trained models learn low-level features like edges and shapes, which are generally transferable to medical images. However, there are concerns about the suitability of ImageNet due to differences between natural and medical images:

<u>Domain Shift</u>*:* Medical images often lack clear global subjects present in natural images and rely more on subtle texture variations for pathology detection. This domain shift can lead to performance degradation when directly transferring from ImageNet.

<u>Data Characteristics</u>**:** Medical datasets tend to be smaller in size, have fewer classes, and have higher resolutions compared to ImageNet. These differences can limit the effectiveness of ImageNet pre-training.



2.2.2: Alternative Of ImageNet: RadImageNet & CheXNet

RadImageNet has emerged as a potential alternative to ImageNet for transfer learning in medical imaging. This comprehensive medical image repository, curated with 1.35 million labeled images across various modalities (CT, MRI, US) and pathologies, surpasses ImageNet in its domain relevance. Mei et al. (2022) demonstrated that RadImageNet based pretrained CNN models consistently outperform ImageNet-based models on eight distinct medical image classification tasks, including thyroid nodule classification and anterior cruciate ligament (ACL) injury detection in MRI. These performance gains were particularly significant for smaller datasets, demonstrating the usefulness of RadImageNet in situations where there is a deficiency of training data.

CheXNet, introduced by Rajpurkar et al. introduced. (2017) was dedicated to tasks related to chest radiography. It takes a similar approach to RadImageNet, but with a focused dataset. Pre-trained on a collection of 112,120 chest X-rays annotated for 14 different pathologies, CheXNet's training data directly reflects the type of images it will be used to analyze later. This targeted approach allows it to learn features specific to chest X-rays, such as lung textures and anatomical landmarks.

Studies by Jang et al. (2018) validate CheXNet's proficiency in identifying pneumonia from chest X-rays, achieving an AUC (Area Under the Curve) of 0.83, surpassing human radiologists in specific cases. The success of CheXNet can be attributed to the close match between the chest X-ray images it was trained on and the chest X-ray images it analyzes during inference.

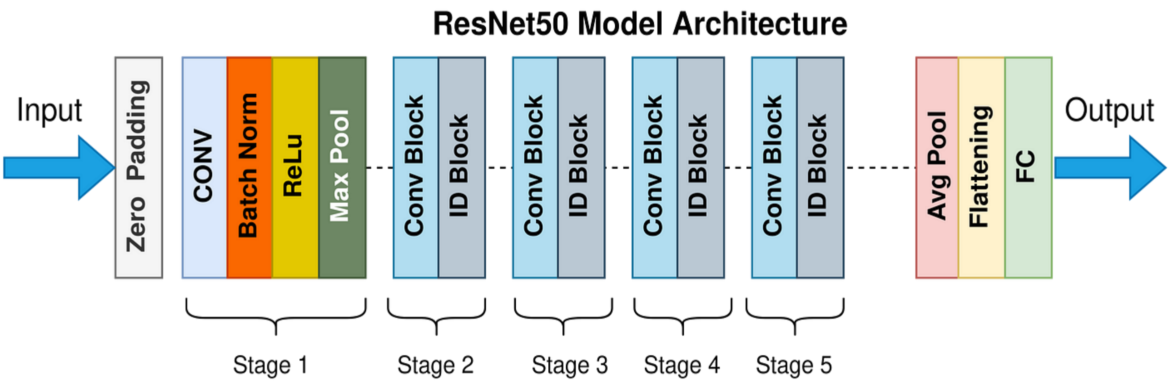
2.3: Pre Trained CNN Models

This part review describes four pre-trained CNN models provided by RadImagenet-based studies: ResNet-50, DenseNet-121, InceptionV3, and IRv2.

2.3.1 ResNet50 (Residual Network):

History: Introduced by Kaiming He et al. in their paper "Deep Residual Learning for Image Recognition" in 2015, ResNet was a breakthrough in deep learning by addressing the vanishing gradient problem through introducing skip connections, allowing for the training of extremely deep neural networks.

Architecture: ResNet50, a variant of the ResNet architecture, consists of 50 layers organized into a series of blocks. The architecture begins with zero padding, followed by convolution (CONV), batch normalization (Batch Norm), ReLU activation, and max pooling (Max Pool). It then incorporates several convolutional blocks (Conv Block) and identity blocks (ID Block), which include shortcut connections that facilitate gradient flow during training, enhancing learning efficiency. The final layers involve average pooling (Avg Pool), flattening, and a fully connected layer (FC). This structure, with its residual blocks, allows for improved gradient propagation and more effective training.

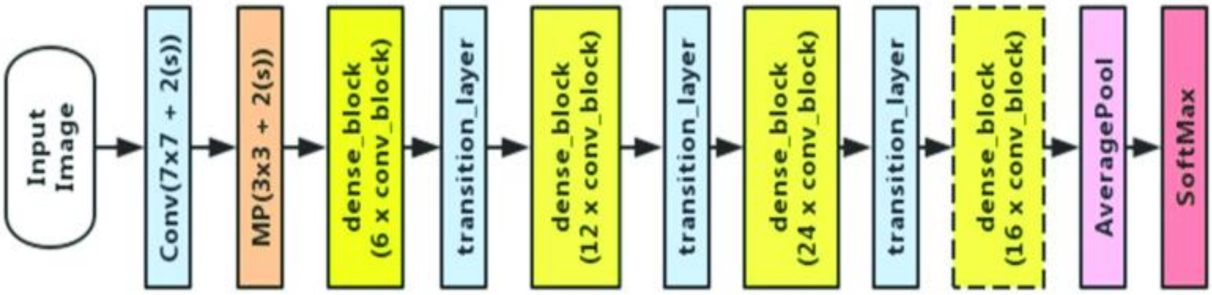


Model Complexity: ResNet50 has a high number of parameters (around 25 million) and requires high amount of computational resources for training and inference.

2.3.2. DenseNet121 (Densely Connected Convolutional Network):

History: DenseNet was introduced by Gao Huang et al. in their paper "Densely Connected Convolutional Networks" in 2017 to address the feature vanishing problem. DenseNet connects each layer to all subsequent layers, promoting feature reuse and removing the vanishing gradient issue. DenseNet121 has 121 convolutional layers and shows competitive performance with fewer parameters compared to ResNet.

Architecture**:** DenseNet utilizes dense convolutional blocks where each layer receives the feature maps from all preceding layers as input. This dense connectivity encourages feature propagation and improves gradient flow. It also includes serveral transition layers between dense blocks.



Model Complexity: DenseNet121 has a lower number of parameters (around 8 million) compared to ResNet50, making it potentially more efficient for deployment on computationally resource-constrained environments.

2.3.3. InceptionV3 (GoogLeNet Inception-v3):

History: InceptionV3 is part of the Inception family of convolutional neural network architectures developed by Google. It was introduced in the paper "Rethinking the Inception Architecture for Computer Vision" by Christian Szegedy et al. in 2015. It is a variant of the Inception architecture that utilizes inception modules for efficient feature extraction.

Architecture: Each inception module in InceptionV3 has multiple parallel convolutional pathways, which enables the network to capture features at various spatial scales. To increase efficiency, it also uses dimensionality reduction and factorized convolutions.

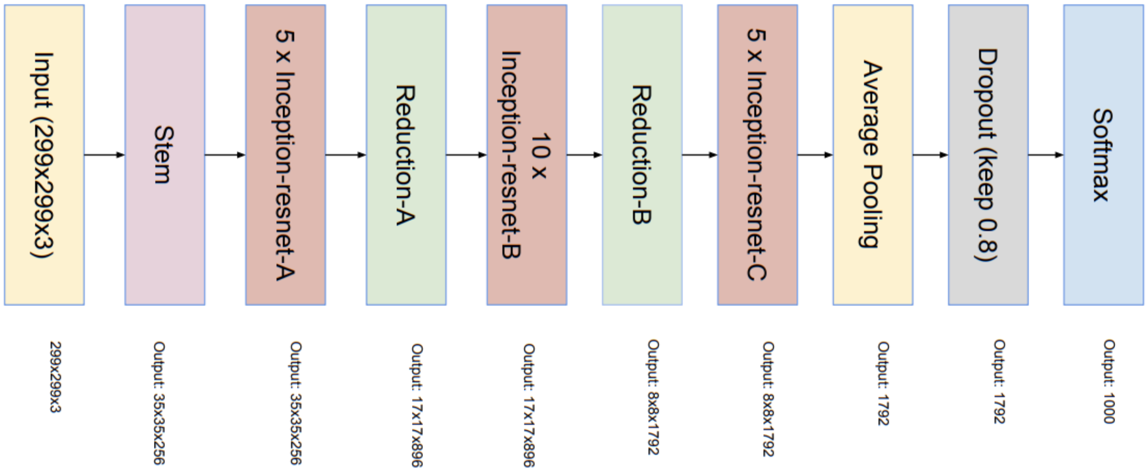
Type	Kernel size/stride	Input size
Convolution	3 × 3/2	299 × 299 × 3
Convolution	3 × 3/1	149 × 149 × 32
Convolution	3 × 3/1	147 × 147 × 32
Pooling	3 × 3/2	147 × 147 × 64
Convolution	3 × 3/1	73 × 73 × 64
Convolution	3 × 3/2	71 × 71 × 80
Convolution	3 × 3/1	35 × 35 × 192
Inception module	Three modules	35 × 35 × 288
Inception module	Five modules	17 × 17 × 768
Inception module	Two modules	8 × 8 × 1,280
Pooling	8×8	8 × 8 × 2,048
Linear	Logits	1 × 1 × 2,048
Softmax	Output	1 × 1 × 1,000

Model Complexity: InceptionV3 has a moderate number of parameters (around 24 million).

