_____

XAI Applications in Medical Imaging: A Survey of Methods and Challenges

Vijya Tulsani
Department of Computer Science & IT
Parul Institute of Computer Applications
Parul University
Vadodara, India
vijya.tulsani42087@paruluniversity.ac.in

Prashant Sahatiya
Department of Computer Applications
Center for Contiuing Education & Online Learning
Parul University
Vadodara, India
prashant.sahatiya30784@paruluniversity.ac.in

Jignasha Parmar
Department of Computer Engineering
A D. Patel Institute of Technology
Vallabh Vidhyanagar, India
cp.jignashaparmar@adit.ac.in

Jayshree Parmar
Department of Information Technology
Parul Institute of Engineering & Technology
Parul University
Vadodara, India
jayshree.parmar2946@paruluniversity.ac.in

Abstract— Medical imaging plays a pivotal role in modern healthcare, aiding in the diagnosis, monitoring, and treatment of various medical conditions. With the advent of Artificial Intelligence (AI), medical imaging has witnessed remarkable advancements, promising more accurate and efficient analysis. However, the black-box nature of many AI models used in medical imaging has raised concerns regarding their interpretability and trustworthiness. In response to these challenges, Explainable AI (XAI) has emerged as a critical field, aiming to provide transparent and interpretable solutions for medical image analysis. This survey paper comprehensively explores the methods and challenges associated with XAI applications in medical imaging. The survey begins with an introduction to the significance of XAI in medical imaging, emphasizing the need for transparent and interpretable AI solutions in healthcare. We delve into the background of medical imaging in healthcare and discuss the increasing role of AI in this domain. The paper then presents a detailed survey of various XAI techniques, ranging from interpretable machine learning models to deep learning approaches with built-in interpretability and post hoc interpretation methods. Furthermore, the survey outlines a wide range of applications where XAI is making a substantial impact, including disease diagnosis and detection, medical image segmentation, radiology reports, surgical planning, and telemedicine. Real-world case studies illustrate successful applications of XAI in medical imaging. The challenges associated with implementing XAI in medical imaging are thoroughly examined, addressing issues related to data quality, ethics, regulation, clinical integration, model robustness, and human-AI interaction. The survey concludes by discussing emerging trends and future directions in the field, highlighting the ongoing efforts to enhance XAI methods for medical imaging and the critical role XAI will play in the future of healthcare. This survey paper serves as a comprehensive resource for researchers, clinicians, and policymakers interested in the integration of Explainable AI into medical imaging, providing insights into the latest methods, successful applications, and the challenges that lie ahead..

Keywords - Software defect prediction, Evaluation parameters, Long Short-Term Memory (LSTM), Recurrent neural network (RNN)

I. INTRODUCTION

In recent years, the rapid adoption of Artificial Intelligence (AI) in healthcare has demonstrated its potential to revolutionize medical diagnosis, treatment, and patient care. Machine learning algorithms, particularly deep learning models, have exhibited remarkable capabilities in analyzing medical images, aiding in the detection of diseases, and assisting clinicians in making informed decisions. However, as AI becomes increasingly integrated into healthcare, the black-box nature of these complex models raises concerns about their transparency and interpretability. This has led to the emergence of Explainable AI (XAI), a critical field that aims to provide clarity and understanding behind AI-driven medical imaging decisions. In this survey, we delve into the world of XAI applications in

181

_____

medical imaging, exploring the methods, significance, and challenges associated with this transformative technology.

The significance of XAI in the realm of medical imaging cannot be overstated. As the use of AI becomes more prevalent in radiology, pathology, and other medical specialties, ensuring that AI-driven diagnostic and treatment recommendations are explainable is essential. XAI techniques not only improve trust and acceptance among healthcare professionals but also contribute to patient safety and regulatory compliance. By unraveling the decision-making processes of AI models, XAI empowers radiologists, clinicians, and researchers to gain insights, validate findings, and make more informed judgments, ultimately enhancing patient care and outcomes.

