PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 2, **Number** 2, July 2025
https://doi.org/10.62671/perfect.v2i2.115

Explainable AI for Medical Imaging: A Taxonomy Based on Clinical Task Requirements

Ali Nadhim Kamber[1], Hussein Alkaabi [2*], Mohammed Al-Rekabi[3], Ali Kadhim Jasim[4]
[1,2]Ministry of Education Iraq, General Direction Of Vocational Education, Al-Najaf, Iraq
[3]Department of Computer Engineering, College of Engineering, University of Al-Shatra, Iraq
[4]Imam Ja'far al Sadiq University – Maysan Branch, Computer Engineering Department, Iraq
[1]ali.nice1986@gmail.com, [2]hussain.njf7@gmail.com, [3]moh.alrekabee85@gmail.com, [4]ali.kadhim89@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 11-07-2025
Accepted: 11-08-2025
PUBLISHED: 14-08-2025

Keywords:
Explainable AI (XAI); Medical Imaging; Clinical Decision Support; Taxonomy; Interpretability.

ABSTRACT

Explainable Artificial Intelligence (XAI) has emerged as a critical enabler for deploying AI-driven medical imaging systems where transparency, trust, and accountability are paramount. However, most current taxonomies of XAI methods categorize techniques based on algorithmic families (e.g., saliency maps, attribution methods), which often fail to reflect the practical requirements of clinical tasks. This paper proposes a novel task-centric taxonomy of XAI in medical imaging that aligns explanation techniques with four key clinical tasks: classification, detection, segmentation, and prognostic assessment. For each task, we analyze how different XAI methods enhance model interpretability, their suitability for clinical decision-making, and their limitations in real-world applications. Our taxonomy aims to provide a practical framework for researchers and practitioners to select appropriate XAI strategies tailored to the specific demands of medical imaging workflows. Furthermore, we highlight the current gaps in task-specific explainability and propose future research directions towards clinically meaningful, task-driven XAI solutions. This work serves as a step towards bridging the gap between technical XAI developments and the functional needs of clinical practice.

INTRODUCTION

Medical imaging has been a cornerstone of modern healthcare for over a century, evolving from the discovery of X-rays in 1895 to the development of advanced modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Ultrasound, and Positron Emission Tomography (PET) [1]. These imaging techniques have revolutionized the diagnosis and management of various diseases, providing clinicians with non-invasive tools for visualizing anatomical structures and pathological conditions with high precision. Despite the significant technological advancements, the interpretation of medical images remains a complex task that requires substantial expertise and is prone to human variability and cognitive biases [2]. Artificial Intelligence (AI), particularly deep learning, has demonstrated remarkable potential in automating and augmenting many applications in recent years, such as healthcare and Image-Based Recognition [3]. AI models have achieved expert-level performance in tasks such as tumor detection, organ segmentation, and disease classification, offering opportunities to enhance diagnostic accuracy and reduce clinician workload [4][5]. However, the "black-box" nature of many AI systems poses significant challenges for clinical adoption, as the lack of transparency and interpretability undermines trust and raises concerns about accountability and ethical decision-making [6]. Explainable Artificial Intelligence (XAI) has emerged as a critical field to address these challenges by developing methods that make AI decision-making processes understandable to human users [7]. In the context of medical imaging, XAI techniques provide visual or textual explanations that highlight the factors influencing a model's predictions, enabling clinicians to verify, contest, or trust the outputs of AI systems [8]. Techniques such as saliency maps, Class Activation Maps (CAM), SHapley Additive exPlanations (SHAP), and Local Interpretable Model-agnostic Explanations (LIME) have been widely explored to enhance the transparency of AI-driven imaging applications [9]. Despite the growing body of research on XAI methods, existing taxonomies predominantly classify these techniques based on algorithmic families or computational principles, with limited consideration of the specific clinical tasks intended to support. This approach overlooks the nuanced requirements of different medical imaging tasks, where the nature of explanations needed for classification differs from those required for segmentation or detection. This paper proposes a task-centric taxonomy of XAI techniques in medical imaging, categorizing methods based on their applicability to key clinical tasks: classification, detection, segmentation, and prognostic assessment. Our objective is to provide a practical framework that guides researchers and healthcare practitioners in selecting suitable XAI strategies tailored to task-specific demands. The paper reviews the strengths and limitations of existing methods

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 2, **Number** 2, July 2025
https://doi.org/10.62671/perfect.v2i2.115

within each task category, identifies current gaps in task-aligned explainability, and outlines future research directions towards more clinically meaningful XAI solutions. The remainder of this paper is organized as follows: Section 2 provides an overview of background concepts and related work in XAI for medical imaging. Section 3 introduces the proposed task-centric taxonomy, detailing how explanation methods align with specific imaging tasks. Section 4 discusses the practical implications, challenges, and emerging trends in task-specific Explainability. Finally, Section 5 concludes the paper and outlines potential directions for future research.

