Multimodal Breast Cancer Classification with Medical Based Transfer Learning CNN Algorithms

Under Guidance of:

Guide:

Dr. Kshipra Chauhan

Assistant Professor

School of Applied Sciences & Technology

Guide:

Prof. Mahesh Panchal

Assistant Professor

Graduate School of Engineering &Technology

PRESENTED BY:

Sanket M. Chaudhary
EN.NO:231451302004
PGDBI

School of Applied Sciences and Technology

Gujarat Technological University

Introduction

- **Breast Cancer: A Global Challenge**

  According to WHO, the most common cancer affecting women worldwide is breast cancer and better treatment depends on its early and accurate detection.

- **Limitations of Single-Modality Diagnosis**

  Individual diagnostic methods have limitations to cover up complexity for detecting cancer with great accuracy. Ex. mammography gives high false-negative results than MRI for detecting cancer in dense breast tissue.

- **Potential of Multimodal Medical Imaging**

  Multimodal CNN models, combines data from several imaging modalities such as ultrasound, thermal images, and mammography can be promising approach for improved breast cancer classification.

AIM & OBJECTIVES

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

- **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.

Background Information

- **Deep Learning**

  A subfield of machine learning that can automatically learn and extract features from data through a process called training. These features are then used to make predictions or classifications.

- **Convolutional Neural Networks (CNNs)**

  A specific type of deep learning architecture particularly well-suited **for image analysis** tasks such as image classification, object detection.

- **3 most basic components of CNN**
- The Convolution Layer
- The Pooling Layer
- The Output Layer (or) Fully Connected Layer