2.3.4. Inception-ResNet-v2 (IRv2):

History: IRv2, introduced by Christian Szegedy et al. in 2017, combines elements of Inception and ResNet architectures. It utilizes inception modules with residual connections, aiming to leverage the strengths of both approaches.

Architecture: IRv2 incorporates residual connections within the Inception modules, allowing for deeper networks with improved gradient flow and feature reuse. It also introduces additional optimizations such as batch normalization and factorized convolutions to enhance performance and efficiency.



Model Complexity: With high number of parameters ,IRv2 has a high model complexity due to its combination of Inception modules and residual connections and offers potentially better performance compared to InceptionV3.

2.4 Multi-Modal Models Medical Image Fusion for Breast Cancer Classification

This portion of review explores the application of multi-modal fusion with Convolutional Neural Networks (CNNs) for breast cancer classification. The advantages of multi-modal fusion, different types of fusion architectures, and their complexities will be discussed and then delve into the various fusion strategies and their implications for different medical image modalities, with a specific focus on breast cancer classification.

2.4.1 Overview

Multimodal Deep Learning is a branch of machine learning that trains models to process and identify correlations among various data types (called modalities), which are usually text, images, video, and audio. A deep learning model can understand its environment more broadly by integrating various modalities because some cues are exclusive to particular modalities. This includes fusion, a process of combining data to carry out a prediction task from two or more modalities.

In context of breast cancer Imaging, multi-modal medical imaging utilizes images from different modalities like digital mammography (DM), ultrasound (US), thermal Images, magnetic resonance imaging (MRI), has emerged as a powerful tool for breast cancer diagnosis. Each modality offers complementary information, and their combined analysis can potentially improve diagnostic accuracy compared to using a single modality alone.

2.4.1 Multi-Modal CNN Architectures

Several CNN architectures have been proposed for multi-modal fusion, each with varying complexity:

Early Fusion**:**

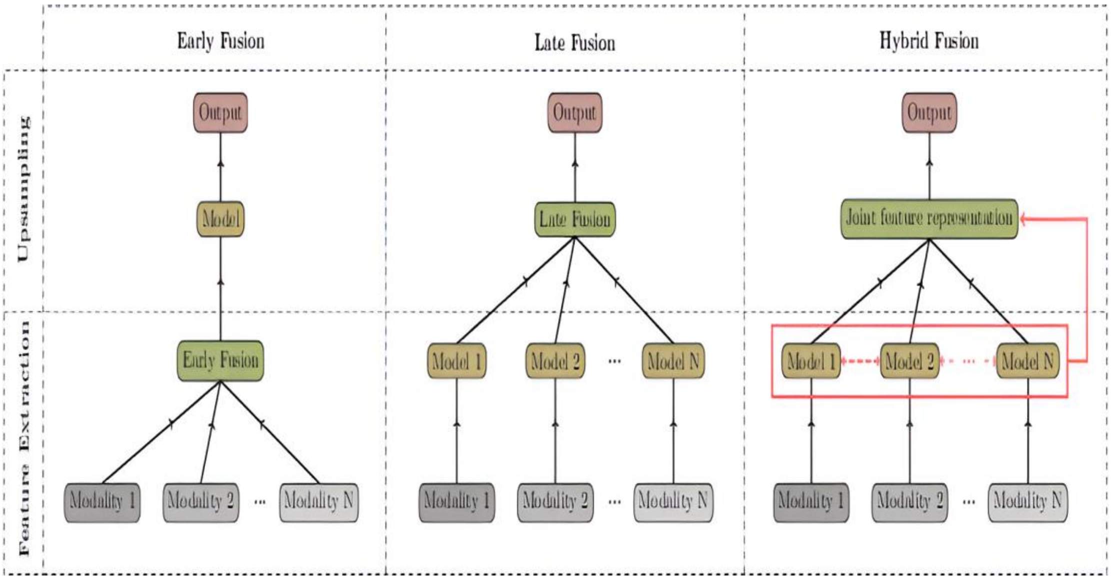
In early fusion, features are extracted from each modality using separate CNNs. These features are then concatenated and fed into a single CNN for classification. This approach is relatively simple but requires careful design to ensure compatibility between the features from different modalities.

Late Fusion**:**

Here, each modality's CNN independently generates its own classification probabilities. These probabilities are then combined using techniques like averaging, weighting, or using another neural network, to arrive at a final classification decision. This approach allows for independent model development for each modality but may not fully capture the inter-modality relationships.

Hybrid Fusion**:**

This approach combines elements of both early and late fusion. It involves strategically merging features from different modalities at various stages within the network architecture. This allows the model to capture both low-level, modality-specific features and higher-level, inter-modality relationships. Hybrid fusion architectures can be more powerful but also more challenging to train due to their increased complexity.



2.4.2 Multi-Modal Image Fusion Techniques

There are 3 main categories of multi-modal fusion techniques, each with its own implications:

<u>Pixel-Level Fusion</u>**:**

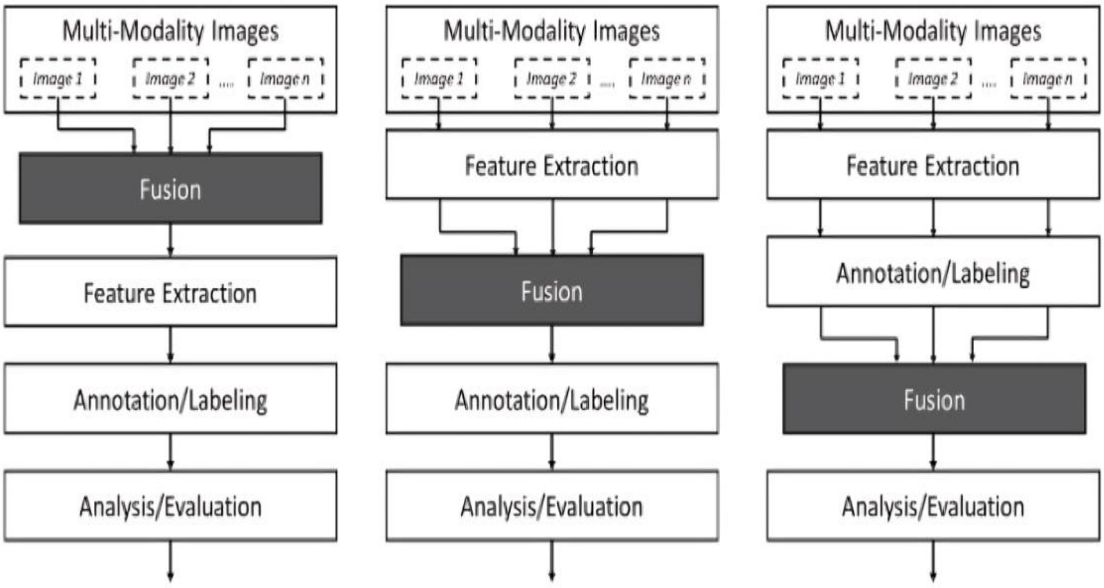
This method directly operates on the pixel intensities of the input images. Techniques like simple averaging, weighted averaging, or principal component analysis (PCA) can be used to combine the pixel values from different modalities into a single fused image. While computationally efficient, this approach may lose spatial information and fail to capture complex relationships between modalities.

<u>Feature-Level Fusion</u>**:**

Features are first extracted from each modality's image using techniques like CNNs. These features are then fused using methods like concatenation, summation, or multiplication. Feature-level fusion offers more flexibility for capturing inter-modality relationships compared to pixel-level fusion. However, designing effective feature fusion methods remains a challenge.

<u>Decision-Level Fusion</u>**:**

In this category, independent classifiers are trained on each modality's data. The resulting classification decisions (e.g., probabilities) are then fused using techniques like voting or rule-based approaches. This approach allows for independent model development for each modality but may not fully exploit the potential benefits of combining information at a lower level.



(a) Pixel-level fusion (b) Feature-level fusion (c) Decision-level fusion

2.4.3 Advantages of Multi-Modal Fusion In Context Of Breast Cancer Classification

Complementary Information**:**

Different imaging modalities capture distinct aspects of tissue properties. DM excels at detecting microcalcifications, US is valuable for assessing tissue vascularity, and MRI provides detailed soft tissue information. Combining these modalities allows for a more comprehensive analysis of the lesion.

Improved Robustness**:**

When dealing with limitations like dense breast tissue or artifacts in individual modalities, information from other modalities can compensate and improve classification accuracy.

Reduced Uncertainty**:**

By leveraging the strengths of different modalities, multi-modal fusion can reduce diagnostic uncertainty and provide radiologists with more confidence in their decisions.

2.5 Role Of Breast Diagnostic Image Modalities In breast cancer diagnosis

2.5.1 Mammography:

Technique**:** X-ray imaging to create detailed pictures of breast tissue.

