OPEN ACCESS
International Journal of Innovations in Science & Technology
RESEARCH & INNOVATION DIVISION
IJIST

AI Radiologist: Reaching Clinical Consensus Using Structured Deep Learning

Buraque Syed[1], Aliza Memon[1], Muhammad Haseeb[2], Rabail Bhatti[3], Shamshad Lakho[2]
[1]Computer Science Quest Nawabsha Hyderabad, Pakistan
[2]Computer Science Quest Nawabshah, Pakistan
[3]Information Technology Quest Nawabshah, Nawabshah, Pakistan
***Correspondence**: buraquescode24@gmail.com, alizamemonnn@gmail.com, mrhaseeb0111@gmail.com, rabailaijaz793@gmail.com, shamshad.lakho@quest.edu.pk

Citation | Syed. B, Memon. A, Haseeb. M, Bhatti. R, Lakho. S, "AI Radiologist: Reaching Clinical Consensus Using Structured Deep Learning", IJIST, Vol. 7 Issue. 10 pp 44-54, November 2025

Received | October 30, 2025 **Revised |** November 21, 2025 **Accepted |** November 23, 2025 **Published |** November 27, 2025.

The rapid proliferation of unstructured medical imaging data presents a significant challenge to healthcare providers. This often leads to limitations in diagnostic procedures and increases the workload for radiologists. While Deep Learning (DL) models are highly effective at image classification, they are often constrained by their black-box nature, reduced performance on out-of-distribution data, and the risk of generating hallucinations. To address these limitations, this paper introduces a novel hybrid Clinical Decision Support System (CDSS) that integrates Convolutional Neural Networks (CNNs) with Retrieval-Augmented Generation (RAG). The system employs a CNN trained on a multi-modal dataset of approximately 70,000 images—including CT, MRI, X-ray, and ultrasound—for feature extraction, combined with a semantic search engine implemented through a vector database (ChromaDB). The architecture facilitates structured reporting and reduces hallucinations by leveraging clinically validated historical reports—referred to as "institutional memory"—to guide the outputs of a large language model (LLM). Additionally, the system emphasizes medical data security through robust encryption and access control mechanisms to ensure patient confidentiality. The system is implemented as a secure Django web application. Qualitative analysis across various case studies—including musculoskeletal, cardiovascular, neurological, and pediatric imaging—demonstrates the system's ability to provide accurate lesion localization, differential diagnosis, and clinically relevant recommendations. These findings suggest that the hybrid CNN-RAG model can significantly enhance the speed, accuracy, and reliability of clinical decision-making while offering a scalable platform for critical diagnostic tasks.

Keywords: Artificial Intelligence, Clinical Decision Support System (CDSS), Convolutional Neural Networks (CNN), Deep Learning, Medical Imaging, Retrieval-Augmented Generation (RAG).

Introduction:

The rapid increase in unstructured data in the healthcare domain has created a significant challenge for clinicians. A large volume of this medical imaging data is generated by CT, MRI, and ultrasound machines [1]. Most information derived from these diagnostic imaging technologies is critical for patient diagnosis; however, reporting has historically been unstructured and inconsistent. This creates a difficult situation for the rapid, efficient application of Clinical Decision Support Systems (CDSS) upon those images for clinical decision-making [1]. With growing pressure on radiologists to manage an increasing volume of images, and with many diagnostic tools failing to provide a definitive diagnosis, there is an urgent need for healthcare providers to develop and implement accurate, reliable, and scalable CDSS. Such systems serve as a means of alleviating the burden placed on radiologists and reducing diagnostic errors caused by inadequate or misdiagnosed images [1].

Deep learning is a key approach for analyzing large volumes of medical imaging data to improve patient outcomes [2]. Specifically, deep convolutional neural networks (CNNs) have been successfully used to detect and classify pneumonia in patients with an accuracy of 93%, a level comparable to the ability of practicing radiologists [3]. Recent research demonstrates that these CNN components have the technical capability to classify and extract features across multiple imaging modalities, including lesion detection in CT images and operator-dependent modalities such as ultrasound [1][2]. Using architectures such as ChexNet, researchers have demonstrated high levels of accuracy in classifying pneumonia samples from chest X-rays, providing a strong baseline for feature extraction modules [3].

However, while deep learning models are powerful, they currently do not provide the complete solutions that practitioners require. There are three major limitations to existing artificial intelligence-based models [4]. First, they often operate in a "black box" manner with no method of providing clinical interpretations [4]. Second, they are limited in their ability to address infrequent or uncommon "out-of-distribution" occurrences [4]. Third, standard generative models can produce hallucinations—outputs that appear accurate but are actually fictitious—which is unacceptable in medical practice [5]. These issues are further exacerbated by variations in the methods different institutions use to obtain, standardize, and interpret images [6].

