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—within 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 224×224×3. 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-uit later 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] J. Irvin, P. Rajpurkar, et. al, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. AAAI Conf. Artif. Intell., 2019, pp. 590–597.image.jpg

[2] S. G. Armato III, G. McLennan, et. al, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of chest CT images and lesion annotations," Medical Physics, vol. 38, no. 2, pp. 915–931, 2011

[3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141.

[4] A. Vaswani, N. Shazeer, P. N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[5] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, 2018

[6] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in Proc. Int. Conf. Learn. Representations (ICLR), 2018.

[7] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Representations (ICLR), 2017.

[8] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. Summers, and R. M. Summers, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7098–7107

[9] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2017, pp. 2980–2988.

[10] M. Yeung, E. Sala, C. B. Schönlieb, and A. Ruyeto, "Unified focal loss: Generalising Dice and cross entropy losses to handle class imbalance," in Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI), 2021, pp. 479–488.

[11] G. King and L. Zeng, "Logistic regression in rare events data," Political Analysis, vol. 9, no. 2, pp. 137–163, 2001.