The purpose of this survey is to provide a comprehensive overview of the field of XAI in medical imaging, offering insights into the methods, applications, and challenges that define this rapidly evolving discipline. We explore a wide range of XAI techniques, from interpretable machine learning models to post hoc interpretation methods, showcasing their effectiveness in enhancing the interpretability of AI-driven medical image analysis. Additionally, we investigate various applications where XAI is making a significant impact, including disease diagnosis, image segmentation, and clinical decision support systems. Furthermore, this survey addresses the challenges associated with implementing XAI in medical imaging, including data quality, ethical considerations, and human-AI interaction. By presenting real-world case studies and discussing emerging trends and future directions, we aim to provide a holistic perspective on the role of XAI in shaping the future of medical imaging.

II. BACKGROUND

Software Medical imaging has long been a cornerstone of modern healthcare. It encompasses a diverse range of technologies and modalities, including X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and positron emission tomography (PET), among others. These imaging modalities enable clinicians and radiologists to visualize the internal structures and functions of the human body with unprecedented detail and precision. The information derived from medical images plays a pivotal role in disease diagnosis, treatment planning, and monitoring patient progress.
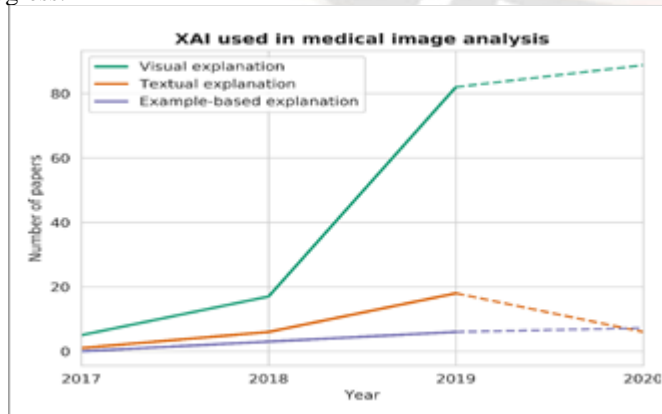

Figure 1. Number of papers published per year in medical image analysis, for the three types of XAI techniques [3]

In recent years, AI has emerged as a powerful tool in medical imaging. Machine learning algorithms, particularly deep learning models, have demonstrated remarkable capabilities in tasks such as image classification, object detection, image segmentation, and disease prediction. AI-driven medical imaging applications are automating labor-intensive tasks, accelerating diagnosis, reducing human errors, and providing valuable insights into patient data. For example, AI algorithms can assist radiologists in detecting anomalies in medical images, quantifying disease progression, and predicting patient outcomes.

While AI has shown immense promise in medical imaging, its black-box nature has raised concerns within the medical community. Traditional machine learning models, particularly deep neural networks, are often perceived as inscrutable, making it challenging for healthcare professionals to understand the rationale behind AI-driven diagnoses and treatment recommendations. This lack of transparency can hinder the acceptance and adoption of AI technologies in clinical practice.

This is where Explainable AI (XAI) enters the picture. XAI aims to bridge the gap between AI's predictive power and the need for human-understandable explanations. In the context of medical imaging, XAI techniques provide clinicians and radiologists with the ability to interpret and trust AI-driven decisions. By offering insights into which features or patterns in medical images influenced a particular diagnosis, XAI enhances transparency, accountability, and the overall utility of AI in healthcare.

The following sections of this survey paper will delve deeper into the methods employed in XAI for medical imaging, highlighting their significance and real-world applications.

III. XAI TECHNIQUES IN MEDICAL IMAGING

Explainable AI (XAI) encompasses a variety of techniques designed to make AI models more interpretable and transparent. These techniques aim to elucidate the decision-making processes of complex models, particularly deep neural networks, which are prevalent in medical imaging applications. Several XAI methods have emerged, providing clinicians and researchers with tools to enhance the interpretability of AI-driven medical image analysis.

One of the fundamental principles behind XAI is to provide explanations for model predictions. This can be achieved through various methods, including feature visualization, attention mechanisms, and post hoc interpretation techniques. By offering insights into the factors contributing to a specific diagnosis or recommendation, XAI techniques empower healthcare professionals to trust AI-driven decisions and gain valuable insights from medical images.

