Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.051363

ARTICLE

Explainable Artificial Intelligence (XAI) Model for Cancer Image Classification

Amit Singhal[1], Krishna Kant Agrawal[2], Angeles Quezada[3], Adrian Rodriguez Aguiñaga[4], Samantha Jiménez[4] and Satya Prakash Yadav[5,*,#]

[1]Department of Computer Science & Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, 201017, India

[2]Department of School of Computing Science and Engineering, Galgotias University, Greater Noida, 203201, India

[3]Tecnológico Nacional de México Campus Tijuana, Baja California, 22414, México

[4]Universidad Autónoma de Baja California, Tijuana, Baja California, 22414, México

[5]Department of Computer Science and Engineering, G.L, Bajaj Institute of Technology and Management (GLBITM), Affiliated to Dr. A. P. J. Abdul Kalam Technical University, Lucknow, 201306, India

*Corresponding Author: Satya Prakash Yadav. Email: satya.yadav_cse@glbitm.ac.in, prakashyadav.satya@gmail.com

#Currently in the School of Computer Science Engineering and Technology (SCSET), Bennett University, Greater Noida, 201310, India

Received: 04 March 2024 Accepted: 22 May 2024 Published: 20 August 2024

ABSTRACT

The use of Explainable Artificial Intelligence (XAI) models becomes increasingly important for making decisions in smart healthcare environments. It is to make sure that decisions are based on trustworthy algorithms and that healthcare workers understand the decisions made by these algorithms. These models can potentially enhance interpretability and explainability in decision-making processes that rely on artificial intelligence. Nevertheless, the intricate nature of the healthcare field necessitates the utilization of sophisticated models to classify cancer images. This research presents an advanced investigation of XAI models to classify cancer images. It describes the different levels of explainability and interpretability associated with XAI models and the challenges faced in deploying them in healthcare applications. In addition, this study proposes a novel framework for cancer image classification that incorporates XAI models with deep learning and advanced medical imaging techniques. The proposed model integrates several techniques, including end-to-end explainable evaluation, rule-based explanation, and user-adaptive explanation. The proposed XAI reaches 97.72% accuracy, 90.72% precision, 93.72% recall, 96.72% F1-score, 9.55% FDR, 9.66% FOR, and 91.18% DOR. It will discuss the potential applications of the proposed XAI models in the smart healthcare environment. It will help ensure trust and accountability in AI-based decisions, which is essential for achieving a safe and reliable smart healthcare environment.

KEYWORDS

Explainable artificial intelligence; artificial intelligence; XAI; healthcare; cancer; image classification

1 Introduction

Cancer is a dreadful disease that affects the lives of many people each year. It is a condition that can be difficult to detect and diagnose, which is why many advancements are being made in the area of cancer detection [1]. Modern cancer detection techniques have allowed for better diagnosis and tracking of certain types of cancer, often at earlier stages [2]. Imaging technology is a powerful tool in the field of healthcare that allows doctors to see inside the body and identify potential health issues. This technique involves various imaging devices, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scanners, which produce detailed images of organs and tissues. These images can be employed to detect abnormalities, such as tumors, and help doctors make accurate diagnoses. Other imaging techniques, such as Positron Emission Tomography (PET) scans, ultrasound scans, and X-rays, are also commonly utilized to detect changes in the body and determine specific types of cancer. Overall, imaging technology plays a crucial role in the early detection and treatment of diseases, allowing for better patient outcomes and improved overall health [3]. Another modern technique is biopsy, a procedure where a sample of the suspicious tissue is taken and analyzed under a microscope. It allows for the accurate diagnosis and determination of the type of cancer in the body. Blood tests are also common in modern cancer detection [4]. These tests search for elevated markers or protein levels, which can indicate cancer in the body. However, they cannot be used to diagnose the disease accurately. Whole-body scans employ advanced imaging and detection technologies to meticulously search for indications of cancer cells throughout the body, making them the most precise method for cancer diagnosis. Molecular and genetic testing aims to identify specific genetic alterations linked to particular forms of cancer. Modern cancer detection techniques have greatly improved the chances of survival for those with cancer, as they can help medical professionals detect and diagnose cancer earlier. However, it is essential to remember that the technology is still in its early stages and should be combined with other methods to have the most accurate diagnosis possible [5].

1.1 Importance of Cancer Detection

The importance of cancer detection in smart healthcare cannot be understated. Early detection is vital to successful cancer treatment and can help save lives. When cancer detection is done in a timely and effective manner, it can significantly reduce the risk of a patient's death from cancer. Smart healthcare enables cancer detection much earlier than traditional methods, allowing patients to start treatments as soon as possible and receive the best chance for a full recovery [6]. CT scans and MRIs provide incredibly detailed images of the body, allowing physicians to detect any potential cancer growth and even check to see if the cancer has spread. Genetic tests also aid in early cancer detection as they detect genetic abnormalities that can otherwise go undetected, helping a patient receive treatments earlier in the process. Smart healthcare can also help improve post-treatment care for cancer patients [7]. Smart healthcare systems can keep track of patient vital signs, capture and analyze patient data, and connect patients with the right doctors. With continued monitoring after treatment, smart healthcare can reduce the risk of cancer recurrence. In addition, because of the increased accessibility to diagnostic tests, calculations, and overall healthcare management, these systems can better support doctors in detecting subtle changes in patient health that could be early signs of cancer [8]. Smart healthcare is revolutionizing the way cancer is detected and managed. Smart healthcare systems help significantly reduce the mortality rate from cancer by optimizing the early detection of cancer and improving post-treatment care. The earlier cancer is detected, the better chance a patient has of fighting and surviving the disease; smart healthcare has made this possible.

1.2 Image Processing in Cancer Detection

Image processing plays a vital role in smart healthcare, particularly in detecting and diagnosing cancer. The powerful image processing techniques enable medical professionals to detect the precise location of tumors in the body and better characterize them in terms of size, shape, and density. It can help inform decisions around the most appropriate treatment plan. Advanced image processing techniques can also provide invaluable insights into the effectiveness of treatment [9]. For example, analyzing images over time can help to indicate the responsiveness of cancerous cells to targeted therapies and to determine when it is necessary to switch to a different course of treatment. Access to precise images and data points makes it possible to determine how effective treatments have been over time [10]. Finally, image processing facilitates the automation of some of the decision-making processes surrounding cancer detection and the setting of diagnosis parameters. It reduces the potential for human error, which can be crucial when assessing potentially life-threatening conditions. Medical professionals can be sure to make informed decisions about treatments and expectations by allowing machines to access and analyze previously collected images and data points [11]. Altogether, image processing is an essential tool for the smart healthcare of the future and will be essential for the successful detection and diagnosis of cancer. The technology offers invaluable support for medical professionals and removes some of the potential for human error and uncertainty.

1.3 Understanding the Decisions and Predictions of Machine Learning Models

Machine learning models are used in cancer detection to analyze data and identify patterns indicative of cancer. These models utilize algorithms to make decisions and predictions based on the input features, such as patient characteristics and test results. The model first undergoes a training phase, learning from a large dataset of cancer and non-cancer cases. After this training, the model can make predictions on new cases by comparing the input features to the patterns it learned. The decisions and predictions made by the model can help medical professionals identify potential cancer cases earlier and improve treatment outcomes [12]. Fig. 1 illustrates the architectural design of the machine learning model in cancer detection.

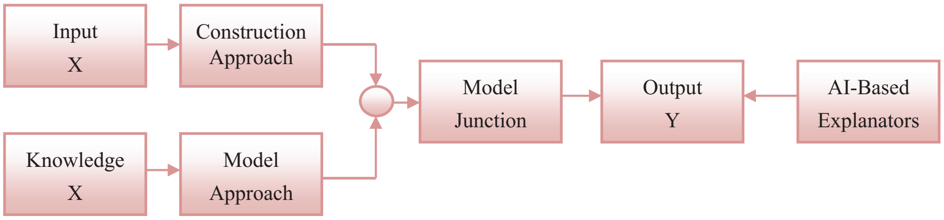

Figure 1: Construction of machine learning model for cancer detection

Input X refers to the data and information that is collected for a specific problem. The construction approach involves designing and developing a system or model to solve the problem. Knowledge X represents the expertise and understanding of the problem domain employed to inform the model approach. The model approach refers to the techniques and methods applied to create and train a model that can accurately solve the problem [13]. Model junction is where input X, construction approach, and knowledge X combine to build and optimize the model. Output Y is the desired result or prediction generated by the model. AI-Based Explainers are tools or algorithms that use artificial intelligence to analyze and interpret the model's results, providing insightful explanations for the predictions made by the model. These explanators help to bridge the gap between technical model outputs and human understanding, making AI more accessible and transparent [14].

1.4 Explainable Artificial Intelligence in Cancer Classification

It is a nascent subfield within machine learning and artificial intelligence (AI) that centers on advancing algorithms and models capable of offering comprehensible and practical elucidations regarding the underlying mechanisms of AI systems. It is increasingly important as AI models and algorithms become increasingly complex and challenging to interpret. The Explainable Artificial Intelligence (XAI) model for cancer image classification is one such system [15]. The XAI model for cancer image classification is designed to provide an AI-guided approach to more accurately detect cancer in medical images. Fig. 2 shows the XAI for cancer classification [16].

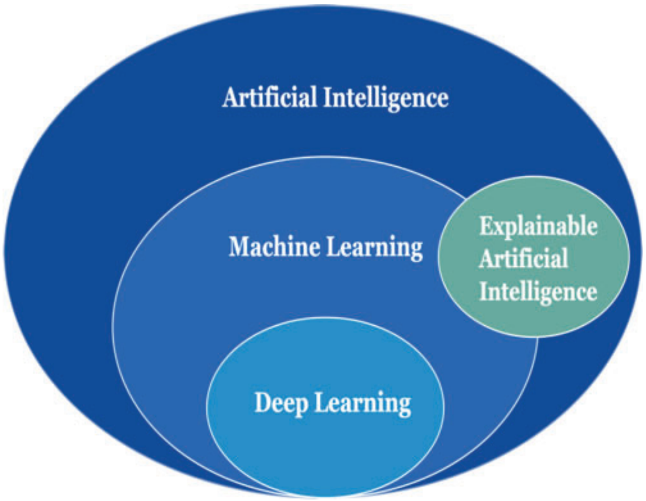

Figure 2: XAI for cancer classification

The proposed model employs a hybrid approach, incorporating convolutional neural networks, decision trees, transfer learning, and various other machine learning approaches to ascertain the existence and magnitude of a specific type of cancer [17]. The objective is to develop an AI system capable of delivering a significantly more comprehensive elucidation of an image, specifically about the presence or absence of cancer inside it. The system takes in medical images and produces an output of probabilities of each different based on predetermined protocols [18]. It helps to reduce the ambiguity of human interpretation while providing an acceptable level of accuracy. The goal is to provide an explainable AI system with a more reliable and accurate cancer diagnosis than traditional methods [19]. The system provides an accurate and explainable interpretation of the image and any potential positive/negative areas of the image. It allows clinicians to make decisions faster and more efficiently. In addition, this enhanced accuracy in cancer detection can help prevent misdiagnosis and decrease mortality rates [20]. The primary contribution of this study encompasses the following aspects:

- Increased Transparency: The proposed framework provides a transparent mechanism for understanding and interpreting the decisions made by AI models. This is especially important in the context of cancer image classification, where an AI model's decisions can significantly impact patient outcomes.
- Interpretable Feature Importance: It can identify and highlight the most important features or regions in an image that contribute to the final classification outcome. This helps clinicians and researchers better understand the characteristics and patterns in cancer images that lead to a specific diagnosis.
- Detection of Bias: It can detect bias in the training data and identify potential sources of bias in the AI model's decision-making process. This is critical in ensuring that the model is not making biased or unfair decisions that could negatively impact specific patient populations.

- Model Explainability: The proposed framework can explain the decisions made by AI models, making them more interpretable and explainable to clinicians and patients. This can help build trust and acceptance of AI technology in the medical field.
- Improvement of Model Performance: Identifying and visualizing the key features and patterns in cancer images can assist in model optimization and improvement. This leads to more accurate and reliable classification results, which are crucial in cancer diagnosis.

2 Related Works

The XAI model utilized for cancer picture classification represents a noteworthy advancement within artificial intelligence. The utilization of this technology has promise in enhancing the precision and dependability of cancer detection through the analysis of medical images, potentially contributing to the amelioration of survival rates. Clinicians and medical practitioners can gain more trust in the results and better understand why certain decisions were made by explaining how an AI system reached its conclusions. It can lead to greater adoption of AI in clinical settings, which will be beneficial in improving diagnosis accuracy and outcomes.

Hassan et al. [21] examined the categorization of prostate cancer based on ultrasound and MRI pictures. They employed a deep learning-based XAI technique to automatically segment and categorize tumors in prostate ultrasound and MRI images. The deep learning model was trained to effectively identify and classify malignant spots in the images, enhancing treatment decisions by furnishing clinicians with more specific information. O'sullivan et al. [22] discussed an XAI as a field of research that seeks to make machine learning algorithms more interpretable, allowing for improved analysis of diagnostic procedures. It is vital for complex, invasive procedures such as medical imaging or digital pathology. XAI involves using algorithms that can explain why the classifier came to a particular conclusion, such as what attributes of the image were employed to identify a tumor or other diagnosis. XAI bridges the gap between automated image analysis and the manual navigation of intricate diagnostic procedures by furnishing the analyst with an added layer of detail. Knapič et al. [23] examined the concept of XAI to enhance human decision support within the medical field. This approach to AI technology aimed to increase transparency in the decision-making process. XAI solutions were formulated and designed with a foundation rooted in the fundamental principles of explainability, interpretability, and trustworthiness. The primary objective of their work was to enhance human decision-making processes by providing comprehensive explanations of the methodology employed by an AI-based medical decision-support system to generate its outcomes. Explanation systems play a crucial role in enhancing medical practitioners' comprehension of the rationales underlying the decisions made by the system. This fosters the development of trust and confidence in the judgments and conclusions rendered by the system. XAI solutions also provide tools that enable easy adjustment of parameters to make them better adapted for individual circumstances, increasing the system's flexibility for practitioners. Sarp et al. [24] discussed an XAI as a novel approach to machine learning and decision-making that seeks to provide insights into the operations and decisions made by AI systems. In particular, it was recognized as an essential tool for chronic wound classification, as understanding the decisions of AI systems is critical in cases of high-stakes medical decision-making. XAI provides crucial explanations regarding why the AI system reached a particular outcome, including what data were important, what other factors can be considered, and how the system filtered and manipulated the data before concluding. Clinicians can extrapolate from the current dataset to make sound decisions by accessing the XAI system's interpretations in future cases. van der Velden et al. [25] examined the concept of XAI in the context of deep learning-based medical image analysis. XAI refers to a form of AI technology that enhances medical professionals' comprehension of the decision-making processes employed by AI systems. The medical AI system describes its decision-making process utilizing various traits and data points to justify its evaluations. In addition, it validates that the artificial intelligence system is making evaluations within accepted bounds. XAI possesses the capacity to serve as a beneficial instrument for medical professionals, offering a profound understanding of the decision-making processes of medical AI systems. It can clarify critical inquiries, including identifying the most significant features inside a given model.