Significance**:** The gold standard for breast cancer screening, particularly for women over 40. It can detect masses, calcifications and architectural distortions, often indicative of cancer.

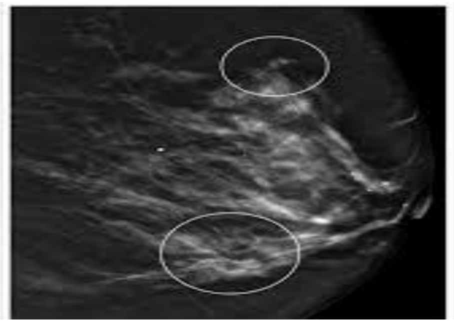


Figure 2.5.1: Breast Mammography Image

Limitations**:** May miss cancers in women with dense breast tissue. Also involves low-dose radiation exposure.

2.5.2 Ultrasound:

Technique**:** Utilizes sound waves to generate real-time images of breast tissue.

Significance**:** Excellent tool for differentiating between solid masses (potentially cancerous) and fluid-filled cysts (benign). Also used to guide biopsies for tissue sampling.

Limitations**:** Image quality can be operator-dependent and may be less effective for women with dense breasts.

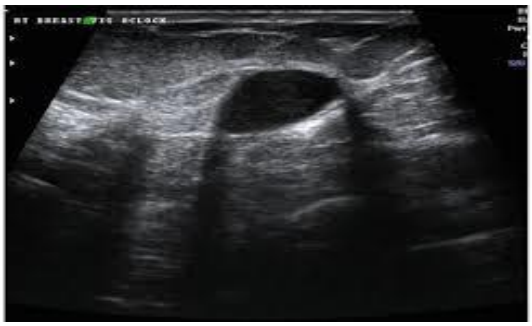


Figure 2.5.2: Breast Ultrasound Image

2.5.3 Magnetic Resonance Imaging (MRI):

Technique: Powerful magnetic fields and radio waves create detailed cross-sectional images of the breast.

Significance: Highly sensitive for detecting breast cancer, especially in high-risk women and those with dense breasts. It can also be used for staging cancer to determine its extent.

Limitations: Expensive, claustrophobic for some patients, and not recommended for routine screening due to high false-positive rates.

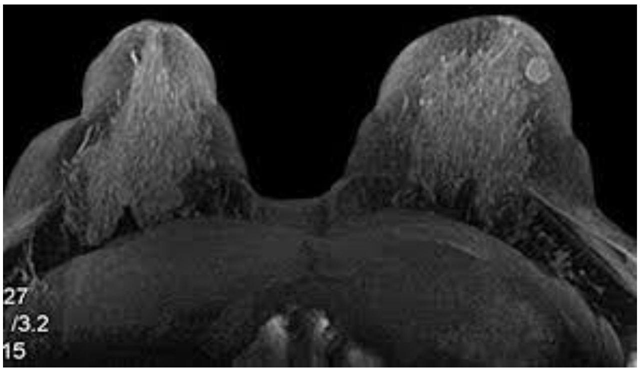


Figure 2.5.3: Breast MRI Image

2.5.4 Thermal Imaging (Thermography):

Technique: Measures infrared radiation emitted from the breast surface to create a temperature map.

Significance: A non-invasive and painless technique that may detect increased blood flow associated with tumors. Can be a helpful tool for younger women with dense breasts.

Limitations: Still considered an emerging technology. Can be influenced by various factors like ambient temperature and needs further research for definitive diagnosis.

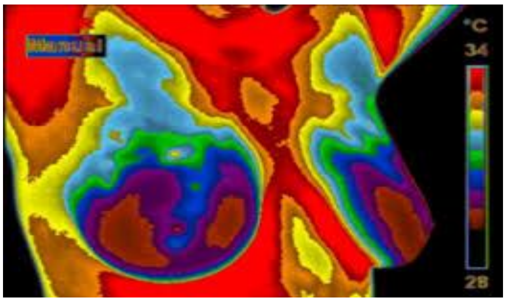


Figure 2.5.4: Breast Thermal Image

2.5.5 Histopathology:

Technique**:** Microscopic examination of biopsied tissue samples by a pathologist.

Significance**:** The definitive diagnostic tool for breast cancer. Examines cell characteristics to confirm the presence and type of cancer.

Limitations**:** Requires a tissue sample obtained through biopsy, which can be invasive.

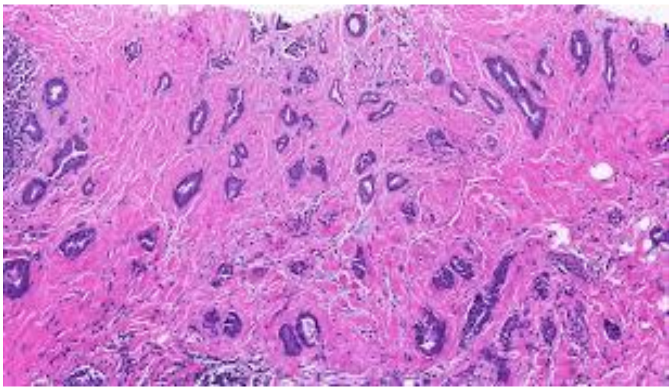


Figure 2.5.5: Breast Histopathology Image

Chapter 3: Methodology

3.1: Proposed Method

In this chapter, methods were proposed to classify breast cancer using unimodal (using a single data type) and multimodal (using multiple data types) approach for binary and multiclass classification for breast cancer diagnostic images.

1. Collecting Datasets From Publicly Available Breast Cancer Database For Mammography, Ultrasound And Thermal Images
2. Data Pre-processing & Augmentation
3. Creating Classification models With RadImagenet & CheXnet Pretrained CNN models
4. Training model (Pretrained model finetuning and hyperparameter-tuning)
5. Model Evaluation

3.2: Datasets Curation & Description

3.2.1 Breast cancer mammogram dataset: King Abdulaziz University Mammogram Dataset (KAU-BCMD)

About Dataset: A publicly available dataset containing mammogram images specifically collected in Saudi Arabia. This dataset addresses the scarcity of local public datasets for breast cancer research in that country.

Source**:** Sheikh Mohammed Hussein Al-Amoudi Center of Excellence in Breast Cancer at King Abdulaziz University, Jeddah, Saudi Arabia

Total Images**:** 5662 (4 per case: CC & MLO views, both breasts with 1416 patient cases)

Classification**:** BI-RADS (1, 2, 4, 5; excludes 3)

In this dataset, there wasn’t enough data for each BI-RADS (Breast Imaging Reporting & Data System) category which motivated to reduce the numbers of class by converting dataset into 3 classes (normal: 1265 , benign: 387, malignant: 102 Images).

Table 3.2.1: The association between the radiologist’s diagnosis (BIRAD) and the system’s designed output class.

BIRAD	Class
{1}	Normal
{2}	Benign Finding
{4,5}	Malignant Finding

3.2.2 Breast Ultrasound Images Dataset

Source: 600 female patients (25-75 years old) examined at Baheya Hospital, Egypt (2018).
Images: 780 total (PNG format, 500x500 pixels).
Classes: Normal, benign (non-cancerous), malignant (cancerous).

Case	Number of images
Benign	487
Malignant	210
Normal	133
Total	780

Table 3.2.2: Number of available images corresponding to their classes

3.2.3 Thermal Image Dataset for Breast Cancer Diagnosis: DMR-IR Dataset

The DMR-IR dataset is a collection of thermal images specifically designed for research on breast cancer diagnosis using infrared thermography.

A portion of the DMR-IR dataset, pre-segmented for targeted cancer areas (regions of interest or ROIs), is available on the Kaggle platform and has been utilized as part of multimodal dataset in this project. This Kaggle subset of DMR-IR contains roughly 1542 thermal images from 56 patients. Notably, half (780 images) were from healthy individuals, while the remaining half (762 images) showed breasts with cancer. All the thermal images share a resolution of 640 x 480 pixels.

3.3 Data Preparation for training

This sub-point contains different methods to prepare the curated datasets ready for neural network training.

3.3.1 Data Augmentation

From three datasets, mammography and ultrasound datasets consisted significant amount of class image data imbalance which need to be removed since CNN prioritizes learning the features of the majority class because of more frequent training examples, leading to the model performing well on the majority class but poorly on the minority class. Data augmentation increases the variety of minority class examples, reducing the Overfitting to Majority Class and increasing the overall generalizability.

There are many techniques for data augmentation, including random reflection, rotations, horizontal or vertical translations, zoom. In this study data augmentation had been applied to benign, malignant classes and normal class not been augmented. The augmentation parameters used are presented in below Table 3.3.1

Parameters	Value
flip	probability=0.5, max_left_rotation=10, max_right_rotation=10
flip_left_right	probability=0.4
flip_top_bottom	probability=0.4
zoom_random	probability=0.4, percentage_area=0.9

3.3.2 Data Preprocessing

1] Before to use the data, resizing of medical images with different pixel resolutions to a standard size had been done so that the CNN can process them consistently. Otherwise, it would have to continuously adjust to the different resolutions of the images, which would reduce the CNN's learning efficiency. Resizing of input images was accomplished by using load_img function from Keras' preprocessing module, which rescale and crope the images to get targeted image dimensions. The pre-trained CNNs used in this thesis all require 224x224 image size, so that was the targeted size.