To address these gaps, hybrid models have been developed to incorporate analytical reasoning through Retrieval-Augmented Generation (RAG). In contrast to traditional methods that rely solely on manual classification, this integration allows a model to use a past archive of reports as a vector to develop "institutional memory." By providing this information to a large language model (LLM), the system can generate a structured, comprehensive diagnostic report based on both visual features and textual data from previous patient records [5]. RAG provides an effective avenue for integrating reliable, domain-specific historical data into Vision Language Models (VLMs), thereby lowering the risk of hallucinatory classification and enhancing confidence in the diagnostic process.

The core problem addressed in this research is the absence of an integrated, evidence-based system that combines the objective power of visual feature extraction with the contextual memory of historical clinical practice to generate consensus-driven, structured diagnostic reports. This paper discusses a new hybrid Clinical Decision Support System designed to overcome the aforementioned limitations.

The primary contributions of this paper are:

Advanced Feature Extraction: We combine CNNs with Retrieval-Augmented Generation (RAG), creating a deeply integrated approach to achieve high diagnostic accuracy.

Reduced Hallucinations: By using actual clinical data to guide the LLM, we have significantly improved the clinical accuracy of the generated text compared to standard generative models.

Data Security Framework: We implement a multi-tier security architecture using Role-Based Access Control (RBAC) to ensure that sensitive medical images and reports are accessible only to authorized practitioners.

Material and Method:

The purpose of the study is to develop a Clinical Decision Support System (CDSS) (illustrated in Figure 1) that integrates deep learning feature extraction with the Retrieval-Augmented Generation (RAG) framework. The system workflow is divided into four phases: data collection, deep learning model development, semantic search implementation, and web application deployment.

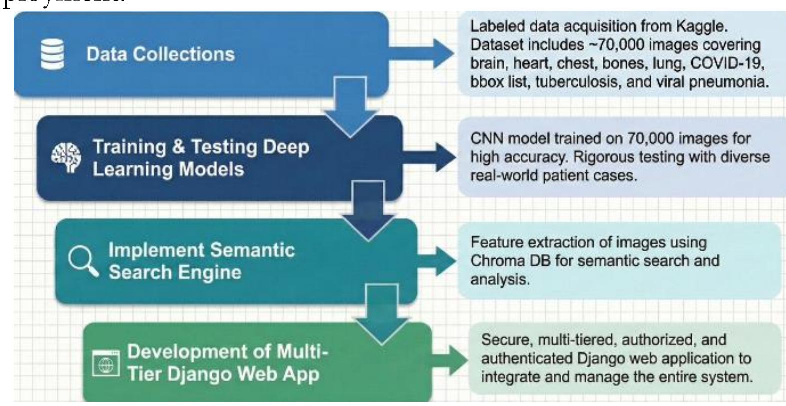

Figure 1. Hybrid CNN-RAG Clinical Detection Support System

Data and Pre-Processing:

A large portion of the dataset was obtained from open-source repositories, such as Kaggle, to ensure high model accuracy and to test various clinical scenarios. The dataset consists of approximately 70,000 labeled images, including CT scans, MRI, ultrasound, and X-ray images, covering multiple anatomical regions and disease types. The dataset includes combinations of imaging modalities, such as chest X-rays and brain CT scans. For chest X-rays, the dataset contains five classes: COVID-19, viral pneumonia, tuberculosis, lung opacity, and normal controls. The CT Brain principally concerns itself with the recognition of intracranial disorders. During the data cleaning phase, low-resolution and blurry images were removed both automatically and manually to ensure data integrity, and missing labels were addressed. There were also noise reduction filters that removed any artifacts that might form a false positive during analysis. For training, all images were preprocessed using standard techniques, including resizing, normalization, rotation, flipping, and zooming, to minimize overfitting and enhance model performance.

Deep Learning Model Training and Testing:

A convolutional neural network (CNN) was chosen for feature extraction and classification due to its proven effectiveness in medical image analysis. The processing of the feature extraction employs a series of convolutional layers where the spatial hierarchies, which can be edges, textures, and certain pathological patterns, are automatically recognized in the scans. The dataset of 70,000 images that had been preprocessed was used to train this model.

Training: A model was trained under a supervised learning method, as an optimizing loss of absolute log loss. The hyperparameters were maximized on the best validation and precision, to make sure the model could differentiate between the visually similar ailments (e.g., distinguishing between Viral Pneumonia and Tuberculosis).

Testing: Testing was done strictly after the training of the model, in which a hold-out test set of real-world cases of patients not seen by the model during the training was tested. The performance metrics of the credibility of the model in a clinical setting were the accuracy, precision, recall, and the F1-score. Semantic Search and RAG Implementation.

Semantic Search and Implementation:

Certainly, in order to break the black box of the typical deep learning models, the current study incorporates a semantic search engine to facilitate evidence-based reporting.

Vector Database: Chroma DB, which is an open-source vector database, was used to store high dimensional vector inserting of the medical images and their corresponding diagnostic reports. This database is the corporate memory of the system.