Alonso et al. [26] discussed a bibliometric analysis of the XAI research field to evaluate the quantity and quality of research conducted in a particular field. It utilizes tools such as citation analysis, impact factor, and content/concept analysis to gain insights into the topics and methods being used within a field. This type of analysis can help understand the state of the field and help guide decision-making about future research investments. In addition, it can help identify areas of research that need more investigation, areas that are not being explored, or even highlight trends that could portend the direction of future research. Dağlarli [27] has discussed an XAI approach using algorithms that can explain to users why decisions are being made. Such approaches are invaluable for allowing users to understand the decisions made by AI algorithms and trust their results. Deep meta-learning models are AI algorithms that learn hierarchically from data to develop a model they can apply to other data. These models are helpful for tasks such as learning to play video games, where they can learn to adapt and make predictions. Muddamsetty et al. [28] investigated the evaluation of expert-level assessments for XAI techniques within the medical field. This evaluation entails examining the accuracy, consistency, and effectiveness of the AI system's outcomes regarding the evaluations of medical experts. It also assesses how well the AI system's explanations can help the medical expert better understand the patient's condition. Finally, it evaluates whether the AI system is compliant with the relevant regulatory or legal requirements in the domain, such as applicable disclosure and consent requirements. Holder et al. [29] provided a comprehensive analysis of XAI, a burgeoning area of study aimed at enhancing the transparency and comprehensibility of AI systems for human users. XAI is a research area dedicated to advancing AI systems that can interact with people by offering comprehensive explanations for their actions and judgments. XAI facilitates enhanced comprehension of an AI system's decision-making process, fostering increased confidence and reliance in said system. XAI demonstrated significant potential in sectors characterized by high stakes, such as the medical field. As a novice cyber analyst within this field, the implementation of XAI will provide the capability to indicate to medical professionals and patients the underlying rationale behind a specific therapy recommendation made by an AI system. Enhancing trust in the AI system can positively impact the reliability and personalization of medical decision-making. In their study, Kakogeorgiou et al. [30] have examined the evaluation of XAI techniques for multi-label deep learning classification tasks in the field of remote sensing, which necessitates the examination of prediction accuracy and the comprehension of system functioning to establish its reliability. XAI techniques, such as Local Interpretable Model-agnostic Explanations (LIME) and Instance-Based Explanations (IBE), can be employed to indicate the decision-making process of a deep learning model when applied to remote sensing data, thereby facilitating a deeper comprehension of the underlying mechanisms at play. The explanations above can then be employed to determine regions that necessitate enhancement and assess the model's performance with greater precision.

Based on Bento et al. [31], XAI is a kind of AI technology that enables robots to communicate their choices to people comprehensibly. It tries to bridge the gap between the black-box nature of neural networks and traditional decision trees. It can improve deep learning performance by identifying relevant features, attributing importance levels to them, and discovering patterns that humans do not easily see. In this way, XAI can optimize and improve a deep learning model to better meet the requirements of a given task. Hauser et al. [32] discussed an XAI as a form of AI specifically designed to explain its actions and decisions. XAI can help improve the understanding of skin cancer diagnosis by providing more insight into why decisions were made. With an XAI model, a doctor or healthcare provider can better understand what factors lead to a specific diagnosis. XAI can also help guide the diagnosis process, allowing for a more informed decision-making process. The goal is for XAI to increase accuracy and reduce misdiagnosis while explaining what drove the decisions to help increase trust in the process. Vilone et al. [33] examined the issue of explainability being of paramount importance in developing AI systems, given that the decision-making processes driven by AI can frequently exhibit complexity and opacity. The concept of explainability pertains to the degree to which AI can offer substantial justifications for its decisions. The evaluation methodologies employed for explainable AI are designed to assess the effectiveness and dependability of an AI system in delivering explanations that are comprehensible to humans in its decision-making processes. These can encompass the evaluation of the accuracy of explanations, along with conducting user tests to assess the effectiveness of an AI system in communicating explanations to users. Nazir et al. [34] explored the application of XAI techniques in biomedical imaging. Specifically, they explore using deep neural networks to analyze medical pictures and derive meaningful insights for decision-making. XAI methodologies employ visual explanations and other interpretable models to show deep learning systems' behavior and decision-making processes. The comprehension of the decision-making processes employed by computers in the context of biomedical imaging is of paramount significance within the area. This understanding facilitates enhanced diagnostic capabilities and informed medical treatment decisions for healthcare practitioners. In addition, utilizing XAI approaches enhances the precision and dependability of these decision-making processes. This is primarily due to the incorporation of XAI, which facilitates the interpretation of the output and enables a more comprehensive understanding of the decision-making mechanisms. Örnek et al. [35] examined the concept of XAI, a category of techniques within the field of AI that seek to enhance the interpretability of machine learning models and decision-making processes. Class Activation Maps (CAMs) are a type of XAI approach employed to interpret image classification models. CAMs achieve this by selectively emphasizing significant regions within an image that have played a crucial role in the model's decision-making process. In neonatal medical thermal imaging, class activation maps (CAMs) are employed to detect and highlight significant areas within the image. This aids the model in accurately categorizing the image according to its specific kind. Visualizing the picture regions utilized by the model for categorization enables medical practitioners to enhance their comprehension of the model's decision-making process.

Cilli et al. [36] provided a discussion on the topic of XAI as a specific sort of AI technology utilized to identify instances of wildfires in natural environments. XAI uses supervised machine learning methods to analyze various types of environmental data, including satellite photography, weather data, and topography information. Based on the provided data, XAI demonstrates a high level of accuracy in detecting and precisely determining the location of wildfire outbreaks by identifying abnormalities within the data. The significance of this technology is growing in its ability to effectively mitigate fire hazards in many settings among the challenges posed by climate change. Zhang et al. [37] provided a comprehensive discussion on XAI, an emerging domain within the field of AI research. XAI pertains to developing AI algorithms and systems that can render decisions and offer explanations in the context of medical diagnoses or surgical procedures. XAI is employed to get insights into machines' decision-making processes and establish transparency in the interaction between medical professionals and AI systems. This can potentially facilitate the elucidation of medical practitioners' decision-making processes during patient diagnosis and provide insights into optimal surgical techniques for specific procedures. XAI can also offer auditing capabilities for medical judgments, assisting physicians in comprehending the underlying rationale behind AI-generated findings. Nigar et al. [38] examined a specific form of artificial intelligence that possesses the ability to provide explanations for its decision-making processes. The methodology employed in this study involves using deep-learning models to identify and categorize skin lesions. The model after that provides an explication of its decision-making process, facilitating enhanced comprehension among medical practitioners and enabling them to make more judicious decisions. This methodology can potentially enhance the precision of skin lesion detection by detecting uncommon lesions that conventional methods might fail to detect. Diagnosing spinal diseases is a significant burden for doctors, as highlighted by Dindorf et al. [39]. Deep learning models are utilized to develop pathology-independent classifiers to classify aberrant and normal postures of the spine. XAI tools are then employed to comprehensively comprehend the classification procedure. XAI techniques facilitate the acquisition of knowledge by physicians regarding the decision-making process of deep learning models, enabling them to comprehend the rationale behind specific results. In addition, XAI techniques provide the means to scrutinize potential inaccuracies that may arise throughout the prediction process. Mehta et al. [40] presented a scholarly examination of a technology-centric strategy for detecting and reducing hate speech on digital platforms. The system employs natural language processing and machine learning techniques to identify and evaluate instances of offensive language within social media content. Explainable AI is employed to indicate the identification of hate speech by elucidating the rationale for flagging specific words or phrases. Enhancing transparency and accountability aids in facilitating the hate speech detection and mitigation process.

Vilone et al. [41] discussed the output formats used for classifying explainable AI-based methods. Natural language explanations provide descriptions of the reasoning behind model predictions, allowing users to understand why the decision was made. Visualizations are graphical representations of the model's outputs that highlight the various components of the model. Causality explanations are explanations that primarily focus on the causality of the model, as opposed to the characteristics of the data. The mechanism explanations reveal the "inner workings" of the model by outlining the algorithms and techniques used to build it. Geetha et al. [42] discussed a technique to accurately sense, identify, and classify structural cracks in concrete structures. This technique is based on deep Convolutional Neural Networks (CNNs) combined with XAI for automated detection and categorizing of structural cracks and their severity. The 1D convolutional neural network is trained on the solution-based approach using extracted features of the cracks' images and their angles. After training, the deep convolutional neural network is used for crack classification, and the XAI technique is employed to explain the model's decisions. It enables engineers to gain knowledge of the functioning of the deep learning model. It helps them to understand the identification process and the predictions made by the model. Loetsch et al. [43] discussed an approach that leverages machine learning and artificial intelligence techniques to understand chronic pain's complex, multi-dimensional phenotype. It utilizes interpretable models to capture the underlying patterns among various phenotypic characteristics and to identify meaningful clusters within a given data set. Medical researchers and clinicians can better diagnose and treat painful conditions by understanding the relationships between pain phenotypes. This approach can be employed to uncover potential drug targets and better understand pain's molecular and physiological mechanisms. Clancey et al. [44] discussed some methods and standards for research on XAI to guide scientists and engineers when designing, developing, and evaluating XAI systems. They guide how to design systems that have explainability as one of their core design objectives, how to monitor and evaluate progress and performance, and how to create an XAI product that is safe and acceptable to end users. Standards include guidelines, reference architectures, and best practices for implementing XAI systems. These methods and standards aim to ensure that XAI systems are explainable, reliable, and safe. Ghnemat et al. [45] discussed an approach to using AI in the medical imaging field. They used deep neural networks capable of learning high-level representations for various image analysis tasks in medical imaging to obtain higher classification accuracy. XAI aids in understanding the system's inner workings and provides insight into the decisions made by the system, which is especially desirable from a clinical perspective. In addition, XAI offers interpretability, which is crucial in the medical domain, where explanations for decisions are provided. Table 1 shows the comprehensive analysis of related works.

Table 1: Comprehensive analysis

Author	Year	Strength	Weakness
Hassan et al. [21]	2022	The inclusion of an explainable AI component enables clinicians to acquire a deeper understanding of the fundamental characteristics and decision-making processes employed by the model.	It could be employed to identify the progression of a disease.
O'sullivan et al. [22]	2022	It provides transparency with regard to AI decision-making processes, enabling medical practitioners to verify and refine patient-centric interventions.	It may be difficult to implement due to the complexion of AI technology and algorithms.
Knapič et al. [23]	2021	It can provide insight into how input data varies and impacts the prediction or outcome of the AI algorithm.	The explainable simulations are often more complex and slower to compute than traditional AI operations.
Sarp et al. [24]	2021	It allows for faster diagnosis of wound-related symptoms, which can lead to quicker and more accurate treatment decisions.	It may lead to doctors relying too much on the technology and less on their judgment, which can lead to incorrect diagnoses or treatments.
Van der Velden et al. [25]	2022	This facilitates enhanced comprehension of the decision-making process employed by deep learning models, enabling the detection and rectification of errors or biases.	This could violate patient privacy laws and lead to legal repercussions.

(Continued)

Table 1 (continued)

Author	Year	Strength	Weakness
Alonso et al. [26]	2018	The analysis can offer a significant understanding of the present condition of the study domain of Explainable AI, encompassing growing patterns and prominent contributors.	It cannot provide a detailed analysis of current Explainable AI research, such as interpretability methods used and research conclusions.
Dağlarli [27]	2020	XAI methods provide insights into the system's decision, which is unperceivable in other learning paradigms.	Utilizing XAI and deep meta-learning models necessitates substantial quantities of data and robust computational resources to deliver precise explanations.
Muddamsetty et al. [28]	2021	Expert-level evaluations can provide a more accurate assessment of trust in XAI methods for medical use.	Expert-level evaluations can be costly and time-consuming to implement.
Holder et al. [29]	2021	It gives the human analyst more capacity to identify potential malicious content due to the AI's ability to process multiple inputs quickly.	Potential for bias and errors due to XAI's inability to correctly interpret and make sense of complex data sets.
Kakogeorgiou et al. [30]	2021	The utilization of this approach can facilitate a more comprehensive comprehension of the fundamental process governing the model's performance in tasks involving the classification of multiple labels.	Explaining the performance of intricate, deep learning models, particularly in the context of remote sensing data, can provide challenges.
Bento et al. [31]	2021	XAI provides more interpretable deep learning models to help identify weaknesses more accurately.	XAI can be computationally expensive and require huge amounts of data to work correctly.
Hauser et al. [32]	2022	Explainable AI can help improve the accuracy of skin cancer recognition by providing a transparent understanding of the decision-making process.	Explainable AI can be more computationally expensive than traditional AI approaches.
Vilone et al. [33]	2021	They can help to ensure transparency and accountability between AI-based decisions and human users.	They can be challenging to evaluate correctly, potentially leading to bias.

(Continued)

Table 1 (continued)

Author	Year	Strength	Weakness
Nazir et al. [34]	2023	Deep neural networks effectively extract features from biomedical images and provide explainable AI results.	Deep neural networks can be computationally expensive and require substantial training data.
Örnek et al. [35]	2021	It can provide a non-expert with a graphical representation and explanation for the model's decision.	Additional resources are required to acquire knowledge related to medical thermal images.
Cilli et al. [36]	2022	XAI can detect wildfire occurrences with more accuracy and precision than non-explainable models by reasoning the cause and effect of the data.	XAI is generally more expensive and complex to set up than traditional AI models, which can make it less accessible to businesses.
Zhang et al. [37]	2022	It allows for a transparent look into the decision-making process to facilitate patient understanding.	It can be challenging to automate the interpretation of complex decision-making processes.
Nigar et al. [38]	2022	The deep learning methodology, which incorporates explainable artificial intelligence, demonstrates the ability to make precise predictions with high accuracy.	Training such models may take longer and require more computing power due to the complexity of deep learning networks.
Dindorf et al. [39]	2021	This technique can help clinicians assess postural health without needing pathology-specific criteria.	The cost of implementing the required technology and data storage infrastructure may be expensive.
Mehta et al. [40]	2022	XAI can provide insight into why specific hate speech is detected, allowing for better decision-making concerning how to address it.	XAI methods are still computationally expensive and can be challenging to implement in production systems.
Vilone et al. [41]	2021	It allows for greater accuracy of results through differentiated output formats.	It can be difficult to apply in practice due to the large number of classification groups.