2] During this process, specific preprocess_input functions were used for specific pretrained models using Keras' Application module. After that images were converted into numpy array using numpy library.

3] Data was then randomly shuffled and split into train, validate and test dataset in 60:20:20 ratio by using train_test_split function from module model_selection from sklearn library. Also one hot encoding was used for y_validation and y_train.

4] Multimodal fusion requires same numbers of images for all three modalities as input for simultaneously stable learning in each epoch. Using numpy library functions three datasets were minimally down sampled (randomly) for training datasets for 1048 samples for each dataset.

3.4 Classification Models

In this project, There were two distinct approaches used; uni and multi modal for classification task using RadImagenet and CheXnet based pretrained CNN models.

3.4.1 Unimodal Testing

For each breast cancer image modality dataset, had been used these individual five medical based pre-trained CNN models; CheXnet(Densenet121) and RadImagenet(Densenet121, InceptionV3, Resnet50, Inception-resnetV2) with all freezed layers to make five distinct models having same model architecture on top level.

<u>Model Architecture:</u> The model takes an image as input, extracts features from the image using a pre-trained CNN. After extracting high dimension features, these features are lowered in dimensions two ways through Global Average Pooling 2D and Flatten. From these two lowered outputs are concatenated into a single dimension. The softmax function outputs a probability distribution over the classes. The final

layer is a dense output layer that represents classification of multiple classes (normal, benign, malignant). For the thermal breast cancer image classification last dense output layer contained two node classes (Normal, Sick) representing binary classification.

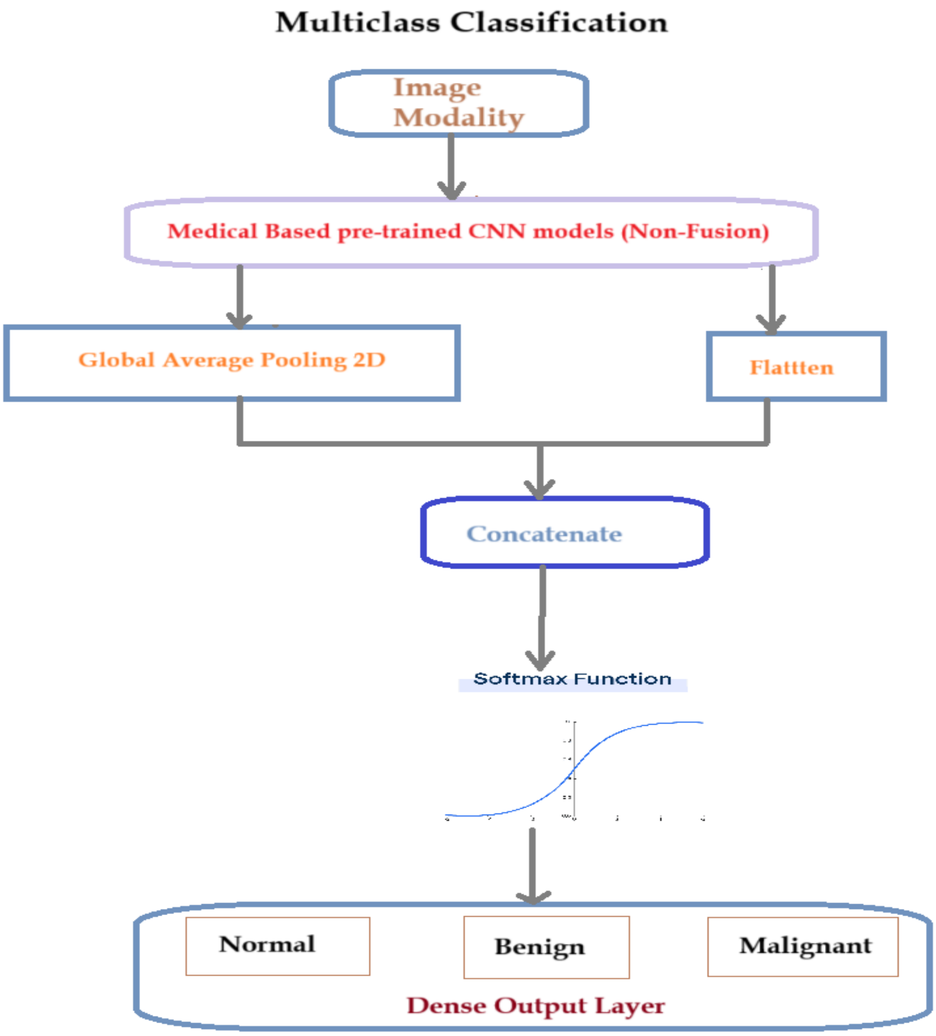


Figure 3.4.1: Unimodel Architecture for multiclass classification

Reason behind using both Global Average Pooling and Flatten was to reduce the need of number of hyper parameters by providing as much features from the pretrained models.

Training Hyperparameters: Hyper parameter tuning been done after experimenting with some parameters and found the best parameters for model's accuracy matrics on validation data. Below table 3.4.1 shows used hyperparameters for training.

Property	Value
Optimizer	Adam
Loss	categorical_crossentropy
Learning Rate	Mammo: 0.0001 , Ultra&Thermal: 0.001
Epoch	10-20
Mini Batch Size	32,40,50,64,96

3.4.2 Multimodal Testing

In Multimodal pretrained CNN models, feature level fusion was implemented to combine all three breast cancer image modalties. The project had used two distinct methods to make Multimodal models where each method contains two alternatives (a&b). All the models were fine-tuned with very low learning rate (0.0001) for last convolutional layers of their pretrained CNN models.

Method_1: Using only most accurate pretrained model found from unimodal testing across modalities, which was Inception-ResnetV2(IRV2) to extract features from all three modalities.

Method_2: In this method, using three distinct pre-trained CNN models for three image modality to extract features from all three modalities. Associated models to datasets shown in below table 3.4.2.1

INCEPTION-RESNET V2 (IRV2)	Mammography Dataset
RESNET50	Ultrasound Dataset
INCEPTION V3	Thermal Dataset

For both methods, two alternative model architectures implemented on top level.

Model Architecture:

In the method1, training datasets from mammography, ultrasound and thermal dataset were the inputs to the IR-V2 models and extracted features individually for each dataset. Conv_7b, a last layer of this CNN model containing more than 3 million parameters were fine-tuned, allowing to train on specific task much faster with low amount of training data. In alternative model 1(a), concatenated low dimensional output directly fed into last layer of three dense outputs whereas there was 48 nodes of dense layer with l2 regularizer between these layers in the other alternative model 1(b). The purpose of adding dense layer in second alternative was to give more parameters to learn since it increased the available parameters to 9 million from 4.

In the method2, IR-V2, the most complex model with highest numbers parameters in last CNN layer, was fine-tuned for mammography dataset which was the image modality giving lowest accuracy in unimodal testing and needed to be trained on more parameters. For ultrasound dataset, Resnet50 was fine-tuned with last layer called 'conv5_block3_3_conv' containing 1 million parameters. But InceptionV3 was not fine-tuned for thermal dataset since it was showing significant classification accuracy without fine-tuning in unimodal testing and fine-tuning may increase the chances for over fitting. For alternative model 2(b), 24 nodes for dense layer were chosen with l2 regularizer, containing same number of parameters (9 million) as method 1(b) model contained for comparison between two different approaches on equal level of parameter complexity.

Training Hyperparameters: Below table 3.4.2.2 shows used hyperparameters for training.

Property	Value
Optimizer	Adam
Loss	SparseCategoricalCrossentropy
Learning Rate	0.0001
Epoch	10-20
Mini Batch Size	64

