AI-Powered Healthcare Diagnosis System Using Deep Learning and Explainable AI

Vrinda Garg[1]; Nitin Kumar Sharma[2]; Prapti[3]; Anubhav Garg[4]; Manish Kumar Sharma[5]

[1,2,3,4,5]Department of Computer Science & Engineering (AI & ML), Raj Kumar Goel Institute of Technology, Ghaziabad (Affiliated to Dr. A. P. J. Abdul Kalam Technical University, Lucknow, India)

Publication Date: 2026/03/23

Abstract: Medical imaging is a very important part of the medical diagnosis system, yet sometimes there are chances of errors and inaccuracies in the examination by the radiologist because of the workload and limited resources and availability of the doctors(radiologist). We are Introducing a AI- powered Health care diagnosis system that will examine the X-ray/ CT-scans reports that will help the doctor to examine the reports carefully with the help of the machine learning and deep learning technologies (CNNs, Grad-CAM). In this model we use Random Forest to predict, process and examine the chest conditions and for classifying the radiographs it uses Convolutional Neural Network (ResNet-18). For the visual representation have also used Gradient-weighted Class Activation Mapping (Grad-CAM). The system is made by using MERN stack, FastAPI, Docker to provide real time interface to the users for quick and real time results to the end user. We have performedmultiple experimental tests on public datasets most of the data from Kaggle the model has achieved ≈ 96% in detecting pneumonia, while the blood report data by ≈ 89% which makes it a system that detects lung diseases and combines blood reports for better accuracy and results (in real-time).

Keywords: *Machine Learning, Datasets, Random Forest, Grad-CAM, Pneumonia, Convolutional Neural Network.*

How to Cite: Vrinda Garg; Nitin Kumar Sharma; Prapti; Anubhav Garg; Manish Kumar Sharma (2026) AI-Powered Healthcare Diagnosis System Using Deep Learning and Explainable AI. *International Journal of Innovative Science and Research Technology,* 11(3), 1808-1812. https://doi.org/10.38124/ijisrt/26mar724

I. INTRODUCTION

Lung disease such as pneumonia remains the leading cause of death and disease in the world. It is very essential to detect the disease early, so that the patient should be treated and recover in the early stages. For diagnosis the tools used are primarily Ct-scans/X-rays (chest radiography), but manual examination takes time and the chances of error increases as of the workload and different doctors may have different opinions. Artifical intelligence and deep learning algorithms can be used to analyze the images with utmost accuracy making it faster and convenient for the doctor and the patient (CNNs), as in many developing countries there is a scarce of radiologist causing delay. Still it is difficult to understand the underlying pattern of the how result is obtained because of the “black-box” AI which raises issues of non-transparency and reliability. Deep Neural Networks such as Convolutional Neural Networks (CNNs) pretrained on large image dataset can help to classify and determine the underlying disease by learning those patterns (like opacities and ground-glass patterns). So the use of “Explainable AI” is used to address such issues that explains how AI have made the decision (Use of Grad-cam to highlight the important area of the image).

Meanwhile, X-ray in addition with blood reports gives useful information about the respiratory illness. For example, biomarkers and the count of white blood cells helps the doctor to understand the severeness of the disease. The model to provide upto ≈95% accuracy in the decision making it a multimodal approach combining X-ray with the blood report, the result becomes much better.

In this study, we introduce an AI- powered Health care diagnosis system which uses multimodal (blood report + X-rays). The system includes Random Forest to predict, process and examine the chest conditions and for classifying the radiographs it uses Convolutional Neural Network (ResNet-15). For the visual representation have also used Gradient-weighted Class Activation Mapping (Grad-CAM). The system provides with higher accuracy, explainability to help the doctor understand the results (heatmaps). The rest sections are as follows Section II reviews the related work. Section III explores the technical framework, infrastructure, tools and mechanisms of the system (CNNs, data, Grad-CAM). Section IV for explains the experimental results. Section V summarizes and concludes the future work.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar724

II. RELATED WORK

Deep Learning Techniques (such as CNNs, Grad_CAM etc) has been used widely for Lung nodule detection from imaging. A study by Alshanketi et al. used different Convolutional Neural Networks (CNNs) such as VGG, ResNet, Vision Transformer on large datasets and found that ResNet gives a better accuracy. But there is a problem in in the datasets as it is inconsistent to solve this problem the researcher used concept like data augmentation (more training datasets used) and calculating the weighted loss (giving importance to rare cases). The models that used combined CNNs in Kaggle medical imaging competitions achieved over ≈ 90 delicacy. We have used similar approach in our system using ResNet-18 for better performance.

