An Explainable AI-Assisted Diagnosis System for Chest, Brain, and Bone Medical Imaging

Prof. Dipika Paranjape[1], Meenal Shelar[2], Aachal Khodape[3], Sumit Borse[4]
[1]Assistant Professor, NMIET PCET's - NMVPM's Nutan Maharashtra Institute of Engineering and Technology Pune, Maharashtra, India
[2,3,4]CSE - Artificial Intelligence PCET's - NMVPM's Nutan College of Engineering and Research Pune, Maharashtra, India

***Abstract*—The concept of deep learning showed that the system implementation includes combining visual and textual explanations to improve understanding. The paper discusses supporting the disease categories among multiple medical imaging tasks. Proposed system diagnosis involves chest and bone X-rays and brain-MRIs using EfficientNet-B0 and DenseNet-like Convolutional Neural Network (CNN). The system's unique characteristic is the use of Explainable AI (XAI) techniques like Gradient Class Activation Mapping (Grad-CAM) to output visual heatmaps by emphasizing relevant regions in medical images. The system takes medical images as the only input and passes them through CNN models based on the selected category. It then analyses images to emphasize key areas in them, while language-based clinical reasoning helps interpret the clear and meaningful insights. Moreover, the project was end-to-end implemented using a Streamlit-based web interface, which gave visual heatmap generation, confidence score, and report generation along with the correct medical references. The system also serves as a decision-support tool, which improves transparency and usability rather than being a replacement for clinical expertise.**

***Index Terms*—Explainable Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Medical Image Analysis, Gradient Class Activation Mapping, decision support tool.**

I. INTRODUCTION

Artificial Intelligence (AI) is increasingly playing an important role in modern healthcare, especially in medical imaging and diagnosis. A revolutionary progress has facilitated auto- mated detection and classification across different approaches, which enhances clinical decision-making at ease [1], [2]. Consistency and efficiency have resulted in a boost in the consistency and efficiency of image-based disease analysis through Convolutional Neural Networks (CNNs) [3]. The inability or the lack of transparency in reasoning, also referred to as the model's 'black box' nature, has always been an ambiguity. Due to this, deep learning models have remained a major barrier to the idea behind clinical adoption. In medical settings, clinicians require open and interpretable systems that are meant to justify predictions using clinically meaningful evidence. As a result, Explainable Artificial Intelligence (XAI) has come up as a key mechanism for encouraging trust, ac-countability, and acceptance in AI-assisted healthcare systems [4], [5]. Thus, the XAI pipelines also support overcoming the gaps between the algorithmic output and human clinical reasoning through the visual and contextual explanations.

Intelligent diagnosis systems are said not to just interpret the visuals or decisions but also speak towards integrated AI frameworks to improve clinical decision support. Modern AI techniques combined with data input like clinical text can produce more informed diagnoses [2], [6]. In the context, the framework integrates a visual model for predictions along with language-driven clinical reasoning for context-aware di-agnostic insights. Such techniques result in a better potential to improve system usability while allowing more intuitive interaction between the clinicians and the diagnostic support tool.

Whether the approaches are effective, neutral, and fair with various groups of patients is hard to determine [6]. The existing systems are unable to align a balance between accuracy and explainability. Others are more concerned with high predictive performance but are not transparent, and thus not easy to understand the justification. Although others offer visual explanations, they can fail to give sufficient context to the explanations. The paper introduces an Explainable AI-assisted diagnosis agent, which helps doctors make diagnoses by blending image analysis, visual explanations, and reasoning using suitable information retrieved from the knowledge base. The framework provides a clear and efficient diagnosis. Research combines information from medical images with the doctor's reasoning and explanations based on a literature review. The system also offers clinicians' prediction confidence scores, while supporting visual explanations via heatmaps, and structured natural language interpretations based on medical knowledge to provide concurrent evidence regarding results.

The key impact of the study can be summarized as:

- Developing a diagnostic system using explainable AI for medical image classification.
- Grad-CAM visualization highlights diagnostically relevant regions in medical images.
- A language-based reasoning module generates structured, textual explanations that correlate with the model predictions.
- A deployable web interface for interactive representations and automated report generation.

The rest of the paper is structured as follows: Section II reviews related work on explainable AI, and Section III broadly describes the proposed system architecture. Section IV discusses the proposed methodology, Section V demonstrates results and discussions, and Section VI concludes the paper and suggests future scope.

II. LITERATURE REVIEW

While Explainable Artificial Intelligence (XAI) is becoming a mandatory part in clinical settings, the goal is to remove 'black-box' AI models. The models need to be transparent, making it easier for the healthcare professionals to understand and trust them. XAI bridges the gap by providing techniques, applications, and solutions to the challenges using deep learning in healthcare [1]. In the domain of medical decision support, AI techniques evolve that combine detailed imaging analysis with its corresponding clinical data [2]. Using XAI is especially useful for brain-related issues. Doctors need to understand and trust the results and evaluations to find them helpful [3]. Deep learning has consistently been practiced in today's medical settings. On detecting the mammography images, it has improved its prediction of cancer risk. Thus, it helps enable transparency and trust. An increase in new learning techniques has considerably improved the prediction of breast cancer risk by analyzing full-field digital mammography images [5]. The use of technologies requires careful consideration of healthcare disparities. The risk of biases exists in AI systems, which may hallucinate outputs. To tackle this, there is a need of large datasets and features for building easier and understandable AI models [6].