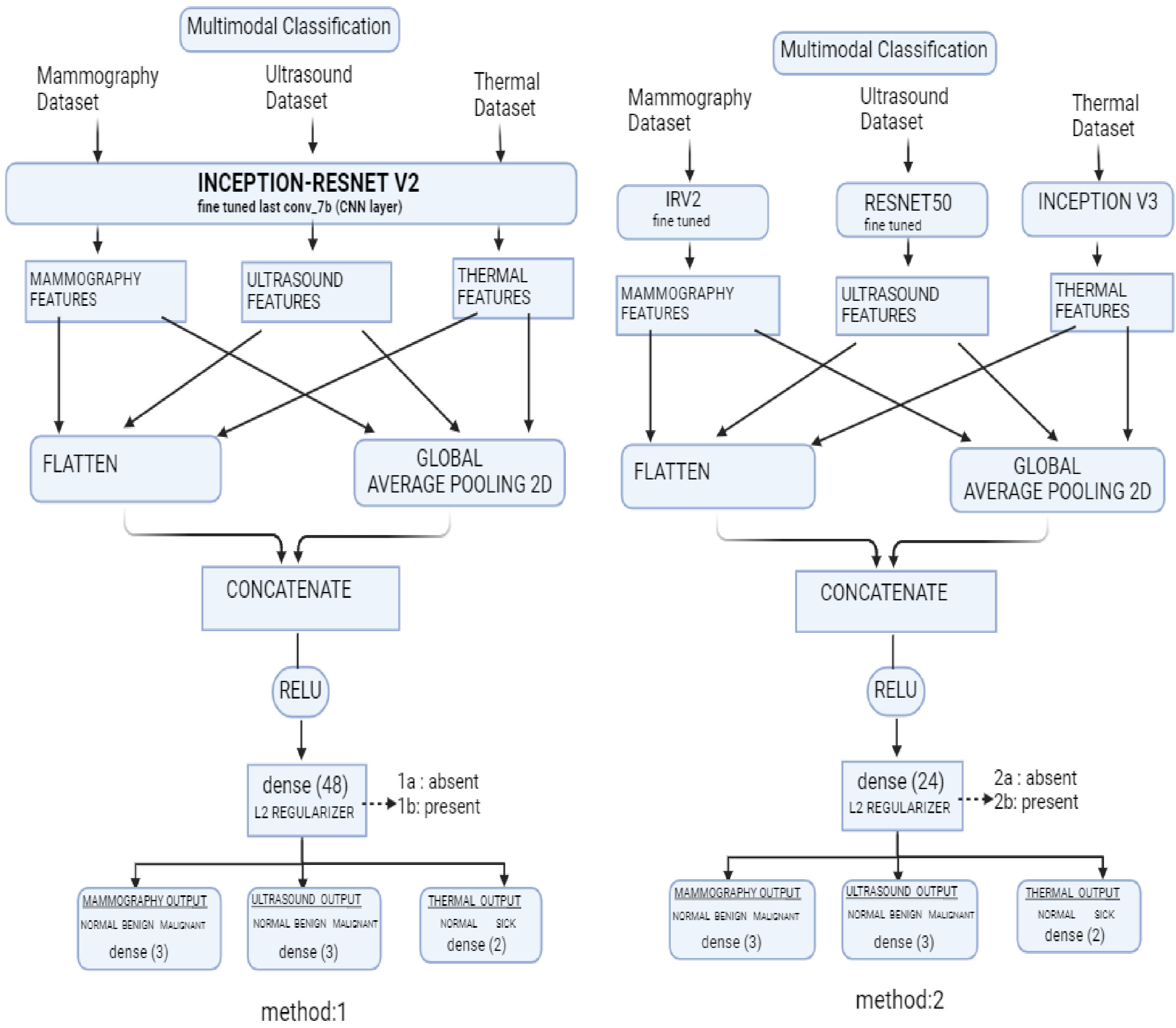


Figure 3.4.2: Method 1 & 2 for Multimodal Classification Architecture

3.5 Model Evaluation Metrics

In this work, the breast cancer classification task had been evaluated using the most descriptive metrics. Specifically, after training, the model were assessed using Precision, Recall, and Accuracy. Precision measures the percentage of individuals who truly have the disease among those diagnosed by the network. Recall indicates the percentage of actual disease cases that are correctly identified by the network as positive. Accuracy represents the overall proportion of correct diagnoses, including both those with and without the disease.

- *True Positive (TP):* No of observations were positive and predicted to be positive.

- *False Negative (FN):* No of observations were positive but predicted negative.

- *True Negative (TN):* No of observations were negative and predicted to be negative.

- *False Positive (FP):* No of observations were negative but predicted positive.

Evaluation metrics	Formula
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$
Precision **Positive Predictive Value**	$\frac{TP}{TP + FP}$
Sensitivity (Recall) **True Positive Rate**	$\frac{TP}{TP + FN}$
Specificity **True Negative Rate**	$\frac{TN}{FP + TN}$
F1 Score	$\frac{2 * Sensitivity * Precision}{Sensitivity + Precision}$

Table 3.5: Evaluation metrics.

3.6 Resources & Computing Tools Used

Curated datasets were used as mentioned before.

1. Python virtual Environment in Ubuntu (WSL):
   - GPU T1000
   - NVIDIA's CUDA and cuDNN libraries
2. Kaggle NoteBook:
   - GPU P100: 35+hrs. Of training
3. Imported Libraries:
   - Tensorflow vers.16
   - OpenCV
   - Keras pretrained Models & Model API
   - Numpy
   - Pillow
   - Sklearn
   - Matplotlib
4. Pretrained Models and Weights:
   - RadImagenet & CheXnet pretrained models

Chapter 4: Results & Discussion

4.1 Introduction

In this chapter, the results of experiments with 15 pretrained models of unimodal testing (five models each for three datasets) and 4 models of multimodal testing (two models each for two methods) had been presented. During training, validation loss was decreased as much could and avoided over fitting by early stopping. For the results, The study had focused on the validation and testing accuracy specifically. The following sections will describe the performances of RadImagenet & CheXnet based pretrained models in uni & multi modal testing.

4.2 Results from Unimodal Testing for Three Image Modalities

4.2.1 Trained Models with Mammography Dataset

Chexnet based Densenet121 performed most poorly compared to RadImagenet based pretrained models. Complex model IR-V2, with large number of parameters performed well than other less complex models.

	Precision	Recall	f1 score	Accuracy
Densenet121(c)	52%	43%	36%	43%
Densenet121(R)	57%	57%	56%	57%
Inception V3	64%	61%	59%	61%
IR-V2	67%	66%	65%	66%
Resnet50	60%	61%	58%	61%

Table 4.2.1: Models result for Mammography test set

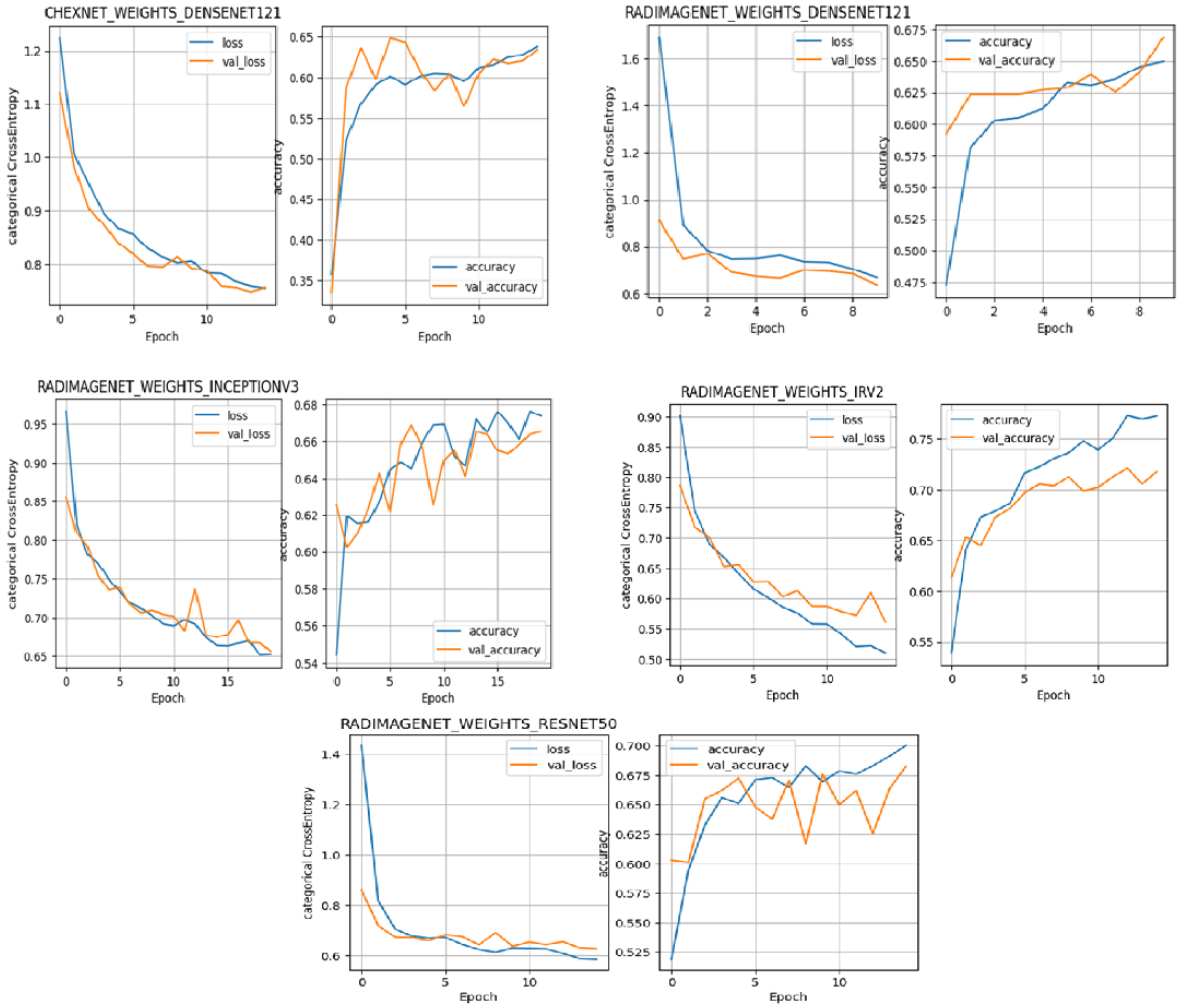


Figure 4.2.1: Training history line graphs for five pretrained models on mammography training dataset

4.2.2 Trained Models with Ultrasound Dataset After fifth epoch, models were overfitting as training loss decreased very quickly compare to validation loss. IR-V2 was best model as per all evaluation metrics shown in table.

