iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71641

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Vision Transformer and Explainable AI for Breast Cancer Detection and Classification: A Review

Laxmi Yadav[1], Girish Chandra[2], Divakar Yadav[3]
[1], [2]Department of Computer science and Engineering, Institute of Engineering and Technology, Lucknow
[3]School of Computer and Information Sciences, Indira Gandhi National Open University, New Delhi

***Abstract:** Vision Transformers (ViTs) and Explainable AI (XAI) are revolutionizing breast cancer detection and classification. ViTs have proven to perform better than conventional Convolutional Neural Networks. XAI techniques are important for enhancing the transparency and trustworthiness of the models, making them more acceptable to healthcare professionals and patients.Current research is exploring self-attention mechanisms within ViTs to generate inherent explanations and using XAI to identify and mitigate biases in models. These advancements are being applied to tasks such as identifying breast cancer subtypes and predicting treatment response. To increase efficiency and speed up training, transfer learning—in which models are pre-trained on huge datasets before being modified for particular breast cancer tasks—is also becoming more common. The lack of huge, high-quality datasets and the high computational expenses of training ViTs are two issues that persist despite advancements. Future studies will concentrate on building more dependable XAI methods, larger and more varied datasets, and more effective ViT designs. User-friendly XAI tools are needed for clinical workflows. Addressing concerns about bias, fairness, and transparency is essential for responsible AI use in healthcare. Data standardization is also needed to ensure consistent results across different locations.*
***Keywords:** Vision Transformers, Explainable AI, Breast Cancer Detection, Breast Cancer Classification, Medical Imaging, Deep Learning, Self-Attention, Transfer Learning*

I. INTRODUCTION

Breast cancer (BC), a pervasive and life-threatening disease, necessitates the development of advanced and accurate diagnostic techniques to improve patient outcomes (Rahman et al., 2022). Traditional methods, while effective to some extent, often suffer from limitations in sensitivity, specificity, and efficiency, thus motivating the exploration of innovative approaches (Madani et al., 2022). BC diagnosis and classification are among the many medical image analysis tasks where deep learning, and in particular convolutional neural networks, have demonstrated impressive effectiveness (Flosdorf et al., 2024). With the ability to get around some of CNNs' drawbacks, the introduction of ViTs—a new class of neural networks modelled after the Transformer architecture in natural language processing—has created new opportunities for medical image analysis (He et al., 2022).In order to identify long-range dependencies in images, the Vision Transformer architecture makes use of self-attention processes, which could result in more reliable and accurate diagnoses. The interpretability and reliability of deep learning models are called into question due to their "black box" character, particularly in crucial applications such as medical diagnostics. Explainable AI approaches seek to solve this problem by shedding light on how these models make decisions, increasing openness, and building patient and clinician trust. This review paper illustrates a brief overview of the application of Vision Transformers and Explainable AI in breast cancer detection and classification, exploring the underlying principles, architectural designs, and performance characteristics of these approaches. By elucidating the strengths and limitations of Vision Transformers in this domain, The goal of the article is to offer a useful tool for practitioners and researchers who want to improve the identification of breast cancer.Self-attention processes are used by the Vision Transformer architecture to identify long-range dependencies in images, which could result in more reliable and accurate diagnoses.

A hybrid computer-aided diagnostic (CAD) system based on deep neural networks that combines the Residual and Inception blocks for binary classification tasks was presented by (Singh et al.,2023). The Breast Histopathology Images (BHI)and BreakHisdatasets were used to assess this architecture. The suggested approach produced consistent performance by utilizing multilayer feature maps and fusing the benefits of residual and inception architectures. In particular, it achieved classification accuracy of 85.21% on the BHI dataset and 80.80% for 40X, 82.76% for 100X, 86.55% for 200X, and 85.80% for 400X magnification levels on the BreakHis dataset.Deep feature transfer learning (TL) was the main mechanism for feature extraction in (Abbasniya et al., 2024).

Because of its hybrid inception-residual architecture, they tested sixteen distinct pre-trained convolutional neural networks before concluding that Inception-ResNet-v2 was the best for feature extraction in breast cancer histology images. LightGBM, CatBoost, and XGBoost were among the ensemble learning strategies they incorporated to improve classification performance. Compared to the most advanced methods currently in use,, their suggested framework, known as IRv2-CXL, performed better when tested on the BreakHis dataset.(Davoudi et al., 2025) used the BreakHis dataset to propose a convolutional neural network (CNN) model for binary classification with a special focus on optimization. They used a genetic algorithm (GA) to refine CNN weights, in contrast to traditional methods that depend on Adam or SGD optimizers. With an accuracy of 85.83% with the Adam optimizer, 69.88% with SGD, and 85.49% for the GA optimizer, the model demonstrated the possibility of applying evolutionary algorithms for deep learning optimization.A novel CNN-based design that uses pooling in the last layer regardless of the histopathological image magnification level was proposed by (Shallu et al., 2022). Their model obtained an average classification accuracy of 85.3% by using data augmentation (DA) approaches to improve image variety and information. CNNs still struggle to capture the global contextual elements that are frequently essential in Breast Cancer Histology (BCH) images, despite recent advancements. Transformer designs, which were first widely used in Natural Language Processing, have lately been modified for computer vision tasks in order to address this issue. In contrast to conventional CNNs, the ViT model adds attention mechanisms that allow the network to model global relationships more successfully.

