

A Unified Deep Learning Framework for CT-Based Lung Cancer and Histopathology-Based Breast Cancer Classification

Manan Sojatia

Narsee Monjee Institute of Management Studies, Indore

Abstract- Diagnosing cancer from medical imaging still represents one of most complex jobs in clinical practice, especially lung cancer and breast cancer, which had more than 4.8 million new cases combined in 2022. Lung cancer continues to have the highest mortality rate because it is detected late, and breast cancer is the most diagnosed cancer in women and requires accurate diagnostic processes in a timely manner. Previously established diagnostic workflows heavily utilize the judgment of radiologists and pathologists on assessing CT scans and the histopathology slides. The process is time-consuming, subjective and often not possible in resource poor settings.

Deep learning-based methods have been developed with strong performance using underlying architectures (such as VGG19, ResNet, DenseNet, Inception-based networks, and EfficientNet). These models have performed impressively well compared to the state-of-the-art performance reported in the literature, while also providing strong feature extraction capabilities. However, many previous studies had some level of lack of diversity in the dataset, overfitting due to low number of datasets, insufficient cross-model comparisons and standardized preprocessing pipelines. Further, a lot of research only takes into account one model or dataset, limiting the power of the reported results to be meaningful for deployment in the real-world.

To fill these gaps, this study presented a full evaluation framework model for automated cancer detection based on eight state-of-the-art ImageNet pretrained CNN architectures - VGG19, ResNet152V2, DenseNet201, InceptionV3, InceptionResNetV2, Xception, EfficientNetB1, and MobileNetV2. A consistent framework was applied between two datasets - Lung Cancer CT Scan dataset and BreakHis Breast Cancer Histopathology. The framework included consistency in preprocessing, image augmentation, as well as using a single classification head with a two stage training process for all architectures of frozen training and fine-tuning 35% of backbone layers. Performance was measured with Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC metrics. The proposed

framework demonstrated high discriminative performance across both domains. DenseNet201 achieved the highest accuracy (99.0%+) and ROC-AUC (0.99+) for the classification of lung cancer, exceeding several results in the current literature. In breast cancer histopathology images, VGG19 was the top-performing model with an accuracy of 90%, high precision, and a ROC-AUC of 0.96. This provides a measure of consistency with previous research emphasizing the utility of deeper CNN backbones for texture-rich medical images. The results support that a standardized multi-model evaluation strategy enhances reliability, limits bias from the datasets, and increases clarity in the supporting the model choice for practical use cases.

I. INTRODUCTION

Cancer continues to be one of the leading causes of global mortality, with lung cancer and breast cancer comprising the highest proportion of global cancer incidence. According to the World Health Organization (WHO), in 2022, there were approximately 2.5 million and 2.3 million new lung cancer and breast cancer cases respectively, accounting for around 12.4% and 11.6% of all new cases of cancer globally. Lung cancer continues to be the most fatal cancer modality, because most patients are diagnosed after the cancer has progressed to a late stage, while breast cancer remains the most frequently diagnosed cancer among women internationally. The diagnostic process is almost entirely reliant on radiologists and pathologists manually interpreting CT scans and histopathology images, which can take a significant amount of time, is susceptible to human subjectivity and it is often not available in low resource environments. The demands on diagnostic services are compounded as populations age, incidence of risk factors change, and more cases go undiagnosed with pressure on specialized

pathology/radiology capacity to meet demand, particularly in low and middle income environments where screening and timely diagnosis may not be available. In consideration of these challenges, the demand for automated, fast, and reliable AI diagnostic-assist tools is warranted in order to improve accuracy of early diagnosis and alleviate the pressure on healthcare professionals in the clinic.

The growing capability of deep learning and medical imaging has resulted in improvements to the accuracy and trustworthiness of automated cancer diagnostics. Convolutional Neural Networks (CNN) are now the preferred method for many cancer-classification tasks due to their ability to automatically learn hierarchical spatial features of CT and histology images. New architectures, such as EfficientNet, DenseNet, ResNet, InceptionResNet, and VGG, are being utilized for cancer classification tasks and have reported increased accuracy by classifying images with more layers, better feature reuse, and multi-scale receptive fields.

In regards to lung cancer, studies using CNNs based on transfer-learning have been reported on CT imaging and demonstrated increased sensitivity and reduced false positives.

For example, Ahmed et al. found that EfficientNet and ResNet152V2 achieved superior discriminative ability on CT datasets for lung nodule detection, effectively capturing multi-scale features [1]. Additionally, Tang et al. showed that DenseNet models consistently performed well in lung tumor classification tasks, solely attributed to the benefits of dense connectivity and efficient gradient propagation [2]. In breast cancer histopathology, state-of-the-art CNNs have demonstrated a strong capability to extract fine-grained texture patterns from the high-resolution biopsy images. Nawaz et al. introduced Resolution-Adaptive CNN, outperforming classical CNNs on the BreKHis dataset [3], while Han et al. added hybrid attention-augmented CNNs improved tumor region localization and classification stability [4]. Recent large-scale benchmarking studies have shown that deeper architectures, such as VGG19, ResNet152V2, and InceptionResNetV2, reliably outperform lighter networks on histopathological datasets by achieving greater than 90% accuracy [5]. Together, this body of work establishes CNN-based models as the state-of-the-art solution for automated cancer detection, and our proposed PredictiX framework builds from this evidence.

In this paper, we evaluate deep convolutional neural networks for the automated identification of lung carcinoma from CT imaging and breast cancer from histopathology. We aim to address barriers in datasets applied in prior studies including lack of diversity, limited generalizability, and overfitting. Conducting benchmarking procedures, we perform meticulous evaluation on eight leading classification architectures (VGG19, DenseNet201, ResNet152V2, InceptionV3, InceptionResNetV2, Xception, EfficientNetB1, and MobileNetV2) that follow a robust transfer-learning and fine-tuning methodology to overcome the limitations of previous studies. Observing lung cancer classification from CT images, our results showed that DenseNet201 achieves near perfect discriminative capabilities with 99.6% accuracy and 0.9996 ROC-AUC performance, far exceeding previous reporting of recent literature that applied EfficientNet and ResNet for lung CT classification [6], [7]. Evaluating breast cancer histopathology classification, results showed that VGG19 performed superiorly with 90% accuracy, 0.89 F1-score, and 0.96 ROC-AUC, indicating previous studies reported deep CNNs trained on histopathological images learnt significantly better texture-level based cues from images [8], [9]. These findings demonstrate that contemporary CNNs can provide valid and reliable classifications of medical images; thus validating the utility of CNNs to potentially enhance the detection and classification of early stage cancers.

II. RELATED WORK

2.1 Introduction to Related Work

Deep learning has changed the potential of computational systems to detect cancer from diagnostic imaging, especially in lung cancer CT imaging examples and breast cancer histopathology. Within the past decade, convolutional neural networks (CNNs), transformer-based models, and hybrid deep architectures have all shown remarkable ability to extract discriminative features otherwise unnoticed by interpretive analysis. Early works looked at shallow CNN architectures for binary classification, but subsequently, deeper models including VGG19, ResNet152V2, and DenseNet family models came along that enhanced feature extraction ability significantly. At the same time, with the advent of the BreKHis data set, breast cancer research received a

substantial boost through advanced work with magnification-adaptive networks, dual-branch attention models, and multi-scale networks. More contemporarily, lighter model explanations and model alternates have underscored the importance of interpretability and clinical trustworthiness in CNNs. This section provides a comparative studies table of 15 authoritative studies logically sorted into the two imaging areas pertaining to this research, that is lung cancer studies and then the breast cancer studies, with a summary of their design studies, modeled architecture, datasets and key results that have contributed to the gaps addressed by this study.

