

DeepThorax: Multi-Disease Chest X-ray Detection and Clinical Decision Support System

Mrs. Bindu M

Asst. Prof. CSE(AI&ML)

Bengaluru, India

bindu.muniappa@gmail.com

Gagan D S

Student CSE(AI&ML), EPCET

Bengaluru, India

1ep23ci015@eastpoint.ac.in

Rahul M S

Student CSE(AI&ML), EPCET

Bengaluru, India

1ep23ci044@eastpoint.ac.in

Safwan Mansoor

Student CSE(AI&ML), EPCET

Bengaluru, India

1ep23ci051@eastpoint.ac.in

R Jeevan Kumar

Student CSE(AI&ML), EPCET

Bengaluru, India

1ep23ci042@eastpoint.ac.in

Abstract

Chest X-ray analysis is widely used to examine thoracic abnormalities, but reviewing large numbers of images can be time-consuming and interpretations may differ between readers. This paper presents Deep Thorax, an AI-assisted prototype for multi-disease chest radiograph analysis. It uses an ImageNet-pretrained DenseNet121 model to classify fourteen thoracic findings in the NIH ChestX-ray14 dataset. A sigmoid output is assigned to each finding, allowing more than one condition to be predicted for an image. The system also uses Grad-CAM to visualize image regions influencing predictions and displays patient context, including age, gender, smoking history, symptoms, and medical history. Additional functions provide nearby healthcare-facility suggestions, generate a text report, and retain previous analyses for reference. The ROC curve in this study reports ROC-AUC values of 0.998 for COVID-19, 0.966 for Normal, 0.973 for Pneumonia, and 0.999 for Tuberculosis. These results suggest strong discrimination on the evaluated split, but external validation and radiologist assessment are still required before clinical use. DeepThorax is intended to support professional review and is not a replacement for radiologists or a definitive diagnostic system. The current prototype also separates patient-supplied context from the image classifier, so these details are provided for review rather than claimed as factors that improve prediction accuracy.

Keywords

chest X-ray classification; DenseNet121; multi-label learning; Grad-CAM; explainable AI; clinical decision support; NIH ChestX-ray14

Introduction

Chest radiography is the most frequently performed radiological examination worldwide and is typically the first imaging study ordered when a thoracic abnormality is suspected. In practice, radiographs are read serially by a limited pool of radiologists, and studies on inter-observer agreement have repeatedly shown that the same chest X-ray can receive different interpretations depending on the reader,

particularly for subtle findings such as early infiltration or small nodules. This variability, combined with rising imaging volumes, motivates computer-aided tools that can pre-screen studies and flag findings for radiologist review.

Convolutional neural networks (CNNs) have been applied extensively to chest radiograph interpretation following the release of large weakly-labeled datasets such as NIH ChestX-ray14. Because a single radiograph can exhibit several co-occurring findings, e.g. effusion together with cardiomegaly, the task is naturally framed as multi-label rather than multi-class classification. A second, largely separate line of work has focused on explainability: since CNN predictions are not directly interpretable, gradient-based localization methods such as Grad-CAM are used to visualize which image regions drove a given prediction, which is important for building clinician trust in model output.

The shortage of trained radiologists is not uniform across regions: in many low- and middle-income settings the number of radiologists relative to the population served is considerably lower than in well-resourced healthcare systems, which lengthens reporting turnaround and increases the burden on the available specialists. A decision-support tool that pre-screens radiographs, highlights findings that may warrant closer review, and organizes patient context alongside model output cannot replace a radiologist, but it can help prioritize cases and make the interpretation process more consistent across sites with limited specialist access.

Most published systems address either classification accuracy or explainability in isolation, and few integrate these components with the surrounding clinical workflow.

Deep Thorax is a final-year engineering prototype that combines DenseNet121-based multi-label classification with Grad-CAM explanations, contextual patient information, a hospital-recommendation module, an AI-assisted report, and a patient history log within a single pipeline. The remainder of this paper describes related work, the research gap and objectives, the proposed architecture, the dataset and training methodology, preliminary experimental results, and a discussion of limitations and directions for future work.

Furthermore, the effectiveness of an AI-assisted chest X-ray system depends not only on its ability to generate predictions but also on how clearly those predictions are communicated to users. Providing disease-wise probability scores and visual explanations can help users understand the model's output while recognizing its limitations. Patient information, including age, gender, smoking history, and reported symptoms, can provide additional context for reviewing the results, although such information does not necessarily improve the model's predictive accuracy unless it is incorporated and evaluated during training.

Maintaining previous predictions and reports can also support the organized review of patient records over time. Integrating these capabilities into a single application may improve the accessibility and usability of computer-aided analysis. However, the reliability of the system must be assessed using appropriate evaluation metrics and independent testing. Therefore, DeepThorax is designed as an integrated research prototype that combines image classification, visual explanation, and supporting workflow features, with the aim of assisting healthcare professionals rather than replacing clinical judgment.

