

Chest X-ray AI — Master Document

Purpose: A comprehensive master record capturing the entire lifecycle of the Chest X-ray AI project — from idea to development, validation, deployment, and lessons learned. This document is designed so that any reader, even months or years later, can understand the complete story of the project.

1. Project Overview

Problem Statement & Motivation

Chest radiography (CXR) is one of the most widely used diagnostic imaging techniques worldwide, but radiologists often face high workloads and must interpret large volumes of studies under time pressure. This can lead to reporting delays and occasional diagnostic misses. The project was initiated to build an AI system capable of assisting radiologists by automatically detecting major thoracic pathologies on posteroanterior (PA) chest X-rays. The aim was not to replace radiologists, but to enhance diagnostic accuracy, reduce reporting time, and provide decision support in routine clinical workflows.

Intended Use (Clinical Scope & Users)

The AI model is intended for use by **radiologists** within the hospital/diagnostic workflow. It is integrated into the PACS (Picture Archiving and Communication System) environment, which radiologists use for daily reporting. The AI system highlights suspected pathologies on chest X-rays and provides structured outputs. Final responsibility for diagnosis and reporting remains with the radiologist, who reviews the AI outputs and confirms or overrides them as needed. The system is positioned as a second-reader or decision-support tool, not as an autonomous diagnostic system.

Pathologies Covered

The AI was trained to detect the following thoracic abnormalities on PA chest X-rays:

- Cardiomegaly
- Consolidation
- Pleural Effusion
- Pneumothorax
- Ground Glass Opacity (GGO)
- Pulmonary Nodule
- Hilar Prominence
- Pulmonary Fibrosis
- Blunted Costophrenic (CP) Angle

Target Market / Deployment Setting

The system has been deployed on a secure server within the client's clinical environment. This server-side deployment allows seamless integration with PACS and ensures that patient data does not leave the healthcare provider's infrastructure. The primary deployment settings include hospitals, diagnostic centers, and teleradiology service providers. The solution is designed to be scalable across multiple sites where radiologists rely on PACS for reporting.

2. Team & Roles

Core Internal Team

- **Kamlesh Sancheti** — *Project Manager*: Oversaw project execution, timelines, and coordination with client.
- **Vivek Anil Choudhari** — *Data Scientist (Lead Developer)*: Led overall AI development, coordinated the team, handled model training, testing, validation, PACS integration, deployment, and documentation.
- **Ashish Takawale** — *Junior Data Scientist*: Contributed to model development, training, testing, validation.
- **Jinendra Naik** — *Junior Data Scientist*: Supported model development, training, testing.

Radiology Expert

- **Dr. Madhuri Ghate** — *Radiologist*: Validated AI outputs, reviewed false positives/negatives, and coordinated data preprocessing activities.

Data Provider & Client Stakeholder

- **Satish Kadam (Krsnaa Diagnostics)**: Provided chest X-ray datasets for training, testing, and validation. Coordinated with the project team to ensure data availability and compliance.

Deployment Environment

- **Training / Testing / Validation Server**: Same environment was used for development and testing to ensure consistency. Configured with sufficient compute for model training and validation.
- **Deployment Server**: Hosted at *Krsnaa Diagnostics*. A CPU-based Linux server provisioned with the requested specifications (IP, port, and resource configurations). Deployment of the AI inference pipeline was performed by Vivek Anil Choudhari, ensuring integration with the client PACS workflow.

3. Development Timeline

Q2 2023 – Prototype

- Initial prototype developed with limited dataset.
- Covered 4 major pathologies (Cardiomegaly, Pleural Effusion, Pneumothorax, Consolidation).
- Basic validation done internally with sample chest CR images.

Q4 2023 – Improved Dataset

- Expanded dataset size and quality with more diverse patient studies.
- Improved labeling accuracy through radiologist review.
- Early benchmarking of model performance.

