© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715093

Explainable AI for Disease Diagnosis Feature-Based and Visual Interpretability in CKD and Lung Disease Detection

PROF. LEENA SHAH[1], RAJ TIWARI[2], SUMIT SINGH[3], SAHIL JADHAV[4], AJAY KUMAWAT[5]
[1]*Professor, Department of Electronics and Telecommunications Engineering Lokmanya Tilak College of Engineering, Navi Mumbai, India*
[2, 3, 4, 5]*Student, Department of Electronics and Telecommunications and Engineering Lokmanya Tilak College of Engineering, Navi Mumbai, India*

Abstract- ***Explainable Artificial Intelligence (XAI) is crucial in medical diagnosis to enhance transparency, interpretability, and trust in AI-driven decision-making. This study explores the application of XAI techniques for Chronic Kidney Disease (CKD) prediction and lung disease detection. For CKD, a machine learning model was developed using patient clinical data, with SHAP (SHapley Additive exPlanations) employed to identify the most influential features affecting predictions. The results demonstrate that attributes such as serum creatinine, blood urea, and hemoglobin levels significantly impact CKD risk assessment. For lung disease detection, a deep learning model was trained on chest X-ray images, and Grad-CAM (Gradient-weighted Class Activation Mapping) was applied to generate visual explanations, highlighting the critical regions influencing model decisions. Experimental results show that both methods improve the interpretability of AI predictions, aiding healthcare professionals in understanding and validating the model's reasoning. The study highlights the importance of integrating explainability into medical AI models to ensure reliability, facilitate clinical adoption, and enhance patient trust. Future work includes expanding the dataset and exploring additional XAI techniques to further improve model transparency.***

Keywords: Explainable AI (XAI), Chronic Kidney Disease (CKD) Prediction, Lung Disease Detection, SHAP (SHapley Additive exPlanations), Grad-CAM (Gradient-weighted Class Activation Mapping), Medical AI Interpretability

I. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) in healthcare has revolutionized diagnostic processes, enabling clinicians to analyze vast amounts of medical data with unprecedented speed and accuracy. However, as these AI systems become more sophisticated, the complexity of their decision-making processes often leads to a significant challenge: the lack of transparency and interpretability. Healthcare professionals frequently encounter AI-generated diagnoses that they cannot fully understand or explain, resulting in hesitancy to trust these systems and integrate them into clinical practice.

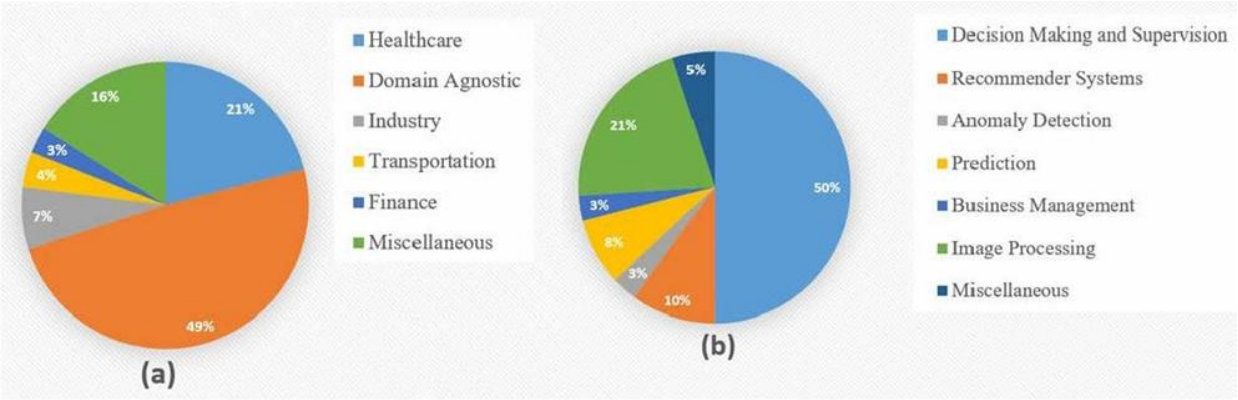

Figure 1 : EXAI research statistics for different applications (a) Domain and (b) Task applied.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715093

Explainable Artificial Intelligence (XAI) emerges as a critical solution to this dilemma, aiming to demystify AI models and provide clear insights into their reasoning. By making the inner workings of AI algorithms more accessible, XAI not only enhances clinician understanding but also fosters accountability and ethical considerations in patient care. This project seeks to develop a robust XAI framework tailored specifically for healthcare diagnostics, addressing the pressing need for clarity and interpretability in AI applications. Through case studies and practical applications, we aim to demonstrate how XAI can empower healthcare providers, improve diagnostic accuracy, and ultimately enhance patient outcomes in an increasingly data-driven medical landscape.