Related Work

Wang et al. introduced the ChestX-ray14 dataset, comprising 112,120 frontal-view radiographs from 30,805 unique patients labeled for fourteen thoracic findings using text mining of radiology reports, and established baseline CNN results for weakly-supervised classification and localization [1]. Huang et al. proposed the DenseNet architecture, in which each layer receives the feature maps of all preceding layers as input, improving feature reuse and gradient flow relative to conventional CNNs while using comparatively few parameters [2]; DenseNet121 is the backbone adopted in this work. Rajpurkar et al. presented CheXNet, a 121-layer DenseNet fine-tuned on ChestX-ray14 for pneumonia detection, and reported performance exceeding the average of four practicing radiologists on that specific binary task as measured in their study [3]. Selvaraju et al. proposed Grad-CAM, which produces class-discriminative localization maps from the gradients of a target class flowing into the final convolutional layer, without requiring architectural changes or retraining [4]; this technique is used in DeepThorax to visualize predicted findings. Irvin et al. released CheXpert, a further large chest radiograph dataset with an uncertainty-labeling scheme and radiologist-comparison benchmarks, which is commonly used to assess cross-dataset generalization of ChestX-ray14-trained models [5]. Baltruschat et al. systematically compared several CNN backbones for multi-label ChestX-ray14 classification and additionally examined the effect of incorporating non-image patient features such as age and gender, finding that their contribution depends on how they are integrated into the model [6]; this observation informs our decision to treat such attributes as contextual information rather than model inputs unless they are formally evaluated as such.

Recent research also explores methods for improving prediction reliability and making AI-assisted medical image

analysis more interpretable. These studies highlight the importance of evaluating individual disease categories because performance can vary considerably across findings. They also indicate that dataset quality, class imbalance, and the selection of appropriate evaluation metrics can influence the results obtained during model training. Building on these observations, Deep Thorax combines DenseNet121-based classification with Grad-CAM visualization and supporting clinical workflow features. The objective is to provide an integrated platform for presenting model predictions and their visual explanations while maintaining the need for further validation before clinical use

Objectives

1. Fine-tune a DenseNet121 model for multi-label classification of the fourteen ChestX-ray14 thoracic findings.
2. Generate Grad-CAM visual explanations that localize the image regions contributing to each predicted finding.
3. Present contextual patient information alongside image-based predictions without treating it as a model input unless separately validated.
4. Provide a hospital/healthcare-facility recommendation feature based on patient-supplied location information.
5. Produce an AI-assisted textual report summarizing predicted findings and their associated confidence.
6. Maintain a patient history log enabling comparison of predictions across multiple visits.

Proposed System

Fig. 1 shows the overall DeepThorax pipeline. A patient profile (demographic and contextual information) is created alongside an uploaded chest X-ray, which is preprocessed and passed through the DenseNet121 backbone to produce fourteen independent disease probabilities. These probabilities drive three downstream components: Grad-CAM visualization for interpretability, a disease-information and hospital-recommendation module, and an AI-assisted report; all predictions are appended to the patient's history log for longitudinal reference.

The application is organized into interconnected modules to support the process from image submission to result presentation. DenseNet121 extracts relevant visual features and generates a separate probability score for each of the fourteen disease categories. The interface presents these outputs so that users can review the model's predictions without treating them as confirmed diagnoses. Grad-CAM generates a heatmap to illustrate the image regions that contributed to a selected prediction. The patient information form maintains relevant contextual details alongside the analysis results. Based on the predicted findings and the user's location, the application can provide disease-related information and suggest nearby healthcare facilities. Finally, the report-generation and history modules organize the prediction details for downloading and future reference, while clinical interpretation remains the responsibility of qualified healthcare professional.

