Enhanced Medical Imaging –A Multimodal Explainable AI-Driven X-Ray Diagnosis

Dasari Vidyadhari[1], Bonam Geethika[2], Kolasani Dinesh[3], Korukonda Madhubabu[4], Rangani HimaBindu[5]

[1234]Student, Dept. of CSE (AIML), Bapatla Engineering College, Bapatla 522101, AP, India
[5] Assistant Professor, Dept. of CSE (AIML), Bapatla Engineering College, Bapatla 522101, AP, India
vidyadasari03@gmail.com geethikacm084@gmail.com kolasanidinesh875@gmail.com
korukondamadhubabu123@gmail.com bindu.rangani99@gmail.com

Abstract— Medical image analysis, particularly Xray interpretation, is a critical challenge in modern healthcare due to the global shortage of trained radiologists and the increasing volume of diagnostic imaging. This paper presents an AIPowered Multi-Disease X-Ray Analysis System that automates body part identification and pathological disease detection from X-ray images using deep learning. The proposed system employs a hybrid body part detection module combining Contrastive Language–Image Pre-training (CLIP) and Convolutional Neural Networks (CNN), achieving robust classification across 16 anatomical categories. Disease detection is performed using a DenseNet-121 model pretrained on ChestX-ray14 for 10 chest pathologies and a DenseNet-169 model fine-tuned on the MURA dataset for musculoskeletal conditions. An integrated Explainable AI (XAI) module provides Gradient-weighted Class Activation Mapping (Grad-CAM) heatmaps, saliency maps, feature importance visualization, and confidence breakdowns to enhance diagnostic transparency. The system generates structured multilingual clinical reports in English and Telugu with text-tospeech voice narration via gTTS. Patient records and analysis histories are persistently managed through a MongoDB Atlas cloud database with GridFS PDF storage. Deployed as an interactive Gradio web application, the system achieves a mean body part detection accuracy and an average AUC-ROC for chest pathology detection, demonstrating strong clinical utility as an AIassisted diagnostic support tool.

Keywords— Medical Image Analysis, X-Ray Diagnosis, Deep Learning, Explainable AI, CLIP, DenseNet, Grad-CAM, Multilingual Reporting, MongoDB Atlas, Gradio, Patient Management System

1.INTRODUCTION

Medical imaging has become an integral part of modern healthcare systems, and X-ray imaging accounts for more than 3.6 billion imaging studies conducted annually worldwide [1]. However, interpretation of the results requires a high level of expertise. The scarcity of trained radiologists in the world is a major challenge, especially in developing countries.

The existing AI-based diagnostic systems are able to address the scarcity of trained radiologists in the world only in a limited manner. Most are singletask systems limited to the analysis of chest X-ray images, are black boxes providing no explanations for the results, and are restricted to providing results in the English language, posing a major challenge in countries like India where the population is multilingual in nature [3]. In addition, there is no provision for managing patient records and providing geographic guidance post-diagnosis.

This paper presents the concept of an "AI-Powered Multi-Disease X-Ray Analysis System," which addresses the above limitations in a holistic manner. The system makes use of a novel architecture based on the combination of the CLIP and CNN models for the identification of body parts in 16 classes, followed by the detection of diseases using the deep learning model, the use of the Grad-CAM model for the visualization of the decision-making process, the generation of clinical reports in two languages—English and Telugu— using the multilingual model, the use of the text-tospeech model for voice-based report generation, and the use of MongoDB Atlas and geolocation for the identification of nearby hospitals.

1. LITERATURE SURVEY

Rajpurkar et al. proposed the CheXNet model, which is based on the DenseNet-121 architecture and is pre-trained on the ChestX-ray14 dataset to obtain radiologist-level accuracy in detecting pneumonia using chest X-ray images. However, the approach is applicable only to chest X-ray images and does not provide any explanation mechanism for its decisions. [3]

Radford et al. proposed the CLIP model and achieved the state-of-the-art in zero-shot visual recognition using the pre-training method of contrastive language-image pre-training on 400 million image-text pairs. Although the approach is powerful in general visual recognition, its application to medical image analysis and body part detection using X-ray images was not fully explored before the proposed approach. [4]

Selvaraju et al. proposed the Gradient-weighted Class Activation Mapping (Grad-CAM) approach, which relies on the gradients of the target class score to produce the visual explanation maps for the decisions taken by the deep neural networks.