LITERATURE REVIEW

Background and Related Work

This section will briefly present the topic's background and review some survey papers in this field.

Medical Imaging Modalities: An Overview

Medical imaging has evolved into diverse modalities, each offering unique advantages for visualizing anatomical structures and pathological conditions. X-ray imaging remains the most widely used modality due to its simplicity and effectiveness in visualizing dense tissues like bones. CT provides higher-resolution cross-sectional images, facilitating the detection of tumors, fractures, and vascular anomalies. MRI excels in soft tissue characterization, making it invaluable in neuroimaging and musculoskeletal diagnostics. Ultrasound imaging, on the other hand, offers real-time imaging capabilities, which are crucial in obstetrics, cardiology, and point-of-care diagnostics. More advanced techniques, such as PET, enable functional imaging by tracing metabolic processes, thereby supporting early cancer detection and monitoring disease progression [10][11].

The Role of AI in Medical Image Analysis

The integration of AI, particularly deep learning, has significantly advanced the field of medical image analysis. Convolutional Neural Networks (CNN) have demonstrated exceptional performance in tasks such as image classification, object detection, and segmentation across various medical imaging datasets [12]. These AI models can learn complex patterns from large-scale annotated data, enabling automated detection of pathologies like tumors, lesions, and anomalies with accuracy comparable to expert radiologists [13]. Furthermore, AI has shown promise in predictive analytics, where imaging data is leveraged to forecast disease progression and patient outcomes. However, the adoption of AI in clinical settings remains cautious, primarily due to the opacity of deep learning models and the absence of transparent decision-making processes [8].

Explainable AI (XAI) Techniques in Medical Imaging

XAI aims to bridge the gap between complex AI models and human interpretability by providing insights into the reasoning behind model predictions. In medical imaging, XAI methods typically generate visual explanations highlighting image regions influencing the model's decision. Gradient-based techniques, such as Saliency Maps and Gradient-weighted Class Activation Mapping (Grad-CAM), have been extensively used to visualize feature importance in CNN models [14]. Model-agnostic methods like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) offer post-hoc explanations by approximating the contribution of individual features to the model's output [15]. Additionally, attention-based mechanisms have been integrated into neural network architectures to provide intrinsic interpretability by focusing on salient regions during inference [16]. Despite their potential, these XAI methods often have limitations when applied to clinical workflows. For instance, saliency maps may produce noisy or clinically irrelevant heatmaps, while model-agnostic methods can be computationally intensive and sensitive to perturbations. Moreover, the suitability of an XAI method is highly dependent on the nature of the imaging task, as the explanatory needs for a segmentation task are fundamentally different from those of a classification or detection task.

METHOD

Task-Centric Taxonomy of XAI in Medical Imaging

This section introduces a task-centric taxonomy that categorizes XAI techniques based on their applicability to four key clinical tasks in medical imaging: classification, detection, segmentation, and prognostic assessment. Unlike traditional taxonomies that group XAI methods by algorithmic principles, our approach emphasizes the functional alignment between explanation techniques and the specific demands of clinical workflows (Figure 1).

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 2, **Number** 2, July 2025
https://doi.org/10.62671/perfect.v2i2.115