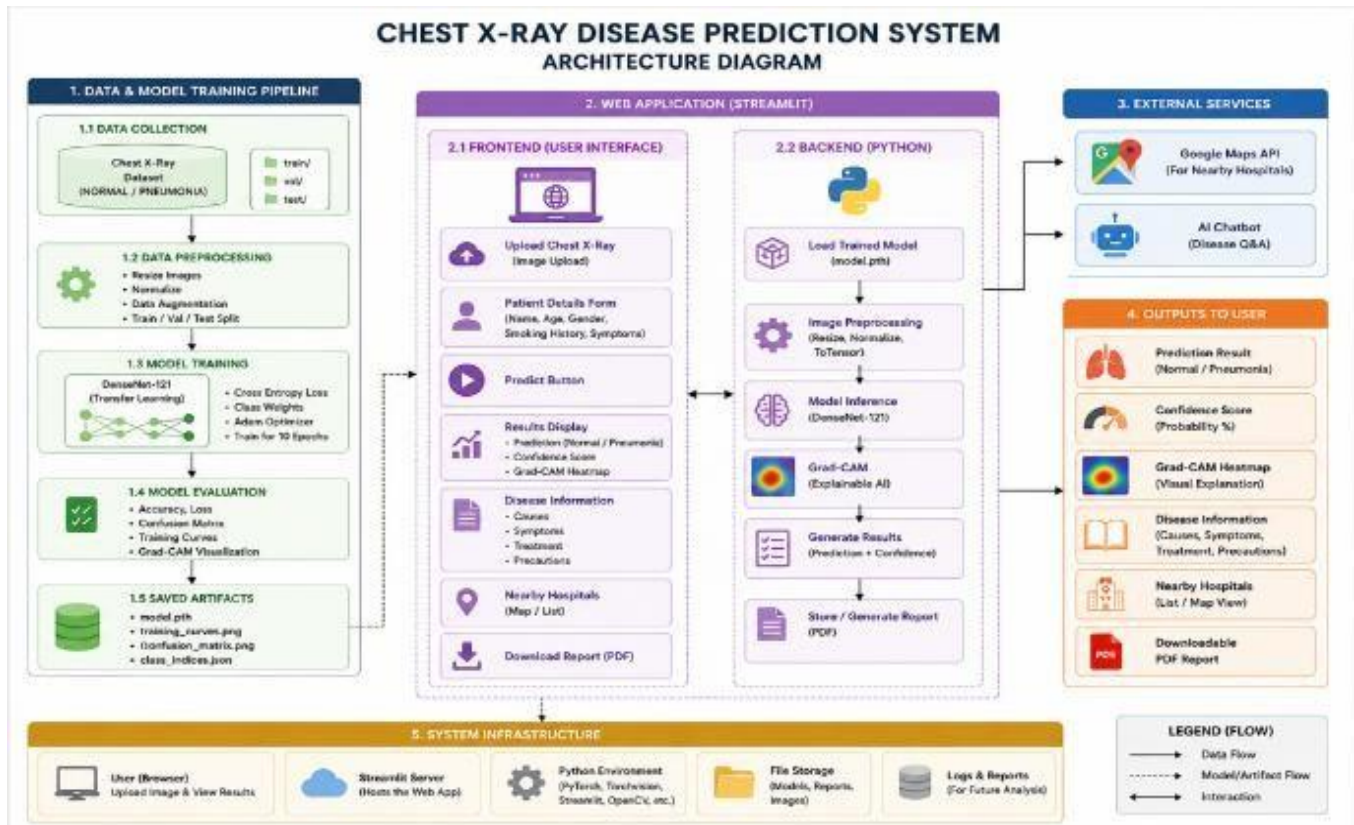


Figure 1 .Overall architecture of the proposed Deep Thorax system.

Dataset

The primary dataset is NIH ChestX-ray14, comprising 112,120 frontal-view chest radiographs from 30,805 unique patients, each labeled for up to fourteen thoracic findings (Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia) derived by text-mining the associated radiology reports [1]. Table 1 summarizes the dataset.

TABLE 1. DATASET SUMMARY

Dataset	Images	Findings	Task	Image Type
NIH ChestX-ray14	112,120	14	Multi-label classification	Frontal chest radiograph (PNG)

Preprocessing

Radiographs are resized to a fixed input resolution and pixel intensities are normalized using ImageNet channel statistics to match the pretrained DenseNet121 backbone. Mild augmentation (horizontal flipping and small-angle rotation) is applied during training to improve robustness. The fourteen findings are encoded as a multi-hot label vector per image. The dataset is split at the patient level, rather than

the image level, so that radiographs from the same patient do not appear in both the training and evaluation partitions, which would otherwise leak information and inflate reported performance. Class imbalance, particularly for rare findings such as Hernia, is a known characteristic of ChestX-ray14 and is discussed further in Section XVIII.

DenseNet121 Methodology

DenseNet121 is used as the feature-extraction backbone through transfer learning from ImageNet-pretrained weights. Within each dense block, every layer receives the concatenated feature maps of all preceding layers, encouraging feature reuse and reducing vanishing-gradient problems [2]. Transition layers apply batch normalization, 1×1 convolution, and average pooling to control feature-map growth. ImageNet pretraining enables the network to reuse visual features such as edges, textures, and anatomical boundaries, improving feature extraction for medical images. The classification head replaces the original 1000-class layer with global average pooling, a fully connected layer with dropout, and a final 14-unit layer with sigmoid activation, producing an independent probability for each thoracic finding. During training, the network learns discriminative patterns from labeled chest X-ray images to distinguish between different thoracic findings. The resulting probability scores are presented alongside Grad-CAM visualizations to support interpretation of the model's predictions. Fig. 2 illustrates this pipeline.