To address the concerns, interactive and explainable models have been developed. For instance, MRI-based models used for diagnosing hepatocellular carcinoma make the diagnostic process more transparent and collaborative [7]. Researchers found ways to make AI's decision-making transparent, where visualization techniques played an important role [8]. For instance, Grad-CAM helps the system identify which part of the area is focused on to make a diagnosis. In addition to finding problems, explainable AI is also being used to suggest treatments for diseases like renal cell carcinoma. This will help doctors to make decisions during team discussions [9]. Combining learning with Natural Language Processing (NLP) has also allowed computers to automatically generate reports for chest X-Ray [10]. AI models must be easy to understand and should work well to be implemented in clinical settings [12]. Many research suggest that there a need of AI systems that are transparent and easy to define the things [8]. Though many have used architectures like EfficientNet with Grad-CAM for classification, still results are to be checked by humans [13]. Grad-CAM has become a tool for making AI models more understandable [14]. It has also been used in models like scale CNNs to classify different lung diseases [15]. The given approaches had tried to reduce the gap between AI models and practical medical diagnosis, specifically in MRI-based diagnosis [16]. Still, it is suggested that while visual explanations are easy to understand, they may not always provide precise insights compared to more detailed methods [17]. Overall, explainable AI is helping to make medical diagnoses more accurate and reliable. It is being used in different areas, including cancer diagnosis and treatment.

Recent studies are also looking at how to make AI systems more useful in life medical situations, yet they're not being used in real-time [18]. Although AI systems and medical experts have strengths, they don't always explain their thinking clearly. We need to trust the decisions made by experts [19]. Thus, we should make the decision-making process more transparent. AI tools for diagnosing diseases like those used in mammography have shown results, but they need to be tested more in real medical settings [20]. The need for reasoning is an increasing need for application in the medical field. Implementing it then helps medical experts increase their trust in AI systems. Many medical fields recognize the need for AI systems to explain how they make decisions [21]. The system helps us understand how the AI system analyses information. Some computer-aided diagnosis systems have been developed for lung diseases. They often have problems like uneven data and bias [22]. Integration of different techniques used for systems like COVID-19 may usually work well. However, its performance is often backed by the limitations of available data [23]. Methods detecting objects in images can highlight certain areas and might neglect the overall context [24].

Surveys on AI mention that there is no standard way to measure how well these systems work. It lacks standardization, which makes it difficult to compare different XAI approaches and assess how well they perform [25]. The concept of explainability is often considered a feature rather than a part of a system. It then led to a lack of evidence and justification, and may affect the trust [26]. Brain tumor detection can be done well using a mapping method, where the exact region is highlighted, aligning the prediction outcome. Though these techniques perform well, they are limited to a small dataset only [27]. These limitations can thereby impact the overall efficiency of the system. Existing classification systems don't justify the reasoning. Therefore, it becomes a problem because doctors need to understand why a system makes a certain diagnosis to trust it. It is harder to be efficient when there is a lack of explanation. Multi-label classification systems for chest X-rays also lack integrated explainability features [28]. In surgical applications, XAI methods show its own potential to generalize well but still face challenges in real-world deployment [29]. Certain methods interpret well for specific image types. However, these vg [30]. AI triage systems could help with infectious diseases, but they need more data to work well. A lack of sufficient data is holding back their potential. AI triage systems for infectious diseases are promising, but again limited by data scarcity [31]. Although there are several frameworks proposed for AI-based diagnosis, many of them are still theoretical and not fully implemented in practice [32].

Despite all these advancements, there are still important research gaps. Most existing systems focus on a single disease or a single organ, and do not provide a unified approach for multi-organ diagnosis. Also, there is very limited interaction between clinicians and AI systems, which makes it difficult to query or better understand predictions. In many cases, systems provide either visual explanations or textual explanations, but not both together in a meaningful way. In this project, we observed that most models are still at a prototype level and lack proper real-world validation. These limitations motivate the proposed system, which aims to combine multi-organ diagnosis with both visual and textual explainability, along with an interactive approach to better support clinical decision-making.

Table I shows the research gaps from existing studies and the goals of the proposed system. It explains the research challenges that each part of the system solves and why it is included in the system design. The objectives of the system are clearly linked to the research gaps; however, the proposed system design is based on the objectives listed.

