Polytechnic Journal

Polytechnic Journal

Volume 15 | Issue 1 Article 1

February 2025

Incorporation of XAI and Deep Learning in Biomedical Imaging: A Review

Sushil Kumar Singh
Associate Professor, Marwadi University, Rajkot, Gujarat, India,,
sushilkumar.singh@marwadieducation.edu.in

Bal Virdee
London Metropolitan University, Centre for Communications Technology, School of Computing and Digital Media, UK,

Saurabh Aggarwal
San Jose State University, USA,

Abhilash Maroju
Research Graduate, Department of Information Technology, University of the Cumberlands, USA,

How to Cite This Article

Singh, Sushil Kumar; Virdee, Bal; Aggarwal, Saurabh; and Maroju, Abhilash (2025) "Incorporation of XAI and Deep Learning in Biomedical Imaging: A Review," *Polytechnic Journal*: Vol. 15: Iss. 1, Article 1.
DOI: https://doi.org/10.59341/2707-7799.1845

Incorporation of XAI and Deep Learning in Biomedical Imaging: A Review

Abstract

Artificial Intelligence (AI) and Deep Learning (DL) technologies have revolutionized disease detection, particularly in Medical Imaging (MI). While these technologies demonstrate outstanding performance in image classification, their integration into clinical practice remains gradual. A significant challenge lies in the opacity of Deep Neural Network (DNN) models, which provide predictions without explaining their structure. This lack of transparency poses severe issues in the healthcare industry, as trust in automated technologies is critical for doctors, patients, and other stakeholders. Concerns about liability in autonomous car accidents are comparable to those associated with deep learning applications in medical imaging. Errors such as false positives and false negatives can negatively affect patients' health. Explainable Artificial Intelligence (XAI) tools aim to address these issues by offering understandable insights into predictive models. These tools can enhance confidence in AI systems, accelerate the diagnostic process, and ensure compliance with legal requirements. Driven by the motivation to advance technological applications, this work provides a comprehensive review of Explainable AI (XAI) and Deep Learning (DL) techniques tailored for biomedical imaging diagnostics. It examines the state-of-the-art methods, evaluates their clinical applicability, and highlights key challenges, including interpretability, scalability, and integration into healthcare. Additionally, the review identifies emerging trends and potential future directions in XAI research, offering a structured categorization of techniques based on their suitability for diverse diagnostic tasks. These findings are invaluable for healthcare professionals seeking accurate and reliable diagnostic support, policymakers addressing regulatory and ethical considerations, and AI developers aiming to design systems that balance innovation, safety, and clinical transparency.

Keywords

Explainable AI (XAI), Deep Neural Networks (DNN), Medical Imaging, Disease Diagnosis, Transparency in AI

REVIEW

Incorporation of XAI and Deep Learning in Biomedical Imaging: A Review

Sushil K. Singh [a,*], Bal Virdee [b], Saurabh Aggarwal [c], Abhilash Maroju [d]

[a] Marwadi University, Rajkot, Gujarat, India
[b] London Metropolitan University, Centre for Communications Technology, School of Computing and Digital Media, UK
[c] San Jose State University, USA
[d] Department of Information Technology, University of the Cumberlands, USA

Abstract

Artificial Intelligence (AI) and Deep Learning (DL) technologies have revolutionized disease detection, particularly in Medical Imaging (MI). While these technologies demonstrate outstanding performance in image classification, their integration into clinical practice remains gradual. A significant challenge lies in the opacity of Deep Neural Network (DNN) models, which provide predictions without explaining their structure. This lack of transparency poses severe issues in the healthcare industry, as trust in automated technologies is critical for doctors, patients, and other stakeholders. Concerns about liability in autonomous car accidents are comparable to those associated with deep learning applications in medical imaging. Errors such as false positives and false negatives can negatively affect patients' health. Explainable Artificial Intelligence (XAI) tools aim to address these issues by offering understandable insights into predictive models. These tools can enhance confidence in AI systems, accelerate the diagnostic process, and ensure compliance with legal requirements. Driven by the motivation to advance technological applications, this work provides a comprehensive review of Explainable AI (XAI) and Deep Learning (DL) techniques tailored for biomedical imaging diagnostics. It examines the state-of-the-art methods, evaluates their clinical applicability, and highlights key challenges, including interpretability, scalability, and integration into healthcare. Additionally, the review identifies emerging trends and potential future directions in XAI research, offering a structured categorization of techniques based on their suitability for diverse diagnostic tasks. These findings are invaluable for healthcare professionals seeking accurate and reliable diagnostic support, policymakers addressing regulatory and ethical considerations, and AI developers aiming to design systems that balance innovation, safety, and clinical transparency.

Keywords: Explainable AI (XAI), Deep Neural Networks (DNN), Medical imaging, Disease diagnosis, Transparency in AI

1. Introduction

Deep Neural Networks (DNNs) have demonstrated remarkable performance in image classification tasks, often surpassing both human experts and traditional artificial intelligence (AI) methods. This success has ignited significant interest in applying AI to biomedical imaging, particularly for tasks such as image segmentation and classification. Among the state-of-the-art models, Convolutional Neural Networks (CNNs) have become the standard for these tasks, while other advanced architectures, such as encoder-decoder frameworks and transformer-based models, are also being actively developed to enhance performance and versatility. Medical image segmentation involves identifying regions of interest (RoI) by labeling each pixel in an image and grouping related areas together, such as distinguishing lesions or anatomical structures. AI models can determine whether an image represents a benign or malignant condition or handle more complex multi-class situations for classification tasks. Biomedical images are acquired using various modalities, including

Received 5 December 2024; accepted 5 December 2024.
Available online 6 February 2025

* Corresponding author.
E-mail addresses: sushilkumar.singh@marwadieducation.edu.in, sushil.sngh1@ieee.org (S.K. Singh), b.virdee@londonmet.ac.uk (B. Virdee), sorav77@gmail.com (S. Aggarwal), doctorabhilashmaroju@gmail.com (A. Maroju).

https://doi.org/10.59341/2707-7799.1845

ultrasounds, X-rays, CT scans, and MRI. AI has emerged as a critical tool in assisting disease diagnosis, enabling more precise analysis of features such as tumour size, texture, and shape—often surpassing the capabilities of human radiologists.

A notable application is the detection of ocular diseases using fundus imaging, where AI systems have shown impressive diagnostic accuracy [1–6]. The growing accessibility of AI tools has empowered domain experts to apply machine learning methods without requiring deep technical expertise in underlying algorithms. This democratization of AI enables better diagnostic outcomes, especially in regions with a shortage of radiologists. For instance, the UK reportedly has only 2000 radiologists, creating a pressing need for automated diagnostic systems to support healthcare delivery. Despite these advancements, ensuring the transparency and reliability of AI predictions remains critical, particularly in regulated industries like healthcare. While simpler, deterministic models such as shallow decision trees offer inherent explainability, more complex models like DNNs are often criticized for their "black box" nature, where the decision-making processes are opaque even to the developers [7]. To build trust and promote the widespread adoption of AI in diagnostics, it is essential to develop explainable AI (XAI) techniques that clarify and validate model predictions. This will improve user confidence and ensure compliance with legal and ethical standards, paving the way for transformative advancements in image-based diagnosis.

Deep learning models are highly effective in generating predictions from complex data, such as those used in medical imaging. These models can classify images, diagnose diseases, and detect patterns that are often difficult for humans to recognize. However, a significant challenge with these models is their tendency to function as a "black box" [8]. This means that while the model provides an output (e.g., predicting whether a tumor is malignant), it does not explain the process by which it arrived at the conclusion. This is akin to using a calculator that shows the result without revealing the steps taken to compute it. This lack of transparency poses a critical issue for medical practitioners, especially when life-and-death decisions are at stake. In medical contexts, understanding not only the prediction but also the reasoning behind it is essential [9]. Errors such as false positives—incorrectly identifying a condition as present—or false negatives—failing to detect an actual condition—can lead to serious harm, eroding trust in the technology [10]. Deep Neural Networks (DNNs) mimic the interconnected structure of neurons in the human brain, but our limited understanding of neural connections makes it challenging to explain how these models generate their results fully. This opacity becomes particularly concerning in high-stakes applications like healthcare, where false positives and negatives can have dire consequences for patient outcomes. Moreover, neural networks require vast amounts of training data and computational resources to achieve reliable results. Their complexity often makes them inaccessible to non-technical users, who may struggle to understand how decisions are made [11].