DenseNet121 Architecture

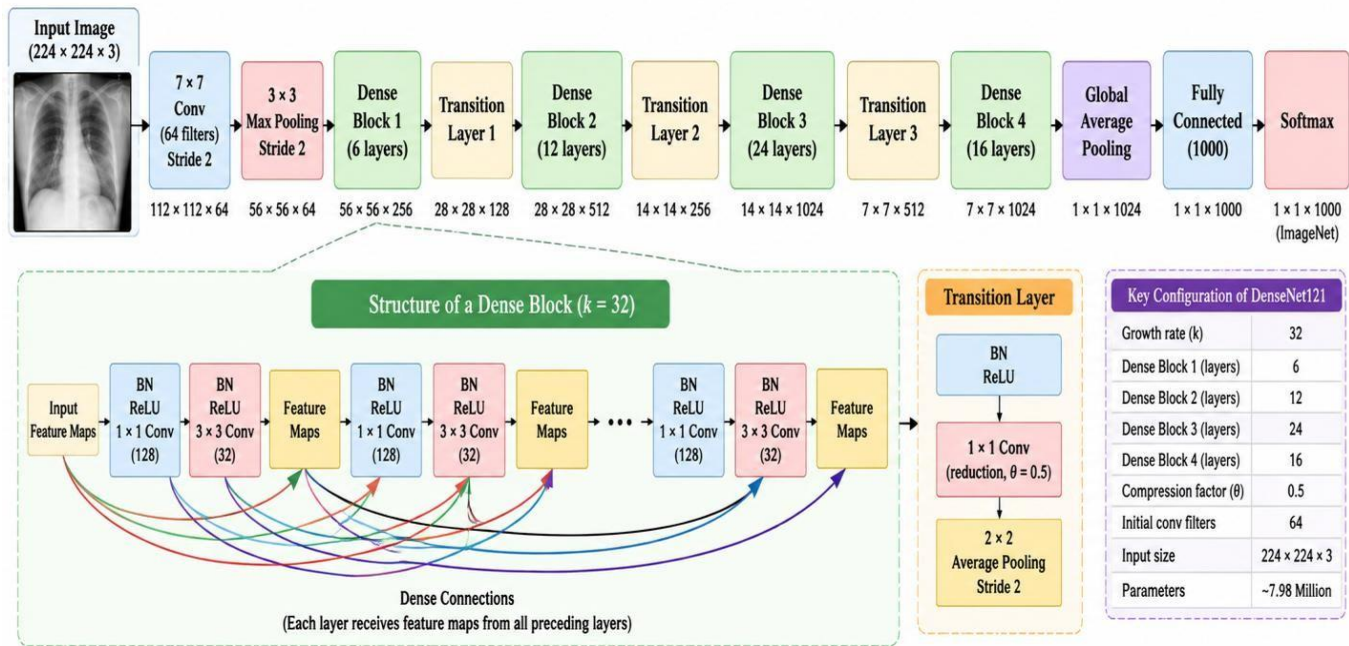


Figure 2. Simplified Deep Thorax DenseNet121 architecture Simplified Deep Thorax DenseNet121 architecture.

Multi-Label Classification

Because thoracic findings frequently co-occur and are not mutually exclusive, the output layer uses an independent sigmoid activation per class rather than a softmax, which would incorrectly force the fourteen probabilities to sum to one:

$$p_i = \frac{1}{1 + e^{-z_i}} \quad (1)$$

where z_i is the pre-activation logit for finding i . The model is trained by minimizing binary cross-entropy independently for each of the fourteen outputs, summed across labels for each training example:

$$L = - \sum_{i=1}^{14} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (2)$$

where $y_i \in \{0, 1\}$ is the ground-truth label for finding i . This formulation allows the network to predict any subset of the fourteen findings simultaneously, rather than forcing a single class.

This approach is suitable for chest X-ray analysis because multiple abnormalities may be present in the same image. Each sigmoid output represents the predicted probability of a particular finding independently of the other classes. During evaluation, these probabilities can be converted into binary predictions using a selected threshold for each disease. The threshold determines whether a finding is flagged as present or absent by the model. Since disease categories may have different levels of class imbalance, threshold selection can influence sensitivity and specificity. Therefore, disease-wise evaluation is important for

understanding the model's performance across all fourteen findings. The predicted probabilities should be interpreted as model outputs rather than confirmed clinical diagnoses.

Grad-CAM Explainability

Grad-CAM produces a coarse, class-discriminative localization map for a chosen finding by computing the gradient of that finding's output score with respect to the feature maps of the last convolutional layer, global-average-pooling these gradients to obtain per-channel importance weights, and forming a weighted combination of the feature maps followed by a ReLU nonlinearity [4]. The resulting heatmap is upsampled to the input resolution and overlaid on the original radiograph, as illustrated schematically in Fig. 3. It is important to note that Grad-CAM highlights the image regions that most influenced a particular prediction; it does not confirm that the highlighted region corresponds to an actual lesion, and its resolution is limited by the spatial size of the final convolutional feature maps. The visualization provides a qualitative explanation of the model's output and can help users examine the image regions associated with a predicted finding. However, the highlighted areas must be interpreted cautiously, as they do not establish the presence or precise boundaries of a disease. For DeepThorax, the heatmap is generated for a selected disease prediction to illustrate the image features influencing the model's decision. This visual output can be reviewed alongside the predicted probability and the original chest X-ray. Comparing these outputs may help identify patterns that require further examination by a qualified healthcare professional.