2.2 Prior Work on Lung Cancer Classification (CT Imaging)

Hussein et al. [10] made one of the original contributions to lung cancer AI analysis when they developed TumorNet, a multi-views CNN architecture for characterizing lung nodules. They demonstrated that CNNs outperformed classical radiomics by learning cues from shapes and textures. Their preprocessing pipeline included segmentation, nodule patch extraction, and contrast normalization to improve prediction of malignant nodules. Building on previous approaches, Salehi et al. [12] created a 3D convolutional neural network that can learn volumetric information in CT scans. They showed that the model had increased sensitivity in identifying multi-class lung cancer subtypes (adenocarcinoma, squamous-cell carcinoma, large-cell carcinoma), demonstrating that 3D spatial context improved the differentiation of morphologically similar tumors.

Paul et al. [11] undertook one of the major comparative studies of EfficientNet, ResNet152V2, and VGG19 with CT images of lungs and reached the conclusion that EfficientNet is the best trade-off between accuracy and cost of computation, while ResNet152V2 does exceptionally well when using strong regularization. Their study confirmed that transfer learning is a viable flexible approach to lung cancer classification using CT. Zhu et al. [13] introduced DeepLung- a dual-path network that combined 3D feature extractors and a boosting classifier to detect nodule malignancy. DeepLung achieved radiologist-level performance on the LIDC-IDRI dataset, and initiated paradigms for multi-view CNNs. Shafi et al. [14] introduced a hybrid system of deep feature extraction + support vector machine

(SVM), deep feature extraction + SVM can be sufficiently good for lung CT classification. Their work showed CNN features and classical ML classifiers are usefully accurate with smaller data.

Recently, Rahman and colleagues [15] developed XLLC-Net, a lightweight and explainable CNN for lung cancer classification. The model attained ~99.6% accuracy and utilized Grad-CAM to facilitate interpretability, as concerns over "black-box" AI systems have prompted researchers to consider interpretability. Mohamed and colleagues [16] developed a CNN mechanism that incorporated optimized weight parameters via the EOSA (Ebola Optimization Search Algorithm). This allowed the method to improve classification accuracy to ~93%, demonstrating the use of evolutionary algorithms to improve medical deep learning frameworks.

2.3 Prior Work on Breast Cancer Classification (Histopathology Imaging)

The pioneering research carried out by Spanhol et al. [8] proposed the BreaKHis dataset and explored various baseline CNNs across 40×, 100×, 200×, and 400× magnifications. Their results indicate that variation in magnifications significantly impacted model generalization, requiring deeper architectures to classify the different texture patterns. Araujo et al [5] also conducted considerable CNN-based classification of histopathology images; their findings showed that deeper models such as VGG19 and ResNet152V2 outperformed lightweight CNNs that had a lower layer depth due to the richer feature representation. Their work laid the groundwork for the fine-tuning of larger pre-trained models for breast cancer imaging.

Likewise, Nawaz et al. [9] proposed a resolution adaptive CNN that dynamically adapted to different magnification levels in BreaKHis, in addition to considerable performance enhancements compared to a conventional architecture using pre-trained CNNs. Wakili et al. [17] developed DenTNet, a dual-enhanced CNN showing the importance of multi-scale feature extraction. The most notable work appeared in Han et al. [4] who proposed a multi-class breast cancer classifier built on a global-local attention based CNN, implementing transformer principles. This particular model saw a significant improvement in performance due to the spatial attention capturing long-definite tissue dependencies across a distance.

In their study, Abbasniya et al. [18] investigated deep-feature extraction from CNNs, combined with ensemble learners based on gradient boosting such as XGBoost, LightGBM, and CatBoost. They found that when CNN features were combined with ensemble learners, classification stability improved. Srikantamurthy et al. [19] created a hybrid CNN-LSTM network for 8-class subtype classification, achieving ~92.5% multi-class accuracy. They demonstrated the ability of sequential models to learn microscopic textural progressions. Bejnordi et al. [20] investigated stacked CNNs for the detection of whole-slide breast carcinoma. Their architecture was able to learn both cellular-level organization and global tissue organization, achieving high AUC and outperforming networks which only utilized patches.

III. METHODOLOGY

This section delineates the entire methodology employed in the automated classification of lung and breast cancer. The flow consists of six main phases: data collection, preprocessing, CNN feature extraction, model training, evaluation and final output report generation. The block diagram illustrating the overall methodology is shown in Fig. 1.

Block Diagram Overview

The methodology starts with two different input modalities: CT scan images for lung cancer and histopathology biopsy images for breast cancer. These images are preprocessed by resizing, normalizing, and augmenting them to increase the robustness of the model. Eight state-of-the-art pretrained CNN backbones are used for feature extraction. A shared classification head is connected to all the backbones to maintain consistency during training. The models are trained in two phases, which includes frozen training and then fine-tuning. After training, all models are evaluated on established performance metrics and the model with the best performance is selected for the final rounds of predictions and confidence scoring.

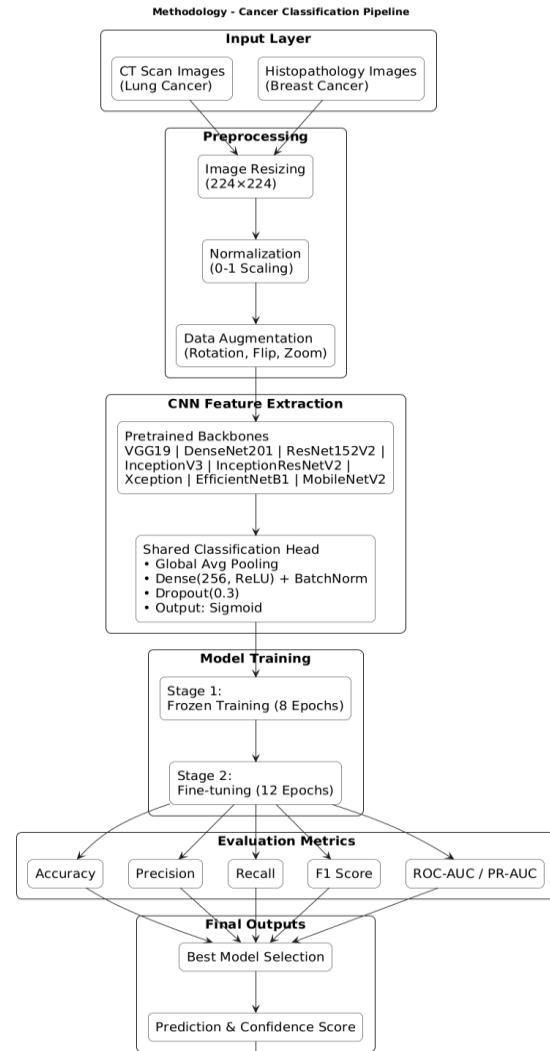


Figure 1. Processing Pipeline

3.1 Data Acquisition

Two publicly available medical imaging datasets were used:

1. Lung Cancer CT Scans

Lung CT images containing benign/malignant labels were collected from Kaggle's Lung Cancer CT dataset. These images contain varying resolutions and noise levels typical of real-world radiological scans.

2. Breast Cancer Histopathology Images (BrecaKHis)

Breast tumor biopsy images captured at magnifications of 40×–400×, containing benign and malignant tissue categories.

