

# MSc Data Science Project

7PAM2002-0901-2024

Department of Physics, Astronomy and Mathematics

## **Data Science**

## **FINAL PROJECT REPORT**

### **Project Title:**

Breast Cancer Classification Using Mammogram Images  
for Enhanced Early Detection

### **Student Name and SRN:**

Enter Your Full Name and Student ID

Supervisor: Enter the full name of your Supervisor

Date Submitted: Enter the date you are submitting this report

Word Count: Enter the word count

GitHub address: Enter your linked GitHub address

# Abstract

---

This research focuses on breast cancer diagnosis using two advanced models, **DenseNet121** and **ResNet50** have been integrated together to form a hybrid model using transfer learning methods to obtain a complete feature vector, along with a simple **neural network** with *three dense layers* for classification purposes. The aim of this *hybrid model* is to combine the ability of DenseNet121 to create low level feature maps which contain smallest of the particulars in the masses, together with the ability of ResNet50 to have a more comprehensive understanding of the mammographs such as the pattern of overall symmetry of the breasts, establishing high-level features maps. By doing so the model will have a more in-depth understanding of the patterns and properties within the images resulting in improved classification accuracy. Once these feature maps are extracted, they are merged together and fed to a neural network of three fully connected layers for classifying the lump as benign or malignant. The dataset used to train this model is also a *combination of three widely used mammographic datasets* in the literature, **INbreast**, **DDSM** and **MIAS**. Although very small in size, INbreast dataset has the highest quality and consistent images, compared to DDSM dataset which contains approximately 10000 screening mammographs making it suitable for deep studies of the patterns but is projected to preprocessing challenges due to inconsistent quality. On the other hand, the MIAS dataset is well-annotated for ML algorithms, however the digitized mammographs do require some normalization and ROI extraction methods to be applied before training. This unique approach of combining feature maps of two distinct models and a diverse dataset allows leveraging the benefits of the individual models and datasets while mitigating their limitations. In diagnostic imaging analysis it is equally important to justify the prediction as well as the accuracy of diagnosis thus GRAD-CAM (Gradient-weighted Class Activation Mapping) is used to visualize how the algorithm is making decisions. This tool creates heatmaps which if laid above the mammographic image can help envisioning the areas which influenced the decision-making process of the ML classifier the most. The motivation behind this study is to create better CAD systems with improved diagnostic accuracy in order to increase the number of cancer survivors through early tumor detection.

# Contents

---

Abstract.....	2
1. Introduction .....	4
1.1. Background .....	4
2. Literature Review .....	6
3. Methodology (Mammography Classification) .....	14
3.1. Overview .....	14
3.2. Methodology Flowchart.....	14
3.3. Setting up Co-lab Environment.....	15
3.4. Dataset .....	15
3.4.1. <i>INbreast Dataset:</i> .....	15
3.4.2. <i>MIAS Dataset:</i> .....	16
3.4.3. <i>DDSM Dataset:</i> .....	17
3.5. Preprocessing.....	19
3.5.1. <i>Resizing</i> .....	19
3.5.2. <i>Normalizing</i> .....	19
3.5.3. <i>Data Augmentation</i> .....	20
3.6. Labelling and storing the images .....	20
3.7. Model Training .....	20
3.7.1. <i>Densenet121</i> .....	20
3.7.2. <i>ResNet50</i> .....	21
3.7.3. <i>Hybrid Model Approach</i> .....	21
3.7.4. <i>Tweaking the models</i> .....	22
3.8. Grad-CAM.....	26
3.8.1. <i>How Grad-CAM works?</i> .....	26
3.8.2. <i>How Grad-CAM is applied</i> .....	26
4. Results.....	28
5. References .....	35

# 1. Introduction

---

Cancer is a genetic disease caused by cell proliferation of abnormal cells in human body (NCI, 2021) and thus breast cancer is a form of cancer in which these anomalous cells grow and multiply creating lumps (also called tumor) in different parts of the breast. Breast cancer is commonly found in women with the highest mortality rate than any other cancer type (Kaushal et al, 2019).

Due to this reason engineers and medical professionals have been working closely together to bring in advanced techniques, especially in *Computer Aided Diagnostic systems* to help in time diagnosis of breast cancer. There are various scanning techniques available for breast cancer detection, *digital mammography* being the most versatile as it allows improved, reliable as well as efficient management of images and also improves the accuracy of diagnosis by enabling the use of advanced image enhancement and manipulation techniques (Dromain et al, 2013).

This research is utilizing three digital mammography datasets namely **INbreast**, **MIAS** and **DDSM**, to train a hybrid machine learning model consisting of **ResNet50** and **DenseNet121** for timely breast cancer detection. This thesis has been carried out in the field of Data Science and Artificial Intelligence in order to not only decrease the instance of false positive or false negative breast cancer diagnosis due to human error or misinterpretation of scans but also accomplish early breast cancer detection so treatment can begin timely overall increasing the life expectancy of cancer patients.

## 1.1. Background

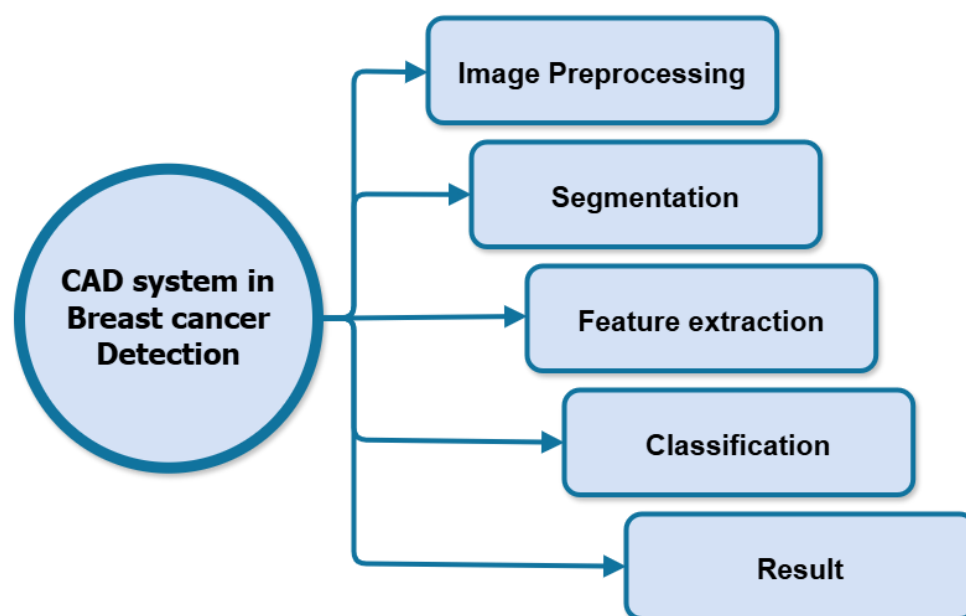
According to Cho, W.C. (2007), there are around 11 million cancer diagnoses each year. In the global cancer statistics report of 2005, there were a total of 7.6 million fatalities due to cancer which makes 13% of the overall 58 million deaths that year. The number of deaths related to cancer is expected to increase by the year 2030 to 11.4 million. Being the most prevalent form of cancer, breast cancer had a high incidence of new cases in 2012, with an extremely high mortality rate as well, 8.2 million deaths out of 14.1 million diagnoses (Cuthrell et al, 2023). By 2040 this trend of breast cancer is expected to double compared to 2012 statistics.

Screening mammography has always been the most traditional method of identifying lumps in breasts and it has significantly decreased the rate of false cancer results (Verma et al, 2010) however it comes with its limitations as the scans are ultimately observed by a radiologist or an oncologist which creates room for misapprehension along with missing details or anomalies which are too minor to be detected by human eye. Mello-Thomas et al, (2003) states that it is estimated that almost among the missed breast cancer cases the majority are thought to be a result of misinterpretation while the rest are ignored lesions. This problem is eliminated by the research and development in artificial intelligence as well as medical imaging including computer-aided diagnosis (CAD) in medicine in order to improve the sensitivity of mammography exams.

The development of CAD systems for breast cancer analysis commenced in 1980's and soon after the first industrial CAD system was authorized as a secondary evaluation for screening mammography by Food and Drug Administration (FDA) (Chan et al, 2019). Consequently by

2016, approximately 92% of American medical professional were making use of this digital facilities (Mello-Thomas et al, 2003). Various studies have been carried out solely to recognize the improvement in cancer diagnosis after the introduction of CAD systems for example Jiang et al, (1999) carried out a research for this purpose and the results showed that average ROC curve grew from the area of 0.61 to 0.75, while for harmful lesions the system increased the number of callbacks to 6.4 ( $P=0.006$ ) and in case of harmless lumps there were reduction in biopsies by 6.0 ( $P=0.003$ ) hence resulting in better positive biopsy yield, sensitivity and specificity.

In conclusion the significant improvement in the reading of mammography scans and ultimately increased cancer diagnostic accuracy, is a key innovation in the application of artificial intelligence for breast cancer diagnosis.



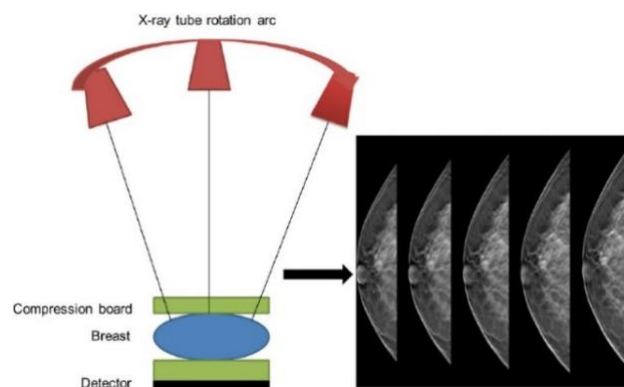
*Figure 1: Different stages of a CAD system. To analyze a mammogram for detection the system first applies some preprocessing techniques on the image, then segments the image, then extracts high-level and low-level features, then classifies the mammogram as having cancerous tumor or not then displaying the findings.*

## 2. Literature Review

*Breast cancer* is the most frequently diagnosed cancer among women worldwide, remaining a primary cause of mortality, especially in resource-limited areas (Jung et al., 2013). In 2020, approximately **2.3 million new cases** of breast cancer were recorded, resulting in an estimated 684,000 deaths globally (Cancer Today, 2020). The development of breast cancer is influenced by a multitude of *risk factors, including age, genetic susceptibility, familial cancer history, and lifestyle variables such as diet, obesity, and exposure to environmental pollutants* (Michael et al., 2022). Typically originating in the ducts or lobules of the breast, breast cancer often progresses gradually in its initial stages before potentially metastasizing (Bernardi et al., 2012). Although many breast lumps are benign, malignant tumors can develop asymptotically, underscoring the critical need for routine screenings, particularly given the absence of early symptoms. Early intervention is pivotal for favorable outcomes, as untreated breast cancer tends to progress, complicating treatment and diminishing survival rates (Sung et al., 2021).

The evolution of medical imaging technologies has significantly enhanced breast cancer diagnostics, enabling earlier and more precise interventions. **Mammography** remains the principal screening modality for breast cancer, with extensive research affirming its capacity to reduce mortality rates by enabling early detection, especially among women aged 40 to 74 (López et al., 2021). Digital mammography has further improved diagnostic accuracy by providing high-resolution imaging, which enhances visualization of overlapping tissues and reduces false negatives (Mall et al., 2017).

**Digital Breast Tomosynthesis (DBT)**, an innovative 3D mammography technology, captures layered views of breast tissue, facilitating the differentiation of benign and malignant structures (Dhungel et al., 2015). Recent studies indicate that DBT offers heightened sensitivity and specificity compared to traditional mammography, especially for patients with dense breast tissue, thereby enabling the detection of early-stage malignancies.