(Continued)

Table 1 (continued)

Author	Year	Strength	Weakness
Geetha et al. [42]	2022	This technology enables expedited and precise detection of fissures in concrete, enhancing operational efficiency.	The financial implications associated with implementing this technology are considerably greater than those of conventional inspection methods.
Loetsch et al. [43]	2021	A more thorough inspection of products, materials, and components, as well as a faster identification of faults.	Developing, setting up, and using the technology can incur additional time and resources.
Clancey et al. [44]	2021	It reveals potential applications of AI-driven technologies and solutions to existing problems.	The discussion of applying these approaches and standards to future use cases is lacking in specificity.
Ghnemat et al. [45]	2023	XAI provides the ability to interpret the results of medical imaging classifications, increasing the trust and accountability in the system.	XAI can be a complex and time-consuming task, which may limit its use in time-sensitive acute situations.

Deep learning algorithms have demonstrated encouraging outcomes in identifying and categorizing diverse cancer forms. The models included in this study have undergone training using datasets comprising several photos classified as malignant or benign. These models demonstrate high accuracy in classifying cells into various forms of cancer using convolutional neural networks and deep learning methods. However, the current performance of deep learning models for cancer classifications remains constrained. Their ability to effectively detect and classify all types of cancer may be limited. With the accumulation of supplementary data and the advancement of research in this domain, the precision of deep learning models is expected to be enhanced. Table 2 below presents the performance of other existing models.

Table 2: Performance of existing models

Author	Year	Model	Accuracy
Hassan et al. [21]	2022	Deep learning	95.28%
O’sullivan et al. [22]	2022	Image analysis	93.26%
Knapič et al. [23]	2021	Human decision support system	88.65%
Sarp et al. [24]	2021	Chronic wound classification	89.71%
Van der Velden et al. [25]	2022	Deep learning	92.57%

(Continued)

Table 2 (continued)

Author	Year	Model	Accuracy
Alonso et al. [26]	2018	Biblio-metric analysis	91.68%
Dağlarli et al. [27]	2020	Deep meta-learning models	84.36%
Muddamsetty et al. [28]	2021	Expert level analysis	92.35%
Holder et al. [29]	2021	Junior cyber analyst	85.69%
Kakogeorgiou et al. [30]	2021	Deep learning classification	90.33%
Bento et al. [31]	2021	Deep learning	84.69%
Hauser et al. [32]	2022	Skin cancer recognition	82.36%
Vilone et al. [33]	2021	Evaluation approach	81.61%
Nazir et al. [34]	2023	Deep neural networks	88.54%
Örnek et al. [35]	2021	Class activation maps	92.65%
Cilli et al. [36]	2022	Deep learning	80.66%
Zhang et al. [37]	2022	Deep learning	90.48%
Nigar et al. [38]	2022	Deep learning	92.65%
Dindorf et al. [39]	2021	Pathology-independent classifier	91.11%
Mehta et al. [40]	2022	Deep learning	90.58%
Vilone et al. [41]	2021	Deep learning	92.55%
Geetha et al. [42]	2022	Deep learning	88.54%
Loetsch et al. [43]	2021	Deep learning	84.56%
Clancey et al. [44]	2021	Deep learning	86.65%
Ghnemat et al. [45]	2023	Deep learning	92.68%

2.1 Research Motivation

- Lack of data regarding biological sequencing of cancer samples, e.g., mutations and gene expression, and lack of experimentally validated image classification models.
- Inaccurate annotation in image datasets and inadequate understanding of the complexity of cancer image features and the relationship to classification outcomes.
- Lack of generalizable and transferrable representations and methods for cancer image classification and insufficient data augmentation techniques for accurate image classification and segmentation.
- Poor understanding of effectively deploying deep learning approaches for cancer image classification and lack of adequate evaluation data sets with clinically relevant annotations.
- Poor benchmarking methodologies for comparison of image classification algorithms and lack of modern interpretability techniques for cancer image classification models.

2.2 Research Contribution

- Increased Accuracy of Diagnoses: XAI models can increase the accuracy of cancer image classification by providing detailed explanations of the model's decisions and recommendations, which can help reduce the burden on medical professionals.

- Improved Patient Experience: XAI models can improve patient experience by providing detailed explanations specific to their diagnoses. It can help patients understand their diagnosis and treatment plan better.
- The transparency of AI decision-making is facilitated by XAI models, which offer insights into the rationale and mechanisms behind the decisions made by AI systems. This capability enhances transparency and accountability in AI decision-making processes.
- Enhanced Interoperability of Systems: XAI models have the potential to enhance interoperability between AI systems and medical experts by offering comprehensive explanations of the decision-making process of the AI system, which can be readily understood and interpreted by medical practitioners.
- Augmented Understanding for Healthcare Professionals: XAI models can offer an augmented understanding of the decision-making process within AI systems, hence providing valuable insights that can assist healthcare professionals in making more informed and improved judgments.

3 Proposed Model

An input image is a digital representation of a visual scene captured by a camera or generated by a computer. It is a two-dimensional array of pixels, where each pixel represents a specific color or intensity. The size and resolution of the input image play an essential role in the performance of computer vision systems. Image resize is the process of changing the size or resolution of an image while preserving its visual content. This technique is commonly employed to adapt images to different display or processing requirements. In computer vision, image resizing is particularly important for efficient processing and extracting features from large datasets. Efficient NETB0 is a convolutional neural network architecture designed for efficient image classification. It uses a combination of depth-wise and point-wise convolutions to reduce the number of parameters, resulting in a lighter and faster model compared to traditional architectures. This makes it suitable for tasks where real-time processing is crucial, such as object detection and recognition. ResNet-50 is a popular Residual Neural Network (ResNet) architecture variant designed to overcome the vanishing gradient problem in deep learning. It utilizes skipping connections and residual blocks to train deep neural networks efficiently. This architecture has been widely used in various computer vision tasks, including image classification and object detection. Fig. 3 shows the block diagram of the proposed model.

Dense NET-201 is a deep learning architecture that uses densely connected convolutional layers to facilitate feature extraction. It connects each layer to every other layer in a feed-forward fashion, creating more complex features and improving accuracy. This architecture has shown promising results in tasks such as image classification and object recognition. ResNet-18 is another variant of the Residual Neural Network architecture designed for efficient training and classification of images. It consists of 18 convolutional layers with fewer parameters than deeper architectures like ResNet-50. This makes it suitable for tasks requiring a lighter and faster model without compromising performance. The support vector machine (SVM) classifier is a popular machine learning algorithm for classification tasks. It works by finding the optimal hyperplane to separate different classes in a dataset. In image classification, SVM classifiers are often employed to classify features extracted from images. They can handle high-dimensional data and are robust to outliers, making them suitable for complex image recognition tasks. Classification is categorizing or labeling data into different classes or categories. In computer vision, classification is employed to identify the content of an input image and assign it to a specific class or category. Different classification models, such as deep learning architectures or traditional machine learning algorithms, can be used for this task, depending on the specific requirements and data characteristics. Accurate classification is essential in various applications, such as object and pattern recognition, face detection, and medical diagnosis.

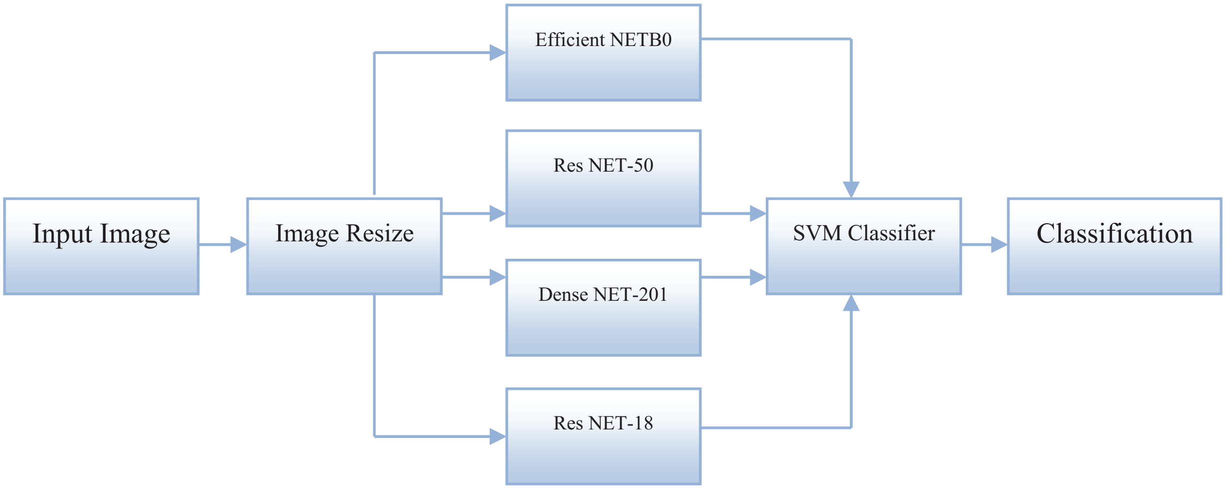

Figure 3: Proposed block diagram

3.1 Dataset Description

The cancer classification dataset was separated as follows:

- No. of samples: 6423
- Training: 80% (5139 samples)
- Testing: 20% (1284 samples)

3.2 Pre-Processing

The pre-processing stage in cancer image classification involves numerous steps. Firstly, the data is retrieved and cleaned to remove any underlying noise or artifacts from the digital images.

$$\frac{df}{de} = \lim_{g \to 0} \frac{g\,(e+f) - g\,(e)}{h} \tag{1}$$

$$\frac{df}{de} = \lim_{g \to 0} \frac{\left(\frac{1}{e+f}\right) - \frac{1}{e}}{h} \tag{2}$$

It helps to ensure that the AI system can obtain a pure representation of the data and interpret it accurately.

$$\frac{df}{de} = \lim_{g \to 0} \frac{\left(\frac{1}{e+f} * \frac{e}{e}\right) - \left(\frac{1}{e} * \frac{e+f}{e+f}\right)}{h} \tag{3}$$

$$\frac{df}{de} = \lim_{g \to 0} \frac{\left(\frac{e-e-g}{(e+g)e}\right)}{h} \tag{4}$$

After that, the images are transformed into subsequent lower dimensions for further classification purposes. An example in this case can be the RGB to Gray transformation, which is often used for tumor segmentation.

$$\frac{df}{de} = \lim_{g \to 0} \frac{\left(\frac{-g}{(e+g) * e}\right)}{h} \tag{5}$$

$$\frac{df}{de} = \lim_{g \to 0} \frac{\left(\frac{-1}{(e+f) * e}\right)}{h} \tag{6}$$

$$f = \frac{-1}{e^2} \tag{7}$$

The input images are then resized to a constant size or cropped to remove any unwanted background. Following that, the images are normalized to ensure that the AI system can differentiate between subtle differences in color intensity. Finally, the images are augmented, creating synthetic versions of the existing images to increase the data size. It helps to make the model robust against over-fitting. The AI system can effectively and accurately classify the cancerous cells in the images by implementing these pre-processing steps.

3.3 Feature Extraction

Feature extraction, commonly known as feature engineering, constitutes a fundamental stage in developing an XAI framework for categorizing cancer images. The procedure above entails extracting pertinent information from individual images to facilitate the detection and classification of cancer using an artificial intelligence model.

$$G = \lim_{e \to 0} \left(\frac{g(f+e) - g(f)}{e}\right) \tag{8}$$

$$G = \lim_{e \to 0} \left(\frac{\left(\frac{1}{f+e-1}\right) - \left(\frac{1}{f-1}\right)}{e}\right) \tag{9}$$

Many algorithms are employed to extract specific characteristics from images during feature extraction. These algorithms include edge detection, color segmentation, contour extraction, and morph metrics. These techniques enable extracting features such as shapes, structures, and colors from the images.

$$G = \lim_{e \to 0} \left(\frac{\left(\frac{1}{f+e-1}\right)\left(\frac{f-1}{f-1}\right) - \left(\frac{1}{f-1}\right)\left(\frac{f+e-1}{f+e-1}\right)}{e}\right) \tag{10}$$

$$G = \lim_{e \to 0} \left(\frac{(f-1) - (f+e-1)}{e(f-1)(f+e-1)}\right) \tag{11}$$

$$G = \lim_{e \to 0} \left(\frac{(-e)}{e(f-1) * (f+e-1)}\right) \tag{12}$$

The extracted features are then combined to form feature vectors used to train the machine-learning model. For example, the feature vectors can be employed to train an SVM to recognize different types of tumors. Feeding the feature vectors into the model allows it to distinguish between similar-looking tumors based on their visual characteristics. Thus, an XAI model for cancer image classification can explain its decisions through the extracted features.

3.4 Segmentation

The segmentation stage of an XAI model in cancer image classification is responsible for localizing the tumor or tumor-affected area in the image and describing the lesion characteristics.

$$G = \lim_{e \to 0} \left(\frac{(-1)}{(f-1) * (f+e-1)} \right) \tag{13}$$

$$G = \left(\frac{(-1)}{(f-e)^2} \right) \tag{14}$$

It is achieved by employing computer vision techniques such as image segmentation and object recognition.

$$g''(e) = \lim_{g \to 0} \frac{g(e+f) - g(e)}{h} \tag{15}$$

$$g''(e) = \lim_{g \to 0} \frac{f^{e+g} - f^e}{h} = \lim_{g \to 0} \frac{f^e f^g - f^e}{h} \tag{16}$$

This stage will involve identifying the regions of interest in the image that are likely to contain cancerous lesions and separating them from other regions in the image that contain healthy tissue.

$$g''(e) = \lim_{g \to 0} \frac{f^e (f^g - 1)}{h} = f^e \lim_{g \to 0} \frac{f^h - 1}{h} \tag{17}$$

$$g''(e) = f^e \lim_{g \to 0} \frac{f^g - 1}{h} \tag{18}$$

This separation is then used as an input to the classifier stage and as a feature to explain the classification result. Specifically, the segmentation stage will provide a more detailed feature vector of the region of interest that is used by the classifier for more accurate predictions. This feature vector may include information about the texture of the region, the shape, and the size of the region. Once these features are extracted, they can be utilized to explain the classification results of the classification stage.