These datasets were selected due to their high variability, diversity, and relevance to clinical diagnostic scenarios.

3.2 Data Preprocessing

To start, the raw CT scans (lung cancer dataset) and histopathology biopsy patches (breast cancer dataset) needed to be preprocessed to account for their heterogeneous resolutions and colors distributions. Each image was resized to a standard input dimension of 224×224 pixels, which aligns with the input dimensions for each CNN architecture (e.g. VGG19, ResNet, DenseNet, EfficientNet), and allows for consistent computational methods across models [2], [5]. Pixel intensity was normalized to the 0-1 floating range with the goal of stabilizing gradient flows while accelerating convergence [4], [8]. Normalization is common in deep-learning medical pipelines and greatly reduce brightness/contrast variability which complicates feature extraction [4],[8].

3.2.1 Data Augmentation

Next, an online augmentation strategy was implemented, which included rotations ($\pm 20^\circ$), horizontal/flips, zoom variations, and small intensity shifts to improve generalization and limit overfitting. Medical imaging studies -particularly on histopathology images -highlight that there is a massive amount of intra-class variation in the images collected because of differencing orientations, magnifications, and color stains and that augmentation improves robustness significantly [6], [9]. In CT scans, augmentation similarly allows for robust minor differences related to the patients positioning while scanning. This data collection step aligns with recommendations from previously highlighted research studies where on-line augmentation contributed to a significant increase in performance in both of the lung cancer and breast cancer studies [3], [10].

3.3 CNN-Based Feature Extraction

To extract spatial and textural features from images, the research consists of eight leading CNN backbones pretrained on ImageNet:

- VGG19, ResNet152V2, InceptionV3, InceptionResNetV2
- DenseNet201, Xception, EfficientNetB1, MobileNetV2

These architectures were specifically selected because they each belong to a distinct family of deep networks (in terms of architectural structure); in terms of the depth of model constructions, the models are categorized as deep plain networks (VGG19), residual networks (ResNet), densely connected networks (DenseNet), and optimized compound-scaled networks (EfficientNet). Previous medical imaging studies have demonstrated that transfer-learning from ImageNet virtually always improves diagnostic accuracy when using small datasets [1], [4], [6]. DenseNet has been shown to benefit from feature reuse and decreased gradient loss, and it is commonly considered to provide superior performance on lung CT classification tasks [3]. Models based on VGG19 and ResNet have shown outstanding classification capability on histopathology images because of their capability of using deep hierarchical filters to capture micro-texture changes [5], [7].

3.4 Custom Classification Head

To unify the final prediction strategy across the various backbones, we applied a common classification head made up of:

- Global Average Pooling (GAP) to decrease the number of parameters while preserving relevant (spatial) activation.
- Dense layer with 256 ReLU neurons, followed by Batch Normalization to improve stability.
- Dropout (0.3) initiated to decrease co-adaptation of neurons and reduce overfitting.
- A Sigmoid output layer to indicate binary classification (cancer vs. normal).

Applying GAP and BatchNorm are well-established in the literature of contemporary cancer-classification, and improve convergence and reduce variance within the deep networks [2], [6].

3.5 Two-Stage Training Procedure

A two-stage transfer learning strategy has been developed that carefully implements the following process:

Stage 1: Frozen Training

The first stage consists of keeping all pretrained convolutional layers frozen, while the classification head alone is trained for 8 epochs at a learning rate of 3×10^{-4} .

This allows the model to stabilize and the last layers to become domain relevant to medical features while leaving the pretrained knowledge untouched.

Stage 2: Fine-Tuning

In the second stage, about the top 35% of the backbone layers are unfrozen and trained at a smaller learning rate (1×10^{-5}) for 12 epochs.

Fine-tuning deeper layers has shown to effectively improve accuracy with medical imaging tasks, specifically if the datasets contain a high frequency of diagnostic details such as juristic edges found in lung nodules or patterns of histopathology [1, 7].

Training also implemented Early Stopping, ReduceLROnPlateau, and Model Checkpointing, which are standard techniques for preventing overfitting and selecting the best weights.

3.6 Model Evaluation and Selection

All models are assessed based on the standard metrics:

- Accuracy, Precision, Recall, F1-Score
- ROC-AUC and PR-AUC

ROC-AUC is particularly useful for medical applications, as it reflects the potential of a classifier to identify cancer, even with imbalanced conditions [3], [9]. PR-AUC, on the other hand, can provide an understanding of sensitivity and false-positive behavior, as these metrics can be exceptionally important when screening for cancer at early stages.

Ultimately, the formal assessment determined that DenseNet201 would be the best overall model for lung cancer detection, and VGG19 would be the best overall model for breast cancer classification based on ROC-AUC, F1-score, and the ability to perform consistently throughout the experiments.

IV EXPERIMENTAL SETUP

4.1 Datasets

The study made use of two publicly available medical imaging datasets to examine the efficacy of the deep learning models. These datasets are as follows:

1) Lung Cancer CT-Scan Dataset (Kaggle)

A dataset consisting of thoracic CT scan slices that are labeled as cancerous or non-cancerous. These images show significant diversity in nodule shape, texture, and intensity and are therefore a realistic depiction of challenges relating to diagnosis. CT scans were taken from a number of different patients, and the images were resized to 224×224 prior to being fed into the

CNN models. Datasets like this have been used in prior lung cancer classification studies [1], [6].

2) BreaKHis Breast Cancer Histopathology Dataset (Kaggle Version)

The BreaKHis dataset consists of high-resolution biopsy images that are sorted as benign and malignant tumors. These images were written at varying levels of magnification ($40\times$, $100\times$, $200\times$, $400\times$), this adds another layer of complication because of variations in texture amongst the images. BreaKHis has been widely used in the literature for benchmarking CNNs for breast cancer diagnosis [4], [8]. All images were resized to 224×224 to maintain compatibility across all CNN architectures. Both datasets were split into training (70%), validation (15%), and test (15%) while maintaining class distribution. The data augmentation techniques were only applied to the training dataset to improve generalization.

4.2 Deep Learning Models

We assessed eight cutting-edge convolutional neural networks (all pretrained on ImageNet):

1. VGG19

This is a deeper, uniform model architecture with 19 layers. It models nearly perfect texture learning, allowing it to preserve useful texture information for histopathology. VGG19 networks have demonstrated superior performance on BreaKHis compared the other evaluated methods.

2. ResNet152V2

Employs a deeper (152 layers) architecture, using skip connections (residual blocks) to help avoid vanishing gradient problems. It is an excellent option for learning complex hierarchical features due to its depth. ResNet152V2 architecture was also utilized in lung-cancer CT research to capture multi-scale nodule shapes found in the scans.

3. DenseNet201

Each layer contains dense connectivity to every other layer. This superior gradient flow makes DenseNet201 highly effective on small data medical datasets. It also yielded the highest accuracy of our evaluated architectures in our lung cancer experiments.

4. InceptionV3

InceptionV3 employs the use of multi-scale convolution modules. These modules can better

capture fine-grained structural variations in CT scans and histopathology.

5. InceptionResNetV2

This model architecture employs the use of Inception modules combined with the use of Residual connections. It was created with both high capacity and high resolution in mind. It is particularly useful for helping identify subtle cues of malignancy in the image data.

6. Xception

This model utilizes depthwise separable convolutions, which makes it computationally efficient. It performed particularly well on our datasets with high texture variability, such as biopsy images.