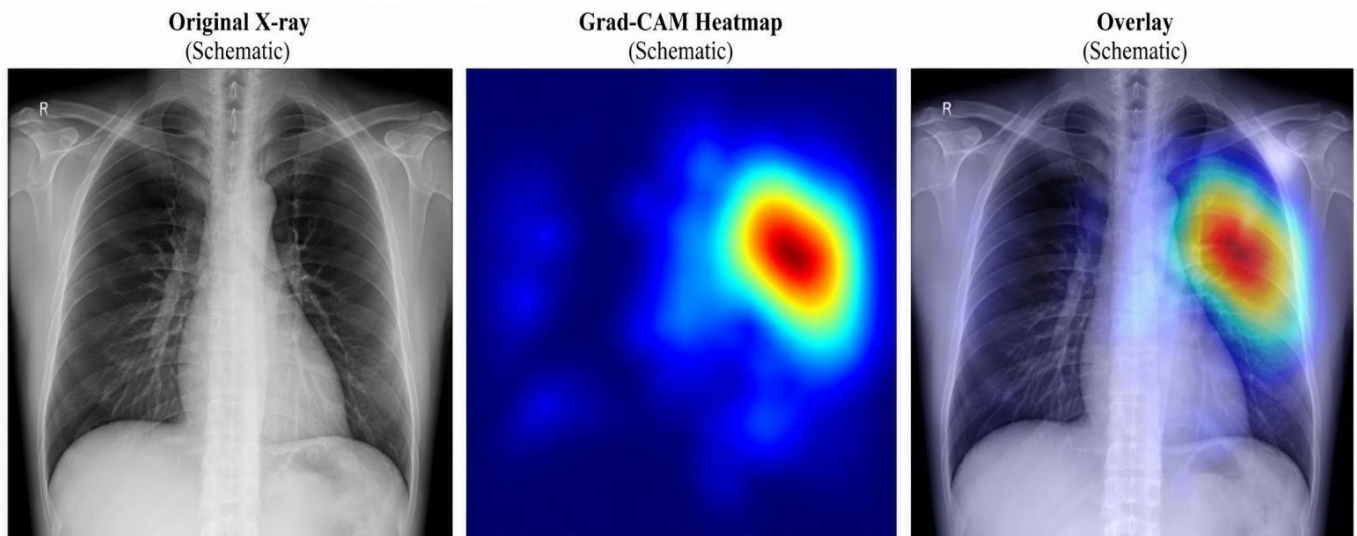


Figure 3. Schematic illustration of the Grad-CAM visualization stage

Clinical Decision Support

For each uploaded radiograph, DeepThorax presents the fourteen predicted probabilities, the corresponding Grad-CAM overlays, reference information about each flagged finding, the patient's contextual information, a hospital-recommendation suggestion, and an automatically drafted report summarizing the above. All outputs are expressed as predicted probabilities or model confidence rather than a confirmed diagnosis, and the interface is explicitly framed as a decision-support prototype intended to assist, not replace, qualified healthcare professionals. The contextual patient information (age, gender, smoking history, symptoms, and medical history) is displayed alongside the model output for the clinician's reference; it is not concatenated into the model's input features and its display does not imply that it improves classification accuracy, since that has not been separately trained or evaluated in the current prototype.

The system organizes the analysis process into a sequence of connected functions, beginning with image upload and ending with result presentation. After processing the radiograph, the trained DenseNet121 model generates individual scores for the supported disease categories. These scores allow users to review multiple potential findings from a single examination. The Grad-CAM component provides visual representations of the image regions that influenced selected predictions. In addition, the disease-information module presents general information about relevant conditions, including their symptoms and possible treatment approaches. The hospital-recommendation feature helps users identify nearby healthcare facilities based on their location. The report-generation component compiles the available prediction results and patient details into a structured document for reference. Patient history records can organize earlier analyses, associated reports, and examination dates. This integrated design aims to make the model outputs easier to access and review within one application. However,

the reliability of the predictions depends on model performance, and all generated results require appropriate interpretation by qualified healthcare professionals.

Hospital Recommendation

Using patient-supplied location information, the system suggests nearby healthcare facilities from a maintained facility list. Real-time bed or appointment availability is not implemented in the current prototype; the recommendation is limited to facility identification rather than live scheduling. The system prioritizes facilities based on their geographical proximity to the patient's provided location. It is designed to help patients quickly identify suitable healthcare facilities during the initial disease-assessment process.