3.5 Detection

The detection stage of the explainable AI model in cancer image classification involves automatically segmenting images to detect, identify, and localize cancerous regions or tumors. This process requires computer vision techniques and deep learning models, which can extract useful features from the images and accurately identify cancerous regions.

$$h = f^e - 1 => f^e = h + 1 => e = \log_g (h+1) \tag{19}$$

$$As, f \to 0 => \log_f (h+1) \to 0 => h + 1 \to 1 => h \to 0 \tag{20}$$

$$g''(e) = f^e \lim_{g \to 0} \frac{g^h - 1}{h} = f^e \lim_{g \to 0} \frac{g + 1 - 1}{\log_f (g+1)} \tag{21}$$

In this process, the model can be trained to identify patterns in the data, such as changes in texture or color, which can be employed to detect the presence of cancer.

$$g''(e) = f^{e} \log_{g} f \tag{22}$$

$$\ln(f) = \ln\left(g^{e}\right) \tag{23}$$

$$\ln(f) = e * \ln(g) \tag{24}$$

Once the tumors are detected, they can be localized and classified using image segmentation techniques.

$$\frac{1}{f} * \frac{df}{de} = \ln(g) \tag{25}$$

$$\frac{df}{de} = f * \ln(g) \tag{26}$$

Substituting Eq. (22) in Eqs. (26), (27) can be derived

$$\frac{df}{de} = g^{e} * \ln(g) \tag{27}$$

Finally, the model can be employed to generate explainable AI results, such as highlighting which part of which image is cancerous and providing a summary of the classifications and results.

3.6 Classification

The classification stage of XAI in cancer image classification is a process that uses a trained machine-learning model to assign a class label to an image based on its characteristics. This classification is based on the features extracted from the image and then classified according to their relevance to the analyzed cancer type.

$$y(e_1) = \frac{1}{Y} \sum_{e_2} \alpha(e_1, e_2) * \sum_{e3} \alpha(e_1, e_3) * \sum_{e_4} \alpha(e_2, e_4) * \sum_{e_5} \alpha(e_3, e_5) * \sum_{e_6} \alpha(e_2, e_5, e_6) \tag{28}$$

$$y(e_1) = \frac{1}{Y} \sum_{e_2} \alpha(e_1, e_2) * \sum_{e3} \alpha(e_1, e_3) * \sum_{e_4} \alpha(e_2, e_4) * \sum_{e_5} \alpha(e_3, e_5) * f_6(e_2, e_5) \tag{29}$$

In some cases, XAI models also perform feature selection, meaning they select only a subset of the available features for the classification.

$$y(e_1) = \frac{1}{Y} \sum_{e_2} \alpha(e_1, e_2) * \sum_{e3} \alpha(e_1, e_3) * f_5(e_2, e_3) * \sum_{e_4} \alpha(e_2, e_4) \tag{30}$$

$$y(e_1) = \frac{1}{Y} \sum_{e_2} \alpha(e_1, e_2) * f_4(e_2) \sum_{e3} \alpha(e_1, e_3) * f_5(e_2, e_3) \tag{31}$$

During the classification process, the model builds a prediction based on the extracted features to determine the most likely class label for the image. The variables such as size, shape, and texture can be employed to refine the model's classification further.

$$y(e_1) = \frac{1}{Y} \sum_{e_2} \alpha(e_1, e_2) * f_4(e_2) * f_3(e_1, e_2) \tag{32}$$

$$y(e_1) = \frac{1}{Y} f_2(e_1) \tag{33}$$

Finally, an evaluation metric measures the model's performance and provides an overall score. This metric will typically comprise accuracy, precision, recall, and other accuracy metrics.

3.7 Proposed Algorithm

XAI algorithms in cancer image classification can help doctors improve diagnostic accuracy by providing a reasoning path and insight into why certain decisions are made. XAI algorithms analyze patterns to identify and classify cancerous cells in images. As part of the classification process, XAI algorithms can explain why a cell is classified as benign or malignant. The proposed flow diagram is shown in Fig. 4.

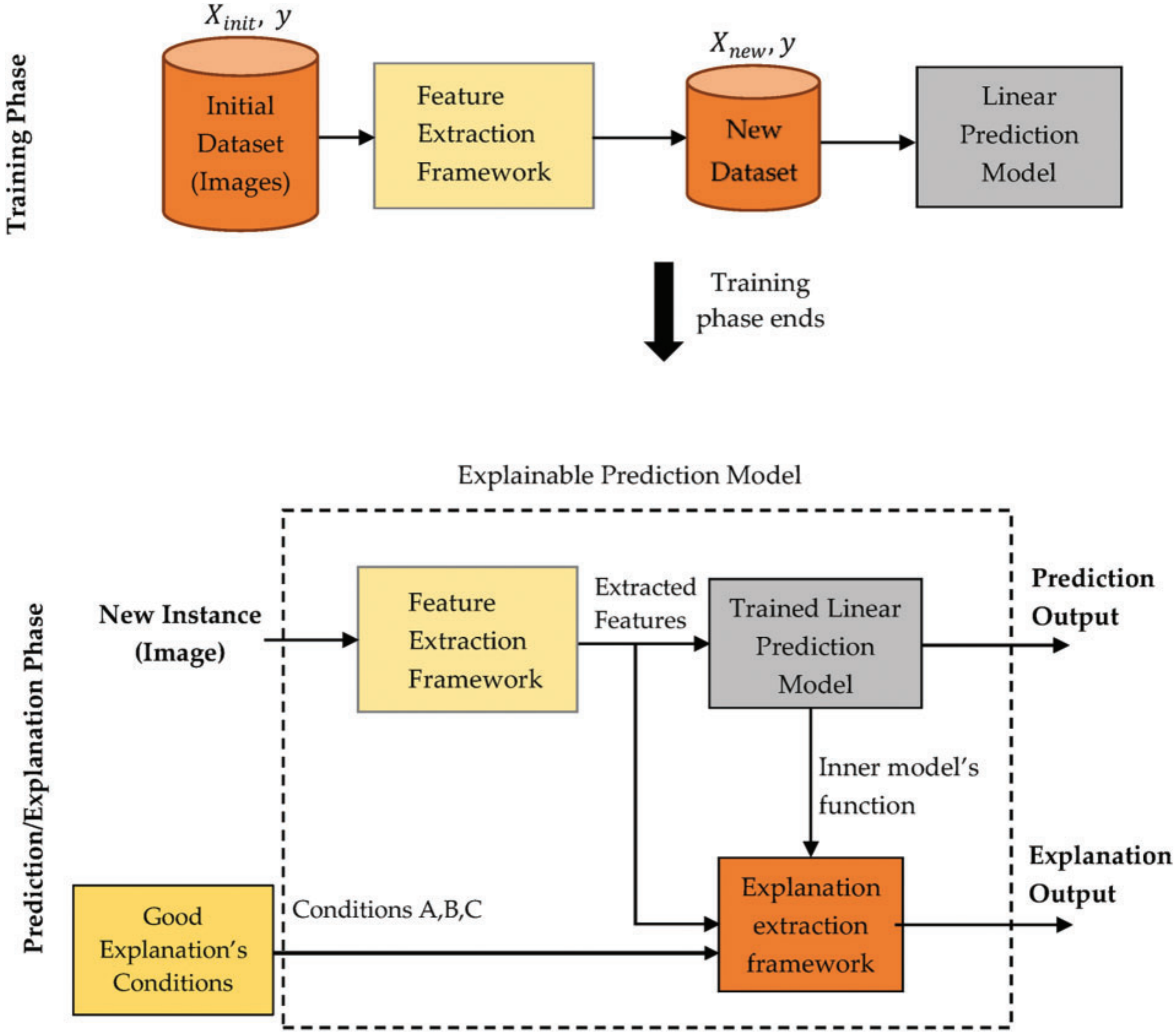

Figure 4: Proposed flow diagram

It helps medical professionals gain further insight into the features that indicate if cancer is present and to adjust classification accordingly. Additionally, XAI algorithms are employed to assess the credibility of the classification in the context of the input image. With its interpretability and transparency capabilities, XAI can be utilized to explain the processes through which an image is classified and how the model's accuracy can be improved.

- The first step is to collect and label cancer images for training and testing, including detailed annotations for each image. It can be done manually or using automated algorithms to detect cancer regions.
- The subsequent phase involves the development of a deep learning model, an artificial intelligence algorithm capable of categorizing photos based on annotated data. The model undergoes training using annotated data, enabling it to perform cancer picture classification subsequently.
- The user's text lacks academic language and structure. Once the training of the model is completed, the use of the XAI algorithm becomes feasible to discern the significant components of the image that play a role in influencing the choice made by the model. The regions above, referred to as feature maps, offer valuable insights into the image's primary features that hold significance for classification.
- The user's text does not contain any information to rewrite. The XAI algorithm ultimately generates explanations that indicate the rationale behind a specific decision made by the model. The analysis can illustrate how the image's characteristics align with the model's choice, offering a more comprehensive comprehension of the underlying process.

Incorporating convolution layers in XAI models for cancer picture classification enhances interpretability by facilitating the identification and extraction of task-relevant characteristics. The operational mechanism of the suggested model is illustrated in Fig. 5.

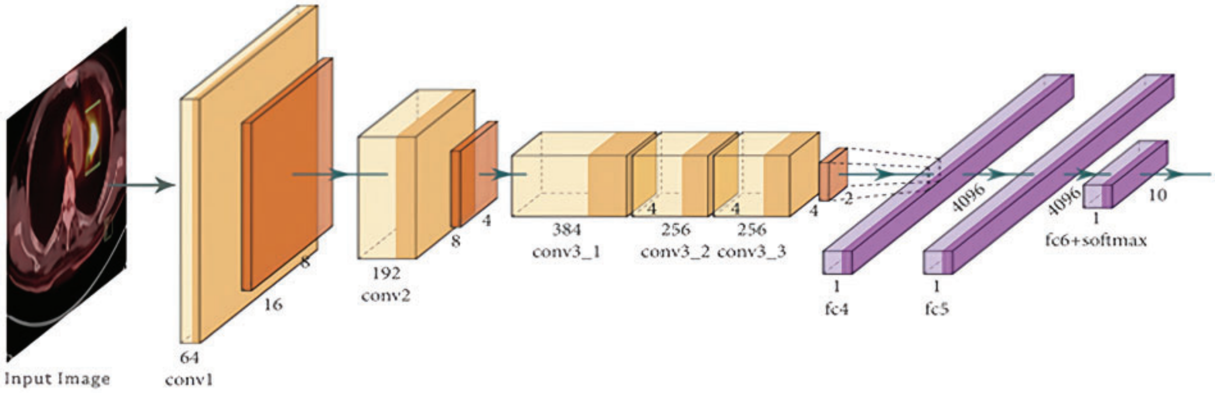

Figure 5: Functional working of the proposed model

Unlike normal convolutional neural networks, XAI models typically contain multiple convolutional layers to identify complex patterns from the images.

$$E = \sum_{f} g_{f} = \sum_{f} g^{-e} * g^{-f} \tag{34}$$

$$E = g^{-e} * \sum_{f} g^{-f} \tag{35}$$

$$g^{-e} = \frac{E}{\sum_{f} g^{-f}} \tag{36}$$

The convolution layers apply a set of learnable filters to the input image, which helps identify distinct features of the image. For example, a convolution layer can be employed to identify the presence of cancerous cells by looking for specific characteristics in the image, such as certain shapes or colors.

$$F = \frac{E}{\sum_{f} g^{-f}} \tag{37}$$

$$E = E_0 * g^{-ft} \tag{38}$$

where, $t = T_p$; $E = \frac{E_0}{2}$.

The performance of the convolution layer-2 in the proposed model is crucial because it provides insight into how the model can identify and classify cancer images. This layer plays a vital role in extracting features from the input images, which the model then uses to make predictions. Analyzing this layer's performance can help better understand the key features and patterns the model uses to distinguish between cancerous and non-cancerous cells. This information is essential to build transparency and trust in the model and to identify any biases or weaknesses that may be present in the classification process. Ultimately, a strong and accurate performance from the convolution layer-2 is crucial for a reliable and explainable model for cancer image classification. The primary function of a pooling layer is to downsampling the input image by reducing its spatial dimensions while retaining the most relevant features. This is achieved through a process called pooling, where the layer takes the maximum or average value from a specific region of the input image. The pooling layer helps to reduce the number of parameters in the network, making it more efficient. In addition, the model can focus on the essential details of the image while ignoring any irrelevant or noisy elements by selecting the most critical features. This allows the model to have a more comprehensive understanding of the images and make more accurate predictions.