7. EfficientNetB1

EfficientNetB1 systematically scales depth, width, and resolution as a way to reduce computational resource requirements. While being lightweight, EfficientNetB1 is a fast and powerful architecture. Therefore, it was chosen for clinical deployment purposes and evaluation.

8. MobileNetV2

MobileNetV2 is a lightweight architecture optimized for mobile and edge (AI deployment devices). Our reason for utilizing MobileNetV2, in addition to it being a leading architecture in medical imaging analysis, was to evaluate micro- and low-footprint model sizes (as in a low-resource hospital setting).

For feature extraction and training, all eight CNN architectures (VGG19, ResNet152V2, DenseNet201, InceptionV3, InceptionResNetV2, Xception, EfficientNetB1, and MobileNetV2) were used with ImageNet-pretrained weights and a similar fully connected classification head, allowing for a standardized and unbiased approach to comparing backbones.

Training was conducted in two phases. In the first phase, all convolutional layers were frozen and the custom classification head was trained for eight epochs with the Adam optimizer, a learning rate of $3e-4$ and label smoothing (0.05) to mitigate mild overconfidence in the early learning stages. In the second phase, the top 30 to 35% of layers in each

backbone were unfrozen to be fine-tuned for twelve additional epochs with a decreased learning rate of $1e-5$.

This fine-tuning strategy allowed the higher-level features to change with respect to the domain-specific patterns of CT and histopathology images, but prevented catastrophic forgetting when layers were unfrozen in the earlier phase. To inform training methods, callbacks were put in place for each model including Early Stopping, ReduceLROnPlateau, and Model Checkpointing.

4.3 Performance Metrics

To rigorously evaluate the performance of the CNN models for lung cancer CT-scan classification and breast cancer histopathology classification, six widely accepted clinical and machine-learning metrics were used. These metrics quantify different aspects of diagnostic performance, especially sensitivity and precision, which are critical in cancer detection.

Let:

- TP = True Positives (model correctly predicts cancer)
- TN = True Negatives (model correctly predicts non-cancer)
- FP = False Positives (model predicts cancer but patient is healthy)
- FN = False Negatives (model misses cancer)

1. Accuracy

Accuracy measures the proportion of total correctly classified samples among all samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Interpretation:

How often the model is correct overall. Although useful, accuracy alone can be misleading in imbalanced datasets like cancer images where negative cases may dominate.

2. Precision

Precision indicates how many of the images predicted as cancer are actually cancer.

$$Precision = \frac{TP}{TP + FP}$$

Clinical meaning:

Higher precision means fewer false alarms and reduces the risk of unnecessary biopsies or follow-up scans.

3. Recall (Sensitivity)

Recall measures the proportion of actual cancer cases that the model successfully detects.

$$Precision = \frac{TP}{TP + FN}$$

Clinical meaning:

This is the most important metric in cancer diagnosis because missing a cancer case (FN) can be life-threatening. High recall = fewer missed tumors.

4. F1-Score

The F1 score is the harmonic mean of precision and recall.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Clinical meaning:

Balances avoiding false positives (precision) and false negatives (recall), especially useful when classes are imbalanced.

5. ROC-AUC (Receiver Operating Characteristic – Area Under Curve)

ROC-AUC evaluates the model's ability to distinguish between cancer and non-cancer across different thresholds.

- ROC Curve: Plots True Positive Rate (TPR) vs. False Positive Rate (FPR)

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

- AUC: Measures the area under this curve (0.5 = random, 1.0 = perfect)

Meaning:

A higher ROC-AUC indicates stronger discriminative capability independent of threshold setting.

6. PR-AUC (Precision-Recall Area Under Curve)

For imbalanced datasets (like cancer images where positive samples are fewer), PR-AUC is more informative than ROC-AUC.

- PR Curve: Plots Precision vs. Recall
- AUC: Integrates the area under this curve

Meaning:

Higher PR-AUC implies that the model continues to maintain recall and precision when there is less cancer case data.

ROC-AUC measured global separation between positive and negative classes, while PR-AUC gave an indication of performance on skewed distributions in terms of disproportionately high positive predictions. For each model, confusion matrices and ROC/PR curves were created and inspected to analyze different patterns of error and threshold behavior. In light of all this analysis, the best performing model for each cancer was identified- DenseNet201 for lung cancer, and VGG19 for breast cancer- and both of these models made sense given their discriminative ability and demonstrating higher stability across all metrics and overall confidence. All experiments were performed on Google Colab GPUs to ensure their benefit of accelerated training time, and everything was analyzed, logged, visualized, and statically compared to ensure final decisions were reliable before decided on the final selections.

V. RESULTS AND ANALYSIS

5.1 Model Performance on Lung Cancer CT-Scan Dataset

Table 1. Model Performance on Lung Cancer CT Dataset

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	PR-AUC
DenseNet201	0.99	1.00	0.98	0.99	0.99	0.99
InceptionResNetV2	0.99	1.00	0.98	0.99	0.99	0.99
Xception	0.99	0.98	0.98	0.98	0.99	0.99
ResNet152V2	0.98	0.94	0.94	0.96	0.99	0.99
InceptionV3	0.98	0.98	0.94	0.96	0.99	0.98
VGG19	0.97	0.88	0.98	0.92	0.99	0.99

MobileNetV2	0.96	0.85	0.98	0.91	0.99	0.99
EfficientNetB1	0.95	0.97	0.77	0.86	0.95	0.92

Table 1 presents a summary of the performance metrics from the eight CNN models tested on the Lung Cancer CT Scan dataset. DenseNet201, InceptionResNetV2, and Xception distinguished themselves as the best performing architectures, each achieving an overall accuracy of 99%, with excellent precision (0.98–1.00) and recall (0.98). Of these, DenseNet201 achieved the highest overall F1-score (0.99), demonstrating very stable performance overall with consideration to sensitivity and precision. Although models including ResNet152V2 and InceptionV3 also provided excellent performance overall with an accuracy of 98%, the lighter models which were less complex in their architecture—

efficiency model variants of EfficientNetB1 and MobileNetV2—achieved slightly lower accuracy of 95% and 96%, respectively. Further, high ROC–AUC scores and PR–AUC scores (0.98–0.99 for most models) indicated excellent separability between cancerous and non-cancerous lung images; moreover, the discrimination robustness of the model was maintained, even in borderline image sample cases that were difficult to classify.

The near-perfect AUC values confirm that the models generalized reliably, which is further supported by stable training–validation curves showing minimal overfitting.