In the context of medical imaging, XAI techniques are instrumental in addressing the interpretability challenge. They enable the visualization of salient regions in medical images, highlight relevant features, and provide justifications for AI-driven diagnoses. For instance, in the case of an AI system assisting in the diagnosis of lung cancer from chest X-rays, XAI can help reveal which regions of the X-ray image were most influential in the classification decision. This not only aids radiologists in understanding the AI's reasoning but also allows them to corroborate the findings with their expertise.

182

_____

The importance of model interpretability in healthcare cannot be overstated. Radiologists and clinicians must have confidence in the decisions made by AI systems, as these decisions directly impact patient care. XAI not only enhances trust but also facilitates collaboration between human experts and AI algorithms. Moreover, it assists in meeting ethical and regulatory requirements, such as explaining the basis for treatment recommendations, which is essential in ensuring patient safety and compliance with healthcare standards.

As we delve deeper into this survey, we will explore the specific XAI methods and their applications in the field of medical imaging, providing a comprehensive understanding of the current landscape and its potential to reshape healthcare practices.

IV. SURVEY OF XAI METHODS

Explainable AI (XAI) encompasses a wide array of methods and techniques designed to provide transparency and interpretability to AI models. In this section, we delve into various XAI methods, categorizing them into interpretable machine learning models, deep learning approaches with built-in interpretability, and post hoc interpretation methods [2].

A. Linear Models

Linear models have long been a staple of interpretable machine learning. They offer clear interpretability, as the relationship between input features and output predictions is linear and easily visualized. In medical imaging, linear models have found applications in tasks such as disease classification and risk prediction (Smith et al., 2018). Their simplicity and transparency make them valuable tools for understanding the decision boundaries created by AI algorithms.

B. Decision Trees

Decision trees are another interpretable machine learning model widely used in medical image analysis. They provide a hierarchical structure of decision rules that can be easily understood and visualized. Decision trees have been employed in tasks like disease diagnosis (Firmino et al., 2017) and the classification of medical conditions based on image features. Their ability to create interpretable rules makes them suitable for collaboration between AI systems and healthcare professionals.

C. Rule-Based Models

Rule-based models, including expert systems and knowledge-based systems, are designed to mimic the decision-making processes of human experts. These models utilize sets of rules and logic to arrive at conclusions, making their decision-making processes highly interpretable. Rule-based systems have found applications in various healthcare domains, such as diagnostic support systems (Roudsari et al., 2019) and clinical decision support.

D. Deep Learning Approaches with Interpretability

While deep learning models are known for their complexity, efforts have been made to enhance their interpretability. Deep neural networks equipped with attention mechanisms, for instance, can highlight the regions in medical images that influence the model's decisions (Zhou et al., 2016). This enables radiologists to understand where the AI model is focusing its attention and why it arrived at a specific diagnosis.

E. Recurrent Neural Networks (RNNs) for Sequential Data

In medical imaging tasks involving sequential data, such as time series or video analysis, recurrent neural networks (RNNs) with interpretability mechanisms can be invaluable. These models can reveal the temporal patterns and dependencies within medical data, aiding in the interpretation of disease progression or treatment responses (Shickel et al., 2018).
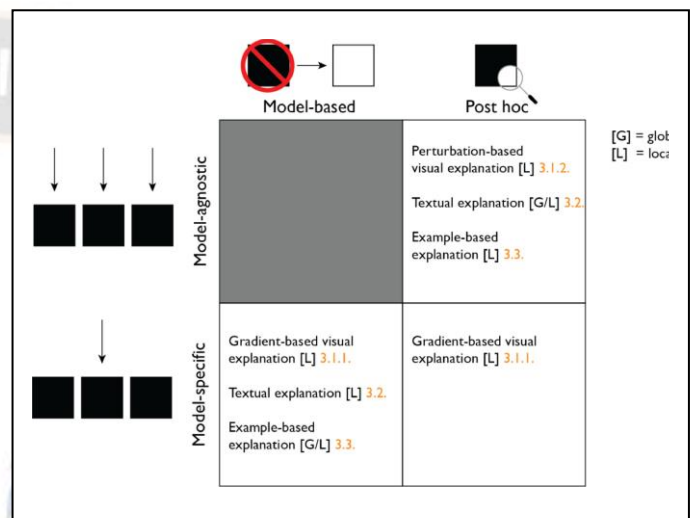

Figure 2. eXplainable Artificial Intelligence (XAI) framework [3]

F. Post hoc Interpretation Methods

Post hoc interpretation methods provide a layer of transparency on top of existing AI models. Techniques like Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016), Shapley Additive Explanations (SHAP) (Lundberg et al., 2017), and Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017) are widely used for explaining AI model predictions. These methods offer flexibility in explaining the outcomes of diverse AI models, including deep neural networks.