Deep learning (DL) models have shown great potential, but their opacity has hampered their clinical acceptance. One major issue is that these models' processing of input images to generate predictions is opaque, which is especially problematic in delicate fields like breast cancer diagnosis. In order to overcome this, Explainable AI (XAI) has become a crucial area that aims to ensure confidence in automated medical judgments and demystify the "black-box" character of DL models. By integrating domain knowledge, XAI not only facilitates human interpretability but also speeds up model development.To make deep learning models easier to understand, a number of XAI approaches have been proposed. To find significant visual characteristics, computer vision professionals frequently employ gradient-based techniques like the saliency maps, guided backpropagation (GBP), and layer-wise relevance propagation (LRP). More sophisticated techniques like Deep Taylor decomposition, pattern attribution, and integrated gradients have improved the visual explanation capabilities even more.

A common method for displaying model decisions is the Class Activation Map (CAM) (Pereira & Hussain, 2024), which creates heatmaps that show the discriminative areas of an image that are used for classification. It accomplishes this by examining how feature activations and related weights interact inside the network.

Using the publicly accessible BreakHis dataset, (Kaplun et al., 2025) created an automated approach for evaluating breast cancer images. They employed the Local Interpretable Model-Agnostic Explanations (LIME) technique to explain their predictions and used simple neural network models for binary classification. LIME improves the transparency of the classification results by identifying and visualizing the most crucial areas in the input photos that influence the model's judgments.

II. VISION TRANSFORMER IN BREAST CANCER DETECTION

ViTs, initially designed for NLP, have illustrate remarkable adaptability and performance in computer vision tasks. Unlike CNNs, Vision Transformers use a self-attention technique to capture global dependencies in an image, in contrast to convolutional processes and local receptive fields (Matsoukas et al., 2023). Vision Transformers can efficiently describe the relationships between various sections of a picture by breaking it up into a series of patches and using each patch as a token. This allows for the extraction of more thorough and context-aware features (Henry et al., 2022). The core of a Vision Transformer consists of multiple layers of transformer encoders, each comprising self-attention mechanisms and feed-forward neural networks, allowing the model to attend to distinct regions and capture long-distance relationships (Kabir et al., 2023). Pre-training is important for transformer applications (Takahashi et al., 2024). Transfer learning strategies have been shown to be superior when compared to feature extraction and SVM classification (Ashraf et al., 2024). In the context of BC detection, Vision Transformers have shown promise in analysing histopathological images, mammograms, and other imaging modalities. Masked attention mechanisms may improve interpretability of vision transformers (Grisi et al., 2024).

One study demonstrated the efficacy of a Vision Transformer-based model in the identification and assessment of HER2 expression in breast cancer, with a 92.6% classification accuracy for tumor patches (Kabir et al., 2023). The model exhibited high precision (92.81%), sensitivity (92.6%), and F1-score (92.69%), suggesting its potential for accurate identification of tumor regions in whole slide images (Kabir et al., 2023). Vision Transformers have shown potential in histopathological image analysis for breast cancer classification, offering an alternative to CNNs that can capture long-range dependencies (Ashraf et al., 2024).

These models excel in capturing long-range dependencies, potentially leading to more accurate diagnoses. Vision transformers can outperform CNN based models when trained with a large number of images (Sharma & Verma, 2023).

III. EXPLAINABLE AI FOR BREAST CANCER CLASSIFICATION

XAI has become a vital field that seeks to improve the interpretability and transparency of intricate machine learning models, especially in high-stakes industries like healthcare. It is difficult to comprehend the reasoning behind the predictions made by many deep learning models due to their "black box" character, which raises questions around responsibility and trust (Shen et al., 2023). By attempting to shed light on the models' decision-making processes, XAI approaches help users comprehend which aspects or areas of an image were most important in formulating a certain forecast. The key areas of the input image that affect the model's output are highlighted in heatmaps produced by gradient-based techniques like Grad-CAM.Image segmentation techniques can be used to identify the region of interest in images, and attention mechanisms can be employed to diagnose breast cancer using magnetic resonance imaging (Kumari & Ghosh, 2023). Explainable AI has the potential to significantly improve physicians' trust in a model's predictions by assisting them in understanding the reasoning behind a specific diagnosis in the context of breast cancer categorization. Clinicians can determine if the model is focusing on pertinent disease traits or if it is being misled by artifacts or irrelevant information by seeing the regions of interest that the model has identified. In order to make sure the model is not producing discriminatory predictions based on patient demographics or other protected features, Explainable AI may also assist in identifying potential biases in the model. Explainable AI methods can be used to visualize, explain, and interpret deep learning models (Samek & Müller, 2019). The use of Explainable AI helps end-users understand the model and interpret the model's inner workings so they can trust and utilize the model's output to make decisions (Rogha, 2023). Explainability increases the utility of machine learning models. (Sadeghi et al., 2023).