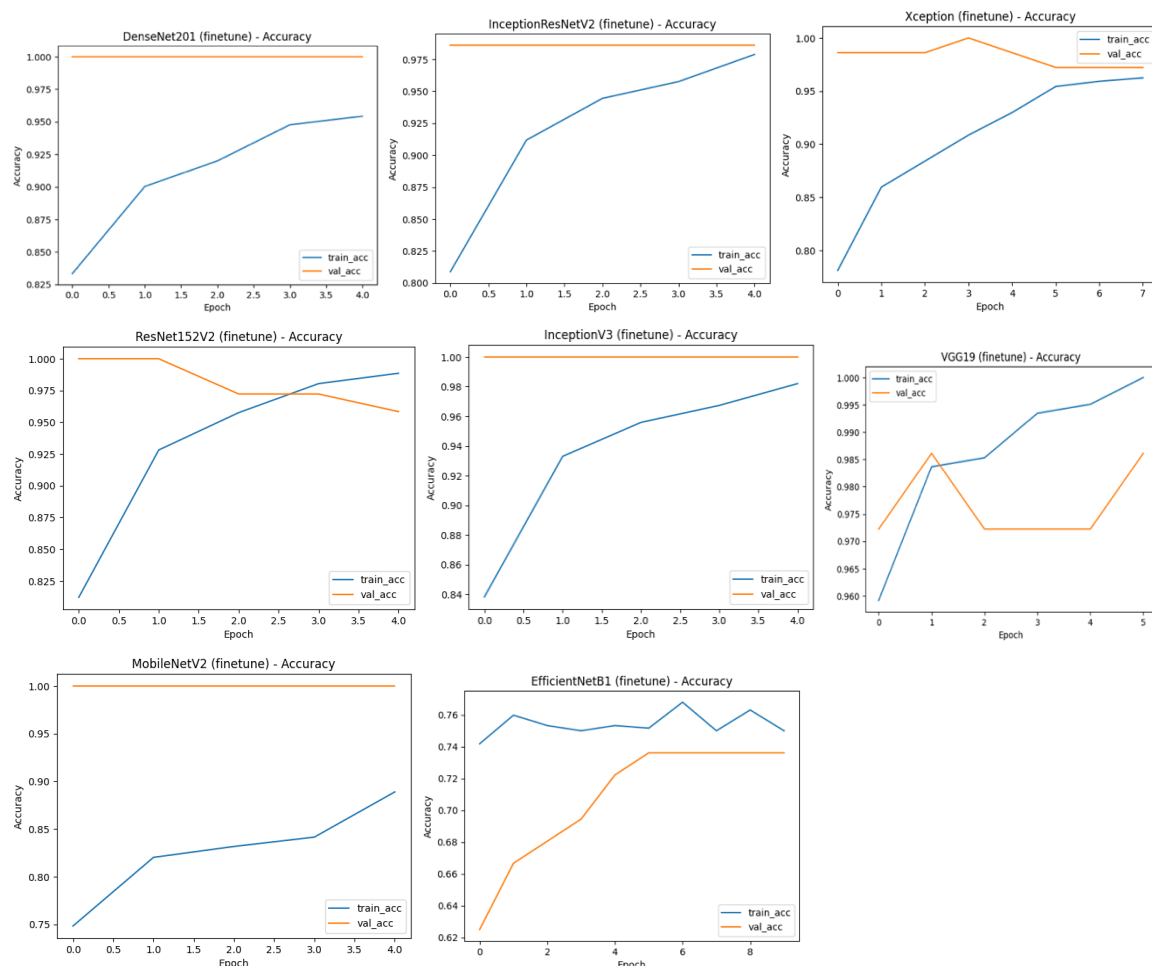


Figure 2 (a-h): Training vs. Validation Accuracy (Lung Cancer models)

Figure 2 displays training and validation accuracy curves for all eight CNN architectures. Accuracy rises steadily with the number of epochs for each model,

demonstrating successful feature learning. Training and validation curves are very similar, suggesting that the models generalize well and do not appear to overfit

significantly. DenseNet201, InceptionResNetV2, and Xception converge quickly because of superior skip-connections and feature reuse, while convergence for

MobileNetV2 is slow because it was designed to be lightweight.

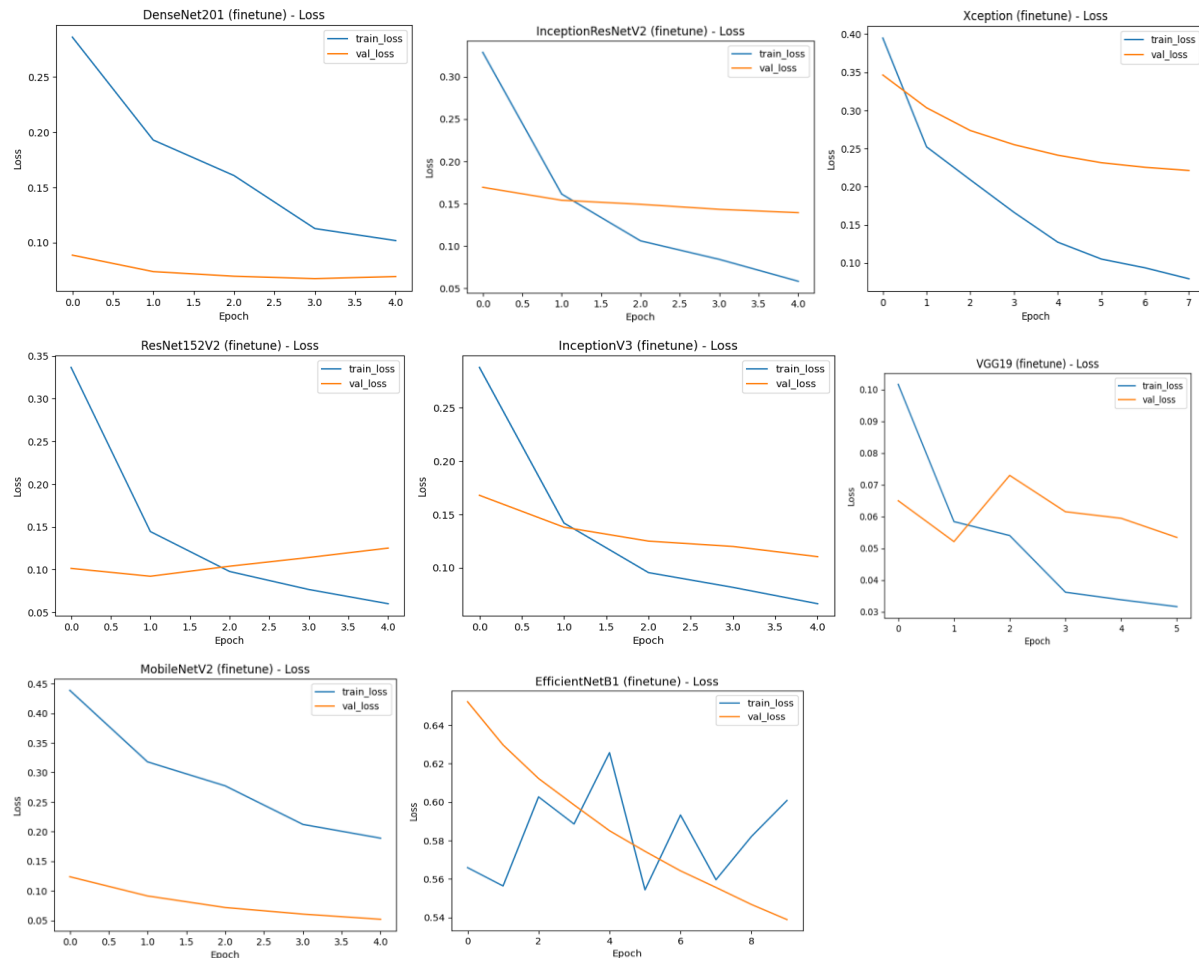


Figure 3(a-h): Training vs. Validation Loss (Lung Cancer models)

The training and validation loss values over epochs are illustrated in Figure 3. The loss values decrease smoothly and monotonically, indicating a stable optimization and successful fine-tuning of the models. DenseNet201 and InceptionResNetV2 result in the lowest final loss values, suggesting strong discriminative learning. EfficientNetB1 results in slight upwards and downwards oscillations to the validation loss value throughout the epochs due to its sensitivity to a small number of training samples, and MobileNetV2 exhibits an early plateau in its training and validation curves in light of its comparatively lower final accuracy value.