III. PROPOSED SYSTEM ARCHITECTURE

In the paper, an approach is presented to help doctors to diagnose medical images like X-rays, CT scans, and MRIs using deep learning techniques and explainable AI. The system combines image analysis to explain how AI helps justify a diagnostic report, along with a language-based perception. Thus, diagnoses are clearer and easier to understand. Fig 1 illustrates how the system processes medical images while turning them into suitable outputs that doctors can use. The system is developed to be explainable, and be able to provide information without delays to make decisions in real-world medical situations.

Table I. Research Gaps and Project Objectives

Reference	Research Limitations (Gap)	Project Objective & Implementation
[10]	Existing systems help generate automated re-ports but lack interaction for clinicians. It may be an obstacle to a delay when taking a second opinion.	To implement an AI-assisted QA module in the system. It enables doctors to ask questions and understand the ex-act generated diagnostic report.
[12]	Current diagnostic systems help with expert explanations, but remain complex for junior doctors and trainees.	To make the doctor available, structured clinical explanations, along with references and clear summaries. Thus, it is also useful in medical education.
[28]	Current systems usually focus mainly on chest X-ray anomaly detection. Thus, the limitations exist in datasets for categories like brain and bone.	To develop a unified multi-Organ diagnosis system that supports chest, brain, and bone.
[31]	Current AI systems follow a single-user work-flow and lacks a collaborative interaction between multiple doctors.	To implement a case discussion room that al-lows doctors to review and discuss cases.
[32]	Different approaches provide conceptual explainability, but lack visual mechanisms that emphasize areas influencing model predictions.	To use the Grad-CAM technique to highlight anomalous image areas that justify why a prediction is made. It also provides confidence scores with a list of possible classes detected.

A. Input and Preprocessing module

The module consists of the initial stages of providing input to the explainable diagnostic system. It consists of the User interface layer, an Image upload and Processing module.

- User interface Layer: It is designed to let people upload medical images and choose a category among chest X-ray or a brain-MRI. The system then shows results, which include what it thinks is behind an image diagnosis in a region. It also provides clinical explanations that align with the heatmap visuals. To make a system accessible, a simple dashboard is set up, which enables smooth interaction for doctors and other healthcare professionals.
- Image processing module: This module is used for checking whether images are in the right format as required by the system or not. It prepares the images for analysis that involves resizing to 224 * 224 pixels. The images are normalized, and preprocessing steps are essential to reduce the variations in input.

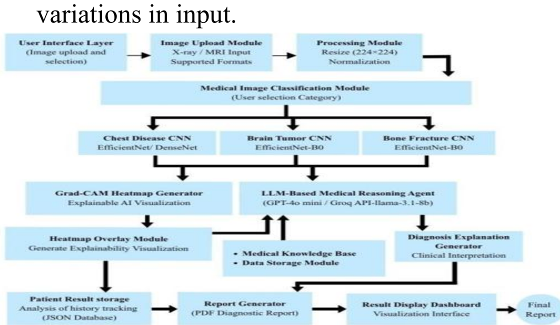

Fig. 1. Overall architecture for the proposed explainable diagnostic system

B. Medical Image Classification Module

The module for classifying medical images is established and consists of specific CNN models, including EfficientNet-B0 for the chest and bone category and DenseNet for brain MRIs. To ensure accuracy, separate CNN pipelines for each category are created.

- The Brain tumor CNN: It is used for analyzing MRI images to detect tumors. The system first checks for an anomaly in the brain scans. It then classifies these areas while inserting them in different categories. It doesn't just predict based on unusual scans, but also gives confidence scores.
- The Bone Fracture CNN: It helps to automatically find fractures in the bone scans. The goal is to look for breaks and unusual shapes. The system then helps us explain what category of fracture it is, along with the justifications for the same.
- The Chest Disease CNN: It helps us to automatically check and identify different diseases from chest X-ray images. It revolves around the idea of finding disease patterns from the chest image that indicate a problem. The system tends to provide a list of possible diseases and their confidence scores instead of just giving one answer.

C. Explainability module

The module consists of the factors affecting a prediction in the output results. It is important to understand how the model makes predictions.

- Grad-CAM heatmap generator: The module helps us create an activation map that shows what parts of the image are likely to make a prediction. It also gives the list of possible classes along with their confidence scores.

- Moreover, it helps highlight the areas in images that are most important for a model's decision. It proves to be helpful to a doctor and other specialists seeking a diagnosis and a judgment concurrently.
- Heatmap Overlay module: To help understand the decision made by a model, we overlapped a Grad-CAM heatmap with that of the original input image. It makes it easier to understand if the model is focusing on the relevant regions required for complete reasoning. Thereby, the approach helps us verify if the model's reasoning aligns with clinically related areas.

D. Clinical Reasoning and Knowledge Module

The module consists of three main components, which includes: LLM-Based Medical Reasoning Agent, Medical Knowledge Base, and Data Storage Module. It helps in interpreting model predictions using language-based reasoning and medical knowledge base.