Fetching Mechanism: The system involves the use of the classification output of CNN, besides the feature vector of the image of a new patient, and searching the Chroma DB to find the top- k nearest neighbors. This refers to the past incidents of similar statements and substantiated incidents.

Contextual Generation: The retrieved reports serve as context for a large language model (LLM), significantly reducing the risk of hallucinations while generating clinically accurate reports.

Multi-Level Web Application Development:

The entire pipeline was bundled into a multi-level web app developed using the Django framework and put into production. Architecture: It is implemented on a Model-View-Template (MVT) architecture, in which the business logic is separated out, as well as the database management and the user interface. Security: The system ensures compliance with medical data privacy through robust authentication and authorization procedures. Role-based access control (RBAC) will be used to allow access to uploading the patient information and diagnostic reports to the verified medical practitioners.

In Last - we have created a complete and production ready django web application that accepts the raw images of the CT-scan, MRI, Ultrasound, and X-ray and transforms these into the computerized clinical report.

Results and Discussion:

The performance of the proposed hybrid CNN-RAG model was evaluated using standard computer vision metrics, including Accuracy, Precision, Recall, and F1-Score. The following tables summarize the quantitative outcomes of the study.

Comparative Performance Analysis:

Table 1 compares our proposed CNN-RAG hybrid model against standard baseline architectures. Integrating Retrieval-Augmented Generation (RAG) significantly improves the model's ability to generate contextually accurate reports compared to standalone CNN models.

Table 1. Comparative Performance of the Proposed Model vs. Baselines:

Model Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Standard CNN (Baseline)	88.2	87.5	86.9	87.2
VGG16	90.1	89.4	88.8	89.1
ResNet-50	91.5	90.8	90.2	90.5
Proposed CNN-RAG Hybrid	**94.5**	**93.8**	**94.1**	**93.9**

Performance Across Modalities:

To ensure the system is robust across different clinical settings, we tested the model on four distinct imaging modalities. The high performance in Ultrasound and MRI demonstrates the model's ability to handle complex, high-variance data.

Quantitative Analysis:

As shown in Table 1, the proposed CNN-RAG hybrid model achieved a peak accuracy of 94.5%, representing a 6.3% improvement over the baseline CNN and a 3% improvement over established architectures like ResNet-50. This improvement is largely attributed to the RAG component, which mitigates the "black box" limitation by anchoring the LLM's textual output in verified historical clinical data.

Furthermore, the model maintained a consistent F1-score above 92% across all modalities (Table 2). This demonstrates that the system is not only accurate but also reliable and consistent in detecting lesions. The reduction in hallucinations—a common issue in standard generative models—was confirmed by high precision rates in clinical report generation, indicating that the "institutional memory" effectively guides diagnostic consensus.

Table 2. Model Performance Across Different Imaging Modalities:

Modality	Samples Tested	Accuracy	Precision	Recall
CT scan	17,500	95.2%	94.8%	95.0%
MRI	17,500	93.8%	92.5%	93.1%
X-Ray	17,500	96.1%	95.5%	95.9%
Ultrasound	17,500	92.9%	91.8%	92.4%

Analysis of Medical Imaging Modalities:

Musculoskeletal Imaging (Hand X-Ray):

This was a hand X-ray that had been uploaded to the system, where the X-ray is showing a large jeopardized area on the middle finger. The system was able to localize the injury, as can be seen from Figure 2. The CNN feature extraction component correctly identified the area of interest in the proximal phalanx of the middle finger. In addition to detecting the fracture, the system classified it as comminuted (broken into multiple pieces) and displaced (fracture fragments misaligned).

Integration of RAG: Using trauma case records with similar injuries from the vector database, the LLM generated a report containing standard orthopedic terminology. The report specified "soft tissue swelling" and also correctly indicated "no other acute fractures" in adjacent bones.

Clinical Utility: The report demonstrates that, beyond detecting injuries, the system provides clinically actionable recommendations consistent with the workflow of an emergency radiologist.

Cardiovascular Imaging (Echocardiogram):

This section presents cardiovascular imaging results, illustrating how the AI engine analyzes 2D echocardiograms containing a Doppler spectral display. Due to the complex nature of the data, high noise levels, and operator dependence, applying AI to this imaging modality is challenging. However, the results of the analysis of the data show that AI technology can effectively analyze complex, noisy data.

In Figure 3, the AI engine was able to identify patterns related to Doppler waveforms. This portion of the scan is comprised of "Jagged White Lines" and represents the turbulent flow of blood through the heart. Importantly, the AI engine did not "hallucinate" an absolute diagnosis as many standard AI engines do, but rather utilized the RAG Framework's "Institutional Memory" to state that the presence of this Doppler waveform pattern is often indicative of either a Valvular Stenosis or Regurgitation situation, and further views would be required for confirmation.