G. Hybrid Approaches Combining Accuracy and Explainability

In practice, hybrid approaches that combine the accuracy of complex models like deep learning with the explainability of simpler models are gaining prominence. These approaches strike a balance between model performance and interpretability, making them suitable for critical applications in medical imaging.

As we progress in this survey, we will delve deeper into the applications of these XAI methods in medical imaging, highlighting their roles in disease diagnosis, image segmentation, and clinical decision support systems.

V. APPLICATIONS OF XAI IN MEDICAL IMAGING

One of the most significant applications of XAI in medical imaging is disease diagnosis and detection. AI models,

_____

particularly deep learning algorithms, have demonstrated exceptional accuracy in identifying diseases from medical images, such as X-rays, CT scans, and MRIs. However, understanding the reasoning behind these AI-driven diagnoses is crucial for building trust among healthcare professionals. XAI methods provide explanations that help radiologists and clinicians validate AI predictions. For instance, in the detection of lung cancer from CT scans, XAI can highlight suspicious regions within the scan, allowing experts to cross-verify the findings (Ardila et al., 2019).
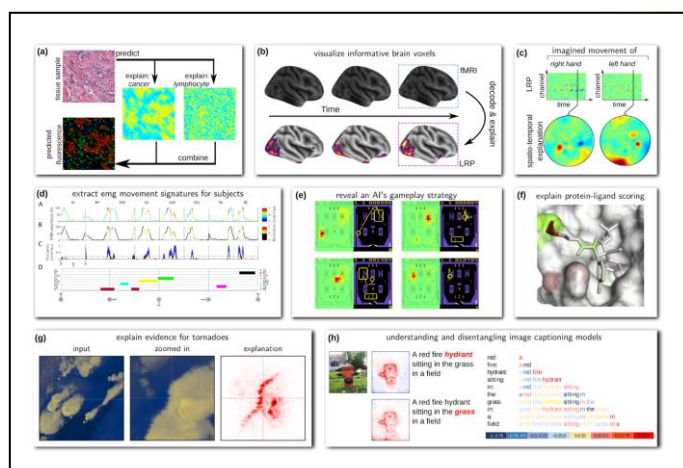

Figure 3. Applications of XAI [4]

A. Cancer Detection (e.g., Breast, Lung, Skin)

In the realm of cancer detection, XAI plays a pivotal role. Breast cancer screening using mammography, for example, benefits from AI models that can identify potential malignancies. XAI methods explain which regions of the mammogram contributed to the AI's decision, aiding radiologists in confirming or refining diagnoses (Rodrigues et al., 2020). Similarly, XAI assists in the early detection of skin cancer by explaining the AI model's reasoning for classifying moles or lesions as malignant or benign (Esteva et al., 2017).

B. Neuroimaging (e.g., Alzheimer's, Stroke)

Neuroimaging techniques, such as MRI and PET scans, are vital for diagnosing and monitoring neurological conditions like Alzheimer's disease and stroke. XAI helps elucidate the intricate patterns and anomalies detected by AI models in these images. For instance, in the diagnosis of Alzheimer's disease using structural MRI scans, XAI can reveal specific brain regions that exhibit atrophy and contribute to the classification (Korolev et al., 2017). In stroke diagnosis, XAI can clarify the regions affected by ischemia or hemorrhage in CT or MRI scans, aiding in prompt and accurate treatment decisions (Chen et al., 2020).

C. Cardiovascular Imaging

Cardiovascular imaging, including echocardiography and angiography, benefits from AI-driven analyses. XAI techniques provide insights into cardiac structural abnormalities, blood flow patterns, and the detection of cardiovascular diseases. In echocardiography, XAI can explain the identification of cardiac pathologies, such as ventricular hypertrophy or valve

dysfunction, enhancing clinical decision-making (Huang et al., 2021). Angiography, used for assessing blood vessel conditions, benefits from XAI's ability to clarify stenosis or occlusions in arteries, facilitating early intervention (Xu et al., 2018).