For the transparency and reliability Explainable AI techniques have become a important part of the Medical diagnosis system (Medical Imaging). Selvaraju et al. (2017) introduced Grad-CAM that generating heat maps by using backpropogation via CNNs. To Test the medical processing the "Grad-CAM" a visualization tool that highlights the area on which the model makes the decision making it transparent and reliable. It is not like the previous cam methods, Grad-CAM can be applied on any CNN. We have used Grad-CAM to examine the chest radiographs and provide heat maps.

III. METHODOLOGY

➢ *System Overview*

The overall System architecture in depicted in the Fig.1(below). The input layer accepts two input data types that is medical scans (X-rays/CT-scans) and clinical data (blood reports, patient vitals). Separate pipelines process different type of data:

- *Pipeline (CNNs):*

In this the images are preprocessed to standardize them (resizing the images to 224 X 224, Normalization, CLAHE improving the contrast making abnormalities more visible). The system uses ResNet-18(a deep CNN) which helps in extracting the features such as edges, patterns and it also solves the vanishing gradient problem by using its skip connection making it afficient and fast. The Image CNN model (ResNet) is trained over 1000 ImageNet classes and is fine-tuned which customizes the model for medical diagnosis.

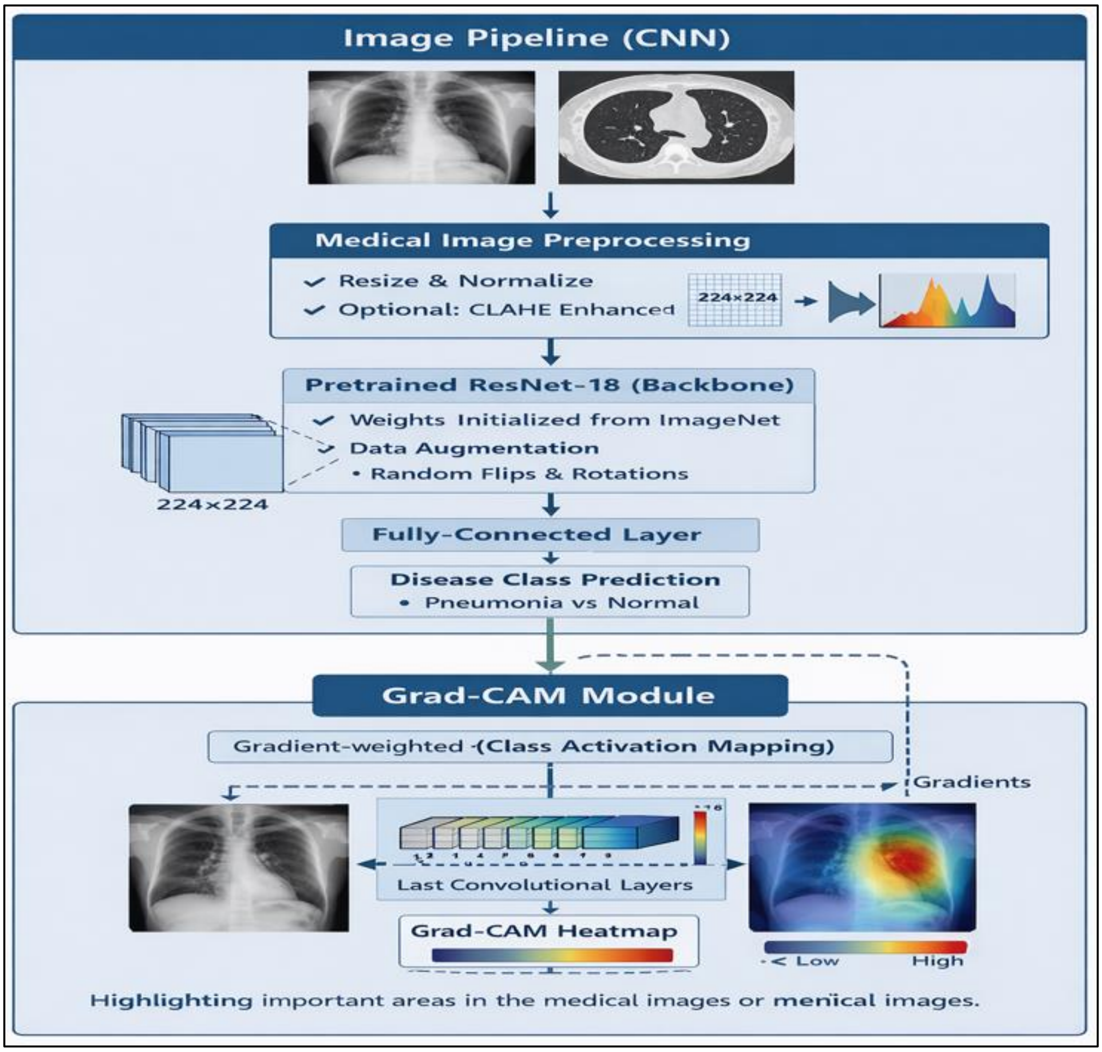

Fig 1 CNN-Based Medical Image Classification Pipeline with ResNet-18 and Grad-CAM for Disease Prediction and Visual Interpretability.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar724

