World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

MEDXAI: A SELF-SUPERVISED EXPLAINABLE MEDICAL IMAGE INTELLIGENCE FRAMEWORK WITH SIMILAR CASE RETRIEVAL, ACTIVE LEARNING, AND CLINICAL DECISION SUPPORT

Bhargavi Jangam, Praharika Reddy Gongidi, Varun Vijay Amgoth and Sai Karthik Sriram *

Department of CSE (AI and ML) of ACE Engineering College, Hyderabad, India.
* Corresponding Author

World Journal of Advanced Research and Reviews, 2026, 32(02), 976–981

Publication history: Received on 05 October 2026; revised on 11 November 2026; accepted on 13 November 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.32.2.2916

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Hospitals generate large numbers of medical images, but only a small portion are labeled by expert doctors, and most existing AI models predict diseases without explaining their decisions. This paper presents MedXAI, a self-supervised medical image intelligence framework that learns useful image representations from unlabeled medical images and fine-tunes them using a small labeled dataset for disease prediction. The framework integrates Explainable AI using Grad-CAM, similar case retrieval using FAISS, Image Quality Assessment, Out-of-Distribution (OOD) Detection, and Active Learning. It also generates confidence scores, analytics, and downloadable PDF reports to support clinical decision-making. The proposed framework aims to reduce dependency on labeled data, improve prediction transparency, and provide a reliable approach for medical image analysis.

Keywords: Self-Supervised Learning; Explainable AI; Grad-CAM; Similar Case Retrieval; Active Learning; Out-of-Distribution Detection; Medical Image Analysis; Deep Learning.

1 INTRODUCTION

Medical imaging plays an important role in disease diagnosis and clinical decision-making. Hospitals generate large amounts of medical images every day, including X-rays, CT scans, MRI scans, and other imaging data. However, labeling these images requires expert medical professionals, making the process expensive and time-consuming.

Traditional medical AI models mainly depend on large labeled datasets for training. Their performance may be limited when only a small number of labeled medical images are available. Furthermore, many existing AI systems function as black boxes, providing predictions without clearly explaining the reasons behind their decisions.

To overcome these limitations, MedXAI proposes a self-supervised medical image intelligence framework that learns meaningful representations from unlabeled medical images and performs disease prediction using a small labeled dataset. The framework integrates Grad-CAM for visual explanations, FAISS for similar medical case retrieval, Image Quality Assessment, and OOD Detection to improve prediction reliability.

The system also incorporates Active Learning to improve model performance through expert feedback. Confidence scores and downloadable PDF reports provide additional information to support doctors in understanding and reviewing the predictions. This approach aims to make medical AI more transparent, data-efficient, and useful for clinical decision support.

2 LITERATURE SURVEY

2.1 Huihui Yu and Qun Dai (2026) – Anatomy-Guided Self-Supervised Distillation Learning for Medical Image Analysis

Yu and Dai (2026) proposed AG-SSD, which incorporates anatomical priors, cross-view consistency, and self-distillation to learn robust representations from 3D CT and MRI images. Their approach improves classification and segmentation performance with limited labeled data. This work demonstrates the importance of anatomical information in self-supervised medical image analysis and supports the representation-learning approach used in MedXAI.

2.1.1 Methodologies and Algorithms

The methodology uses anatomical priors and cross-view consistency to guide representation learning. Self-distillation is applied to improve the quality of learned features from medical images. The framework focuses on understanding anatomical structures and learning useful representations from limited labeled medical imaging data.

2.2 Munib Mesinovic and Tingting Zhu (2026) – Clinical XAI: A Tutorial on Concepts, Methods, and Modalities from Linear Models to LLMs

Mesinovic and Zhu (2026) presented a tutorial on Clinical Explainable AI methods, evaluation metrics, and explainability techniques across electronic health records, medical imaging, genomics, and multimodal data. Their work explains modern Clinical XAI concepts and practical evaluation approaches. It provides useful background for developing transparent medical image intelligence systems such as MedXAI

2.2.1 Methodologies and Algorithms

The study discusses different explainability methods and evaluation metrics used in clinical AI applications. It covers approaches ranging from traditional linear models to large language models and examines their applications across different healthcare data modalities.