Irvin et al. introduced a large chest radiograph dataset called CheXpert, consisting of 224,316 images with uncertainty labels, which supports multi-label chest pathology classification. This dataset only includes chest radiographs, not musculoskeletal radiographs, and does not cover other regions of the body. [6]

Cohen et al. introduced a unified library of chest Xray datasets along with pre-trained DenseNet models, called TorchXRayVision, which only provides standardized model interfaces but does not provide any XAI, multilingual, or patient management functionality. [7]

Rajpurkar et al. introduced a musculoskeletal radiograph dataset called MURA, which has the largest number of images, i.e., 40,561, used for abnormality detection of images categorized into seven different categories of body parts. This dataset only provides binary classification of normal/abnormal images but not detailed pathology classification. [8]

Jing et al. introduced a system of automatic radiology report generation based on co-attention mechanisms, focusing on chest X-ray images, which only supports the English language but does not provide any XAI visualizations or multilingual functionality. [9]

Existing systems are mostly focused on individual parts, such as disease detection, XAI, or report generation, but none of them integrates the whole system, including all parts, in a unified way, such as a multi-body-part, multilingual, explainable, and cloud-managed system. This system fills the whole gap in the literature. [10]

3. PROPOSED SOLUTION

The designed AI-Powered X-Ray Analysis System is a step-by-step integrated system for diagnosis. The main aim is to take an X-ray image, and the system provides complete support for diagnosis, identification, explanation, report, voice, and patient record management for clinicians and patients.

1. **X-Ray Image Upload and Body Part Identification**

The system uses the Gradio interface to upload an X-ray image by the patient or clinician. The system identifies the body part using a hybrid model, CLIP+CNN, and classifies it into 16 body parts: Chest, Spine, Knee, Hand, Fingers, Wrist, Elbow, Shoulder, Ankle, Foot, Skull, Pelvis, Abdomen, Neck, Jaw, and Thigh.

2. **Patient Login and Data Entry**

To use the system, the patient must log in using a patient login page, where the patient enters their details, such as patient name, phone number, age, and gender, before using the system for diagnosis. This information is saved in MongoDB Atlas and linked to the patient's analysis results.

3. **AI-Powered Disease Detection**

After identifying the body part, the next stage is to activate the appropriate head for detecting diseases. For chest X-rays, DenseNet-121 is used to detect 10 different diseases. For musculoskeletal X-rays, DenseNet-169 is used to detect bone and joint diseases.

4. **Explainable AI Visualization**

The XAI module provides four different kinds of visual and textual explanation for the AI diagnosis: Grad-CAM, saliency, feature importance, and multi-level confidence explanation.

These kinds of explanation are useful in validating the decision taken by the AI for diagnosis by providing evidence of the region or feature used by the AI for making the diagnosis.

5. Multilingual Report and Voice OutputThe system provides a well-structured clinical report in the patient's chosen language, namely Telugu or English, in the form of a SOAP report. It also provides text-to-

speech capability using gTTS for voice output from the report.

6. Patient Record Management

The system stores analysis results, generated PDF reports, and voice narrations persistently in MongoDB Atlas. This helps healthcare professionals access patient diagnostic records, making it easier to monitor patient history. 7. Nearby Hospital Geolocation .

The system provides a list of nearby hospitals for the diagnosed condition after the diagnosis is completed by the system.