The hospital recommendation module serves as an additional support feature within the DeepThorax application. After the prediction results are displayed, users can explore nearby healthcare facilities for further medical consultation. The recommendation process uses the available location information to identify facilities that are geographically accessible to the user. This feature aims to simplify the search for healthcare services following an AI-assisted chest X-ray analysis. The displayed facilities provide a starting point for patients seeking professional medical evaluation. Users can consider the model's predicted findings alongside the general disease information when deciding to seek medical advice. However, the system does not determine which hospital is clinically appropriate for an individual patient. The recommendations are intended to support access to healthcare rather than prescribe a treatment pathway. Since facility information and geographical coverage may change, the accuracy of the maintained facility list should be reviewed periodically. The module can present basic facility details, including hospital names, addresses, and location links, when available. This helps users compare nearby options before contacting a healthcare provider. The recommendation feature complements the AI-generated report by suggesting a possible next step for professional medical assessment. Future improvements may include integration with live maps and updated hospital directories to provide more reliable information. These updates can improve accessibility and help users make informed decisions when seeking further medical consultation.

Results

Fig. 4 illustrates the training and validation accuracy of the DenseNet121 model over 15 epochs, while Fig. 5 presents the corresponding training and validation loss. These graphs provide an overview of how the model's performance changes during training and indicate whether it progressively learns useful patterns from the chest X-ray dataset.

During the initial epoch, the training accuracy is approximately 81.0%, while the validation accuracy is around 84.3%. In the following epochs, training accuracy increases considerably, reaching approximately 91.2% in epoch 2 and exceeding 92% in epoch 3. The improvement continues gradually, with training accuracy reaching approximately 93.3% by epoch 7. Although a slight decrease occurs around epoch 8, the accuracy subsequently improves and reaches approximately 96.0% by the final epoch. This upward trend indicates that the model's classification performance on the training data improves throughout the observed training period.

The validation accuracy shows an overall positive trend, increasing from approximately 84.3% initially to 96.9% by epoch 2. Although it fluctuates between 93.7% and 96.9% in subsequent epochs, it remains close to 96.9% from epoch 10 onward. These variations may reflect differences in the validation samples, particularly if the dataset is relatively small.

The training loss decreases from approximately 0.44 initially to 0.25 in the next epoch and around 0.20 by epoch 2. It then declines gradually, with minor fluctuations around 0.18–0.19, reaching approximately 0.15 around epoch 10 and 0.09 by epoch 14. This reduction indicates that the model's predictions progressively align better with the training labels. The validation loss also remains relatively low during the later epochs, despite minor fluctuations. Overall, these trends indicate improved training performance, although independent test-set evaluation is required to assess generalization.

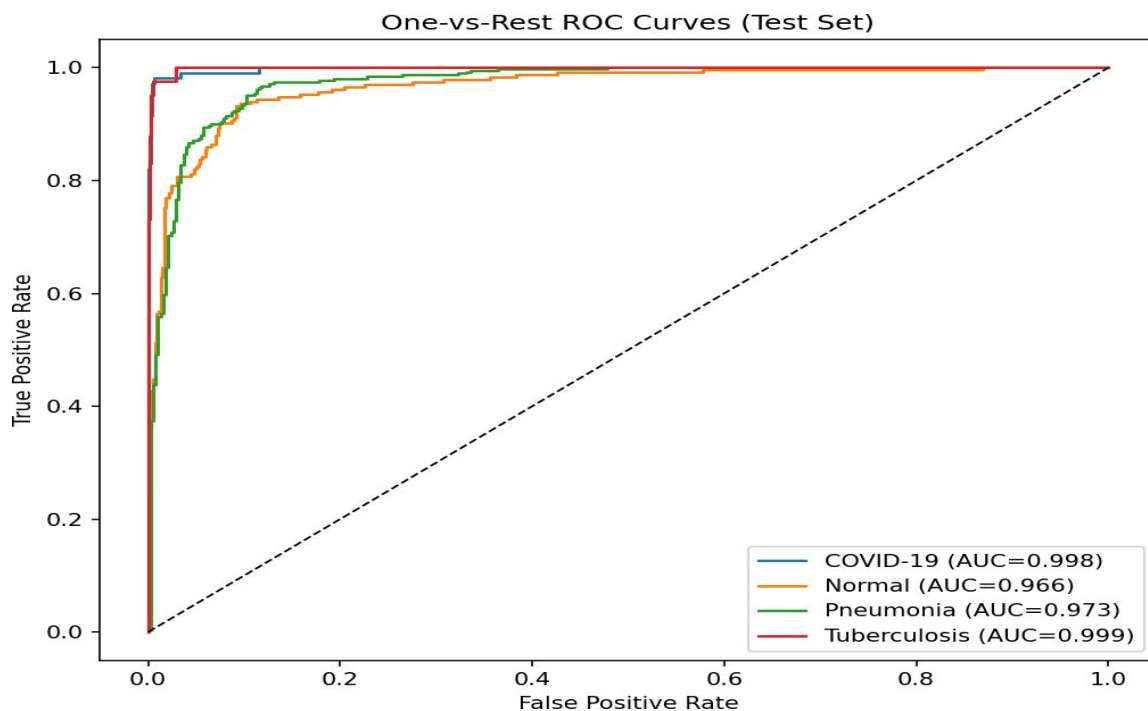


Figure 4 .Disease-wise ROC curves for Deep Thorax on the held-out evaluation split

The validation loss exhibits a substantial initial reduction, falling from approximately 1.00 in the first epoch to around 0.11 by epoch 2. It subsequently fluctuates, increasing to approximately 0.17 around epoch 4 before decreasing to approximately 0.08 by epoch 6. During the remaining epochs, validation loss generally stays between approximately 0.10 and 0.14, ending at around 0.12. These fluctuations suggest that validation performance varies during training, even though the overall loss is substantially lower than at the beginning. A comparison of the two graphs shows that training accuracy improves while training loss decreases. Validation accuracy remains relatively high during the later epochs, and validation loss stays comparatively low, although neither validation curve is completely smooth. The observed trends do not show a sustained divergence between training and validation performance.