2.3 Mohammad Ennab and Hamid Mcheick (2025) – Enhancing Pneumonia Diagnosis Through AI Interpretability: Comparative Analysis of Pixel-Level Interpretability and Grad-CAM on X-ray Imaging With VGG19

Ennab and Mcheick (2025) compared Pixel-Level Interpretability and Grad-CAM techniques for explaining pneumonia classification using the VGG19 model on chest X-ray images. Their work improves model transparency and helps clinicians understand AI-generated predictions through visual explanations. This study supports the use of Grad-CAM in MedXAI for interpreting disease prediction results.

2.3.1 Methodologies and Algorithms

The methodology compares Pixel-Level Interpretability and Grad-CAM to explain the predictions of a VGG19-based pneumonia classification model. Visual explanation techniques are used to identify important image regions contributing to the classification results.

2.4 Dariush Lotfi, Mohammad-Ali Nikouei Mahani, Mohamad Koohi-Moghadam and Kyongtae Ty Bae (2026) – Safeguarding AI in Medical Imaging: Post-Hoc Out-of-Distribution Detection with Normalizing Flows

Lotfi et al. (2026) proposed a post-hoc Normalizing Flow framework for detecting out-of-distribution medical images without retraining existing AI models. Their approach improves the reliability and safety of medical AI systems by identifying unknown or abnormal images. This concept is relevant to MedXAI, which incorporates OOD Detection to identify unsupported scan types before performing disease prediction.

2.4.1 Methodologies and Algorithms

The methodology uses post-hoc Normalizing Flow techniques to identify medical images that differ from the distribution of supported training data. This helps detect unfamiliar or unsupported inputs and improves the reliability of AI-based medical image analysis.

Table 1: Comparison Table of Literature Survey

S. No	Author(s)	Title	Methodology Used	Findings from the Reference Paper
1.	Huihui Yu and Qun Dai (2026)	Anatomy-Guided Self-Supervised Distillation Learning for Medical Image Analysis	Used anatomical priors, cross-view consistency, and self-distillation to learn representations from 3D CT and MRI images.	Improves classification and segmentation with limited labeled data and provides better anatomical understanding.
2.	Munib Mesinovic and Tingting Zhu (2026)	Clinical XAI: A Tutorial on Concepts, Methods, and Modalities from Linear Models to LLMs	Reviewed clinical XAI methods and explainability techniques for healthcare applications.	Covers modern Clinical XAI concepts and practical evaluation methods for trustworthy healthcare AI.
3.	Mohammad Ennab and Hamid Mcheick (2025)	Enhancing Pneumonia Diagnosis Through AI Interpretability: Comparative Analysis of Pixel-Level Interpretability and Grad-CAM on X-ray Imaging With VGG19	Compared Pixel-Level Interpretability and Grad-CAM techniques for explaining pneumonia classification using the VGG19 model on chest X-ray images.	Improves model transparency and helps clinicians understand AI predictions through visual explanations.
4.	Dariush Lotfi et al. (2026)	Safeguarding AI in Medical Imaging: Post-Hoc Out-of-Distribution Detection with Normalizing Flows	Proposed a post-hoc Normalizing Flow framework to detect out-of-distribution medical images without retraining existing AI models.	Improves reliability and safety by identifying unknown or abnormal medical images.

2.5 System Architecture

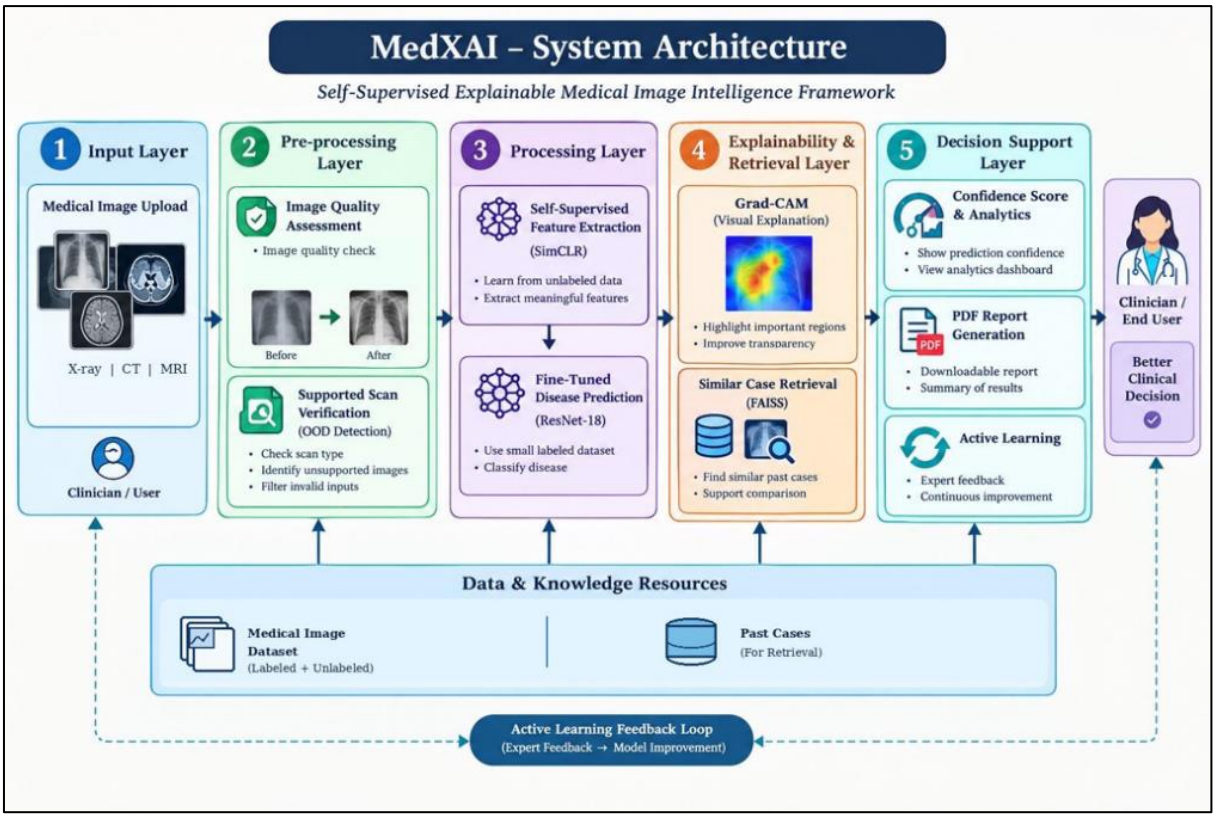


Figure 1: System Architecture

The system architecture of MedXAI is designed to process medical images through multiple stages, including image quality assessment, supported scan verification, self-supervised feature extraction, disease prediction, explainability, and similar case retrieval.

The Input Layer receives medical images uploaded by users. The uploaded images are passed to the Image Quality Assessment module, which checks image quality before further processing. The supported scan verification module performs OOD Detection to identify unsupported or unfamiliar medical images.

The Processing Layer performs feature extraction and disease prediction. Self-Supervised Learning is used to learn meaningful representations from unlabeled medical images. The extracted features are used by the fine-tuned disease prediction model, which is based on the proposed ResNet-18 and SimCLR technology stack.

The Explainability and Retrieval Layer generates visual explanations using Grad-CAM and retrieves similar medical cases using FAISS. These components help users understand the prediction and compare the uploaded image with previously available cases.

The Decision Support Layer generates prediction confidence scores and provides analytics and downloadable PDF reports. Active Learning supports continuous model improvement through expert feedback. The overall framework is designed to assist doctors in reviewing AI-generated results and making informed clinical decisions.

3 PROPOSED METHODOLOGY

3.1 Medical Image Upload

The first step involves uploading a medical image into the MedXAI framework. The system accepts supported medical image inputs for analysis. The uploaded image acts as the primary input for the subsequent processing stages.

3.2 Image Quality Assessment

After uploading, the image is passed to the Image Quality Assessment module. This module checks whether the uploaded medical image satisfies the required quality conditions for further analysis. Poor-quality images are identified to reduce the possibility of unreliable predictions.

3.3 Supported Scan Verification and OOD Detection

The system performs Out-of-Distribution Detection to verify whether the uploaded image belongs to the supported image categories. This stage identifies unfamiliar or unsupported medical images before disease prediction. It helps improve system reliability and prevents unsupported inputs from being treated as valid diagnostic cases.

3.4 Self-Supervised Feature Extraction

The proposed framework uses Self-Supervised Learning to learn meaningful representations from unlabeled medical images. The model first learns useful image features without requiring extensive manual annotations. SimCLR is included in the proposed technology stack for self-supervised representation learning.

3.5 Fine-Tuned Disease Prediction

The learned image representations are used for disease prediction through fine-tuning with a small labeled dataset. ResNet-18 is included as the model architecture in the proposed system. This approach aims to reduce dependency on large labeled datasets while supporting medical image classification.

3.6 Explainable AI Using Grad-CAM

Grad-CAM is used to generate visual explanations for the disease prediction results. It highlights important regions of the medical image that contribute to the model's prediction. This allows doctors to understand which image regions influenced the AI-generated result and improves model transparency.

3.7 Similar Case Retrieval Using FAISS

The framework integrates FAISS for retrieving similar medical cases. The extracted image features are used for similarity-based retrieval from available medical image representations. The retrieved cases provide additional reference information that can help doctors compare the uploaded image with similar cases.

3.8 Confidence Score Generation and Active Learning

The system generates confidence scores along with the disease prediction results to indicate the model's confidence in its output. Active Learning is incorporated to improve the model through expert feedback and selected data for further learning. This supports continuous improvement of the medical image analysis framework.

3.9 PDF Report Generation and Clinical Decision Support

Finally, the framework generates a downloadable PDF report containing the prediction results and related analysis information. The system also provides analytics and similar case information to support clinical decision-making. The generated information is intended to assist doctors rather than replace their professional judgment.

4 ALGORITHMIC FLOW OF THE SYSTEM

- The working of the proposed MedXAI framework is explained through the following steps:
- Initialize the MedXAI medical image intelligence framework.
- Upload the medical image into the system.
- Perform Image Quality Assessment to check the quality of the uploaded image.
- Verify whether the uploaded image belongs to the supported scan categories using OOD Detection.
- Extract meaningful image representations using Self-Supervised Learning.
- Fine-tune the learned representations using a small labeled dataset for disease prediction.
- Generate disease prediction results using the classification model.
- Apply Grad-CAM to generate visual explanations for the prediction.
- Retrieve similar medical cases using FAISS-based similarity search.
- Generate confidence scores for the prediction results.
- Use Active Learning and expert feedback to support model improvement.
- Generate a downloadable PDF report containing the analysis results.
- Present the prediction and related information for clinical decision support.

5 CONCLUSION

The proposed MedXAI framework presents an integrated approach to medical image intelligence by combining Self-Supervised Learning, Explainable AI, Similar Case Retrieval, Image Quality Assessment, OOD Detection, and Active Learning. The framework aims to reduce dependency on large labeled medical image datasets while improving the transparency and reliability of AI-based disease prediction.

By integrating Grad-CAM for visual explanations and FAISS for similar case retrieval, the system provides additional information that helps doctors understand and review prediction results. Image Quality Assessment and OOD Detection further support reliable image processing, while confidence scores and PDF reports provide useful information for clinical decision support.

The proposed work provides a foundation for developing intelligent and explainable medical image analysis systems. Future development and evaluation can further improve its performance, reliability, and practical applicability in healthcare environments.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] H. Y. Zhou, C. Lu, C. Chen, S. Yang, and Y. Yu, "A Unified Visual Information Preservation Framework for Self-supervised Pre-Training in Medical Image Analysis," 2023.

[2] H. Yu and Q. Dai, "Anatomy-Guided Self-Supervised Distillation Learning for Medical Image Analysis," 2026.

[3] H. Chu, F. Li, Y. He, and T. Guan, "Generative and Contrastive Based Self-Supervised Learning Model for Histopathology Image Analysis," 2023.

[4] Q. Sun, A. Akman, and B. W. Schuller, "Explainable Artificial Intelligence for Medical Applications: A Review," 2025.

[5] M. Mesinovic and T. Zhu, "Clinical XAI: A Tutorial on Concepts, Methods, and Modalities from Linear Models to LLMs," 2026.

[6] M. Ennab and H. Mcheick, "Enhancing Pneumonia Diagnosis Through AI Interpretability: Comparative Analysis of Pixel-Level Interpretability and Grad-CAM on X-ray Imaging With VGG19," 2025.

[7] D. Lotfi, M. A. Nikouei Mahani, M. Koohi-Moghadam, and K. T. Bae, "Safeguarding AI in Medical Imaging: Post-Hoc Out-of-Distribution Detection with Normalizing Flows," 2026.

[8] A. Ahmed and S. J. Malebary, "Query Expansion Based on Top-Ranked Images for Content-Based Medical Image Retrieval," 2020.

[9] L. Tang et al., "GAN-Guided Few-Shot Attention Network for Medical Images Fusion Quality Assessment," 2025.

[10] W. Liu et al., "Explainable AI for Medical Image Analysis in Medical Cyber-Physical Systems: Enhancing Transparency and Trustworthiness of IoMT," 2025.

[11] Python Software Foundation, "Python Programming Language," Available: https://www.python.org/

[12] PyTorch, "PyTorch Deep Learning Framework," Available: https://pytorch.org/