Chronic Kidney Disease (CKD) is a progressive condition that can lead to kidney failure if not diagnosed early. Predictive models using machine learning have shown promise in identifying high-risk individuals based on clinical data. However, to ensure their reliability in medical practice, it is essential to understand which features (e.g., serum creatinine, blood urea, hemoglobin levels) contribute to the predictions. SHAP (SHapley Additive exPlanations) is an effective XAI technique that provides feature importance, enabling better decision-making in CKD prediction.

Similarly, Viral Pneumonia and Lung Opacity are critical lung conditions that require early and accurate detection. Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated high accuracy in identifying these diseases from chest X-ray images. However, without interpretability, their clinical acceptance remains limited. Grad-CAM (Gradient-weighted Class Activation Mapping) is utilized to generate heatmaps that highlight the most relevant regions in the X-ray images, providing a visual explanation of the model's decisions.

Recent research has primarily focused on improving the accuracy of AI models in medical diagnosis. However, there remains a gap in integrating explainability techniques to make these models more interpretable and clinically useful. This study addresses this gap by combining SHAP for CKD prediction and Grad-CAM for Viral Pneumonia and Lung Opacity detection, demonstrating how XAI enhances transparency and trust in AI-powered medical diagnostics.

II. METHODOLOGY

This study implements an Explainable AI (XAI) approach for disease diagnosis by integrating interpretability techniques with machine learning and deep learning models. The methodology consists of multiple stages, including data acquisition, preprocessing, model training, evaluation, and explainability analysis using SHAP and Grad-CAM.

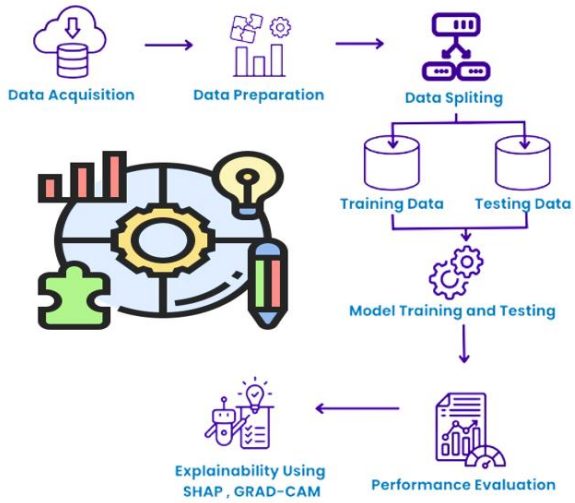

Figure 2: Methodology

1. Data Acquisition:

- Chronic Kidney Disease (CKD) Prediction: A structured clinical dataset containing biochemical and physiological parameters was used. Key attributes include blood pressure, hemoglobin, serum creatinine, and blood urea levels.
- Lung Disease Detection: A dataset of chest X-ray images was utilized, consisting of samples labeled as Normal, Lung Opacity, and Viral Pneumonia.

2. Data Preprocessing

- CKD Data: Missing values were handled using mean imputation for numerical features and mode imputation for categorical features. The dataset was standardized to improve model performance.
- Lung Disease Data: Images were resized and normalized to ensure uniformity before being fed into the deep learning model. Data augmentation

techniques, such as rotation, flipping, and contrast adjustments, were applied to enhance model generalization.

3. Model Training and Testing

- CKD Prediction Model: A machine learning-based classification model was developed using algorithms like Random Forest and XGBoost, with hyperparameter tuning performed to optimize accuracy.
- Lung Disease Detection Model: A custom Convolutional Neural Network (CNN) architecture based on VGG16 was trained on chest X-ray images to classify the conditions into Normal, Lung Opacity, and Viral Pneumonia.

4. Performance Evaluation

- The models were evaluated using accuracy, precision, recall, and F1-score to measure classification effectiveness.
- A confusion matrix was generated to analyze misclassification patterns.

5. Explainability using XAI Techniques

To enhance interpretability, the following explainability techniques were employed:

- SHAP (SHapley Additive exPlanations): Applied to the CKD prediction model to determine the feature importance of clinical parameters. A force plot and summary plot were generated to illustrate the contribution of each parameter to the model's decision.
- Grad-CAM (Gradient-weighted Class Activation Mapping): Used for lung disease detection to visualize the important regions in X-ray images that influenced the model's classification. Heatmaps were generated to highlight affected lung areas, making AI-based diagnoses more transparent for radiologists.