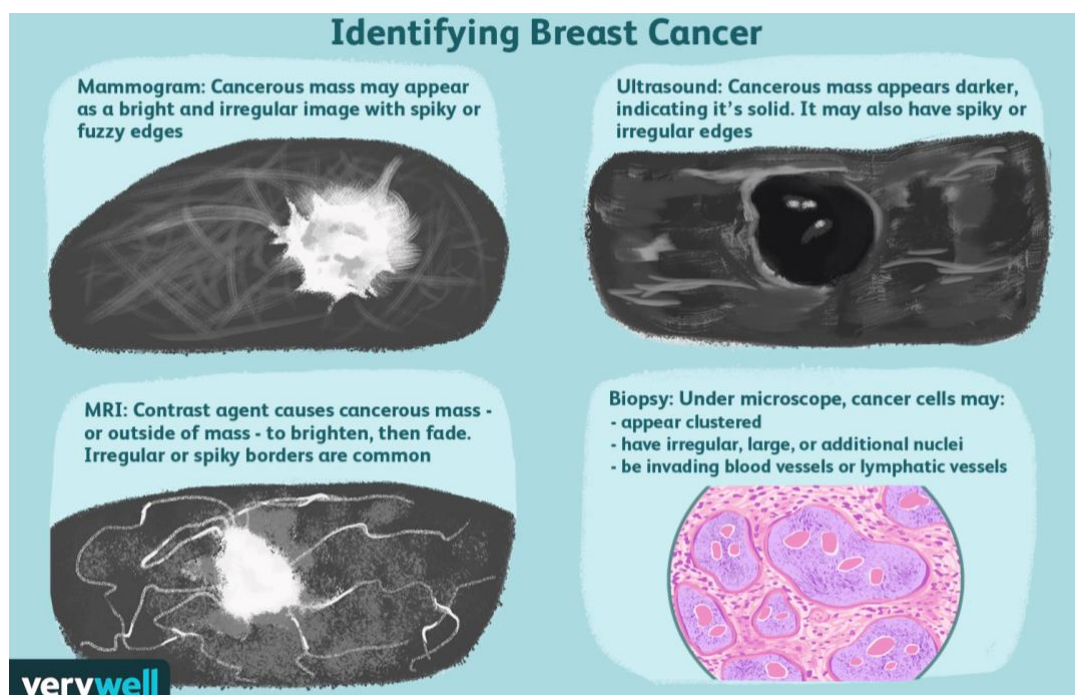
**Figure 2: Working of tomosynthesis. The x-ray tube rotates around the breasts which is compressed between the detector and the compression and thus takes the imaging from all angles**

In addition to mammography, supplementary imaging techniques such as **ultrasound** and **Magnetic Resonance Imaging (MRI)** have become indispensable in breast cancer diagnosis. Ultrasound, which uses high-frequency sound waves to create breast images without ionizing radiation, is particularly useful for dense breast tissue and for screening younger women at elevated risk (Jiang et al., 2024). However, ultrasound's dependency on the operator can affect the consistency of results, and it generally lacks the specificity required for independent screening. MRI, with its high sensitivity for detecting minute tissue

abnormalities, has proven efficacious in identifying invasive breast cancers in women with genetic predispositions, providing detailed visualizations of soft tissues (Alotaibi et al., 2023). However, MRI's high cost and limited accessibility constrain its widespread application in general population screening, making it more appropriate for high-risk individuals.

<i>Techniques</i>	<i>Sensitivity</i>	<i>Tumor size corresponding to sensitivity</i>	<i>Advantages</i>	<i>Disadvantages</i>
<b>Mammography</b>	85	$\leq 2\text{cm}$	Improved image resolution, widely available	Limited sensitivity in dense breast tissue, exposure to radiation
<b>Ultrasound</b>	82	2cm	No ionizing radiation, suitable for dense breasts and implant imaging	Operator-dependent, limited specificity
<b>MRI</b>	95	$< 2\text{cm}$	images small details of soft tissues	Expensive
<b>Diffused Optical Tomography</b>	92	1cm	images small details of soft tissues	Ill-posed problem during reconstruction

*Table 1: Comparison of Imaging Techniques for Tumor Detection.*

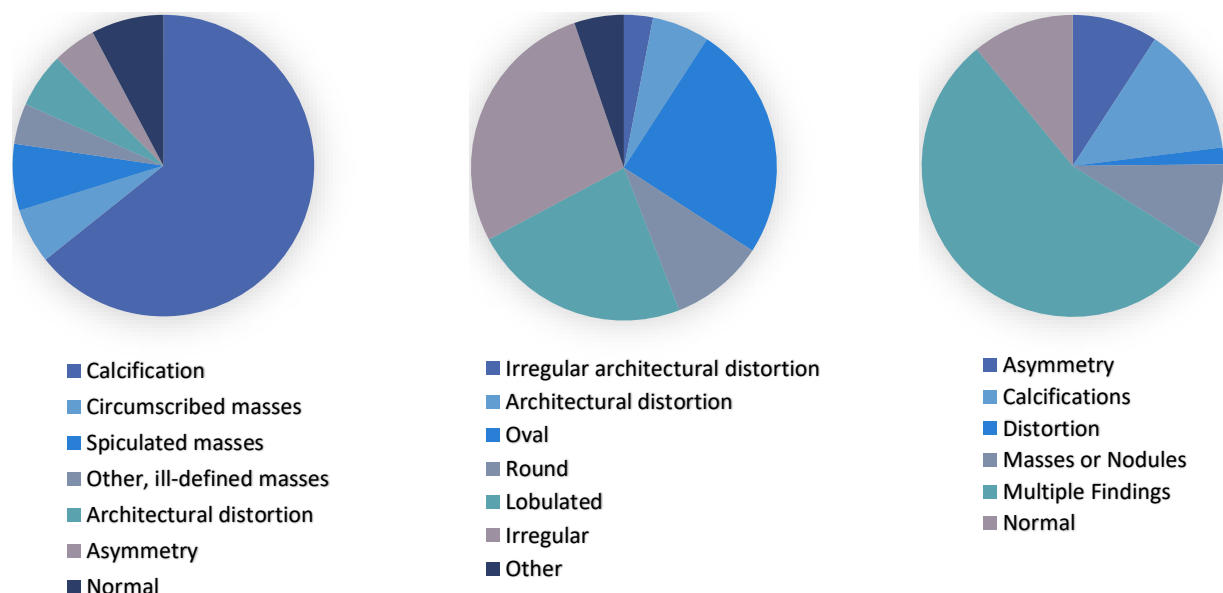


*Figure 3: Contrast between various techniques (Davis, L., 2022)*



The availability of extensive, well-annotated mammography datasets has been instrumental in advancing breast cancer diagnostics, particularly in fostering artificial intelligence (AI) and machine learning-based methodologies. The Mammographic Image Analysis Society (MIAS) dataset, established in 1994, contains labeled mammograms of both benign and malignant lesions, serving as an early cornerstone for automated breast cancer detection research (Suckling et al., 1994). The Digital Database for Screening Mammography (DDSM), developed in the early 2000s, comprises thousands of mammogram images annotated by experts, significantly aiding the development of computer-aided detection (CAD) systems (Heath et al., 2001). CAD systems trained on these datasets have demonstrated notable improvements in detecting microcalcifications and masses, expediting diagnosis and enhancing accuracy.

More contemporary datasets, such as INbreast (2012), provide high-quality full-field digital mammography (FFDM) images, reflecting current clinical standards and proving instrumental for training convolutional neural networks (CNNs) and other machine learning models in breast cancer detection (Moreira et al., 2012). The INbreast dataset offers detailed annotations on breast density, mass shape, and calcifications, facilitating the training of models to recognize nuanced cancer indicators. CBIS-DDSM (Curated Breast Imaging Subset of DDSM), developed through the National Cancer Institute's Cancer Imaging Archive, includes digital images and has become a valuable resource for training advanced deep learning models (Lee et al., 2017). The availability of these comprehensive datasets has enabled further innovations in machine learning, thereby improving detection rates and reducing false positives.



**Figure 4: Types of lesions present in MIAS, CBIS-DDSM and INbreast datasets. (left to right). In the MIAS dataset major type of lesion is calcification. The DDSM dataset contains tumors of irregular shape while in INbreast the major mammograms belong to calcifications**

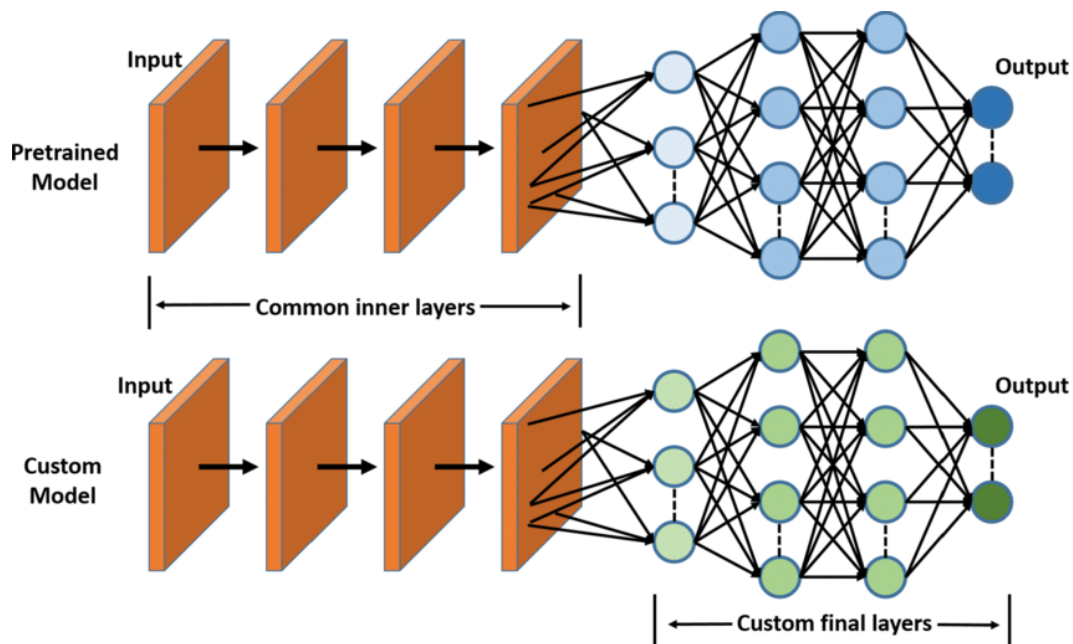


Computer-aided detection (CAD) systems have become essential in breast cancer diagnosis by alleviating radiologists' workload and enhancing diagnostic accuracy. Typically, CAD systems involve stages of image preprocessing, feature extraction, feature selection, and classification (Godavarty et al., 2015). Image preprocessing techniques improve image clarity, allowing for better visualization of potential tumor regions by minimizing noise (Singh et al., 2020). The detection of regions of interest (ROIs), often achieved through techniques like fuzzy logic and clustering, is critical in identifying suspicious areas within images for further analysis (Wang et al., 2019). Recent advancements in image preprocessing and feature extraction methods have allowed CAD systems to analyze images with higher precision, resulting in more accurate identification of abnormal tissue structures.

**Feature extraction** in CAD systems captures key image characteristics, such as *texture, shape, and intensity*, enabling classifiers to distinguish between benign and malignant regions. Traditional machine learning algorithms, like support vector machines (SVMs) and decision trees, have achieved high classification accuracy in identifying malignant lesions (Dhungel et al., 2017). However, these approaches are constrained by their reliance on manually engineered features, which may not fully capture the complexity of mammographic images. Modern CAD systems harness deep learning to automate feature extraction and classification, enhancing diagnostic accuracy and efficiency (Zhu et al., 2019).

Deep learning, especially through convolutional neural networks (CNNs), has revolutionized breast cancer diagnostics by significantly improving feature extraction and classification capabilities. CNNs utilize large datasets to learn hierarchical representations of medical images, refining their capacity to identify complex patterns indicative of cancer. Advanced CNN architectures, such as ResNet, DenseNet, and Inception, have further enhanced diagnostic accuracy, as these models use residual connections and multi-scale feature extraction to identify fine-grained details within images (Litjens et al., 2017). Studies by Esteva et al. (2019) have demonstrated that CNN-based models can achieve accuracy levels comparable to radiologists in identifying malignancies, a promising development for breast cancer diagnostics.

Transfer learning, a method in which pre-trained models are adapted for new tasks, has proven especially beneficial in medical imaging, where labeled data is scarce. Transfer learning enables researchers to use models trained on extensive general datasets, adapting them for specific tasks like breast cancer detection (Fletcher and Elmore, 2013). By leveraging transfer learning, deep learning models achieve enhanced accuracy with less annotated data, making this approach particularly valuable in medical applications.



**Figure 5: Diagrammatic view of working of Transfer Learning (Gilik et al, 2022). The initial layers of the pretrained models are retained but the last layers, are modified according to the use case**

Optimization of hyperparameters, such as learning rate and batch size, is also crucial for fine-tuning deep learning models to improve diagnostic accuracy in breast cancer detection. Studies have indicated that careful hyperparameter tuning can significantly increase model performance, as demonstrated by (Litjens et al., 2017). Recent research has also focused on feature reduction techniques, which streamline model efficiency by discarding redundant features, thereby improving classification precision (Wellings, Vassiliades and Abdalla, 2016). Incorporating feature selection algorithms and dialectical frameworks into deep learning further enhances robustness and accuracy.

To evaluate AI model performance in breast cancer detection, various metrics are used, including the Dice Similarity Index, Jaccard Coefficient, and F1 Score. These metrics provide critical insights into model accuracy, reliability, and predictive power. Studies employing advanced models like Mask R-CNN and MS-ResCU-Net have achieved high sensitivity and specificity in segmentation tasks, indicating the potential for these systems in clinical applications (Global Burden of Disease Cancer Collaboration et al., 2015). However, real-world challenges persist, including model interpretability, patient-specific variability, and the need for further validation across diverse patient populations.

Gradient-weighted Class Activation Mapping (Grad-CAM) plays a crucial role in explainable AI (XAI), particularly in medical imaging, where it helps identify regions within an image that significantly influence a model's predictions. This capability enhances the interpretability of deep learning models, allowing healthcare professionals to trust and better understand AI-driven diagnostic tools. Grad-CAM has shown remarkable promise in breast cancer detection. Talaat et al. (2024) demonstrated its integration with a 3D Inception-ResNet V2 model, achieving a high accuracy of 98.53% and an AUROC of 0.9933, while effectively

highlighting malignant areas in mammograms. Wang, 2024 showcased Grad-CAM's ability to lower false-positive rates in breast cancer diagnostics, particularly in mammography and ultrasound imaging. Despite these advancements, challenges persist, such as variability in localizing malignant regions across different datasets and the limited application of hybrid modeling approaches in XAI.