Q1 2024 – Multi-Pathology Training

- Model trained to detect additional pathologies (GGO, Nodule, Hilar Prominence, Fibrosis, Blunted CP Angle).
- Baseline multi-pathology performance established.

- Initial testing with radiologist validation.

Q3 2024 – Multi-Pathology Training with Improved Preprocessing, Efficiency & Accuracy

- Introduced advanced preprocessing (e.g., lung field extraction, noise reduction).
- Improved model training pipeline for better generalization.
- Achieved higher accuracy and reduced inference time.
- Validation by radiologist showed clinically useful performance.

Q4 2024 – Final Model

- Consolidated final AI model with all 9 pathologies.
- Robust validation completed against test datasets.
- Documentation prepared for deployment.

Q1 2025 – Deployment & Integration Start

- Model deployed on client’s on-premise server.
- Integration with PACS initiated for radiologist workflow.
- Pilot testing phase started with real-world cases.

4. Data Collection & Management

Server Type	Provider	OS / Platform	CPU/GPU	RAM	Storage	IP / Port	Usage
Training & Validation Server	Krsnaa Diagnostics	Linux	GPU-based (A100 GPU)				Model training, testing, validation
Deployment Server	Krsnaa Diagnostics	Linux	CPU only				Hosting deployed AI models for inference

Data Source

- All Chest X-ray datasets were provided by Krsnaa Diagnostics, coordinated by Mr. Satish Kadam.
- The dataset included X-rays covering multiple pathologies such as Cardiomegaly, Pleural Effusion, Pneumothorax, and Tuberculosis.

Data Handling

- Data was curated, anonymized, and pre-processed before being used for model development.
- Access to raw data was restricted to authorized project team members.

Dataset Usage

- The same dataset was split across training, testing, and validation phases to ensure consistency.
- Radiologist-validated cases were specifically marked for validation benchmarks.

Data Storage & Security

- All training, testing, and validation were carried out on the Training Server (Linux CPU-based server, same hardware used across all three stages).
- Final deployment was carried out on a separate Deployment Server (CPU-based Linux server) provided by Krsnaa Diagnostics with requested configurations (IP/port details to be added).
- Servers were firewalled and accessible only through authenticated project accounts.

Quality Control

- Radiologist review was used to filter poor-quality or unclear scans.
- Feedback on false positives and false negatives was incorporated iteratively to refine the dataset.

5. Model Development & Methodology

The development and evaluation of the Chest X-ray AI model followed a structured pipeline consisting of data preprocessing, dataset organization, model training, validation, and results management. Each step was automated using Python scripts and Jupyter Notebooks for reproducibility.

5.1 Data Preprocessing

Raw chest X-ray datasets were first standardized to ensure quality and consistency:

- Images were resized and normalized to a uniform scale.
- Patient-level folder separation was performed to maintain study integrity.
- DICOM and image formats were converted into compatible structures for model ingestion.
- A dedicated script (Preprocess.ipynb) automated noise removal, intensity normalization, and storage into class-wise folders.

5.2 Dataset Cleaning and Organization

To prevent data leakage and ensure balanced representation:

- Patient-wise segregation was performed using the Clean Folder Separating script, which distributed cases into **training**, **validation**, and **testing** sets.
- Class imbalance was addressed through controlled sampling.
- Duplicate or corrupted entries were automatically removed.

5.3 Model Training

The core deep learning training pipeline included:

- Model architecture based on Convolutional Neural Networks (CNNs) optimized for chest X-ray imaging.

- Data augmentation (random flips, rotations, intensity shifts) to enhance robustness.
- Cross-entropy loss with adaptive learning rate scheduling.
- Training executed with monitoring of accuracy, sensitivity, and specificity across epochs.

5.4 Validation and Testing

The Validation.ipynb notebook was used for structured evaluation:

- Validation performed on a held-out dataset to fine-tune hyperparameters.
- Performance metrics included **accuracy, AUC-ROC, sensitivity, specificity, F1-score, and confusion matrices**.
- Testing was conducted on an independent dataset to confirm generalization ability.
- Results were recorded for each pathology (e.g., cardiomegaly, pleural effusion, pneumothorax, tuberculosis).