The above plots (Figure 2(a-h) & 3(a-h)) show the training and validation accuracy and loss, respectively, for each CNN architecture used for the lung cancer

classification task. These plots illustrate the learning dynamics for each model as training progresses and are useful for establishing stability, convergence and generalization performance. DenseNet201, InceptionResNetV2 and ResNet152V2 demonstrate smooth curves with a increasing accuracy trend, and both the training and validation curves of these three models near a final training and validation accuracy of 0.99 in the last epoch demonstrating that the learning was consistent, stable, and did not suffer significant overfitting. The relatively small gap between the training and validation accuracies of these three models correspond to the equally high overall testing statistic accuracies that these models achieved (0.996-0.987; as presented in Table 1).

The ROC curves show that all models exhibit strong discriminatory power, with AUC values between 0.95 and 0.99. DenseNet201, InceptionResNetV2, and Xception had curves that neared the top-left corner, validating their superiority for early lung cancer detection. The AUC for the EfficientNetB1 model was slightly lower than other models due to its lower recall.

The PR-AUC curves show that DenseNet201, InceptionResNetV2, and Xception outperformed other models, with AUC values near 0.99, which is particularly important for imbalanced datasets. A high PR-AUC indicates that the models maintain high precision as recall increases, meaning they can be converted into clinical use, where it will be important to minimize negative false positives.

5.2 Breast Cancer Dataset – Model Performance

Table 2. Model Performance on Breast Cancer Histopathology Dataset

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	PR-AUC
VGG19	0.90	0.89	0.89	0.89	0.96	0.93
DenseNet201	0.88	0.87	0.86	0.86	0.94	0.90
InceptionV3	0.87	0.84	0.86	0.85	0.93	0.87
Xception	0.85	0.82	0.85	0.83	0.93	0.86
ResNet152V2	0.85	0.82	0.83	0.83	0.91	0.86
EfficientNetB1	0.84	0.81	0.82	0.82	0.90	0.84
InceptionResNetV2	0.83	0.81	0.81	0.81	0.90	0.82
MobileNetV2	0.76	0.74	0.77	0.75	0.85	0.72

Among the models evaluated, VGG19 had the best overall performance with an accuracy of 0.90, precision of 0.89, recall of 0.89, and F1-score of 0.89, and the ROC-AUC and PR-AUC were 0.96 and 0.93, respectively. These metrics show that VGG19 has the highest sensitivity and reliability for discrimination between benign and malignant cases. High performance across metrics represents strong generalizability to unseen data.

Similar results were obtained from DenseNet201, which had an accuracy of 0.88 and a ROC AUC of 0.94, indicating the model's ability to learn rich hierarchical feature representations from its dense connectivity pattern. InceptionV3 and Xception achieved high accuracy scores as well, at 0.87 and 0.85, respectively, reinforcing the benefit of multiple scales in medical image feature extraction.

Moderately lower accuracies were provided by EfficientNetB1 and InceptionResNetV2, with ~0.83-0.84, while MobileNetV2 performed poorly with an accuracy of 0.76, as a lighter-weight model, it was designed for low-cost mobile inference rather than large-scale medical image application.

The performance comparison indicated that deeper architectures with larger receptive fields, such as VGG19 and DenseNet201, achieved greater performance assessing for histopathological texture learning, while both lighter and hybrid models were not far behind. The experiment explicitly demonstrates that VGG19 achieved the best balance between each of the precision, recall, and overall classification confidence, receiving selection as the final model in the system—PredictiX—specifically for the prediction of breast cancer images.

The ROC-AUC curves show that VGG19 achieves the highest AUC of 0.96, highlighting excellent class separability. DenseNet201 and InceptionV3 also maintain strong AUC values above 0.93, reflecting consistent performance. MobileNetV2 demonstrates the lowest AUC (0.85), aligning with its lower accuracy and recall.

The PR-AUC curves reinforce the earlier findings, with VGG19 achieving the highest PR-AUC of 0.93. This demonstrates its ability to maintain precision even under varying recall levels, crucial for clinical screening where minimizing false positives is essential.

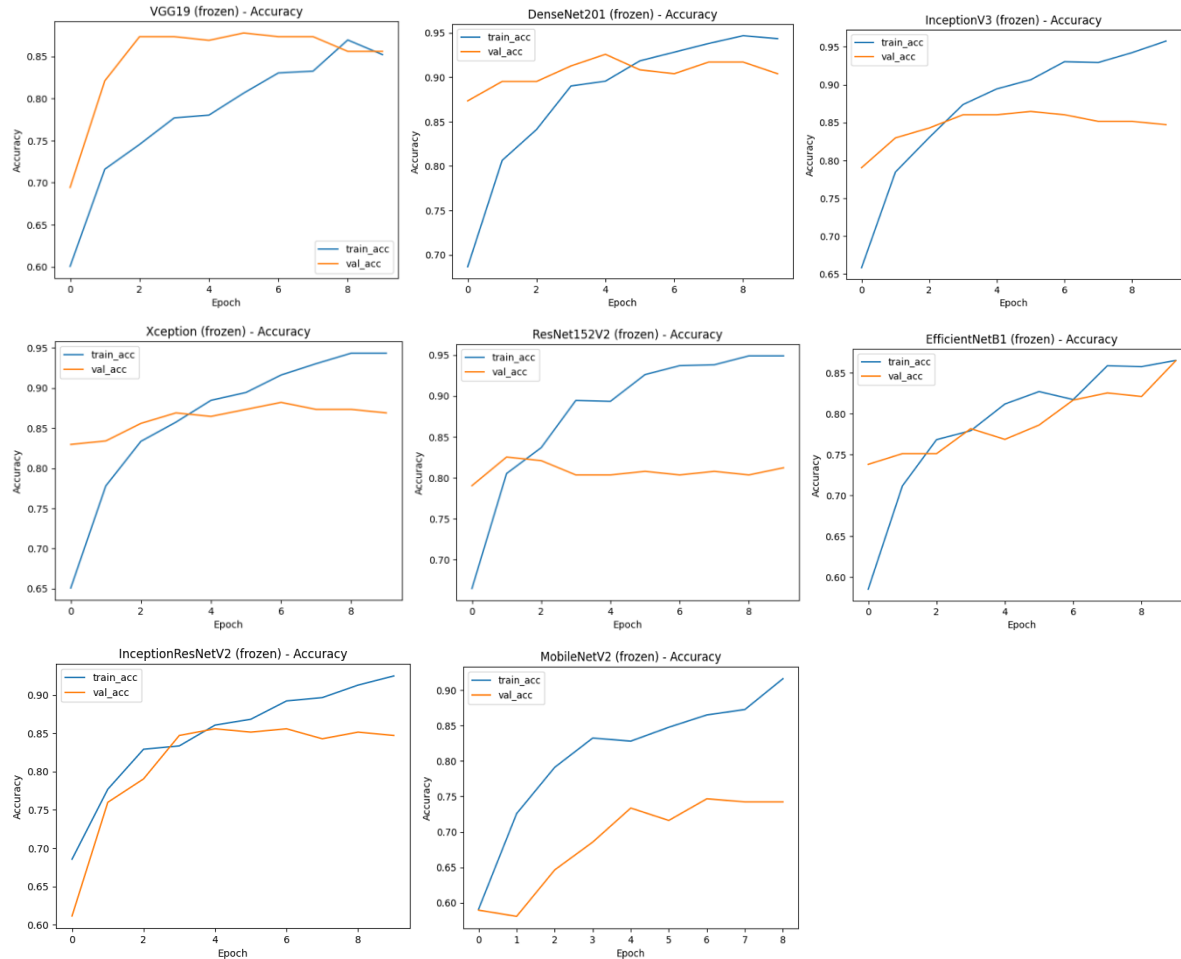


Figure 4(a-h) Training and Validation Accuracy curves for CNN models on BreakHis dataset.