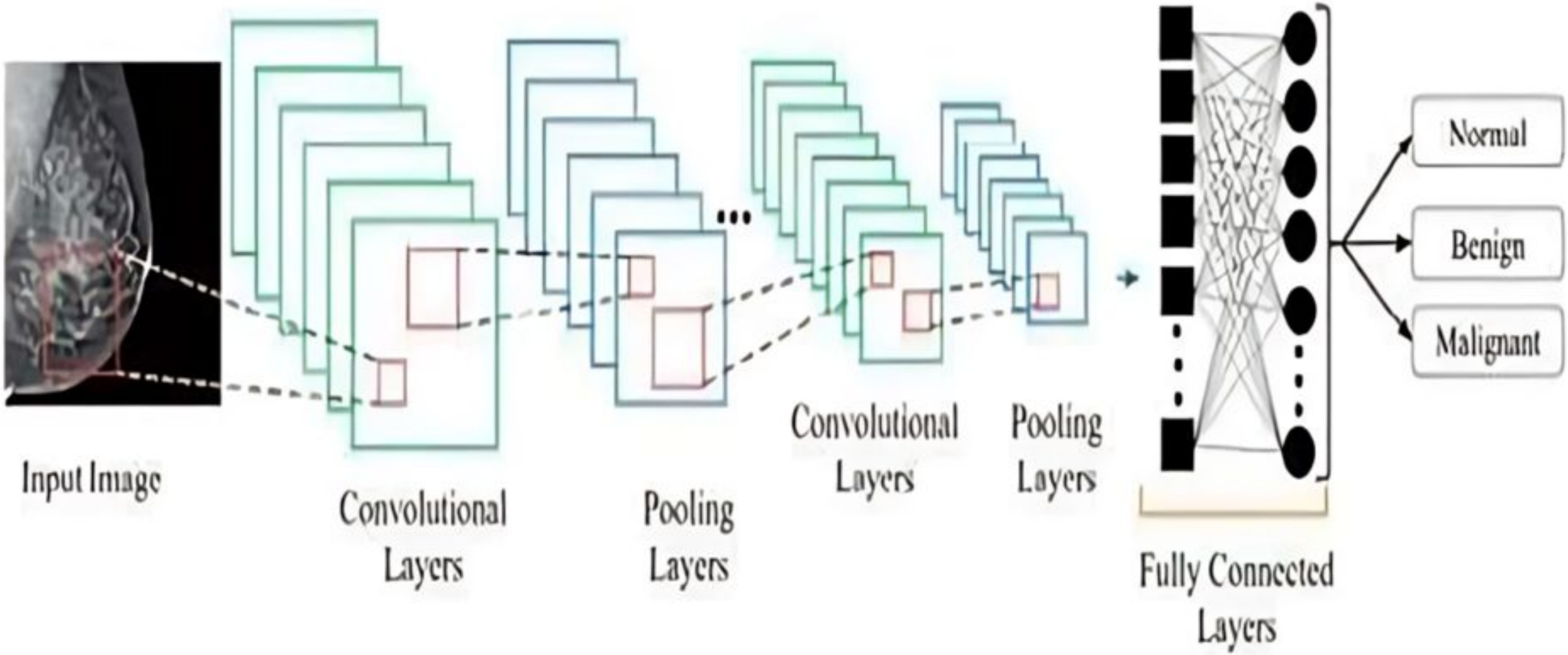
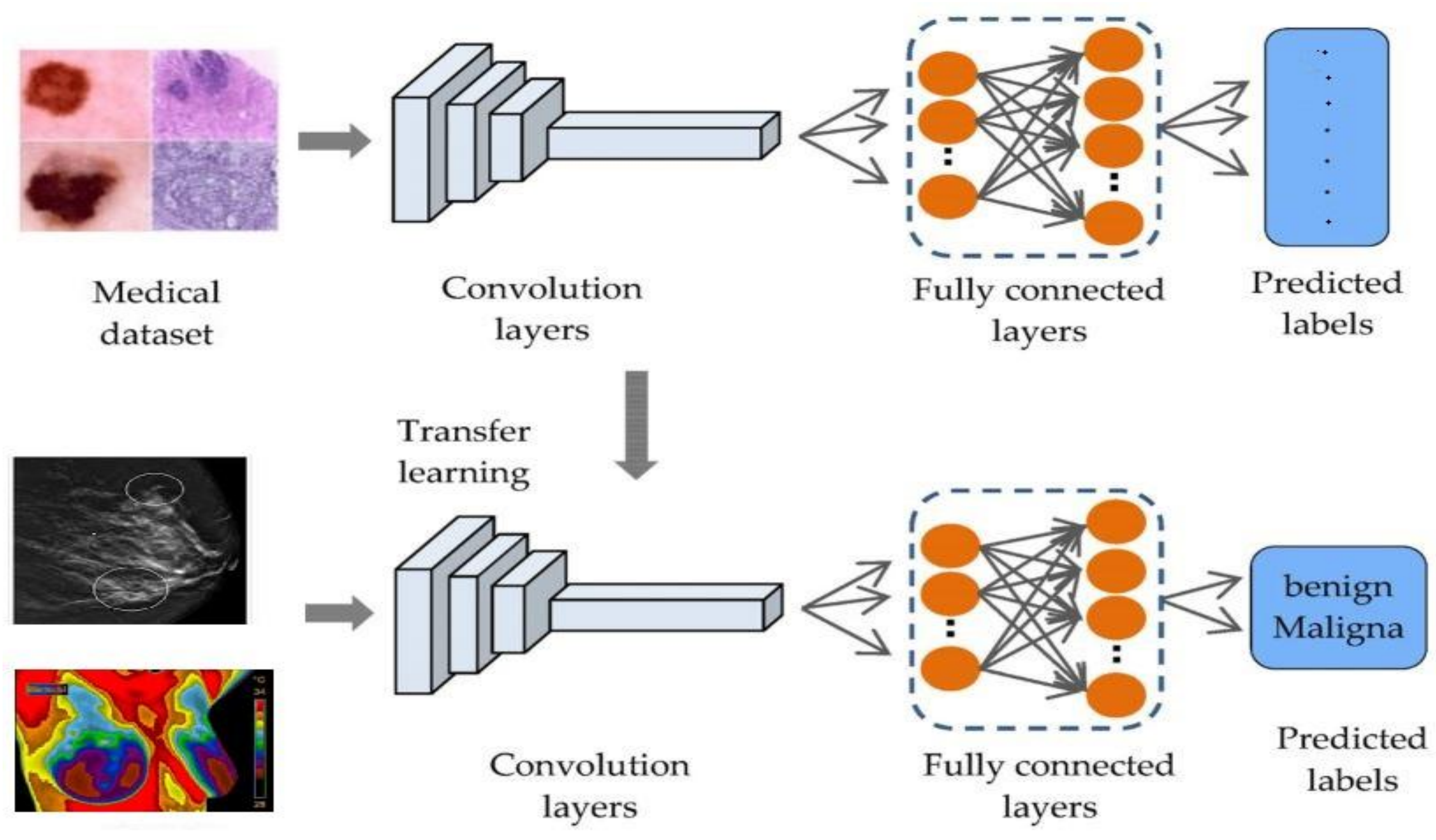