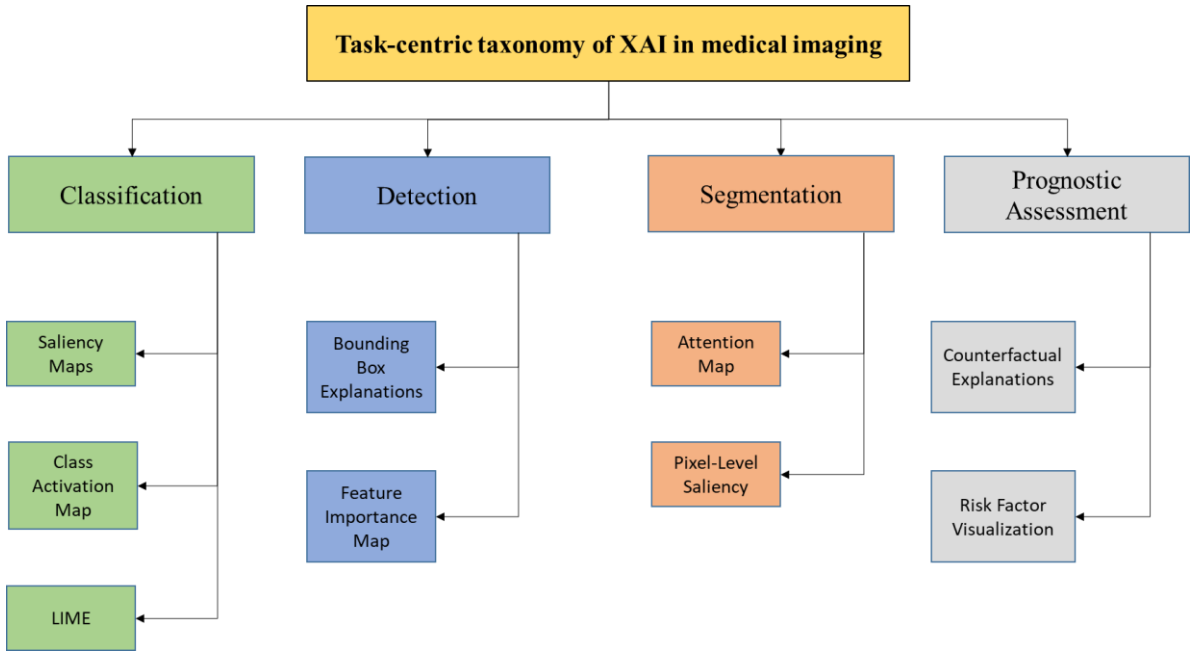

Figure 1. Task-centric taxonomy of explainable AI (XAI) methods in medical imaging, illustrating the alignment between clinical tasks and explanation modalities.

Classification Tasks

Classification tasks in medical imaging involve assigning a diagnostic label to an entire image or region of interest, such as distinguishing between benign and malignant tumors. In this context, XAI methods highlight image regions that contribute most significantly to the classification decision, thereby enabling clinicians to assess the rationale behind AI predictions. Saliency Maps are widely used for visualizing pixel-level importance scores by computing the gradient of the output with respect to input features [17]. However, saliency maps often produce diffuse and noisy heatmaps that lack clinical precision. Class Activation Maps (CAM) and their variants (e.g., Grad-CAM) improve upon this by localizing discriminative regions associated with specific classes through the use of global average pooling layers in CNNs [18]. Although initially designed for tabular data, LIME has been adapted to medical imaging by perturbing superpixels and approximating the model locally with an interpretable surrogate model [19]. While these methods enhance transparency, their explanations are inherently qualitative and may not always align with clinical reasoning, especially in complex multi-label classification tasks. Moreover, the resolution of visual explanations is constrained by the network's architecture, often limiting interpretability at fine-grained anatomical levels.

Detection Tasks

Detection tasks aim to localize and identify specific pathological entities within an image, such as nodules, lesions, or fractures. Unlike classification, detection requires explanations that justify the detected objects' presence and spatial location. Bounding Box Explanations are frequently used in object detection models (e.g., Faster R-CNN, YOLO) to highlight regions where anomalies are identified [20]. To augment interpretability, feature attribution methods can be overlaid within these bounding boxes to indicate which features contributed most to the detection outcome. Feature Importance Maps, derived from methods like SHAP, offer quantitative assessments of pixel or region-wise contributions, providing a more nuanced understanding of the model's focus [21]. One challenge in detection explainability is ensuring that explanations are precise (spatially accurate) and clinically meaningful. Saliency-based methods may highlight large regions that lack diagnostic relevance, while fine-grained explanations often suffer from sensitivity to model perturbations.

