

# **Progress Seminar Report**

(Jan-Jul 2026 session)

*Submitted by*

Kirti Swagat Mohanty

Registration no. 21-3-04-068

*Under the guidance of*

**Dr. Prabina Pattanayak (Supervisor)**

Assistant Professor, ECE Dept, NIT Silchar

**Dr. Chandrajit Choudhury (Co-supervisor)**

Assistant Professor, ECE Dept, NIT Silchar



Department of Electronics & Communication Engineering

National Institute of Technology Silchar

## I. Introduction

Medical image analysis using deep learning has significantly advanced automated diagnosis of thoracic diseases. Convolutional Neural Networks (CNNs) have achieved expert-level performance in detecting abnormalities from chest X-ray images and lung cancer from CT scans [1], [2]. However, most existing approaches rely on **single-modality learning**, which limits diagnostic reliability due to incomplete representation of anatomical and pathological information.

Chest X-ray imaging is widely used for screening due to its low cost and accessibility, but it suffers from projection overlap and limited sensitivity for early-stage diseases. In contrast, CT imaging provides high-resolution volumetric details suitable for accurate lung cancer characterization, albeit at higher cost and radiation exposure [3]. Recent studies highlight that **multi-modal deep learning**, which jointly learns from complementary imaging modalities, improves robustness and clinical reliability [4].

Motivated by these findings, this work implements a **dual-modal deep learning architecture** that simultaneously processes chest X-ray and CT images using pretrained CNN backbones. The system performs **multi-label thoracic disease classification** from X-ray images and **multi-class lung cancer classification** from CT scans within a unified multi-task learning framework.

## II. Problem Statement and Objectives

### Problem Statement

Most deep learning models for thoracic diagnosis operate independently on either chest X-ray or CT data, ignoring the complementary diagnostic cues across modalities. Additionally, handling heterogeneous prediction tasks—multi-label disease detection and multi-class cancer classification—with a single unified architecture remains challenging.

## Objectives

The objectives of this work are:

1. To design a **dual-modal CNN-based encoder** for joint learning from chest X-ray and CT images.
2. To perform **multi-label classification** of thoracic diseases from chest X-ray images.
3. To perform **multi-class classification** of lung cancer types from CT scans.
4. To implement **feature-level fusion** for learning complementary representations.
5. To evaluate the model using **Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics**.

## III. Methodology & System Architecture

### A. Dual-Modal Encoder Design

The implemented architecture consists of two independent CNN encoders based on EfficientNet-B0, a state-of-the-art scalable convolutional architecture known for superior accuracy-to-parameter efficiency [5].

- X-ray Encoder:  
Processes chest X-ray images resized to  $224 \times 224 \times 3$ . Global average pooling is applied, followed by dense layers with batch normalization and dropout for regularization.
- CT Encoder:  
Processes CT scan images using an identical EfficientNet-B0 backbone and feature refinement layers.
- Refinement: Each modality features a sequential refinement block consisting of Dense layers (1024 units), ReLU activation, Batch Normalization, and Dropout (0.3) to enhance discriminative capacity.

Both encoders are trainable, enabling fine-tuning on medical imaging data.

## B. Feature Fusion Strategy

The extracted 512-dimensional feature vectors from X-ray and CT encoders are concatenated and passed through fully connected layers to form a shared fused representation. This combined vector is processed through a fusion network consisting of:

- **Concatenation Layer:** Merges X-ray and CT feature maps.
- **Fusion Dense Block:** A 1024-unit layer followed by a 512-unit bottleneck layer to produce a shared latent embedding.

Feature-level fusion has been shown to outperform decision-level fusion in multi-modal medical image analysis [4].

## C. Multi-Task Learning Framework

The fused representation is used for two parallel prediction tasks:

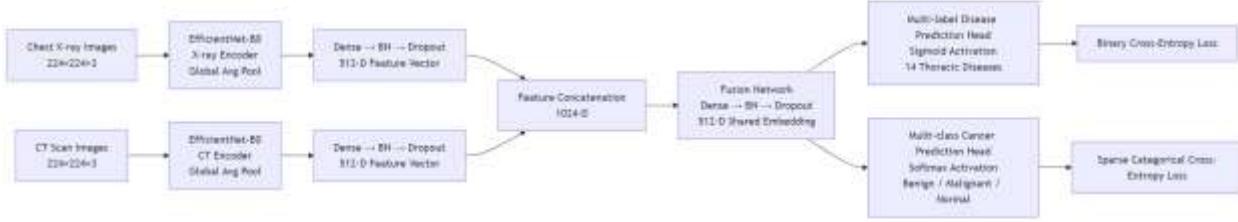
- **Disease Prediction Head:**  
Employs **Sigmoid activation** for **multi-label** classification of multilabel diseases (e.g., Atelectasis, Cardiomegaly) [6].
- **Cancer Classification Head:**  
Employs **Softmax activation** for multi-class classification of CT categories.

This joint learning approach improves generalization by sharing representations across related tasks [7].

## D. Dataset Preparation and Training

- Chest X-ray images are loaded from a folder-structured dataset with multi-label annotations.

- CT images are loaded using class-wise directory organization.
- Random pairing of X-ray and CT samples is used during training.
- The model is trained using Adam optimizer with binary cross-entropy and sparse categorical cross-entropy losses.
- Early stopping, learning-rate scheduling, and model checkpointing are employed to prevent overfitting.



## IV. Key Accomplishments (July-Dec 2025)

1. Implemented a dual-modal EfficientNet-based encoder architecture.
2. Developed a multi-task learning pipeline supporting heterogeneous prediction objectives.
3. Designed custom data loaders for multi-label and multi-class datasets.
4. Integrated ROC-AUC, F1-score, and confusion matrix evaluation for robust performance analysis.
5. Successfully trained and validated the system using TensorFlow and Keras.

## V. Preliminary Results

Initial experimental evaluation demonstrates:

- Stable convergence of training and validation loss curves.
- Effective learning of multi-label disease patterns from chest X-ray images.
- High class separability in CT-based cancer classification, as observed from ROC-AUC analysis.
- Improved representational robustness due to feature-level fusion and multi-task learning.

## References

- [1] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

<https://doi.org/10.1016/j.media.2017.07.005>

- [2] S. Rajpurkar *et al.*, “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” *arXiv*, 2017.

<https://arxiv.org/abs/1711.05225>

- [3] H. Kang *et al.*, “Deep learning-based lung cancer diagnosis using CT images,” *Computer Methods and Programs in Biomedicine*, vol. 193, 2020.

<https://doi.org/10.1016/j.cmpb.2020.105444>

- [4] T. Zhou *et al.*, “Multi-modal deep learning for lung cancer diagnosis,” *Medical Image Analysis*, vol. 68, 2021.

<https://doi.org/10.1016/j.media.2020.101873>

- [5] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” *ICML*, 2019.

<https://proceedings.mlr.press/v97/tan19a.html>

- [6] J. Irvin *et al.*, “CheXpert: A large chest radiograph dataset with uncertainty labels,” *AAAI*, 2019.

<https://doi.org/10.1609/aaai.v33i01.3301590>

- [7] R. Caruana, “Multitask learning,” *Machine Learning*, vol. 28, pp. 41–75, 1997.

<https://doi.org/10.1023/A:1007379606734>