III. MODELING AND ANALYSIS

1. Model Description

This study utilizes two distinct AI-based models for disease diagnosis, incorporating Explainable AI (XAI) techniques for interpretability.

(A) Chronic Kidney Disease (CKD) Prediction Model

- Model Type: Machine Learning Classifier
- Algorithms Used: Random Forest, Naive Bayes, SVM, XGBoost (best-performing model selected)
- Input Features: Clinical parameters such as blood pressure, hemoglobin, serum creatinine, blood urea, and other biochemical markers.
- Output: Binary classification (0 - No CKD, 1 - CKD)
- Explainability: SHAP (SHapley Additive exPlanations) to analyze feature contributions.

(B) Lung Disease Detection Model

- Model Type: Deep Learning (CNN-based Image Classifier)
- Architecture Used: Custom CNN model based on Inception V3
- Input Data: Chest X-ray images
- Output Classes:
  - 0 - Normal
  - 1 - Lung Opacity
  - 2 - Viral Pneumonia
- Explainability: Grad-CAM (Gradient-weighted Class Activation Mapping) for heatmap visualization.

2. Dataset Description

(A) CKD Dataset

Feature Name	Description	Type
id	Unique identifier for the patient	Integer
age	Age of the patient in years	Numeric
bp	Blood pressure (in mm/Hg)	Numeric
sg	Specific gravity (urine concentration)	Numeric
al	Albumin levels in urine (proteinuria)	Categorical (0-5)
su	Sugar levels in urine	Categorical (0-5)
rbc	Red Blood Cells count	Categorical (normal/abnormal)
pc	Pus Cells in urine	Categorical (normal/abnormal)

pcc	Pus Cell Clumps in urine	Categorical (present/not present)
ba	Bacteria in urine	Categorical (present/not present)
bgr	Blood Glucose Random (mg/dL)	Numeric
bu	Blood Urea (mg/dL)	Numeric
sc	Serum Creatinine (mg/dL)	Numeric
sod	Sodium levels in blood (mEq/L)	Numeric
pot	Potassium levels in blood (mEq/L)	Numeric
hemo	Hemoglobin levels (g/dL)	Numeric
pcv	Packed Cell Volume (%)	Numeric
wc	White Blood Cell Count (cells/cu mm)	Numeric
rc	Red Blood Cell Count (millions/cu mm)	Numeric
htn	Hypertension status	Categorical (yes/no)
dm	Diabetes Mellitus status	Categorical (yes/no)
cad	Coronary Artery Disease status	Categorical (yes/no)
appet	Appetite status	Categorical (good/poor)
pe	Pedal Edema (swelling in feet)	Categorical (yes/no)
ane	Anemia status	Categorical (yes/no)
class	CKD status (target variable)	Categorical (ckd/notckd)

Dataset Link

(B) Lung Disease X-ray Dataset

Class Label	Description	Number of Images
0	Normal	1125
1	Lung Opacity	1250
2	Viral Pneumonia	1100

Dataset link

3. Model Architecture

(A) CKD Model Architecture

- Input Layer: Processes numerical and categorical features.
- Feature Selection: Important attributes identified using SHAP.
- Classification Layer: Uses Random Forest for final prediction.
- Output Layer: Binary classification (CKD/No CKD).

(B) CNN Model for Lung Disease Detection model Architecture

The proposed model is a deep convolutional neural network designed for medical image classification, leveraging a MobileNet-based architecture.

1. Input Layer
   - Input: (256, 256, 3) RGB Chest X-ray images
2. Feature Extraction (MobileNet Backbone)
   - Conv2D Layer: 1 (3x3 kernel, stride 2, ReLU activation)
   - Depthwise Separable Convolutions: 13 layers
   - Batch Normalization: 13 layers (after each convolution)
   - ReLU Activation: 13 layers (after each batch norm)
3. Global Feature Aggregation
   - Global Average Pooling (GAP): 1 layer
4. Fully Connected Layers
   - Dense Layer: 1 (128 neurons, ReLU activation)
   - Dropout Layer: 1 (rate = 0.3 to prevent overfitting)
   - Output Dense Layer: 1 (3 neurons, Softmax activation)
5. Optimization & Training
   - Optimizer: Adam
   - Loss Function: Categorical Cross-Entropy

Total Layer Count: 30 (including convolutions, batch norm, activations, pooling, dense, and dropout layers).