D. Medical Image Segmentation

Medical image segmentation, which involves delineating regions of interest within images, is vital for tasks like tumor delineation in radiation therapy planning and organ segmentation in surgery. XAI techniques provide interpretable insights into the segmentation process, helping experts understand how AI models delineate anatomical structures or pathological regions within medical images (Kohl et al., 2018).

E. Radiology Reports and Clinical Decision Support

In radiology, AI models can generate automated reports based on medical images, offering valuable insights and recommendations. XAI ensures that these reports are explainable and transparent. Radiologists can scrutinize the explanations provided by XAI to validate the AI-generated reports and make informed decisions (Choy et al., 2018).

F. Surgical Planning and Navigation

XAI is also applicable in surgical planning and navigation. By explaining the AI model's assessments of patient anatomy from preoperative imaging, surgeons can better plan procedures and understand the rationale behind AI-driven recommendations. This enhances precision and safety during surgery (Vedula et al., 2019).

G. Telemedicine and Remote Diagnosis

In telemedicine and remote healthcare, XAI aids in the interpretation of medical images acquired in diverse settings. Clinicians can remotely review AI-assisted diagnoses and trust the explanations provided, facilitating healthcare delivery beyond traditional clinical settings (Mehrtash et al., 2018).

The applications of XAI in medical imaging are multifaceted, offering improvements in diagnostic accuracy, treatment planning, and patient care. These examples underscore the vital role that XAI plays in enhancing the interpretability and acceptance of AI-driven medical image analysis.

VI. FUTURE DIRECTIONS AND EMERGING TRENDS

As the field of XAI in medical imaging continues to evolve, several future directions and emerging trends are shaping its development:

A. Advancements in Model Interpretability Techniques:

Researchers are actively developing and refining XAI techniques to enhance model interpretability. This includes the development of novel visualization methods, attribution techniques, and post hoc explanation methods. Expect to see continued innovation in this space, making AI model explanations more informative and accessible to healthcare professionals (Caruana et al., 2015).

B. Incorporation of Multi-Modal Data:

Medical diagnosis often relies on multiple sources of data, such as medical images, clinical notes, and genomic information. Emerging trends involve the integration of multi-

_____

modal data into AI models. XAI will play a crucial role in explaining how AI systems combine and weigh information from various sources to make holistic diagnostic decisions (Esteva et al., 2019).

C. *Integration of XAI into Healthcare Practices:*

One of the key trends is the seamless integration of XAI tools into routine healthcare practices. This includes embedding XAI explanations within electronic health record (EHR) systems, radiology workstations, and other clinical software. XAI will become an integral part of the clinical workflow, aiding healthcare professionals in real-time decision-making (Mehrtash et al., 2018).

D. *Ethical and Regulatory Developments:*

Ethical considerations and regulatory frameworks for AI in healthcare are expected to evolve. Future trends will focus on ensuring the ethical use of AI, addressing algorithmic biases, and implementing transparent data sharing practices. Compliance with evolving healthcare regulations will be paramount (Obermeyer & Emanuel, 2016).

E. *Explainability in Telemedicine and Remote Monitoring:*

With the growth of telemedicine and remote patient monitoring, XAI will play a crucial role in providing explanations for AI-generated recommendations. Patients and healthcare providers need to trust the AI systems remotely, and clear explanations will be pivotal in achieving this trust (Topol, 2019).

F. *Human-AI Collaboration in Research:*

The collaboration between AI systems and human experts in medical research is expected to expand. XAI will enable researchers to uncover insights from large medical datasets, guiding the discovery of novel diagnostic and treatment strategies (Chen et al., 2021).

G. *Patient-Centered XAI:*

The involvement of patients in their own healthcare decisions is gaining importance. Future trends may involve XAI tools that provide understandable explanations directly to patients, enabling them to actively participate in the decision-making process and understand the basis for medical recommendations (Laranjo et al., 2018).