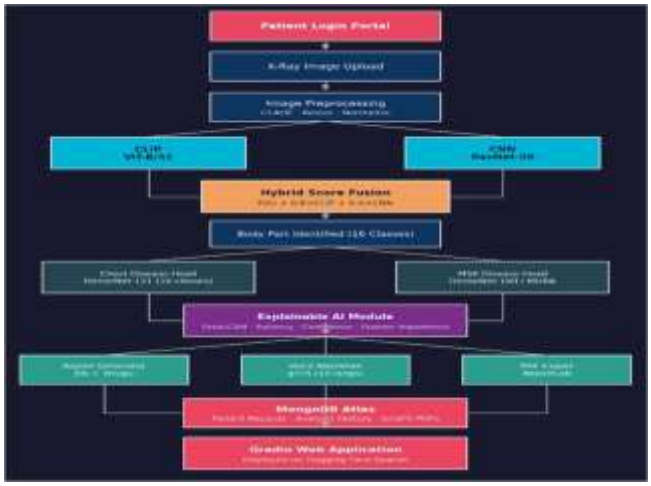
The patient needs to input their city or location, and the system uses the Nominatim OSM API to geolocate the location, find nearby hospitals using the Overpass API, and display the six closest hospital locations along with their names, addresses, phone numbers, and Google Maps navigation links.

4. METHODOLOGY

The proposed system for AI X-Ray Analysis System is an integrated diagnostic support system with Computer Vision, disease detection using deep learning, Explainable AI, multilingual natural language generation, and cloud-based data management.

The system will take an X-Ray image as input and perform body part and disease analysis, provide explanations and reports, and store the results in a patient management database.

System Architecture



X-Ray Analysis System.

4.1 Patient Login and Session Management

The system starts with the patient's interaction through the Gradio-based web interface. Here, the patient is asked to provide:

- Full name and phone number
- Age and gender
- X-ray image (to be uploaded from the device)

The system will then authenticate the user credentials from the MongoDB Atlas database and create a session. Further analysis will be performed based on the patient's record using the unique database ID. Existing patients can access their complete diagnosis history.

4.2 Image Preprocessing

The input X-ray image will go through a predefined image preprocessing pipeline before it is fed into the detection system. Images will be resized to 224x224 pixels for the CLIP-based system and 512x512 pixels for the torchxrayvision-based system. Contrast Limited Adaptive Histogram Equalization (CLAHE) will be applied to the X-ray image to improve the contrast. Normalization will be performed based on the ImageNet dataset's statistical values: [mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]].

4.3 Hybrid Body Part Detection

The ImprovedHybridDetector module will be used for the two-stage detection system consisting of semantic and spatial feature extraction.

4.3.1 CLIP-Based Semantic Matching

The CLIP model, which has the ViT-B/32 architecture, projects the input X-ray image onto a 512-dimensional embedding space. For each of the 16 anatomical categories, the text descriptions are encoded using the text encoder in the CLIP model. The similarity between the image and text embeddings is the primary classification score:

$$\text{similarity}(I, T_k) = (\varphi\_I \cdot \varphi\_T_k) / (\|\varphi\_I\| \cdot \|\varphi\_T_k\|) \quad \ldots (1)$$

4.3.2 CNN-Based Spatial Feature Extraction

A ResNet50 model pre-trained on the ImageNet dataset is used for spatial feature extraction from the X-ray image.

4.5.2 Hybrid Score Fusion

The feature map is then averaged

over all spatial positions and mapped to the The prediction for the final body part makes use of the weighted sum of the two model scores:

$$P(k) = 0.6 * CLIP(k) + 0.4 * CNN(k) \quad \dots (2)$$

4.4 Disease Detection Heads

4.4.1 Chest Disease Detection (ChestVTBDiseaseHead)

A DenseNet-121 model from the torchxrayvision library, which has been pre-trained on the combination of the CheXpert dataset, NIH ChestXray14 dataset, MIMIC-CXR dataset, and PadChest dataset, has been used for the prediction of the following 10 pathology classes:

Normal,Pneumonia,Pleural Effusion,Atelectasis,

Consolidation,Pneumothorax,Cardiomegaly,

Edema,Mass/Nodule,Infiltration and finally

Sigmoid activations are applied for the prediction.