Two main limitations are present in all of these works: (1) explainability tailored to mammography is not given enough attention, as most studies focus on general medical imaging or radiology applications; and (2) there are no standardized, quantitative metrics to assess the clinical relevance and efficacy of XAI techniques.A number of reviews highlight the advantages of transparency in AI, but they do not offer any consistent frameworks for evaluation. For instance, explainability approaches are examined by Groen et al. [31], who focus on dimensionality reduction techniques like t-SNE, feature maps, and visualization tools like Class Activation Maps (CAM). Although Srinivasu et al. [67] address interpretability with methods such as LIME and SHAP, their research is still comprehensive and covers general medical images. Although Chaddad et al. [20] broaden the scope of the topic to cover techniques like LIME,Grad-CAM,ProtoPNet, SHAP, TCAV, and attention mechanisms, they concentrate on a number of clinical areas rather than mammography specifically. Furthermore, these investigations typically focus on visual inspection and model interpretation, paying little attention to the creation and validation of evaluation criteria that are vital for interpretability and clinical judgements.

IV. INTEGRATION OF VISION TRANSFORMER AND EXPLAINABLE AI

The integration of Vision Transformers and Explainable AI holds immense promise for advancing breast cancer detection and classification. Vision Transformers offer powerful feature extraction capabilities, while Explainable AI provides the means to interpret and understand the model's decision-making processes. Self-eXplainable AI methods offer inherent explanations in line with their inherent choices, improving the trustworthiness and resilience of AI systems in healthcare applications (Hou et al., 2024). By combining these two approaches, In addition to achieving great accuracy, researchers can create models that give clinicians important information about the variables influencing the model's predictions. Using Explainable AI approaches to show the attention maps produced by Vision Transformers and emphasize the areas of the image that the model is focusing on during diagnosis is one such strategy. This can assist physicians in determining which characteristics the model deems most crucial for categorization, allowing them to verify the logic of the model and spot possible problem areas. By identifying areas of weakness and identifying hidden patterns, explainable AI can be utilized to assess and enhance performance (Muhammad & Bendechache, 2024). The development of more clear, reliable, and clinically applicable AI systems for breast cancer diagnosis and classification is made easier by the combination of Vision Transformer and Explainable AI, which eventually improves patient outcomes (Rane et al., 2023). The summary of the Vision Transformer and XAI is illustrated in Table 1.

Table1 : Summary of Vision Transformer and XAI Integration in Histopathology Image Analysis

Component	Description	Benefits	Challenges	Potential Solutions
Vision Transformer	Transformer-based architecture that divides images into patches	Captures global context, superior performance on	Requires large datasets and high computational	Use of pre-trained models and data augmentation.

Component	Description	Benefits	Challenges	Potential Solutions
(ViT)	and uses self-attention for feature learning.	high-res images.	power.	
Explainable AI (XAI)	Methods to interpret model decisions, e.g., attention maps, Grad-CAM, LIME, SHAP.	Improves trust, transparency, and clinical acceptance.	Existing XAI methods are often tailored to CNNs, not ViTs.	Develop ViT-specific XAI tools like attention rollout or token attribution.
Attention-Based Visualization	Uses attention weights from ViT to identify important image patches.	Highlights key diagnostic regions in histopathological images.	May provide low-resolution or abstract explanations.	Combine with other XAI methods (e.g., saliency maps) for richer interpretability.
Clinical Utility	Application in breast cancer diagnosis and decision support.	Assists pathologists, supports faster and consistent diagnoses.	Limited validation in real-world clinical settings.	Perform extensive validation and collaborate with clinicians for feedback.
Hybrid XAI Approach	Combines multiple XAI techniques (e.g., attention + gradient-based) for enhanced insights.	Provides comprehensive and human-understandable explanations.	Increased computational complexity.	Optimize post-hoc analysis and focus on real-time interpretability.

In a healthcare environment, AI systems need to be dependable and safe. The application of XAI in computer vision tasks often involves generating heatmaps that pinpoint image regions or pixels significantly influencing the AI model's output (Shortliffe et al., 1975). The growing complexity of modeling approaches further supports the need of Explainable AI for end users (Rozario & Čevora, 2023). XAI can improve transparency and faith in AI systems by explaining how complex models produce outcomes (Kale et al., 2022). Explainable AI enhances the understanding of complex models.