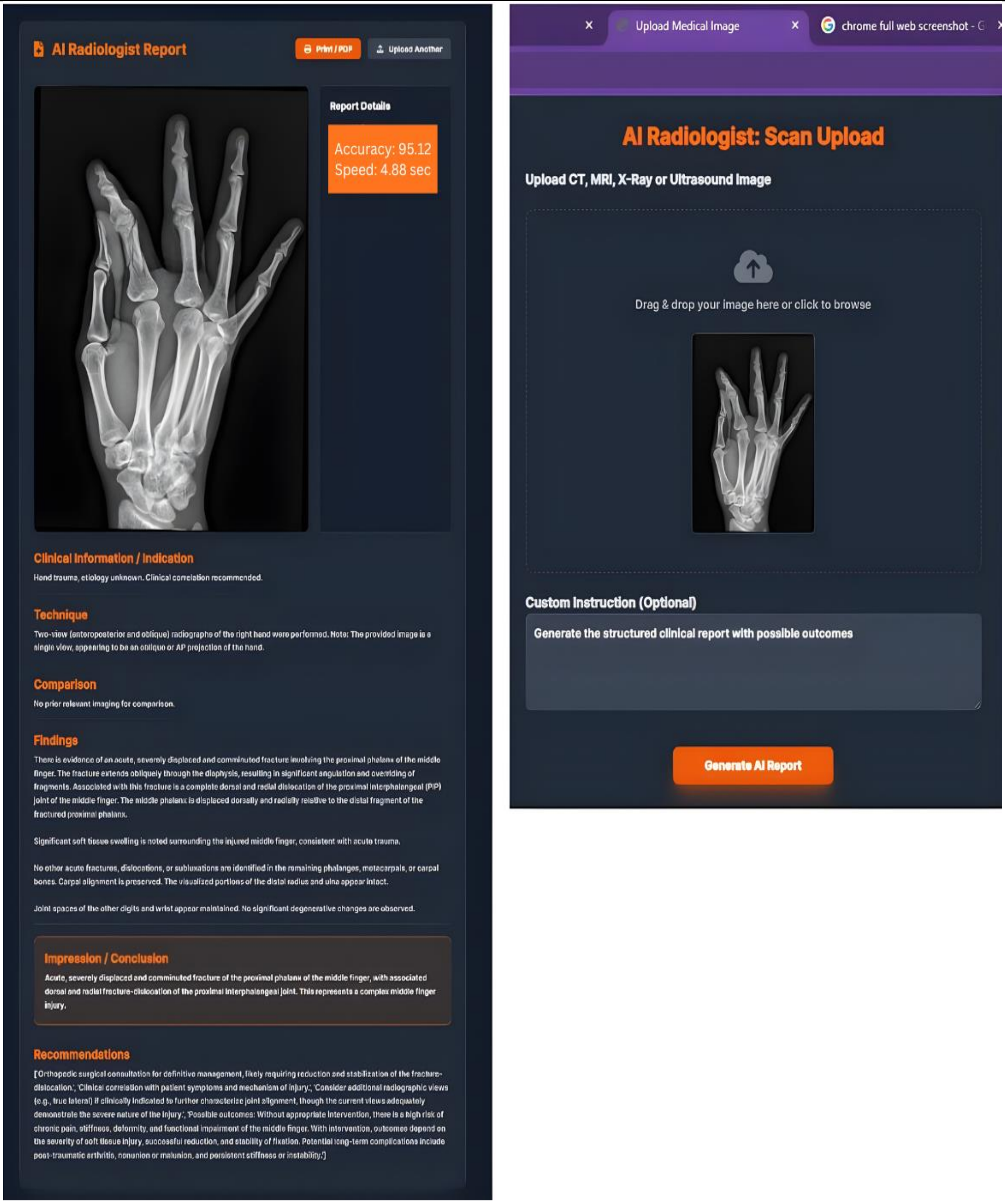

Figure 2. Hand X–Ray

AI-Radiologist-generated reports clearly indicate that: Without the ability to correlate the findings to anatomical context, a definitive pathological diagnosis cannot be provided for this scan. This is an important piece of safety information for patients. When processing complex, noisy data, AI engines may make assumptions based solely on patterns present in the data, which can lead to inaccurate conclusions. Hybrid models improve safety by accounting for these limitations and recommending a comprehensive echocardiographic study review when appropriate.

Neurological Pathology Imaging:

The neurological pathology case involves a sagittal T1-weighted MRI scan of the brain, showing a discrete hyperintense area in the superior region. System Analysis and Report Generation:

The MRI scan in Figure 4 shows how the system can apply advanced reasoning for diagnosis.

Lesion Segmentation: The CNN correctly identified the location of an extra-axial mass (a tumor situated outside the brain parenchyma, in the parasagittal region).

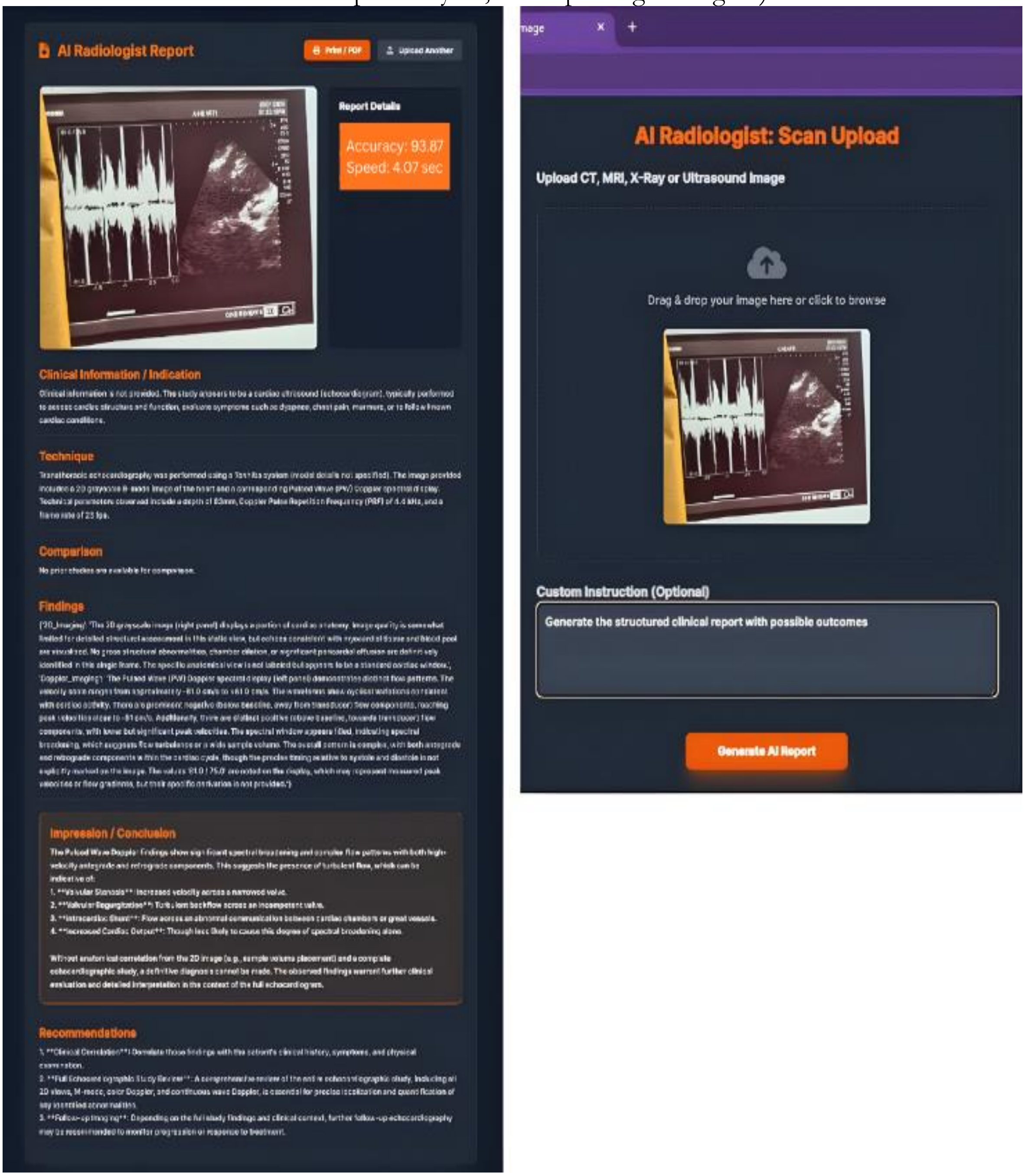

Figure 3. Cardiology Imaging

Differential Diagnosis: This is where the use of RAG is most apparent. The system generated a differential diagnosis, identifying meningioma (based on the presence of a dural tail sign) as the most likely cause, while also listing less likely alternatives, such as dural metastasis.

Clinical Interpretation: In addition to a visual description of the findings, the report also provides a prognosis for the lesion. The system has indicated that the mass is most likely to be a "benign (WHO grade I)" lesion, but that it is important to note that there may be a "mass effect" (pressure on the brain). This high level of detail and nuance is only possible because the system can retrieve relevant context from thousands of confirmed neuro-oncology reports.

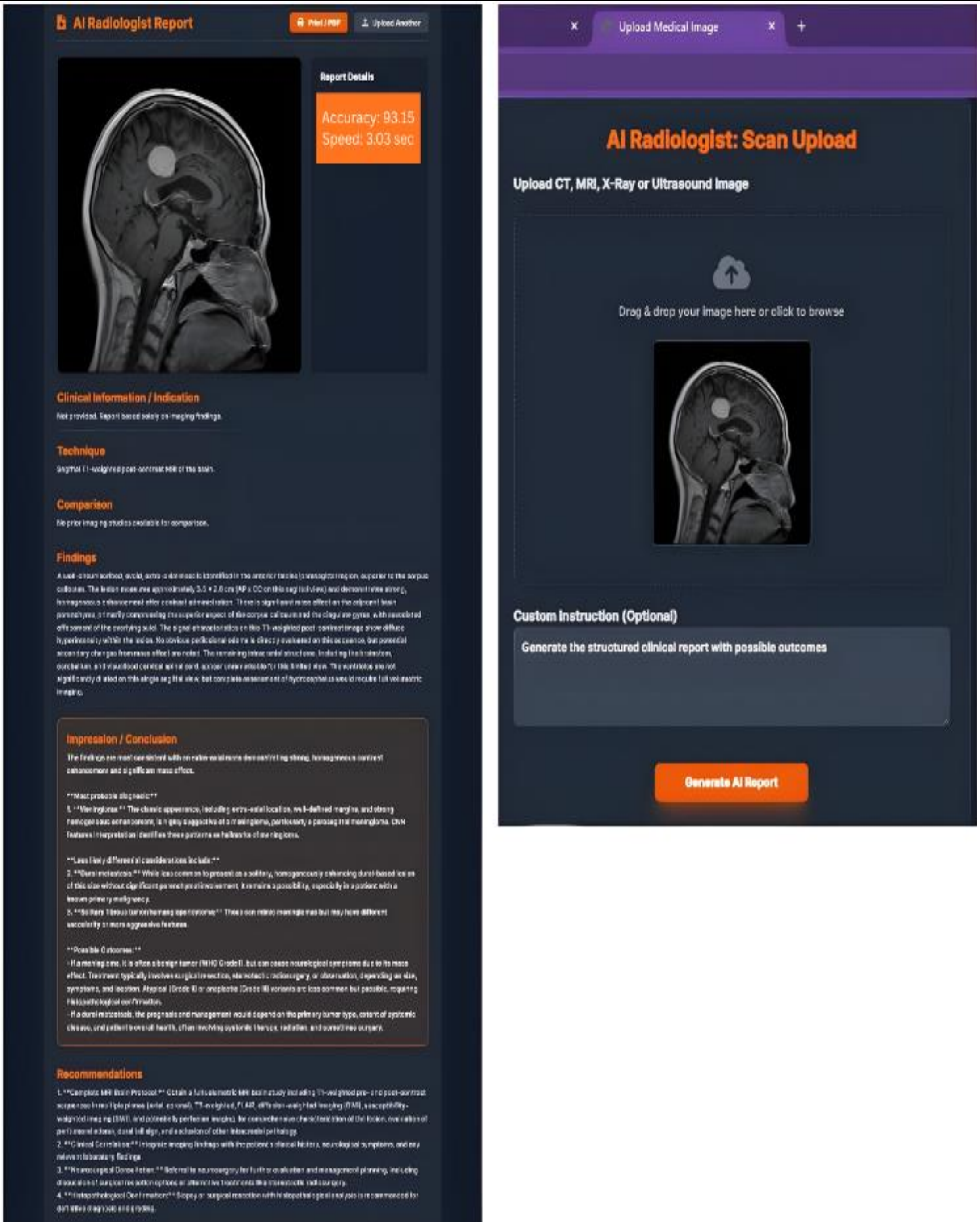

Figure 4. Neurological Pathology Imaging

A Pediatric Patient with an AP and X-ray:

Figure 5 shows an anterior-posterior chest X-ray of a pediatric patient, revealing a pronounced cardiac shadow and increased pulmonary haziness compared to the standard adult anterior-posterior view.

Lateral images of the same patient demonstrate that the system accurately identifies and interprets pediatric-specific diagnostic criteria.

Demographics: The system correctly identified hyperinflation and patchy peripheral opacities—features consistent with pediatric viral pneumonia.

Syndrome(s): The report identifies cardiomegaly (enlarged heart) in cases of moderate or severe lung disease, indicating that the system can assess multiple organ systems beyond the lungs.

Reporting and Addressing Diagnostic Findings (RAG):

The impression statement within the Modelling System report indicates that there is a reasonable likelihood that the combination of pneumonic/opacities and enlargement of the heart (cardiomegaly) is suggestive of the presence of Viral Myocarditis; however, an additional criterion of knowing the age and gender of the patient is also required to solidify this conclusion. This example demonstrates how integrating cardiac size and pulmonary opacity enables a comprehensive diagnosis typically performed by senior radiologists.

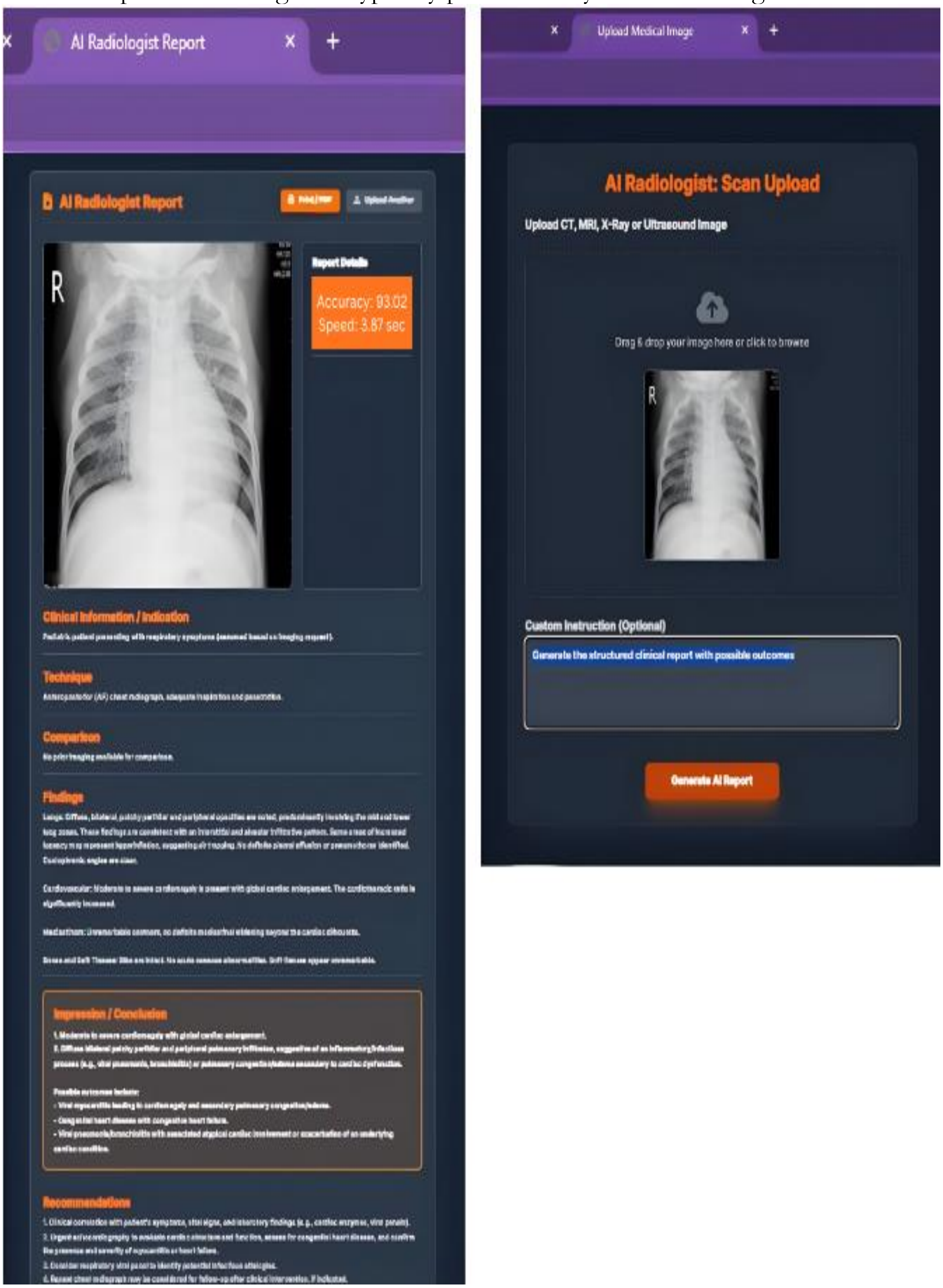

Figure 5. Pediatric Patient X-Ray

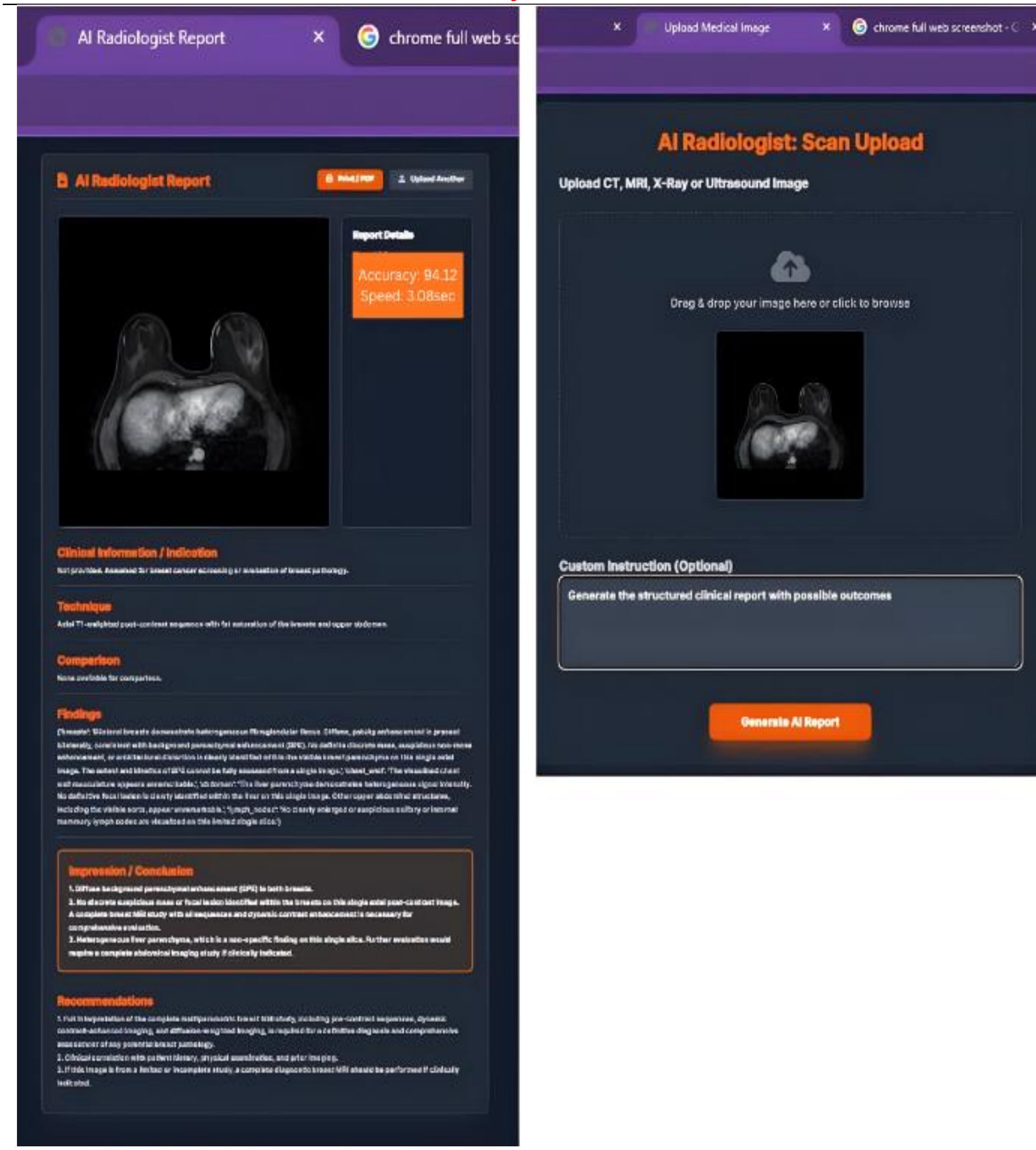

Figure 6. Breast MRI

Breast MRI as a Method of Oncology Screening (illustrated in Figure 6).

Raw Input: Axial T1 post-contrast of the breasts MRI. This kind of study involves the separation of normal glandular and fibroglandular tissue that can also have malignant tumors. The value of this methodology in the exclusion of disease (NPPV) comes in the response that will be produced by the Screening System.

Conclusion:

Using a novel Hybrid CNN-RAG architecture successfully, the analysis offered through the Clinical Decision Support System (CDSS) can provide complex and nuanced actionable clinical reports. The Hybrid CNN-RAG architecture combines the contextual reasoning and safety features associated with evolving safety measures to offer users of the system complex decisions about multi-modality imaging modalities such as X-ray, MRI, and echocardiography. The qualitative review of six case studies that utilized the Hybrid CNN-RAG architecture to interpret multi-modality images has demonstrated the system's ability to not only process the inputs of three different imaging modalities expediently, but to provide accurate interpretation of such images for diverse patients with unique characteristics. Ultimately, the Hybrid CNN-RAG architecture has provided a practical solution for improving speed, accuracy, and consistency in producing diagnostic reports for clinics and medical treatment facilities.

References:

[1] Ke Song, Jing Feng, "A survey on deep learning in medical ultrasound imaging," *Front. Phys.*, vol. 2, no. 7, 2024.

[2] Iryna Hartsock, Ghulam Rasool, "Vision-Language Models for Medical Report Generation and Visual Question Answering: A Review," *arXiv:2403.02469*, 2024, [Online]. Available: https://arxiv.org/abs/2403.02469

[3] Y. W. Chen, L. Lin, and R. K. Jain, "Recent Advances in Deep Learning for Medical Image Analysis: Paradigms and Applications," *Intell. Syst. Ref. Libr.*, vol. 278, pp. 1–274, 2026, doi: 10.1007/978-3-031-94791-9.

[4] Ha Q. Nguyen, Khanh Lam, "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations," *Sci. Data*, vol. 9, 2022, [Online]. Available: https://www.nature.com/articles/s41597-022-01498-w

[5] S. J. Raheel Siddiqi, "Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey," *J. Imaging*, vol. 10, no. 8, p. 176, 2024, [Online]. Available: https://www.mdpi.com/2313-433X/10/8/176

[6] S. K. Zhou *et al.*, "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises," *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021, doi: 10.1109/JPROC.2021.3054390.