After extracting these features from the image, they can subsequently be employed to classify the image as either including malignant cells or not.

$$g''(e) = e^{e} * \frac{1}{\ln f} \tag{39}$$

In addition, by examining the outcomes derived from the convolution layer, it becomes feasible to acquire a deeper understanding of the specific characteristics inside the image that were utilized to facilitate the classification process. Therefore, this enhances the comprehensibility and interpretability of the model.

$$\frac{E_0}{2} = E_0 * g^{-ft} \tag{40}$$

$$2 = g^{-fT_p} \tag{41}$$

The primary function of the first convolutional layer in an XAI model designed for cancer picture classification is to extract relevant information from the input image. The model benefits from its ability to detect edges, lines, and shapes, enhancing its capacity to accurately perceive the shape and structure of cancer cells depicted in the image. Enhancing the precision of the cancer picture categorization task is facilitated.

$$\log_e 2 = fT_p \tag{42}$$

$$E = F * \alpha \tag{43}$$

In addition, the first convolution layer aids in diminishing the volume of input data that the model must handle, hence mitigating the processing resources demanded by the model. Consequently, the model's total performance is enhanced.

$$E = \frac{4}{3}\pi H^3 * \alpha \tag{44}$$

The performance of convolution layer-2 for the XAI model for cancer image classification is significant in understanding how the model classifies cancer images. By looking at the layer's output, we can gather insights as to why the model makes its decisions.

$$F_e = \sqrt{\frac{2A * \left(\frac{4}{3}\pi H^3 * \alpha\right)}{G}} = \sqrt{\frac{8A\pi H^2 * \alpha}{3}} \tag{45}$$

$$F_e = 2H * \sqrt{\frac{2\pi A\alpha}{3}} = H * \sqrt{\frac{8\pi A\alpha}{3}} \tag{46}$$

The convolution layer-2 is responsible for automatically extracting features from the input image, where each feature is a combination of pixels that are "activated" by the convolution filter. This layer can be considered a feature detector that looks for edges, lines, or textures within the cancer cells.

$$f_a = \{f_1, f_2, f_3 \ldots\ldots, f_n\} \tag{47}$$

$$g_b = \{g_1, g_2, g_3 \ldots\ldots, g_n\} \tag{48}$$

Looking at the output from layer-2 can gain insights into what features the model considers essential for classifying the cancer cell. This layer can be fine-tuned to increase the accuracy and robustness of the model's classification accuracy.

$$Q = f_a + g_b \tag{49}$$

$$Q = \{f_1, f_2, f_3 \ldots\ldots, f_n\} + \{g_1, g_2, g_3 \ldots\ldots, g_n\} \tag{50}$$

$$Q_n = \sum_{n=1}^{\infty} f_{a(n-1)} + g_{b(n-1)} \tag{51}$$

The weights associated with the convolution filters can be adjusted depending on the features in the input image to which the model should pay more attention. Doing so can enable the model to extract the necessary features from the image more accurately, allowing for a more accurate classification of cancer cells.

$$g(e) = e_1(f) * e_2(f) \tag{52}$$

$$g(h) = h_1(g) * h_2(g) \tag{53}$$

The convolution layer-3 of the XAI model for cancer image classification performs the role of feature extraction by mining data for patterns and extracting features from the images. It takes the raw images as inputs and performs convolution operations to detect edges, shapes, and other meaningful patterns.

$$Q = \left\{ \frac{g(e) + g(f)}{g(e,f)} \right\} \tag{54}$$

$$G = \left\{ \frac{(e_1(f) * e_2(f)) + (g_1(h) * g_2(h))}{e(f,h) * f(g,h)} \right\} \tag{55}$$

The results of the convolution operations are then passed through an activation function, which determines the importance of certain features compared to others. This layer also helps reduce the images' dimensionalities while retaining relevant information. Hence, the model can focus on essential details in the image and ignore irrelevant features. It enables the model to make better and more accurate predictions about the cancer images.

$$g(e) = \{F_1 * h_1(e) + F_2 * h_2(e) + \cdots\cdots + F_d * f_d(e)\} \tag{56}$$

$$F = \int_{e=1}^{n} \frac{a(e_1 * e_n)}{b(e_1 * e_n)} \tag{57}$$

$$h_1(e) = \left\{ \frac{p(e_1)}{q(e_1)} \right\} \tag{58}$$

The inclusion of a pooling layer is a crucial element in the categorization of cancer images. The purpose of this technique is to minimize the number of variables involved in the calculation process, achieved by reducing the spatial dimensions inside the model. Utilizing the pooling layer enables the model to effectively reduce the parameters required for representing equivalent information while maintaining high-performance accuracy.

$$G_1 = \sum_{e=1}^{m} \sum_{f=1}^{n} g_{f1}^{(e1)} - h_{e1}^2 \tag{59}$$

$$G_2 = \sum_{e=1}^{m} \sum_{f=1}^{n} g_{f2}^{(e2)} - h_{e2}^2 \tag{60}$$

It reduces the complexity of the model, making it easier to explain. Additionally, the pooling layer allows the model to identify scale-invariant patterns in the high-level feature maps, which helps differentiate between cancerous and non-cancerous cells.

$$g(G_1|G_2) = \frac{g(G_2|G_1) * g(G_1)}{g(G_2)} \tag{61}$$

Reducing the spatial dimensions of the model makes it easier for the XAI model to identify and explain why the cancerous and non-cancerous cells are different. Furthermore, the pooling layer can be employed to reduce the overfitting of the model, making it more robust and reliable.

$$G = \left| \frac{\left(\vec{s}_1 * \vec{s}_2\right) * r_1 \vec{*} r_2}{\left(\vec{s}_1 * \vec{s}_2\right)} \right| ; where \left| \left(\vec{s}_1 * \vec{s}_2\right) \right| \neq 0 \tag{62}$$

If the values $\left| \left(\vec{s}_1 * \vec{s}_2\right) \right| = 0$

The performance of the Softmax layer for the XAI model for cancer image classification refers to the model's overall accuracy in accurately classifying images into cancer or non-cancer classes.

$$G = \left| \frac{\left(\vec{s}_1\right) * r_1 \overset{\rightarrow}{*} r_2}{\left(\vec{s}_1\right)} \right| \tag{63}$$

The final layer of a Convolutional Neural Network (CNN) model is often the Softmax layer. Its purpose is to provide a probability distribution across the output classes. This is achieved by calculating the normalized exponential of the scores obtained from the preceding layer. The Softmax layer subsequently predicts by identifying the class with the highest probability.

$$G = \frac{\sqrt{(r_{12} * f_1 - g_{12} * a_1)^2 + (f_{12} * g_1 - a_{12} * r_1)^2 + (f_{12} * r_1 - g_{12} * a_1)^2}}{\sqrt{g_1^2 + r_1^2 + f_1^2}} \tag{64}$$

The performance of the Softmax layer is even more critical because it considers the prediction accuracy and gives additional insights into how the model works.

$$r_{12} = r_1 - r_2 \tag{65}$$

$$f_{12} = f_1 - f_2 \tag{66}$$

$$g_{12} = g_1 - g_2 \tag{67}$$

It is done by extracting the activations of each neuron and visualizing the output in terms of what classes it is most and least likely to classify the input image. It helps identify regions in the image that contribute the most to the output. Therefore, the performance of the Softmax layer for the XAI model for cancer image classification is significant because it helps provide additional insights into how the model works and is used to predict the class of the input image.

3.8 Integration Techniques

The proposed XAI framework integrates different techniques to facilitate explainability in AI models. These techniques include end-to-end explainable evaluation, rule-based explanation, and user-adaptive explanation.

3.8.1 End-to-End Explainable Evaluation

The end-to-end explainable evaluation technique evaluates the AI model's performance and generates explanations at different stages of the model's decision-making process. This includes assessing the input data, the model's internal representations, and the final predictions. The explanations are generated using different methods such as feature importance, saliency maps, and local interpretability techniques. The end-to-end explainable evaluation module is implemented using a combination of various libraries and algorithms such as SHAP (Shapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations), and Integrated Gradients.

3.8.2 Rule-Based Explanation

The rule-based explanation technique generates explanations by extracting the rules the AI model uses to make decisions. These rules can be generated using Sequential Covering Rule-based Learning algorithms, Association Rule Mining, or Decision Trees. The rule-based explanations provide a transparent representation of the model's decision-making process and can be easily understood by non-technical users. The rule-based explanation module uses open-source libraries such as Explainable Boosting Machine (EBM) or RuleFit.

3.8.3 User-Adaptive Explanation

The user-adaptive explanation technique tailors the explanations to each user's specific needs and preferences. This is achieved by integrating user feedback into the explanation generation process. The user feedback can include explicit user preferences or implicit feedback from the user's interactions with the AI model. The user-adaptive explanation module is implemented using collaborative filtering, reinforcement learning, or Bayesian optimization techniques.

The XAIM framework also includes a feedback loop that continuously updates the explanations based on new data and user feedback. This ensures that the explanations remain up-to-date and relevant. Additionally, the XAIM framework provides a user interface that allows it to interact with the explanations, provide feedback, and customize them according to their needs.

4 Results

The proposed XAI model was compared to the existing deep learning-based medical image analysis (DLMIA), deep meta-learning models (DMLM), multi-label deep learning classification (MDLC), and Deep Learning Based Medical Imaging Classification (DLMIC).

4.1 Computation of Accuracy

The accuracy of an explainable AI algorithm for cancer image classification is determined by measuring the degree of agreement between its predictions and the actual labels of the cancer images. It is typically performed by comparing the classification results of the AI algorithm to a gold standard or ground truth label associated with the images. Image classification accuracy is determined by dividing the count of accurately categorized photos by the total count of images in the dataset. The accuracy will be expressed as a percentage, with higher values indicating better performance. Table 3 shows the computation of accuracy.

Table 3: Computation of accuracy (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	49.48	54.59	73.62	86.10	97.93
200	48.19	53.84	69.00	82.70	97.83
300	48.44	53.87	69.00	83.06	97.76
400	48.57	54.69	69.47	84.25	97.72
500	48.49	54.78	69.67	84.12	97.68
600	48.50	54.91	69.93	84.10	97.65
700	48.85	55.33	70.56	84.53	97.63

Fig. 6 illustrates the comparison of accuracy. In the context of computational methods, the currently available DLMIA obtained 48.57%, DMLM obtained 54.69%, MDLC reached 69.47%, and DLMIC reached 84.25% accuracy. The proposed XAIM reached 97.72% accuracy.

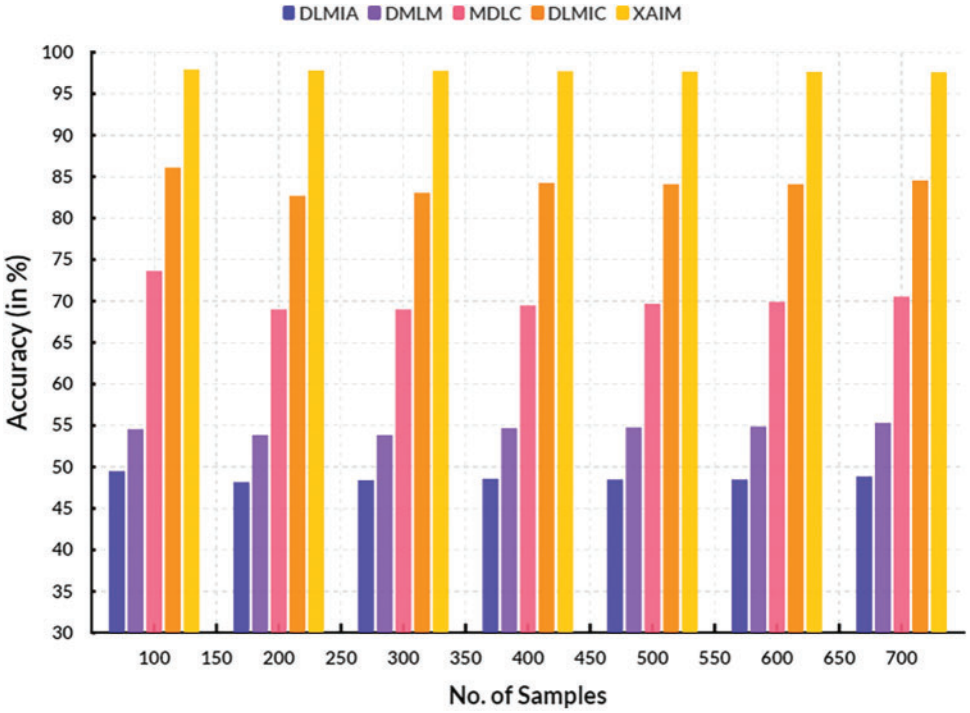

Figure 6: Accuracy

4.2 Computation of Precision

Precision is an essential measure of accuracy for XAI algorithms in cancer image classification. The calculation involves determining the proportion of correctly identified cancer images (true positive, TP) with the overall number of anticipated positive classifications (true positive plus false positive, TP + FP). Accuracy is a metric that quantifies the model's capacity to precisely categorize cancer images as positive or negative. By quantifying the precision of the XAI model, it is possible to assess its ability to detect cancerous conditions in photos accurately. The computation of precision is presented in Table 4.

Table 4: Computation of precision (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	48.62	53.53	70.29	82.78	90.94
200	48.12	53.53	69.20	82.52	90.83
300	47.37	52.70	68.06	81.95	90.77
400	47.37	53.43	68.42	83.09	90.72
500	48.42	54.54	69.95	84.11	90.68
600	48.70	54.94	70.59	84.35	90.65
700	47.98	54.37	70.01	83.70	90.63

Fig. 7 illustrates the comparison of precision. In the context of computational techniques, the currently available Deep Learning for Medical Image Analysis (DLMIA) obtained 47.37%, DMLM obtained 53.43%, MDLC reached 68.42% and DLMIC reached 83.09% precision. The proposed XAIM reached 90.72% precision.

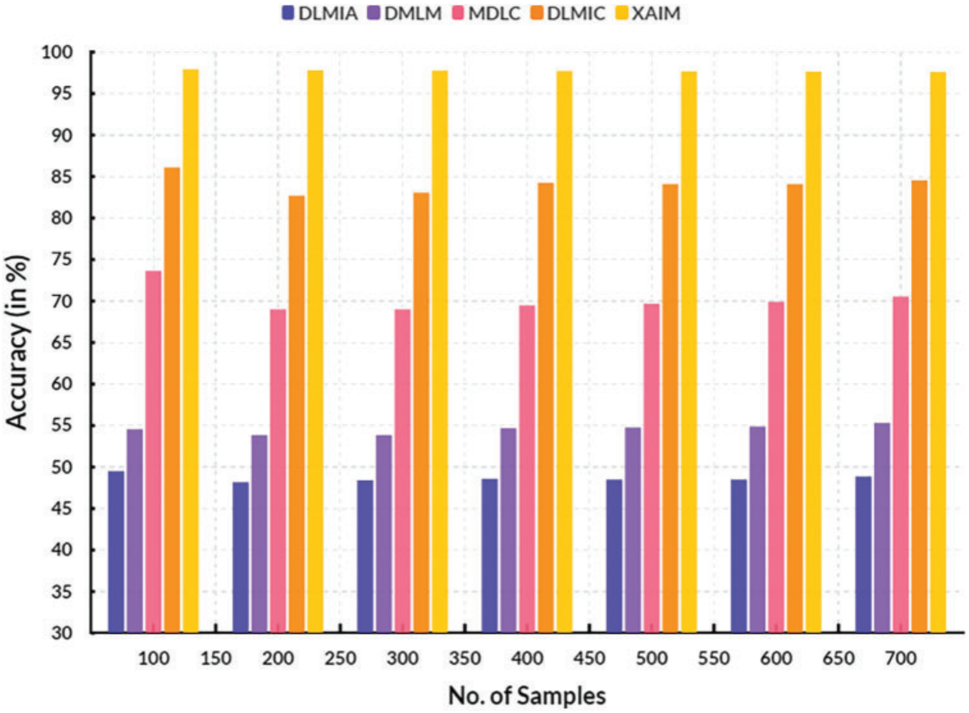

Figure 7: Precision

4.3 Computation of Recall

The recall for explainable AI algorithms in cancer picture classification refers to the ratio of correctly identified actual positive cases to the total number of actual positive instances. The term "it" can also be characterized as the quotient obtained by dividing the count of true positives (instances correctly classified as positive) by the sum of all positive cases (true positives plus false negatives). The recall rate of explainable AI algorithms in the context of cancer picture classification is commonly determined by utilizing a collection of annotated or labeled images. The process entails identifying and documenting all the regions within each image that exhibit positive indications of malignancy, as well as the equivalent regions that exhibit negative indications. Once the annotation process is finalized, the recall rate can be determined by dividing the number of true positives detected in the photos by the total count of all positive cases. This computation enables the evaluation of the AI algorithm's performance in accurately detecting all affirmative cases. The computation of recall is listed in Table 5.