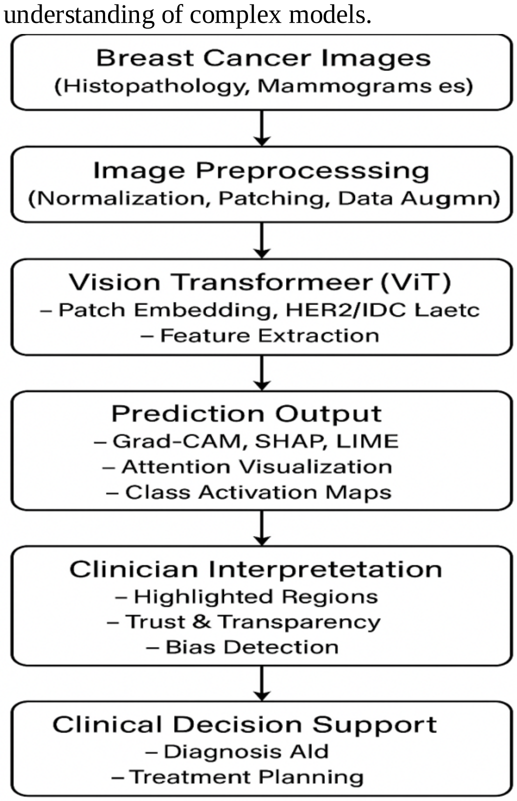


Fig. 1. Vision Transformers (ViT) and XAI for breast cancer identification and categorization

Flowchart in fig.1 illustrating the integration of Vision Transformers (ViT) and Explainable AI (XAI) for BCidentification and categorization.

1) Breast Cancer Images
- Input Data: The process begins with acquiring medical images such as:
  - Histopathology images – stained tissue slides captured via microscopy.
  - Mammograms – X-ray images of the breast.
- These images can vary in resolution and magnification (e.g., 40X, 100X, 200X, 400X from datasets like BreakHis).
- These inputs often contain critical visual cues related to tumor characteristics like size, shape, and spread.

2) Image Preprocessing
- The raw images go through several preprocessing steps to prepare them for the model:
  - Normalization: Adjust pixel values to standard ranges.
  - Patching: The image is divided into smaller patches (ViTs treat each patch as a "token" like words in NLP).
  - Data Augmentation: Techniques like flipping, rotating, zooming, and color jitter are applied to increase data variability and reduce overfitting.
- These steps are crucial because ViTs expect fixed-size patch sequences and benefit from rich and diverse data during training.

3) Vision Transformer (ViT)
- The pre-processed image patches are fed into the Vision Transformer architecture:
  - Patch Embedding: A vector representation is created by flattening and embedding each patch.
  - Self-Attention Mechanism: Every patch takes care of every other patch to model global relationships.
  - Feature Extraction: The transformer encoder layers extract hierarchical and contextual features from across the entire image.
- In this stage, the model might be trained or fine-tuned to recognize patterns related to specific types of breast cancer (e.g., HER2 expression, IDC).

4) Prediction Output
- The ViT model provides predictions such as:
  - Tumor type (benign vs malignant or subtypes like HER2+).
  - Location or severity markers.
- Alongside predictions, Explainable AI tools are used to interpret model behavior:
  - Grad-CAM: Highlights regions influencing the classification by using gradient information.
  - SHAP & LIME: Model-agnostic techniques that estimate the contribution of each feature/region.
  - Attention Maps: From ViT's attention layers, they show which patches were most influential in the decision.
  - Class Activation Maps (CAMs): Heatmaps showing discriminative regions used for classification.

5) Clinician Interpretation
- The visual explanations and attention maps are passed to radiologists or pathologists:
  - Highlighted Regions: Show tumor cells, abnormal tissues, or artifacts the model attended to.
  - Trust & Transparency: Helps clinicians understand why a diagnosis was made.
  - Bias Detection: Ensures the model isn't influenced by irrelevant image features (e.g., artifacts, background).

6) Clinical Decision Support
- The final step integrates model outputs into clinical workflows:
  - Diagnosis Aid: Helps in confirming or suggesting possible diagnoses.
  - Treatment Planning: Results may inform personalized treatment options (e.g., chemotherapy for HER2+ cases).

V. CURRENT RESEARCH AND APPLICATIONS

Current research in BCidentification and classification using Vision Transformers and Explainable AI is focused on developing novel architectures, training strategies, and Explainable AI techniques to improve performance and interpretability.