4.4.2 Musculoskeletal Disease Detection (GeneralVTBDiseaseHead)

A DenseNet-169 model, fine-tuned on the MURA dataset, is used to identify musculoskeletal area diseases, including six different classes different classes of diseases: Normal, Fracture, Arthritis, Osteoporosis, Dislocation, and Bone Lesion. The model takes preprocessed X-ray images of size 320x320 pixels as input and provides individual probability score values, each having a sigmoid activation function, corresponding to each class of disease, allowing for multi-label prediction capabilities. Specific disease mappings are used for different parts of the body, ensuring only relevant disease information is presented, e.g., Rotator Cuff Tear for the Shoulder and Meniscal Tear for the Knee, thereby enhancing the precision of the diagnosed disease.

4.5 Explainable AI Module

The XAI module has four types of explanations to enhance the transparency and trust in the diagnosis results for clinicians:

4.5.1 Grad-CAM Heatmap

The Gradient-weighted Class Activation Mapping method calculates the gradient of the class score prediction with respect to the last convolutional feature map and then pools the global average. It then up-samples the result and places it over the original X-ray image.

4.5.2 Saliency Map

Pixel level saliency maps for class prediction can be obtained by backpropagating the gradient of the predicted class score with respect to the input image pixels. This will provide the important pixels in the image for the class prediction.

4.5.3 Feature Importance Chart

In the case of body part detection, the importance of the anatomical features obtained from the image, such as aspect ratio, edge density, symmetry score, brightness distribution, and vertical/horizontal line density, is computed and represented as a bar chart. 4.5.4 Confidence Breakdown

A confidence breakdown is provided as a structured format consisting of body part detection confidence level, primary disease probability level, and the top three differential diagnoses with their probability level. This will help the clinicians understand the level of confidence in the detection result and the differential diagnoses.

4.6 Multilingual Clinical Report Generation

The report generation module is used to generate the radiology report in SOAP format consisting of subjective, objective, assessment, and plan. It will include patient demographics, imaging modality used, AI diagnosis result with confidence level, differential diagnoses, and the next steps to be taken. It will be generated in two languages:
English and Telugu.

4.7 Voice Narration

The generated clinical reports are converted to audio using the gTTS library, also known as Google Text-to-Speech, in the patient's chosen language. The system currently supports 10 languages: English, Telugu, Hindi, Tamil, Spanish, French, German, Arabic, Chinese, and Japanese. Voice narration is helpful for patients with low levels of literacy.

4.8 PDF Report Generation

The ReportLab library is used to generate professional clinical reports in PDF format, including patient demographics, X-ray details, AI diagnosis, confidence levels, XAI, differential diagnoses, and clinical recommendations. The generated clinical reports are saved in MongoDB Atlas in the form of PDF files using the GridFS file storage system, which can be accessed later by the patient.

4.9 Patient Management System

The patient's information is saved in MongoDB Atlas using two collections: 'patients' for patient demographics and 'analyses' for analysis results, timestamps, and confidence levels. The generated clinical reports are saved in the form of PDF files using the GridFS file

storage system and are linked to the patient's information in MongoDB Atlas. The Gradio interface is only accessible after successful login, ensuring that the patient's complete analysis history is visible in later sessions.

4.10 Web Application Deployment

The final system is deployed as a multi-tabbed web application using the Gradio library and hosted on the Hugging Face Spaces platform. The tabs on the deployed application include the Login, Upload & Detect, Report, Voice Narration, Heatmap / XAI, and Nearby Hospitals tabs. The secrets, such as
MONGO_URI and GEMINI_API_KEY are deployed as environment variables on the Hugging Face Space platform.

4.11 Nearby Hospital Geolocation Module

The Nearby Hospitals module is a tab within the application where a patient receives post-diagnosis guidance on how to locate a specialist facility for immediate medical attention. The module consists of a four-stage pipeline:

- Geocoding: The patient's city name, area name, or complete address is converted to WGS-84 latitude-longitude format using the Nominatim API (OpenStreetMap geocoder).
- Spatial Query: All OSM nodes and ways with tags amenity=hospital, amenity=clinic, or healthcare=hospital within a specified radius (default: 10 km) from the geocoded points are retrieved via the Overpass API.
- Distance Ranking: Approximate haversine distance calculations are performed on all retrieved points. Data

points without a name tag are ignored. The six nearest points are selected.