Table 5: Computation of recall (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	56.91	45.00	84.80	81.62	93.94
200	58.57	50.86	77.96	87.80	93.83

(Continued)

Table 5 (continued)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
300	59.02	49.72	76.67	89.29	93.77
400	54.33	50.86	74.53	92.53	93.72
500	53.94	51.74	76.10	91.81	93.68
600	54.10	52.94	77.72	91.68	93.65
700	54.84	54.59	79.52	92.95	93.63

Fig. 8 illustrates the contrast of recall. In the context of computational methods, the currently available Deep Learning for Medical Image Analysis (DLMIA) obtained 54.33%, DMLM obtained 50.86%, MDLC reached 74.53% and DLMIC reached 92.53% recall. The proposed XAIM reached 93.72% recall.

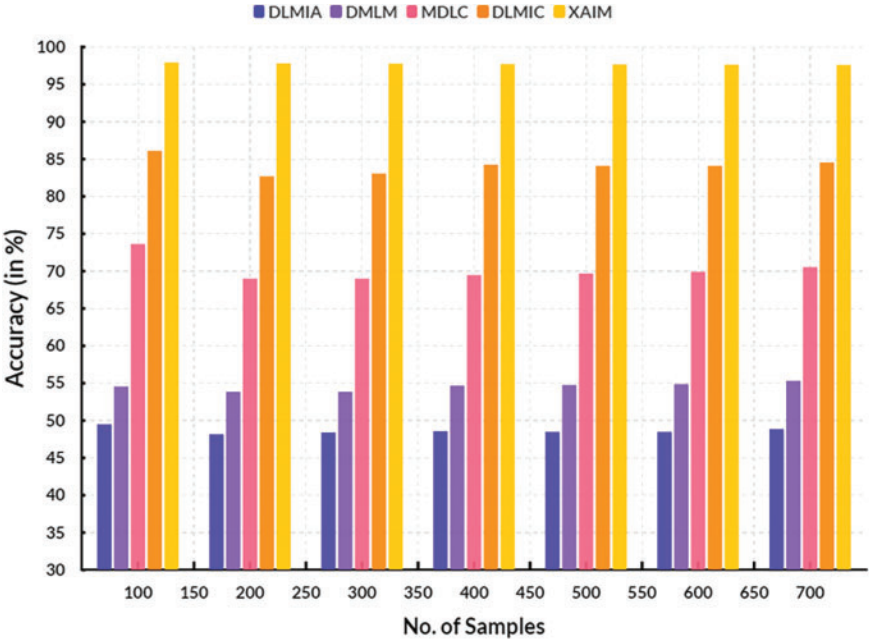

Figure 8: Recall

4.4 Computation of F1-Score

The F1-score in cancer picture classification for an XAI method is computed as the harmonic mean of precision and recall, representing a combined evaluation of two distinct metrics. Precision is a metric that quantifies the algorithm's capacity to correctly classify photos into their respective classes with high accuracy. On the other hand, recall evaluates the algorithm's effectiveness in successfully identifying all pertinent images. This statistic provides a more comprehensive assessment of the algorithm's performance by considering both the successfully categorized cases and the misclassified examples. The calculation of the F1-score involves determining the precision and recall values of the XAI algorithm, followed by computing the harmonic mean of these values to obtain the final score. The F1-score is obtained by calculating the precision and recall scores, summing their reciprocals, and dividing the result by two. The F1-score serves as a valuable statistic in assessing the efficacy of an XAI algorithm in the classification of cancer images. Considering the algorithm's precision and recall provides a more comprehensive evaluation of its overall performance. Table 6 lists the calculation of the F1-score.

Table 6: Computation of F1-score (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	52.97	49.46	69.55	82.20	96.94
200	53.68	52.22	69.76	85.24	96.83
300	53.61	51.24	68.50	85.78	96.77
400	51.00	52.16	67.31	88.08	96.72
500	51.27	53.16	68.89	88.14	96.68
600	51.49	53.95	70.14	88.18	96.65
700	51.55	54.48	71.05	88.59	96.63

Fig. 9 presents a visual representation of the comparison of F1-scores. In computational techniques, the currently available Deep Learning for Medical Image Analysis (DLMIA) obtained 51.00%, DMLM obtained 52.16%, MDLC reached 67.31% and DLMIC reached 88.08% F1-score. The proposed XAIM reached 96.72% F1-score.

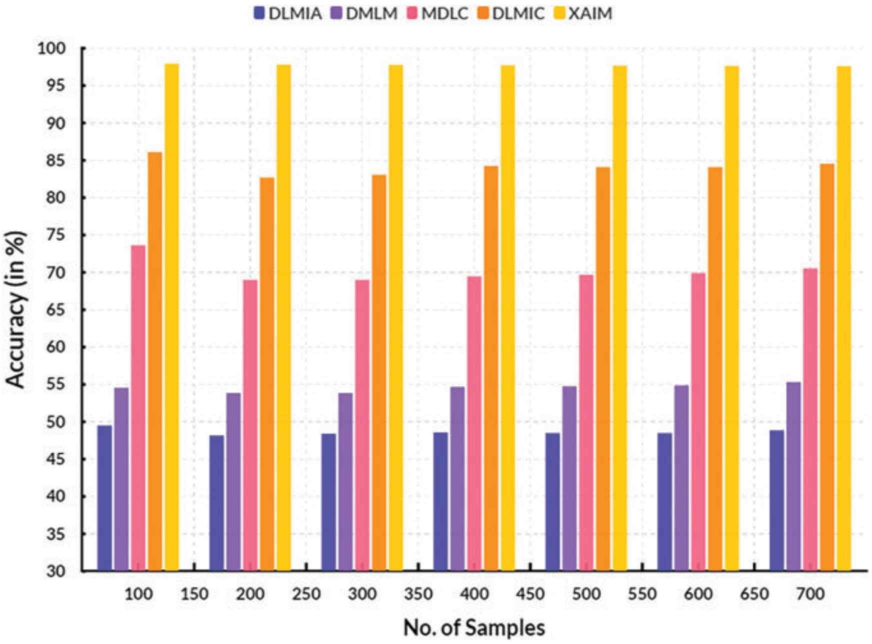

Figure 9: F1-score

4.5 Computation of False Discovery Rate

The False Discovery Rate (FDR) is a metric to evaluate the efficacy of explainable AI algorithms in categorizing cancer images. This metric aims to quantify the occurrence of false positive errors generated by the AI system while categorizing photos as either malignant or non-cancerous. The calculation of the False Discovery Rate (FDR) involves dividing the count of False Positives (FP) by the sum of the false positives and True Positives (TP).

$$\mathrm{FDR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TP}) \tag{68}$$

A high False Discovery Rate (FDR) indicates a diminished level of accuracy in the AI system's ability to correctly classify malignant images as cancerous and non-cancerous images as non-cancerous. A decrease in False Discovery Rates (FDRs) indicates enhanced precision in AI algorithms for correctly classifying malignant pictures as cancerous. Table 7 presents the calculation of the False Discovery Rate.

Table 7: Computation of false discovery rate (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	25.78	19.23	22.88	25.41	14.02
200	24.11	18.10	19.95	24.15	11.55
300	22.16	17.75	18.41	22.26	10.75
400	20.17	15.80	16.38	21.06	9.55
500	17.59	15.03	15.48	29.50	8.91
600	15.60	14.65	13.51	17.75	7.65
700	13.58	13.52	12.04	16.82	6.63

Fig. 10 illustrates the comparative analysis of the False Discovery Rate. In computational methods, the currently available DLMIA obtained 20.17%, DMLM obtained 15.80%, MDLC reached 16.38%, and DLMIC reached 21.06% False Discovery Rate. The proposed XAIM achieved a 9.55% False Discovery Rate.

4.6 Computation of False Omission Rate

False Omission Rate (FOR) is a performance measure for explainable AI algorithms in cancer image classification. It is a metric to assess how successful an AI algorithm is at correctly classifying cancer images. It is calculated by taking the percentage of missed cancer image classification cases out of the total number of cancer image examples. The FOR typically ranges between 0 and 1, with 0 indicating perfect accuracy and 1 indicating that all cases are misclassified. A higher FOR indicates poor performance. Table 8 shows the computation of the False Omission Rate.

Fig. 11 illustrates the contrast of the False Omission Rate. In the context of computational techniques, the currently available Deep Learning for Medical Image Analysis (DLMIA) obtained 20.79%, DMLM achieved 20.41%, MDLC reached 16.22% and DLMIC reached 21.38% False Omission Rate. The proposed XAIM reached a 9.66% False Omission Rate.

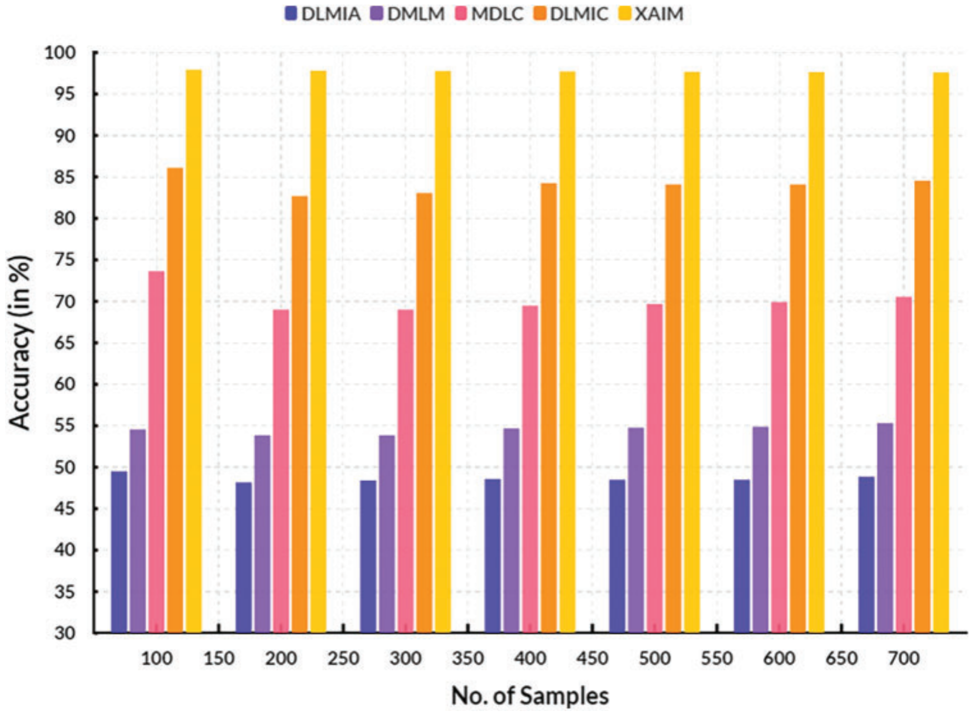

Figure 10: False discovery rate

Table 8: Computation of false omission rate (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	24.19	25.30	21.63	26.05	12.85
200	22.56	23.56	20.05	24.63	11.56
300	22.08	21.22	17.85	23.37	10.55
400	20.79	20.41	16.22	21.38	9.66
500	18.68	18.12	15.08	18.91	9.29
600	17.19	16.19	12.88	17.47	7.65
700	15.38	14.46	11.73	15.75	7.28

4.7 Computation of Diagnosis Odd Ratio

The Diagnostic Odd Ratio (DOR) measures how accurately an XAI algorithm can classify cancer images compared to human experts. DOR measures the odds that a misclassification is made by the algorithm when compared to a gold standard. This measure is meant to assess the accuracy of AI algorithms in comparison to the accuracy of a human expert. The AI algorithm is compared to the gold standard (human expert) by producing two sets of diagnoses (positive and negative) to compute the Diagnostic Odd Ratio (DOR). The ratio of correct positive predictions by the AI is divided by the rate of correct negative predictions made by the AI concerning the gold standard. The DOR can be defined as the ratio of the likelihood of an accurate positive prediction made by the AI to the probability of a correct negative prediction. The formula utilized for the diagnostic odds ratio (DOR) computation is as follows:

$$\text{DOR} = (\text{TP}/\text{FN}) \,/\, (\text{FP}/\text{TN}) \tag{69}$$

where TP is true positives (correct positive predictions by AI), FN is false negatives (missed positive predictions by AI), FP is false positives (incorrect positive predictions by AI), and TN is true negatives (correct negative predictions by AI). Researchers can assess the accuracy of an AI algorithm in comparison to a gold standard by computing the DOR. A higher DOR indicates a higher accuracy. Table 9 shows the computation of the Diagnostic Odd Ratio.

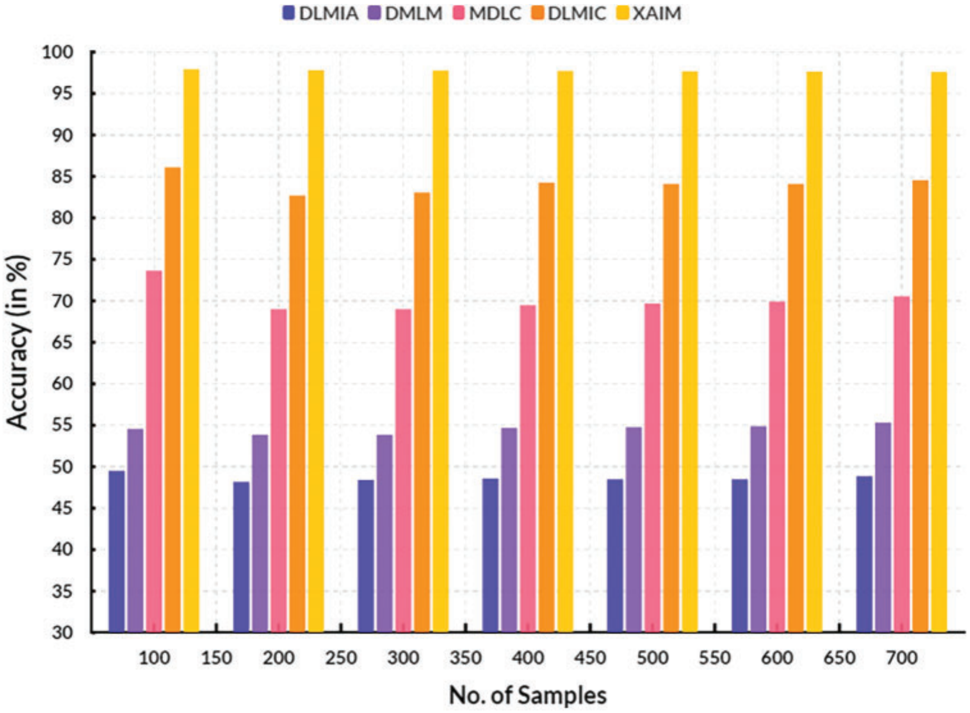

Figure 11: False omission rate

Table 9: Computation of diagnostic odd ratio (in %)

No. of samples	DLMIA	DMLM	MDLC	DLMIC	XAIM
100	54.08	61.20	71.47	85.04	94.85
200	52.59	59.23	69.05	82.84	92.86
300	51.79	58.10	68.64	82.04	91.66
400	49.46	56.89	67.04	81.37	91.18
500	48.45	56.52	64.72	79.94	89.75
600	47.81	54.99	63.47	78.85	88.59
700	47.15	54.49	60.74	78.37	87.82

Fig. 12 illustrates the comparison of the Diagnostic Odd Ratio. In computational techniques, the currently available Deep Learning for Medical Image Analysis (DLMIA) obtained 49.46%, DMLM achieved 56.89%, MDLC reached 67.04% and DLMIC reached 81.37% Diagnostic Odd Ratio. The proposed XAIM reached 91.18% Diagnostic Odd Ratio.