Our project addresses these limitations by combining Grad-CAM with a hybrid architecture that integrates DenseNet121 and ResNet50. This strategy leverages the complementary strengths of both models for more robust feature extraction and accurate classification of tumors.

<i>Year</i>	<i>Name</i>	<i>Authors</i>	<i>Evaluation Dataset</i>	<i>Noise Removal Method</i>	<i>Performance Metrics (Results)</i>
<b>2015</b>	CRF	Wentao Zhu, Xiaohui Xie	INbreast, DDSM-BCRP	N/A	89% Dice Index
<b>2017</b>	Mammogram Classification Using Support Vector Machine	Youssef Ben Youssef, Elhassane Abdelmounim, Belaguid Abdelaziz	INbreast	N/A	Accuracy: 92%
<b>2018</b>	Adversarial Deep Structural Networks for Mammographic Mass Segmentation	Wentao Zhu, Xiaohui Xie	INbreast, DDSM-BCRP	N/A	97.0% segmentation rate
<b>2020</b>	Mammographic CAD based on pseudocolor and Mask RCNN	Hang Min, Devin Wilson, Yinhuang Huang	INbreast	Morphological filters	Dice Similarity Index: 0.88, True Positive Rate (TPR): 0.90
<b>2021</b>	Evaluation of U-net-based Image Segmentation Model to Digital Mammography	Priscilla Cho, Hong-Jun Yoon	CBIS-DDSM, INbreast	Adaptive median filter	Dice Coefficient (DC): 0.80, 10-fold cross-validation

<b>2021</b>	DenseNet for Breast Tumor Classification in Mammographic Images	Yuliana Jiménez Gaona, María José Rodríguez-Alvarez, Hector Espinó Morató	BCDR	N/A	Sensitivity: 99%, Specificity: 94%, Accuracy: 97.7%, AUC: 97%
<b>2024</b>	YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms	Ghada Hamed Aly, Mohammed Marey, Safaa Amin El-Sayed, Mohamed Fahmy Tolba	INbreast	N/A	Detection rate: 98.96%, MCC: 97.62%, F1 Score: 99.24%
<b>2024</b>	Grad-CAM Enabled Breast Cancer Classification with a 3D Inception-ResNet V2: Empowering Radiologists with Explainable Insights	Fatma M. Talaat ,Samah A. Gamel ,Rana Mohamed El-Balka, Mohamed Shehata and Hanaa ZainEldin	CBIS-DDSM	NA	Accuracy: 98.53% Recall:0.98 Precision: 0.984 AUROC Score: 99.33

**Table 2: Performance Comparison of Various Deep Learning Models for Tumor Detection and Classification in Mammography Datasets**

The integration of advanced imaging technologies and AI has catalyzed significant improvements in breast cancer detection and diagnostics. While mammography remains widely used, its limitations in dense breast tissue underscore the necessity for multimodal imaging approaches. Emerging techniques, such as DBT, MRI, and DOT, show promise for more comprehensive and accurate diagnostics, especially when combined with AI methodologies.

Hybrid models have emerged as a promising approach in mammography classification, effectively combining multiple deep learning techniques to capture diverse features and enhance generalization. For example, a hybrid model that integrates DenseNet<sup>121</sup> with an enhanced deep randomized vector functional link (edRVFL) network, has demonstrated remarkable performance in breast cancer detection. (Qasrawi et al., 2024, ). These preprocessing methods significantly improve diagnostic accuracy by enhancing image quality and detail extraction. The inclusion of Grad-CAM analysis further enhances interpretability by providing visual explanations of the model's decisions, making it more comprehensible for healthcare professionals. This hybrid approach achieved a clinical

validation accuracy of 98.6% with a processing time of 0.75 seconds, outperforming many existing methods and underscoring its potential for reliable and efficient breast cancer detection.

AI, and particularly deep learning, has shown substantial promise in enhancing diagnostic accuracy and minimizing human error in breast cancer detection. Nonetheless, challenges such as data quality, class imbalance, and interpretability remain. Further research may focus on developing hybrid models that integrate AI with conventional imaging and personalized medical data, paving the way for more precise, individualized screening protocols. Future advancements in dataset quality, model interpretability, and interdisciplinary collaboration will be essential to unlocking AI's full potential in improving breast cancer outcomes and survival rates.

For this project, I selected ResNet50 and DenseNet121 due to their proven effectiveness in capturing complex image features and their ability to mitigate issues like vanishing gradients. ResNet50, with its residual learning framework, enables the model to train deeper networks without performance degradation, making it particularly well-suited for extracting hierarchical features from mammography images. The architecture's skip connections facilitate the flow of gradients, enhancing training stability and accuracy. On the other hand, DenseNet121, with its dense connections between layers, ensures that each layer receives feature maps from all previous layers, promoting feature reuse and reducing the number of parameters.

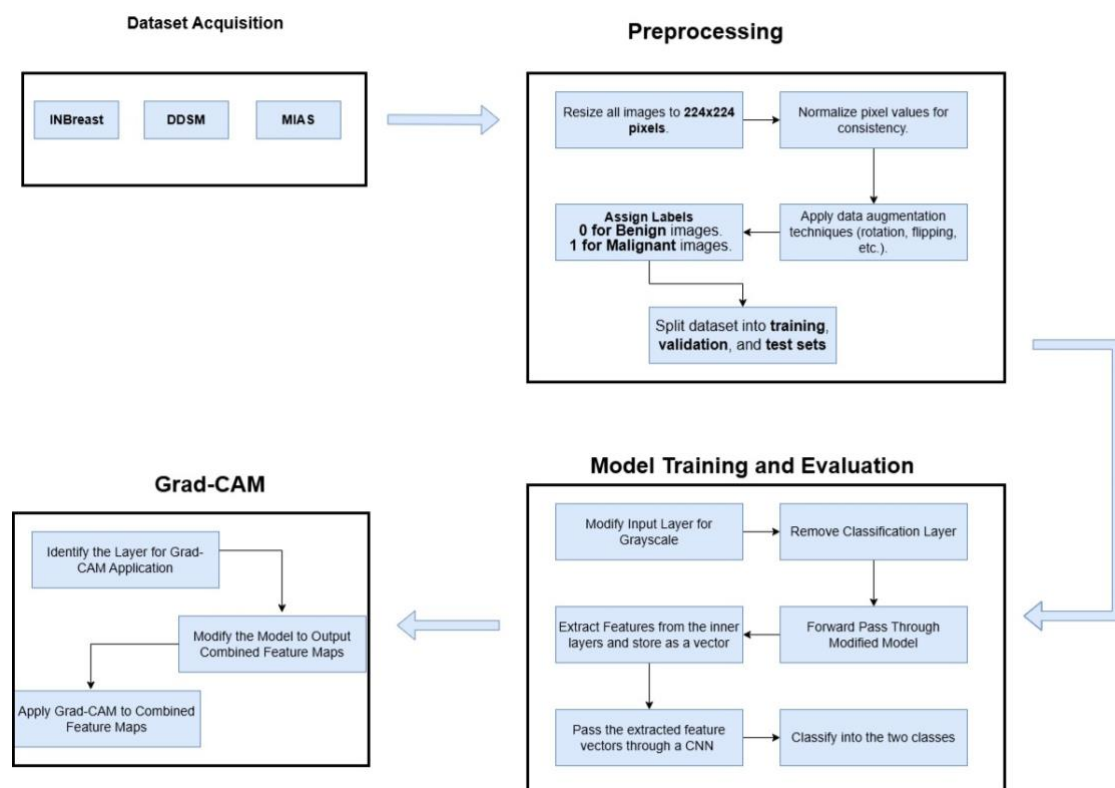
## 3. Methodology (Mammography Classification)

### 3.1. Overview

In this project, the goal is to create a hybrid deep learning model that can classify mammography images as either benign or malignant. To do this, we use two convolutional neural networks: DenseNet121 and ResNet50. DenseNet121 is good at reusing features from earlier layers, while ResNet50 uses residual connections to solve the problem of vanishing gradients. We take the features from both models and combine them into one, which helps capture more detailed patterns from the images and improves the accuracy of the classification.

The combined features are then passed through fully connected layers to perform binary classification. To make the model more interpretable, we use Gradient-weighted Class Activation Mapping (Grad-CAM), which creates heatmaps showing which parts of the image are most important for the model's decision. This helps us understand how the model is making its predictions. Overall, the methodology covers everything from preparing the data, extracting features, training the model, and evaluating its performance, all while ensuring the results are both accurate and easy to interpret.

### 3.2. Methodology Flowchart



**Figure 6: The summary of AI workflow carried out in this study which consists of acquiring the dataset, applying various preprocessing methods on the dataset, training a hybrid model through transfer learning, then evaluating the model and visualizing its decisions via Grad-CAM.**

### 3.3. Setting up Co-lab Environment

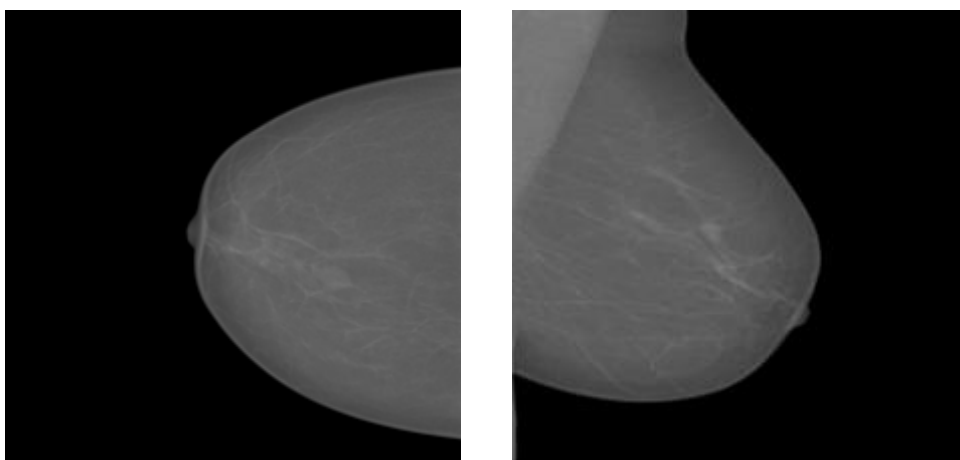
To begin the project, I first set up a Google Colab environment, which provides an easy-to-use platform with free access to GPUs, making it ideal for deep learning tasks. After opening a new Colab notebook, I mounted Google Drive to access the datasets stored there. Once the drive was successfully mounted, I navigated to the specific directory containing the mammography dataset. The dataset is organized into two main subfolders: "Benign Masses" and "Malignant Masses," each containing the respective images. I used the `os` module in Python to load the paths to these folders, enabling me to load the images and labels for further processing. With the environment set up and the data loaded, I was ready to move on to the next steps in the project, including data preprocessing and model training.

### 3.4. Dataset

This project uses the Mendeley Mammogram dataset, which combines images from three major sources: INbreast, MIAS (Mammographic Image Analysis Society), and DDSM (Digital Database for Screening Mammography). The dataset is used to train a deep learning model that classifies mammography images as benign or malignant, which aids in the early identification of breast cancer. The combination of these three datasets yields a broad and well-annotated collection of photos, each with unique attributes and characteristics, ideal for training robust models.

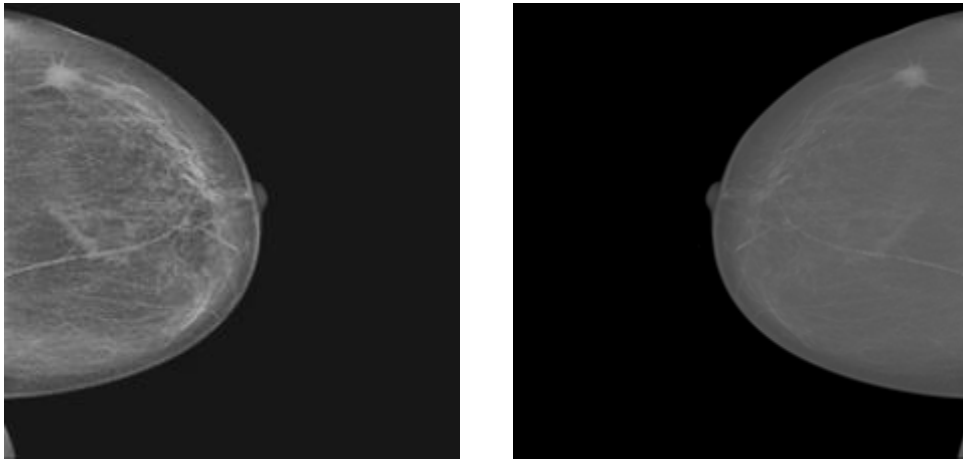
#### 3.4.1. INbreast Dataset:

The INbreast dataset is a collection of high-quality digital mammogram images from 41 patients, containing both benign and malignant cases. Each image is labeled with relevant annotations, such as the presence of masses or microcalcifications. Additionally, the dataset provides details about breast density, which can influence the ability to detect abnormalities. These annotations allow the dataset to be used for tasks involving the identification of different types of breast tissue patterns, which is crucial for accurate classification.