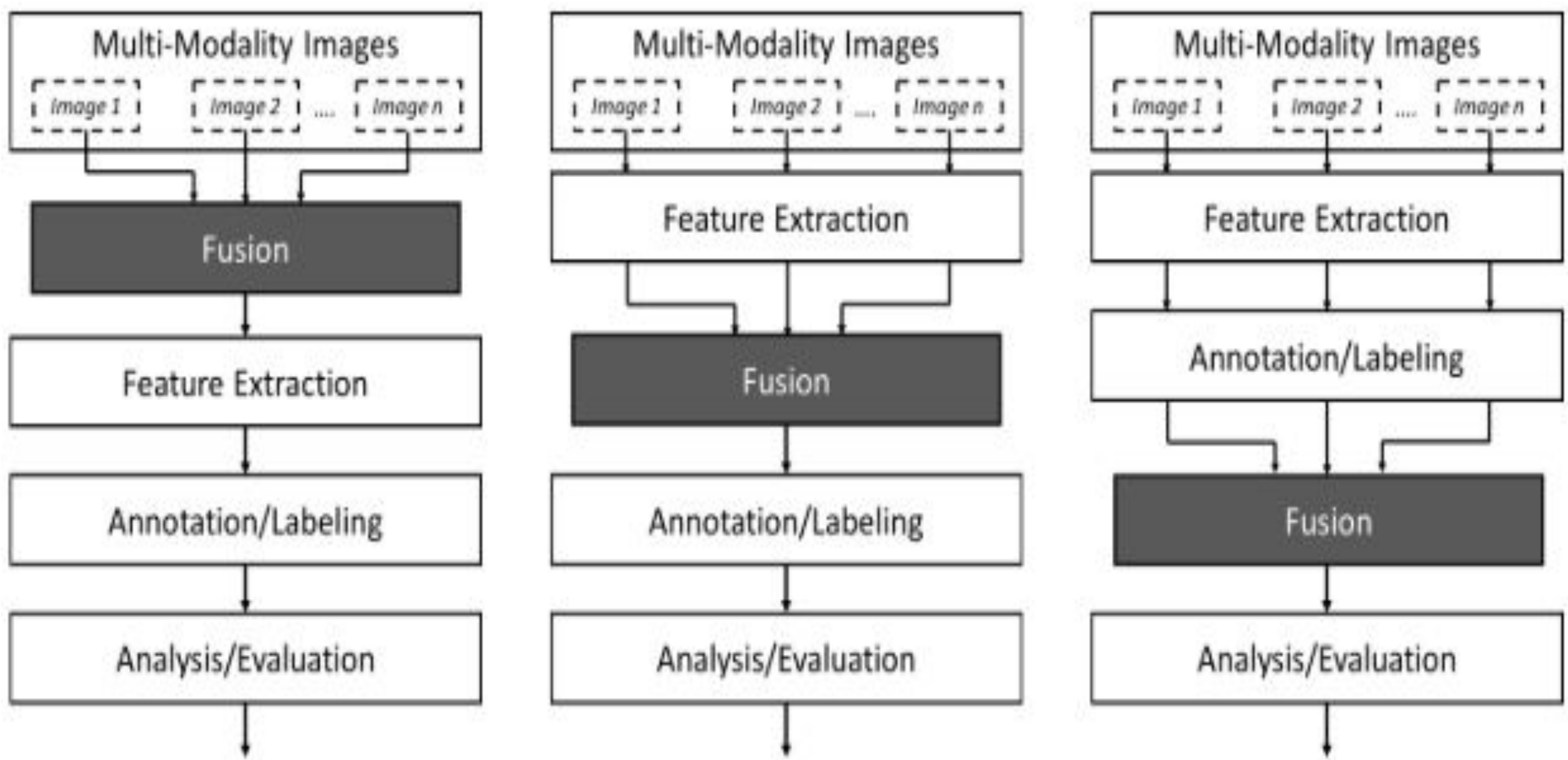