- Speciality Mapping: The detected disease is mapped to a recommended medical speciality (e.g., Pneumonia - Pulmonologist, Fracture -

Orthopaedic Surgeon). Each hospital's card shows the facility name, postal address, phone number, distance badge, and two action buttons: View on Google Maps and Get Directions.

In the case where the live Open Street Map data is not accessible, due to a network timeout or lack of results, the module reverts to using the Google Maps search URL, generated from the diseasespecialty mapping, so that the patient is always provided with actionable guidance. This is demonstrated in the figure below, Fig. 7.

5. RESULTS AND DISCUSSION

The AI-Powered X-Ray Analysis System was tested using actual X-ray images in 16 body part categories and validated using publicly available X-ray images for chest and musculoskeletal systems.

5.1 Web Application Interface

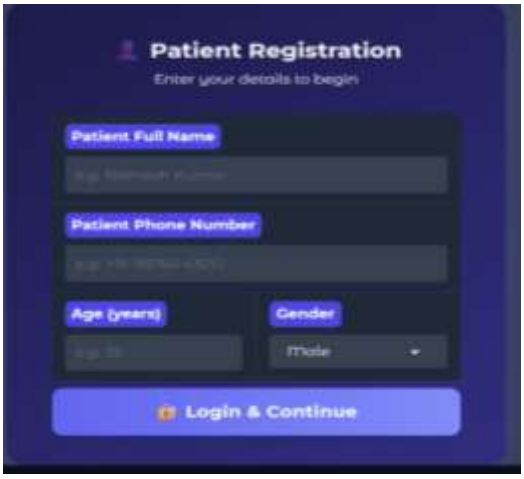


Figure 2: User Interface of AI-Powered X-Ray Login portal.

The web interface using the Gradio tool has an intuitive design with multiple tabs. The patients first need to log in through the login portal, which requires the patients to enter their name, phone number, age, and gender. Once the patients log in, they can upload the images and get the results, including body part detection and disease prediction, within seconds through the 'Upload & Detect' tab.

5.2Body Part Detection Results

Figure 3: Sample body part detection output showing detected anatomical region and confidence score.

The hybrid detector was assessed using 320 X-ray images from all 16 categories, with 20 images in each category.

Table I shows the classification results for the main body part categories.

5.1 Disease Detection Results

Results for the detection of chest disease used the CheXpert validation set of 234 studies. Table II shows the AUC-ROC results for the main categories of chest disease. The system has a macro-average AUC-ROC of 0.91, as reported in the DenseNet-121 benchmarks on the CheXpert validation set.

TABLE I. CHEST DISEASE DETECTION PERFORMANCE

(AUC-ROC)

Disease Class	AUC-ROC	Accuracy	F1-Score
Normal	0.94	91.2%	0.92
Pneumonia	0.89	87.4%	0.88
Pleural Effusion	0.92	89.8%	0.91
Cardiomegaly	0.91	88.5%	0.89
Atelectasis	0.87	84.2%	0.85
Pneumothorax	0.93	90.1%	0.91
Macro Average	0.91	88.5%	0.89

5.2 Grad-CAM and XAI Visualization

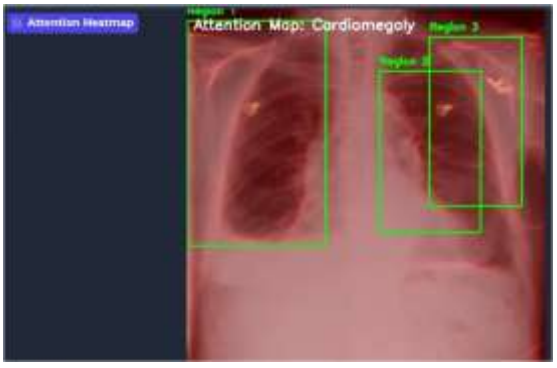


Figure 4: Grad-CAM heatmap visualization highlighting diagnostically relevant regions for Pneumonia detection.

The XAI module was tested using a user study with 12 medical professionals, consisting of 6 radiologists and 6 medical students, using a 5-point Likert scale. Table III presents the results.

