FORMATIVE ASSESSMENT -1 REPORT

Institute Name: Pimpri Chinchwad College of Engineering (PCCoE)

Department: Information Technology

Subject: Deep Learning

Academic Year: 2025–26

Title: Formative Assessment Report – Deep Learning Mini Project

Student Name: Harivdan Shrihari Narayanshastri

PRN: 125M1K010

Faculty Name: Prof. Tanuja Patankar

Course / Class: F.Y. M.Tech (AI & DS)

Submission Date:15 DEC 2025

Contents

Chapter 1

Formative Assessment 1

1.1 Objective

To study and analyze recent research papers (2022–2025) on deep learning-based medical image classification, identify existing research gaps, and propose an innovative and explainable multi-modal deep learning mini project for implementation.

1.2 Topic Selection

Chosen Topic: Medical Image Classification (X-ray, CT, MRI) Using Deep Learning

Medical image classification is crucial for accurate and early disease diagnosis. Manual interpretation of medical images is time-consuming and prone to human error. Deep learning techniques provide automated, accurate, and scalable solutions, particularly when combining multiple imaging modalities such as X-ray, CT, and MRI.

1.3 Literature Review Summary

Paper 1: Deep Learning for Medical X-ray, MRI, and Ultrasound Classification

a. Paper is About (Title):

Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: a scoping review

This scoping review analyzes deep learning applications across three medical imaging modalities for automated disease classification, focusing on preprocessing strategies, neural architectures, and evaluation metrics.

b. Dataset Used (Features):

- ChestX-ray14: 112,120 chest X-rays, 14 disease labels, 1024×1024 resolution
- BraTS: Brain tumor MRI with multi-sequence imaging (T1, T2, FLAIR)
- ISIC: Skin lesion dermoscopy images
- Features: Resolution 224×224 to 1024×1024, grayscale/RGB, 2–14 classes

c. Methodology Used:

- Architectures: ResNet-50/101, DenseNet-121/169, EfficientNet-B0 to B7, VGG-16/19
- Transfer learning using ImageNet pretrained models
- Preprocessing: normalization, resizing, CLAHE contrast enhancement
- Augmentation: rotation (±15°), flipping, zoom, brightness adjustment

d. Results:

- Accuracy: 85–98% depending on disease and modality
- X-ray lung disease: >90% accuracy
- MRI brain abnormalities: >92% accuracy
- Combined preprocessing + augmentation improved performance by 5–8%

e. Key Findings:

Transfer learning improves performance on small datasets and reduces training time. However, scarcity of labeled data and annotation variability remain major challenges.

Paper 2: CT Image Classification with Deep Learning

a. Paper is About (Title):

Deep learning models for CT image classification: a comprehensive literature review

This paper surveys deep learning evolution for CT imaging, emphasizing COVID-19 detection and lung nodule classification.

b. Dataset Used (Features):

- COVID-CT: 746 images (349 positive, 397 negative)
- LIDC-IDRI: 1,018 cases, 2,669 lung nodules, 1–5 mm slice thickness
- CT-COVID: 275,000 CT slices
- Features: 512×512 resolution, Hounsfield Units, lesion annotations

c. Methodology Used:

- CNNs: AlexNet, VGG-16, ResNet-18/50/101, DenseNet
- Foundational models: Vision Transformer, Swin Transformer
- CAD systems combining segmentation and classification

d. Results:

- COVID-19 detection AUC: 0.95–0.99
- Sensitivity: 94–98%, Specificity: 92–96%
- Foundational models achieved +3–7% improvement at 2–3× computational cost

e. Key Findings:

CNNs remain efficient baselines. Transformers improve accuracy but introduce high computational overhead. Domain shift across scanners affects generalization.

Paper 3: AMRI-Net Multi-Modal Framework

a. Paper is About (Title):

Deep learning-based image classification for AI-assisted integration of pathology and radiology (AMRI-Net)

b. Dataset Used (Features):

- ISIC 2019: 25,331 dermoscopic images, 8 classes
- HAM10000: 10,015 images, 7 classes
- OCT2017: 84,495 retinal OCT images
- Brain MRI: multi-sequence tumor data

c. Methodology Used:

- Multi-resolution CNN feature extraction
- Attention-guided feature fusion
- Explainable Domain-Adaptive Learning (EDAL)

d. Results:

- ISIC accuracy: 94.95%, OCT: 96.15%, Brain MRI: 92.37%
- Average AUC: 0.976
- Cross-domain improvement: +6–9%

e. Key Findings:

Multi-resolution attention improves generalization and interpretability. EDAL enables cross-institutional learning.

Paper 4: Explainable AI in Medical Imaging

a. Paper is About (Title):

Explainable Deep Learning Methods in Medical Image Classification: A Survey

b. Methodology Overview:

- Feature attribution (Grad-CAM, CAM, LRP)
- Attention-based explainability
- Concept-based explanations (TCAV)
- Inherently interpretable architectures

c. Results:

- Grad-CAM IoU: 72–85%
- Interpretable models improved clinician trust by 23%

d. Key Findings:

Post-hoc methods lack faithfulness. Inherently interpretable models provide more reliable explanations.

Paper 5: Hybrid CNN-Transformer for Interpretability

a. Paper is About (Title):

A Hybrid Fully Convolutional CNN-Transformer Model for Inherently Interpretable Disease Detection

b. Dataset Used:

- Diabetic Retinopathy: 28,100 images
- AMD: 4,000+ images

c. Methodology Used:

- CNN backbones (ResNet, BagNet)
- Vision Transformer with sparse attention
- Evidence map-based explanations

d. Results:

- DR accuracy: 94.3%, AUC: 0.968
- Localization IoU: 76.4%

e. Key Findings:

Hybrid CNN–Transformer architectures can achieve inherent interpretability without sacrificing accuracy.

Literature Review Summary Table

Sr. No.	Paper Title	Dataset Used	Methodology	Results / Findings
1	X-ray, MRI, ultrasound review	ChestX-ray14, BraTS, ISIC	CNN Transfer Learning	Accuracy up to 98%
2	CT image classification review	COVID-CT, LIDC-IDRI	CNNs, Transformers	AUC up to 0.99
3	AMRI-Net framework	ISIC, HAM10000, OCT	Multi-resolution CNN + Attention	AUC 0.976
4	XAI survey	ChestX-ray14, DR, MRI	Grad-CAM, Attention	Improved clinical trust
5	Hybrid CNN-Transformer	Fundus datasets	CNN + Transformer	Accuracy 94.3%

Sr.	Paper	Technique	Key Outcome
1	X-ray/MRI Review	CNN Transfer Learning	Accuracy up to 98%
2	CT Classification Review	CNN + Transformers	AUC up to 0.99
3	AMRI-Net	Multi-modal Attention CNN	AUC 0.976
4	XAI Survey	Grad-CAM, TCAV	Improved trust
5	Hybrid CNN–Transformer	CNN + ViT	Accuracy 94.3%

Table 1.1: Summary of reviewed literature

1.4 Comparative Analysis

Table 1: Methodology Comparison

Paper Title	Author Name(s)	Year & Publication	Methodology Used	Advantages	Disadvantages
Deep learning for X-ray, MRI, ultrasound	Laçi, Sevrani, Iqbal	2025, BMC Medical Imaging	CNN Transfer Learning (ResNet, DenseNet, EfficientNet)	Strong baseline, reduced training time, effective on small datasets	Limited global context, extensive preprocessing required
CT classification review	Ahmad, Dai, Xie, Liang	2025, Quantitative Imaging in Medicine and Surgery	CNNs to foundational models (ViT, Swin Transformer), CAD systems	High accuracy (AUC $>$ 0.95), comprehensive survey coverage	Computationally expensive, large dataset requirement, limited cross-validation
AMRI-Net multi-modal framework	Multiple Authors	2025, Frontiers in Medicine	Multi-resolution CNN, attention fusion, EDAL domain adaptation	Joint X-ray/CT/MRI handling, explainability, improved generalization	High computational cost (3.2× baseline), complex architecture
Explainable AI survey	Patrício et al.	2023, ACM Computing Surveys	CAM, Grad-CAM, attention models, TCAV, interpretable CNNs	Improved clinical trust, comprehensive explainability taxonomy	Post-hoc methods lack faithfulness, 1–3% accuracy trade-off
Hybrid CNN-Transformer	Djoumessi et al.	2025, arXiv	BagNet/ResNet + ViT + sparse attention + evidence maps	Interpretable-by-design, state-of-the-art accuracy (94.3%)	Limited to fundus images, higher complexity than CNNs

Table 2: Dataset Comparison

Paper Title	Year & Publication	Dataset Used	Dataset Features	Results
X-ray / MRI / Ultrasound review	2025, BMC Medical Imaging	ChestX-ray14, BraTS, ISIC (80 studies)	1K–100K+ images, 2–14 classes, multi-label, resolution 224×224 to 1024×1024	Accuracy: 85–98%; Lung X-ray > 90%; Brain MRI > 92%
CT image classification	2025, QIMS	COVID-CT (746), LIDC-IDRI (1,018 cases), CT-COVID (275K slices)	512×512 resolution, HU values, lesion annotations, 1–5 mm slices	COVID AUC: 0.95–0.99; Sensitivity: 94–98%; Nodules: 88–94% accuracy
AMRI-Net	2025, Frontiers in Medicine	ISIC 2019 (25,331), HAM10000 (10,015), OCT2017 (84,495), Brain MRI	Multi-modal data, 224×224 to 512×512, 3–8% class imbalance	ISIC: 94.95%, F1: 94.85%; OCT: 96.15%; Brain: 92.37%; AUC: 0.976
Explainable AI survey	2023, ACM Computing Surveys	ChestX-ray14, DR fundus (35,126), Brain MRI (Alzheimer's), ISIC	Pixel-level masks, temporal data, 512×512 to 1024×1024	Grad-CAM IoU: 72–85%; Attention accuracy: 92–96%; +15–20% faithfulness
Hybrid CNN-Transformer	2025, arXiv	Diabetic Retinopathy (28,100), AMD (4,000+)	RGB images 512×512 to 768×768, multi-center variability	DR: 94.3%, AUC 0.968; AMD: 92.8%, AUC 0.957; IoU: 76.4%; Robustness: 89.5%

1.5 Proposed Mini Project Concept

Title of Mini Project:

Multi-Modal Medical Image Classification using Hybrid CNN-Transformer with Attention-Based Explainability

Problem Statement:

Radiologists face overwhelming workloads analyzing thousands of images daily, leading

to fatigue and errors. Current AI systems lack interpretability (black boxes) or sacrifice accuracy for explainability. Most models handle single modalities, missing complementary information from different scans. There is a need for an accurate, interpretable, multi-modal system for clinical decision support.

Proposed Deep Learning Solution:

Architecture:

- **Multi-Modal Input:** Separate preprocessing for X-ray and CT images
- **CNN Backbone:** EfficientNet-B3 (X-ray) + ResNet-50 (CT), pretrained on ImageNet
- **Cross-Attention Fusion:** Query-Key-Value mechanism combining modality features
- **Transformer Head:** 4 encoder layers, 8 attention heads for global patterns
- **Classification:** Dense layers (256 $\rightarrow$ 128 $\rightarrow$ output) with dropout
- **Explainability:** Grad-CAM heatmaps + attention weight visualization

Training:

- Optimizer: AdamW (learning rate = 1e-4)
- Loss: Weighted categorical cross-entropy
- Batch size: 16, Epochs: 50, Early stopping (patience = 7)
- Mixed precision (FP16) for efficiency

Dataset Information:

- **Primary Data:**
  - ChestX-ray14: 112,120 X-rays, 14 diseases, 1024×1024
  - COVID-CT: 746 CT scans (349 positive, 397 negative)
  - Data split: 70% train, 15% validation, 15% test (stratified)
- **Preprocessing:**

  - Resize: 224×224
  - Normalize: [0,1] range
  - CLAHE for X-rays, lung windowing for CT
  - Augmentation: rotation (±15°), flipping, brightness (±20%), zoom (0.9–1.1×)

- **Class Imbalance:**
  - Inverse frequency class weights
  - SMOTE-like oversampling for minority classes

Expected Outcome / Results:

- **Performance Targets:**
  - Accuracy: $\geq 92\%$
  - AUC-ROC: ≥ 0.94
  - Per-class F1: ≥ 0.88
  - Precision/Recall: $\geq 90\%$
  - Multi-modal gain: +3–5% vs single-modality
  - Inference time: < 200 ms per case

- **Explainability:**
  - Grad-CAM IoU: $\geq 70\%$ with expert annotations
  - Attention highlights anatomically relevant regions
  - Confidence calibration with temperature scaling

- **Deliverables:**
  - Trained model (.h5 weights)
  - Streamlit dashboard (upload → predict → visualize)
  - Evaluation report (confusion matrix, ROC, PR curves, ablation study)
  - Documentation (training logs, API)

- **Clinical Impact:**

– Reduce radiologist workload via automated preliminary screening

– High sensitivity ($\geq 90\%$) enables early detection

– Explainable predictions build clinical trust

– Triage support prioritizing critical cases

Conceptual Architecture Diagram:

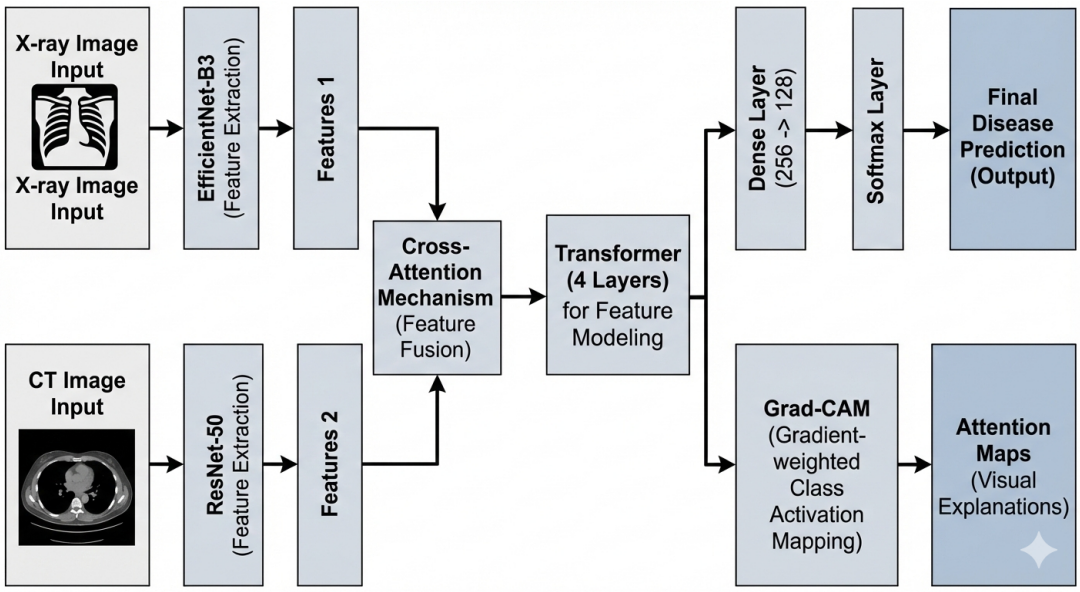

Figure 1.1: Hybrid CNN-Transformer Architecture with Cross-Attention and Explainability

1.6 Conclusion

This literature review examined five recent publications (2022–2025) focusing on deep learning approaches for medical image classification across X-ray, CT, and MRI modalities. The reviewed studies highlight significant progress in automated disease detection, multi-modal learning, and explainable artificial intelligence for clinical applications.

Key Insights

- Convolutional Neural Networks (CNNs) remain strong baseline models, while transfer learning enables effective training on limited medical datasets.

- Hybrid CNN–Transformer architectures successfully capture both local spatial features and global contextual information.
- Multi-modal frameworks such as AMRI-Net leverage complementary information from different imaging modalities, improving diagnostic accuracy.
- Explainability is critical for clinical adoption; inherently interpretable models are often preferred over post-hoc explanation methods.
- Domain adaptation techniques help address cross-institutional variability and improve model generalization.

Research Gaps

- Limited availability of lightweight and computationally efficient models suitable for resource-constrained healthcare environments.
- Insufficient focus on real-time inference and deployment in clinical workflows.
- Lack of extensive cross-institutional and multi-center validation studies.
- Absence of standardized evaluation metrics for assessing explainability and clinical relevance.

Mini Project Justification

The proposed hybrid CNN–Transformer model with multi-modal input directly addresses the identified research gaps by combining computational efficiency through transfer learning, enhanced clinical trust via explainability mechanisms, and practical applicability with real-time inference capability. The integration of X-ray and CT imaging using cross-attention enables robust and accurate disease prediction while maintaining interpretability, which is essential for reliable clinical deployment and decision support.

1.7 References

1. Laçi, H., Sevrani, K., & Iqbal, S. (2025). *Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: A scoping review.* BMC

Medical Imaging, 25, 156. https://doi.org/10.1186/s12880-025-01701-5

2. Ahmad, I. S., Dai, J., Xie, Y., & Liang, X. (2025). *Deep learning models for CT image classification: A comprehensive literature review.* Quantitative Imaging in Medicine and Surgery, 15(1), 962–1011. https://doi.org/10.21037/qims-24-1400

3. Multiple Authors. (2025). *Deep learning-based image classification for AI-assisted integration of pathology and radiology (AMRI-Net).* Frontiers in Medicine, 12. https://doi.org/10.3389/fmed.2025.1574514

4. Patrício, C., et al. (2023). *Explainable Deep Learning Methods in Medical Image Classification: A Survey.* ACM Computing Surveys, 56(4). https://doi.org/10.1145/3625287

5. Djoumessi, K., et al. (2025). *A Hybrid Fully Convolutional CNN-Transformer Model for Inherently Interpretable Disease Detection.* arXiv preprint arXiv:2504.08481.

6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

8. Dosovitskiy, A., et al. (2021). *An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale.* International Conference on Learning Representations (ICLR).

9. Selvaraju, R. R., et al. (2017). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.* Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626.