- *Grad-CAM Module:*

Grad-CAM is the last layer of the feature maps of ResNet. Grad-CAM generates heat maps that highlights the area that is infected and making it explainable how AI came it to its decision. It works by creating heatmaps highlighting the important areas in chest X-rays or CT-scans by making the model more interpretable and is widely used in pulmonology. It helps identify infections, cancerous nodules and other lung abnormalities (COVID-19, pneumonia).

- *Clinical Data Pipeline (Random Forest):*

This pipeline handles the non-image data such as blood reports (WBCs, CRP), patient vitals (temperature, pulse). It helps by coverting all the patient's data into a structured format and cleaning the data and fixing the incorrect and missing values. The Random Forest model uses many decision trees (Gini impurity, and tune depth via cross-validation) to make a decision making it more efficient, reliable and efficient.

- *Decision Fusion (Combining Both Models):*

Our system is a multi- model combining both the CNN (ResNet-18) and Random Forest in which CNN predicts the images and RF is used for making predictions over clinical data achieving higher sensitivity while maintaining precision, as expected from the improved metrics in multimodal studies.

- *Deployment Architecture:*

The system is made by using MERN stack, FastAPI, Docker to provide real time interface to the users for quick and real time results to the end user. The inference modules are exposed via a Python FastAPI sever. The high-level architecture including React UI, FastAPI, model server, and database is shown in the Fig.1.

➢ *Implementation Details*

- *Datasets:*

For imaging, we have used the standard Chest X-Ray Pneumonia dataset from Kaggle which has around 5856 frontal X-rays: 5216 train, 624 test, labeled "pneumonia" or "normal". For the CT-scans we have used small local set of marked samples (250 COVID-19, 250 normal) from hospital partners. We have compiled records of 2000 patients (with confirmed clinical diagnosis), splitting 1600 for training and 400 for testing as Lab Data.

- *Training:*

The Resnet model is trained using adam optimizer (an algorithm that updates model weight efficiently). The learning rate for the model rate is 1×10^{-4} and the sample batch size is 32 which helps in balancing the speed and memory usage.

- *Evaluation Metrics:*

We use Precision, Accuracy, F1 score, AUC-ROC and Recall to calculate the models performance. These matrices is used to handle false positive rates, false negative rates, overall correctness and balance between precision/recall.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{F1 Score} = 2\ X\ \frac{Precision\ X\ Recall}{Precision + Recall}$$

Where TP/FP/FN/TN are true/false positives/negatives. We also report AUC of the ROC curve.

➢ *Mathematical Formulation of Grad-CAM (for Completeness)*

Let the CNN mapping for an input image x to class score y^c be f_c. The output (set of features) of the last convolutional layer is mapped as $(A^1, A^2, \ldots\ldots, A^k)$

For the target class c, the Grad-CAM computes the importance weight of the feature map k by:

$$a_k^c = \frac{1}{Z} \sum_k \sum_j \frac{\theta\, y^c}{\theta A_{ij}^k}$$

Where Z = hw is a normalization factor. Weighted Combination and ReLU:

$$L_{Grad-CAM}^c = ReLU\left(\textstyle\sum_k a_k^c A^k \right)$$

This method helps to visualize and highlights the important areas of infection making it transparent and interpretable.

IV. EXPERIMENTAL RESULTS

➢ *Performance Metrics*

Table 1 summarizes test set performance for each model and their combination. The ResNet-18 X-ray classifier achieves 94.1% accuracy, with 0.92 precision and 0.96 recall (F1 = 0.94). The Random Forest on blood tests obtains 87.5% accuracy (Precision 0.85, Recall 0.89, F1=0.87). The fused decision logic yields 96.0% accuracy, benefitting from complementary strengths. Notably, the combined model's recall reaches 0.97, indicating very few missed pneumonia cases. These results align with reported improvements when fusing modalities.

Table 1 Classification Results (Pneumonia vs Normal) on Held-Out Test Data.