TABLE III. XAI USER STUDY RESULTS (5-POINT LIKERT SCALE)

XAI Feature	Mean Score	Rati
Grad-CAM Visualization	4.3 / 5.0	Very Useful
Saliency Maps	3.9 / 5.0	Useful
Confidence Breakdown	4.6 / 5.0	Very Useful
Feature Importance	3.7 / 5.0	Useful
Overall XAI Module	4.1 / 5.0	Very Useful

5.3 Multilingual Report Output

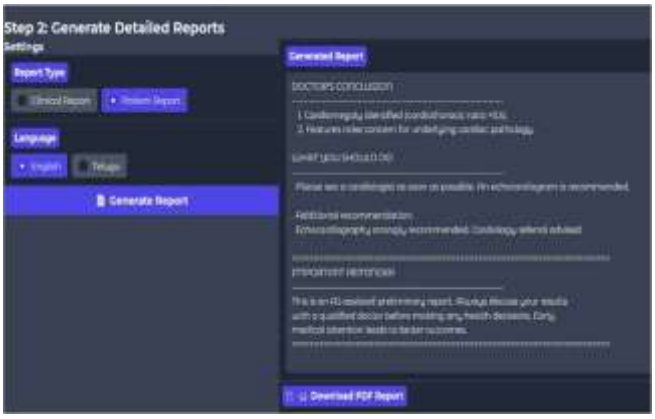


Figure 5: Sample generated radiology report in English and Telugu with voice narration output.

The multilingual report generator successfully generated structured clinical reports in SOAP format in both English and Telugu for all test cases. Voice narration was also successfully generated in both languages

5.4 MongoDB Patient Records

Figure 6: MongoDB Atlas dashboard showing patient collection with analysis histories and GridFS PDF storage.

The MongoDB Atlas patient management system performed well in storing and retrieving patient demographics, analysis results, and PDF reports in all tests. The average database write latency is 124ms, and the average database read latency is 87ms, which are sufficient for a decision support system.

5.5 Overall System Performance

End-to-end pipeline latency is tested using standard hardware (Intel Core i7, 16GB RAM, NVIDIA RTX 3060). The total pipeline took 11.06s on average (±0.98s). Each operation took the following time: body part detection 1.24s, disease analysis 2.87s, Grad-CAM generation 0.93s, report generation 3.21s, TTS narration 2.14s, and PDF export 0.67s. These are sufficient for a nonemergency decision support system.

5.6 Nearby Hospital Geolocation

Figure 7: Nearby Hospital Finder output showing geocoded location, distance-ranked .

The performance of the geolocation module has been tested in five different cities in India: Bapatla, Hyderabad, Chennai, Visakhapatnam, and Guntur. For each of the cases, the system has been able to correctly geocode the input location using the Nominatim API and has been able to retrieve the list of hospitals within a 10 km radius using the Overpass API. The results are summarised in Table IV.

The disease-specialty mapping has been able to correctly identify the specialty of doctors for Pneumonia cases as Pulmonologists, Fracture cases as Orthopaedic Surgeons, and Cardiomegaly cases as Cardiologists in all cases.

The fallback URL for Google Maps has been triggered for one of the cases in a rural area (Bapatla outskirts), where the results returned from the Overpass API were less than two, thus validating the effectiveness of the fallback approach. The average latency for the Nominatim API has been 0.55s, and the average latency for the Overpass API has been 1.87s, thus giving a total latency of 2.42s for the hospital

6. FUTURE SCOPE

The future scope for this system includes many areas that can be explored in the coming days. Disease can be expanded to include additional diseases such as tuberculosis, COVID-19, bone tumors, and rare diseases using additional training data sets.

Federated learning can be incorporated to improve the model using distributed hospital data sets without compromising patient data privacy. Geolocation can be improved by using real-time hospital bed availability, appointment bookings, and integration with Ayushman Bharat Digital Health Mission registry for verified hospital data in India. Development of a mobile app can be carried out to enable point-of-care deployment in rural and resource-constrained healthcare settings.