Segmentation Tasks

Segmentation tasks involve delineating anatomical structures or pathological regions at the pixel level, such as outlining tumor boundaries or segmenting organs [22]. In these tasks, explainability must address not just the presence of a feature but its exact spatial extent and boundaries. Attention Maps integrated within U-Net and Transformer-based architectures have proven effective in guiding the network to focus on relevant regions during segmentation. These attention mechanisms inherently provide a level of interpretability by visualizing the areas of interest during the

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 2, **Number** 2, July 2025
https://doi.org/10.62671/perfect.v2i2.115

learning process [23]. Adapted for segmentation models, Pixel-level Saliency Techniques offer explicit visual cues by highlighting pixels with the most significant influence on the segmentation outcome [24]. A major limitation in segmentation explainability lies in balancing interpretability with model performance. Overly smoothed explanations can obscure critical boundary information, while high-resolution saliency maps may amplify noise, leading to misleading interpretations.

Prognostic Assessment Tasks

Prognostic assessment tasks leverage imaging data to predict future clinical outcomes, such as disease progression, treatment response, or survival rates. Explanations in this domain require articulating how specific imaging features correlate with long-term patient outcomes. Counterfactual Explanations have gained traction in prognostic applications by illustrating how minimal changes in imaging features could alter the predicted prognosis. This approach offers clinicians actionable insights into the factors influencing risk stratification [25]. Additionally, Risk Factor Visualization Techniques that integrate imaging biomarkers with clinical variables provide holistic explanations, aiding in comprehensive patient assessments [26]. However, prognostic explainability faces the challenge of temporal complexity, as predictions often depend on dynamic and longitudinal data. Ensuring that explanations remain coherent and clinically valid over time requires sophisticated integration of multi-modal data sources and robust validation against real-world outcomes. Table 1 compares clinical imaging tasks and the corresponding explainability requirements, outlining how specific XAI techniques align with task-driven interpretability needs. The table emphasizes both the functional strengths of these methods and their inherent limitations when deployed in real-world medical imaging workflows.

Table 1. Task-centric mapping of XAI techniques aligned with clinical imaging requirements, summarizing their strengths and inherent limitations across classification, detection, segmentation, and prognostic tasks.

Clinical Task	Representative XAI Methods	Explanation Focus	Task-Specific Advantages	Key Challenges and Limitations
Image Classification	Saliency Maps, Class Activation Maps (CAM), LIME	Highlighting discriminative image regions	Provides intuitive visual justifications for AI predictions	Low spatial resolution; may highlight non-pathological areas
Object Detection	Bounding Box Visualization, SHAP, Perturbation-based	Localizing detected abnormalities	Correlates object presence with feature importance	Lacks precise boundary delineation; sensitive to noise
Anatomical Segmentation	Attention Mechanisms, Pixel-wise Saliency Maps	Emphasizing fine-grained anatomical structures	Facilitates boundary-level interpretability for segmentation tasks	Trade-off between resolution and explanation stability
Prognostic Risk Assessment	Counterfactual Explanations, Feature Attribution Maps	Explaining the impact of imaging features on outcomes	Supports scenario-based and personalized risk interpretations	Difficult to validate longitudinally; requires multi-modal data

DISCUSSION

The growing integration of AI-driven systems in medical imaging has increased the need for robust and clinically meaningful explainability methods. Although many XAI techniques have been developed and used for various imaging tasks, the match between explanation strategies and specific clinical goals remains imperfect. The common trend in literature categorizing XAI methods based on algorithmic traits (gradient-based, perturbation-based, surrogate models) often overlooks the contextual details linked to different clinical tasks like classification, detection, segmentation, and prognostic evaluation. This mismatch can cause the use of explanation methods that, while technically correct, do not satisfy the interpretability needs of clinical end-users. This paper suggests that a task-focused approach can close this gap by directly linking XAI methods to the functional demands of imaging tasks. For example, classification tasks are better served by simple, intuitive region-based explanations, whereas segmentation tasks need detailed, pixel-level insights that follow anatomical boundaries. Likewise, detection tasks require object-level attributions that balance spatial localization with feature relevance, while prognostic evaluations need scenario-specific explanations that connect imaging biomarkers with clinical outcomes. Though a task-oriented taxonomy has conceptual benefits, several challenges remain in applying this framework within clinical workflows. One major challenge involves balancing explanation accuracy and usability [27]. Techniques offering detailed, high-resolution explanations like pixel-level saliency maps often face noise sensitivity and can overwhelm clinicians with excessive details. On the other hand, methods that produce more straightforward explanations, such as bounding box visualizations, might miss crucial diagnostic information, risking oversight. Additionally, the subjective nature of interpretability presents another hurdle.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 2, **Number** 2, July 2025
https://doi.org/10.62671/perfect.v2i2.115