Explainable AI is being explored to improve transparency in clinical decision support systems across medical fields such as radiology, pathology, cardiology and oncology (Rane et al., 2023). Some studies are exploring the use of self-attention mechanisms within Vision Transformers to generate inherent explanations of the model's decisions, eliminating the need for post-hoc Explainable AI techniques (Hou et al., 2024). Others are investigating the use of XAI to recognize and address the model's flaws, ensuring that it is making fair and equitable predictions across different patient populations. Researchers are also exploring the use of Vision Transformers and Explainable AI for various tasks beyond binary classification, such as identifying different subtypes of breast cancer or predicting treatment response.To detect invasive ductal carcinoma, for example, deep learning pre-trained models such as ResNet-50 and DenseNet-161 have been employed (Rahman et al., 2022). The application of transfer learning, which refines models already trained on huge datasets for particular breast cancer tasks, is also gaining popularity as a way to improve performance and reduce training time (Kabir et al., 2023). Explainable AI is used to enhance the understanding of medical imaging data by providing visual and textual explanations of model predictions (Jeyaraman et al., 2023).

These studies are opening the door for the creation of AI systems for breast cancer diagnosis and classification that are more precise, dependable, and interpretable.

These systems could revolutionize clinical practice and enhance patient outcomes. Nonetheless, some contend that the objectives of patient-level decision assistance are unlikely to be satisfied by the explainability techniques currently in use(Ghassemi et al., 2021). Explainable AI facilitates the discovery of new relationships between variables and outcomes in complex datasets (Loh et al., 2022). Explainable AI helps end-users determine whether the model is suitable for its intended use and whether its limitations are acceptable.

VI. CHALLENGES AND FUTURE DIRECTIONS

Even with the tremendous advancements in breast cancer classification and detection with the use of Explainable AI and Vision Transformers, a number of obstacles still need to be overcome. The lack of substantial, high-quality datasets for training and assessing these models is one of the main obstacles. Another challenge is the computational cost associated with training Vision Transformers, which can be prohibitive for some researchers and institutions. Machine learning techniques' opaque nature also restricts clinical use (Singh et al., 2020). This tension between accuracy and interpretability is an ongoing issue, as complex models can be challenging to interpret (Xu & Shuttleworth, 2023).

Explainable AI can identify performance gaps and uncover hidden patterns to assess and enhance performance. Moreover, the lack of standardized evaluation metrics for Explainable AI makes it difficult to compare different techniques and assess their effectiveness. Another challenge lies in the inherent complexity of Vision Transformers, It may make it challenging to comprehend and analyze their choices, even with Explainable AI techniques. GridSearch along with cross-validation takes prohibited amounts of time (Prusty et al., 2023). The protection of patients is one of the main needs of AI in healthcare.

Going forward, future research directions include the development of more efficient Vision Transformer architecturesthe development of more extensive and varied datasets, as well as more resilient and trustworthy Explainable AI techniques (Zhou et al., 2020). Explainable AI may also be used to provide customized training and classification options based on user needs, potentially enhancing adaptability and effectiveness in diverse scenarios. Explainable AI systems are becoming more and more necessary, particularly those that adhere to ethical standards. Future studies should concentrate on creating Explainable AI technologies that are easier for healthcare workers with less technical knowledge to use and that can be seamlessly incorporated into clinical workflows. To ensure the responsible and equitable application of AI in healthcare, it is imperative to address concerns regarding algorithmic bias, fairness, and transparency (Kelly et al., 2019; Yang et al., 2021).

Lack of data standardization may result in inconsistent outcomes across different locations. AI systems are susceptible to errors because they are human-designed (Mudgal et al., 2022). Explainable AI builds trust in AI systems by showing how they make decisions, but trust isn't the only thing that matters for adoption (Sendak et al., 2019). The success of these AI models is contingent on the availability of high-quality data, the careful selection of hyperparameters, and the mitigation of overfitting (Prusty et al., 2023).

VII. CONCLUSION

the integration of Vision Transformers (ViTs) and Explainable Artificial Intelligence (XAI) into breast cancer histopathology analysis is an important step forward in the creation of intelligent, reliable, and clinically meaningful computer-aided diagnosis (CAD) systems.

Vision Transformers have proven to be more effective than conventional convolutional neural networks (CNNs) at capturing global contextual features and long-range relationships. Their application in breast cancer classification has the potential to significantly enhance diagnostic accuracy across varying image magnifications and data complexities.At the same time, the growing emphasis on XAI reflects a critical shift toward interpretability and trustworthiness in deep learning models. In sensitive domains such as medical imaging, it is essential not only for models to perform well but further to give clear and logical justification for their forecasts. XAI methodsranging from saliency maps and Grad-CAM to more advanced tools like SHAP and LIMEare essential in helping to close the gap between clinical applicability and black-box algorithms by emphasizing pertinent information and offering numeric or visual explanations.Despite these advancements, several challenges remain. Current XAI research still lacks standardized, domain-specific evaluation metrics and often fails to address the unique requirements of breast imaging. Moreover, many existing studies focus primarily on model performance while giving limited attention to clinical integration and validation.Nevertheless, ongoing research continues to push the boundaries of what is possible, promising the future development of AI-driven systems that are not only accurate but also transparent, interpretable, and ultimately beneficial to patient care and clinical decision-making.