Figure showcasing portions of CNN Architectures from Breast Mammography Image

- Medical Based Transfer Learning

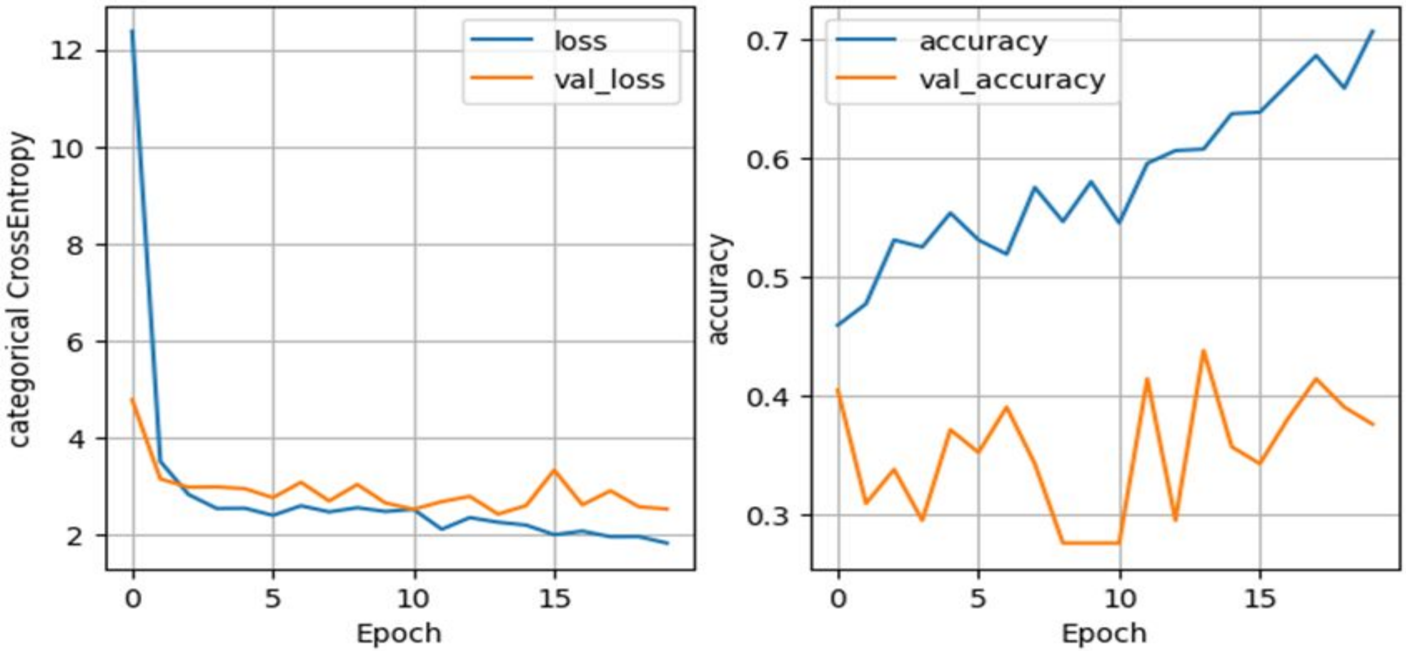


- Feature Level Fusion



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

Why to Even Use Pretrained Models: Self Created CNN Model



- Self-Created model 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.

Methodology & Results

1. Datasets Curation & Description

2. Data Pre-processing & Augmentation

3. Unimodal Models Training & Results

4. Multimodal Models Training & Results

5. Used Model Evaluation metrics

1.Datasets Curation & Description

- **Breast Cancer Mammograms: 1754 Images**

  Source : King Abdulaziz University Mammogram Dataset

  Classes: 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 me to reduce the numbers of class by converting dataset into 3 classes (normal: 1265 , benign: 387, malignant: 102 Images).

- **Breast Ultrasound Images Dataset : 1578 files**

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

  Classes: Normal, benign, malignant.

- **Thermal Images for Breast Cancer Diagnosis: 1542 files**

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

  Classes: Normal (780 Images), sick (762 Images)

2. Data Pre-processing & Augmentation

- **Image Resizing (used 224*224)**
  Used dataset contains images with different dimensions and without resizing, model would have to continuously adjust to the different resolutions of the images, which would reduce the CNN's learning efficiency.

- **Data Augmentation**
  To Increase the diversity of the training dataset by applying transformations such as rotation, flipping, scaling, and cropping to the images. It also helped to reduce classification class data imbalance and improved generalizability across all classes.

- **Data Split**
  Data was then randomly shuffled and split into train, validate and test dataset in **60:20:20** ratio.

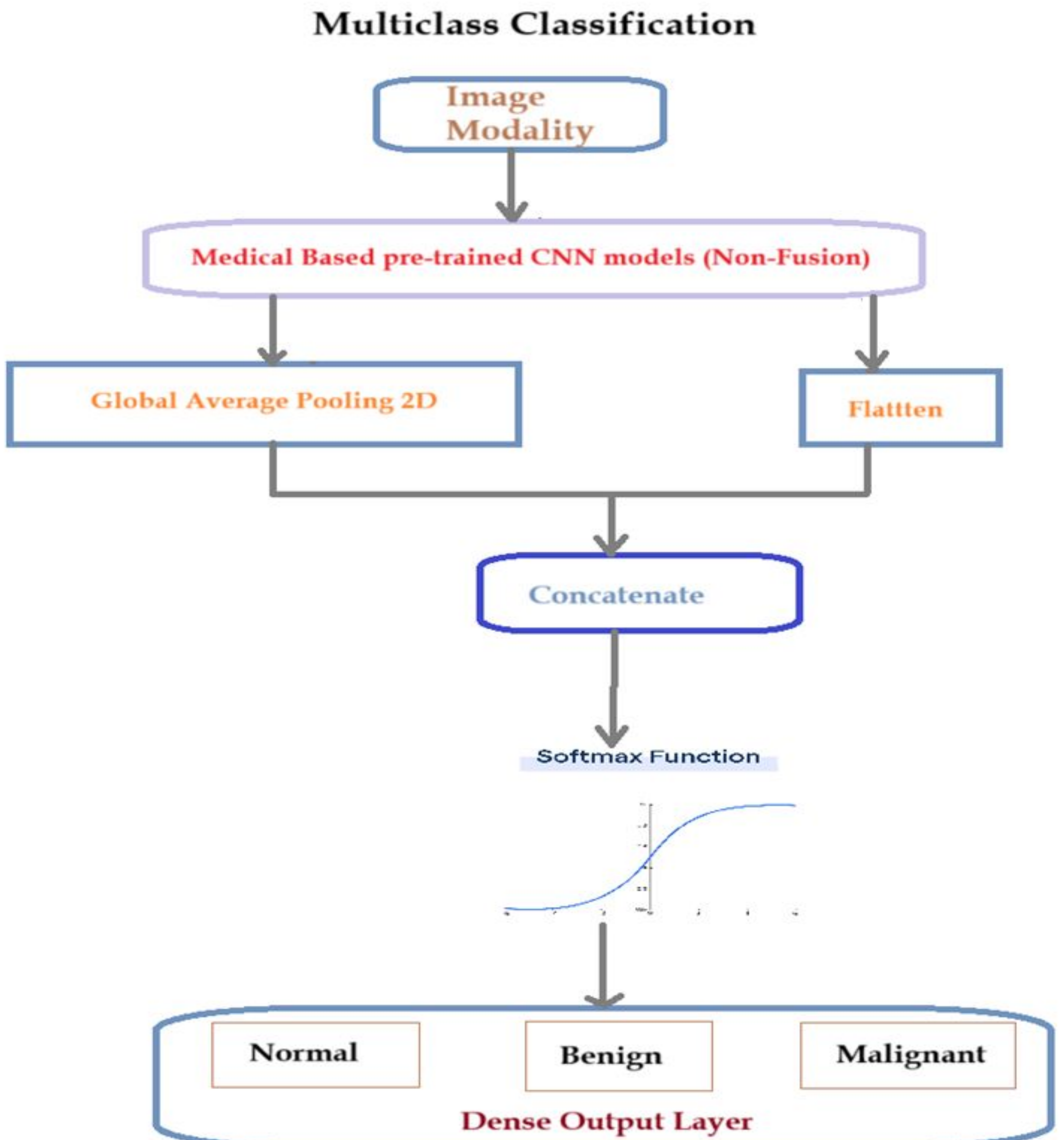
- **Down Sample**
  Multimodal fusion requires same numbers of images for all three modalities so, three datasets were minimally down sampled (randomly) for training datasets for 1048 samples for each dataset.