*Figure 7: Benign sample images (INbreast)*

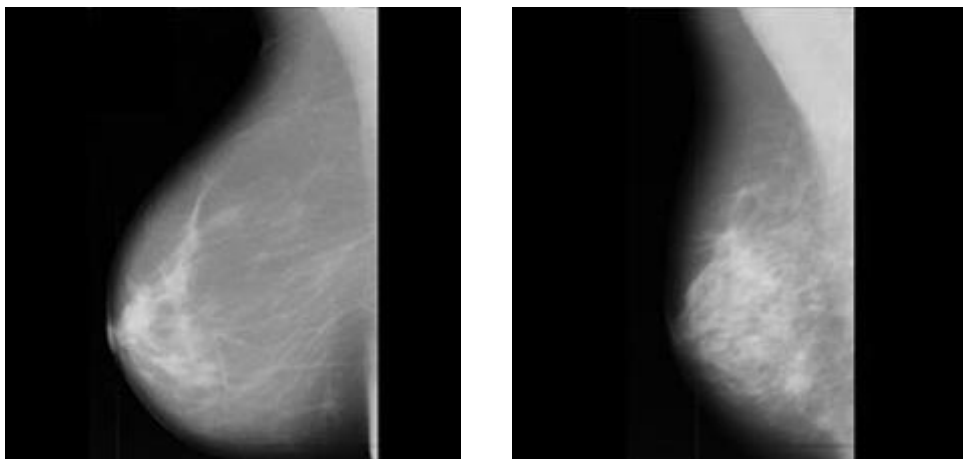




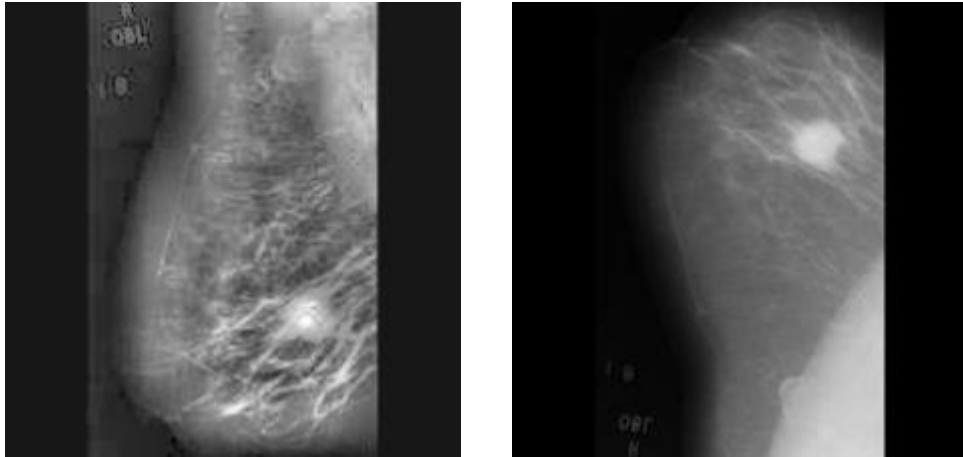
*[Figure 8: Malignant sample images \(INbreast\)](#)*

### 3.4.2. MIAS Dataset:

The MIAS dataset contains 322 mammogram images from 161 patients. These images are annotated with different types of abnormalities, such as masses, calcifications, and asymmetries, providing essential information for training classification models. One of the unique features of the MIAS dataset is its variety in image resolution, making it ideal for testing models that need to handle different qualities and sizes of mammograms. This helps ensure that the model is robust across varied input data.



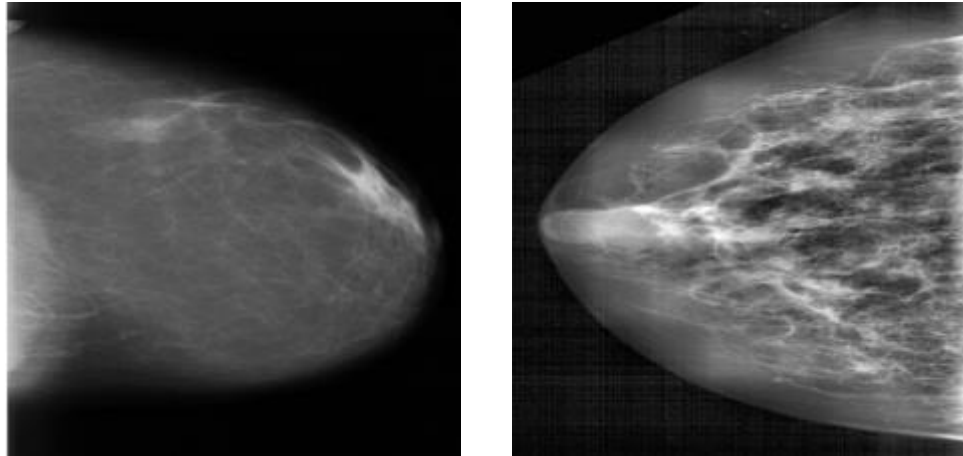
*[Figure 9: Benign sample images \(MIAS\)](#)*



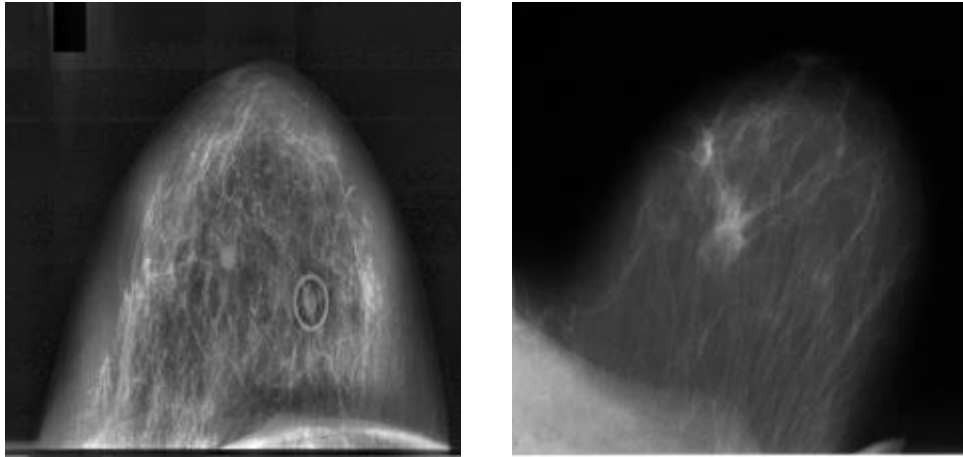
*[Figure 10: Malignant sample images \(MIAS\)](#)*

### 3.4.3. DDSM Dataset:

The DDSM dataset includes over 500 mammogram images from a diverse set of patients. It contains both standard and digitized film mammograms, offering a wide variety of cases to help train the model on different image types. The dataset is extensively annotated with information on masses, asymmetries, and calcifications, which are key features for breast cancer detection. The DDSM is one of the largest and most widely used mammography datasets, providing valuable data for developing deep learning models that require large and varied datasets for training.



*[Figure 11: Benign sample images \(DDSM\)](#)*



***Figure 12: Malignant sample images (DDSM)***

<b><i>Attributes</i></b>	<b>INbreast</b>	<b>CBIS-DDSM</b>	<b>DDSM</b>
<b><i>Total Images</i></b>	410	3,000	10,480 images
<b><i>Image Resolution</i></b>	High (up to 3328 × 4084 pixels)	Medium to High (varies, generally high for digitized)	Low to Medium (varies significantly)
<b><i>Annotations</i></b>	Yes (masks, detailed ROI, pathology)	Yes (detailed ROI, pathology, bounding boxes)	Limited (some ROI annotations, varies)
<b><i>Classes</i></b>	Benign, Malignant	Benign, Malignant	Benign, Malignant
<b><i>File Format</i></b>	DICOM	DICOM	LJPEG (digitized versions in DICOM available)
<b><i>Dataset Focus</i></b>	Digital mammography	Digitized film mammography	Screen-film mammography (digitized)
<b><i>Availability</i></b>	Restricted (license required)	Public (via TCIA, with approval)	Public (with request, through TCIA or collaborators)
<b><i>Region of Interest (ROI)</i></b>	Yes, detailed lesion annotations	Yes, detailed annotations	Limited or absent, depending on version

<b>Augmentation Provided</b>	No	No	No
<b>Strengths</b>	High-quality images, consistent format	Large size, well-annotated for ML tasks	Large size, useful for deep learning studies
<b>Challenges</b>	Small dataset size	Preprocessing required (e.g., ROI extraction, normalization)	Inconsistent quality and preprocessing challenges

**Table 3: Comparison of Mammography Datasets: INbreast, CBIS-DDSM, and DDSM. INbreast dataset has the highest quality and consistent images but very small in size. DDSM data has the most screening mammographs but is projected inconsistent quality. The MIAS dataset is well-annotated for ML algorithms, but they require some normalization and ROI extraction methods to be applied**

## 3.5. Preprocessing

The dataset is organized into two main categories: Benign and Malignant. These labels indicate the presence or absence of cancerous growths in the breast tissue. The images are primarily grayscale, with pixel values representing the intensity of the tissue. Before training the model, the dataset needs to undergo several preprocessing steps to ensure it is in the right format.

### 3.5.1. Resizing

First, the images are resized to a uniform size (224x224). Resizing is a crucial preprocessing step to ensure that all images in the dataset have the same dimensions. Since deep learning models like ResNet50 and DenseNet121 require input images of a fixed size, resizing all images to a standard dimension (such as 224x224 pixels) ensures compatibility with these models. Resizing also helps reduce computational load by standardizing the input, which allows the model to process the data more efficiently. This step is essential because images in the dataset might come in varying resolutions, and having consistent image dimensions improves the stability and performance of the model during training.

### 3.5.2. Normalizing

Normalization is another important step in the data preprocessing pipeline. It involves scaling the pixel values of images to a *range of 0 to 1* by dividing the pixel values by 255 (since pixel values typically range from 0 to 255 in an 8-bit image). This helps the model learn more effectively by standardizing the input features and preventing large variations in pixel intensity from interfering with the learning process. Normalization also speeds up convergence during training by ensuring that the gradient updates during backpropagation are consistent across all input features, which ultimately helps the model to converge faster and achieve better performance.

### 3.5.3. Data Augmentation

Data augmentation is a powerful technique used to artificially expand the training dataset by applying random transformations to the original images. This helps improve the model's generalization capability by exposing it to a wider variety of image variations, preventing overfitting. In the case of the Mendeley Mammogram dataset, augmentation techniques such as *rotation*, *zooming*, and *flipping* are commonly applied. These transformations simulate different perspectives and conditions, making the model more robust when it encounters new, unseen images.

## 3.6. Labelling and storing the images

After preprocessing, the mammography images were saved as NumPy arrays to simplify storage and retrieval, allowing for more efficient data handling during model training. By transforming the photos to NumPy arrays, I ensured that the data could be easily retrieved and loaded, eliminating the need to repeat preprocessing processes each time the model was trained. This strategy dramatically increased the workflow's overall efficiency because NumPy arrays are fast and lightweight when processing massive datasets. Labels were assigned to each image based on its class, with '0' indicating benign masses and '1' indicating malignant masses. This label assignment helped to categorize the photos, which is necessary for training a supervised classification model.

After labelling the images, the images and their corresponding labels were saved as two separate NumPy arrays—one for the images and one for the labels. These arrays were then uploaded to Google Drive, which ensured that the data would be saved even if the session ended and made it easy to access when needed. Storing the data this way also made it quicker to load the images and labels during training, so we didn't have to process them again. Plus, saving the data as NumPy arrays helped with memory management, making it easier to load the entire dataset into memory when training and testing the model.

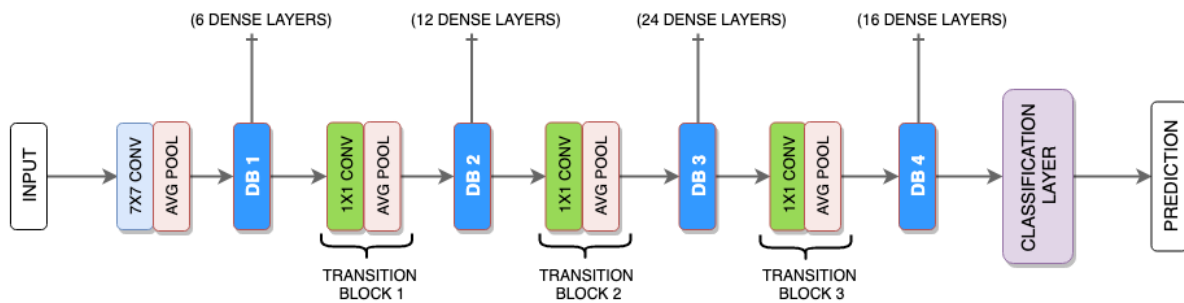
## 3.7. Model Training

Following the preprocessing, the mammography images were processed using convolutional neural networks (CNNs) to extract and classify features. We used the preprocessed dataset, which was scaled, normalized, and saved as numpy arrays, to make it easier to put into models. For this, we used two well-known models: DenseNet121 and ResNet50. These models were chosen because they are exceptionally good at detecting patterns and features in medical images. After extracting features from both models, we integrated them into a single set of features that allowed us to classify the photos as benign or malignant.