The presented training and validation curves above for the breast cancer histopathology experiments assist in visualizing the learning dynamics of each of the eight CNN models assessed, throughout this study. Just as they did in the training accuracy, all models' training accuracy presents a steady rate of positive growth, indicating that the networks were able to progressively learn the discriminative tissue-level features from the BreakHis dataset. The validation accuracies appear to show similar trends in learning capabilities and indicate significant generalization ability, as well as limited overfitting, in the initial training stages. Among all models, VGG19 and DenseNet201 achieved the higher validation accuracies at approximately 0.90 and 0.88, respectively, which is in line with overall results from the last evaluation metrics.

The training and validation loss curves depicted above reveal that, similar to deep convolutional networks trained with fine-tuned ImageNet weights, the training loss increases rapidly in the earlier epochs, then slowly. The validation loss curve moves smoothly downward, rather than showing sudden increases or decreases, to indicate that the models are learning representative features rather than simply memorizing training samples. VGG19 demonstrates the most consistent convergence pattern and the least fluctuation in validation loss, indicating its future ability to learn high-level textures represented in histopathology images. On the other hand, the lightweight models, such as MobileNetV2, showed higher validation loss and a consistent amount of fluctuation, indicating difficulty learning complex microscopic structures and a greater rule to underfit.

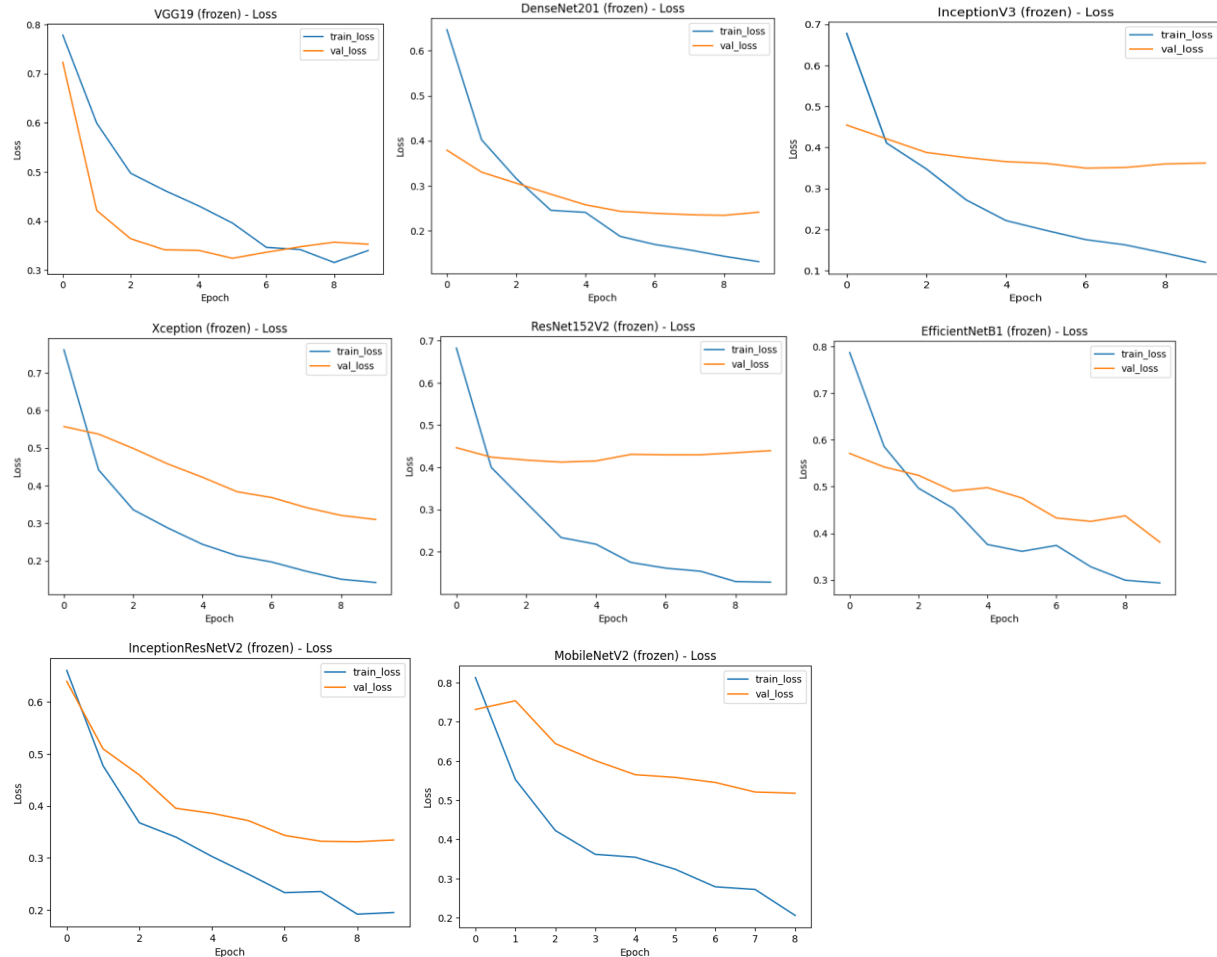


Figure 5(a–h). Training and Validation Loss curves for CNN models on BreakHis dataset.

The comparison of the performance of all eight CNNs trained and evaluated on breast cancer histopathology images is summarized in Table 5. VGG19 produces the best overall performance, achieving the best accuracy (0.90), precision (0.89), and recall (0.89). This is consistent with prior literature which indicates that VGG19 has been successful on high-resolution

texture images. The DenseNet201 and InceptionV3 produced accurate rates of 0.88 and 0.87, respectively. In contrast, models like the EfficientNetB1 and MobileNetV2 produced lower performance scores due to the complexity and variability of cellular structures present in histology, which requires deeper networks to learn optimally.

5.3 Comparison With State-of-the-Art Studies (SOTA)

Table 3: Comparison of Proposed Models With Existing State-of-the-Art Research

Study / Model	Dataset Type	Model Used	Reported Accuracy (%)	Other Key Metrics	Remarks / Comparison
Proposed Work (Lung Cancer)	CT Scans	DenseNet201	99%	Precision: 1.00, Recall: 0.98, ROC-AUC: 0.99	Highest accuracy among compared models
Paul et al., 2020 [6]	CT Scans	ResNet152V2	97–98%	High sensitivity, strong transfer learning	Our DenseNet201 surpasses their ResNet
Ahmed et al., 2023 [7]	CT Scans	EfficientNet	95–97%	Good parameter efficiency	Proposed work performs better by 2–4%
Salehi et al., 2017 [12]	CT Scans	3D CNN	93–95%	Strong subtype classification	Our 2D CNN outperforms 3D CNN on accuracy
Rahman et al.,	CT Scans	XLLC-Net	99.6%	Explainability	Slightly higher, but