3.Unimodal Models Training

- For each breast cancer 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.
- **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.

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

Unimodal model Architecture



Results for three Datasets

- **Mammography Dataset**

	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%

- **Ultrasound dataset**

	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%

- **Thermal Dataset**

	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%

Overall Results Analysis:

- 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, since 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.

4. Multimodal Models Training

All the models were **fine-tuned** with very <u>low learning rate (0.0001)</u> for last convolutional layers of their pretrained CNN models.

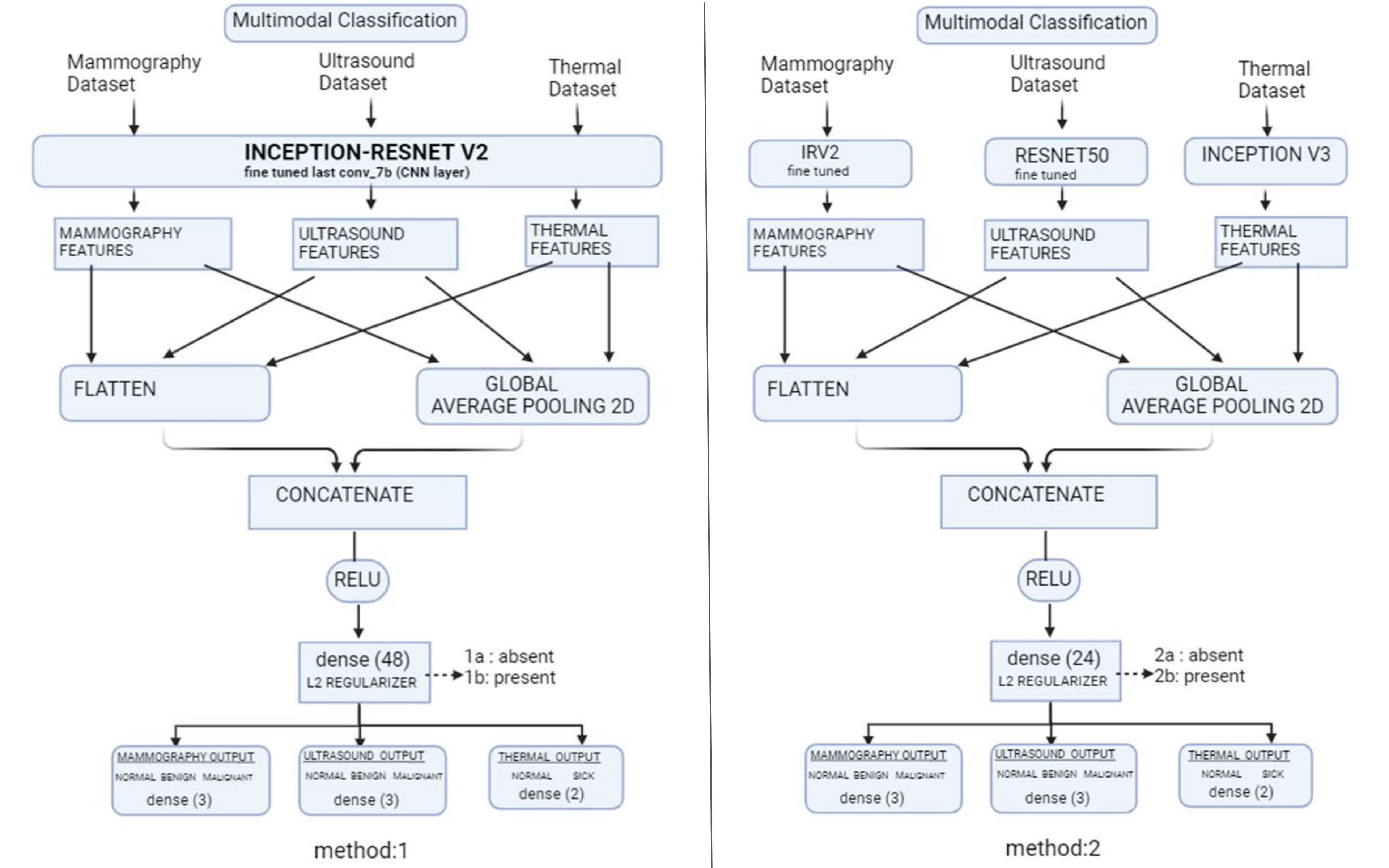
<u>Method1:</u> Using **only most accurate pretrained model** found from unimodal testing across modalities, which was **Inception-ResnetV2(IRV2)** to extract features from all three modalities.