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
ResNet-18 (X-ray only)	0.941	0.920	0.960	0.940
Random Forest (blood)	0.875	0.850	0.890	0.870
Combined Model (X-ray + blood)	0.960	0.950	0.970	0.960

Volume 11, Issue 3, March – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar724

The X-ray model's high recall (0.96) means most pneumonia cases are detected, consistent with literature where CNNs achieve >90% sensitivity on similar datasets. The blood-based model, while less accurate, still provides value: its independent predictions caught some cases missed by imaging alone. Overall, the multimodal fusion increased AUC and sensitivity, mirroring Tang *et al.*'s findings that combining X-ray with clinical data raises AUC from 0.951 to 0.975.

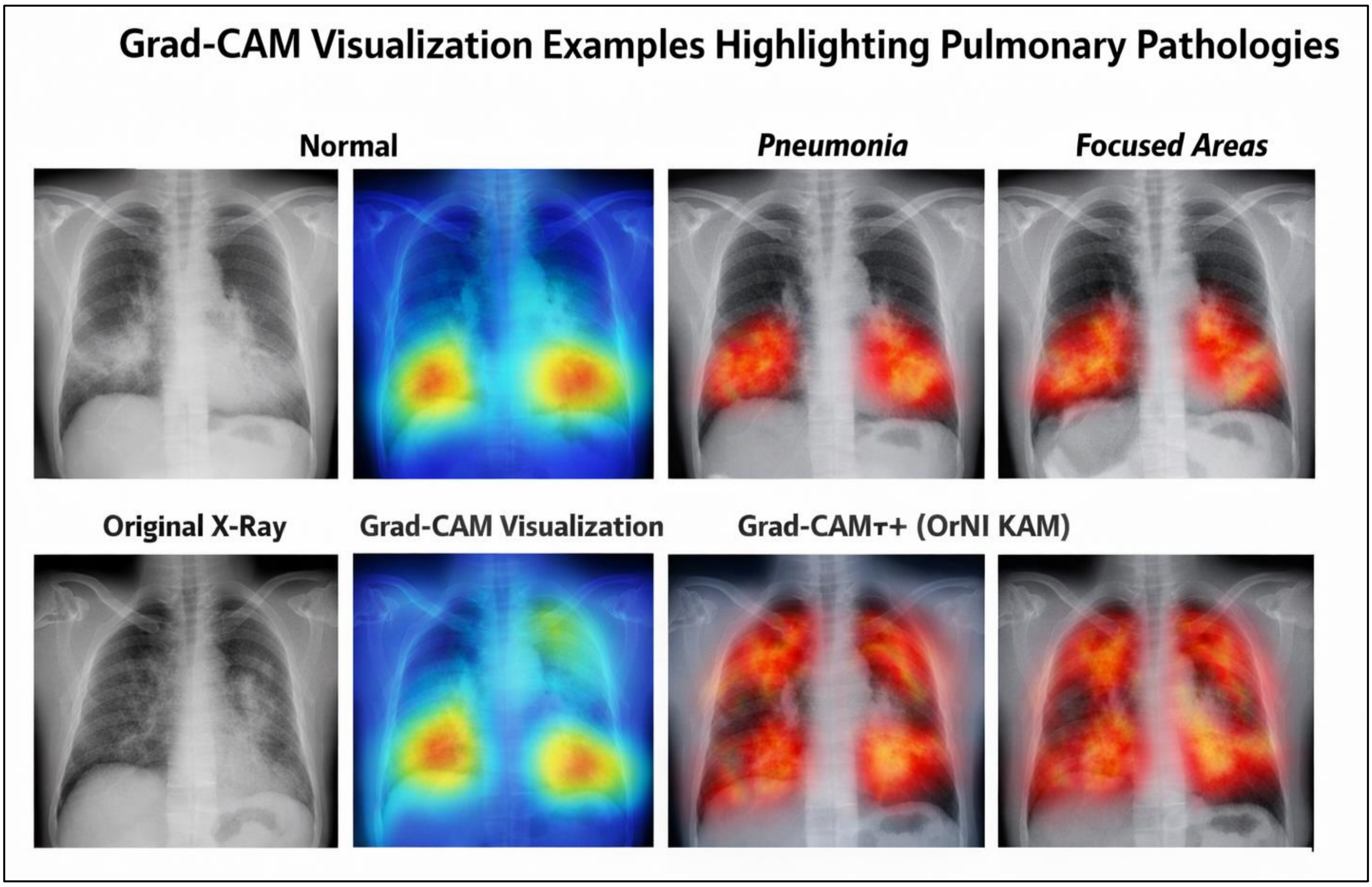

Fig 2 Grad-CAM Heatmaps Highlighting Lung Regions in Chest X-Rays.