	Precision	Recall	f1 score	Accuracy
Densenet121(c)	63%	60%	58%	60%
Densenet121(R)	76%	75%	76%	74%
Inception V3	75%	73%	75%	75%
IR-V2	80%	79%	82%	81%
Resnet50	77%	76%	75%	76%

Table 4.2.2: Models result for Ultrasound test set

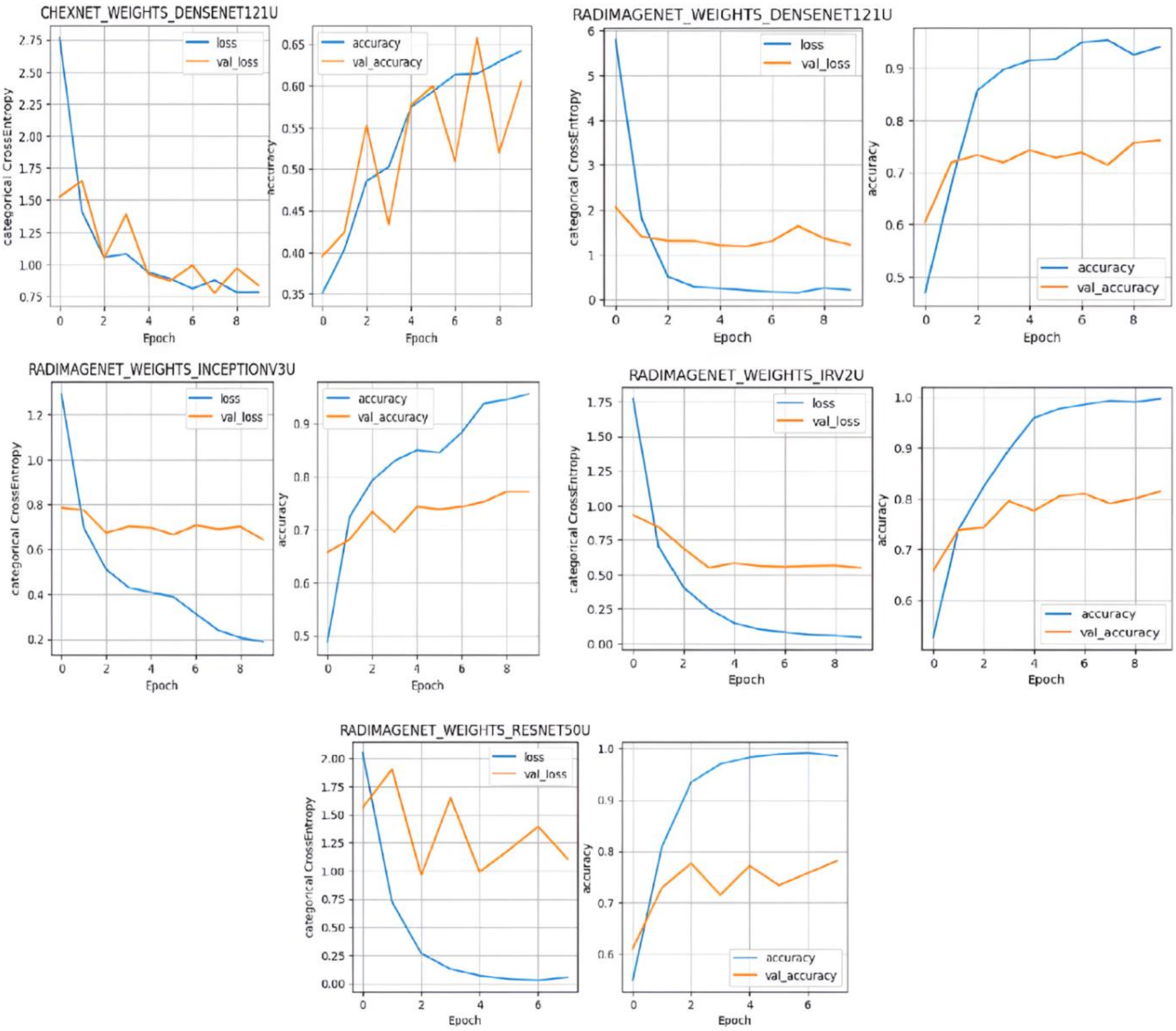


Figure 4.2.2: Training history line graphs for five pretrained models on ultrasound training dataset

4.2.3 Trained Models with Thermal Dataset

Thermal dataset had shown the best results than other two image datasets. Notably, Chexnet Densenet121 gave 87% accuracy which was higher than RadImagenet based Densenet121 model.

	Precision	Recall	f1 score	Accuracy
Densenet121(c)	88%	87%	87%	87%
Densenet121(R)	86%	82%	81%	82%
Inception V3	93%	92%	92%	92%
IR-V2	99%	98%	98%	99%
Resnet50	91%	92%	89%	91%

Table 4.2.3: Models result for Thermal test set

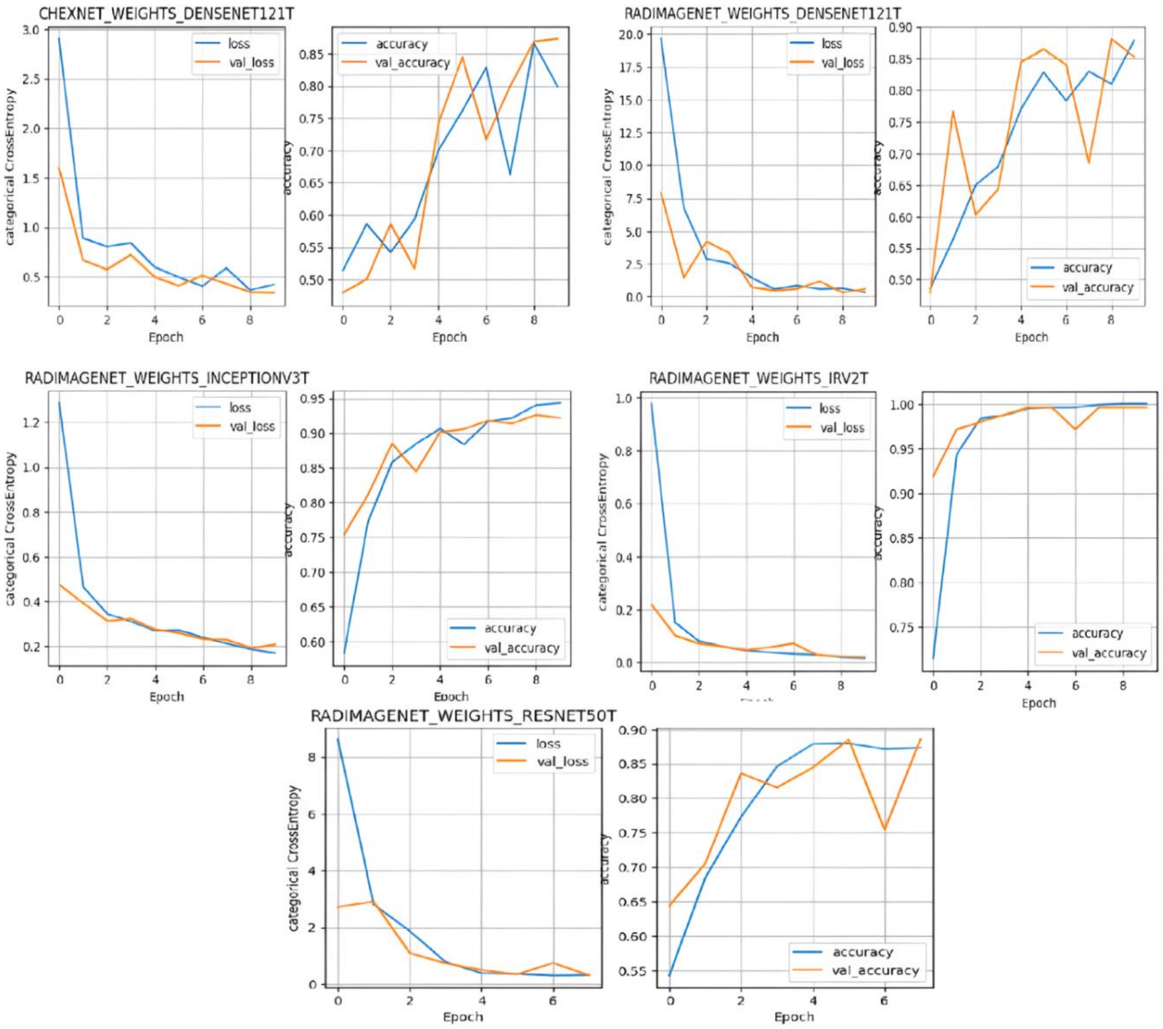


Figure 4.2.3: Training history line graphs for five pretrained models on Thermography training dataset