The need for interpretable models is especially pressing in fields involving complex problems, such as simulating cognitive processes in children or diagnosing rare medical conditions. In such scenarios, the demand for transparency and interpretability serves as a guiding principle. Interpretable AI (XAI) aims to address these challenges by elucidating how predictions are made, which is particularly important in regulated industries like healthcare. XAI techniques not only help simplify decision-making but also foster stakeholder trust and ensure patient safety by providing actionable insights into the model's decision-making process. As deep learning continues to gain traction in healthcare and other regulated domains, the integration of XAI has become increasingly vital. By offering interpretable explanations, XAI bridges the gap between model performance and user understanding, ensuring that advanced AI systems can be deployed responsibly and effectively [12]. Such advancements are key to aligning cutting-edge technologies with the ethical and practical demands of critical applications like medical diagnostics. XAI model for Explainable Segmentation is shown in Fig. 1.

Various techniques and tools have been developed to enhance the explainability of Deep Neural Networks (DNNs) explainability. One common approach is visualizing layer-wise activations and features of trained networks, offering insight into how models process input data. Additionally, the interplay between computer vision and natural language processing (NLP) has been explored to make AI models more interpretable. Methods such as visual question-answering systems and image annotation tools aim to provide human-friendly explanations by combining visual and textual information. Industry-academia collaborations have launched Explainable AI (XAI) challenges to advance explainability methods for applications in domains like finance and healthcare [13]. In healthcare, the need for transparency extends beyond imaging systems to include contextual language models such as Bidirectional Encoder

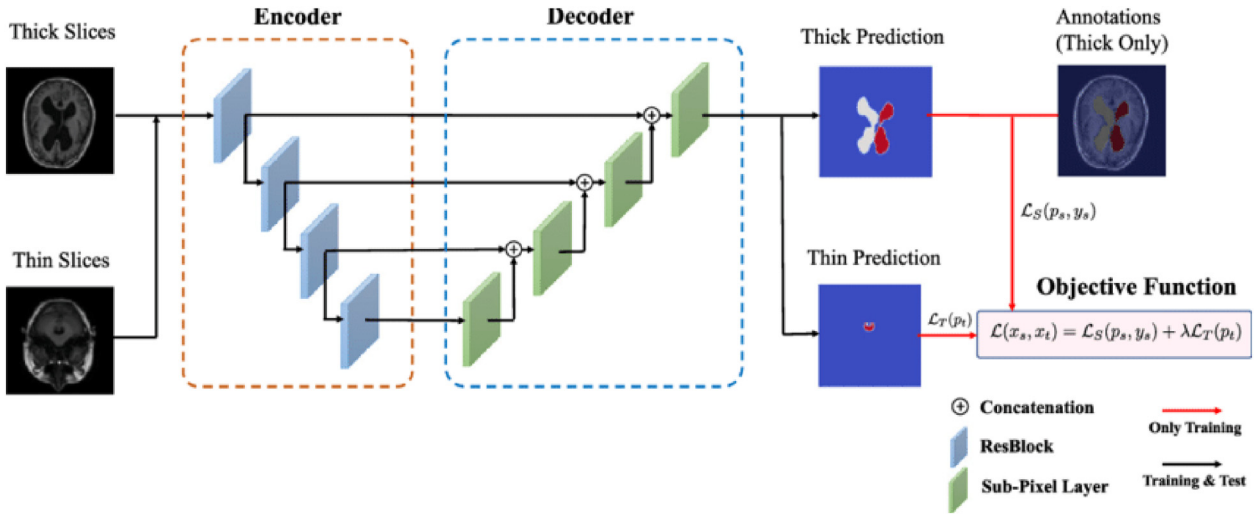

Fig. 1. XAI model for Explainable Segmentation [8].

Representations from Transformers (BERT). While DNNs often struggle to provide comprehensible justifications for their decisions, symbolic AI methods—such as knowledge graphs—offer a more interpretable alternative by leveraging structured, human-readable representations.

In regulated industries like healthcare, where understanding not just *how* a model made a decision but also *why* it failed is essential, integrating XAI into deep learning models has become indispensable [14]. Addressing the explainability gap is crucial for improving model performance, fostering user trust, and ensuring compliance with ethical and legal standards. Recent research has focused on bridging this gap, proposing taxonomies and evaluation frameworks for XAI, as well as exploring visual analytics techniques to demystify the inner workings of deep learning models. Post-hoc explainability approaches have gained significant attention, generating explanations after model predictions. These include strategies for assessing trustworthiness in DNNs, improving interpretability in machine learning systems, and enhancing the transparency of decision-making processes. Specific studies have examined the application of XAI in medical imaging, radiology, and clinical decision support systems (CDSS), highlighting its importance in healthcare contexts [15]. Moreover, emerging research emphasizes integrating explainability techniques directly into model architectures, enabling real-time, interpretable predictions. As healthcare increasingly adopts AI-driven tools, the development of domain-specific XAI solutions tailored for diverse tasks—such as tumor detection, disease progression analysis, and patient outcome prediction—has become a priority. Such advancements ensure that AI systems can support critical decision-making processes while maintaining safety, reliability, and user trust.

In this study, we use the term "Explainable Artificial Intelligence" (XAI) as an umbrella concept encompassing related notions such as "interpretable AI," "understandable AI," and "trustworthy AI." For Convolutional Neural Network (CNN) models, explainability refers to any techniques or supplementary information that help elucidate the underlying decision-making processes of these models, enabling users to comprehend *how* and *why* specific predictions are made. The literature reviewed in this study is collected from diverse and reputable sources, including Web of Science, ScienceDirect, IEEE Xplore, Google Scholar, and other scholarly databases. The focus was on studies addressing reliable, interpretable, and explainable AI solutions specifically for biomedical applications [16]. These applications span various domains, such as medical imaging, clinical diagnostics, and patient outcome predictions, reflecting the increasing integration of AI into healthcare workflows. This comprehensive review also emphasizes the importance of aligning XAI approaches with the unique demands of biomedical contexts. For instance, explainability methods in AI for healthcare must address both technical and ethical challenges, such as minimizing bias, ensuring fairness, and maintaining patient privacy. Furthermore, the study explores how integrating domain knowledge—such as medical ontologies or expert annotations—can enhance the interpretability of AI systems and build trust among practitioners. By systematically analyzing existing research, this study highlights critical gaps in current XAI methodologies, such as

limited real-time applicability, insufficient focus on multi-modal data integration, and challenges in communicating complex AI decisions to non-technical users. Addressing these gaps is crucial for developing next-generation AI systems that are not only high-performing but also transparent and trustworthy. Ultimately, this work is a foundation for advancing XAI practices in biomedical applications, supporting a future where AI-powered tools can safely and effectively assist in critical healthcare decision-making processes.

The main contributions of this research are as follows:

- This manuscript comprehensively reviews Explainable Artificial Intelligence and Deep Learning for Biomedical Imaging Diagnostics.
- Provides a thorough assessment of XAI approaches, with an emphasis on their applicability to various imaging modalities and medical specialties, in contrast to earlier surveys.
- A thorough analysis of XAI methods for comprehending DNN models, as well as a classification of these methods according to their range, suitability, and applications.
- Discuss existing challenges in Biomedical Imaging Diagnostics based on the related studies.
- Identifying future directions according to the categorizing XAI techniques.
- The findings are crucial for healthcare professionals, policymakers, and AI engineers striving to integrate advanced technology into clinical workflows while ensuring accuracy, reliability, and ethical compliance.

The manuscript's organization is as follows: Section 2 discusses related studies based on the seminal title of the manuscript, with limitations and merits. Section 3 introduces explainability techniques for DNN. Section 4 offers Explainability for Biomedical Imaging with categories of BI. Section 5 discusses the application and Challenges of XAI in Biomedical Imaging. Finally, Section 6 concludes the manuscript.

2. Related works

This section delivers a detailed discussion of existing research studies, emphasizing their descriptions, methodologies, and the challenges they address in the context of biomedical imaging applications. Key focus areas include image classification, image segmentation, automated diagnosis, predictive modeling, clinical decision support systems (CDSS), and the role of Explainable AI (XAI) in enhancing model interpretability and trustworthiness. Applications of Deep Neural Networks (DNN) in medical contexts is shown in Table 1.