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715093

IV. RESULTS AND DISCUSSION

The experimental results are presented for both Chronic Kidney Disease (CKD) detection and Lung Disease Prediction. The models were evaluated using accuracy and loss metrics, along with confusion matrices to illustrate performance.

1. Chronic Kidney Disease Detection (Text Data)
The classification models used for CKD prediction and their corresponding accuracy and loss values are shown in Table 1.

Table 1: CKD Detection Model Performance

Model	Accuracy	Loss
Naïve Bayes	0.91	2.89
Random Forest	1.00	0.03
SVM	0.98	0.14
XGBoost	0.98	0.02

The confusion matrix (Table 2) shows the correct and incorrect predictions for CKD classification.

Table 2: Confusion Matrix for CKD Detection

Model	Predicted 0	Predicted 1
Actual 0	Naïve Bayes: 67	Naïve Bayes: 9
	Random Forest: 76	Random Forest: 0
	SVM: 71	SVM: 4
	XGBoost: 70	XGBoost: 5
Actual 1	Naïve Bayes: 6	Naïve Bayes: 8
	Random Forest: 2	Random Forest: 44
	SVM: 1	SVM: 43
	XGBoost: 1	XGBoost: 28

2. Lung Disease Prediction (Image Data)
For lung disease classification, a Custom CNN model and Inception V3 were used. Table 3 presents their accuracy and loss values.

Table 3: Lung Disease Prediction Model Performance

Model	Accuracy	Loss
Custom Model	0.7883	0.4657
Inception V3	0.9243	0.3882

Table 4 provides the confusion matrix for lung disease prediction.

Table 4: Confusion Matrix for Lung Disease Prediction

Actual Class	Predicted (Inception V3)	Predicted (Custom Model)
0 (Lung Opacity)	248	243
1 (Normal)	58	52
2 (Viral Pneumonia)	30	25

The results demonstrate that Inception V3 outperformed the custom CNN model in accuracy, achieving 92.43% accuracy compared to 78.83% for the custom model. However, the custom model shows promise for further improvement.

Explainable AI Analysis: SHAP and Grad-CAM Visualizations
This visualization showcases Explainable AI techniques: SHAP plots highlight feature importance in Chronic Kidney Disease prediction, while Grad-CAM heatmaps provide interpretability for lung disease detection from X-ray images.

1. SHAP Force Plot: Individual Prediction Explanation

Figure 3 : Force Plot

The force plot explains how different features contribute to the AI model's prediction for an individual case in Chronic Kidney Disease (CKD) detection.

- The base value (0.3749) represents the average model output before considering individual feature effects.
- Red features (e.g., age, sg, al, dm, htn) push the prediction towards a higher probability of CKD.
- Blue features (e.g., hemo, sc, pcv, rc, sod) push the prediction towards a lower probability of CKD.
- The final predicted value (0.20) is computed by adjusting the base value according to these feature contributions.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715093

- This plot helps in understanding why the model made a specific decision for an individual patient.

2. SHAP Summary Plot: Feature Importance Analysis

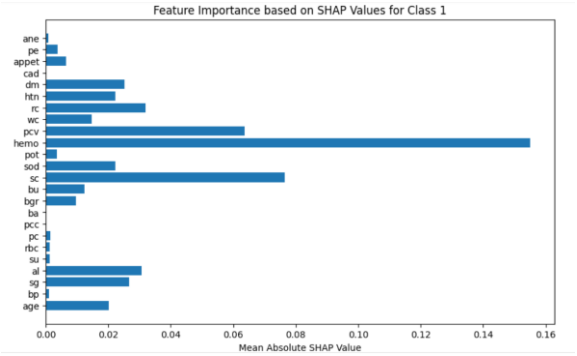

Figure 4: Summary Plot

The summary plot shows the average impact of each feature on the model's predictions across multiple patients.

- X-axis: Represents the Mean Absolute SHAP Value, which indicates how much a feature contributes to the model's decision.
- Y-axis: Lists the input features, ranked from most to least influential.
- Key observations:
  - Hemoglobin (hemo) has the highest impact on CKD prediction.
  - Packed Cell Volume (pcv) and Serum Creatinine (sc) are also significant.
  - Features like bacteria (ba) and pus cell clumps (pcc) have relatively low importance.

This plot provides an overall global interpretability of the AI model, helping clinicians understand which factors influence CKD predictions the most.