<u>Method2:</u> Using **3 distinct pre-trained CNN models** for three image modality to extract features from all three modalities. Associated models to datasets shown in below table

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

Training Hyperperparameters:

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


Multimodal Classification
Mammography Dataset
Ultrasound Dataset
Thermal Dataset
INCEPTION-RESNET V2
fine tuned last conv_7b (CNN layer)
MAMMOGRAPHY FEATURES
ULTRASOUND FEATURES
THERMAL FEATURES
FLATTEN
GLOBAL AVERAGE POOLING 2D
CONCATENATE
RELU
dense (48)
L2 REGULARIZER
1a : absent
1b: present
MAMMOGRAPHY OUTPUT
NORMAL BENIGN MALIGNANT
dense (3)
ULTRASOUND OUTPUT
NORMAL BENIGN MALIGNANT
dense (3)
THERMAL OUTPUT
NORMAL SICK
dense (2)
method:1
Multimodal Classification
Mammography Dataset
Ultrasound Dataset
Thermal Dataset
IRV2
fine tuned
RESNET50
fine tuned
INCEPTION V3
MAMMOGRAPHY FEATURES
ULTRASOUND FEATURES
THERMAL FEATURES
FLATTEN
GLOBAL AVERAGE POOLING 2D
CONCATENATE
RELU
dense (24)
L2 REGULARIZER
2a : absent
2b: present
MAMMOGRAPHY OUTPUT
NORMAL BENIGN MALIGNANT
dense (3)
ULTRASOUND OUTPUT
NORMAL BENIGN MALIGNANT
dense (3)
THERMAL OUTPUT
NORMAL SICK
dense (2)
method:2