Table 1. Applications of Deep Neural Networks (DNN) in medical contexts.

Application	Description	Challenges	Example
Image classification	DNNs automatically learn features from medical images and classify them (e.g., benign vs malignant tumours) [17].	Interpretability issues; Black-box nature of DNNs makes it difficult to understand how decisions are made.	IBM Watson for health used DNNs to aid oncologists in diagnosing cancers.
Image segmentation	DNNs segment medical images, marking regions of interest like lesions or tumors, useful in treatment planning.	Errors such as false positives and false negatives can lead to misdiagnosis or unnecessary treatment.	Deep learning-based chest X-ray image segmentation for pneumonia or COVID-19 detection.
Automated diagnosis	AI systems diagnose diseases by processing large datasets, identifying patterns, and predicting outcomes, reducing clinician workload.	Bias in the training data can lead to incorrect diagnosis for certain patient groups [18].	Watson's oncology system helps suggest treatments for cancer patients but faced limitations in different regions.
Predictive modeling	DNNs predict disease progression and patient outcomes based on image and patient data.	Poor generalizability if the model is not trained on diverse datasets; overfitting on internal hospital data.	AI models predicting the likelihood of pneumonia or COVID-19 progression based on medical imaging.
Assisting in clinical decision support	DNN models assist clinicians by providing additional insights into potential diagnoses or treatment plans.	Lack of transparency in how decisions are derived poses trust issues for clinicians and patients [19].	Clinical decision support systems (CDSS) use DNNs to help in diagnosing eye diseases based on fundus images [22].
XAI for explainability	XAI techniques are needed to make predictions understandable and trustworthy, helping clinicians interpret the decisions of DNN models [20].	Trade-off between accuracy and interpretability; post-hoc methods can provide explanations, but they may not always reflect the true decision-making process [21].	DARPA's XAI program focuses on creating explainable DNN models for medical applications [23].

- **Image Classification:** In image classification, studies have illustrated the effectiveness of deep learning models in differentiating between various medical conditions, such as distinguishing benign from malignant tumors or pinpointing disease states from X-rays, MRIs, and CT scans, despite advancements, challenges persist, such as addressing imbalanced datasets, improving generalizability across diverse patient inhabitants, and minimizing false positives and negatives.
- **Image Segmentation:** In image segmentation, research has focused on delineating provinces of interest (PoIs), such as tumors, lesions, or organs, with pixel-level accuracy. Techniques like U-Net, Mask R−CNN, and transformer-based models (T-Model) have shown favourable results, yet challenges remain in segmenting complex structures, mitigating artifacts in imaging data, and executing robust performance across different imaging modalities.
- **Automated Diagnosis:** For automated diagnosis, AI models aim to provide genuine and rapid diagnostic predictions, decreasing the workload of medical practitioners. However, integrating these tools into clinical workflows is problematized by concerns over transparency, liability, and the need for validation in real-world scenarios.
- **Predictive Modeling:** In predictive modeling, studies seek to forecast disease progression, treatment outcomes, or patient survival based on multimodal data, including imaging, genomics, and electronic health records (EHRs). There are challenges and limitations, which include handling missing or noisy data, ensuring model interpretability, and managing computational complexity in large datasets.
- **Clinical Decision Support System:** For clinical decision support systems (CDSS), AI-powered tools assist healthcare providers by offering insights based on vast amounts of medical data. While these systems can enhance decision-making, they face hurdles in ensuring real-time performance, user-friendliness, and clinician acceptance.
- **Explainable AI:** Finally, Explainable AI (XAI) has emerged as a crucial component for building trust in AI systems by offering transparent and interpretable predictions. Studies in this area explore methods such as saliency maps, attention mechanisms, and rule-based systems to clarify model decisions. XAI has the following limitations and issues: balancing explainability with model accuracy, ensuring scalability across applications, and tailoring explanations to the needs of diverse stakeholders.

This section underscores the significance of managing these challenges to unlock AI's full potential in biomedical imaging. Future research must focus on developing robust, explainable, and ethically sound AI systems that can seamlessly fuse into healthcare settings to enhance patient consequences and optimize clinical workflows.

3. Explainability techniques for DNN

This section discusses the explainability techniques for utilizing deep neural networks. These techniques make these models' forecasts and decision-making processes translucent and interpretable. Fig. 2 shows the hierarchical order of Explainability techniques for DNN. They are categorized into Intrinsic Interpretability Methods, Post-hoc Interpretability Methods, and Model-specific versus Model-agnostic Approaches.

i) Intrinsic Interpretability Methods

These techniques are created into the standard's architecture to produce it intrinsically interpretable.

- **Attention Mechanisms:** It highlights essential parts of the input data that contribute most to a decision. For instance, in medical imaging, attention maps can show specific regions of an X-ray that are critical to the prediction.
- **Self-Explaining Networks:** These networks are designed to produce explanations as part of their outputs. For example, prototype networks classify inputs by comparing them to learned prototypes, providing human-understandable reasoning.

ii) Post-hoc Interpretability Methods

These methods analyze and explain the predictions of pre-trained models without altering their architectures.

- **Visualization Techniques:** Saliency Maps and feature Visualization are essential techniques for posthoc interpretability Methods. In saliency maps, highlight the regions in input data (e.g., image pixels) that most influence the model's prediction. Techniques include Grad-CAM (Gradient-weighted Class Activation Mapping) and Integrated Gradients. Feature Visualization explores the patterns and features learned by individual neurons or layers of the network. This helps understand what kind of input activates a specific neuron or layer.

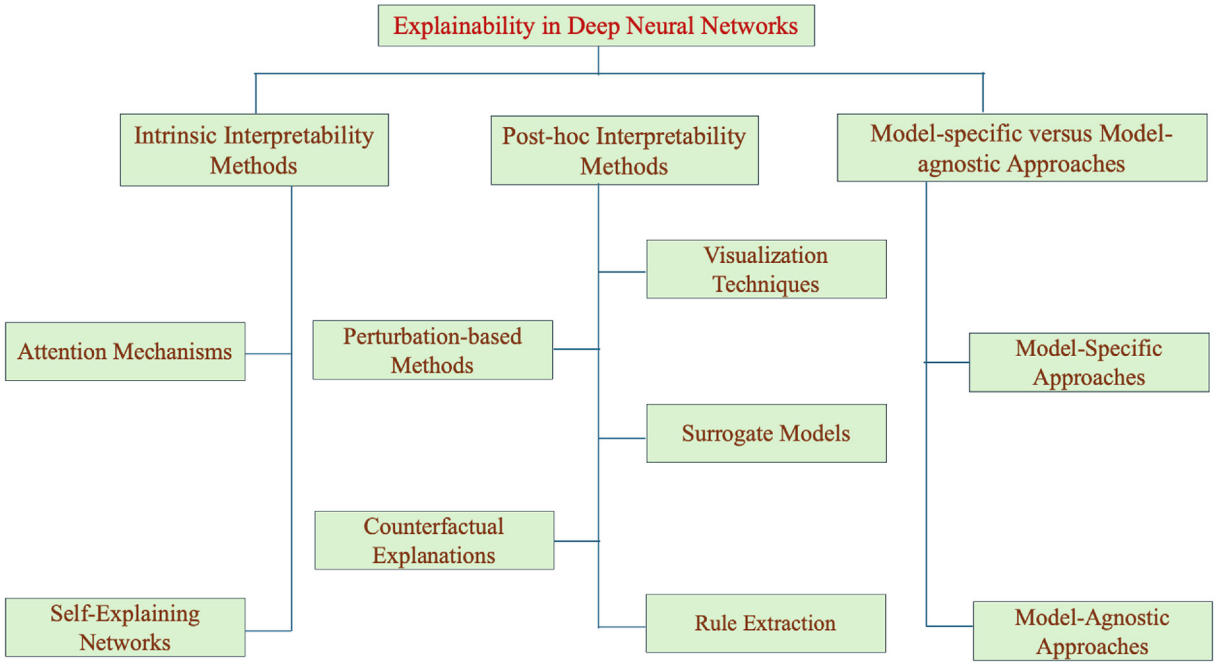

Fig. 2. Explainability techniques for DNN.