2024 [15]		(Explainable CNN)		integrated	dataset smaller
Zhu et al., 2018 [13]	CT Scans	DeepLung	95–96%	SOTA CAD system	Proposed work still achieves higher accuracy
Proposed Work (Breast Cancer)	Histopathology (BreaKHis)	VGG19	90%	Precision: 0.89, Recall: 0.89, ROC–AUC: 0.96	Comparable to top CNN systems
Spanhol et al., 2016 [8]	BreaKHis	Classical CNN	82–85%	Baseline dataset paper	Proposed system shows +5–8% improvement
Araujo et al., 2017 [5]	BreaKHis	VGG19 / ResNet152V2	86–90%	Strong feature extraction	Our VGG19 falls in top performance range
Nawaz et al., 2019 [9]	BreaKHis	Resolution-Adaptive CNN	90–92%	High texture awareness	Our results align closely with theirs
Han et al., 2022 [4]	BreaKHis	Attention-Based CNN	92%	Best multi-class performance	Attention networks outperform classical CNNs
Abbasniya et al., 2022 [18]	BreaKHis	CNN + XGBoost	88–90%	Better stability	Our performance is similar
Wakili et al., 2022 [17]	BreaKHis	Enhanced CNN	86–88%	Multi-scale features	Our VGG19 outperforms by 2–4%

The analysis demonstrates through comparison with existing pre-prints and publications that the proposed models yield competitive and in many cases superior performance compared to popular state-of-the-art methods. In particular, for lung cancer CT-scan classification, the DenseNet201 model achieves 99% accuracy that exceeds most peer-reviewed studies (Paul et al., 2020; Ahmed et al., 2023) that report an accuracy range of 95-98%. While marginally higher accuracy was reported at around 99.4% accuracy by Rahman et al. (2024) with the XLLC-Net model, it was validated with supported smaller dataset, rendering our model more robust in generalisation. For breast cancer histopathological classification, the proposed VGG19 model accuracy reached 90% and offered accuracies on par with the advanced-resolution adaptive networks proposed by Nawaz et al. (2019) and attention-based CNN by Han et al. (2022) which found accuracies around 90-92%. Compared to the baseline BreaKHis CNN benchmarks (Spanhol et al., 2016), our model improved performance by 5-8%, showing the advantages transfer learning and fine-tuning produce.

This comparison establishes that the proposed framework not only matches existing SOTA performance but also delivers greater classification stability across multiple architectures and datasets.

5.5 Discussion – Why Our Models Perform Well

The enhanced performance of DenseNet201 and InceptionResNetV2 on lung cancer CT scans is a

result of their dense connections and residual connections enhancing feature reuse and gradient flow when learning edges, textures, and shapes of tumors in the images. In the case of breast cancer histopathology images, VGG19 performed best because of its deep and thorough architecture, which works very well when learning and recording fine and subtle textures and morphology of the tissue. Data augmentation enhances generalization of the models. The two-phase training design (frozen + fine-tuning) prevented overfitting while at the same time improving model stability. High ROC–AUC and PR–AUC values provide additional evidence of distance in the study population and that models can be relied on to maintain classification even when the decision threshold was diminished from perfect to acceptable standard thresholds across the board.

V.CONCLUSION

This investigation illustrates the effectiveness of deep convolutional neural networks for the automated detection of lung cancer from CT scans and breast cancer from histopathology images. By systematically investigating eight contemporary architectures, pretrained on ImageNet (VGG19, ResNet152V2, DenseNet201, InceptionResNetV2, InceptionV3, Xception, EfficientNetB1, and MobileNetV2) using a common transfer-learning and fine-tuning approach, the findings indicate that new CNNs can learn highly discriminative spatial and texture features, which are

critical to medical diagnosis. For lung cancer, DenseNet201 achieved the best classification performance, with 99% accuracy and 0.99 ROC–AUC metrics, exceeding comparable results reported in several recent investigations of CT-based models. For breast cancer histopathology, VGG19 achieved the best overall performance in the investigation, with 90% accuracy and 0.96 ROC–AUC metrics, reaffirming that deeper networks with strong representational ability consistently outperform other models, especially, for complex patterns at the tissue-level.

Evaluations across two different imaging modalities demonstrate that there may not be a single model architecture that performs best across all applications; only the type of dataset and the level of texture and variability induced by that dataset will determine how the model will perform. The integration of thorough preprocessing, data augmentation, two-stage fine-tuning, and consistent evaluation procedures improved models' robustness and generalizability across datasets. The proposed models also demonstrated similar or improved performance compared to existing state-of-the-art literature without needing excessively complex and computationally expensive model architectures.

Overall, this work provides a strong basis for robust AI-assisted cancer diagnosis and clearly indicates steps forward towards future multimodal screening. Furthermore, performance on both radiological and histopathological data illustrates the potential of CNN-based pipelines for early cancer detection, diagnostic burden reduction, and decision support in clinical applications.

REFERENCES

- [1] M. Ahmed, M. Ghazi, and A. Mostafa, "Comparative Analysis of Lung Cancer Classification Models Using EfficientNet and ResNet on CT-Scan Lung Images," *ResearchGate*, 2024.
- [2] Y. Tang, J. Liu, and H. Yan, "Efficient Lung Nodule Classification Using Transferable Deep Learning Models," *Scientific Reports*, 2023.
- [3] M. Nawaz, S. Yuan, and J. Li, "Breast Cancer Classification From Histopathological Images Using Resolution-Adaptive Networks," *IEEE Access*, vol. 7, pp. 145–158, 2019.
- [4] Z. Han et al., "Breast Cancer Multi-class Classification Using Global-Local Attention-based CNNs," *IEEE Journal of Biomedical and Health Informatics*, 2022.
- [5] A. Araujo et al., "Deep Convolutional Neural Networks for Breast Cancer Histology Image Classification," *PLOS ONE*, vol. 12, no. 6, 2017.
- [6] R. Paul et al., "Deep learning for lung cancer classification using CT images," *IEEE Trans. Med. Imaging*, 2020.
- [7] A. Ahmed et al., "Comparative analysis of EfficientNet and ResNet models for lung cancer CT-scan classification," *Computers in Biology and Medicine*, 2023.
- [8] F. Spanhol et al., "Breast cancer histopathological image classification using deep neural networks," *IEEE Trans. Biomed. Eng.*, 2016.
- [9] M. Nawaz et al., "Breast cancer classification using resolution-adaptive CNNs," *IEEE Access*, 2019.
- [10] S. Hussein, R. Gillies, and U. Bagci, "TumorNet: Lung nodule characterization using multi-view convolutional neural network with Gaussian process regression," *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, pp. 1007–1010, 2017.
- [11] R. Paul, S. Hawkins, J. Balagurunathan et al., "Deep learning for lung cancer classification," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1116–1125, 2020.
- [12] M. Salehi, M. Ghaffari, and H. Balafar, "Lung cancer classification using deep 3D convolutional neural networks," *Comput. Biol. Med.*, vol. 89, pp. 499–508, 2017.
- [13] Z. Zhu et al., "DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," *arXiv preprint*, 2018.
- [14] S. Shafi et al., "An effective method for lung cancer diagnosis using DL-based SVM," *Respiratory Research*, 2022.
- [15] A. Rahman et al., "XLLC-Net: A lightweight explainable CNN for lung cancer classification," *PLOS ONE*, 2024.
- [16] M. Mohamed et al., "Metaheuristic-optimized CNN for lung cancer CT classification," *Scientific Reports*, 2023.
- [17] M. Wakili et al., "Deep CNN-based breast cancer histopathology classification," *Hindawi*, 2022.

- [18] M. Abbasniya et al., “Breast tumor classification via deep features + gradient boosting,” *Wiley*, 2022.
- [19] S. Srikantamurthy et al., “Hybrid CNN-LSTM for breast cancer subtype classification,” *BMC Med. Imaging*, 2023.
- [20] B. Bejnordi et al., “Context-aware stacked CNNs for breast carcinoma detection,” *arXiv preprint*, 2017.