3. Original Chest X-ray Image , Grad-CAM Heatmap and Overlayed Grad-CAM

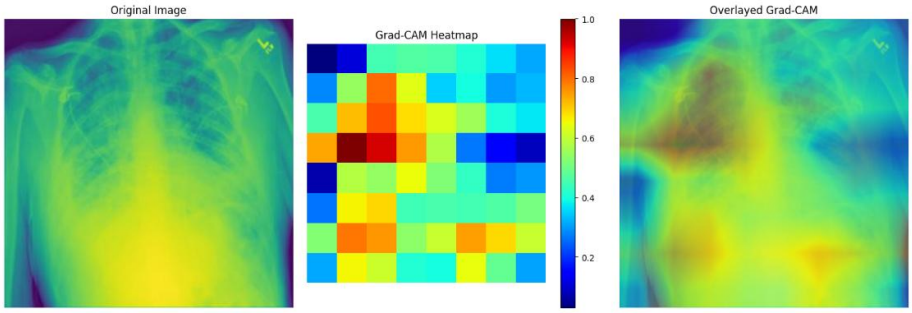

Figure 5: Original Chest X-ray Image , Grad-CAM Heatmap and Overlayed Grad-CAM

Original X-ray Image

- This is the raw chest X-ray image used for diagnosing lung diseases like Viral Pneumonia and Lung Opacity.
- The AI model processes this image to detect abnormalities.

Grad-CAM Heatmap

- This heatmap shows the activation regions of the AI model when predicting the disease.
- Red and yellow areas indicate the most important regions for the model's decision, while blue areas contribute less.

Overlayed Grad-CAM on X-ray Image

- This visualization overlays the Grad-CAM heatmap on the original X-ray to show where the model focused when making its decision.
- Brighter regions (red/yellow) indicate potential lung abnormalities, helping in explainability for clinicians.

CONCLUSION

This study presents an AI-driven approach for diagnosing Chronic Kidney Disease (CKD) and Lung Diseases using text and image data, respectively. Multiple machine learning models were evaluated for CKD detection, with Random Forest and XGBoost achieving the highest accuracy. For lung disease classification, Inception V3 outperformed the custom

CNN model, demonstrating its effectiveness in handling medical image classification.

The results highlight the potential of explainable AI (XAI) in medical diagnosis, providing transparency in predictions. The CKD model effectively classifies patient records based on clinical parameters, while the lung disease model distinguishes between different conditions from X-ray images.

Despite promising results, challenges such as dataset limitations and model generalization remain. Future work will focus on enhancing model performance through data augmentation, advanced deep learning architectures, and further explainability techniques. The integration of AI into clinical workflows could assist healthcare professionals in early diagnosis, decision-making, and personalized treatment recommendations.

V. ACKNOWLEDGEMENTS

I remain immensely obliged to Prof. Leeena shah for providing me with the idea of the topic, for her invaluable support in gathering resources, and for her guidance and supervision, which made this work successful.

I would like to extend my sincere gratitude to the Head of the EXTC Department, Dr. Ravindra Duche, and Principal Dr. Subhash Shinde for their continuous encouragement and support throughout this research.
I am also thankful to the faculty and staff of the EXTC Department and Lokmanya Tilak College of Engineering, Navi Mumbai, for their invaluable support.

It has indeed been a fulfilling experience working on this research, and I appreciate everyone who contributed to its success.

REFERENCES

[1] Madhuri, Nandini S, and Mrs. Mona, "Diagnosis of Chronic Kidney Disease Using Machine Learning," International Research Journal of Engineering and Technology (IRJET), vol. 09, no. 07, July 2022, pp. 642. e-ISSN: 2395-0056, p-ISSN: 2395-0072.

[2] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, and R. Sharma, "Explainable AI for Healthcare 5.0: Opportunities and Challenges," IEEE Access, vol. 2022, pp. 1-16, Aug. 2022. DOI: 10.1109/ACCESS.2022.3197671.

[3] R. Chelghoum, I. Ameur, A. Hameurlaine, and S. Jacquir, "Transfer Learning Using Convolutional Neural Network Architectures for Brain Tumor Classification from MRI Images," IFIP Advances in Information and Communication Technology, in Artificial Intelligence Applications and Innovations, May 2020, pp. 189-200. DOI: 10.1007/978-3-030-49161-1_17.

[4] P. Devi, S. Sagar, and H. Rohil, "Prediction of Lung Disease Using Machine and Deep Learning Techniques: A Review," Chaudhary Devi Lal University, 2022.