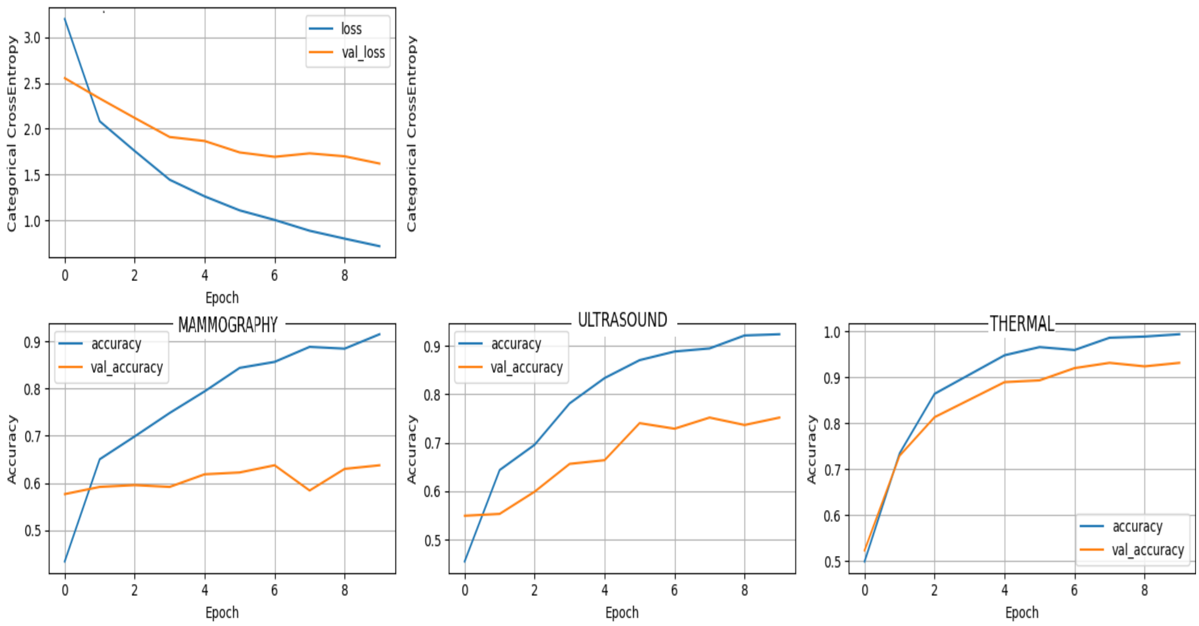
From unitest, it was proved that IR-V2 model was best performing pretrained model across three datasets.

4.3 Results from Unimodal Testing

4.3.1 Multimodal models with Method1

4.3.1(a) With Absence of Dense Layer

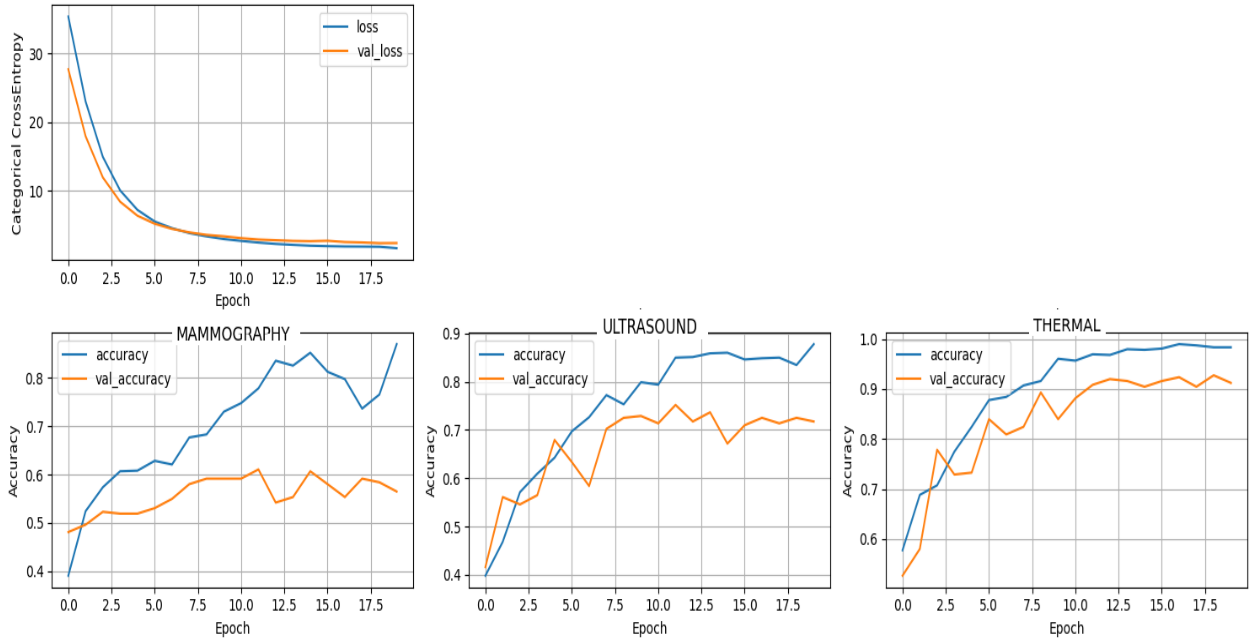
Test Loss: 1.70
Test Accuracy (Mammography): 0.577
Test Accuracy (Ultrasound): 0.935
Test Accuracy (Thermal): 0.695



4.3.1(b) With Presence of Dense Layer

A dense layer present between concatenate layer and final classification dense layer.

Test Loss: 2.3783106803894043
Test Accuracy (Mammography): 0.593
Test Accuracy (Ultrasound): 0.942
Test Accuracy (Thermal): 0.695

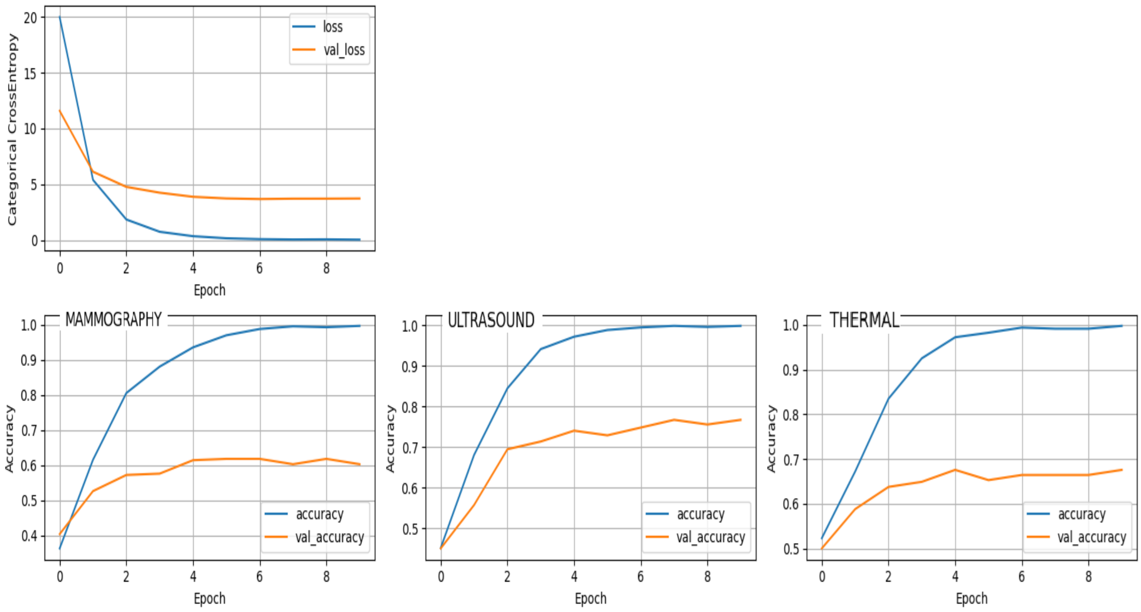


In method1, model with presence of dense layer had comparatively performed better to its alternative model(b) across all three modalities except for thermal images where both models had got same accuracy of 69.5%.

4.3.2 Multimodal models with Method2

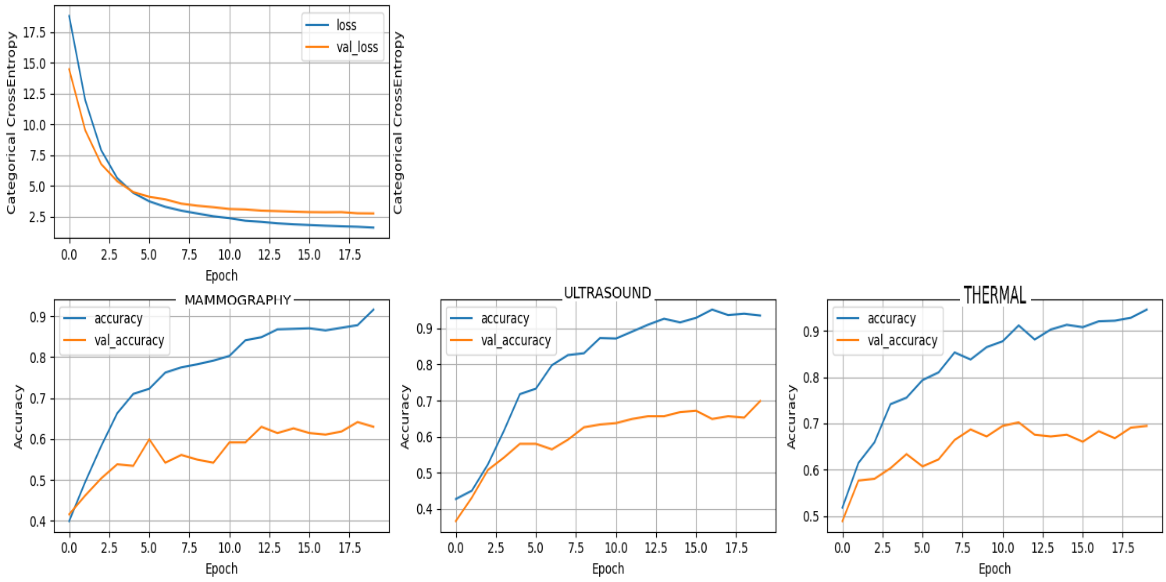
4.3.2(a) Absence of Dense Layer

Test Loss: 3.910769462585449
Test Accuracy (Mammography): 0.585
Test Accuracy (Ultrasound): 0.711
Test Accuracy (Thermal): 0.741



4.3.2(b) Presence of Dense Layer

Test Loss: 2.7040276527404785
Test Accuracy (Mammography): 0.60
Test Accuracy (Ultrasound): 0.714
Test Accuracy (Thermal): 0.684



Multi models with method 2 showed accuracy of 71.1% and 71.4% for Ultrasound datasets which was notably low in compare to multi models of method 1.