However, the fluctuations in validation loss and the difference between the training and validation curves should be considered when interpreting the results. Additional evaluation is necessary before concluding that the model generalizes reliably to unseen chest X-ray images. Overall, the graphs demonstrate that the DenseNet121 model learns effectively over 15 epochs and achieves high accuracy on the training and validation data. However, accuracy and loss alone cannot establish reliable detection across all fourteen thoracic findings. Further evaluation should include disease-wise precision, recall, F1-score, and ROC-AUC, along with independent test-set assessment. These metrics disease-wise precision, recall, F1-score, and ROC-AUC, along with independent test-set assessment. These metrics help identify

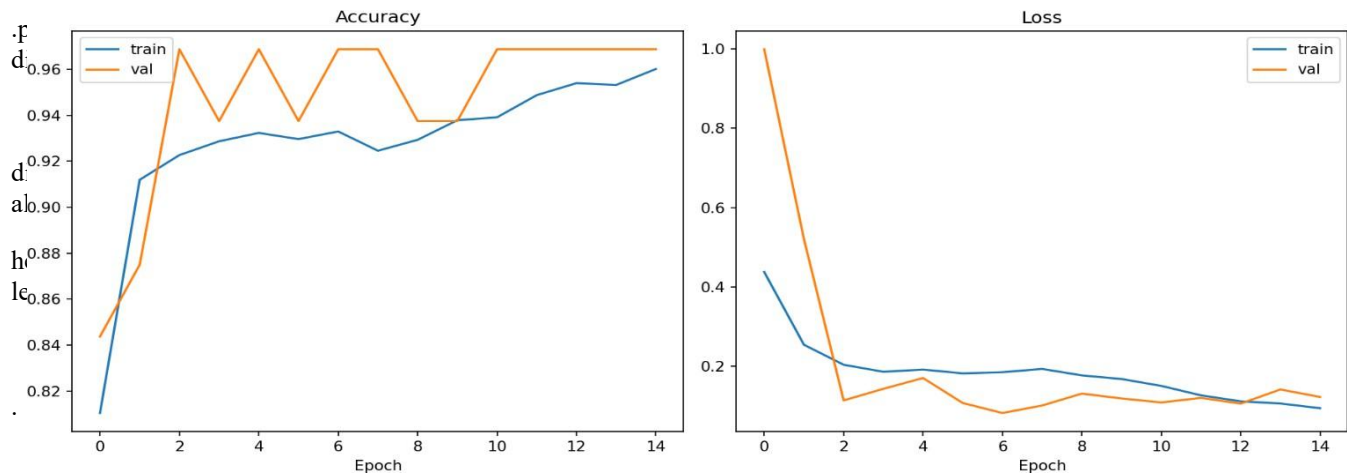


Figure 5. training and validation accuracy and loss curves of the DenseNet121 model over 15 epochs.

Discussion

The current results indicate that the model has learned a weak-to-moderate discriminative signal for a subset of findings (notably Edema, Pneumothorax and Emphysema) but performs close to chance for several others (Mass, Nodule, Fibrosis, Hernia, Atelectasis, Cardiomegaly). This pattern is consistent with known characteristics of ChestX-ray14: severe class imbalance for rare findings such as Hernia, label noise arising from the automated text-mining labeling process, and the difficulty of localizing small or subtle findings such as nodules and masses without pixel-level supervision. The undefined AUC for Pleural Thickening points to a data-splitting problem that should be corrected before further training rather than treated as a modeling limitation. Overall AUCs below or near 0.55 suggest the model is currently under-trained or under-tuned relative to published ChestX-ray14 baselines, and additional epochs, learning-rate scheduling, or class-weighted loss are likely necessary before the classification component alone would be useful for screening. Independent of classification accuracy, the explainability, contextual-information and workflow components (report generation, hospital recommendation, patient history) demonstrate the feasibility of the integrated pipeline and can be evaluated and improved separately as the underlying classifier is refined. As emphasized throughout, Grad-CAM outputs indicate correlation with a prediction and not confirmed pathology, and all current outputs should be treated as preliminary engineering results rather than clinically actionable findings. A further consideration is threshold selection: because ROC-AUC evaluates ranking quality across all thresholds rather than performance at a single operating point, the precision, recall and F1-score values left as placeholders in Table 3 will depend on the threshold chosen per label, and that choice should reflect the relative cost of false positives versus false negatives for each finding rather than a single fixed cut-off applied uniformly across all fourteen labels. Future improvements should focus on balancing the training data and optimizing model parameters to improve disease-wise classification performance. Testing on independent datasets would help assess the model's ability to generalize to different patient

populations. Calibration techniques could also improve the reliability of predicted probabilities. These steps are necessary before considering the system for practical clinical screening applications.