### 3.7.1. Densenet121

DenseNet121 is a convolutional neural network known for its innovative architecture that connects each layer to every other layer in a feed-forward manner. This unique structure, called dense connectivity, allows the model to pass information and gradients directly across layers, which significantly reduces the risk of vanishing gradients, a common challenge in deep networks. By reusing features throughout the network, DenseNet121 minimizes redundancy and improves efficiency. This characteristic makes it particularly useful for medical image analysis, where fine details and subtle patterns in images, such as masses or microcalcifications in mammograms, play a critical role in diagnosis. Its ability to

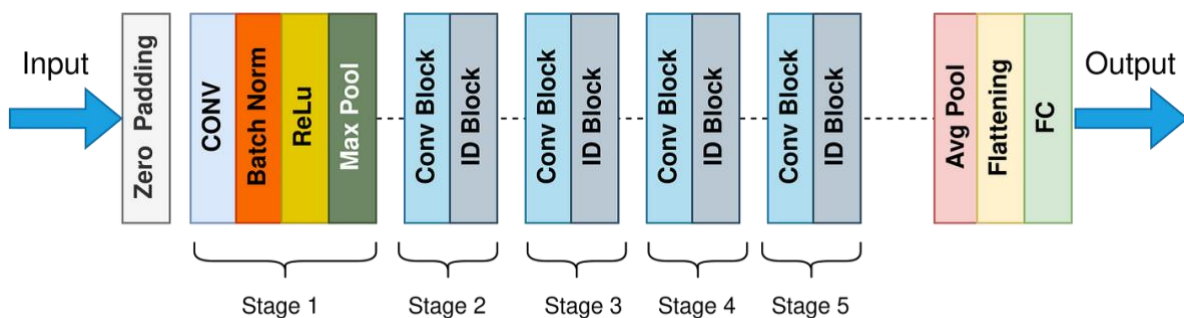
focus on intricate image features while maintaining computational efficiency is why it is a strong choice for this project.



**Figure 13: Densenet121 architecture.** It consists of 5 blocks of dense layers each containing 6, 12, 24, 16 dense layers respectively. It also has three transition blocks which are combination of convolution layer and average pooling.

### 3.7.2. ResNet50

ResNet50, on the other hand, is a highly effective model based on the residual learning principle. Training can get difficult in deeper networks as gradients decrease, leading earlier layers to learn slower. ResNet50 addresses this issue by creating shortcut connections that skip specific layers, allowing the network to focus on learning the residuals, which are the differences between the present and desired outputs. This approach enables the model to learn deeper and more complex properties while maintaining efficiency. In mammography classification, ResNet50's capacity to identify high-level patterns, such as the overall structure of masses or asymmetries in breast tissue, complements DenseNet121's granular emphasis.

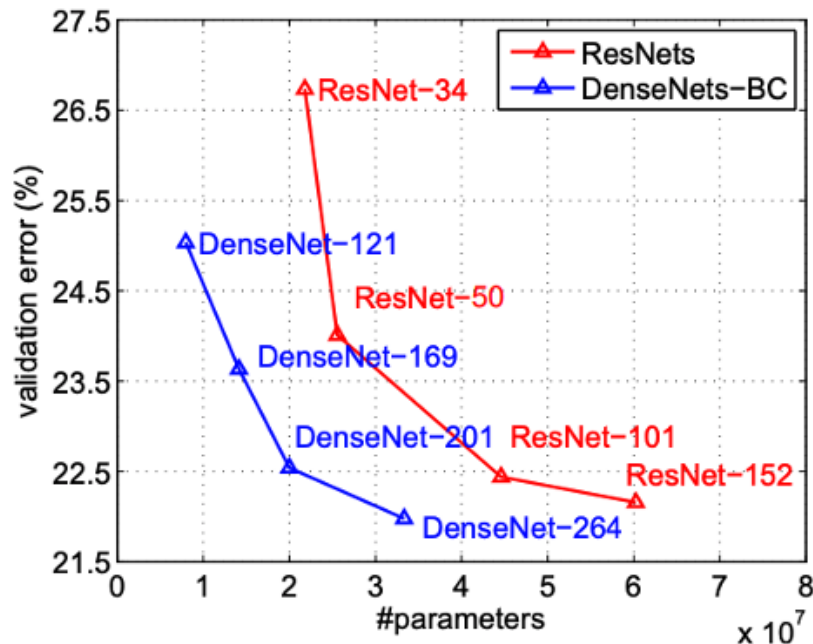


**Figure 14: ResNet50 architecture (Mukherjee, 2022)** containing a conv layers, batch normalization layers, ReLu activation followed by max pooling. Then comes the 4 residual blocks which are the pair of conv block and id block. In the end it also has average pooling layer, and flattening layer and fully connected layer.

### 3.7.3. Hybrid Model Approach

We combine DenseNet121 and ResNet50 to make use of their respective capabilities, resulting in a more successful hybrid strategy. DenseNet121 excels at capturing fine-grained, low-level features required for medical imaging, such as subtle details in masses or microcalcifications in mammograms. ResNet50, on the other hand, is particularly good at extracting high-level, abstract properties like the overall shape or asymmetry of breast tissue. By combining the features extracted from these two models, we may obtain a more complete description of the data. This hybrid approach enables us to assess both local and

global patterns in mammography images, yielding a model that is better able to distinguish between benign and malignant instances. The fusion of these feature sets not only enhances the classification accuracy but also ensures a more reliable diagnostic performance, making it a valuable method for addressing the challenges of medical image analysis.



**Figure 15: Validation error vs. No of parameters of DenNet121 and ResNet50. DenseNet121 has lower validation error as compared to ResNet50 for a given no of parameters hence DenseNet121 has better performance as it requires lesser computation expense.**

In our project, we aimed to utilize the pre-trained DenseNet121 model for classifying mammogram images. DenseNet121 is a deep convolutional neural network typically trained on color images with three channels (RGB), but our dataset consists of single-channel grayscale images (mammograms). To use this pre-trained model, we needed to modify its input layer to handle single-channel images, while still leveraging the model's pre-trained weights.

Before passing the images through the model, the first step was to preprocess the mammogram images. We resized all images to a consistent size of 224x224 pixels, which is the input size expected by DenseNet121. Additionally, the pixel values were normalized by dividing them by 255.0, which scales the pixel values from the original range of 0 to 255 to a normalized range between 0 and 1. This normalization step is essential for faster model convergence during training, as it ensures that all features are on a similar scale.

### 3.7.4. Tweaking the models

DenseNet121, by default, expects an input shape of (224, 224, 3) to process three-channel RGB images. However, since our dataset consists of grayscale images with only a single channel, we needed to adjust the model to accept this format. Directly changing the input shape to (224, 224, 1) would cause issues because DenseNet121 is designed to work with three channels. To address this, we added a custom convolutional layer at the beginning of the model. This layer replicates the single-channel input into a three-channel format,



effectively converting the grayscale images into an artificial RGB format. The convolution layer used a 3x3 kernel and three filters, ensuring the grayscale data is expanded to three channels while keeping the original information intact.

After modifying the input layer, we loaded the pre-trained DenseNet121 model using the *Keras DenseNet121 function*, with the *include\_top=False* argument. This removed the final classification layers of the model, as they were originally designed for a different task. The goal was to use DenseNet121 as a feature extractor and add our own custom classification layers for our specific binary classification problem: identifying whether a mammogram image is benign or malignant. We used the pre-trained weights from ImageNet to initialize the model's convolutional layers, which helps the model detect general patterns in images like edges, textures, and shapes.

To adapt the model to our task, we added a *Global Average Pooling (GAP)* layer after the feature extraction layers. The GAP layer condenses the output from the DenseNet121 model into a single vector, representing the most important features from the entire image. Following this, we added a dense layer with a sigmoid activation function to predict a probability value between 0 and 1, indicating whether the image is benign or malignant. This architecture is well-suited for binary classification tasks, where the output is a probability representing the likelihood of the image belonging to one of the two classes.

Finally, the modified model was compiled with the *Adam optimizer* and *binary cross-entropy loss*. Adam is an efficient optimizer that adapts the learning rate during training, and binary cross-entropy is the appropriate loss function for binary classification tasks. We also tracked the model's accuracy during training to monitor its performance. This approach allowed us to fine-tune the pre-trained DenseNet121 model for our specific mammogram classification task, benefiting from its ability to capture detailed features while adapting it to the grayscale format of our images.

By making these adjustments, we were able to effectively use DenseNet121 for the task of mammogram classification, leveraging the power of transfer learning while customizing the model to handle single-channel images. These modifications ensured that the model could still recognize and classify subtle patterns in the mammogram images, improving diagnostic accuracy.

Like DenseNet121, ResNet50 typically expects an input shape of (224, 224, 3) for processing three-channel RGB images. However, since our dataset consists of grayscale images with only one channel, we needed to adjust the model accordingly. Directly changing the input shape to (224, 224, 1) would cause a mismatch with the model's architecture, so we used an alternative approach. We added a custom convolutional layer at the start of the model. This layer took the single-channel grayscale input and expanded it to a three-channel format, simulating RGB images. To achieve this, we used a 3x3 kernel with three filters, allowing the model to process the grayscale data while maintaining the original information.

Once the input layer was modified, we loaded the pre-trained ResNet50 model using the *ResNet50 function* from Keras, with the argument *include\_top=False*. This excluded the original classification layers, as they were not suited for our task. Our goal was to utilize

ResNet50 as a feature extractor, taking advantage of its ability to learn complex and high-level features, and then adding custom classification layers for our binary classification problem: distinguishing between benign and malignant mammograms. By using the pre-trained weights from ImageNet, we could leverage general features like edges, textures, and shapes, which are relevant for medical image analysis.

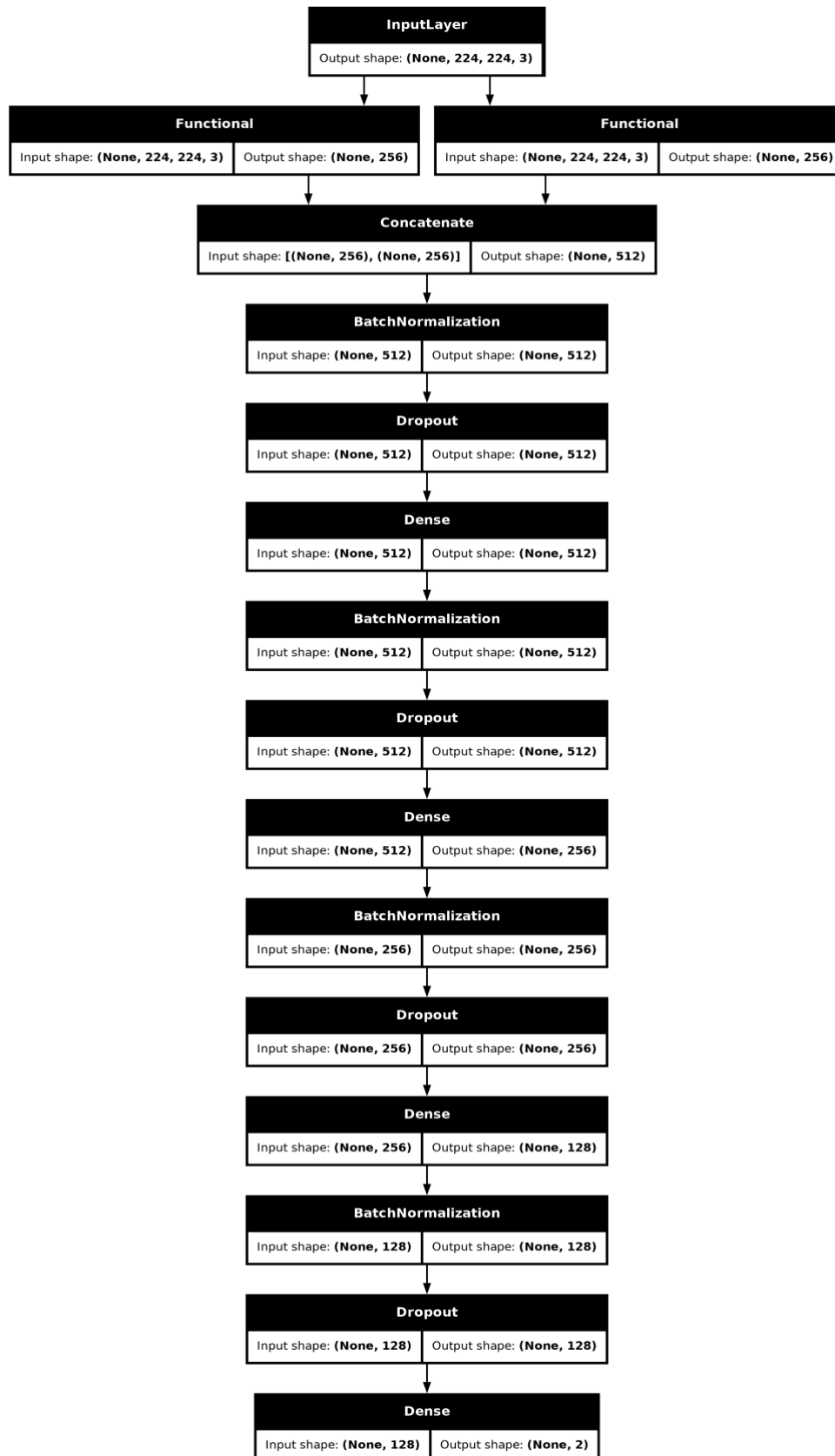
To adapt ResNet50 for our specific task, we inserted a Global Average Pooling (GAP) layer after the feature extraction layers. The GAP layer aggregates the feature maps into a single vector, representing the most important information from the image. After this, we added a fully connected dense layer with a sigmoid activation function, which outputs a probability value between 0 and 1, classifying the image as either benign (0) or malignant (1). This setup is well-suited for binary classification, as the output probability indicates the likelihood of the image belonging to either of the two classes.