4.4 Overall Results Analysis and Discussion

As expected, in unimodal testing, IR-V2 gave best accuracy of 66%, 81%, 99% for Mammograms, Ultrasound and Thermal dataset respectively. RadImagenet based Densnet121 was more accurate than Chexnet except for thermal dataset, showcasing the difference between medical image data used to train these pretrained models; Chexnet pretrained model been trained only on chest X-rays data whereas multi organs and modalities based medical image datasets been used to train the RadImagenet based pretrained models.

In multimodal testing, for mammography, ultrasound, thermography dataset, method 2(b) with 60%, method 1(b) with 94.2% and method 2(a) with 74.1% gave the best accuracy respectively. In consideration of overall accuracy (average accuracy for 3 modalities), method 1(b) was the best performing model with 74.33% accuracy. Interestingly, best two models with best overall accuracy were from method 1, reaching small training loss in comparison to method 2 models. Also, method 1 gave approximately 20% more accuracy for ultrasound dataset than method 2. These statistics performance concluded method 1 as a better technique for multimodal classification.

Classification accuracy on mammography was the poorest among other modalities. Although this accuracy was low than expected, that may be due to image noise or unknown problem associated with preprocessing for this dataset.

Chapter 5: Conclusion

The project successfully demonstrated the potential of using medical domain-specific pre-trained CNN models for breast cancer classification across different imaging modalities with both uni & multi-modal based approaches. The individual performance of models on mammograms, ultrasound, and thermal images was significantly enhanced by using pre-trained models & weights and appropriate data augmentation techniques. The RadImagenet based Inception-Resnet-v2 model consistently outperformed other architectures in both uni & multi modal testing, indicating its potential suitability for breast cancer classification specifically in CNN domain. Project had showed new approach to reduce number of hyper parameters to train CNN models by using Global Average Pooling 2D and Flatten combined.

RadImagenet and CheXnet based pretrained models are very recent developments, and in future, medical domain indeed will get more medical domain specific pretrained models which will be highly specialized for its application with more accuracy since more medical data will be available to train deep learning models.

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

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

[3] CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. https://doi.org/10.48550/arXiv.1711.05225

[4] Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep learning for breast cancer diagnosis from mammograms—a comparative study. Journal of Imaging, 5(3), 37. https://doi.org/10.3390/jimaging5030037

[5] RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. https://doi.org/10.1148/ryai.210315

[6] DMR -IR Data Set From http://visual.ic.uff.br/en/proeng/thiagoelias/

[7] King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD) https://doi.org/10.3390/data6110111/

[8] Adeshina S. A., Adedigba A. P., Adeniyi A. A., Aibinu A. M. (2018). "Breast Cancer Histopathology Image Classification with Deep Convolutional Neural Networks". 2018 14th International Conference on Electronics Computer and Computation (ICECCO), pages 206- 212.

[9]] Fernando B., Fromont E., Muselet D., Sebban M. (2012). "Discriminative feature fusion for image classification." 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pages 434–3441.

[10] Gandomkar, Z., Brennan, P. C., & Mello-Thoms, C. (2018). "MuDeRN: Multi-category classification of breast histopathological image using deep residual networks." Artificial Intelligence in Medicine, 88, pages 14–24.

[11] Houssein, E. H., Emam, M. M., Ali, A. A., & Suganthan, P. N. (2021). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications, 167, 114161. https://doi.org/10.1016/j.eswa.2020.114161

[12] Falconi, L. G., Perez, M., Aguila, W. G., & Conci, A. (2020). Transfer learning and fine tuning in breast mammogram abnormalities classification on cbis-ddsm database. Advances in Science, Technology and Engineering Systems Journal, 5(2), 154-165. doi:10.25046/aj050220.

[13] Wu, N., Phang, J., & Shen, Y. (2020). Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE Transactions on Medical Imaging.

Appendix

Appendix A: To access Used Code & Dataset in This Project

To access full code and datasets used in the project visit github repository
https://github.com/Sanket10j/Medical-Based-Transfer-Learning-Approach-For-Uni-And-Multi-Modal-Breast-Cancer-Classification

Appendix B: Code of Self-Created CNN Model

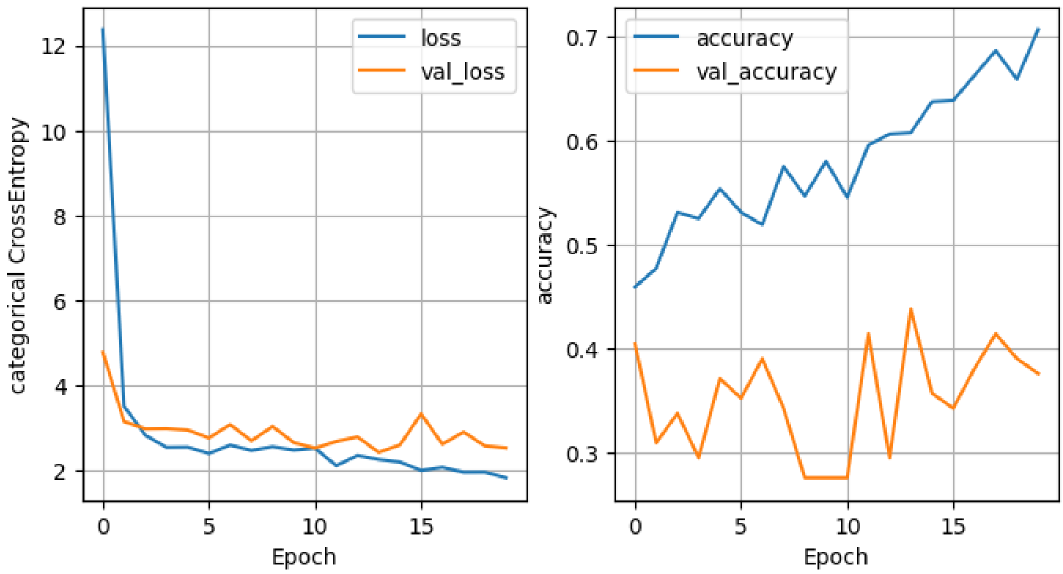


Figure Appendix B: Training history line graph of self-created model

Self-Created model was trained on mammography dataset. It showed very high amount of overfitting to data even though Dropout and BatchNormalization was used to reduce it. Result clearly indicates lack of training data and this problem had been solved with approach of Transfer learning as shown in project.

Appendix C: Ethical Concerns for Medical Implications of Deep Learning for Breast Cancer Classification

This part will discuss ethical concerns surrounding a deep learning model designed to classify breast cancer tumors. Here's a concise breakdown of the key points:

Safety and Training Data

- *Mislabeled data:* Inaccurate labels in training datasets can lead the model to make wrong classifications, jeopardizing patient safety.
- *Data bias:* Datasets lacking diversity in ethnicity or breast density can lead to falsely high accuracy and missed diagnoses.
- *Opacity of deep learning:* The deep learning models contain black box nature, making it difficult to understand how it arrives at diagnoses.

Risks and the Medical Professional

- *Liability concerns:* Radiologists using the model lack clear justification for diagnoses, making it challenging to explain misinterpretations.
- *Overreliance on results:* High accuracy might lead to overlooking the model's limitations and potential misdiagnoses.
- *Compassionate care vs. efficiency:* The model prioritizes accuracy, potentially conflicting with a doctor's duty to provide compassionate care based on individual needs.

Informed Consent

- *Patient privacy:* While using open-source data protects immediate privacy, collecting new data requires informed consent from patients.
- *Doctor's knowledge gap:* Medical professionals might not fully understand the model's workings, hindering proper explanation to patients.
- *Nudging patients:* Presenting the model as the default option or failing to clearly explain its limitations could influence patient decisions unfairly.

Conclusion

Deep learning models for breast cancer diagnosis offer potential benefits, but ethical considerations regarding training data, interpretability, and patient autonomy require careful attention.