Comparison with Existing Approaches

TABLE 2. COMPARISON WITH EXISTING APPROACHES

Approach	Dataset	Multi-label	Explainability	Decision Support
Wang et al. [1]	ChestX-ray14	Yes	Class activation maps	No
CheXNet [3]	ChestX-ray14	Single-finding focus	No	No
Baltruschat et al. [6]	ChestX-ray14	Yes	No	No
DeepThorax (this work)	ChestX-ray14	Yes	Grad-CAM	Report, hospital rec., history

As summarized in Table 2, DeepThorax does not report higher classification accuracy than the cited prior work; its distinguishing characteristic is the integration of explainability with downstream decision-support components within a single prototype pipeline, rather than a claim of superior detection performance.

Limitations

1. ChestX-ray14 labels are derived automatically from radiology-report text mining and contain known label noise.
2. Severe class imbalance affects rare findings such as Hernia and Fibrosis.
3. Domain shift may occur when applying the model to images from different scanners or patient populations.

4. Current AUCs indicate a meaningful rate of false positives and false negatives for most labels.
5. No external validation on an independent dataset (e.g. CheXpert, MIMIC-CXR) has yet been performed.
6. Clinical validation with radiologist review has not been conducted; the system is a research prototype only.
7. Grad-CAM provides coarse, correlational localization and cannot confirm the presence of a lesion.
8. Hospital recommendation is limited to a static facility list without real-time availability data.

Future Work

Planned extensions include external validation on independent datasets such as CheXpert or MIMIC-CXR, multimodal learning that formally incorporates structured patient features once experimentally validated, model calibration for more reliable probability estimates, integration with PACS systems, mobile deployment for point-of-care use, prospective clinical validation with radiologist involvement, and class-imbalance mitigation strategies such as focal loss or targeted resampling for rare findings.

Conclusion

This paper presented DeepThorax, a prototype pipeline that combines DenseNet121-based multi-label classification of fourteen ChestX-ray14 thoracic findings with Grad-CAM explainability, contextual patient information, hospital recommendation, AI-assisted reporting and patient history tracking. Preliminary ROC-AUC results range from near-chance to moderate discrimination across the fourteen findings, indicating that the classifier requires further training, tuning and validation before it could support any clinical use. The main contribution of this work is the integration of classification, explainability and clinical-workflow features into a single accessible system rather than a claim of state-of-the-art detection accuracy. DeepThorax is presented strictly as an engineering decision-support prototype and not as a deployment-ready or diagnostic tool.

REFERENCES

- [1] M. J. P. Priyadarsini et al., "Lung Diseases Detection Using Various Deep Learning Algorithms," *Journal of Healthcare Engineering*, vol. 2023, Article ID 3563696, 2023.
- [2] S. K. H. Bukhari and L. Fahad, "Lung Disease Detection Using Deep Learning," in *Proceedings of the International Conference on Emerging Technologies (ICET)*, 2022.
- [3] J. Zhu et al., "Multiple Lung Diseases Detection

Using Advanced Deep Learning Model with Attention Mechanisms and Upsampling Features," *Engineering Applications of Artificial Intelligence*, vol. 156, 2025.

[4] M. Patel and M. Shah, "Deep Ensemble Learning Model for Diagnosis of Lung Diseases from Chest X-Ray Images," *Indian Journal of Science and Technology*, vol. 17, no. 8, pp. 702–712, 2024.

[5] A. Nadal-Martínez et al., "Lung Disease Classification Using Deep Learning and ROI-Based Chest X-Ray Images," *Technologies*, vol. 14, no. 1, 2026.

[6] P. S. Kulkarni et al., "Lung Disease Detection Using Deep Learning Algorithms," *Biomedical Engineering Open Access*, vol. 1, no. 1, 2025.

[7] P. Kwaśniewska et al., "Pulmonary Diseases Identification: Deep Learning Models and Ensemble Learning," *Applied Computer Science*, vol. 21, no. 3, pp. 38–58, 2025.

[8] M. Irtaza et al., "Multi-Label Classification of Lung Diseases Using Deep Learning," *IEEE Access*, vol. 12, 2024.

[9] S. N. Pranavi and V. Sreedevi, "Lung Disease Detection Using Deep Learning," Bachelor of Engineering project report, Sathyabama Institute of Science and Technology, 2023.

[10] Y. S. Kumari, "Deep Learning Approaches for Lung Disease Detection Through Voice Analysis," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no. 10, 2024