- **Perturbation-based Methods:** Occlusion Sensitivity is an essential part of perturbation-based methods, in which input data are systematically occluded (e.g., by masking regions of an image) to observe how the prediction changes. This identifies which input features are crucial for the model's decision.
- **Surrogate Models:** LIME (Local Interpretable Model-agnostic Explanations and SHAP (Shapley Additive Explanations) are part of Surrogate Models. LIME generates a simple interpretable model (like linear regression) locally around a specific prediction to approximate the complex model's decision boundary and significance scores to each input feature based on its contribution to the model's prediction, inspired by cooperative game theory in SHAP.
- **Counterfactual Explanations:** It delivers alternative techniques that would change the standard's prediction. For instance, in medical diagnostics, "If the tumour size were 1 cm shorter, the forecast would be innocent."
- **Rule Extraction:** Extracts logical rules from the model's forecasts. For instance, it might develop if-then regulations corresponding to the model's decision boundaries.

iii) Model-Specific vs. Model-Agnostic Approaches

These techniques depend on the category of the deep learning methods.

- **Model-Specific Approaches:** Tailored to specific architectures, such as CNNs, techniques like Grad-CAM are explicitly designed for convolutional layers in image processing.
- **Model-Agnostic Approaches:** General techniques that can be applied to any machine learning model, such as LIME and SHAP.

Understanding the decision-making procedures of Deep Neural Networks (DNNs) is intrinsically challenging due to their complexity and opacity, leading to their designation as "black-box" models. The intricate layers of computation within DNNs make it hard to interpret how specific projections are generated directly. Visualization approaches such as heatmaps are commonly employed to highlight the significance of different features and indicate which parts of the input data the model relies on most. Heatmaps provide a visual representation of feature significance by assigning colors to various areas of the input based on their contribution to the model's predictions. For instance, warmer colors may indicate regions of high importance, while cooler colors highlight less relevant areas. Despite their widespread use, the quality of heatmaps must be validated using techniques like perturbation analysis, where input data is systematically altered (e.g., by occlusion or adding noise) to evaluate how changes affect model outputs [24]. In addition to these approaches, improvements in hybrid explainability techniques are emerging,

combining perturbation-based and gradient-based methods to enhance explanations' fidelity and usability. Such techniques are particularly relevant in critical applications like medical imaging, where comprehending the model's rationale is as essential as its accuracy. As the field progresses, developing standardized evaluation metrics for explainability techniques, such as fidelity, stability, and human interpretability, is becoming increasingly necessary. These metrics help ensure that visual explanations are accurate and understandable to end-users, including clinicians and domain experts, thereby fostering trust and adoption of AI systems in high-stakes environments like healthcare [25].

Explaining by example is another important technique to increase the explainability of a model. This approach validates the model's predictions by providing comparable examples or by creating more descriptive elements, such as text. This approach links the model's predictions with additional data or well-known examples. It can help consumers understand the reasoning behind their predictions. Interpretation techniques can be divided into groups according to their nature, scope, and application. Meanwhile, global processes attempt to explain the overall model performance. Local methods focus on defining specific results. Moreover, the techniques can be model-agnostic. This means those techniques work with all model types, particular models, or certain architectures. There are also two types of annotation approaches: Post-hoc. Where the description is developed after the model is created and within which the model can be interpreted from the start [26–36].

Visualization techniques are important for understanding how image-based models make decisions. Improved insights can be obtained by extending techniques such as Individual Conditional Expectation Plans (ICE) and Partial Dependency Plans (PDP), which help to visualize feature importance. It also uses forward feature dependency to identify discriminative image areas that affect predictions. The model predictions provided are interpreted through perturbation-based techniques such as local model-agnostic description (LIME), which approximates the local model around the projections and often agrees with them. Human instinct is very good [37]. Expository AI (XAI) visualization is supported by various tools and packages. For example, the Descriptor visual analysis framework integrated into TensorBoard allows users to understand model limitations and improve performance. XAI technologies have also been created by major cloud platforms. Vertex XAI from Google Cloud explains AutoML and custom-trained models; Amazon SageMaker Clarify generates heatmaps for feature significance based on PDP and SHAP, which aids in understanding the model. Analogously, Microsoft's InterpretML toolbox offers information on both black-box and white-box models. As XAI approaches advance, it is more crucial than ever to assess explanation quality statistically and qualitatively to ensure the explanations satisfy various applications' unique requirements. It has been suggested that quantitative measures, including relevance mask correctness, be used to evaluate the effectiveness of visual explanations in an impartial manner [38]. Last but not least, recent developments such as Grad-CAM, Grad-CAM++, and Integrated Grad-CAM have greatly enhanced confidence and interpretability by emphasizing key areas in input pictures and improving the visual explanation of CNNs. By bridging the gap between human comprehension and black-box models, these techniques increase AI systems' transparency and the trustworthiness of their conclusions.

4. Explainability for Biomedical Imaging

Biomedical imaging comprises various modalities, such as ultrasound and traditional X-rays. Medical picture capture usually requires expensive equipment and might take a long period. The substantial obstacles that researchers in the field confront are partly caused by this high expense and time commitment. However, research into lung illnesses and breast cancer has accelerated due to the increasing availability of datasets for chest radiographs, mammograms, and thoracic computed tomography (CT). The amalgamation of explainable AI (XAI) methodologies with deep neural network (DNN) models has facilitated extensive illness research in diverse specialties, such as ophthalmology and radiology. Interestingly, radiology has become a prominent discipline, indicating the urgent necessity to use XAI in this sector [7]. XAI has enabled the identification of biases and potential errors in DNN predictions, a crucial step in ensuring equitable and reliable healthcare delivery. XAI-driven models, by offering clear and interpretable outputs, are better positioned to comply with these regulations while enhancing the confidence of clinicians and patients in AI-powered diagnostics.

Integrating Explainable AI (XAI) techniques with Deep Neural Network (DNN) models has further revolutionized biomedical imaging, allowing improved research across specialties such as ophthalmology, cardiology, and radiology. In ophthalmology, for instance, XAI-enhanced DNNs have been used to detect diabetic retinopathy from

retinal fundus images, providing high diagnostic accuracy and transparent insights into the features contributing to the diagnosis. Similarly, in cardiology, AI systems are used to analyze echocardiograms and predict cardiovascular risk with interpretable outputs that can aid clinicians in decision-making. The combination of biomedical imaging advancements and XAI methodologies is reshaping the landscape of disease diagnosis and treatment planning. As the availability of imaging datasets continues to grow and explainability techniques evolve, the integration of XAI into medical imaging workflows promises to improve accuracy, trust, and accessibility across a broad range of specialties. XAI Evaluation Techniques are shown in Table 2.

4.1. Conventional imaging/X-rays

A readily available and reasonably priced imaging modality that has been widely used for COVID-19 diagnosis is the conventional chest X-ray (CXR). For example, research used CXR pictures to categorize pediatric pneumonia caused by viruses and bacteria using a customized and fine-tuned VGG16 model. Critical locations within the photographs were highlighted by applying techniques such as Grad-CAM and LIME to visualize the model activations. Using an ensemble of five distinct DNN models, another study produced an area under the curve (AUC) of 0.92 for CXR classification on the XrPP dataset, showing higher performance than transfer learning. To improve interpretability, heatmaps were created by combining the separate heatmaps from each model.

Additionally, to help medical personnel comprehend AI-driven predictions, Bayesian Teaching was used to produce explanations for pneumothorax diagnosis using CXR. This study examined the correlation between AI classifications and human diagnostic predictions using saliency maps and Bayesian approaches with eight radiologists. Another example shows how visual characteristics contribute to nodule malignancy by using capsule vector weights to predict malignancy in lung cancer datasets.

Additionally, cardiac hypertrophy and cardiomegaly have been diagnosed using CXR pictures. Convolutional neural networks (CNN) that use explainable feature maps enhance interpretability. While explainability in algorithms may not always be able to replicate the complex knowledge of human radiologists, XAI is essential for determining the underlying causes of certain AI conclusions improving clinical decision-making.

4.2. Computed tomography (CT)

LIME has shown to be helpful in the field of CT imaging for separating COVID-19 from normal pictures using eight CNN models. Another method employed t-distributed stochastic neighbor embedding (t-SNE) in conjunction with Grad-CAM to exhibit clearly defined clusters for both COVID-19 and non-COVID-19 instances. Furthermore, U-Net models achieved better metrics in accuracy, precision, specificity, and AUC in the classification of COVID-19 CT images than current state-of-the-art techniques like COVID-Net and COVNet. Here, super-pixel contributions were quantitatively

Table 2. XAI evaluation techniques.