Lastly, we compiled the modified model using the Adam optimizer and binary cross-entropy loss. Adam is a popular optimizer known for adjusting the learning rate throughout training, which helps in speeding up convergence and avoiding overfitting. Binary cross-entropy was used as the loss function, which is standard for binary classification problems. We also tracked the model's accuracy during training to evaluate its performance. This combination of transfer learning with fine-tuning allowed us to tailor the pre-trained ResNet50 model to our mammogram classification task, effectively adapting it to handle single-channel images. Through these adjustments, we successfully leveraged ResNet50 for the mammogram classification task. The model was able to extract high-level features from the images while being adapted to handle grayscale input.

After obtaining the features from both ResNet50 and DenseNet121, the next step is to save them for later use in the feature fusion and classification stage. Since these models are used for feature extraction, we will store the output features (which are typically in the form of vectors after the Global Average Pooling layer) in a structured format that can be easily loaded for further processing. The next step is to combine the features extracted from both models. Since the features from ResNet50 and DenseNet121 are both in the form of feature vectors (e.g., each model outputs a vector of size 2048), we can concatenate them to form a combined feature vector that captures information from both models.

Now that we have the combined features, we can use them to train a classification model. Since our task is binary classification (benign vs malignant), a simple neural network or logistic regression model can be used for this task. Here, we'll use a basic *fully connected (dense) neural network model*, which is commonly used for classification tasks after feature extraction.

While training the model I noticed that the validation loss continued to increase sharply causing the model to reach its optimal state too soon, the reason being the high learning rate. To cater for this problem a scheduler was added to model training which reduced the learning rate by half whenever the validation loss would not improve for three epochs. This approach brought significant improvement in accuracy and loss. After training the model, we can evaluate its performance on a separate test dataset or validation set to see how well it generalizes to new data.



*Figure 16: Flowchart representation of the model used in this study. The images are passed on to the DenseNet121 and ResNet50 model (Functional) which give out the feature maps. These features are concatenated into a single vector and then a simple 3- layered NN is used for classification.*

## 3.8. Grad-CAM

Next Step is to implement GRAD-CAM for visualization. A potent visualization method for comprehending and interpreting convolutional neural networks' (CNNs') decision-making process, especially in image classification applications, is Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM gives information on the areas of an image that are most crucial for a model's predictions. It is particularly helpful for comprehending the decision-making process of deep learning models, which are sometimes regarded as "black-box" systems. This openness is crucial in domains such as medical image analysis, where credibility and validation may depend on the capacity to justify predictions.

### 3.8.1. How Grad-CAM works?

Grad-CAM works by analyzing the gradients flowing through the CNN during the prediction process. It calculates the importance of each feature map in the final convolutional layer, and then creates a heatmap showing which areas of the image influenced the model's decision the most. This heatmap can be overlaid on the original image, helping us understand which regions the model used to make its prediction.

Here's a breakdown of the Grad-CAM process:

1. **Forward Pass:** The input image is passed through the CNN model to generate predictions.
2. **Compute Gradients:** After the forward pass, Grad-CAM calculates the gradients of the output for the target class (benign or malignant) with respect to the feature maps of the final convolutional layer. These gradients show how much each feature map contributes to the final prediction.
3. **Global Average Pooling (GAP):** The gradients are averaged across the spatial dimensions to get a set of important weights for each feature map.
4. **Weighted Sum of Feature Maps:** These weights are then used to perform a weighted sum of the feature maps from the final convolutional layer.
5. **Generate Heatmap:** The weighted feature maps are passed through a ReLU activation function, which ensures that only positive importance values are kept. This generates a heatmap showing which regions of the image are most important for the model's decision.
6. **Overlay Heatmap:** Finally, the heatmap is overlaid on the original image, highlighting the regions that influenced the model's decision the most.

### 3.8.2. How Grad-CAM is applied

To apply Grad-CAM to our mammogram classification task, we need to slightly modify the pre-trained ResNet50 and DenseNet121 models, as we are using them for feature extraction and binary classification (benign vs. malignant). Before using Grad-CAM, we need to make sure the models are set up properly. Both ResNet50 and DenseNet121 were fine-tuned to classify mammograms as either benign or malignant. Grad-CAM will be applied to the last convolutional layers that we added after the model features were saved

Here's how we prepare the model for Grad-CAM:

- **Choose the Layer Before Final Classification:** Grad-CAM works by focusing on the last convolutional layer, where high-level features like shapes or textures are captured. Since we are using combined feature vectors from both DenseNet121 and ResNet50, we need to identify the layer just before the final classification layer in our model, which is the point where both feature sets are merged. This will be the layer we apply Grad-CAM to, as it holds the most relevant information for the final decision.
- **Modify the Model to Output Combined Feature Maps:** Since Grad-CAM uses feature maps to generate the heatmap, we need to modify the model so that it outputs the combined feature maps from both DenseNet121 and ResNet50 before they are passed to the final fully connected layers for classification. This way, Grad-CAM can focus on the feature vectors derived from both networks, capturing the most important features before they are classified as benign or malignant. This setup will allow Grad-CAM to highlight the regions of the image that influenced the combined decision from both networks.

Once the Grad-CAM heatmap is generated, it can be overlaid on the original mammogram image. The heatmap shows the areas that contributed the most to the model's prediction. Hotter colors like red and yellow indicate regions of the image that were most influential for the decision. This visualization allows us to see where the model focused its attention when classifying the image.

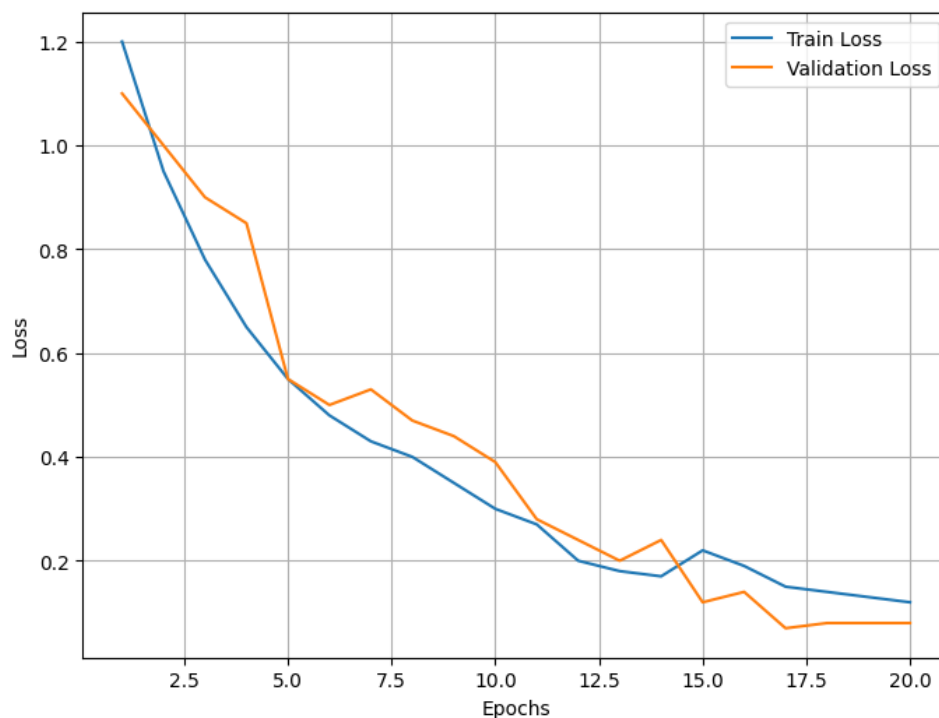
For example, if the model predicts a malignant tumor, the heatmap will likely highlight the suspicious areas (e.g., masses or calcifications) in the mammogram. This helps clinicians understand which regions the model considers important, making the model's decision more interpretable.

## 4. Results

This chapter discusses the outcome of the training and testing of the model established to classify a lump as cancerous or non-cancerous by analyzing the digital mammographs.

### 4.1. Train and validation loss and accuracy

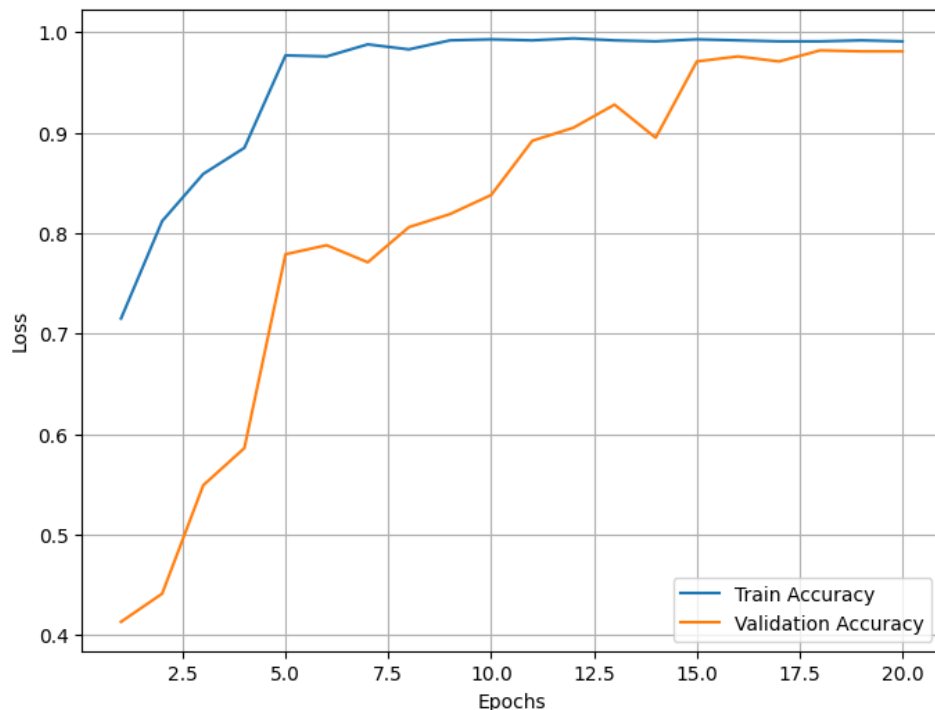
As shown below, the training and validation loss were highest at the start of training and as model progressed its learning the losses decreased at a healthy rate. This proves that the model successfully learned from the training set and also learned to generalize well from the unseen validation split. The training loss decreased from 1.2 to 0.11 while the validation loss dropped from the value of 1.1 to 0.9 in the span of 20 epochs. However, both these trends were fairly proximate during the training process thus indicating a good generalization and absence of overfitting. There can be seen a few sharp peaks in validation graph during epochs 12 to 16 which is normal can occur due to different sampling of validation set and noise in the images.



**Figure 17: Training and Validation Loss during the model's training. This basically explains how well the model is performing on validation data and how well the model is learning from the training data.**

On the contrary if we look at the accuracy trends for both training and validation sets then we can see a similar pattern in terms of the closeness of both lines at the end of the training indicating that the model is performing very well such that the accuracy reaches significantly close to 1. The plot exhibits a positive trend since there is not a major gap between the two accuracies during the course of the fitting of the model since there are no signs of overfitting or underfitting. Again, fluctuations can be seen in the validation accuracy in the same epochs as in validation loss. To improve this several changes can be made to the

model such as using a regularization technique, tuning parameters like batch size, no of epochs or learning rate, using better preprocessing and increasing the size of validation set. Since these results were satisfactory for the objective of my research, no further processing was done to improve the accuracies.



**Figure 18: Train and Validation accuracy.** This graph shows how accurately the model classified the images in train and validation set during the model's training phase.

## 4.2. Test accuracy and loss

Once the evaluation of training process was completed, I moved on to running the model on test data to judge how it is performing. Since this is a classification problem, to evaluate the model's performance on test data, apart from test accuracy and test loss, traditional metrics commonly associated with classification NN were used, such as Recall, Precision and F1 scores. Recall measures the ratio of correct predicted labels of a class to the number of actual labels of that class for example in this case the total samples of benign class in the test data are 1653 and the model classified 1650 images as benign but only 1600

mammographs out of the 1650 were actually benign in the test data, hence  $\text{Recall} = \frac{1600}{1653}$ .