FUNDING DECLARATION
No funding information is available.
DATA AVAILABILITY
All datasets used in this research are publicly available.
CONFLICT OF INTEREST
The authors declare no conflict of interest. The research is conducted independently, and no competing financial or personal interests influenced the outcomes of this study.

REFERENCES

[1] Apon, T. S., Hasan, Md. M., Islam, A., & Alam, Md. G. R. (2021). Demystifying Deep Learning Models for Retinal OCT Disease Classification using Explainable AI. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). https://doi.org/10.1109/csde53843.2021.9718400
[2] Ashraf, F. B., Alam, S. M. M., & Sakib, S. M. (2024). Enhancing breast cancer classification via histopathological image analysis: Leveraging self-supervised contrastive learning and transfer learning. Heliyon, 10(2). https://doi.org/10.1016/j.heliyon.2024.e24094
[3] Davenport, T. H., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94. https://doi.org/10.7861/futurehosp.6-2-94
[4] Dhopte, A., & Bagde, H. (2023). Smart Smile: Revolutionizing Dentistry with Artificial Intelligence [Review of Smart Smile: Revolutionizing Dentistry with Artificial Intelligence]. Cureus. Cureus, Inc. https://doi.org/10.7759/cureus.41227
[5] Flosdorf, C., Engelker, J., Keller, I. C., & Mohr, N. (2024). Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2407.18554
[6] Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11). Elsevier BV. https://doi.org/10.1016/s2589-7500(21)00208-9
[7] Grisi, C., Litjens, G., & Laak, J. van der. (2024). Masked Attention as a Mechanism for Improving Interpretability of Vision Transformers. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2404.18152
[8] He, K., Chen, G., Li, Z., Rekik, I., Yin, Z., Wen, J., Gao, Y., Wang, Q., Zhang, J., & Shen, D. (2022). Transformers in Medical Image Analysis: A Review [Review of Transformers in Medical Image Analysis: A Review]. arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2202.12165
[9] Henry, E. U., Emebob, O., & Omonhinmin, C. A. (2022). Vision Transformers in Medical Imaging: A Review [Review of Vision Transformers in Medical Imaging: A Review]. arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2211.10043
[10] Hou, J., Liu, S., Bie, Y., Wang, H., Tan, A., Luo, L., & Chen, H. (2024). Self-eXplainable AI for Medical Image Analysis: A Survey and New Outlooks. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2410.02331
[11] Jeyaraman, M., Balaji, S., Jeyaraman, N., & Yadav, S. (2023). Unraveling the Ethical Enigma: Artificial Intelligence in Healthcare [Review of Unraveling the Ethical Enigma: Artificial Intelligence in Healthcare]. Cureus. Cureus, Inc. https://doi.org/10.7759/cureus.43262
[12] Kabir, S., Vranić, S., Saady, R. M. A., Khan, M. S., Sarmun, R., Alqahtani, A., Abbas, T. O., & Chowdhury, M. E. H. (2023). The utility of a deep learning-based approach in Her-2/neu assessment in breast cancer. Expert Systems with Applications, 238, 122051. https://doi.org/10.1016/j.eswa.2023.122051
[13] Kale, A., Nguyen, T., Harris, J. S., Li, C., Zhang, J., & Ma, X. (2022). Provenance documentation to enable explainable and trustworthy AI: A literature review [Review of Provenance documentation to enable explainable and trustworthy AI: A literature review]. Data Intelligence, 5(1), 139. The MIT Press. https://doi.org/10.1162/dint_a_00119
[14] Kaufman, R., & Kirsh, D. (2023). Explainable AI And Visual Reasoning: Insights from Radiology. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2304.03318
[15] Kelly, C., Karthikesalingam, A., Suleyman, M., Corrado, G. S., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 17(1). https://doi.org/10.1186/s12916-019-1426-2
[16] Kumari, V., & Ghosh, R. (2023). A magnification-independent method for breast cancer classification using transfer learning. Healthcare Analytics, 3, 100207. https://doi.org/10.1016/j.health.2023.100207

[17] Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022) [Review of Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)]. Computer Methods and Programs in Biomedicine, 226, 107161. Elsevier BV. https://doi.org/10.1016/j.cmpb.2022.107161

[18] S. Singh, R. Kumar, Breast cancer detection from histopathology images with deep inception and residual blocks, Multimed. Tools Appl. 81 (4) (2022) 5849–5865.

[19] M.R. Abbasniya, et al., Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods, Comput. Electr. Eng. 103 (2022) 108382.

[20] Davoudi, K.; Thulasiraman, P. Evolving convolutional neural network parameters through the genetic algorithm for the breast cancer classification problem. Simulation 2021, 97, 511–527.

[21] Shallu and Mehra, Rajesh, “Automatic magnification independent classification of breast cancer tissue in histological images using deep convolutional neural network” Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2 (2019) 772–781.