Technique	Focus	Description	Ref
Insertion, deletion	Evaluation of interpretability	Involves the insertion and removal of pixels to assess interpretability.	[39]
CLEVR	Quantitative evaluation of ten methods	Employs a Visual Question Answering (VQA) framework to evaluate multiple methods through proposed metrics.	[40]
GWAP	Improve human understanding of AI	Uses game-based methodologies to generate valuable data from gameplay experiences.	[41]
Methodology saliency methods	Determine suitability of methods	Focuses on assessing the effectiveness of different methods for specific tasks.	[42]
Taxonomy	Evaluation of interpretability	Proposes three approaches for evaluating interpretability.	[43]
CMIE	Customizable model evaluation	Utilizes in-model and post-model information to generate multidimensional interpretability evaluations.	[44]
Modality-specific Feature Importance (MSFI) metric	Multi-modal medical imaging	Evaluates 16 post-hoc methods across various imaging modalities.	[38]
Z-inspection	Trustworthy AI	Proposes a general inspection process to evaluate AI outputs.	[39]

estimated using SHAP, and predictions were explained by LIME through picture segmentation.

A complete COVID-19 detection system that integrates segmentation and classification networks was developed to improve diagnosis accuracy. Because of the efficient lung and lobe segmentation provided by this technology, classification results increased by an impressive 6 %. Additionally, employing CT scans, a Joint Classification and Segmentation (JCS) approach showed effectiveness in diagnosing and explaining COVID-19, with activation maps helping prediction interpretability. The analysis was strengthened by the addition of pixel-level annotations to the COVID-CS dataset.

Deep learning frameworks have also been used to forecast the likelihood of cancer in lung nodules on CT scans. Interestingly, research that used a CNN ensemble identified which nodules will probably be identified as cancer within two years with an AUC of 90.29 %. The results are more interpretable since Grad-CAM visualizations were used to clarify model activations associated with input pictures. Radiologists' trust in clinical diagnosis was greatly boosted by using soft activation mapping methods, which improved the localization and categorization of lung nodules within low-dose CT images [27].

4.3. Magnetic Resonance Imaging (MRI)

The widespread neurodegenerative illness known as Alzheimer's disease, which affects millions of people worldwide, has also benefited from improved imaging methods. Researchers combined gene expression and image data for multimodal identification, using several classifiers for gene data and CNN and SpinalNet structures for MRI analysis. The identification of relevant genes to improve explainability was made possible in large part by LIME. The use of Layer-wise Relevance Propagation (LRP) for CNN decision process visualization was largely made possible by the Alzheimer's Disease Neuroimaging Initiative (ADNI) MRI dataset. This concept may be used for comparable conditions and has been useful in clarifying diagnostic predictions linked to Alzheimer's disease.

By examining the interactions between voxels in various brain areas, a novel method was also used to improve clinical accuracy in the early identification of Alzheimer's disease by giving crucial discriminative information. This validation was carried out using a weighted MRI dataset and showed better performance in binary and multiclass classifications when compared to conventional techniques like Regional Mean Volume (RMV) and Hierarchical Feature Fusion (HFF) [28]. White matter lesions have also been found using MRI, and CNNs have produced heatmaps to show the voxel contributions to classification choices. Studies that compared attribution algorithms found that DeepLIFT was a better fit than LRP and that the most meaningful voxel relevance was found around venous arteries. LRP facilitated using 3D CNNs to diagnose multiple sclerosis, illustrating how the framework may increase the openness of decision-making processes.

4.4. Ultrasound imaging

Compared to other modalities, ultrasound imaging has been used less in clinical AI analysis despite its promise. Various factors, including operator dependence and intrinsic acoustic shadows, impede image quality management. Still, there have been some significant breakthroughs. For example, a study that used MRI and ultrasound to classify prostate cancer combined shallow machine-learning techniques with data fusion from many pre-trained deep-learning models. LIME made explainability easier, which showed that the fusion method improved model performance. Ultrasound images have also been used for breast lesion categorization; the BreastMNIST dataset with CycleGAN activation maximization has been used to improve model prediction visualization. Compared to previous techniques, the classifier's accuracy, specificity, and AUC were shown to be enhanced by the qualitative assessment. Furthermore, the susceptibility of a breast cancer ultrasound dataset to adversarial assaults was investigated, and Multi-Task Learning (MTL) was used to improve classification accuracies [29–36].

In response to the high false positive rates in breast cancer diagnostics, the DREAM challenge was launched, and CNN architectures were used to produce the winning solutions. Moreover, multimodal multiview ultrasound pictures have been used to large-scale datasets to assess breast cancer risk. Grad-CAM was used to create heatmaps, which effectively directed expert assessments. The suggested SONO design used a timeline resembling a barcode to improve interpretability and make it easier to identify underlying substructures in foetal ultrasound video analysis. Furthermore, despite significant feature discrepancies across different architectures, datasets such as EyePACS and DIARETDB1 showed consistent classification performance across multiple CNN models [37].

4.5. Cancer detection

Recent advancements in cancer diagnosis, especially for deep neural network applications,

including MRI, demonstrate the promise of explainable AI techniques. Early diagnosis is crucial for prostate cancer, and multi-modal fusion methodologies that include deep learning models that have been trained from both ultrasound and MRI modalities have produced encouraging results in terms of classification accuracy. LIME made it easier to identify key characteristics that distinguish benign from malignant tumors.

Using MRI images with CNN models for lesion classification Researchers have used the influence function to evaluate and validate the importance of specific imaging features in diagnosing liver cancer and AI interpretation methods are useful for interpreting histopathology images. Using the Patch Camelyon (P-CAM) dataset, the CNN model was able to accurately detect lymph node metastasis. The generated LIME descriptions showed good agreement with clinician opinions. This indicates that the CNN model successfully learned to associate tumors' presence with specific visual elements [38–43].

As the field of medical imaging grows Integrating Interpretable AI approaches with various methods is therefore critical to increasing diagnostic accuracy and clinical confidence in AI-driven systems. These approaches have the potential to transform care. Health through ongoing research and development. It enables faster identification and more personalized treatment for patients.

5. XAI in biomedical imaging: applications and challenges

In this section, we discuss theoretically the applications and challenges of XAI in biomedical imaging. XAI in Biomedical Imaging: Applications and Challenges is shown in Fig. 3. Overview of the Feasibility and Requirements for Each Section is described in Table 3.

5.1. Applications of XAI in medical imaging

Medical imaging consists of various techniques which are important for diagnosis. Grad-CAM is the most widely used XAI approach, followed by LRP, LIME, and SHAP. These methods are common in the field of ophthalmology. Respiratory science and neurology. Adding additional information to the model during training can result in better interpretation and more accurate predictions. Better diagnostic results may also be achieved by applying

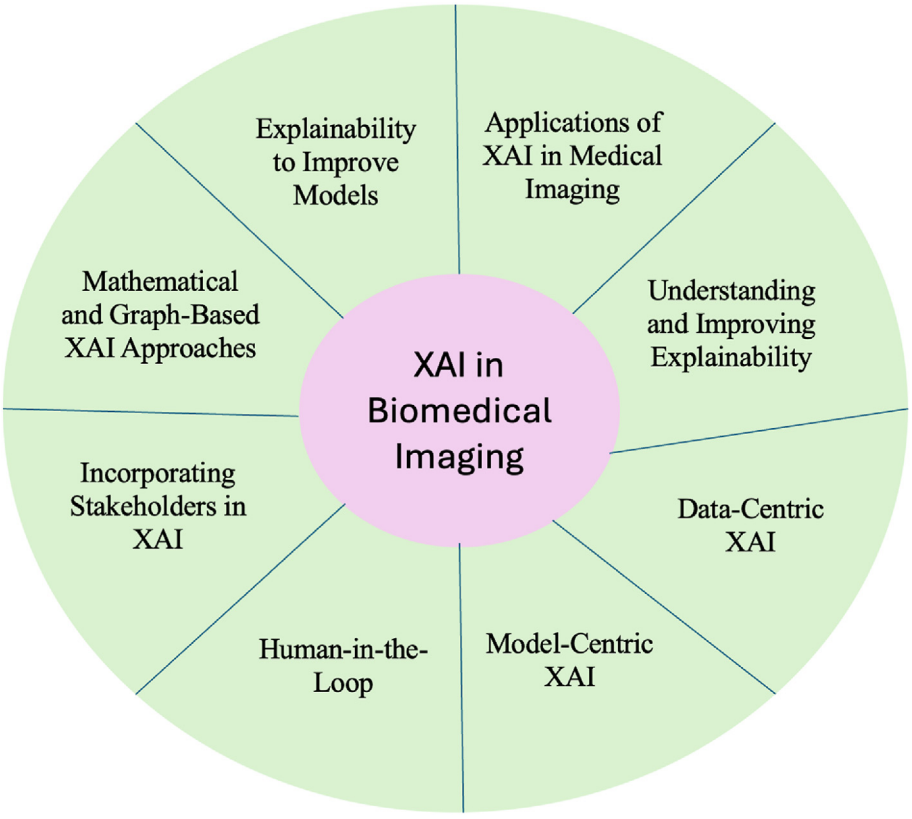

Fig. 3. XAI in biomedical imaging: Applications and challenges.