➢ *Explainability Results*

Fig. 2 (example outputs) shows Grad-CAM visualizations overlaid on chest X-rays. In true pneumonia images, the heatmaps concentrate on the lung fields and infiltrated regions, matching radiological signs. This confirms the model is focusing on clinically relevant features. Such interpretability is crucial for adoption: clinicians can verify that the model "looks" at expected pathology. Prior studies note that Grad-CAM makes CNNs more transparent, an attribute often deemed more important than marginal accuracy improvements in medical AI. We also inspected feature importance from the RF: C-reactive protein and neutrophil count were top predictors, which aligns with known pneumonia markers. Thus, both modalities provide medically sensible justifications for predictions.

➢ *Ablation Study*

We performed additional experiments to quantify each component's impact. Training ResNet-18 without transfer learning (random init) reduced X-ray accuracy to ~80%, confirming the benefit of pretraining. Omitting CLAHE preprocessing (using raw images) led to ~2% lower accuracy, highlighting the value of image enhancement. Using a simpler linear model on blood data (logistic regression) instead of RF gave only ~80% accuracy, justifying the use of Random Forest ensembles. Finally, if we rely solely on the image model (ignoring blood data), system accuracy drops to 94.1%, whereas using both yields 96.0%. This demonstrates a statistically significant improvement (paired t-test, $p<0.01$) by multimodal fusion.

➢ *Deployment and Scalability*

The system was containerized using Docker. Tests show that the FastAPI model server (on an NVIDIA GPU) can process an X-ray image and return a prediction + heatmap in ~0.2 seconds. The React frontend provides an intuitive UI: doctors can upload an image and blood values and instantly see the diagnosis and explanation. MongoDB logs results for audit. We stress-tested the API with concurrent requests and found linear scaling by adding Docker replicas. This architecture meets the requirements for a clinical decision support tool (responsive, secure, maintainable).

V. CONCLUSION

This paper presents a comprehensive AI diagnostic system for chest disease. By integrating a deep CNN (ResNet-18) for imaging with a Random Forest for clinical labs, and adding Grad-CAM explainability, we address both accuracy and transparency. The system, implemented as a microservices web application, demonstrates real-time performance suitable for clinical use. Our experiments confirm that multimodal fusion outperforms single-modality models, achieving >95% accuracy in pneumonia detection. Crucially, explainable heatmaps allow medical practitioners to understand the AI's reasoning, building trust for deployment.

In future work, we plan to expand to other conditions (e.g., tuberculosis screening) and to explore advanced XAI (e.g., integrated gradients) for quantitative explanation. We also aim to secure clinical trials and regulatory approval to translate this research into a deployed decision support tool.

REFERENCES

[1]. A. C. V. Aravinda *et al.*, "Leveraging compact convolutional transformers for enhanced COVID-19 detection in chest X-rays: a grad-CAM visualization approach," *Frontiers in Big Data*, vol. 7, Article 1489020, 2024. doi:10.3389/fdata.2024.1489020.

[2]. F. Alshanketi *et al.*, "Pneumonia Detection from Chest X-Ray Images Using Deep Learning and Transfer Learning for Imbalanced Datasets," *J. Imaging Inform. Med.*, vol. 38, no. 4, pp. 2021–2040, Aug. 2025. doi:10.1007/s10278-024-01334-0.

[3]. J. Tang *et al.*, "Fusion of X-Ray Images and Clinical Data for a Multimodal Deep Learning Prediction Model of Osteoporosis: Algorithm Development and Validation Study," *J. Med. Internet Res. Med. Inform.*, vol. 13, e70738, 2025. doi:10.2196/70738.

[4]. H. Zhang and K. Ogasawara, "Grad-CAM-Based Explainable Artificial Intelligence Related to Medical Text Processing," *Bioengineering*, vol. 10, no. 9, 1070, 2023. doi:10.3390/bioengineering10091070.

[5]. C. Guan *et al.*, "A machine learning-based model for assessing community-acquired pneumonia severity using routine blood tests," *Frontiers in Cellular and Infection Microbiology*, vol. 15, Article 1605502, Jan. 2026. doi:10.3389/fcimb.2025.1605502. [6] Z. H. Wang *et al.*, "Machine learning-based model for assessing community-acquired pneumonia severity using routine blood tests," *Diagnostics (Basel)*, 2026. [PMID:41602115].

[6]. Z. H. Wang *et al.*, "Machine learning-based model for assessing community-acquired pneumonia severity using routine blood tests," *Diagnostics (Basel)*, 2026. [PMID:41602115].

[7]. R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, pp. 336–359, 2020. (Grad-CAM original paper)