[22] G. Yang, Q.H. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond, Inform Fusion 77 (2022) 29–52, https://doi.org/10.1016/j. inffus.2021.07.016 18.

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, ‘‘Learning deep features for discriminative localization,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2921–2929.

[24] D. Kaplun, et al., Cancer cell profiling using image moments and neural networks with model agnostic explainability: A case study of breast cancer histopathological (BreakHis) database, Mathematics 9 (20) (2021) 2616.

[25] Madani, M., Behzadi, M. M., & Nabavi, S. (2022). The Role of Deep Learning in Advancing Breast Cancer Detection Using Different Imaging Modalities: A Systematic Review. Cancers, 14(21), 5334. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/cancers14215334

[26] Marey, A., Arjmand, P., Alerab, A. D. S., Eslami, M., Saad, A. M., Sanchez, N., & Umair, M. (2024). Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. The Egyptian Journal of Radiology and Nuclear Medicine, 55(1). https://doi.org/10.1186/s43055-024-01356-2

[27] Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2020). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics, 113, 103655. Elsevier BV. https://doi.org/10.1016/j.jbi.2020.103655

[28] Matsoukas, C., Haslum, J. F., Söderberg, M., & Smith, K. A. (2023). Pretrained ViTs Yield Versatile Representations for Medical Images. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2303.07034

[29] Mudgal, S. K., Agarwal, R., Chaturvedi, J., Gaur, R., & Ranjan, N. (2022). Real-world application, challenges and implication of artificial intelligence in healthcare: an essay. PubMed, 43, 3. National Institutes of Health. https://doi.org/10.11604/pamj.2022.43.3.33384

[30] Muhammad, D., & Bendechache, M. (2024). Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis. Computational and Structural Biotechnology Journal, 24, 542. Elsevier BV. https://doi.org/10.1016/j.csbj.2024.08.005

[31] Pereira, G. A., & Hussain, M. A. (2024). A Review of Transformer-Based Models for Computer Vision Tasks: Capturing Global Context and Spatial Relationships. arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2408.15178

[32] Prusty, S., Patnaik, S., Dash, S. K., & Prusty, S. G. P. (2023). SEMeL-LR: An improved modeling approach using a meta-learning algorithm to classify breast cancer. Engineering Applications of Artificial Intelligence, 129, 107630. https://doi.org/10.1016/j.engappai.2023.107630

[33] Rahman, M. M., Khan, Md. S. I., & Babu, H. Md. H. (2022). BreastMultiNet: A multi-scale feature fusion method using deep neural network to detect breast cancer. Array, 16, 100256. https://doi.org/10.1016/j.array.2022.100256

[34] Rane, N., Choudhary, S., & Rane, J. (2023). Explainable Artificial Intelligence (XAI) in healthcare: Interpretable Models for Clinical Decision Support. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4637897

[35] Rogha, M. (2023). Explain To Decide: A Human-Centric Review on the Role of Explainable Artificial Intelligence in AI-assisted Decision Making. arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2312.11507

[36] Rozario, S., & Čevora, G. (2023). Explainable AI does not provide the explanations end-users are asking for. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2302.11577

[37] Sadeghi, Z., Alizadehsani, R., Çifçi, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alataş, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalo, P. M. (2023). A Brief Review of Explainable Artificial Intelligence in Healthcare.https://doi.org/10.2139/ssrn.4600029

[38] Samek, W., & Müller, K. (2019). Towards Explainable Artificial Intelligence. In Lecture notes in computer science (p. 5). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-28954-6_1

[39] Sendak, M., Elish, M. C., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S., & O’Brien, C. (2019). “The Human Body is a Black Box”: Supporting Clinical Decision-Making with Deep Learning. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1911.08089

[40] Sharma, A., & Verma, N. K. (2023). A Novel Vision Transformer with Residual in Self-attention for Biomedical Image Classification. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2306.01594

[41] Shen, Y., Guo, P., Wu, J., Huang, Q., Le, N., Zhou, J., Jiang, S., & Unberath, M. (2023). MoViT: Memorizing Vision Transformers for Medical Image Analysis. Lecture Notes in Computer Science, 205. https://doi.org/10.1007/978-3-031-45676-3_21

[42] Shortliffe, E. H., Davis, R., Axline, S. G., Buchanan, B. G., Green, C. C., & Cohen, S. N. (1975). Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system. Computers and Biomedical Research, 8(4), 303. https://doi.org/10.1016/0010-4809(75)90009-9

[43] Singh, A., Sengupta, S., & Lakshminarayanan, V. (2020). Explainable deep learning models in medical image analysis. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2005.13799