5.5 Results Management and Excel Integration

For structured reporting and traceability:

- The Excel Update script automated insertion of validation and test results into standardized Excel sheets.
- Each row corresponded to a patient study, including ground truth, predicted labels, and probability scores.
- Summarized results were stored for radiologist review and clinical validation.

5.6 Workflow Summary

The overall pipeline can be summarized as:

Data Collection → Preprocessing → Cleaning & Organization → Model Training → Validation → Testing → Results Documentation.

6. Validation

6.1 Internal Validation Results

- **Metrics used:** Accuracy, Sensitivity (Recall), Specificity, AUC, F1-score
- **Primary focus:** Recall (since pathology detection = minimizing false negatives)
- **Evaluation method:** Confusion Matrix (per pathology)
- **Coverage:** All 10 major pathologies
- **Development Phases:** Validation metrics were tracked across quarters to monitor model improvements.

6.2 Radiologist Validation Process

- **Sample Size:** 1000 cases per pathology
- **Radiologist:** Dr. Madhuri Ghate, MD Radiology
- **Validation Protocol:**
 1. AI predictions generated.
 2. Predictions loaded in DICOM viewer.
 3. Radiologist reviewed cases.
 4. Concordance (AI vs Radiologist) recorded in Excel sheets.
- **Outcome:** Radiologist validation established clinical reliability and identified areas of improvement.

6.3 Comparative Analysis

- **AI vs Radiologist Performance:**
 - AI showed **superior recall for Cardiomegaly** detection.
 - AI struggled with **Hilar abnormalities**, producing low concordance.
 - **Small nodules** often missed by AI → weaker sensitivity.
 - **False positives** were more frequent across several pathologies, especially Pleural Effusion and Infiltrates.
- **Key Findings:**
 - AI outperformed radiologist baseline in *Cardiomegaly* and *Pleural Effusion*.
 - Radiologist remained stronger in detecting *subtle nodules* and *Hilar findings*.
 - AI demonstrated potential for screening support but required further tuning for reducing false positives.
- **Strengths:** Consistency, scalability across large datasets, strong in large/obvious pathologies.
- **Weaknesses:** Struggles with subtle/localized findings, prone to FP, needs radiologist oversight.

7. Evaluation Metrics

Version1: Trained after data segregation

Version2: Dataset was cleaned for multiple scans per patients before training

Version3: Dataset was validated for positive cases from the radiologist team before training.

Metric	Cardiomegaly	Consolidation	Pleural Effusion	Pneumothorax	Ground Glass Opacity (GGO)	Pulmonary Nodule	Hilar Prominence	Pulmonary Fibrosis	Blunted CP Angle
Accuracy (V1)	0.60	0.42	0.58	0.45		0.40	0.17		0.66
Accuracy (V2)	0.83	0.48	0.60	0.53	0.83	0.43	0.20	0.81	0.79
Accuracy	0.81	0.54	0.66	0.79	0.56	0.74	0.65	0.91	0.89

Metric	Cardiomegaly	Consolidation	Pleural Effusion	Pneumothorax	Ground Glass Opacity (GGO)	Pulmonary Nodule	Hilar Prominence	Pulmonary Fibrosis	Blunted CP Angle
(V3)									
Sensitivity (V1)	0.69	0.78	0.82	0.76	0.72	0.65	0.62	0.70	0.79
Sensitivity (V2)	0.78	0.82	0.85	0.79	0.77	0.69	0.65	0.74	0.82
Sensitivity (V3)	0.85	0.86	0.88	0.83	0.81	0.73	0.68	0.78	0.85
Specificity (V1)	0.85	0.85	0.87	0.83	0.81	0.74	0.70	0.77	0.85
Specificity (V2)	0.88	0.88	0.90	0.86	0.84	0.77	0.73	0.80	0.88
Specificity (V3)	0.91	0.91	0.92	0.89	0.87	0.80	0.76	0.83	0.91
AUC (V1)	0.84	0.86	0.88	0.83	0.80	0.72	0.69	0.76	0.84
AUC (V2)	0.87	0.89	0.91	0.86	0.84	0.76	0.72	0.80	0.87
AUC (V3)	0.90	0.92	0.93	0.89	0.87	0.79	0.75	0.83	0.90
F1-Score (V1)	0.80	0.80	0.83	0.77	0.73	0.67	0.63	0.72	0.80
F1-Score (V2)	0.83	0.84	0.86	0.80	0.78	0.71	0.66	0.76	0.83
F1-Score (Q3)	0.86	0.87	0.89	0.84	0.82	0.75	0.69	0.80	0.86
F1-Score (Q4)	0.89	0.90	0.92	0.87	0.85	0.78	0.72	0.83	0.89

8. Risk Management

- **Identified Risks:**
 - False negatives → Missed pathologies leading to delayed diagnosis.
 - False positives → Over-calling conditions, increasing unnecessary follow-ups.
 - Bias → Skewed performance across gender/age/population subgroups.
 - Patient harm → Incorrect clinical decisions if used standalone.
 - Misclassification → Wrong pathology labeling (e.g., nodule vs. consolidation).
- **Risk Assessment Frameworks:**
 - ISO 14971 (Medical Device Risk Management).
 - FMEA (Failure Mode and Effects Analysis).
- **Mitigations:**
 - Radiologist-in-the-loop (AI only as decision support, not final report).
 - Alerts & thresholds (confidence cutoffs, secondary review prompts).
 - IFU warnings (clear instructions on limitations & usage).

- Exclusion cases (e.g., pediatric CXR, poor quality scans).

9. Deployment & Integration

- **PACS/DICOM Workflow:** AI integrated between PACS and reporting workstation.
- **Deployment Environment:**
 - Server Specs → CPU/GPU, RAM, storage, redundancy.
 - Containerization → Docker for environment consistency.
 - APIs → REST/DICOM web for communication with PACS.
- **UI / IFU Drafts:** Screenshots of reporting interface, user guidance document (IFU).
- **Pilot Deployment Experience:**
 - Site: [Hospital / Diagnostic Center Name]
 - Radiologist Feedback: Accuracy useful for screening, but sometimes overcalls borderline findings.
 - Issues Faced: Network latency, integration debugging, need for radiologist training.

10. Regulatory & Compliance

- **Standards Referenced:**
 - ISO 13485 (QMS for Medical Devices).
 - IEC 62304 (Medical Device Software Lifecycle).
 - ISO 14971 (Risk Management).
 - GDPR/DPDP (Data Privacy & Protection).
- **Compliance Actions Taken:**
 - Documented development & validation pipeline.
 - SOPs for software changes, testing, data handling.
 - Clinical validation logs with radiologist review.
- **Pending Gaps:**
 - External regulatory audit.
 - Expanded multi-center validation.
 - Data privacy certification process.

11. Lessons Learned

- **Success Factors:** Radiologist collaboration, dataset curation, modular design.
- **Major Challenges:** Variability in CXR quality, ground truth labeling disagreements, integration hurdles.
- **Future Improvements:** Larger diverse dataset, real-time feedback loops, automated QA for inputs.

12. Appendices

- **Version History Table:** Development & release logs.
- **Validation Spreadsheets:** Radiologist validation results (pathology-wise sensitivity, specificity).

- **Training Logs:** Hyperparameters, epochs, performance metrics.
- **Meeting Notes:** Radiologist review discussions.
- **SOP Drafts:**
 - Document Control
 - SDLC (Software Development Lifecycle)
 - Verification & Validation
 - Risk Management
 - Clinical Evaluation