Clinical users, radiologists, surgeons, and oncologists may have diverse expectations about explanations' depth, detail, and format. Current XAI methods also lack adaptive systems that customize explanations to user preferences and clinical contexts. Another significant challenge is the absence of standardized metrics for evaluating task-specific explainability in medical imaging [28]. While visual plausibility is often used as a qualitative measure, there is an increasing need for quantitative standards that assess explanation reliability, clinical relevance, and influence on decision-making. To realize the full potential of XAI in medical imaging, future research must focus on developing hybrid explanation frameworks that integrate multiple explanation modalities, such as visual, textual, and quantitative, and dynamically adapt to task-specific and user-specific requirements. Fostering human-AI collaboration through interactive explanation interfaces can enhance user trust and facilitate more informed clinical decisions.

CONCLUSION

Explainable Artificial Intelligence (XAI) has emerged as a pivotal enabler for the trustworthy adoption of AI-driven medical imaging systems in clinical practice. Despite significant progress in developing various explanation methods, the prevailing algorithm-centric taxonomies fail to address the nuanced interpretability demands of distinct imaging tasks. This paper presented a task-centric taxonomy that aligns XAI techniques with key clinical tasks, including classification, detection, segmentation, and prognostic assessment. By contextualizing XAI methods within the functional objectives of these tasks, the proposed taxonomy aims to guide researchers and practitioners in selecting explanation strategies that enhance clinical relevance and usability. However, bridging the gap between technical explainability and clinical interpretability remains a multifaceted challenge. Current XAI methods often entail trade-offs between explanation fidelity and cognitive usability, and they lack adaptive mechanisms to cater to diverse clinical roles and preferences. Moreover, the absence of standardized evaluation metrics hinders the objective assessment of explanation quality, limiting the scalability and reliability of XAI implementations in real-world healthcare environments. Future research directions should focus on developing hybrid, multimodal explanation frameworks that combine visual, textual, and quantitative cues and are tailored to the specific demands of clinical tasks and user contexts. Additionally, advancing interactive human-AI collaboration interfaces will foster trust and facilitate iterative feedback loops between clinicians and AI systems. Establishing domain-specific benchmarks for evaluating explanation effectiveness, in alignment with clinical decision-making processes, is another critical avenue that warrants concerted effort. Realizing the vision of transparent and accountable AI in medical imaging necessitates a paradigm shift from generic, model-centric explanations to task-driven, user-centric interpretability solutions seamlessly integrating into clinical workflows.

REFERENCES

[1] Jain, L., Singh, P.: A historical and qualitative analysis of different medical imaging techniques. International Journal of Computer Applications 107(15) (2014)

[2] Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: Overview, challenges and the future. In: Classification in BioApps: Automation of Decision Making, pp. 323–350 (2017)

[3] Jabbooree, A.I., Alkaabi, H., Kamber, A.N.: Facial Expression Recognition Using Fused Features: A Comparison of Deep and Machine Learning. J. Comput. Netw. Archit. High Perform. Comput. 7(3), 684–699 (2025)

[4] Hua, D., Petrina, N., Young, N., Cho, J.G., Poon, S.K.: Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: A scoping review. Artif. Intell. Med. 147, 102698 (2024)

[5] Song, Y., Liu, Y., Lin, Z., Zhou, J., Li, D., Zhou, T., Leung, M.F.: Learning from AI-generated annotations for medical image segmentation. IEEE Trans. Consum. Electron. (2024)

[6] Khan, M.M., Shah, N., Shaikh, N., Thabet, A., Belkhair, S.: Towards secure and trusted AI in healthcare: a systematic review of emerging innovations and ethical challenges. Int. J. Med. Inform. 195, 105780 (2025)

[7] Minh, D., Wang, H.X., Li, Y.F., Nguyen, T.N.: Explainable artificial intelligence: a comprehensive review. Artif. Intell. Rev. 55(5), 3503–3568 (2022)

[8] Fontes, M., De Almeida, J.D.S., Cunha, A.: Application of example-based explainable artificial intelligence (XAI) for analysis and interpretation of medical imaging: a systematic review. IEEE Access 12, 26419–26427 (2024)

[9] Ahmed, S., Kaiser, M.S., Hossain, M.S., Andersson, K.: A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions. IEEE Access 13, 37370–37388 (2024)

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 2, **Number** 2, July 2025
https://doi.org/10.62671/perfect.v2i2.115

[10] Islam, S.M.S., Nasim, M.A.A., Hossain, I., Ullah, D.M.A., Gupta, D.K.D., Bhuiyan, M.M.H.: Introduction of medical imaging modalities. In: Data Driven Approaches on Medical Imaging, pp. 1–25. Springer, Cham (2023)

[11] Najjar, R.: Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics 13(17), 2760 (2023)

[12] Mohammed, A.A., Abdulwahhab, A.H., Ibraheem, I.K.: Detection Lung Nodules Using Medical CT Images Based on Deep Learning Techniques. Baghdad Sci. J. 22(5), 1596–1608 (2025)

[13] Esmaeilzadeh, P.: Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations. Artif. Intell. Med. 151, 102861 (2024)

[14] Singh, S.K., Virdee, B.S., Aggarwal, S., Maroju, A.: Incorporation of XAI and deep learning in biomedical imaging: a review. Polytech. J. 15(1), 1–15 (2025)

[15] Ashraf, K., Nawar, S., Hosen, M.H., Islam, M.T., Uddin, M.N.: Beyond the Black Box: Employing LIME and SHAP for Transparent Health Predictions with Machine Learning Models. In: 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), pp. 1–6. IEEE (2024)

[16] Wollek, A., Graf, R., Čečatka, S., Fink, N., Willem, T., Sabel, B.O., Lasser, T.: Attention-based saliency maps improve interpretability of pneumothorax classification. Radiol. Artif. Intell. 5(2), e220187 (2023)

[17] Kanglong, F.A.N., Ma, C., Peng, Y., Fang, Y., Ma, K.: Decision Rules are in the Pixels: Towards Pixel-level Evaluation of Saliency-based XAI Models. (Preprint)

[18] Rguibi, Z., Hajami, A., Zitouni, D., Elqaraoui, A., Bedraoui, A.: Cxai: Explaining convolutional neural networks for medical imaging diagnostic. Electronics 11(11), 1775 (2022)

[19] Guluwadi, S.: Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with ResNet-50. BMC Med. Imaging 24(1), 1–19 (2024)

[20] Sarp, S., Catak, F.O., Kuzlu, M., Cali, U., Kusetogullari, H., Zhao, Y., et al.: An XAI approach for COVID-19 detection using transfer learning with X-ray images. Heliyon 9(4) (2023)

[21] Prasad Koyyada, S., Singh, T.P.: An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images. Healthc. Anal. 4, 100206 (2023)

[22] Avazov, K., Mirzakhalilov, S., Umirzakova, S., Abdusalomov, A., Cho, Y.I.: Dynamic focus on tumor boundaries: A lightweight U-Net for MRI brain tumor segmentation. Bioengineering 11(12), 1302 (2024)

[23] Farrag, A., Gad, G., Fadlullah, Z.M., Fouda, M.M., Alsabaan, M.: An explainable AI system for medical image segmentation with preserved local resolution: Mammogram tumor segmentation. IEEE Access 11, 125543–125561 (2023)

[24] Thiagarajan, J.J., Thopalli, K., Rajan, D., Turaga, P.: Training calibration-based counterfactual explainers for deep learning models in medical image analysis. Sci. Rep. 12(1), 597 (2022)

[25] Akinsiku, A.M.: Literature Review on Explainable Artificial Intelligence (XAI): Techniques, Tools, and Applications. Tech-Sphere J. Pure Appl. Sci. 2(1) (2025)

[26] Bibi, N., Courtney, J., McGuinness, K.: Enhancing brain disease diagnosis with XAI: a review of recent studies. ACM Trans. Comput. Healthc. 6(2), 1–35 (2025)

[27] Phillips, V.: A counterintuitive approach to explainable AI in healthcare: balancing transparency, efficiency, and cost. AI Soc. (2025)

[28] Saadoun, O.N., Allayith, R.A., Lafta, M.K., Shakir, K.H., Hussein, A.S., Al-Farouni, M., Shareef, H.: AI Deployment Susceptibility: Challenges in Clinical Decision-Aid Implementation. In: 2024 International Conference on IoT, Communication and Automation Technology (ICICAT), pp. 597–602. IEEE (2024)