Precision is also calculated for each class that gives information on how many predictions of a class were correct compared to the total predictions of that class by the modal. Again, looking at the results of benign class, according to the model there are 1650 mammographs with benign tumor however in reality only 1600 of those predictions were correct thus

$\text{Precision} = \frac{1600}{1650}$ . By calculating the harmonic mean of recall and precision we get another

statistical metric called F-1 score which combines the precision and recall into a single value ensuring that both are regarded equally. Apart from these, the test accuracy is valued at

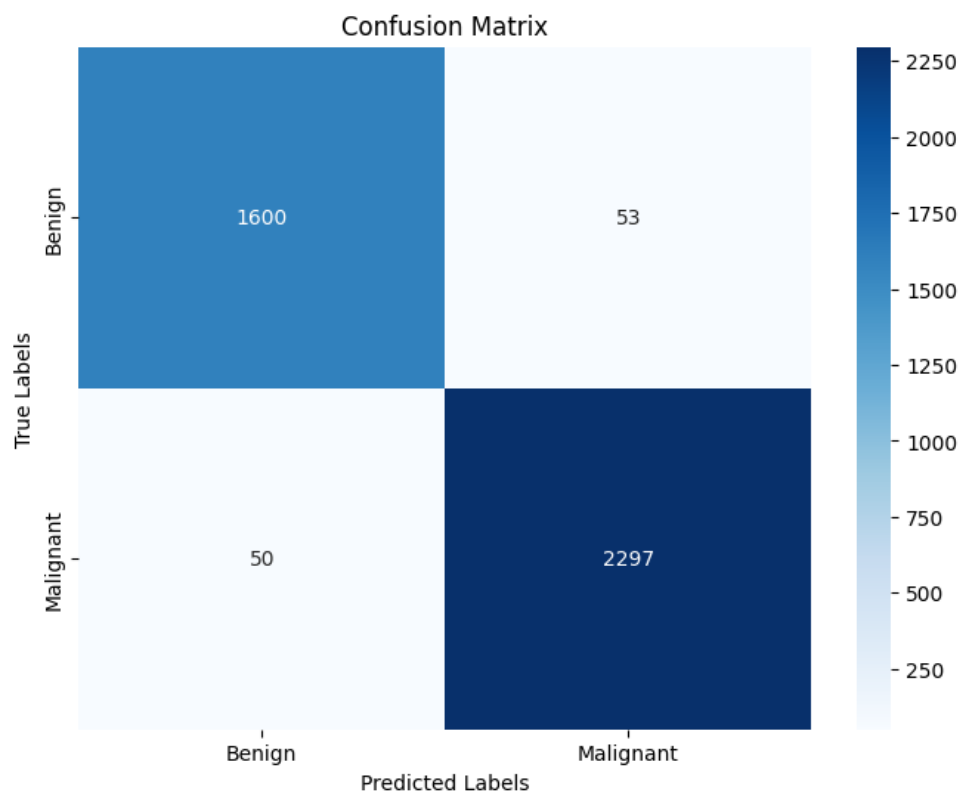


0.9791 in contrast to test loss which is equal to 0.0582 and the overall accuracy is 0.97. These results are summarized in the table below.

	Precision	Recall	F-1 Score	No of samples
Benign	0.97	0.97	0.97	1653
Malignant	0.98	0.98	0.98	23473

**Figure 19: Classification report showing various performance metrics used to evaluate the model's performance on test data. It can be seen that the model's accuracy is close to 1 (100%).**

Similar information on the performance of hybrid model can be visualized using the confusion matrix. It shows the true positive (Were malignant and model predicted malignant), true negative (were benign and model predicted benign as well), false positive (Were benign but model predicted malignant), and false negative (were malignant but model predicted benign) predictions of the algorithm in a colored blocks format.



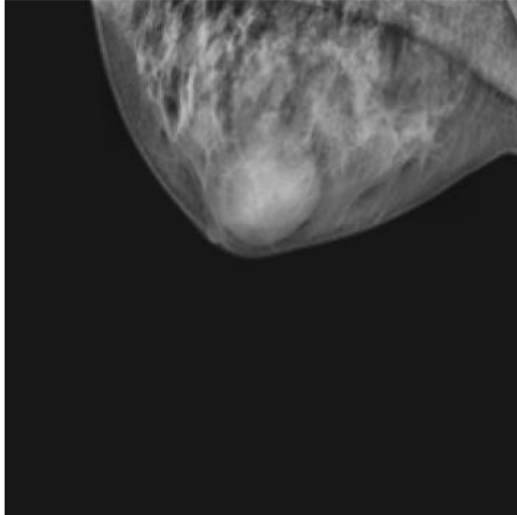
**Figure 20: Distribution of true and predicted labels. 1600 images were correctly predicted benign while 50 were predicted benign by the model but were malignant. 2297 images were correctly guessed by the algorithm as malignant while model made mistake in labeling 53 malignant mammographs as benign.**

### 4.3. Grad-CAM Heatmaps

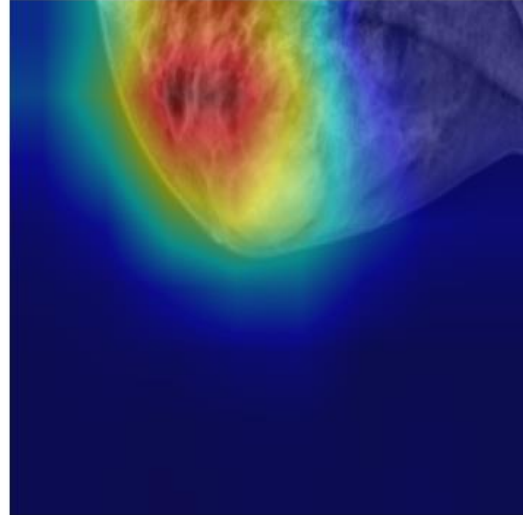
In medical image analysis, it is strictly important to have a system that does not miss a positive diagnosis along with understanding how did the system came to the positive or negative conclusion for it to be a reliable aid for radiologists. As explained in chapter 3.8, the interpretability tool, Grad-CAM allows the engineers to overall understand CNN and how

neural networks make their predictions by creating heatmaps of images and showing important areas involved in the decision making of the algorithm. By placing each mammogram beneath its respective heatmap, the areas which are most significant in model's decision can be located through different color variations. Grad-Cam plays a vital role in the field of explainable AI as it allows human beings to have a better understanding of inexplainable algorithms. The following heatmaps were generated using Grad-CAM.

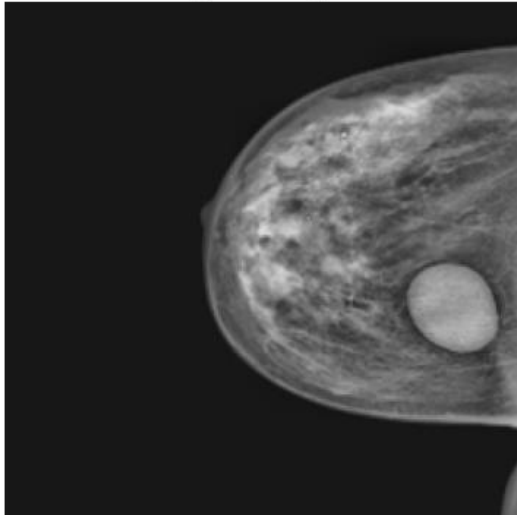
Original Image 1



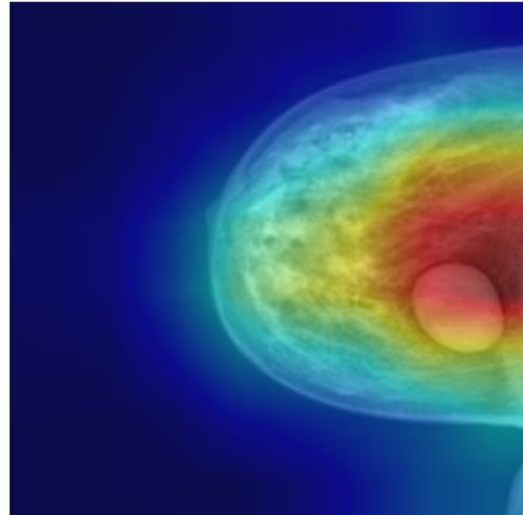
Grad-CAM Heatmap 1



Original Image 2



Grad-CAM Heatmap 2



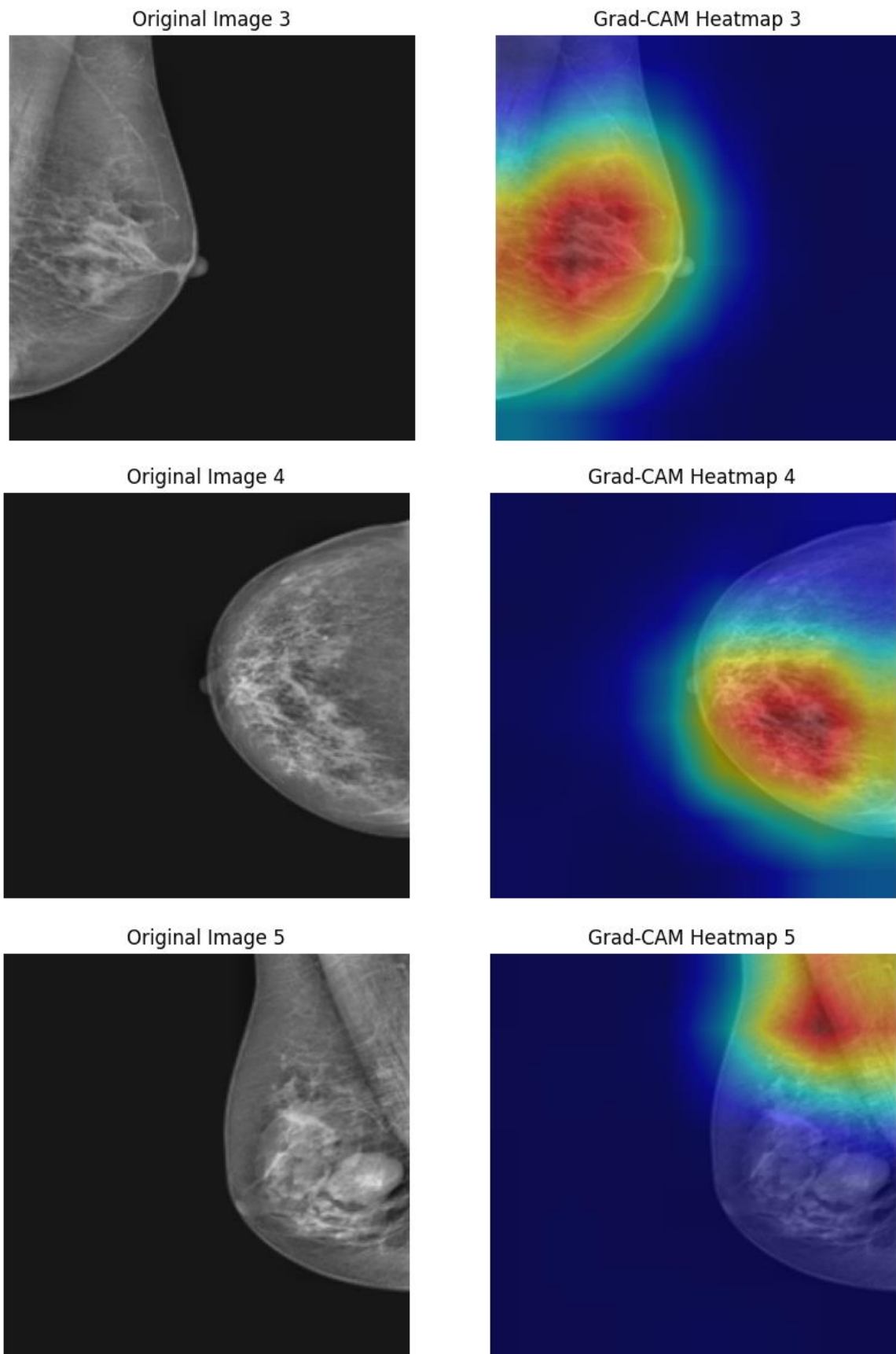


Figure 21: Original images from the dataset (on the left) and their heatmaps generated by Grad-CAM overlaid on top of the images (right). Red area indicated the regions which are the most relevant in the model's prediction.

On the left are the raw mammograms which are the original input to the model (test data) which is used by the model to analyze dense regions, breast tissue, shape of the lump, textures, patterns etc. in order to classify it as benign or malignant. While on the right are heatmaps placed above these images to see how the model is processing these images. The red regions of the heatmap indicate those regions on the original image which carry high significance in the model's thinking and label prediction. In each image the area under the red shade has influenced the most in the model concluding whether or not the subject has breast cancer. The yellow region had reasonable influence on the prediction while the blue region had zero or no relevance when the model was processing the unseen test data.

It can be seen that the model was successfully able to identify the breast region apart from the background as the blue region in all the images is the black background of the mammograph. All the important regions, shaded red, lie inside the shape of the breast and the next most relevant area, yellow shaded region, lies around the red shade thus it can be said that the algorithm smartly learned the area it is required to focus on, i.e., the breast tissue. However, it can be assured with full confidence that the model has yet the ability to identify the lump within the breast with full accuracy. In image 3 and 4 there are no visible lumps or calcification thus it is safe to say that red area lies above the correct region. However, in images 1, 2 and 5 despite being very visible breast the red shade slightly deviated from the lump in images 1 and 2 or completely deviated from the tumor as in the last image.

## 5. Conclusion

---

To conclude, by using extensive data management approaches, deep machine learning and artificial intelligence have the capacity to correctly evaluate tumor microenvironment models (Malherbe, 2021). It can significantly reduce the instances of misdiagnosis by being an aid to diagnostic experts and help in early detection of breast cancer, increasing the mortality rate of cancer patients. According to Shen et al (2021), the use of an AI system improved radiologists' ability to diagnose breast cancer as the use of AI resulted in a notable decrease in biopsy referrals of 27.8% and false-positive rates of 37.3%. Many researchers, engineers and medical experts are under discussion on how to fully grasp the full potential of AI in diagnostic application especially for malignant tumors. In this study similar work has been done on digital mammograms by implementing a hybrid model classifier for breast tumor diagnosis which proved to be successful as it yielded great results.