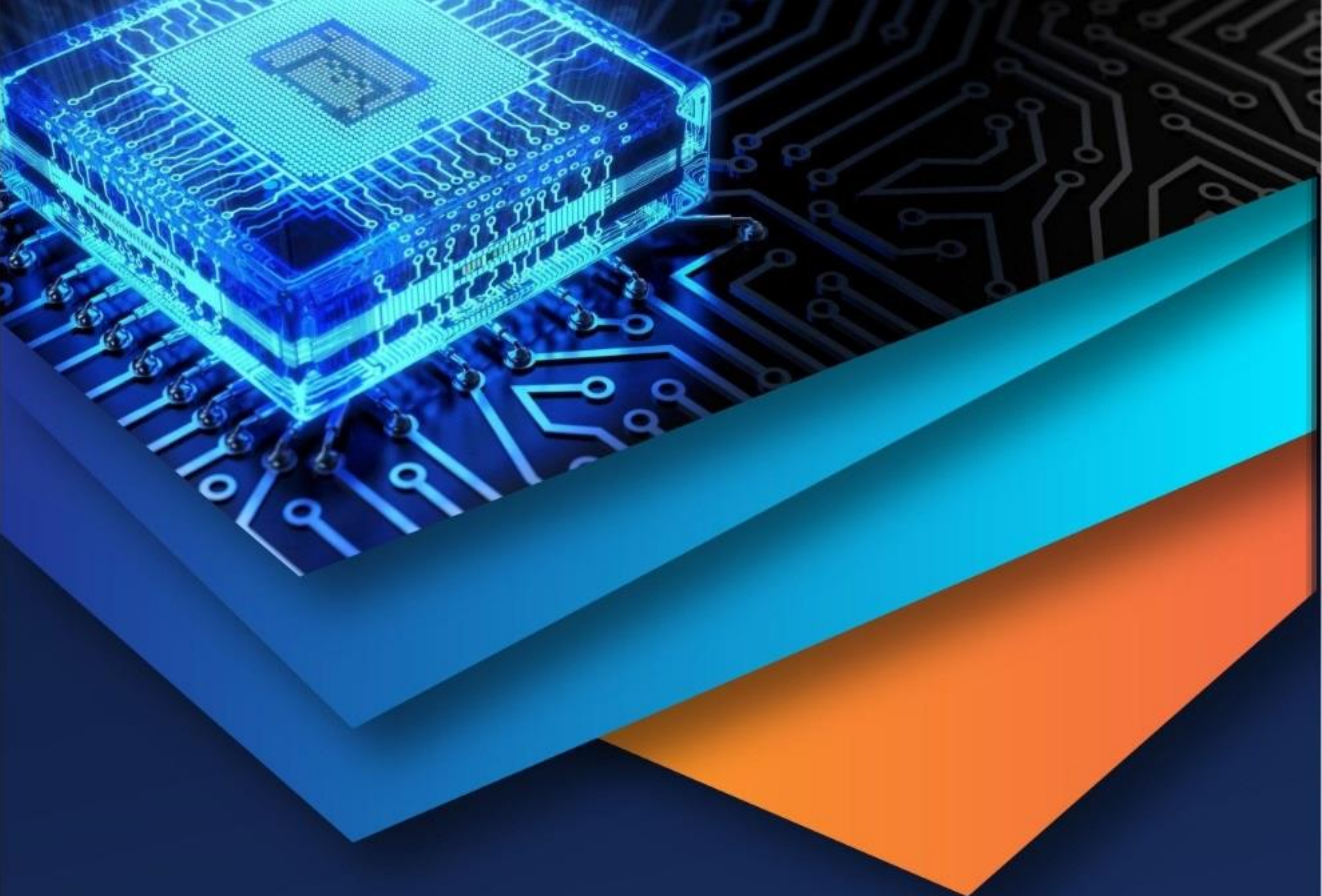
[44] Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., Aoyama, R., Teraya, N., Bolatkan, A., Shinkai, N., Machino, H., Kobayashi, K., Asada, K., Komatsu, M., Kaneko, S., Sugiyama, M., & Hamamoto, R. (2024). Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review [Review of Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review]. Journal of Medical Systems, 48(1). Springer Science+Business Media. https://doi.org/10.1007/s10916-024-02105-8

[45] Xu, H., & Shuttleworth, K. M. J. (2023). Medical artificial intelligence and the black box problem: a view based on the ethical principle of "do no harm." Intelligent Medicine, 4(1), 52. https://doi.org/10.1016/j.imed.2023.08.001

[46] Yang, G., Ye, Q., & Xia, J. (2021). Unbox the Black-box for the Medical Explainable AI via Multi-modal and Multi-centre Data Fusion: A Mini-Review, Two Showcases and Beyond. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2102.01998

[47] Zhang, Y., Weng, Y., & Lund, J. N. (2022). Applications of Explainable Artificial Intelligence in Diagnosis and Surgery, Diagnostics, 12(2), 237. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/diagnostics12020237

[48] Zhou, X., Guo, Y., Shen, M., & Yang, G. (2020). Artificial Intelligence in Surgery. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2001.00627



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY