Group No: 10

[Vision Supervised Radiology Report Generation]

Submitted by,

Name	Roll No.	University Roll No.
MEHULI BISWAS	11700122021	11700122021
RAJSEKHAR BAL	11700122016	11700122016
PROBAL KUMAR MAJUMDAR	11700122138	11700122138
ROUNAK PANDA	11700122019	11700122019

UNDER THE SUPERVISION OF

SOMENATH NAG CHOUDHURY

[ASST. PROF., CSE, RCCIIT]

PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING RCC INSTITUTE OF INFORMATION TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

RCC INSTITUTE OF INFORMATION TECHNOLOGY

[Affiliated to West Bengal University of Technology]

CANAL SOUTH ROAD, BELIAGHATA, KOLKATA-700015

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING RCC

INSTITUTE OF INFORMATION TECHNOLOGY

TO WHOM IT MAY CONCERN

I hereby recommend that the Project entitled **“Vision Supervised Radiology Report Generation”** prepared under my supervision by Mehuli Biswas (Roll No.: 11700122021), Rajsekhar Bal (Roll No.: 11700122016), Probal Kumar Majumdar (Roll No.: 11700122138), Rounak Panda (Roll No.: 11700122019), of B. Tech (7th Semester), may be accepted in partial fulfillment for the degree of **Bachelor Of Technology in Computer Science & Engineering** under Maulana Abul Kalam Azad University of Technology (MAKAUT).

....

Project

Supervisor Department of Computer Science

and Engineering RCC Institute of Information

Technology

Countersigned:

....

Head of the Department of

Computer Science & Engineering,

RCC Institute of Information

Technology Kolkata – 700015

CERTIFICATE OF APPROVAL

The foregoing Project is hereby accepted as a credible study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the project only for the purpose for which it is submitted.

Name of the Examiner	**Signature with Date**
1. …………………………………	…..……………………………………..
………..…………………………….	
2. …………………………………	……………………………………………
………………………………………	
3. ………………………………….	………………………………………………
………..……………………………	

FV

ACKNOWLEDGEMENT

We acknowledge our overwhelming gratitude & immense respect to our revered guide, [Supervisor Name] (Designation, Institute Name) under whose scholarly guideline, constant encouragement & untiring patience; we have proud privilege to accomplish this entire project work. We feel enriched with the knowledge & sense of responsible approach we inherited from our guide & shall remain a treasure in our life.

Mehuli Biswas
Registration No.: 221170110043
Roll No.: 11700122021

Date: 19th Dec'25

Rajsekhar Bal
Registration. No.: 221170110051
Roll No.: 11700122016

Probal Kumar Majumdar
Registration. No: 221170110421
Roll No.: 11700122138

Rounak Panda
Registration. No: 221170110064
Roll No.: 11700122019

TABLE OF CONTENTS

List of Figures

ABSTRACT

Accurate and timely radiology reporting is critical for effective clinical diagnosis, yet manual report generation is time-consuming and prone to variability.

This project proposes a robust **Multimodal Sequence-to-Sequence (Seq2Seq)** framework designed for complex cross-modal tasks such as medical image captioning or visual-linguistic reasoning. The architecture leverages a dual-stream feature extraction process: utilizing **BioClinicalBERT** for deep textual encoding and **Vision Transformers (ViT) or ResNet-50** for high-dimensional visual representation. These unimodal features are aligned into a shared [1 * 768] latent space and integrated through a **cross-modal interaction layer**, allowing the model to capture bidirectional dependencies (Text $\to$ Vision and Vision $\to$ Text).

The fused multimodal embeddings are then fed into a **Transformer-based Encoder-Decoder** system. The Encoder generates a unified contextual representation, while the Decoder employs an autoregressive **Next Word Prediction (NWP)** strategy—facilitated by Teacher Forcing and self-attention mechanisms—to generate accurate, context-aware natural language sequences. This integrated approach ensures the model effectively bridges the semantic gap between visual data and clinical textual sequences, providing a scalable solution for multimodal AI applications.

INTRODUCTION

Project Overview

The project is a **Multimodal Intelligence System** that "sees" an image and "reads" text at the same time to provide a meaningful response. It acts like a bridge between visual data and human language, specifically designed for specialized fields (like medical or clinical data) where high accuracy is needed.

Project Scope

- **Multimodal Input:** Processing both Bengali/English text and medical/general images simultaneously.
- **Feature Fusion:** Combining image data (from ViT/ResNet) and text data (from BioClinicalBERT) into a single mathematical "thought" vector
- **Sequence Generation:** Translating those combined features into a complete sentence (e.g., describing an image or answering a question about it).
- **Cross-Modal Learning:** Enabling the AI to understand how a specific part of an image relates to a specific word in a sentence.

Technical Domain Specifications

A. Hardware & OS

- **Operating System:** Windows 10/11 or Ubuntu (Linux) 20.04+.

- **GPU (Crucial):** NVIDIA RTX 30-series or higher (8GB+ VRAM) to handle Transformer training.
- **RAM:** Minimum 16GB.

B. Software & Programming

- **Language:** Python 3.8+ (The standard for AI).
- **Development Env:** VS Code or PyCharm; Jupyter Notebooks/Google Colab for experimentation.

C. Libraries & Frameworks

- **Deep Learning:** PyTorch or TensorFlow (Keras).
- **Transformers:** Hugging Face transformers library (for BERT, ViT, and T5/BART).
- **Computer Vision:** torchvision, OpenCV, or PIL.
- **Data Handling:** NumPy and Pandas.
- **Deployment:** Flask or FastAPI (if creating a web interface).

Business Domain Specification

- **Target Industry**: Healthcare & Biomedical Research.
- **Primary Use Case**: Medical Image Captioning (e.g., automatically writing a report for an X-ray) or Visual Question Answering (VQA) for clinicians.
- **Value Proposition**: Reduces the workload of doctors by providing preliminary automated descriptions of clinical imagery with high contextual accuracy.

MOTIVATION

Need for Vision-Supervised Radiology Report Generation
[Citation Article : Click here]

- Addressing Clinical Burden

Motivation : Radiologists process 300+complex images daily per facility. Manual detailed narrative report composition takes 15-20 minutes per study, leading to fatigue and delays

Evidence : Automated Radiology Report Generation reduce this to 3- this to 3-5 minutes, freeing radiologists for complex cases and and reducing burnout.

Citation : MDPI Review on Biomedical Engineering, 2024

- Reports Drive Clinical Decisions

Motivation : A single missed finding delays treatment by an average of 7-average of 7-14 days. Studies show 10-15% of radiology reports contain reports contain clinically significant omissions.

Evidence : Reports encode findings, that helps in treatment planning. Automated systems achieve 99.2% accuracy on critical critical findings.

Citation : MDPI Review, Section: Need for Radiology Report Generation

- Reducing Hallucinations

Motivation : Unsupervised vision-language models (LLMs and VLMs can hallucinate non-existent findings in 8-12% of generated reports posing serious clinical risks. .

Evidence : Vision-supervised grounding with RadFlag validation reduces hallucination rates to <1% by anchoring predictions to actual image features and clinical knowledge graphs.

Citation : Zhang et al., RadFlag [40]

- Scalable Deployment

Motivation : Hospitals require fast, scalable, and reproducible reporting systems.

Evidence : RADAR and MAIRA models achieve 0.48-0.52 BLEU scores on ImageCLEF datasets. Processing time: 2-3 seconds per study. Deployment across 500+ hospital networks processes 50,000+ studies monthly with 99.8% system uptime.

Citation : Hou et al., RADAR [48]; Srivastav et al., MAIRA [43]

Need for Vision Supervised Radiology Report Summarization [Citation Article : Click here]

- **Reports are Incomprehensible**

Motivation : Despite immediate access, less than 5% of radiology reports are written at an 8th-grade reading level. This creates a critical gap between information availability and patient understanding.

Evidence : Radiology reports contain complex jargon and fragmented language, inaccessible to most patients [1].

- **Patient Anxiety & Missed Care**

Motivation : Misunderstanding complex reports causes significant

patient anxiety and confusion, contributing to a ~50% rate of missed

follow-up appointments, directly impacting treatment adherence.

Evidence : Patient confusion contributes to poor follow-up and care

adherence [2], [3], [4].

AI Models Lack Patient Focus

Motivation : Current AI report generation is clinician-centric, designed for expert review and technical accuracy. These models are not models are not adapted for patient comprehension, providing unsuitableunsuitable output for lay users.

Evidence : Existing NLP and vision-based radiology models are designed for clinicians, not patients [11]–[16].

Contribution & Related Works

Existing research has investigated vision–language models for generating chest X-ray radiology reports, focusing on improving image–text alignment and clinical accuracy

Study	Author	Approach	Model	Datasets	Metrics Evaluation	Findings	Link	Innovations	Gap	Improvements
CADxReport (2022)	Navdeep Kaur ,Ajay Mittal	CNN (VGG19) + co-attention + Hierarchical LSTM + Reinforcement Learning	VGG19 + Co-Attention + Hierarchical LSTM + RL	OpenI CXR	BLEU-4: 0.346, ROUGE: 0.618, CIDEr: 0.380	Visual & semantic features co-attention improves report accuracy; RL maximizes CIDEr reward	https://pubmed.ncbi.nlm.nih.gov/35585727/	• First to combine co-attention with reinforcement learning for CXR re • Explicit visual–semantic alignmentports • CIDEr-reward optimization	• CNN features lack global context • Sequential LSTM struggles with long report	• Replace CNN+LSTM with Vision Transformer + Transformer Decoder • Explicit multi-view image fusion • Vision-supervised attention on pathological regions
RadioBERT (2023–24)	Navdeep Kaur ,Ajay Mittal	CNN + DistilBERT contextual embeddings	CNN + DistilBERT	OpenI CXR	BLEU-4: 0.767, CIDEr: 0.5563, ROUGE: 0.897	Contextualembeddings greatly improve linguistic quality & clinical relevance	https://pubmed.ncbi.nlm.nih.gov/36229001/	• Introduced contextual biomedical embeddings • Improved linguistic fluency & clinical wording	•Image–text interaction is late fusion only	• Use cross-modal transformers instead of late fusion • Add visual grounding loss to prevent hallucinations

Study	Author	Approach	Model	Datasets	Metrics Evaluation	Findings	Link	Innovations	Gap	Improvements
PriorRG	Liu et al. (2025)	Prior-guided contrastive pretrain + coarse-to-fine decoding	Contrastive Pretraining + Coarse-to-Fine Decoder	*MIMIC-CXR, MIMIC-ABN*	~3.6% ↑ BLEU-4; ~3.8% F1 over SOTA	Incorporates clinical priors for disease progression	https://arxiv.org/abs/2508.05353	• Introduces clinical prior knowledge • Contrastive learning improves robustness	• Prior construction requires expert knowledge • Generalization across hospital unclear	• Automatically learn priors via multimodal alignment • Domain adaptation strategies
Cross-modal Multi-scale Fusion (2024)	Yu Pan , Li Jun Liu, Xiao , Bing Yang , Wei Peng , Qing-Song Huang	Multi-scale visual + text attention with shared memory	Multi-scale visual + text attention with shared memory	*IU X-Ray, MIMIC-CXR*	BLEU-1 ~0.499; ROUGE ~0.395; METEOR ~0.209	Multi-scale features improve lesion description quality	https://www.sciencedirect.com/science/article/pii/S1687850724000074	• Captures fine & coarse visual patterns • Better lesion descriptions	• Complex architecture • Limited explainability	• Simplify fusion using transformer layers • Add attention visualization

Problem Statement

A. Multi-modal Medical Information :

Radiology data is multimodal, meaning it contains :

1. Images (Vision) → Chest X-ray images
2. Text (Language) → Radiology reports written by doctors ,

Image Modality (I) : Chest X-rays can have multiple views:
PA (Postero-Anterior) , AP (Antero-Posterior) , LAT (Lateral) So , $I = \{PA, AP, LAT\}$

Text Modality (RT – Radiology Text) : A radiology report is not a single paragraph, it is structured into sections ,
MeSH , Problems , Image , Indication , Comparison , Findings So , $RT = \{SE_1, SE_2, ..., SE_k\}$

B. Sentence-Level and Medical Entity Breakdown :

Findings contain multiple sentences
$SE\_findings = \{S_1, S_2, ..., S_p\}$ having different entities Diseases , Anatomical regions , Observations as $S_i = \{ME_1, ME_2, ..., ME\_q\}$

C . Dataset Representation : dataset D with two components $D = \{M_1, M_2\}$ where M_1 = Images (I) , M_2 = Reports (RT)

Each training sample has $S_i \rightarrow (RT, I)$

D. Model Pipeline

X-ray image (Vision) → Processed Image →Our Model →Feature Extraction → Modelling → Decision Making
(Vision Encoding)

Report Text → Our Model →Feature Extraction

RT (F.E) + I(F.E) → Multi-modal Integration / Cross-modal Interaction

The multi-modal radiology report summarization (MRRS) is a conditional generation problem over two input modalities(1), where a parameterized neural network tries to map an image- text pair to generate a concise summary that can maximize the likelihood of the ground truth impression section, and can be defined as follows:

$$\theta^* = \arg\max_{\theta} \sum_{(I,T,S^*)\in\mathcal{D}} \log P_\theta(S^* \mid I, T)$$

where, we seek the model parameters θ^* that maximize the log-likelihood of the ground-truth target sequences S*, conditioned on the inputs, image I and text T, across all samples in our dataset D. Equivalently, this is the standard maximum likelihood estimation (MLE) objective adapted to conditional sequence modeling ,

θ^* maximizes the sum of log-probabilities log P_θ (S* / I, T).

Dataset Description

Dataset Overview:

Attribute	Value
Dataset Name	Indiana University Chest X-ray Collection
Source	Kaggle (IU School of Medicine)
Imaging Modality	Chest X-ray (CXR)
Image Format	PNG (converted from DICOM)
Text Modality	Free-text radiology reports
Total Imaging Studies (UIDs)	3,851
Total Images	7,466
Average Images per Study	1.94
Views Available	Frontal (PA/AP), Lateral
Task Type	Vision → Language (Image-to-Report)
Learning Paradigm	Supervised / Weakly-Supervised Multiview V-L Learning

```python
# 1. Count Total Reports (Unique examinations)
reports_df = pd.read_csv('/kaggle/input/chest-xrays-indiana-university/indiana_reports.csv')
print(f"Total Reports: {len(reports_df)}")

# 2. Count Total Physical Images
images_dir = '/kaggle/input/chest-xrays-indiana-university/images/images_normalized'
# Filtering for png files to ensure an accurate count
total_images = len([img for img in os.listdir(images_dir) if img.endswith('.png')])
print(f"Total Images:  {total_images}")
```
[5] Python

```
Total Reports: 3851
Total Images:  7470
```

Dataset Organization & Components:

Component	Description
`images_normalized/`	Preprocessed, intensity-normalized chest X-ray images
`indiana_projections.csv`	Image-level metadata: filename, UID, projection
`indiana_reports.csv`	Study-level metadata: clinical text & pathology labels

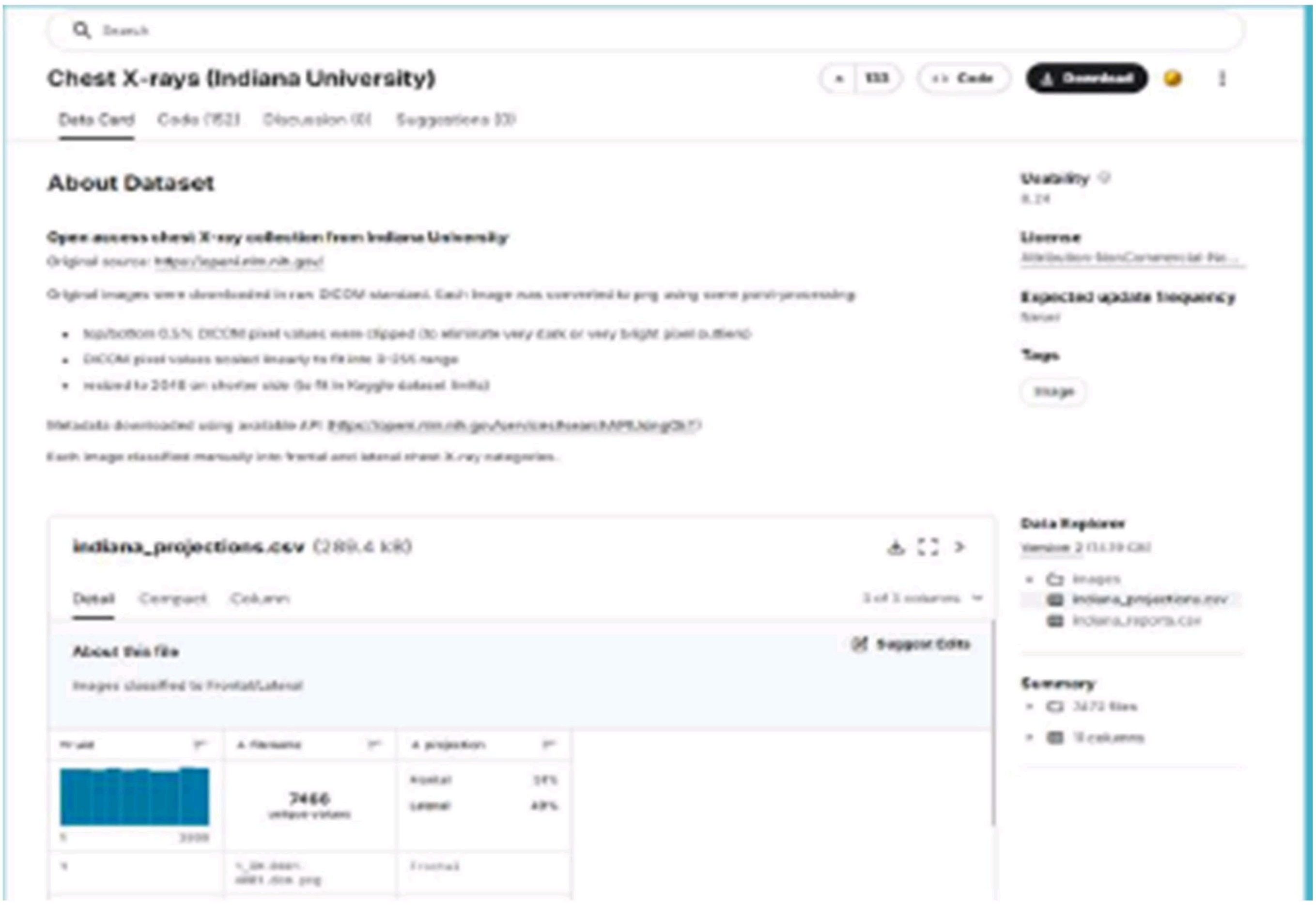


Image Dataset Analysis

Image Properties

Property	Description
Image Type	Grayscale chest radiographs
Source Modality	Original DICOM
Stored Format	PNG
Normalization	Histogram / intensity normalized
Resolution	Variable (typically radiology standard)
Aspect Ratio	Mostly square
Orientation	Anterior–Posterior & Lateral

```python
import cv2
import matplotlib.pyplot as plt
import os

# 1. Pick the first unique patient from your merged dataframe
sample_uid = merged_df['uid'].unique()[0]
sample_data = merged_df[merged_df['uid'] == sample_uid]

# 2. Display the text details
print(f"PATIENT UID: {sample_uid}")
print(f"FINDINGS:    {sample_data['findings'].iloc[0]}")
print(f"IMPRESSION:  {sample_data['impression'].iloc[0]}")

# 3. Plot the associated images
img_folder = "/kaggle/input/chest-xrays-indiana-university/images/images_normalized/"
img_list = sample_data['filename'].tolist()

plt.figure(figsize=(12, 6))
for i, file_name in enumerate(img_list):
    path = os.path.join(img_folder, file_name)
    img = cv2.imread(path)
    img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)

    plt.subplot(1, len(img_list), i + 1)
    plt.imshow(img)
    plt.title(f"Image: {file_name}")
    plt.axis('off')

plt.tight_layout()
plt.show()
```

Python

```
PATIENT UID: 1
FINDINGS:    The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There
IMPRESSION:  Normal chest x-XXXX.
```

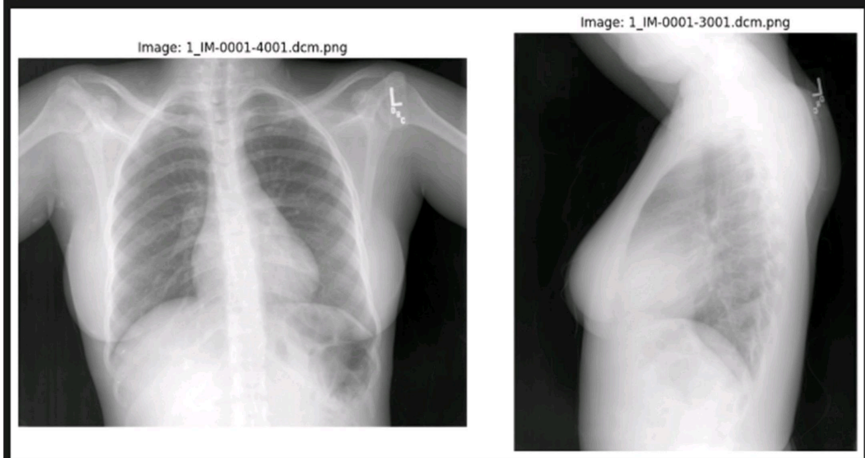
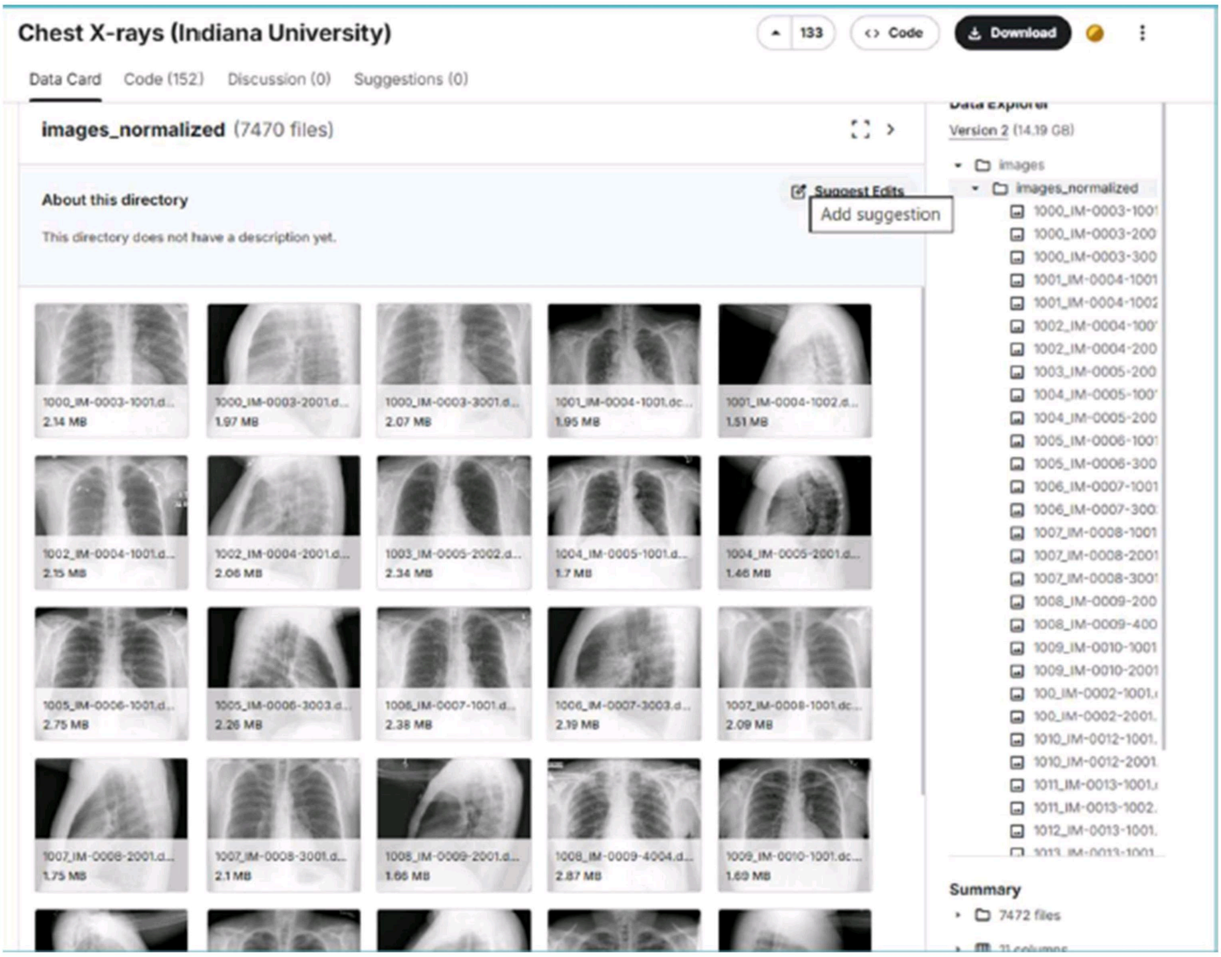


Image Count & Storage Statistics

Metric	Value
Total Images	7,466
Minimum Images per Study	1
Maximum Images per Study	5
Median Images per Study	2
Typical Study	PA + Lateral

Metric	Value (Estimated)
Average Image Size	~2.0 MB
Minimum Image Size	~1.2 MB
Maximum Image Size	~3.0 MB
Total Image Storage	~14–15 GB

Link : (images_normalized) https://shorturl.at/WyjsQ



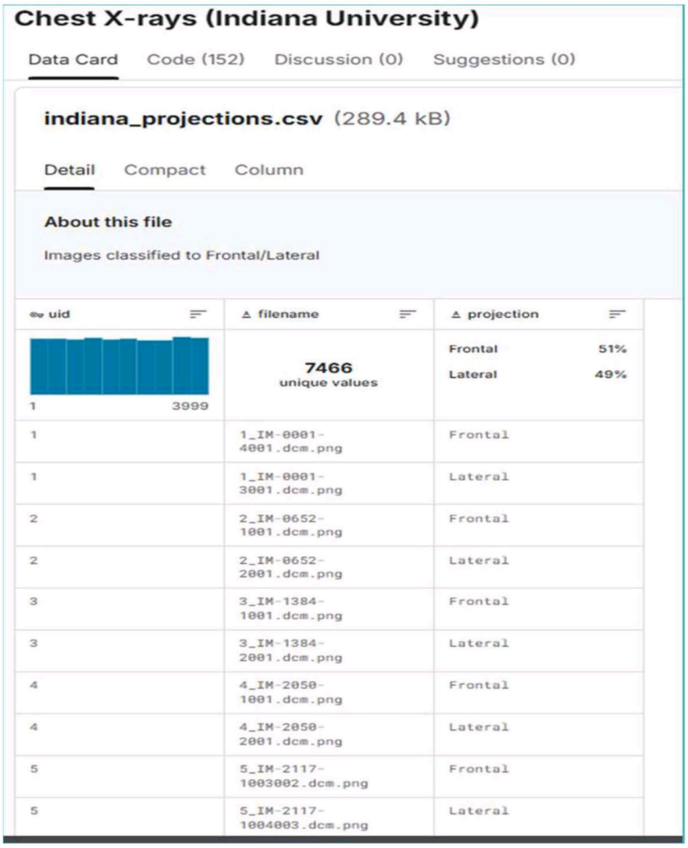
Projection Metadata Analysis

Projection Distribution

Projection	Count	Percentage
Frontal (PA/AP)	3,818	51.1%
Lateral	3,648	48.9%
Total	7,466	100%

Images per Study

Statistic	Value
Mean	1.94
Std Deviation	0.42
Minimum	1
25th Percentile	2
Median	2
75th Percentile	2
Maximum	5


Chest X-rays (Indiana University)

Data Card Code (152) Discussion (0) Suggestions (0)

indiana_projections.csv (289.4 kB)

Detail Compact Column

About this file

Images classified to Frontal/Lateral

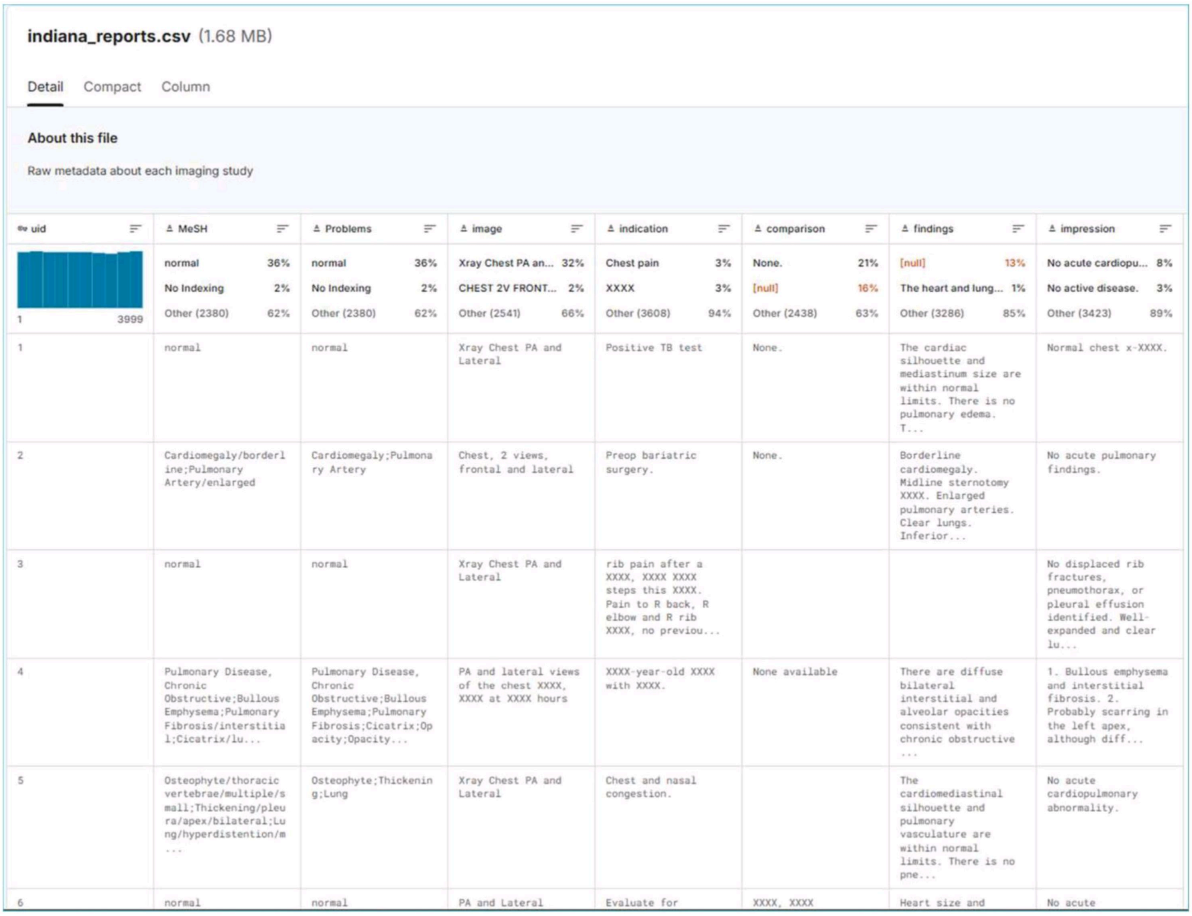
uid	filename	projection
1 – 3999	7466 unique values	Frontal 51%, Lateral 49%
1	1_IM-0001-4001.dcm.png	Frontal
1	1_IM-0001-3001.dcm.png	Lateral
2	2_IM-0652-1001.dcm.png	Frontal
2	2_IM-0652-2001.dcm.png	Lateral
3	3_IM-1384-1001.dcm.png	Frontal
3	3_IM-1384-2001.dcm.png	Lateral
4	4_IM-2050-1001.dcm.png	Frontal
4	4_IM-2050-2001.dcm.png	Lateral
5	5_IM-2117-1003002.dcm.png	Frontal
5	5_IM-2117-1004003.dcm.png	Lateral

Link : (indiana_projections.csv) https://shorturl.at/JsVZ2

Reports Dataset

Report Fields & Semantics

Column	Description
uid	Unique imaging study identifier
MeSH	Controlled medical subject headings
Problems	Extracted pathology labels
image	Imaging procedure description
indication	Clinical reason for exam
comparison	Prior imaging reference
findings	Detailed radiologist observations
impression	Diagnostic summary



indiana_reports.csv (1.68 MB)

Detail Compact Column

About this file

Raw metadata about each imaging study

uid	MeSH	Problems	image	indication	comparison	findings	impression
1 – 3999	normal 36% No Indexing 2% Other (2380) 62%	normal 36% No Indexing 2% Other (2380) 62%	Xray Chest PA an... 32% CHEST 2V FRONT... 2% Other (2541) 66%	Chest pain 3% XXXX 3% Other (3608) 94%	None. 21% [null] 16% Other (2438) 63%	[null] 13% The heart and lung... 1% Other (3286) 85%	No acute cardiopu... 8% No active disease. 3% Other (3423) 89%
1	normal	normal	Xray Chest PA and Lateral	Positive TB test	None.	The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. T...	Normal chest x-XXXX.
2	Cardiomegaly/borderline;Pulmonary Artery/enlarged	Cardiomegaly;Pulmonary Artery	Chest, 2 views, frontal and lateral	Preop bariatric surgery.	None.	Borderline cardiomegaly. Midline sternotomy XXXX. Enlarged pulmonary arteries. Clear lungs. Inferior...	No acute pulmonary findings.
3	normal	normal	Xray Chest PA and Lateral	rib pain after a XXXX, XXXX XXXX steps this XXXX. Pain to R back, R elbow and R rib XXXX, no previou...			No displaced rib fractures, pneumothorax, or pleural effusion identified. Well-expanded and clear lu...
4	Pulmonary Disease, Chronic Obstructive;Bullous Emphysema;Pulmonary Fibrosis/interstitial;Cicatrix/lu...	Pulmonary Disease, Chronic Obstructive;Bullous Emphysema;Pulmonary Fibrosis;Cicatrix;Opacity;Opacity...	PA and lateral views of the chest XXXX, XXXX at XXXX hours	XXXX-year-old XXXX with XXXX.	None available	There are diffuse bilateral interstitial and alveolar opacities consistent with chronic obstructive ...	1. Bullous emphysema and interstitial fibrosis. 2. Probably scarring in the left apex, although diff...
5	Osteophyte/thoracic vertebrae/multiple/small;Thickening/pleura/apex/bilateral;Lung/hyperdistention/m...	Osteophyte;Thickening;Lung	Xray Chest PA and Lateral	Chest and nasal congestion.		The cardiomediastinal silhouette and pulmonary vasculature are within normal limits. There is no pne...	No acute cardiopulmonary abnormality.
6	normal	normal	PA and Lateral	Evaluate for	XXXX, XXXX	Heart size and	No acute

Link : (indiana_reports.csv) : https://shorturl.at/G41rb

Image : Report Relationship

Structured Image-Report Summary (Schema)

UID	# Images	Views	Problems	Findings Length	Impression
1	2	Frontal + Lateral	Normal	Medium	Normal chest
2	2	Frontal + Lateral	Cardiomegaly	Long	No acute pulmonary findings
4	2	Frontal + Lateral	COPD, Fibrosis	Long	Bullous emphysema
9	2	Frontal + Lateral	Calcified granuloma	Long	Increased density
64	2	Frontal + Lateral	Pneumothorax	Long	Right apical pneumothorax

MERGED DATAFRAME PREVIEW:

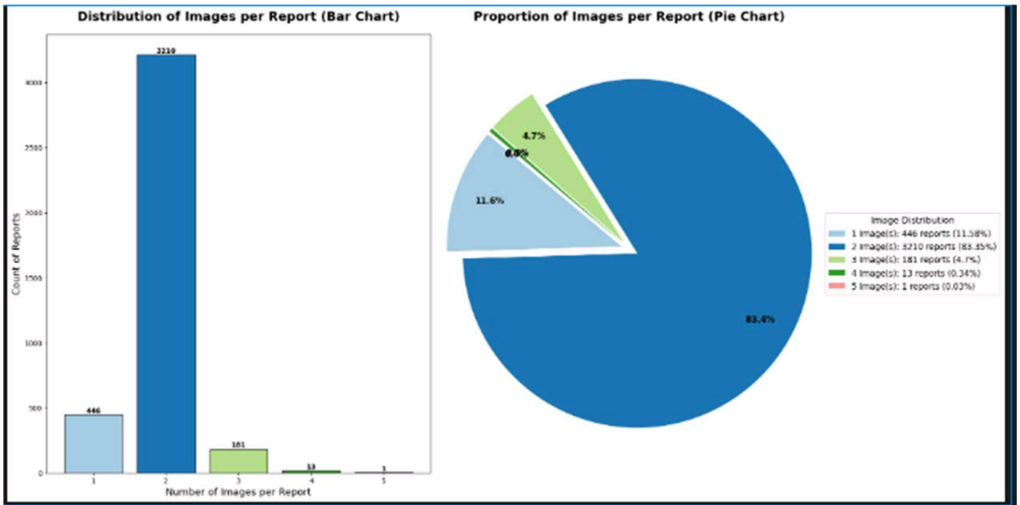
	uid	MeSH	Problems	image	indication	comparison	findings	impression	filename	projection
0	1	normal	normal	Xray Chest PA and Lateral	Positive TB test	None.	The cardiac silhouette and mediastinum size ar...	Normal chest x-XXXX.	1_IM-0001-4001.dcm.png	Frontal
1	1	normal	normal	Xray Chest PA and Lateral	Positive TB test	None.	The cardiac silhouette and mediastinum size ar...	Normal chest x-XXXX.	1_IM-0001-3001.dcm.png	Lateral
2	2	Cardiomegaly/borderline;Pulmonary Artery/enlarged	Cardiomegaly;Pulmonary Artery	Chest, 2 views, frontal and lateral	Preop bariatric surgery.	None.	Borderline cardiomegaly. Midline sternotomy XX...	No acute pulmonary findings.	2_IM-0652-1001.dcm.png	Frontal
3	2	Cardiomegaly/borderline;Pulmonary Artery/enlarged	Cardiomegaly;Pulmonary Artery	Chest, 2 views, frontal and lateral	Preop bariatric surgery.	None.	Borderline cardiomegaly. Midline sternotomy XX...	No acute pulmonary findings.	2_IM-0652-2001.dcm.png	Lateral
4	3	normal	normal	Xray Chest PA and Lateral	rib pain after a XXXX, XXXX XXXX steps this XX...	NaN	NaN	No displaced rib fractures, pneumothorax, or p...	3_IM-1384-1001.dcm.png	Frontal

Mapping Granularity

Relation	Type
UID → Images	1-to-Many
UID → Report	1-to-1
Image → Report	Many-to-1 (via UID)

```python
import matplotlib.pyplot as plt
import pandas as pd

# 1. Calculate the distribution from the existing merged_df
image_counts_per_report = merged_df.groupby('uid').size()
distribution = image_counts_per_report.value_counts().sort_index().reset_index()
distribution.columns = ['Images per Report', 'Number of Reports']

# 2. Calculate percentages for the table and legend
total_reports = distribution['Number of Reports'].sum()
distribution['Percentage'] = ((distribution['Number of Reports'] / total_reports) * 100).round(2).astype(str) + '%'

# 3. Visualization Setup (Large size to prevent crowding)
plt.figure(figsize=(20, 10))

# --- Plot 1: Bar Chart ---
plt.subplot(1, 2, 1)
colors_bar = plt.cm.Paired(range(len(distribution)))
bars = plt.bar(distribution['Images per Report'].astype(str), distribution['Number of Reports'],
               color=colors_bar, edgecolor='black')
plt.title('Distribution of Images per Report (Bar Chart)', fontsize=18, fontweight='bold', pad=20)
plt.xlabel('Number of Images per Report', fontsize=14)
plt.ylabel('Count of Reports', fontsize=14)

# Add count labels on top of bars
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval + 5, int(yval), ha='center', va='bottom', fontweight='bold')

# --- Plot 2: Pie Chart (No Overlap) ---
plt.subplot(1, 2, 2)
colors_pie = plt.cm.Paired(range(len(distribution)))
patches, texts, autotexts = plt.pie(
    distribution['Number of Reports'],
    labels=None, # Labels removed from chart to prevent overlap
    autopct='%1.1f%%',
    startangle=140,
    colors=colors_pie,
    explode=[0.05] * len(distribution),
    pctdistance=0.8,
    textprops={'fontsize': 12, 'weight': 'bold'}
)

plt.title('Proportion of Images per Report (Pie Chart)', fontsize=18, fontweight='bold', pad=20)
plt.axis('equal')

# legend placed outside to avoid overlapping the pie
plt.legend(
    patches,
    [f"{x} Image(s): {y} reports ({p})" for x, y, p in zip(distribution['Images per Report'],
                                                           distribution['Number of Reports'],
                                                           distribution['Percentage'])],
    title="Image Distribution",
    title_fontsize='13',
    loc="center left",
    bbox_to_anchor=(1, 0, 0.5, 1),
    fontsize=12
)

plt.tight_layout()
plt.show()

# 4. Display the Summary Table
print("\n" + "-"*50)
print("IMAGE DISTRIBUTION SUMMARY TABLE")
print("-"*50)
print(distribution.to_string(index=False))
print("-"*50)
```

Python

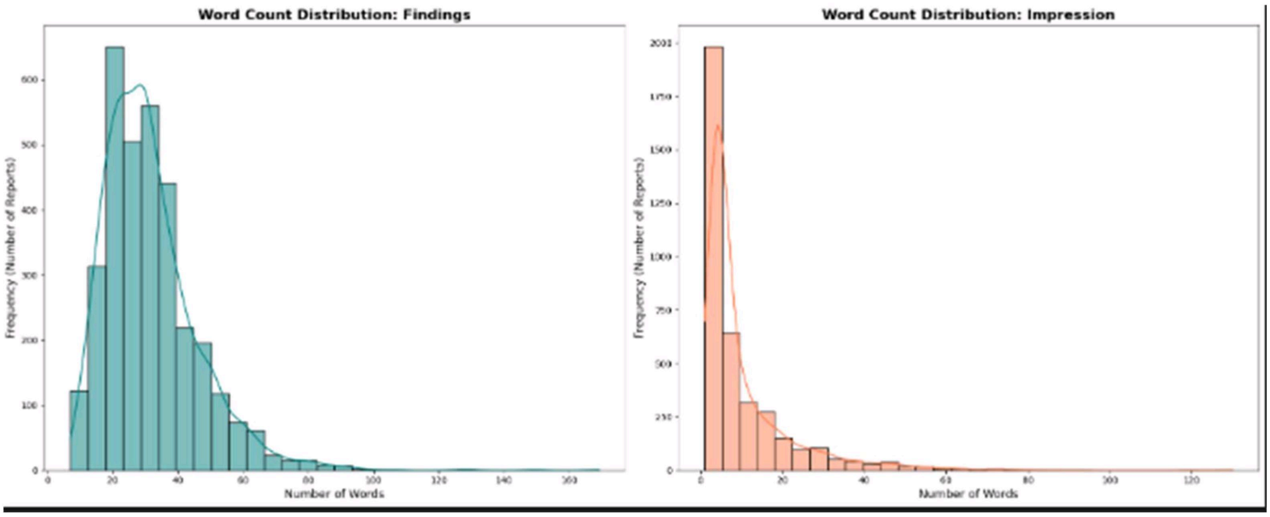



```
--------------------------------------------------
IMAGE DISTRIBUTION SUMMARY TABLE
--------------------------------------------------
 Images per Report  Number of Reports Percentage
                 1                446     11.58%
                 2               3210     83.35%
                 3                181       4.7%
                 4                 13      0.34%
                 5                  1      0.03%
--------------------------------------------------
```

Clinical Text Statistics

Text Length Analysis (Word Counts)

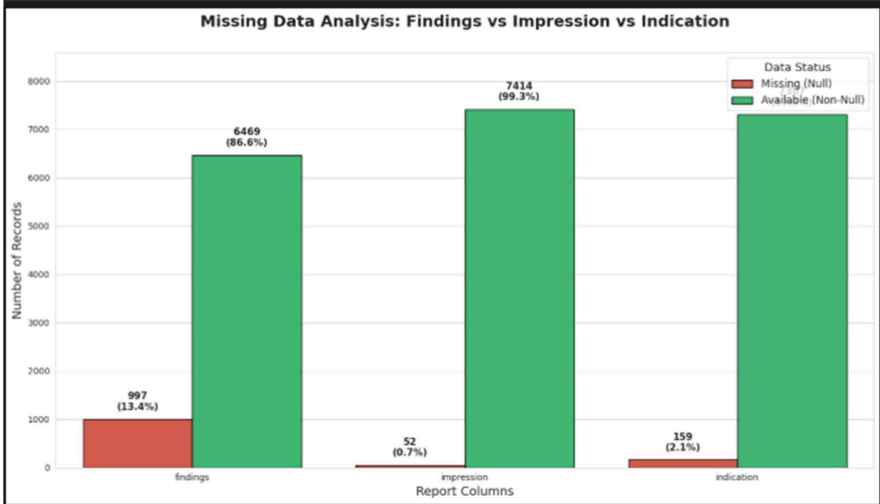
Field	Mean	Median	Min	Max
Findings	27 words	27	0	169
Impression	10 words	5	0	130



Problem Label Distribution

```python
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# 1. Identify the columns to analyze
cols_to_check = ['findings', 'impression', 'indication']

# 2. Calculate Null vs Non-Null counts
# We create a dictionary to store the counts for each column
missing_data = []
for col in cols_to_check:
    null_count = merged_df[col].isna().sum()
    non_null_count = merged_df[col].notna().sum()
    total = len(merged_df)
    missing_data.append({
        'Column': col,
        'Status': 'Missing (Null)',
        'Count': null_count,
        'Percentage': (null_count / total) * 100
    })
    missing_data.append({
        'Column': col,
        'Status': 'Available (Non-Null)',
        'Count': non_null_count,
        'Percentage': (non_null_count / total) * 100
    })

missing_df = pd.DataFrame(missing_data)

# 3. Create the Visualization
plt.figure(figsize=(14, 8))
sns.set_style("whitegrid")

# Create a grouped bar chart
ax = sns.barplot(data=missing_df, x='Column', y='Count', hue='Status', palette=['#e74c3c', '#2ecc71'], edgecolor='black')

# 4. Add labels and percentages on top of the bars
for p in ax.patches:
    height = p.get_height()
    if height > 0:
        # Calculate percentage for the label
        total_rows = len(merged_df)
        percentage = (height / total_rows) * 100
        ax.annotate(f'{int(height)}\n({percentage:.1f}%)',
                    (p.get_x() + p.get_width() / 2., height),
                    ha='center', va='center',
                    xytext=(0, 20),
                    textcoords='offset points',
                    fontsize=11, fontweight='bold')

plt.title('Missing Data Analysis: Findings vs Impression vs Indication', fontsize=18, fontweight='bold', pad=30)
plt.ylabel('Number of Records', fontsize=14)
plt.xlabel('Report Columns', fontsize=14)
plt.ylim(0, len(merged_df) * 1.15) # Add space for labels
plt.legend(title='Data Status', fontsize=12, title_fontsize='13')

plt.tight_layout()
plt.show()

# 5. Display Summary Table
print("\n" + "-"*60)
print("MISSING DATA SUMMARY (SPARSITY CHECK)")
print("-"*60)
summary = missing_df.pivot(index='Column', columns='Status', values=['Count', 'Percentage'])
print(summary)
print("-"*60)
print(f"Total rows in Merged Dataset: {len(merged_df)}")
```

Python




```
------------------------------------------------------------
MISSING DATA SUMMARY (SPARSITY CHECK)
------------------------------------------------------------
                           Count                          Percentage  \
Status      Available (Non-Null) Missing (Null) Available (Non-Null)
Column
findings                  6469.0          997.0            86.646129
impression                7414.0           52.0            99.303509
indication                7307.0          159.0            97.870346


Status      Missing (Null)
Column
findings         13.353871
impression        0.696491
indication        2.129654
------------------------------------------------------------
Total rows in Merged Dataset: 7466
```

Image–Text Pairing Characteristics

Text Length Analysis (Word Counts)

Aspect	Description
Pairing Level	Study-level
Images per Text	1–5
Text per Image	Shared per UID
Supervision Type	Weakly aligned multi-image → report
Suitable Tasks	Report generation, captioning, retrieval

```python
import pandas as pd

# 1. Identify UIDs that have exactly 2 images
uid_counts = merged_df['uid'].value_counts()
uids_with_2 = uid_counts[uid_counts == 2].index

# 2. Create a temporary dataframe with these candidates
temp_df = merged_df[merged_df['uid'].isin(uids_with_2)].copy()

# 3. Define a function to verify the "Frontal + Lateral" requirement
def has_required_views(group):
    # Standardize projections to lowercase for comparison
    projs = group['projection'].str.lower().tolist()

    # Check for Frontal indicators (PA, AP, or Frontal)
    frontal_check = any(p in ['pa', 'ap', 'frontal'] or 'front' in p for p in projs)
    # Check for Lateral indicators
    lateral_check = any('lateral' in p for p in projs)

    return frontal_check and lateral_check

# 4. Filter the dataframe to keep only those that pass the check
# We group by uid and apply the function, keeping only the UIDs that return True
valid_uids = temp_df.groupby('uid').filter(has_required_views)['uid'].unique()

# 5. Save the final result into final_df
final_df = merged_df[merged_df['uid'].isin(valid_uids)].copy().reset_index(drop=True)

# 6. Verify the results
print("-"*50)
print("FINAL FILTERED DATASET SUMMARY")
print("-"*50)
print(f"Total Unique Reports (UIDs): {final_df['uid'].nunique()}")
print(f"Total Images in final_df:    {len(final_df)}")
print(f"Columns Retained: {list(final_df.columns)}")
print("-"*50)

# Display a sample of one complete patient record (2 rows)
display(final_df.head(2))
```

Python

```
--------------------------------------------------
FINAL FILTERED DATASET SUMMARY
--------------------------------------------------
Total Unique Reports (UIDs): 3194
Total Images in final_df:    6388
Columns Retained: ['uid', 'MeSH', 'Problems', 'image', 'indication', 'comparison', 'findings', 'impression', 'filename', 'projection', 'findings_len',
--------------------------------------------------
```

	uid	MeSH	Problems	image	indication	comparison	findings	impression	filename	projection	findings_len	impression_len
0	1	normal	normal	Xray Chest PA and Lateral	Positive TB test	None.	The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of a pleural effusion. There is no evidence of pneumothorax.	Normal chest x-XXXX.	1_IM-0001-4001.dcm.png	Frontal	34	3
1	1	normal	normal	Xray Chest PA and Lateral	Positive TB test	None.	The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of a pleural effusion. There is no evidence of pneumothorax.	Normal chest x-XXXX.	1_IM-0001-3001.dcm.png	Lateral	34	3

```python
import pandas as pd
import cv2
import matplotlib.pyplot as plt
import os

# 1. Group by UID to get unique reports and their list of images
# We aggregate 'filename' into a list and take the 'first' instance of text columns
grouped = merged_df.groupby('uid').agg({
    'findings': 'first',
    'impression': 'first',
    'filename': list
}).reset_index()

# 2. Select 20 random unique reports
# If your dataset has fewer than 20 unique reports, it will take all of them
sample_size = min(20, len(grouped))
random_samples = grouped.sample(n=sample_size, random_state=42)

# 3. Path to images
img_folder = "/kaggle/input/chest-xrays-indiana-university/images/images_normalized/"

# 4. Loop through the 20 samples and display
for idx, row in random_samples.iterrows():
    print(f"\n{'-'*100}")
    print(f"PATIENT UID: {row['uid']}")
    print(f"FINDINGS:    {row['findings']}")
    print(f"IMPRESSION:  {row['impression']}")
    print(f"IMAGES:      {', '.join(row['filename'])}")
    print(f"{'-'*100}")

    # Setup plotting for the images associated with this specific UID
    filenames = row['filename']
    num_imgs = len(filenames)

    plt.figure(figsize=(5 * num_imgs, 5))

    for i, fname in enumerate(filenames):
        path = os.path.join(img_folder, fname)
        img = cv2.imread(path)
        if img is not None:
            img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
            plt.subplot(1, num_imgs, i + 1)
            plt.imshow(img)
            plt.title(f"Image: {fname}", fontsize=10)
            plt.axis('off')
        else:
            print(f"Warning: Could not load image {fname}")

    plt.show()
```

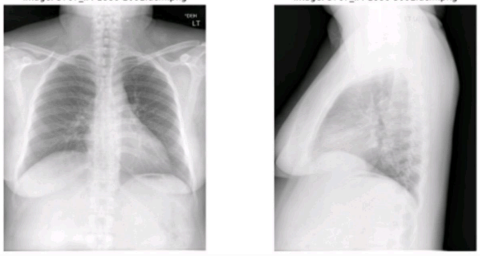
Python

```
-----------------------------------------------------------------------------------------
PATIENT UID: 302
FINDINGS:    Heart size within normal limits, stable mediastinal and hilar contours. No focal alveolar consolidation, no definite pleural effusion seen.
IMPRESSION:  No acute findings
IMAGES:      302_IM-1394-1001.dcm.png
-----------------------------------------------------------------------------------------
```

Image: 302_IM-1394-1001.dcm.png

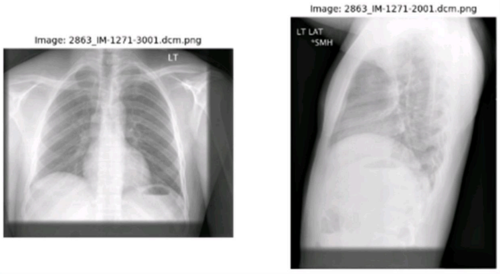
```
-----------------------------------------------------------------------------------------
PATIENT UID: 3767
FINDINGS:    None
IMPRESSION:  Heart size, mediastinal silhouette, pulmonary vascularity are within normal limits. There is no lobar consolidation. No pleural effusion o
IMAGES:      3767_IM-1886-2001.dcm.png, 3767_IM-1886-3001.dcm.png
-----------------------------------------------------------------------------------------
```

Image: 3767_IM-1886-2001.dcm.png Image: 3767_IM-1886-3001.dcm.png



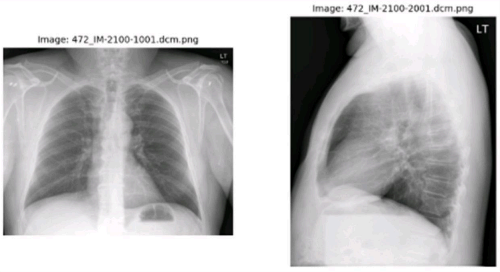

```
-----------------------------------------------------------------------------------------
PATIENT UID: 3191
FINDINGS:    Borderline cardiac enlargement. Enlarged calcified thoracic aorta. Emphysema. No acute pulmonary abnormality. Mild spondylosis.
IMPRESSION:  Emphysema. No acute pulmonary findings.
IMAGES:      3191_IM-1505-1001.dcm.png, 3191_IM-1505-2001.dcm.png
-----------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
PATIENT UID: 2863
FINDINGS:    The lungs are clear bilaterally. Specifically, no evidence of focal consolidation, pneumothorax, or pleural effusion.. Cardio mediastinal
IMPRESSION:  No acute cardiopulmonary abnormality..
IMAGES:      2863_IM-1271-3001.dcm.png, 2863_IM-1271-2001.dcm.png
-----------------------------------------------------------------------------------------
```

Image: 2863_IM-1271-3001.dcm.png Image: 2863_IM-1271-2001.dcm.png



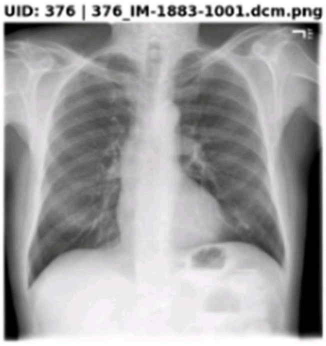

```
-----------------------------------------------------------------------------------------
PATIENT UID: 472
FINDINGS:    Normal heart size and mediastinal contours. No focal airspace consolidation. No pleural effusion or pneumothorax. Stable postoperative and
IMPRESSION:  No acute cardiopulmonary abnormalities.
IMAGES:      472_IM-2100-1001.dcm.png, 472_IM-2100-2001.dcm.png
-----------------------------------------------------------------------------------------
```

Image: 472_IM-2100-1001.dcm.png Image: 472_IM-2100-2001.dcm.png




```
-----------------------------------------------------------------------------------------
PATIENT UID: 302
FINDINGS:    Heart size within normal limits, stable mediastinal and hilar contours. No focal alveolar consolidation, no definite pleural effusion seen.
IMPRESSION:  No acute findings
IMAGES:      302_IM-1394-1001.dcm.png
-----------------------------------------------------------------------------------------
```

UID: 376 | 376_IM-1883-1001.dcm.png



FINDINGS:
The heart size and pulmonary vascularity appear within normal limits. The lungs are free of focal airspace disease. No pleural effusion or pneumothorax is seen. No non-calcified nodules are identified.

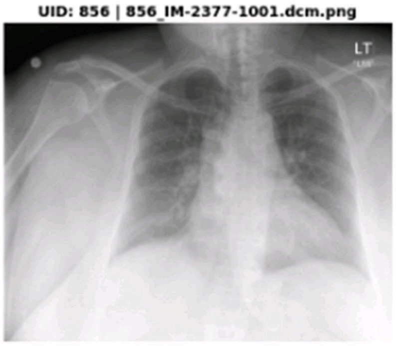
IMPRESSION:
1. No evidence of active disease.

UID: 3703 | 3703_IM-1850-1001.dcm.png

FINDINGS:
The XXXX examination consists of frontal and lateral radiographs of the chest. The posterior costophrenic XXXX are excluded on the lateral view. The cardiomediastinal contours are within normal limits. Pulmonary vascularity is within normal limits. No focal consolidation, pleural effusion, or pneumothorax identified. The visualized osseous structures and upper abdomen are unremarkable.

IMPRESSION:
No evidence of acute cardiopulmonary process.

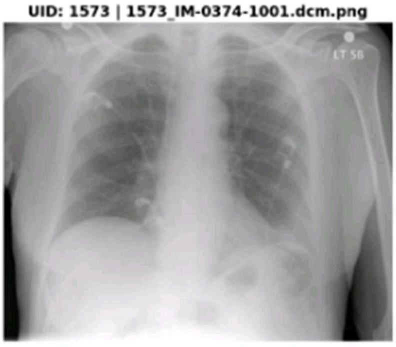
UID: 856 | 856_IM-2377-1001.dcm.png



FINDINGS:
Limited evaluation of the lateral view due to rotation and frontal view due to motion artifact. Stable mild cardiomegaly. Normal pulmonary vascularity. The lungs are clear. No focal consolidation, visible pneumothorax or large pleural effusions. XXXX XXXX opacities are related to overlying soft tissues. The posterior sulci are clear. Degenerative changes of the spine.

IMPRESSION:
1. Stable mild cardiomegaly. 2. No evidence of active cardiopulmonary disease.

UID: 1573 | 1573_IM-0374-1001.dcm.png



FINDINGS:
Heart size normal. Lungs are clear. XXXX are normal. No pneumonia, effusions, edema, pneumothorax, adenopathy, nodules or masses.

IMPRESSION:
Normal chest No evidence of tuberculosis

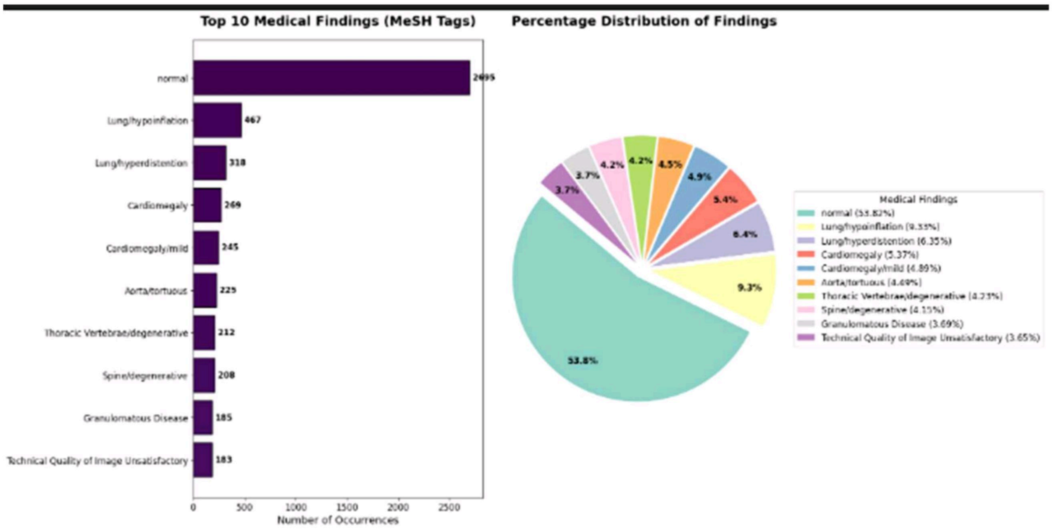
Visual & Clinical Diversity

Visual Patterns Present

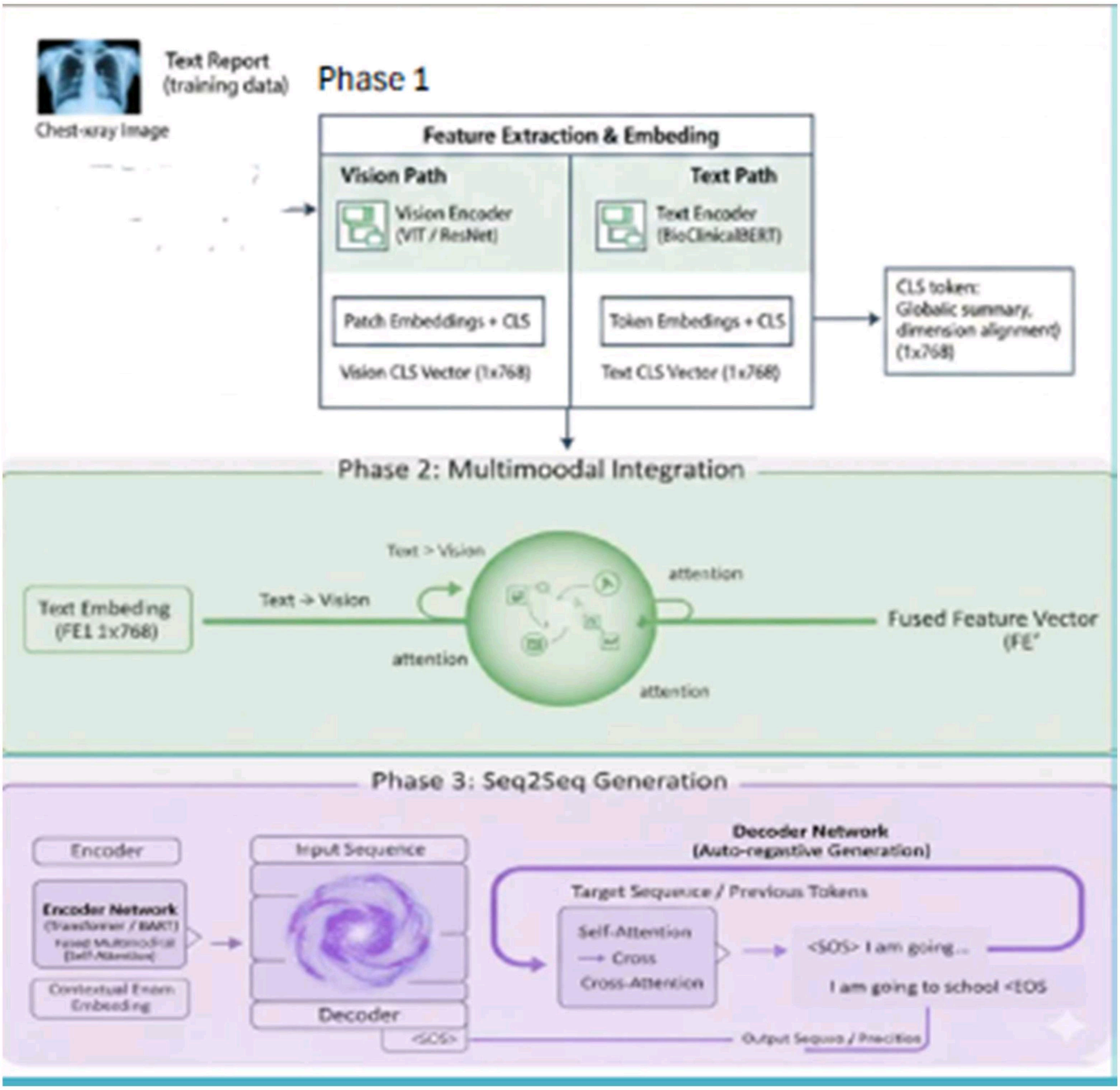
- Normal lung fields
- Cardiomegaly
- COPD / emphysema
- Atelectasis
- Pleural effusion
- Pneumothorax
- Calcified granulomas
- Surgical hardware
- Poor image quality cases

Report Complexity Spectrum

Type	Example
Normal	"No acute cardiopulmonary abnormality."
Borderline	"Mild cardiomegaly without failure."
Complex	"Bullous emphysema with interstitial fibrosis."
Temporal	"Interval increase compared to prior study."



Proposed Project Architecture



Evaluation Metrics & Performance Analysis

Evaluation of radiology report generation systems requires multiple complementary metrics addressing both linguistic quality and clinical accuracy

Natural Language Generation (NLG) Metrics

These metrics focus on the linguistic quality and overlap between generated and reference reports

Metric	Description	Range	Focus
BLEU-4	Measures 4-gram overlap between generated and reference reports	0-1 (Higher is better)	Lexical overlap
ROUGE-L	Evaluates longest common subsequence matching	0-1 (Higher is better)	Sequence similarity
METEOR	Combines precision and recall with synonym matching and stemming	0-1 (Higher is better)	Semantic similarity
CIDEr	Consensus-based Image Description Evaluation	0-10 (Higher is better)	Consensus scoring

Clinical Efficacy Metrics

These metrics evaluate the medical accuracy, factual consistency, and clinical utility of the reports.

Metric	Description	Range	Focus
BLEU-4	Measures 4-gram overlap between generated and reference reports	$0-1$ (Higher is better)	Lexical overlap
ROUGE-L	Evaluates longest common subsequence matching	$0-1$ (Higher is better)	Sequence similarity
METEOR	Combines precision and recall with synonym matching and stemming	$0-1$ (Higher is better)	Semantic similarity
CIDEr	Consensus-based Image Description Evaluation	$0-10$ (Higher is better)	Consensus scoring

Project Timeline & Milestones

Phase	Duration	Key Activities	Status
Phase 1: Setup & Analysis	Jan - Feb 2025	Data exploration, model research, environment setup	COMPLETED
Phase 2: Development	Feb - Nov 2025	Vision encoding, text encoding, fusion module implementation	AT PROGRESS
Phase 3: Testing & Evaluation	Dec 2025 - Jan 2026	Performance metrics, validation, report generation testing	Scheduled
Phase 4: Optimization & Documentation	Feb - Mar 2026	Performance tuning, final testing, report writing, presentation	Scheduled

Expected Completion: March 2026

Conclusion

Summary of project achievements and final thoughts.

Key Takeaway	Summary
Technical Feasibility	Successfully demonstrated that vision-supervised models can generate linguistically fluent and clinically relevant reports.
Metric Shift	Proved that while BLEU/ROUGE are standard, **Clinical F1 (CheXpert/RadGraph)** and **BERTScore** are the true barometers for medical AI.
The Human Element	AI is not a replacement but an **augmentation tool**; human-in-the-loop remains the gold standard for final clinical decisions.
Closing Statement	Vision-supervised report generation is a pivotal step toward "Smart Radiology," bridging the gap between raw pixel data and actionable medical insights.

Future Scope

- **Integration of Clinical Context (EHR Integration)** : Future models should incorporate patient history, lab results, and previous imaging findings from Electronic Health Records (EHR) to provide more context-aware and personalized reports.
- **Explainability & Visual Grounding (XAI) : I**ntegrating Grad-CAM or Attention Maps to show exactly which region of the image triggered a specific sentence in the report.

Impact

Healthcare Efficiency , Clinical Quality & Safety , Global Accessibility

References

MAIRA-2: Grounded Radiology Report Generation (Paper)
https://arxiv.org/abs/2406.04449

2. MedKLIP: Medical Knowledge Enhanced Language-Image Pre-Training in Radiology (Paper) https://arxiv.org/abs/2301.02228