- LLM-Based Medical Reasoning Agent: The language model takes numbers and categories as input to change them into text. It is important to understand what the model predicted. Doctors or junior trainees can find it useful since it can change the results from model pre-dictions into ready-to-understand clinical explanations. Thus, it helps to improve understanding of the diagnostic outcomes.
- Medical Knowledge Base: The concept 'Knowledge base' is known to store medical information. It includes clinical references, articles, and data from sources like PubMed. The knowledge base, thus, helps a reasoning agent work better to generate more accurate explanations.
- Data Storage Module: The system uses a data storage module to handle all the data created during processing. It is an important module since it helps data move easily between different system parts. Thus, it makes it easy to store records of our analyses. It makes doctors to check old results, helpful for long-term studies or training new models.

E. Diagnosis and Reporting Layer

This module helps in compiling the predictions, visual explanations, and reasoning a diagnostic report. It allows easy retrieval for follow-ups and maintains a history of analyzed medical cases.

- Diagnosis Explanation Generator: It helps combine results from different components within a system to create a structured medical report. There is a need to create understandable explanations for medical diagnoses that help doctors understand why a certain decision was made. Factors, including prediction scores and heatmaps into easy-to-read text.
- Patient Result Storage: The proposed system helps us maintain a record for each patient, including the prediction scores, key regions impacting prediction, and clinical reasoning. It is helpful for easy retrieval of old records while tracking a patient's health over time.
- Report Generator: It combines different types of med-ical reports into a single structured report. They are designed to be useful in a real-world clinical setting. Each report includes predictions about possible diseases, along with a confidence score that indicates its certainty. Reports also highlight the areas in medical scans that influenced a prediction. Thus, it is beneficial to easily understand the summary of the diagnostic process.

F. Result visualization Module

This module provides an interactive interface for reviewing predictions, heatmaps, and final reports.

- Result display dashboard: It combines model predictions, Grad-CAM heatmaps, and textual explanations into a simple Streamlit dashboard. It is because of its simplicity in the quick and easy deployment of a web interface without any complex coding. It is simple to integrate well with Python models to easily display representations and different reports. It also helps check the outputs efficiently.
- Final report interface: It helps to keep together a structured medical summary while combining pictures, texts, and references into an easy-to-interact interface. The patient's diagnostic reports can be saved for future reference. Thus, the Final Report Interface lets you under-stand in a visually appealing way while presenting them in a combined manner.

IV. PROPOSED SYSTEM METHODOLOGY

The proposed system follows a structured, phase-wise pipeline for transforming raw medical images into clinically interpretable diagnostic outputs. The given framework integrates deep learning–based disease detection, explainable artificial intelligence for enhancing transparency and usability in clinical environments. Publicly available medical imaging datasets are used for system development and evaluation. It includes chest X-ray images for disease classification, brain MRI datasets for tumor and neurological abnormality detection, and bone X-ray datasets for fracture identification. Data augmentation techniques are also used like rotation, flipping, and contrast enhancement are applied for the improvement of model generalization and to become a robust.

Phase 1: Medical Image Input The given system begins with a user interface where doctors can upload medical images having X-rays, CT scans, or MRIs as per user requirements that accepts different formats like JPG, PNG, DICOM, and NIfTI. Then a user can select the image category, such as chest, bone or brain and the image is sent to the appropriate model. The system first verifies if the uploaded image is in a valid format before processing it.

Phase 2: Image Preprocessing The images are preprocessed to improve model analysis. They are resized to a smaller size (224×224 pixels), and brightness is adjusted. To make a picture clear, normalization and resizing are used. The OpenCV and PIL libraries are used for these steps to make sure the model gets standard and uniform inputs.

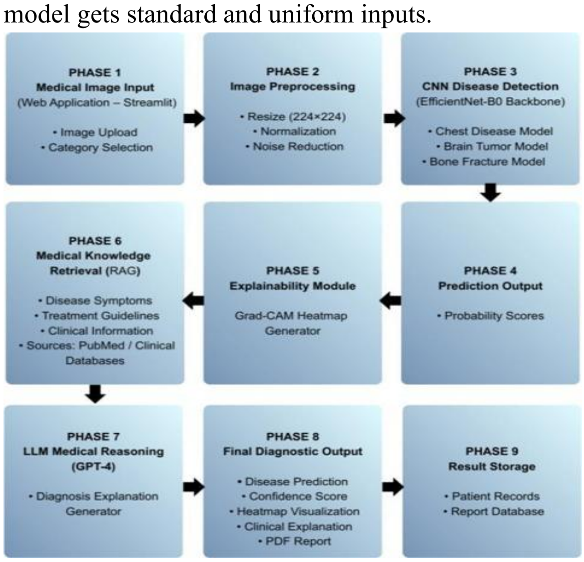

Fig. 2. Operational workflow for the proposed system

Phase 3: CNN Disease Detection Convolutional Neural Networks (CNNs) are used by the system to sort medical images into groups of illnesses. The design has models called EfficientNet-B0 and DenseNet. The EfficientNet-B0 model was picked because it has fewer parameters and works well for medical imaging tasks. To see how efficiently EfficientNet-B0 works, DenseNet is used as a standard model. Transfer learning is used with weights that have already been learned, which helps get better feature extraction even with small medical datasets. The CNN models use the images to learn hierarchical features, which are then used to make the final disease forecasts.