Table 3. Overview of the feasibility and requirements for each section.

Section	Standardization Feasible?	Data-Dependent?	User-Involvement Required?	Applicable in Clinical Settings?	Improves Model Trust?
Applications of XAI in medical imaging	No	Yes	Yes	Yes	Yes
Understanding and improving explainability	No	No	Yes	Yes	Yes
Data-Centric XAI	Yes	Yes	No	Yes	Yes
Model-Centric XAI	Yes	Yes	No	Yes	Yes
Human-in-the-Loop	No	No	Yes	Yes	Yes
Incorporating stakeholders in XAI	No	No	Yes	Yes	Yes
Mathematical and graph-based XAI approaches	No	No	No	Yes	Yes
Explainability to improve models	Yes	Yes	No	Yes	Yes

multimodal Deep Neural Networks (DNNs), which enable the identification of hidden patterns in segmentation and classification tasks. The usage of mobile applications for mass screening, such as glaucoma diagnosis, that leverage DNN segmentation and classification is one example of this, which improves accessibility to healthcare.

5.2. Understanding and improving explainability

Since there isn't a common definition for the word "explanation" in the context of XAI, attempts are being made to standardize its meaning. While much current research mostly uses post-hoc explainability methodologies, using XAI in clinical settings may be improved by developing intrinsic explanation methods. X-Caps is one technique that aims to increase explainability by using high-level visual features to produce ratings for malignancies used by medical professionals. Furthermore, XNN and scalable Bayesian rule lists are examples of white-box models that provide interpretable insights with less accuracy trade-offs, making them more useful in healthcare applications [11].

5.3. Data-centric XAI

A model's performance and ensuing explainability largely depend on the caliber and amount of training data. To prevent overfitting, DNNs need large, high-quality datasets; this is especially important in medical imaging, where it might be challenging to gather samples of sick cases. Data quality may be greatly improved by applying several denoising techniques and picture pre-processing approaches, improving model predictions, and decreasing the explainability gap. Improved DNN resilience against noise and irrelevant input can increase the overall efficacy of medical picture analysis.

5.4. Model-centric XAI

Diverse data types may be used in DNN models to reduce bias and enhance generalisability. Sadly, selective bias results from many models being trained on data from a particular facility or unit. Biases in the training data and the models themselves may be revealed using counterfactual algorithms. Furthermore, visuals offered by XAI approaches can aid skilled doctors in spotting any incorrect or erroneous conclusions made by the DNN models. Adversarial attack vulnerabilities must also be addressed because altered pictures have the potential to confuse models and require strong detection methods to guarantee correct interpretations.

5.5. Human-in-the-Loop

DNN models can only partially replace human knowledge in medical diagnosis. While it is possible to create predictions on par with human experts, DNN and XAI methods should help medical experts get better ideas. The interaction between the doctor and the model is facilitated by utilizing "human-in-the-loop," which allows for adjustments that enable advanced predictions and appropriate interpretations. Methods such as Bayesian Teaching can help medical professionals analyze AI-driven diagnoses. To an eventual diagnosis. Their process has a better understanding of the matter.

5.6. Incorporating stakeholders in XAI

The current use of DNN models in clinical practice may need to be improved by a continued emphasis on the needs of system developers over end users. Therefore, XAI systems must be designed with their needs in mind. Of the user, especially doctors and patients. It is considered to be of utmost

importance. Building acceptance and trust for AI in healthcare requires a multidisciplinary strategy considering social, legal, and ethical issues. Working with physicians throughout the design and evaluation process helps ensure that the solutions that build It is consistent with what users consider reliable and understandable [42].

5.7. Mathematical and graph-based XAI approaches

Although mathematical methods can improve the interpretability of models, they don't always provide real-world examples that end users can relate to. Graph Neural Networks (GNN) can be used to show complex connections in data, capturing many types of interactions. Instead of relying solely on an ordered network structure, Studying complex diseases such as cancer in which various rays interact throughout the diagnosis and treatment process. Where this expertise is very useful, The improved model is evaluated on synthetic and real-world datasets. To prove how well they understand interactions based on graphs [43].

5.8. Explainability to improve models

Explainability may act as a catalyst for model development, starting a feedback loop where improved comprehension motivates better model construction, which stimulates the development of more potent explainability strategies. For example, Layer-wise Relevance Propagation (LRP) model pruning can assist in locating important filters and weights inside a model. The network topology of CNN layers may be visualized using techniques like deconvolution, Grad-CAM, and Guided Grad-CAM. This helps with hyperparameter optimization and overall performance improvement. Explainability enhances classification accuracy and furthers continuous model development by tying saliency maps to the data supporting model predictions [44].

5.9. Comparison of deep learning and explainable AI (XAI) based on biomedical imaging applications

By concentrating on the same issues of accountability and transparency, it will be easier to see how deep learning in medical imaging and liability problems in autonomous vehicle accidents are comparable. To make crucial judgments, like navigating highways in autonomous vehicles or identifying illnesses through medical imaging, both domains mostly rely on sophisticated deep learning models, sometimes called “black-box models.” When mistakes are made, though, the opaqueness of the decision-making process raises serious questions. Deep learning is used by these AI algorithms in autonomous cars to make snap judgments regarding braking, speed, and obstacle avoidance. Determining who is at fault is the main problem when an accident occurs due to a system malfunction, such as the AI failing to recognize a pedestrian. Who is it—the owner, the software developer, or the automaker? Liability is made more difficult because deep learning models don't always explain their choices, making it difficult to pinpoint the failure's primary cause.

In medical imaging, deep learning models are also used to diagnose conditions, such as detecting tumors or fractures in images. Diagnosis errors, such as false negatives (missing a hazardous illness) or false positives (incorrectly classifying a benign problem as cancer), can have serious health repercussions for patients [45,46]. The same dilemma arises: who is to blame for the mistake—the hospital that used the technology, the clinician who depended on the AI, or the AI system's creators? It is challenging to comprehend how the AI arrived at its conclusion in both domains due to the opacity of deep learning models. This black-box characteristic, or lack of interpretability, makes accountability more difficult. A single error, such as an accident or a misdiagnosis, can have potentially fatal outcomes in medical imaging and driverless automobiles. The fundamental problem remains the same: it is challenging to completely trust AI systems as they are not yet transparent enough to communicate their decision-making procedures in a way that is intelligible to humans [47].

Explainable AI (XAI) can help with this. XAI approaches seek to shed light on how AI models make judgments to assist stakeholders and doctors in comprehending the rationale behind forecasts. XAI can contribute to greater confidence in AI systems in both domains by enhancing the interpretability of deep learning models. In medical imaging, for instance, XAI technologies can assist physicians in comprehending the reasons for an AI's flagging of a certain area as problematic, enabling them to make better judgments. Similarly, XAI may help explain why an AI made a certain braking or route choice in autonomous driving, making it more straightforward to assign blame during a failure [48,49].

6. Conclusion and future scope

The importance of regulatory frameworks in the development and uptake of Explainable Artificial

Intelligence (XAI) in biomedical imaging is highlighted in this paper. Transparency, interpretability, and dependability in AI models are becoming increasingly important as AI is further incorporated into healthcare, especially for tasks like picture segmentation and classification. Because they can establish centralized criteria for creating and comparing XAI algorithms, regulatory organizations play a crucial role in directing this process. Before being used in clinical settings, these guidelines will guarantee that all models satisfy uniform quality and safety requirements. Additionally, XAI provides a mechanism for AI model creators to respond to the healthcare industry's growing ethical and regulatory requirements. Developers may make it simpler to comply with health standards and obtain permission for clinical usage by integrating explainability into AI systems, which gives developers clear insights into how choices are made. In addition to fostering trust among healthcare providers, this openness will guarantee that AI models comply with patient safety regulations, lowering the possibility of incorrect diagnoses or forecasts.