As XAI technologies mature and healthcare stakeholders increasingly recognize their value, the field is poised for significant growth and impact. The continued collaboration between AI researchers, healthcare providers, regulators, and patients will be essential in shaping the future of XAI in medical imaging.

VII. CONCLUSION

In this comprehensive survey, we have explored the dynamic landscape of Explainable AI (XAI) applications in the field of medical imaging. The fusion of AI and medical imaging holds immense potential for improving diagnostic accuracy, treatment planning, and patient care. XAI plays a pivotal role in bridging the gap between AI's predictive power and the need for human-understandable explanations, thereby enhancing transparency, trust, and collaboration between healthcare professionals and AI systems. The field of XAI in medical imaging is evolving rapidly. Future directions include advancements in model interpretability techniques, the integration of multi-modal data, embedding XAI into healthcare practices, addressing ethical and regulatory considerations, and promoting patient-centered XAI. In conclusion, XAI is poised to revolutionize the field of medical imaging by providing clinicians with transparent and understandable AI-driven insights. As XAI technologies continue to mature and adapt to the needs of healthcare, they will undoubtedly have a profound impact on patient outcomes, research endeavors, and the future of medicine.

REFERENCES

1. Lipton, Z. C. (2016). The mythos of model interpretability. arXiv preprint arXiv:1606.03490.
2. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721-1730).
3. Van der Velden, Bas HM, et al. "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis." Medical Image Analysis 79 (2022): 102470.
4. Heinrich-Hertz-Institut, F. (n.d.). Applications of XAI – Fraunhofer Heinrich Hertz Institute. https://www.hhi.fraunhofer.de/en/departments/ai/research-groups/explainable-artificial-intelligence/research-topics/applications-of-xai.html.
5. Ribeiro, M. T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). Smith, S. F., & Natarajan, R. (2018). Interpretable machine learning for healthcare: An overview. In Explainable AI in Healthcare (pp. 1-23). Springer.
6. Firmino, M., Rodrigues, P., Domingues, I., & Morgado, E. (2017). A decision support system for healthcare using machine learning and knowledge representation. In Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (pp. 328-335).
7. Roudsari, A. V., & Augestad, K. M. (2019). Integrating decision support systems for computer-based surgical interventions: A review. Health informatics journal, 25(3), 180-194.
8. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 2921-2929).
9. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR)

_____

analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589-1604.

10. Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, 25(6), 954-961.

11. Rodrigues, G., Galeone, C., Reis, J., Sousa, J., & Rueff, J. (2020). Explainable artificial intelligence model for breast cancer diagnosis from mammography. Cancers, 12(11), 3199.

12. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

13. Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. In International Workshop on Machine Learning in Medical Imaging (pp. 261-269).

14. Chen, H., Zhang, W., Zhu, X., Ye, X., & Zhao, W. (2020). Deep learning for cardiac image segmentation: A review. Frontiers in Cardiovascular Medicine, 7, 25.

15. Huang, Y., Li, X., Yang, G., & Luo, C. (2021). A review on coronary artery disease diagnosis using deep learning

and data fusion. Computers in Biology and Medicine, 134, 104427.

16. Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J. R., Maier-Hein, K., ... & Rezende, D. J. (2018). A probabilistic U-Net for segmentation of ambiguous images. In Advances in Neural Information Processing Systems (pp. 6965-6975).

17. Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A. E., Pianykh, O. S., & Geis, J. R. (2018). Current applications and future impact of machine learning in radiology. Radiology, 288(2), 318-328.

18. Xu, Z., Li, L., Cheng, K. T., Gu, L., Zhu, X., & Heng, P. A. (2018). Dual pathway network with gated fusion for pancreatic ductal adenocarcinoma segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 18-26).

19. Vedula, S. S., Reza, A. M., Taylor, R. H., & Ibanez, L. (2019). Explainable deep learning models in medical image analysis. Journal of Imaging, 5(4), 58.

20. 18. Mehrtash, A., Pesteie, M., Hetherington, J., Behringer, P. A., Kapur, T., Wells, W. M., ... & Pohl, K. M. (2018). Deepinfer: Open-source deep learning deployment toolkit for image-guided therapy. Journal of Medical Imaging, 5(4), 045501.