Phase 4: Prediction Output CNN models come up with probability numbers that show how likely it is that a disease will happen by listing all the possible categories. For multiclass classification, the right activation functions like softmax are used. Sigmoid is used to figure out how confident a forecast is. Then, these outputs are sent to the next parts in the system to be processed further.

Phase 5: Explainability Module To make things clearer, the system has an AI part that can be explained using the Grad-CAM method. By looking at the gradients from the convolutional feature maps, it makes activation maps that are unique to each class.

This method helps find important parts of the picture without changing how CNN was built in the first place. The heatmaps that were made show the places that affect the model's prediction. The heatmaps are put on top of the original image so that they are easy to see and understand.

Phase 6: Medical Knowledge Retrieval: This involves the use of the Retrieval-Augmented Generation (RAG) approach to improve contextual understanding. A retrieval module searches indexed biomedical data using semantic similarity to find relevant clinical information. The retrieved data includes details of disease characteristics, its symptoms and diagnostic insights which help support the reasoning process.

Phase 7: LLM Medical Reasoning: A large language model is used to generate clinically meaningful interpretations. The module takes inputs such as predicted disease names, confidence scores, and retrieved medical context. Then the model gives structured diagnostic explanations and clinical interpretations to connect AI outputs with human understanding.

Phase 8: Final Diagnostic Output: The outputs from the classification, explainability, and reasoning modules are com-bined to produce a complete diagnostic result. The result includes predicted disease labels, confidence scores, heatmap visualizations, and clinical explanations. A structured report is then generated for documentation and clinical use in PDF format.

Phase 9: Result Storage The system stores diagnostic outputs, including uploaded images, predictions, and generated reports in a JSON-based database. This allows case tracking with historical analysis, and easy future reference to support long-term clinical decision-making.

V. RESULTS AND DISCUSSIONS

This part looks at how well the planned explainable diagnostic system works and how easy it is to use in real life. This study looks at how well different types of medical scans can diagnose, how confident we can be in our predictions, and how easy they are to understand. The findings show that using explainable AI (XAI) methods along with deep learning models makes things clearer. During research, it also keeps up a reliable diagnostic performance.

A. Quantitative Diagnostic Performance

The proposed framework achieves good diagnostic accuracy across multiple imaging types, including chest X-rays, brain MRI scans, and bone X-ray images. The EfficientNet-B0-based architecture significantly learns important visual features from different medical datasets using transfer learning.

Table II summarizes the performance of the proposed models across different diagnostic modules.

Table II summarizes the performance of the proposed models across different diagnostic modules. The chest X-ray classification shows effective multi-label detection of common thoracic abnormalities, which is consistent with earlier transfer learning methods used in radiology. Brain MRI analysis provides a reliable identification of tumor regions and abnormal brain structures. In the same way, the bone fracture detection module performs well in identifying structural breaks present in fracture patterns. The results also demonstrate how the system achieves strong diagnostic accuracy. It also includes explainability methods that improve model transparency.

Table II. Performance Evaluation of the Proposed Models

Model	Accuracy	Precision	Recall	F1 Score
Brain Tumor Detection	86%	87%	85%	86%
Chest Disease Detection	75%	76%	74%	75%
Bone Fracture Detection	88%	89%	87%	88%

B. Confidence Estimation

For medical images across multiple modalities, the EfficientNet-B0-based classifier produces probabilistic predictions having confidence scores. These values indicate how certain the model is about each predicted diagnostic class. For multiclass classification tasks the model uses the Softmax activation function to convert raw model outputs into probability values.

$$\sigma(z_i) = \frac{e}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

where zi represents the raw score corresponding to classic, and K denotes the total number of diagnostic classes. For binary classification scenarios, the Sigmoid activation function is used:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

These probabilistic outputs help clinicians to evaluate the reliability of model predictions. They support informed clinical decision-making in AI-assisted diagnostic systems. They also help identify uncertain cases where further medical review or tests may be required.

C. Explainability Analysis Using Grad-CAM

To improve the interpretability of deep learning predictions, Gradient-weighted Class Activation Mapping (Grad-CAM) is used to visualize the regions within the images that influence the model's decisions. It helps generate heatmaps that high-light the regions influencing the predicted class. This helps clinicians check whether the model is focusing on medically relevant regions during diagnosis.

The important weight for each feature map is calculated using the gradient of the target class score:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{3}$$

where yc denotes the score for class c, Akij represents the activation at spatial location (i, j) in the kth feature map, and Z denotes the normalization factor equal to the number of pixels in the feature map.