Furthermore, establishing biological reference datasets that are openly accessible and backed by regulatory organizations would improve developers' capacity to assess and improve their XAI models. These datasets can serve as a benchmark for evaluating the explainability and performance of AI systems, guaranteeing steady advancement. Another potential direction for the future is federated learning, which allows hospitals to provide useful data for model training while protecting patient privacy. Furthermore, by using automated, explainable models to streamline decision-making, XAI integration into clinical workflows might lessen the workload for medical professionals like radiologists and ophthalmologists. These technologies guarantee that physicians maintain oversight while gaining insights from AI, improving the decision-making process's efficiency by keeping a human-in-the-loop approach.

The role of XAI in guaranteeing compliance will be crucial as healthcare rules continue to change. AI systems utilized in healthcare contexts are increasingly required by legal and ethical standards to be efficient, open, responsible, and transparent. By offering interpretable models that can be evaluated, validated, and trusted by regulatory agencies and medical practitioners, XAI is uniquely positioned to meet these objectives. XAI can promote AI's safe and efficient integration into healthcare by upholding these regulatory frameworks and guaranteeing compliance, eventually improving patient outcomes and boosting trust in AI-based medical devices. In the future, we will utilize an advanced version algorithm of XAI and Deep Learning for Biomedical Imaging Applications to mitigate essential issues and challenges such as segmentation, dataset size, speed, cost, security, and privacy.

Ethical statement

All the authors declare that they have no competing interests and no objections to publish this manuscript.

Author contributions

Writing—review & editing, Sushil Kumar Singh; Writing—original draft, Sushil Kumar Singh; Methodology, Sushil Kumar Singh, Saurabh Agarwal; Implementation, Abhilash Maroju; Validation, Saurabh Agarwal; Resources, Saurabh Agarwal, Sushil Kumar Singh; Visualization, Abhilash Maroju, Saurabh Agarwal; Formal analysis, Sushil Kumar Singh; Supervision, Bal Virdee; Project Administration, Sushil Kumar Singh and Saurabh Agarwal; Funding acquisition, Bal Virdee.

AI usage declaration

AI usage declaration in this scientific work, generative artificial intelligence (AI) was not used.

Acknowledgement and Funding

This research was collaborated by Marwadi University, Rajkot, Gujarat, India, London Metropolitan University, London, UK, San Jose State University, USA, and University of the Cumberlands, USA.

Conflict of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Oksuz I. Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation. IEEE Trans Med Imag 2020;39(12):4001–10. https://doi.org/10.1109/TMI.2020.3008930.

[2] Leynes AP, Deveshwar N, Nagarajan SS, Larson PEZ. Scan-specific self-supervised bayesian deep non-linear inversion for undersampled MRI reconstruction. IEEE Trans Med Imag 2024;43(6):2358–69. https://doi.org/10.1109/TMI.2024.3364911.

[3] Huang W. Arterial spin labeling images synthesis from sMRI using unbalanced deep discriminant learning. IEEE Trans Med Imag 2019;38(10):2338–51. https://doi.org/10.1109/TMI.2019.2906677.

[4] Ekanayake Z, Harandi M, Egan G, Chen Z. Improving deep learning MRI reconstruction with contrastive learning

pretraining. In: 2024 IEEE international symposium on biomedical imaging (ISBI); 2024. p. 1–4. https://doi.org/10.1109/ISBI56570.2024.10635794.
[5] Wu G, Kim M, Wang Q, Munsell BC, Shen D. Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans Biomed Eng 2016;63(7):1505–16. https://doi.org/10.1109/TBME.2015.2496253.
[6] Yi Z. Fast and calibrationless low-rank parallel imaging reconstruction through unrolled deep learning estimation of multi-channel spatial support maps. IEEE Trans Med Imag 2023;42(6):1644–55. https://doi.org/10.1109/TMI.2023.3234968.
[7] Chen Z, Chen Y, Xie Y, Li D, Christodoulou AG. Data-consistent non-cartesian deep subspace learning for efficient dynamic MR image reconstruction. In: IEEE 19th international symposium on biomedical imaging (ISBI); 2022. p. 1–5. https://doi.org/10.1109/ISBI52829.2022.9761497.
[8] Guan Y. Subspace model-assisted deep learning for improved image reconstruction. IEEE Trans Med Imag 2023;42(12):3833–46. https://doi.org/10.1109/TMI.2023.3313421.
[9] Lieffrig EV. Multi-task deep learning and uncertainty estimation for pet head motion correction. In: IEEE 20th international symposium on biomedical imaging (ISBI); 2023. p. 1–5. https://doi.org/10.1109/ISBI53787.2023.10230791.
[10] DiSpirito A. Reconstructing undersampled photoacoustic microscopy images using deep learning. IEEE Trans Med Imag 2021;40(2):562–70. https://doi.org/10.1109/TMI.2020.3031541.
[11] Gunturu V, Maiti N, Toure B, Kunekar P, Banu SB, Sahaya Lenin D. Transfer learning in biomedical image classification. In: International conference on advances in computing, communication and applied informatics (ACCAI); 2024. p. 1–5. https://doi.org/10.1109/ACCAI61061.2024.10601862.
[12] Liang P, Chen J, Zheng H, Yang L, Zhang Y, Chen DZ. Cascade decoder: a universal decoding method for biomedical image segmentation. In: IEEE 16th international symposium on biomedical imaging (ISBI 2019); 2019. p. 339–42. https://doi.org/10.1109/ISBI.2019.8759430.
[13] Thimmiaraja GG,J, Shelke CJ, Pavithra G, Sharma VK, Verma D. Deep learning with unsupervised and supervised approaches in medical image analysis. In: 2nd international conference on advance computing and innovative technologies in engineering (ICACITE); 2022. p. 1580–4. https://doi.org/10.1109/ICACITE53722.2022.9823491.
[14] Jung H. Integration of deep learning and graph theory for analyzing histopathology whole-slide images. In: 2018 IEEE applied imagery pattern recognition workshop (AIPR); 2018. p. 1–5. https://doi.org/10.1109/AIPR.2018.8707424.
[15] Obayya M, Saeed MK, Alruwais N, Alotaibi SS, Assiri M, Salama AS. Hybrid metaheuristics with deep learning-based fusion model for biomedical image analysis. IEEE Access 2023;11:117149–58. https://doi.org/10.1109/ACCESS.2023.3326369.
[16] Brdnik S, Šumak B. Current trends, challenges and techniques in XAI field; A tertiary study of XAI research. In: 47th MIPRO ICT and electronics convention (MIPRO); 2024. p. 2032–8. https://doi.org/10.1109/MIPRO60963.2024.10569528.
[17] Venugopal A, Farnaghi M, Zurita-Milla R. Comparative evaluation of XAI methods for transparent crop yield estimation using CNN. In: IGARSS 2024 - 2024 IEEE international geoscience and remote sensing symposium, Athens, Greece; 2024. p. 7478–82. https://doi.org/10.1109/IGARSS53475.024.10641426.
[18] Bargagna F, De Santi L, Santarelli M, Positano V, Vanello N. Bayesian XAI methods towards a robustness-centric approach to deep learning: an ABIDE I study. In: IEEE international symposium on medical measurements and applications (MeMeA), Eindhoven, Netherlands; 2024. p. 1–5. https://doi.org/10.1109/MeMeA60663.2024.10596826.
[19] Liang C, Wang M, Liu H, Chong KK. Exploring meibomian gland dysfunction grading: a comparison of machine learning and XAI approaches. In: 2024 IEEE 2nd international conference on image processing and computer applications (ICIPCA), Shenyang, China; 2024. p. 2021–4. https://doi.org/10.1109/ICIPCA61593.2024.10709187.
[20] Arnaud E, Elbattah M, Pitteman A, Dequen G, Ghazali D, Moreno-Sánchez P. Consistency of XAI models against medical expertise: an assessment protocol. In: 2024 IEEE 12th international conference on healthcare informatics (ICHI), Orlando, FL, USA; 2024. p. 732–6. https://doi.org/10.1109/ICHI61247.2024.00116.
[21] Kumar A, Singh U, Pradhan B. Enhancing interpretability in deep learning-based inversion of 2-D ground penetrating radar data: an explainable AI (XAI) strategy. Geosci Rem Sens Lett IEEE 2024;21:1–5. https://doi.org/10.1109/LGRS.2024.3400934.
[22] Gehlot N, Malik S, Jena A, Vijayvargiya A, Kumar R. XAI-driven sEMG feature analysis for hand gestures. In: 2024 third international conference on power, control and computing technologies (ICPC2T), Raipur, India; 2024. p. 19–24. https://doi.org/10.1109/ICPC2T60072.2024.10475061.
[23] Sarkar MS, Lifat MA, Hasan R, Hassan MM. Discovering the depths of cotton leaf disease detection: integrating hypertuned residual networks with GradCAM XAI for enhanced understanding and diagnosis. In: 2024 3rd international conference on advancement in electrical and electronic engineering (ICAEEE), Gazipur, Bangladesh; 2024. p. 1–6. https://doi.org/10.1109/ICAEEE62219.2024.10561737.
[24] Thiruvenkadam K, Ravindran V, Thiyagarajan A. Deep learning with XAI based multi-modal MRI brain tumor image analysis using image fusion techniques. In: 2024 international conference on trends in quantum computing and emerging business technologies, Pune, India; 2024. p. 1–5. https://doi.org/10.1109/TQCEBT59414.2024.10545215.
[25] Gizzini AK, Medjahdi Y, Mabrouk MB. GRACE: gradient-based XAI scheme for channel estimation in wireless communications. In: 2024 IEEE international mediterranean conference on communications and networking (MediTCom), Madrid, Spain; 2024. p. 572–7. https://doi.org/10.1109/MeditCom61057.2024.10621232.
[26] Bourokba A, El Hamdi R, Njah M. A Shapley based XAI approach for a turbofan RUL estimation. In: 2024 21st international multi-conference on systems, signals & devices (SSD), Erbil, Iraq; 2024. p. 832–7. https://doi.org/10.1109/SSD61670.2024.10548499.
[27] Rui L, Gadyatskaya O. Position: the explainability paradox - challenges for XAI in malware detection and analysis. In: 2024 IEEE European symposium on security and privacy workshops (EuroS&PW), Vienna, Austria; 2024. p. 554–61. https://doi.org/10.1109/EuroSPW61312.2024.00067.
[28] Patil P, Pamali SK, Devagiri SB, Sushma AS, Mirje J. Plant Leaf disease detection using XAI. In: 2024 3rd international conference on artificial intelligence for internet of things (AIIoT), Vellore, India; 2024. p. 1–6. https://doi.org/10.1109/AIIoT58432.2024.10574617.
[29] Wang C. An interpretable and accurate deep-learning diagnosis framework modeled with fully and semi-supervised reciprocal learning. IEEE Trans Med Imag 2024;43(1):392–404. https://doi.org/10.1109/TMI.2023.3306781.
[30] Mp G, Vs B, H A. Advanced sound detection and behavior examination for real-time intruder detection using deep learning: a comprehensive security framework. In: 2nd international conference on artificial intelligence and machine learning applications theme: healthcare and internet of things (AIMLA), Namakkal, India; 2024. p. 1–6. https://doi.org/10.1109/AIMLA59606.2024.10531591.
[31] Degadwala S, Krishnamurthy VND, Vyas D. DeepSpine: multi-class spine X-ray conditions classification using deep learning. In: 3rd international conference on sentiment analysis and deep learning (ICSADL), Bhimdatta, Nepal; 2024. p. 8–13. https://doi.org/10.1109/ICSADL61749.2024.00008.

[32] Wang X. Deep reinforcement learning: a survey. IEEE Trans Neural Networks Learn Syst 2024;35(4):5064–78. https://doi.org/10.1109/TNNLS.2022.3207346.
[33] Kartheeban K. Real time poisoning attacks and privacy strategies on machine learning systems. In: 2024 3rd international conference on sentiment analysis and deep learning (ICSADL), Bhimdatta, Nepal; 2024. p. 183–9. https://doi.org/10.1109/ICSADL61749.2024.00036.
[34] Jaffer Sumia H, Ghaeb Nebras H. Important features of EMG signal under simple load conditions. Polytechnic J 2017;7(1):2. https://doi.org/10.59341/2707-7799.1749.
[35] Khidir Hiwa S, Dizayee Saud J, Ali Sangar H. Prevalence of root canal configuration of mandibular second molar using cone-beam computed tomography in a sample of Iraqi patients. Polytechnic J 2021;11(1):5. https://doi.org/10.25156/ptj.v11n1y2021.pp22-26.
[36] Gumma YR, Peram S. Review of cybercrime detection approaches using machine learning and deep learning techniques. In: 2024 3rd international conference on applied artificial intelligence and computing (ICAAIC), Salem, India; 2024. p. 772–9. https://doi.org/10.1109/ICAAIC60222.2024.10575058.
[37] Li Y. Progress in the application of deep learning in natural language processing and its impact on English teaching translation software system. In: 3rd international conference on artificial intelligence and autonomous robot systems (AIARS), Bristol, United Kingdom; 2024. p. 134–7. https://doi.org/10.1109/AIARS63200.2024.00030.
[38] Tiwari RG, Kumar A. Bean leaf lesions image classification: a robust ensemble deep learning approach. In: 2024 ASU international conference in emerging technologies for sustainability and intelligent systems (ICETSIS), Manama, Bahrain; 2024. p. 986–93. https://doi.org/10.1109/ICETSIS61505.2024.10459697.
[39] Zhong Y. Unsupervised fusion of misaligned PAT and MRI images via mutually reinforcing cross-modality image generation and registration. IEEE Trans Med Imaging 2024;43(5):1702–14. https://doi.org/10.1109/TMI.2023.3347511.
[40] Fu M. OIF-net: an optical flow registration-based PET/MR cross-modal interactive fusion network for low-count brain PET image denoising. IEEE Trans Med Imaging 2024;43(4):1554–67. https://doi.org/10.1109/TMI.2023.3342809.
[41] Lin L. Pediatric TSC-related epilepsy classification from clinical MR images using quantum neural network. In: IEEE international symposium on biomedical imaging (ISBI), Athens, Greece; 2024. p. 1–5. https://doi.org/10.1109/ISBI56570.2024.10635849.
[42] Li Z. Medical multimodal image transformation with modality code awareness. IEEE Trans Radiat Plasma Med Sci 2024;8(5):511–20. https://doi.org/10.1109/TRPMS.2024.3379580.
[43] Alenezi AM, Aloqalaa DA, Singh SK, Alrabiah R, Habib S, Islam M, et al. Multiscale attention-over-attention network for retinal disease recognition in OCT radiology images. Front Med 2024;11:1499393.
[44] Su T. Super resolution dual-energy cone-beam CT imaging with dual-layer flat-panel detector. IEEE Trans Med Imag 2024;43(2):734–44. https://doi.org/10.1109/TMI.2023.3319668.
[45] Singh SK, Kumar M, Khanna A, Virdee B. Blockchain and FL-based secure architecture for enhanced external Intrusion detection in smart farming. IEEE Internet Things J 2024. https://doi.org/10.1109/JIOT.2024.3478820.
[46] Singhal S, Betgeri S, Singh SK. Strategies for mitigating security concerns in IoT-enabled smart cities. In: Secure and intelligent IoT-enabled smart cities. IGI Global; 2024. p. 239–73.
[47] Zhang S, Liu J, Zheng Ning G, Zhao Z, Liao H. Advancing optical chromoendoscopy: augmented pseudo-color fusion leveraging feature disparities for superior image contrast. In: 2024 IEEE international symposium on biomedical imaging (ISBI), athens, Greece; 2024. p. 1–5. https://doi.org/10.1109/ISBI56570.2024.10635798.
[48] Singh, S. K., Chauhan, S., Alsafrani, A., Islam, M., Sherazi, H. I., & Ullah, I. Optimizing healthcare data quality with optimal features driven mutual entropy gain. Expet Syst, e13737.
[49] Singh SK, Lee C, Park JH. CoVAC: a P2P smart contract-based intelligent smart city architecture for vaccine manufacturing. Comput Ind Eng 2022;166:107967.