Results for method 1

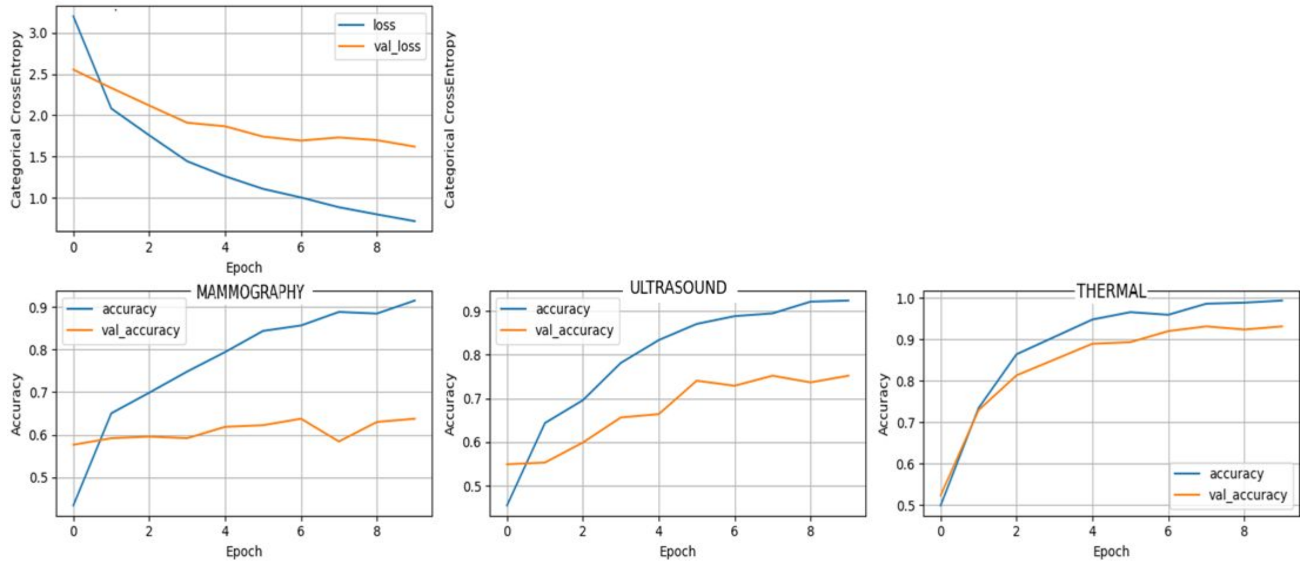
(a) With Absence of Dense Layer

Test Loss: 1.70

Test Accuracy (Mammography): **0.577**

Test Accuracy (Ultrasound): **0.935**

Test Accuracy (Thermal): **0.695**



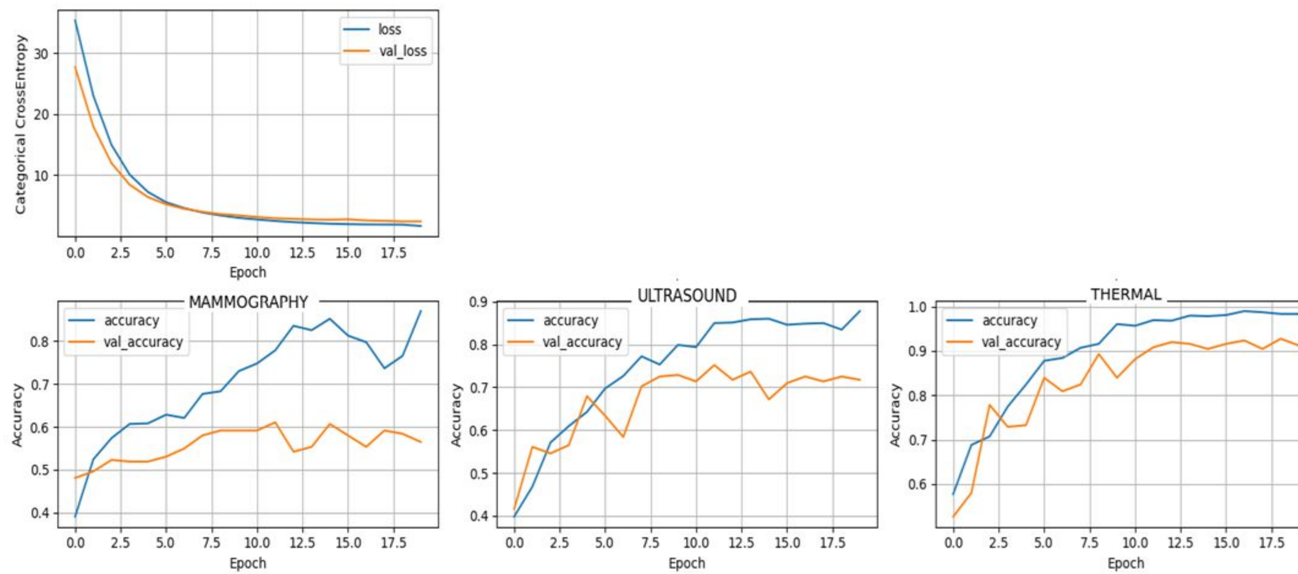
(a) With Presence of Dense Layer

Test Loss: 2.38

Test Accuracy (Mammography): **0.593**

Test Accuracy (Ultrasound): **0.942**

Test Accuracy (Thermal): **0.695**



Results for method 2

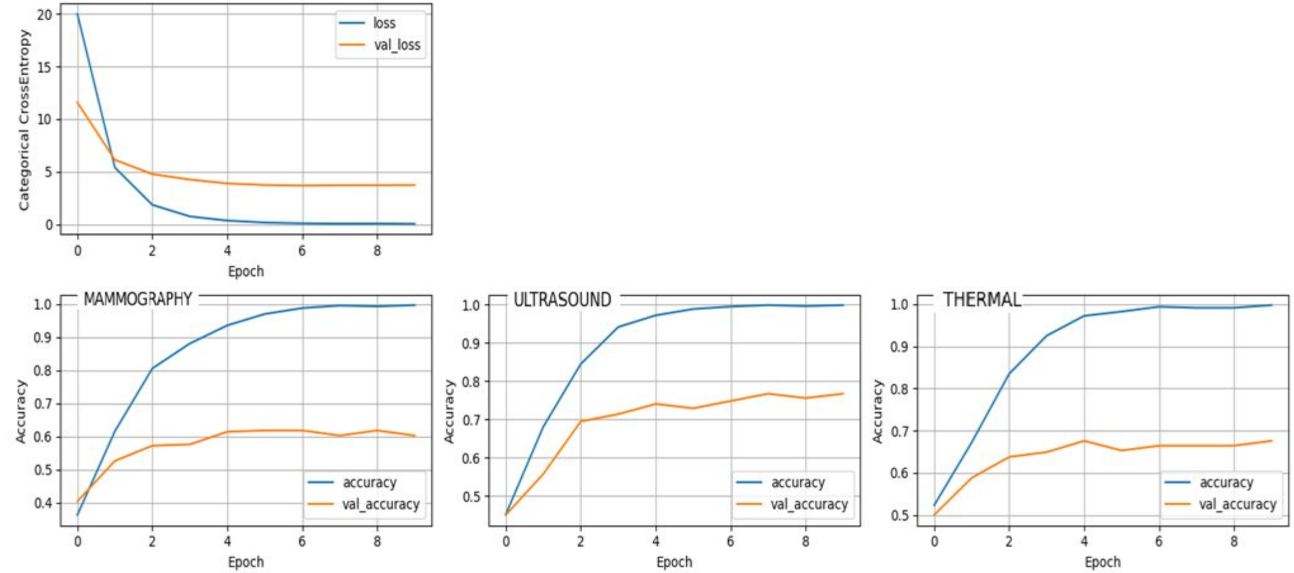
(a) With Absence of Dense Layer

Test Loss: 1.70

Test Accuracy (Mammography): **0.577**

Test Accuracy (Ultrasound): **0.935**

Test Accuracy (Thermal): **0.695**



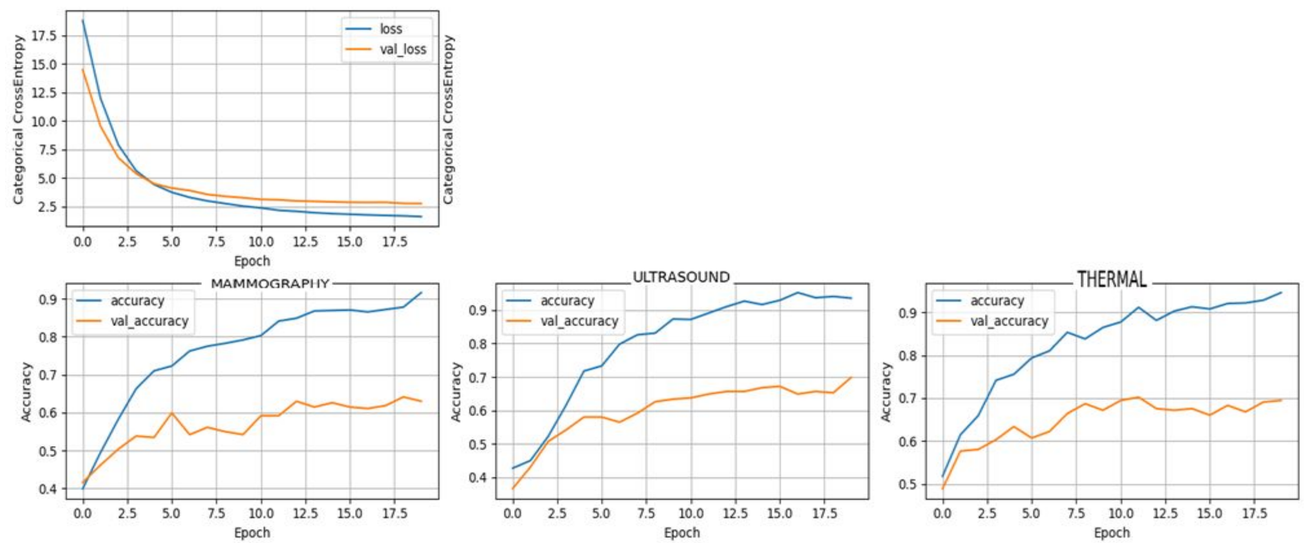
(a) With Presence of Dense Layer

Test Loss: 2.38

Test Accuracy (Mammography): **0.593**

Test Accuracy (Ultrasound): **0.942**

Test Accuracy (Thermal): **0.695**



Overall Results Analysis:

- For mammography, ultrasound, thermography dataset, <u>method 2(b) with 60%</u>, <u>method 1(b) with 94.2%</u> and <u>method 2(a) with 74.1% gave the best accuracy</u> respectively.
- In consideration of overall accuracy (average accuracy for 3 modalities), **method 1(b)** was the **best performing** model with **74.33%** accuracy.
- 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 problems associated with preprocessing for this dataset.

5.Used Model Evaluation metrics in the project

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.

- ***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.

Conclusion

- The project 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.

- 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.

Resources & Computing Tools Used

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: @article{doi:10.1148/ryai.210315,
  author = {Mei, Xueyan and Liu, Zelong and Robson, Philip M. and Marinelli, Brett and Huang, Mingqian and Doshi, Amish and Jacobi, Adam and Cao, Chendi and Link, Katherine E. and Yang, Thomas and Wang, Ying and Greenspan, Hayit and Deyer, Timothy and Fayad, Zahi A. and Yang, Yang}

Thank You