The Grad-CAM localization map is then calculated as:

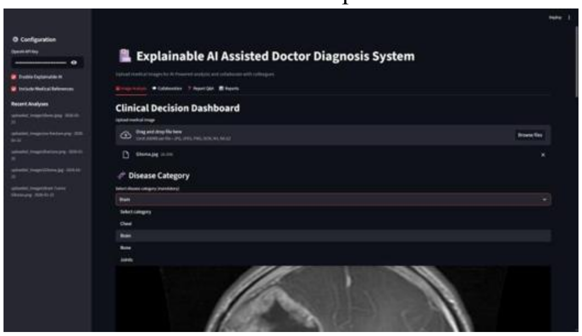

Fig. 3. Primary interface of the clinical decision dashboard for submitting medical images.

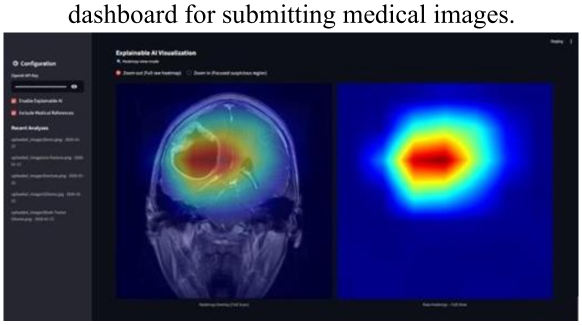

Fig. 4. Explainability visualization using Grad-CAM heatmaps for full medical image analysis.

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \tag{4}$$

The generated heatmaps are overlapped on the original medical image to highlight important regions. Experimental results show that the model consistently focuses on relevant anatomical areas in MRI and X-ray images. This supports the interpretability of the diagnostic predictions.

D. Implementation Results and System Interface

The section represents the implementation results of the proposed system through illustrative figures of the developed modules. These visuals show how the components within an integrated system work together in a clinical decision-support environment.

In the given Fig. 3 shows the Clinical Decision Dashboard that serves as the main interface for doctors to upload medical images. The system supports multiple medical image formats, including DICOM, NIfTI, MRI scans, X-rays, and standard image formats such as JPG, JPEG, and PNG. Before submitting the image, the user must select a disease category such as Brain, Chest, or Bone. This ensures that the image is directed to the appropriate CNN model tailored for that specific organ and diagnostic task.

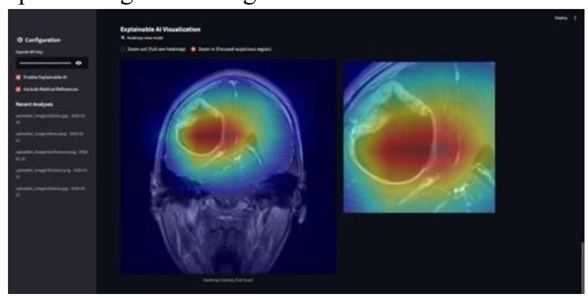

Fig. 5. Focused visualization highlighting suspicious regions for detailed examination.

Fig. 6. Multi-doctor collaboration system for case discussion and shared analysis.

Fig. 4 presents the Explainable AI Visualization module in full-image view mode Grad-CAM is used by the system to make a picture of the whole scan. The heatmap shows the areas that had the most impact on the model's estimate. Such visualization helps clinicians understand which areas of the image were important for the AI's decision-making.

Fig. 5 shows a focused view of the anomalous region identified by the model. In this view, the heatmap is concentrated around the detected abnormality like a brain tumor which enables clinicians to closely examine the region that is considered by the model. This also allows them to verify whether its decision-making aligns with the actual medical condition.

Fig. 6 shows the Collaboration System that allows multiple clinicians to discuss a case together. Users can create or join case discussions, such as those related to disease detections. Within this discussion space, clinicians can share opinions, review AI results, and may have access to insights provided by an AI assistant. Any interaction with the AI for a specific case is recorded and is visible to users who join the discussion later, ensuring continuity of information. Thus, it supports collaborative diagnosis to facilitate second opinions.

Fig. 7 shows the Medical Report Q&A System, where clinicians can ask questions related to the diagnostic report. The AI provides explanations based on stored medical knowledge. In the example shown, the system explains the causes of Tuberculosis. This feature helps convert technical AI results into clear medical explanations that can be used for communication.

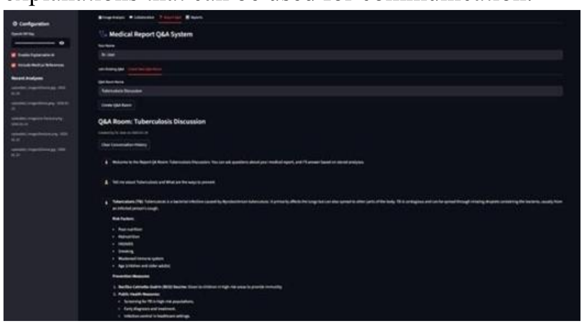

Fig. 7. AI-based medical report question-answer system for clinical explanations.