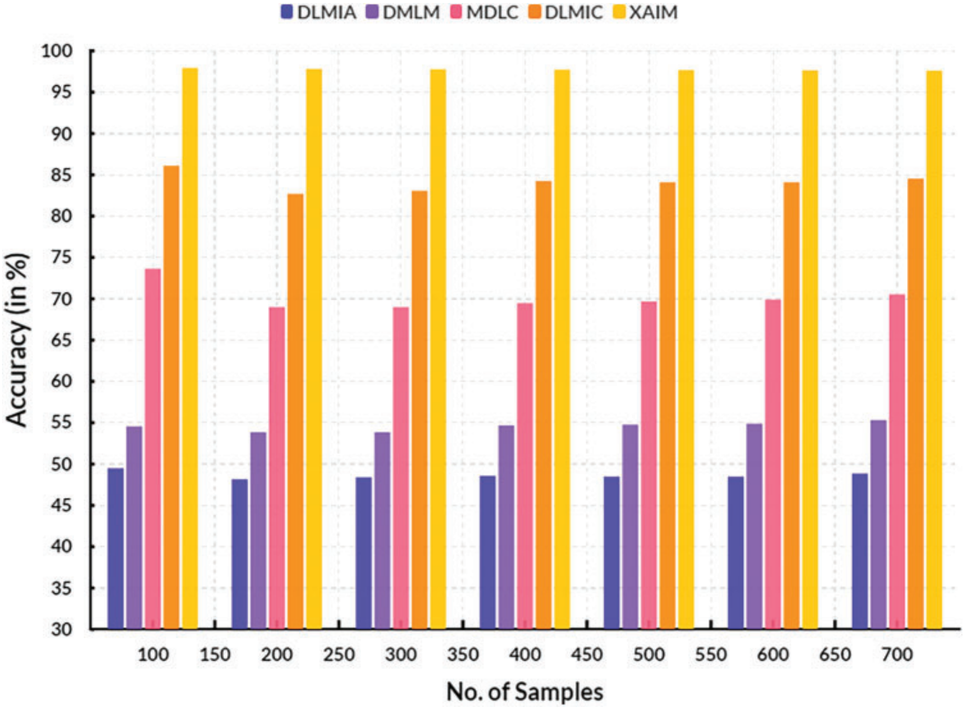

Figure 12: Diagnostic odd ratio

5 Discussion

XAI is a specific branch of AI developed to help healthcare professionals comprehend the fundamental reasoning behind AI-driven decision support systems. The potential of XAI to enhance healthcare decision support systems can yield enhanced patient outcomes, increased accuracy in diagnoses, and the facilitation of more ethical decision-making processes.

5.1 Convergence of Performance

The convergence of performance for explainable AI algorithms in cancer imagery classification is the improvement of accuracy and predictive power as these algorithms are tested against an increasing amount of data. In such AI-based applications, the goal is to bring all data into a common format and represent it understandably to create valuable knowledge that can be used in an automated form. As AI algorithms are constantly fed more data, they can train and adjust to learn more accurately. When the number of features available for AI algorithms is increased, they can detect finer nuances within the data, enabling them to produce a more precise classification. Table 10 shows the convergence of performance.

Table 10: Convergence of performance (in %)

Parameters	DLMIA	DMLM	MDLC	DLMIC	XAIM
Accuracy (A)	48.57	54.69	69.47	84.25	97.72
Precision (P)	47.37	53.43	68.42	83.09	90.72
Recall (R)	54.33	50.86	74.53	92.53	93.72
F1-score (F1)	51.00	52.16	67.31	88.08	96.72
False discovery rate (FDR)	20.17	15.80	16.38	21.06	9.55
False omission rate (FOR)	20.79	20.41	16.22	21.38	9.66
Diagnostic odd ratios (DOR)	49.46	56.89	67.04	81.37	91.18

Fig. 13 illustrates the convergence of performance, while Fig. 14 depicts the convergence of performance specifically for the proposed XAIM. In terms of comparison, the XAIM model achieved an accuracy of 97.72%, precision of 90.72%, recall of 93.72%, F1-score of 96.72%, false discovery rate (FDR) of 9.55%, false omission rate (FOR) of 9.66%, and diagnostic odds ratio (DOR) of 91.18%. XAI refers to a specific domain within AI that concentrates on developing AI systems capable of exhibiting transparency in their internal mechanisms and providing correct explanations for their decision-making processes. The significance of this matter is particularly pronounced within the realm of medical image analysis, particularly in the domain of cancer picture categorization. As an illustration, contemporary deep learning algorithms employed in cancer picture classification have demonstrated the ability to accurately identify lesions with significant precision. However, they have not adequately explained the rationale behind designating a specific image as malignant. XAI algorithms can deduce a collection of interpretable characteristics that have significance in picture recognition. As a result, these algorithms can clarify the classification outcome, facilitating medical professionals in comprehending the underlying rationale behind the model's decision-making process.

5.2 BLA Analysis

XAI-assisted systems can provide information on cancer progression, helping physicians identify the right moment to change the course of action. It can help researchers design better clinical trials by providing objective evaluation and predictions.

Benefits:

- XAI can make it easier to understand the cancer image classification decisions made by the algorithm.
- A clinician can more easily understand the predictions generated by the algorithm by offering deeper insights into the factors and features the algorithm considers.
- The insights offered by XAI can also be utilized to improve the algorithm's accuracy by further optimizing its parameters and variables.

Limitations:

- XAI algorithms can be slightly complex, making them difficult to interpret.
- It cannot be easy to draw accurate conclusions from the algorithms' insights due to the inherent uncertainties of the data and algorithms.
- The algorithm must be tested carefully before being used in a real-world context.

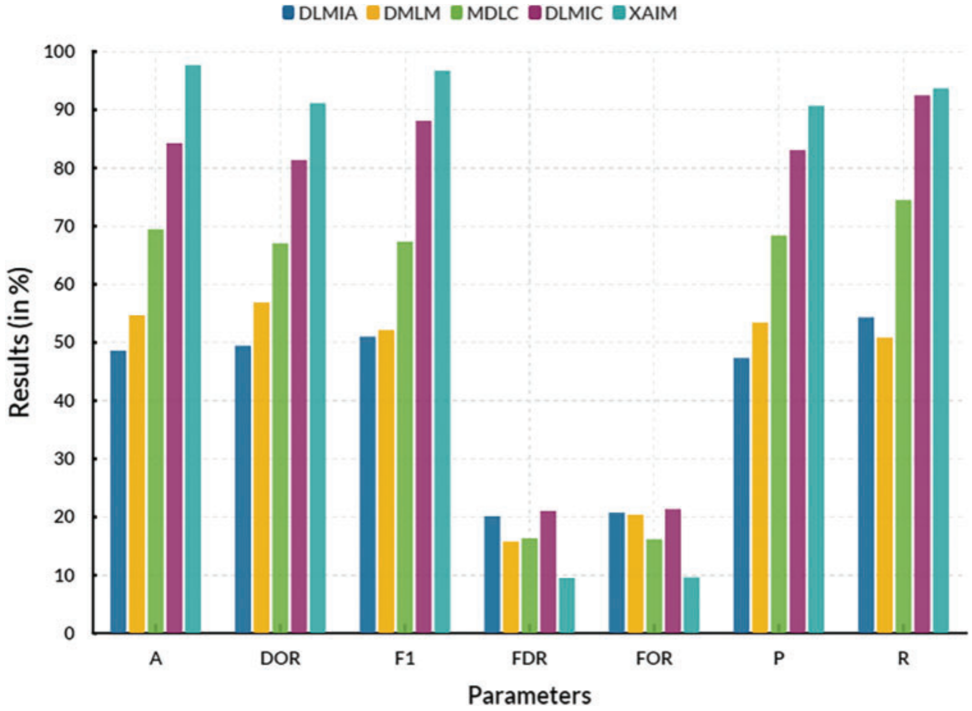

Figure 13: Convergence of performance

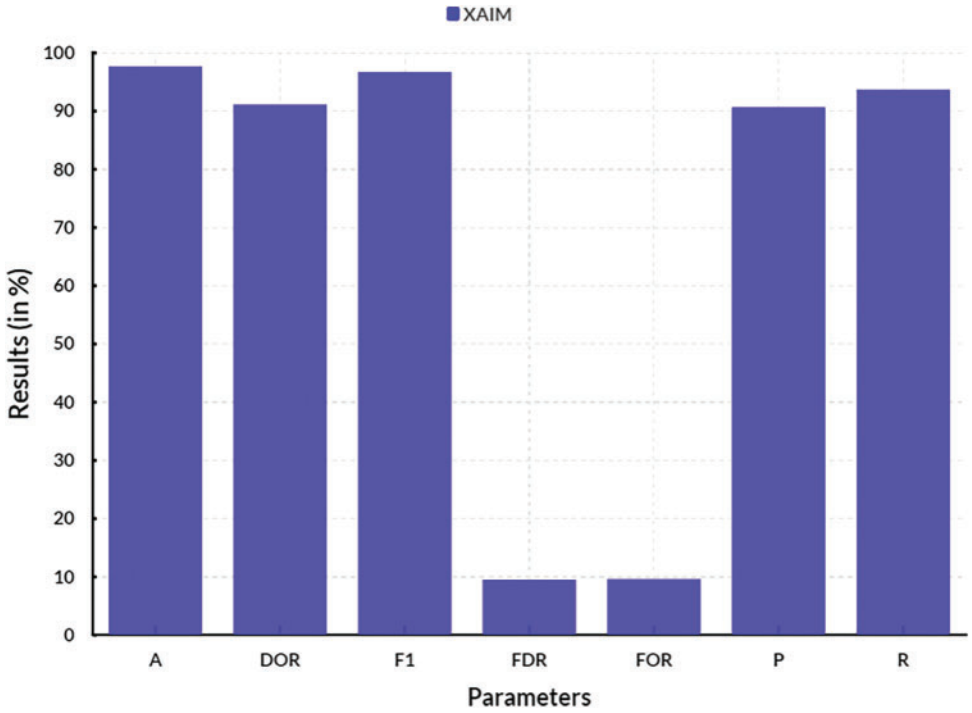

Figure 14: Convergence of performance of proposed XAIM

Applications:

- Explainable Artificial Intelligence can be used in cancer image classification to help detect cancer earlier and more accurately.
- XAI can potentially augment clinical decision-making by providing doctors with more profound insights into the predictions generated by the algorithm.
- In addition, the algorithm's accuracy can be enhanced by optimizing its parameters and variables.
- Clinicians may be able to detect cancer sooner and provide better treatment options.

Interpretable decisions in cancer detection refer to the ability of a decision-making model to explain the reasoning behind its predictions clearly and understandably. This is important for medical personnel as it helps them understand the factors and variables that influence the model's decisions and assess its trustworthiness and reliability. The following are some ways in which the proposed model can provide interpretable decisions to medical personnel in cancer detection:

- Feature importance: The model can list the most important features or variables contributing to its predictions. This will enable medical personnel to understand which factors are most significant in predicting cancer and to evaluate the biological and clinical relevance of these factors.
- Visual explanations: The model can generate visual representations to highlight the regions or features in the input data most important for a particular prediction. This will help medical personnel visualize the model's decision-making process and identify patterns contributing to the prediction.
- Diagnostic accuracy: The model can be evaluated for its diagnostic accuracy, sensitivity, and specificity. These measures can help medical personnel to understand the model's performance in correctly identifying cancer cases and to assess its reliability in making decisions.

The development and implementation of the proposed model for healthcare is a collective effort that requires the involvement of all stakeholders. It faces various challenges, including technical computational demands, data privacy concerns, and scalability in diverse healthcare settings. Addressing these challenges will require further research, collaboration, and regulatory considerations.

- Technical Computational Demands: Developing the proposed model for healthcare applications may require significant computational resources, as it often involves using complex algorithms and techniques. The processing and analysis of large datasets can also be time-consuming and computationally intensive. This can pose a challenge for implementation in real-world healthcare settings, which often need more resources and capacity for data processing.
- Data Privacy Concerns: Healthcare data is susceptible and subject to strict privacy regulations. This model, which often relies on large amounts of patient data, may raise concerns about privacy and data protection. However, explainable AI techniques can help mitigate these concerns by providing transparency and accountability in decision-making.
- Scalability in Diverse Healthcare Settings: While healthcare systems and settings vary widely in terms of resources, infrastructure, and data availability, the proposed model for healthcare is designed to be adaptable. It can be customized to healthcare contexts, ensuring its feasibility and effectiveness. This requires a careful consideration of diverse data sources and the development of transferable models.

5.3 Ethical Considerations

The proposed XAI model has been applied in clinical settings through various studies and trials to investigate its impact on decision-making. For instance, in a study looking at the use of XAI in predicting heart failure patients' readmission risk, the model was found to improve decision-making significantly and was well-received by clinicians. Another analysis focusing on XAI in breast cancer diagnosis found that the model outperformed traditional methods and provided explanations for its predictions, leading to more informed and confident decision-making.