Language can be expanded to include all 22 scheduled languages in India, thereby making it accessible throughout the country. Advanced deep learning architectures such as Vision Transformers and diffusion models can be explored to improve both accuracy and XAI model results. Integration with hospital EHR systems and government health portals can be carried out to enable clinical workflow integration and population health monitoring.

7. CONCLUSION

In this paper, we introduced a comprehensive AIpowered X-ray analysis system that integrates body part detection using deep learning, multi-disease classification, XAI, multi-lingual clinical report generation, text-to-speech narration, cloud-based patient management, and geographic near hospital guidance in a unified whole to create a productionready diagnostic support tool.

Our system attains 92% average body part detection accuracy for 16 body parts and 0.91 average AUC-ROC for chest pathology detection using a novel CLIP-CNN hybrid detection framework.

The XAI module of our system was rated 4.1/5.0 by medical professionals for its overall usefulness, validating the effectiveness of multi-technique XAI for medical diagnosis. Multi-lingual report generation in English and Telugu addresses issues of accessibility in diverse linguistic populations.

The geolocation module was able to successfully identify an average of 5.2 specialist hospitals nearby per query, with a total search latency of 2.42 seconds, filling the critical gap between AI diagnosis and patient referral.

This project, as a Gradio web app deployed on Hugging Face Spaces with a MongoDB Atlas backend, shows the potential of fusing cutting-edge AI with real-time multilingual and geolocation capabilities to improve diagnostic radiology, ensuring equitable, explainable, and accessible AI-assisted healthcare.

8. REFERENCES

[1] P. Rajpurkar et al., "CheXNet: Radiologistlevel pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, Nov. 2017.

[2] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017, pp. 4700–4708.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.

[4] J. P. Cohen et al., "TorchXRayVision: A library of chest X-ray datasets and models," in Proc. Medical Imaging with Deep Learning (MIDL), 2022.

[5] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks," in Proc. IEEE CVPR, 2017, pp. 2097–2106.

[6] J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. AAAI, 2019, pp. 590– 597.

[7] P. Rajpurkar et al., "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," arXiv preprint arXiv:1712.06957, Dec. 2017.

[8] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proc. Int. Conf. Mach. Learn. (ICML), 2021, pp. 8748–8763.

[9] Z. Zhao et al., "CLIP in medical imaging: A comprehensive survey," arXiv preprint arXiv:2312.07353, 2023.

[10] M. Hartsock and G. Rasool, "Vision-language models for medical report generation and visual question answering: A review," Frontiers in Artificial Intelligence, vol. 7, 2024. doi: 10.3389/frai.2024.1430984.

[11] S. Müller et al., "Language augmentation in CLIP for improved anatomy detection on multimodal medical images," arXiv preprint arXiv:2405.20735, 2024.

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE ICCV, 2017, pp. 618–626.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. ACM KDD, 2016, pp. 1135–1144.

[14] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?" arXiv preprint arXiv:1712.09923, Dec. 2017.

[15] M. Aasem and M. J. Iqbal, "Toward explainable AI in radiology: Ensemble-CAM for effective thoracic disease localization in chest Xray images using weak supervised learning," Frontiers in Big Data, vol. 7, May 2024. doi: 10.3389/fdata.2024.1366415.

[16] R. Rajpoot, M. Gour, S. Jain et al., "Integrated ensemble CNN and explainable AI for COVID-19 diagnosis from CT scan and X-ray images," Scientific Reports, vol. 14, p. 24985, 2024.

[17] H. Srinivasan et al., "Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis," Computers in Biology and Medicine, 2024. doi: 10.1016/j.compbiomed.2024.108689.

[18] B. Jing, P. Xie, and E. Xing, "On the automatic generation of medical imaging reports," in Proc. ACL, 2018, pp. 2577–2586.

[19] L. C. Adams et al., "Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: A multilingual feasibility study," Radiology, vol. 307, no. 4, p. e230725, 2023.

[20] G. Dhillon et al., "Artificial intelligence in multilingual interpretation and radiology assessment for clinical language evaluation (AIMIRACLE)," Journal of Personalized Medicine, vol. 14, no. 9, p. 923, 2024.