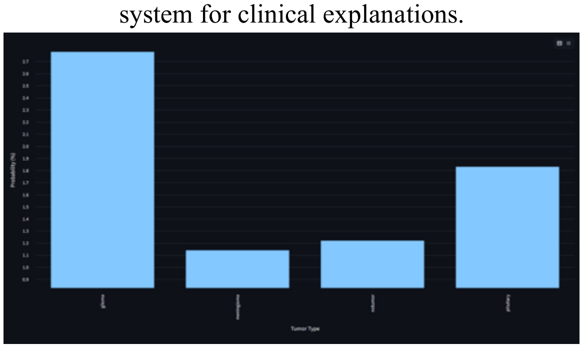

Fig. 8. Probability distribution of predicted brain tumor classes for an input MRI image.

Fig. 8 shows the probability scores assigned by the model to different brain tumor classes including glioma, meningioma, no tumor, and pituitary tumor. In this glioma has the highest probability, which shows its probable likelihood for the given input image. The other classes have comparatively lower probability values which highlights lower confidence scores.

This is useful in understanding not only the final prediction and how confident the model is across all possible classes. Such probability-based outputs support clinical interpretation and allow clinicians to assess uncertainty in the prediction.

These implementation results help us to demonstrate how the proposed framework integrates features to support collaborative decision-making. These features may consist of deep learning analysis, explainable AI visualization, clinical reasoning tools, and using knowledge sources like PubMed. Moreover, the system improves usability by providing interactive modules at different stages of diagnosis. It ensures a better understanding of AI outputs by supporting informed clinical workflow.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, the authors present a fully automated medical image diagnostic system based on state-of-the-art deep learning methodologies across various medical imaging modalities: magnetic resonance imaging (MRI), chest X-ray, and com-puted tomography (CT). Also, it extracts the features from the images and classifies them using the EfficientNet-B0. Other explainable AI techniques explain the predictions to make it more transparent. Grad-CAM visualizes activation maps, showing the areas classified as essential in making particular decisions. Hence, the model may obtain convincing results in classification while being quite interpretable. Certainly, such suitable explanations will allow clinicians to visualize whether the model looks at medically relevant regions when scanning medical images. The developed system demonstrated high accuracy and interpretability, as shown in most of the existing literature, which positions these systems as supporting tools for clinicians in diagnostic decision-making. The system requires further validation and extension on a wider variety of local medical datasets to ensure good generalization. Other imaging modalities, such as ultrasound and histopathology, could also be explored. More recent deep learning architectures, such as EfficientNetV2 or transformers, could be used for feature representation and prediction. Using the best methods for explainability in practice in real life will make AI even more reliable and useful for helping doctors with diagnosis.

REFERENCES

[1] J. Smith, A. Johnson, and M. Lee, “Explainable Artificial Intelligence (XAI) in Medical Imaging: Techniques, Applications, Challenges, and

Future Directions," *Advanced Mechanical and Mechatronic Systems*, vol. 1, no. 1, 2026.

[2] R. Patel and S. Kumar, "The Future of Multimodal Artificial Intelligence Models for Integrating Imaging and Clinical Metadata: A Narrative Review," *Diagnostic and Interventional Radiology*, 2025.

[3] A. Gupta and P. Singh, "Explainable AI in Diagnostic Radiology for Neurological Disorders: A Systematic Review, and What Doctors Think About It," *Diagnostics*, vol. 15, no. 2, 2025.

[4] K. Zhao and L. Li, "How Genomics and Multi-Modal AI Are Reshaping Precision Medicine," *Frontiers in Medicine*, 2025.

[5] S. Wang and J. Liu, "Deep Learning in Breast Cancer Risk Prediction: A Review of Recent Applications in Full-Field Digital Mammography," *Frontiers in Oncology*, 2025.

[6] Y. Li, P. Parikh, and R. Costa, "A Critical Look into Artificial Intelligence and Healthcare Disparities," *Frontiers in Artificial Intelligence*, vol. 8, 2025.

[7] R. Thompson and V. Patel, "Interactive Explainable Deep Learning Model for Hepatocellular Carcinoma Diagnosis at Gadoxetic Acid-Enhanced MRI: A Retrospective, Multicenter Diagnostic Study," *Radiology: Imaging Cancer*, 2025.

[8] A. Chaddad, Y. Hu, Y. Wu, B. Wen, and R. Kateb, "Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview," *arXiv preprint*, 2025.

[9] J. Kufel *et al*., "Multi-Label Classification of Chest X-Ray Abnormalities Using Transfer Learning Techniques," *Journal of Personalized Medicine*, vol. 13, no. 10, 2023.

[10] R. S. Jadhav *et al*., "Chest X-Ray Image Analysis and Report Generation Using Advanced Deep Learning Techniques and NLP," in *Proc. 2025 12th Int. Conf. Emerging Trends Engineering Technology, Signal and Information Processing (ICETET-SIP)*, 2025.

[11] N. Y. Gharaibeh, "Enhancing Interpretability in Brain Tumor Detection: Leveraging Grad-CAM and SHAP for Explainable AI in MRI-Based Cancer Diagnosis," *Applied Computer Science*, 2025.

[12] Y. Singh *et al*., "Beyond Post Hoc Explanations: A Comprehensive Framework for Accountable AI in Medical Imaging Through Transparency, Interpretability, and Explainability," *Bioengineering*, vol. 12, no. 8, 2025.

[13] A. M. Alqhatni *et al*., "Automated Classification and Explainable AI Analysis of Lung Cancer Stages Using EfficientNet and Gradient-Weighted Class Activation Mapping," *Frontiers in Medicine*, 2025.

[14] A. M. Fayyaz *et al*., "Grad-CAM (Gradient-Weighted Class Activation Mapping): A Systematic Literature Review," *Computers in Biology and Medicine*, 2025.

[15] O. Sarkar *et al*., "Multi-Scale CNN: An Explainable AI-Integrated Unique Deep Learning Framework for Lung-Affected Disease Classification," *Technologies*, vol. 11, no. 5, 2023.

[16] W. Nhlapho, M. Atemkeng, Y. Brima, and J.-C. Ndogmo, "Bridging the Gap: Exploring Interpretability in Deep Learning Models for Brain Tumor Detection and Diagnosis from MRI Images," *Information*, vol. 15, no. 4, 2024.

[17] E. Mohammad and M. Hamid, "Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models," *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, 2025.

[18] S. S. Band *et al*., "Explainable Artificial Intelligence for Medical Imaging Systems Using Deep Learning: A Comprehensive Review," *Informatics in Medicine Unlocked*, 2023.

[19] P. Roy and S. Mehta, "AI in Medical Diagnosis: AI Prediction and Human Judgment," *Artificial Intelligence in Medicine*, 2024.

[20] L. Green and M. Black, "AI-Integrated Screening to Replace Double Reading of Mammograms: A Population-Wide Accuracy and Feasibility Study," *Radiology: Artificial Intelligence*, 2024.

[21] A. A. Biswas, "A Comprehensive Review of Explainable AI for Disease Diagnosis," *Array*, 2024.

[22] R. Mukhlis *et al*., "A Novel CAD Framework with Visual and Textual Interpretability: Multimodal Insights for Predicting Respiratory Diseases," in *Proc. 2024 8th Int. Artificial Intelligence and Data Processing Symposium (IDAP)*, 2024.

[23] F. Ahmmed *et al.*, “An Improved Hybrid Deep Learning Approach with Explainable AI for COVID-19 Detection,” in *Proc. European Conf. Cognitive Ergonomics*, 2025.

[24] P. P. Sarma Borah *et al.*, “A Comprehensive Study on Explainable AI Using YOLO and Post Hoc Method on Medical Diagnosis,” *Journal of Physics: Conference Series*, 2024.

[25] Y. Yukta, A. P. Biswas, and S. Kashyap, “Application of Explainable Artificial Intelligence in Medical Health: A Systematic Review of Interpretability Methods,” in *Proc. 2024 OPJU Int. Tech. Conf. (OTCON)*, 2024.

[26] G. Kumar and V. Singh, “Evaluation of Artificial Intelligence Techniques in Disease Diagnosis and Prediction,” *Discover Artificial Intelligence*, vol. 3, 2023.

[27] T. Hussain and H. Shouno, “Explainable Deep Learning Approach for Multiclass Brain Magnetic Resonance Imaging Tumor Classification and Localization Using Gradient-Weighted Class Activation Mapping,” *Information*, vol. 14, no. 12, 2023.

[28] “Development of an Artificial Intelligence-Generated, Explainable Treatment Recommendation System for Urothelial Carcinoma and Renal Cell Carcinoma to Support Multidisciplinary Cancer Conferences,” *Journal of Precision Medicine*, 2025.

[29] A. Sharma and P. Desai, “Applications of Explainable Artificial Intelligence in Diagnosis and Surgery,” *Diagnostics*, vol. 12, no. 2, 2022.

[30] N. I. Papandrianos *et al.*, “An Explainable Classification Method of SPECT Myocardial Perfusion Images in Nuclear Cardiology Using Deep Learning and Grad-CAM,” *Applied Sciences*, vol. 12, no. 15, 2022.

[31] J. Li, Q. Zhang, and Y. Wang, “Artificial Intelligence–Enabled Rapid Diagnosis of Patients with COVID-19,” *Nature Medicine*, 2020.

[32] M. Ahmed and K. Reddy, “Medical Diagnostic Systems Using Artificial Intelligence Algorithms: Principles and Perspectives,” *IEEE Access*, vol. 8, 2020.