Ethical considerations related to using proposed XAI in cancer image classification and broader healthcare applications include algorithmic bias, data security, and patient consent. Algorithmic bias is the potential for machine learning algorithms to replicate or amplify existing societal biases, leading to discriminatory or unfair outcomes for individuals or groups. In healthcare, this can result in inequitable treatment or diagnosis for specific populations, especially those from underrepresented or marginalized communities. This could worsen health disparities and exacerbate existing social inequalities. One potential solution to address algorithmic bias in XAI is to ensure representative and diverse training data that accurately reflects the demographics of the population being served. This can also be achieved through ongoing monitoring and auditing algorithm performance for biases or discrimination. In addition, healthcare professionals must be involved in developing and evaluating XAI systems to ensure clinical relevance and equity. Data security is another important ethical consideration when using XAI in healthcare. XAI systems rely on sensitive patient data, including personal health information, medical images, and diagnosis outcomes. This data must be carefully and ethically collected, stored, and protected to ensure patient privacy and confidentiality. It is essential to have comprehensive data governance policies to prevent data breaches and unauthorized access to patient information. In addition, patient consent is critical for the ethical use of XAI in healthcare. Patients must be fully informed and give explicit consent for their data to be used in XAI systems. This includes transparency about the purpose of data usage, the collected data types, and any potential risks or benefits. Patients should also have the right to withdraw their consent at any time. In addition to these specific ethical considerations, there are broader concerns surrounding the deployment of XAI in healthcare. These include potential workforce displacement of healthcare professionals, data ownership and monetization, and the impact on doctor-patient relationships. It is essential to closely monitor and address these issues to ensure the responsible and ethical use of XAI in healthcare. The use of XAI in cancer image classification and broader healthcare applications holds great potential to improve patient outcomes and advance medical research. However, ethical considerations regarding algorithmic bias, data security, and patient consent must be addressed to ensure equitable and responsible use. Ongoing collaboration between healthcare professionals, data scientists, and ethicists is necessary to navigate these complex ethical implications and ensure the moral development and deployment of XAI in healthcare.

The development and deployment of XAI models require collaboration among computer scientists, medical professionals, and ethicists due to their interdisciplinary nature. Integrating XAI into healthcare systems can significantly improve patient outcomes, increase the efficiency of healthcare processes, and reduce costs. However, deploying these models also brings ethical concerns, such as the potential for biased decisions and lack of transparency. Therefore, a collaborative effort is necessary to ensure that XAI models are accurate and practical but also ethical and accountable. One of the main reasons collaboration among computer scientists, medical professionals, and ethicists is crucial in developing XAI models is the complexity of the healthcare domain. Healthcare data is multifaceted and requires a deep understanding of medical practices, terminology, and standards. In addition, healthcare services involve sensitive and personal information, making it essential to consider ethical principles and privacy concerns when designing and implementing XAI models. This requires input from medical professionals and ethicists with the relevant expertise who can guide the potential impact on patients and possibly legal and ethical implications.

6 Conclusion

XAI is gradually increasing its relevance in the medical field, especially in cancer detection. In this regard, it is employed to aid human experts in understanding the findings of AI-powered technologies. The platform offers decision-support technologies that aid clinicians in comprehending intricate patterns and supplying substantiated data, facilitating more informed decision-making in diagnosing and treating cancer patients. In addition, it has the potential to enhance image analysis by offering evidence-based predictions to radiologists, enhancing their accuracy and precision. The proposed XAIM reached 97.72% accuracy, 90.72% precision, 93.72% recall, 96.72% F1-score, 9.55% FDR, 9.66% FOR and 91.18% DOR. The explainable AI algorithms in cancer image classification are that they help medical professionals make better decisions by providing a clear and transparent explanation of the thought process that led to a classification. The explanations consist of an overview of the classification results and essential characteristics utilized to make the decision. The explanations can help medical professionals to understand why and how specific diagnoses were made, allowing medical professionals to make more informed decisions.

7 Future Scope

Explainable Artificial Intelligence algorithms have the potential to revolutionize cancer image classification by allowing oncologists to diagnose and treat patients. In the future, AI algorithms can identify and analyze tumors better and more accurately. It will enable oncologists to be better equipped to make informed decisions for treatments that are personalized and tailored to each patient. AI algorithms can be employed to monitor cancer progression and assist in early cancer diagnosis. AI algorithms can also be utilized to predict the course of treatment with greater accuracy by identifying the most suitable treatments for any given patient based on the predictive analysis. The AI algorithms can be further enhanced by implementing deep learning models to enhance accuracy and reduce the time of diagnosis. It can be employed to analyze and compare the effectiveness of different treatments over time. Explainable AI algorithms can be utilized to ensure the safety and ethical use of data when using AI for cancer image classification.

Acknowledgement: We are here to acknowledge funds, and Angeles Quezada will pay the journal APC/Fee.

Funding Statement: This work was supported by the CONAHCYT (Consejo Nacional de Humanidades, Ciencias y Tecnologías).

Author Contributions: Conceptualization, Amit Singhal, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez Aguiñaga, Samantha Jiménez, and Satya Prakash Yadav; methodology, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez Aguiñaga, and Samantha Jiménez; software, Amit Singhal, Krishna Kant Agrawal, and Satya Prakash Yadav; validation, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez Aguiñaga, Samantha Jiménez and Amit Singhal, formal analysis, Krishna Kant Agrawal, Angeles Quezada, and Adrian Rodriguez Aguiñaga; investigation, Amit Singhal, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez Aguiñaga, and Samantha Jiménez; resources, Amit Singhal, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez

Aguiñaga, Samantha Jiménez, and Satya Prakash Yadav; draft and writing, Amit Singhal, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez Aguiñaga, writing—review and editing, Angeles Quezada, Adrian Rodriguez Aguiñaga, Samantha Jiménez and Satya Prakash Yadav; visualization, Krishna Kant Agrawal, Angeles Quezada, Adrian Rodriguez Aguiñaga and Samantha Jiménez; supervision, Samantha Jiménez and Satya Prakash Yadav. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The cancer classification dataset: https://www.kaggle.com/datasets/andrewmvd/cancer-inst-segmentation-and-classification. (accessed on 10/04/2024)

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

References

1. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115.
2. Gunning D. Explainable artificial intelligence (XAI). Defense advanced research projects agency (DARPA). AI Mag. 2017;2(2):1.
3. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60.
4. Gunning D, Aha D. DARPA's explainable artificial intelligence (XAI) program. AI Mag. 2019;40(2):44–58.
5. Zhang YD, Satapathy SC, Guttery DS, Górriz JM, Wang SH. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. Inf Process Manag. 2021;58(2):102439.
6. Ren Z, Zhang Y, Wang S. LCDAE: data augmented ensemble framework for lung cancer classification. Technol Cancer Res Treat. 2022;21: doi:10.1177/15330338221124372.
7. Speith T. A review of taxonomies of explainable artificial intelligence (XAI) methods. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency; 2022; Seoul, Republic of Korea. p. 2239–50.
8. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion. 2023;99:101805.
9. Lötsch J, Kringel D, Ultsch A. Explainable artificial intelligence (XAI) in biomedicine: making AI decisions trustworthy for physicians and patients. BioMedInformatics. 2021;2(1):1–17.
10. Dong J, Chen S, Miralinaghi M, Chen T, Li P, Labi S. Why did the AI make that decision? Towards an explainable artificial intelligence (XAI) for autonomous driving systems. Transp Res Part C: Emerg Technol. 2023;156:104358.
11. Chamola V, Hassija V, Sulthana AR, Ghosh D, Dhingra D, Sikdar B. A review of trustworthy and explainable artificial intelligence (XAI). IEEE Access. 2023;11:78994–9015. doi:10.1109/access.2023.3294569.
12. Madhav AS, Tyagi AK. Explainable artificial intelligence (XAI): connecting artificial decision-making and human trust in autonomous vehicles. In: Proceedings of Third International Conference on Computing, Communications, and Cyber-Security; 2022; Singapore, Springer Nature Singapore. p. 123–36.
13. Rosenfeld A. Better metrics for evaluating explainable artificial intelligence. Bar-Ilan University: Israel; 2021, https://cris.biu.ac.il/en/publications/better-metrics-for-evaluating-explainable-artificial-intelligence. [Accessed 2024].

14. Hulsen T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare. AI. 2023;4(3):652–66.
15. Islam MR, Ahmed MU, Barua S, Begum S. A systematic review of explainable artificial intelligence in terms of different application domains and tasks. Appl Sci. 2022;12(3):1353.
16. Ong JH, Goh KM, Lim LL. Comparative Analysis of Explainable Artificial Intelligence for COVID-19 Diagnosis on CXR Image. In: 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA); 2021; Kuala Terengganu, Malaysia. p. 185–90. doi:10.1109/ICSIPA52582.2021.9576766.
17. Taylor JET, Taylor GW. Artificial cognition: how experimental psychology can help generate explainable artificial intelligence. Psychon Bull Rev. 2021;28(2):454–75.
18. Letzgus S, Wagner P, Lederer J, Samek W, Müller KR, Montavon G. Toward explainable artificial intelligence for regression models: a methodological perspective. IEEE Signal Process Mag. 2022;39(4): 40–58.
19. Saranya A, Subhashini R. A systematic review of Explainable Artificial Intelligence models and applications: recent developments and future trends. Decis Anal J. 2023;7:100230. doi:10.1016/j.dajour.2023.100230.
20. de Vries BM, Zwezerijnen GJ, Burchell GL, van Velden FH, Boellaard R. Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review. Front Med. 2023;10:1180773. doi:10.3389/fmed.2023.1180773.
21. Hassan MR, Islam MF, Uddin MZ, Ghoshal G, Hassan MM, Huda S, et al. Prostate cancer classification from ultrasound and MRI images using deep learning based explainable artificial intelligence. Future Gener Comput Syst. 2022;127(1):462–72. doi:10.1016/j.future.2021.09.030.
22. O'sullivan S, Janssen M, Holzinger A, Nevejans N, Eminaga O, Meyer CP, et al. Explainable artificial intelligence (XAI): closing the gap between image analysis and navigation in complex invasive diagnostic procedures. World J Urol. 2022;40(5):1125–34. doi:10.1007/s00345-022-03930-7.
23. Knapič S, Malhi A, Saluja R, Främling K. Explainable artificial intelligence for human decision support system in the medical domain. Mach Learn Knowl Extr. 2021;3(3):740–70. doi:10.3390/make3030037.
24. Sarp S, Kuzlu M, Wilson E, Cali U, Guler O. The enlightening role of explainable artificial intelligence in chronic wound classification. Electronics. 2021;10(12):1406. doi:10.3390/electronics10121406.
25. van der Velden BH, Kuijf HJ, Gilhuijs KG, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal. 2022;79:102470. doi:10.1016/j.media.2022.102470.
26. Alonso JM, Castiello C, Mencar C. A bibliometric analysis of the explainable artificial intelligence research field. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations. Cham: Springer International Publishing; 2018; p. 3–15.
27. Dağlarli E. Explainable artificial intelligence (xAI) approaches and deep meta-learning models. Adv Appl Deep Learn. 2020;79. doi:10.5772/intechopen.87786.
28. Muddamsetty SM, Jahromi MN, Moeslund TB. Expert level evaluations for explainable AI (XAI) methods in the medical domain. In: Pattern Recognition. ICPR International Workshops and Challenges. Cham: Springer International Publishing; 2021; p. 35–46.
29. Holder E, Wang N. Explainable artificial intelligence (XAI) interactively working with humans as a junior cyber analyst. Hum-Intell Syst Integr. 2021;3(2):139–53. doi:10.1007/s42454-020-00021-z.
30. Kakogeorgiou I, Karantzalos K. Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. Int J Appl Earth Obs Geoinf. 2021;103(7):102520. doi:10.1016/j.jag.2021.102520.
31. Bento V, Kohler M, Diaz P, Mendoza L, Pacheco MA. Improving deep learning performance by using explainable artificial intelligence (XAI) approaches. Discov Artif Intell. 2021;1(1):1–11. doi:10.1007/s44163-021-00008-y.

32. Hauser K, Kurz A, Haggenmueller S, Maron RC, von Kalle C, Utikal JS, et al. Explainable artificial intelligence in skin cancer recognition: a systematic review. Eur J Cancer. 2022;167:54–69.

33. Vilone G, Longo L. Notions of explainability and evaluation approaches for explainable artificial intelligence. Inf Fusion. 2021;76:89–106.

34. Nazir S, Dickson DM, Akram MU. Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. Comput Biol Med. 2023;156:106668. doi:10.1016/j.compbiomed.2023.106668.

35. Ornek AH, Ceylan M. Explainable Artificial Intelligence (XAI): classification of medical thermal images of neonates using class activation maps. Trait Signal. 2021;38(5):1271–9. doi:10.18280/ts.380502.

36. Cilli R, Elia M, D'Este M, Giannico V, Amoroso N, Lombardi A, et al. Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe. Sci Rep. 2022;12(1):16349.

37. Zhang Y, Weng Y, Lund J. Applications of explainable artificial intelligence in diagnosis and surgery. Diagnostics. 2022;12(2):237. doi:10.3390/diagnostics12020237.

38. Nigar N, Umar M, Shahzad MK, Islam S, Abalo D. A deep learning approach based on explainable artificial intelligence for skin lesion classification. IEEE Access. 2022;10:113715–25. doi:10.1109/ACCESS.2022.3217217.

39. Dindorf C, Konradi J, Wolf C, Taetz B, Bleser G, Huthwelker J, et al. Classification and automated interpretation of spinal posture data using a pathology-independent classifier and explainable artificial intelligence (XAI). Sensors. 2021;21(18):6323. doi:10.3390/s21186323.

40. Mehta H, Passi K. Social media hate speech detection using explainable artificial intelligence (XAI). Algorithms. 2022;15(8):291. doi:10.3390/a15080291.

41. Vilone G, Longo L. Classification of explainable artificial intelligence methods through their output formats. Mach Learn Knowl Extr. 2021;3(3):615–61. doi:10.3390/make3030032.

42. Geetha GK, Sim SH. Fast identification of concrete cracks using 1D deep learning and explainable artificial intelligence-based analysis. Autom Constr. 2022;143:104572.

43. Loetsch J, Malkusch S. Interpretation of cluster structures in pain-related phenotype data using explainable artificial intelligence (XAI). Eur J Pain. 2021;25(2):442–65.

44. Clancey WJ, Hoffman RR. Methods and standards for research on explainable artificial intelligence: lessons from intelligent tutoring systems. Appl AI Lett. 2021;2(4):e53.

45. Ghnemat R, Alodibat S, Abu Al-Haija Q. Explainable artificial intelligence (XAI) for deep learning based medical imaging classification. J Imaging. 2023;9(9):177.