A pair of DenseNet121 and ResNet50 was utilized to extract local, small-grained details features along with global, large-scale patterns. The dataset included in this study is a combination of three extensively used datasets in research, namely INbreast, DDSM and MIAS. From the training, validation and test loss along with their accuracies stated in the previous section it is obvious that the model performed very well. The accuracy of each individual class is almost 1 (100%) as well as each performance indicator that is also obtained very near to 1 (100%) while the losses are all close to 0 (0%). The predictive model only misclassified 52 mammographs out of 1653 benign and 50 images from a total of 2347 malignant tumors.

### 5.1. Future work

Even though the algorithmic system involved in this study gave exceptional result, there are still some limitations that can be explored in the future work

- In case of this application, it is preferred for the AI to give additional false positives rather than misclassifying true positives. The future work on this model can include identifying why the 3% of accuracy is not being met, especially in the case of malignant tumors being identified as benign. Fine tuning the model hyper-parameters or improving the preprocessing can be helpful in increasing the recall.
- For further advancement a radiologist or a diagnostic expert can be made a part of this study in order to have a better understanding of the mammograms and diagnostic process. Involving a medical practitioner will also help improve the results of Grad-CAM where the model can be seen not completely focusing on the lump itself. Understanding why is this happening can also be addressed in future work
- Making a high-quality dataset with a large number of samples will also help in improving the accuracy and performance of the AI algorithm thus a very important part of further studies in this project is to create a huge dataset of digital mammograms with high quality and annotated images.

## 6. References

---

1. Alotaibi, M., Aljouie, A., Alluhaidan, N., Qureshi, W., Almatar, H., Alduhayan, R., Alsomaie, B. and Almazroa, A. (2023) 'Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images', *Heliyon*, 9(11), e22406.
2. American Cancer Society, n.d. *Breast Cancer*. Available at: <https://www.cancer.org>.
3. Bernardi, D., Ciatto, S., Pellegrini, M., Anesi, V., Burlon, S., Cauli, E., Depaoli, M., Larentis, L., Malesani, V., Targa, L., Baldo, P. and Houssami, N. (2012) 'Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time', *British Journal of Radiology*, 85(1020), pp. e1174–e1178. doi: 10.1259/bjr/19385909.
4. Chan, H.P., Samala, R.K. and Hadjiiski, L.M., 2019. CAD and AI for breast cancer—recent development and challenges. *The British journal of radiology*, 93(1108), p.20190580.
5. Cho, P., & Yoon, H.-J. (2021). *Evaluation of U-net-based Image Segmentation Model to Digital Mammography*. Presented by Emory University and Oak Ridge National Laboratory.
6. Cho, W.C. (2007) Contribution of oncoproteomics to cancer biomarker discovery. *Mol Cancer*, Volume 6, 25. (Available at: <https://doi.org/10.1186/1476-4598-6-25>)
7. Cuthrell, K.M. and Tzenios, N., 2023. Breast Cancer: Updated and Deep Insights. *International Research Journal of Oncology*, 6(1), pp.104-118.
8. Davis, L., (2022) *Breast Masses: Cancerous Tumor or Benign Lump? [Online]* [Available at: <https://www.verywellhealth.com/breast-cancer-tumors-or-benign-masses-430277>] [Accessed 5 December 2024].
9. Dhungel, N., Carneiro, G. and Bradley, A.P. (2015) 'Automated mass detection in mammograms using cascaded deep learning and random forests', in *Proceedings of the 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Adelaide, SA, Australia, 2015, pp. 1–8. doi: 10.1109/DICTA.2015.7371234.
10. Dromain, C., Boyer, B., Ferré R., Canale, S., Delalogue, S., Balleyguier, C., (2013) Computed-aided diagnosis (CAD) in the detection of breast cancer, *European Journal of Radiology*, Volume 82, Issue 3, Pages 417-423, (Available at: <https://doi.org/10.1016/j.ejrad.2012.03.005>)
11. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. and Dean, J. (2019) 'A guide to deep learning in healthcare', *Nature Medicine*, 25(1), pp. 24–29. doi: 10.1038/s41591-018-0316-z.
12. Fletcher, S.W. and Elmore, J.G. (2003) 'Clinical practice. Mammographic screening for breast cancer', *New England Journal of Medicine*, 348(17), pp. 1672–1680. doi: 10.1056/NEJMcp021804.
13. Gilik, A., Ogrenci, A.S. and Ozmen, A., 2022. Air quality prediction using CNN+ LSTM-based hybrid deep learning architecture. *Environmental science and pollution research*, pp.1-19.
14. Global Burden of Disease Cancer Collaboration, Fitzmaurice, C., Dicker, D., Pain, A., Hamavid, H., Moradi-Lakeh, M., MacIntyre, M.F., Allen, C., Hansen, G., Woodbrook, R.,



- Wolfe, C., Hamadeh, R.R., Moore, A., Werdecker, A., Gessner, B.D., Te Ao, B., McMahon, B., Karimkhani, C., Yu, C., Cooke, G.S., Schwebel, D.C., Carpenter, D.O., Pereira, D.M., Nash, D., Kazi, D.S., De Leo, D., Plass, D., Ukwaja, K.N., Thurston, G.D., Yun Jin, K., Simard, E.P., Mills, E., Park, E.K., Catalá-López, F., deVeber, G., Gotay, C., Khan, G., Hosgood, H.D. 3rd, Santos, I.S., Leasher, J.L., Singh, J., Leigh, J., Jonas, J.B., Sanabria, J., Beardsley, J., Jacobsen, K.H., Takahashi, K., Franklin, R.C., Ronfani, L., Montico, M., Naldi, L., Tonelli, M., Geleijnse, J., Petzold, M., Shrimme, M.G., Younis, M., Yonemoto, N., Breitborde, N., Yip, P., Pourmalek, F., Lotufo, P.A., Esteghamati, A., Hankey, G.J., Ali, R., Lunevicius, R., Malekzadeh, R., Dellavalle, R., Weintraub, R., Lucas, R., Hay, R., Rojas-Rueda, D., Westerman, R., Sepanlou, S.G., Nolte, S., Patten, S., Weichenthal, S., Abera, S.F., Fereshtehnejad, S.M., Shiue, I., Driscoll, T., Vasankari, T., Alsharif, U., Rahimi-Movaghar, V., Vlassov, V.V., Marcenes, W.S., Mekonnen, W., Melaku, Y.A., Yano, Y., Artaman, A., Campos, I., MacLachlan, J., Mueller, U., Kim, D., Trillini, M., Eshрати, B., Williams, H.C., Shibuya, K., Dandona, R., Murthy, K., Cowie, B., Amare, A.T., Antonio, C.A., Castañeda-Orjuela, C., van Gool, C.H., Violante, F., Oh, I.H., Deribe, K., Soreide, K., Knibbs, L., Kereselidze, M., Green, M., Cardenas, R., Roy, N., Tillmann, T., Li, Y., Krueger, H., Monasta, L., Dey, S., Sheikhabaei, S., Hafezi-Nejad, N., Kumar, G.A., Sreeramareddy, C.T., Dandona, L., Wang, H., Vollset, S.E., Mokdad, A., Salomon, J.A., Lozano, R., Vos, T., Forouzanfar, M., Lopez, A., Murray, C. and Naghavi, M. (2015) 'The global burden of cancer 2013', *JAMA Oncology*, 1(4), pp. 505–527. doi: 10.1001/jamaoncol.2015.0735. Erratum in: *JAMA Oncology*, 1(5), p. 690. doi: 10.1001/jamaoncol.2015.2892.
15. Godavarty, A., Rodriguez, S., Jung, Y.J. and Gonzalez, S. (2015) 'Optical imaging for breast cancer prescreening', *Breast Cancer (Dove Medical Press)*, 7, pp. 193–209. doi: 10.2147/BCTT.S51702.
  16. INbreast (n.d.) *INbreast: A full-field digital mammography database for breast cancer detection*. Available at: <https://www.inbreast.org>
  17. Jung, R.E., Fenner, N., Hämmerle, C.H. and Zitzmann, N.U. (2013) 'Long-term outcome of implants placed with guided bone regeneration (GBR) using resorbable and non-resorbable membranes after 12-14 years', *Clinical Oral Implants Research*, 24(10), pp. 1065–1073. doi: 10.1111/j.1600-0501.2012.02522.x.
  18. Jiang, B., Bao, L., He, S. et al. (2024) 'Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis', *Breast Cancer Research*, 26, p. 137. Available at:
  19. Kaushal, C., Bhat, S., Koundal, D., Singla, A., (2019) Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images, *IRBM*, Volume 40, (Issue 4), Pages 211-27, (Available at: <https://doi.org/10.1016/j.irbm.2019.06.001>)
  20. Khanna, M., (2019), Paper review: DenseNet -Densely Connected Convolutional Networks, *Medium* [Online] (Available at: <https://towardsdatascience.com/paper-review-densenet-densely-connected-convolutional-networks-acf9065dfefb>) [Accessed on 10 December 2024]
  21. Kuttan, G.O. and Elayidom, M.S. (2023) 'Review on computer aided breast cancer detection and diagnosis using machine learning methods on mammogram image', *Current Medical Imaging*, 19(12), pp. 1361–1371. doi: 10.2174/1573405619666230213093639.



22. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B. and Sánchez, C.I. (2017) 'A survey on deep learning in medical image analysis', *Medical Image Analysis*, 42, pp. 60–88. doi: 10.1016/j.media.2017.07.005.
23. López, N.C., García-Ordás, M.T., Vitelli-Storelli, F., Fernández-Navarro, P., Palazuelos, C. and Alaiz-Rodríguez, R. (2021) 'Evaluation of feature selection techniques for breast cancer risk prediction', *International Journal of Environmental Research and Public Health*, 18(20), p. 10670. doi: 10.3390/ijerph182010670.
24. Malherbe, K., 2021. Tumor microenvironment and the role of artificial intelligence in breast cancer detection and prognosis. *The American journal of pathology*, 191(8), pp.1364-1373.
25. Mall, S., Lewis, S., Brennan, P., Noakes, J. and Mello-Thoms, C. (2017) 'The role of digital breast tomosynthesis in the breast assessment clinic: a review', *Journal of Medical Radiation Sciences*, 64(3), pp. 203–211. doi: 10.1002/jmrs.230.
26. Mello-Thoms, C., 2003. Perception of breast cancer: eye-position analysis of mammogram interpretation. *Academic radiology*, 10(1), pp.4-12.
27. Michael, E., Ma, H., Li, H. and Qi, S. (2022) 'An optimized framework for breast cancer classification using machine learning', *Biomedical Research International*, 2022, Article ID 8482022. doi: 10.1155/2022/8482022.
28. Mukherjee, S., (2022), The Annotated ResNet-50, *Medium [Online]* (Available at: <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758>) [Accessed on 10 December 2024]
29. NCI or National Cancer Institute (2021) *What is Cancer?* [Online] (Available at: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>) [Accessed 8 December 2024]
30. Qasrawi, R., Daraghme, O., Qdaih, I., Thwib, S., Vicuna Polo, S., Owienah, H., Abu Al-Halawa, D., & Atari, S. (2024). Hybrid ensemble deep learning model for advancing breast cancer detection and classification in clinical applications. *Heliyon*, 10(19), e38374.
31. Sawyer-Lee, R., Gimenez, F., Hoogi, A., & Rubin, D. (2016). Curated Breast Imaging **Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]**. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2016.7002S9CY>
32. Shen, Y., Shamout, F.E., Oliver, J.R., Witowski, J., Kannan, K., Park, J., Wu, N., Huddleston, C., Wolfson, S., Millet, A. and Ehrenpreis, R., 2021. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. *Nature communications*, 12(1), p.5645.
33. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) 'Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209–249. doi: 10.3322/caac.21660.
34. Talaat, F.M., Gamel, S.A., El-Balka, R.M., Shehata, M., & ZainEldin, H., 2024. Grad-CAM Enabled Breast Cancer Classification with a 3D Inception-ResNet V2: Empowering Radiologists with Explainable Insights. *Cancers (Basel)*, 16(21), p.3668. doi: 10.3390/cancers16213668. PMCID: PMC11544836.

35. Verma, B., McLeod, P. and Klevansky, A., (2010) Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert systems with applications*, 37(4), pp.3344-3351
36. Wang, L., 2024. Mammography with deep learning for breast cancer detection. *Frontiers in Oncology*, 14, p.1281922. Available at: <https://doi.org/10.3389/fonc.2024.1281922> .
37. Wang, Z., et al. (2022). Environmental Risk Factors in Breast Cancer. *Cancer Reviews*, 14(6), 112-123.
38. Wellings E, Vassiliades L, Abdalla R. Breast Cancer Screening for High-Risk Patients of Different Ages and Risk - Which Modality Is Most Effective? *Cureus*. 2016 Dec 28;8(12):e945. doi: 10.7759/cureus.945. PMID: 28133583; PMCID: PMC5268380.