An Explainable Hybrid CNN–Transformer Framework for Breast Cancer Detection from Histopathology Images

Anandhi K.[1], Karuppasamy K.[2], Sinduja R.[3]

[1]PG Student, [2]Professor, [3]Assistant Professor, Department of Computer Science, RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India

E-mail: [1]anandhipsg91@gmail.com, [2]karuppasamyrvs@gmail.com, [3]sindujabtech42@gmail.com

Abstract

Cancer is one of the most serious health challenges in the world and hence pathologists need accurate and clinically reliable diagnostic support. Even though deep learning algorithms have proven to be very effective in analyzing histopathology images, most of the models currently are not interpretable and prone to false negatives. In this paper, a hybrid CNN-Transformer model detect breast cancer automatically and it combines the local and global contextual modelling of features to capture the complex tissue patterns. The standard preprocessing and augmentation approaches are used to make experiments on large-scale histopathology image dataset robust. Accuracy, precision, recall, F1-score and AUC-ROC are used to evaluate the proposed model, but the main focus of the model is on recall and reduction of false negatives. Grad-CAM and transformer attention maps as explainable AI methods are used to visualize regions of diagnostic interest to enhance model transparency and clinical trust. The experimental outcomes prove that the hybrid framework is more efficient in comparison to independent CNN and transformer models which provide a sensitive, reliable and interpretable solution to AI-assisted histopathology based breast cancer detection.

Keywords: Breast Cancer Detection, Hybrid Deep Learning, CNN–Transformer, Histopathology, Explainable AI, Grad-CAM, Vision Transformer.

DOI: https://doi.org/10.36548/jucct.2026.1.003
Received: 16.12.2025, received in revised form: 19.01.2026, accepted: 07.02.2026, published: 20/02/2026

1. Introduction

Cancer is one of the main causes of death around the world and early detection plays an important role for increasing the survival of patients rates [1]. Histopathology of tissue samples is the improved standard in cancer diagnosis, but manual evaluation takes time and inaccurate with a significant risk of diagnostic variations. The increasing number of histopathology images in hospital requires development of automated and reliable diagnostic tools to help pathologists in making clinical decisions [8,9]. Convolutional neural networks (CNNs) demonstrated better results in the categorization of histopathology images by learning discriminating local variables such as local texture and cellular morphology [7]. However, CNNs have limitations in their capacity to explain long-distance spatial connections are required for modelling global tissue structure. Recently, self-attention-based transformer-based architectures, particularly Vision Transformers (ViTs) [4] shows capable of modelling interactions among global environments. Transformer models also have limitations like inaccurate to capture local specific details are critical for effective cancer detection [15].

This issue explain deep learning models is an additional challenge to their deployment in clinical practice with performance limits. Particularly, False negatives have serious clinical impacts, which makes diagnostic systems must be accurate and transparent. ResNet-18 was assigned for its residual learning capabilities, better local feature extraction behavior, medium parameter size and shown stability on medical imaging tasks. Its convolutional inductive bias may capture effective fine-grained morphological characteristics in histopathology images, and it has effective computation [2, 3]. Furthermore, the proposed work is lightweight when it combines Vision Transformer's global historical modeling with the hybrid fusion architecture. This work shows a hybrid CNN-Transformer model for detecting breast cancer automatically to handle the challenges. The suggested methodology can achieve required accuracy, reduced false negatives and high accessibility by integrating equivalent local and global feature representation with explainable AI (XAI) methods [10] utilized to develop reliable clinical decision support [14].

2. Related Work

Deep learning has made major progress in automated histopathological image analysis to detect cancer [5]. Convolutional neural networks (CNNs) become popular to learn discriminative local features in tissue images and recently, transformer-based models have

been studied to learn global contextual dependencies [11]. Hybrid CNN-Transformer models have been shown to be better than the two representations as they are complementary [12].

According to recent studies, explainable AI (XAI) is critical in medical imaging to enhance transparency and clinical trust. Deep learning prediction interpretations in histopathology have been interpreted using techniques like Grad-CAM and attention-based visualizations [6]. Most of the current strategies are focussed with predictive accurateness and limited integration in terms of sensitivity and reduction of false negatives. This work reduces the limitations by providing a hybrid CNN-Transformer framework is improved with the XAI to detect breast cancer in a sensitive and interpretable way [13].

3. Proposed Work

The proposed system uses the local feature extraction of CNN and the global modelling of Vision Transformers (ViTs) automatically detect breast cancer using histopathological images that handle the limitations of current models. Figure 1 represents the hybrid architecture uses CNN to detect fine-grained structural and textural features. The transformers capture extended connections across tissues enable description for complex histopathological structures. A feature-level fusion method applied and the deep embedded representations of ResNet-18 and Vision Transformer combined to create single 1280 dimensional representation. The CNN captures fine-grained morphological patterns and transformer represents the extended contextual interactions. This approach lacks the additional evaluation process and relevant data from both designs provide results with increased prediction of cross-cancer detection.

This model is trained using cross-entropy and the Adam optimizer using learning rate of 1e-4 and a batch size of 16. The data augmentation helps to improve generalization. The evaluation performed on an NVIDIA T4. This system integrates explainable AI (XAI) techniques to increase clinical treatment and transparency. Grad-CAM trained using CNN feature maps and attention maps are compiled on the transformer to identify regions with increased impact on predictions. Visual explanation in real tissue patches focused on clinically significant values provides description of model's decision process and support pathologists in evaluation the results.

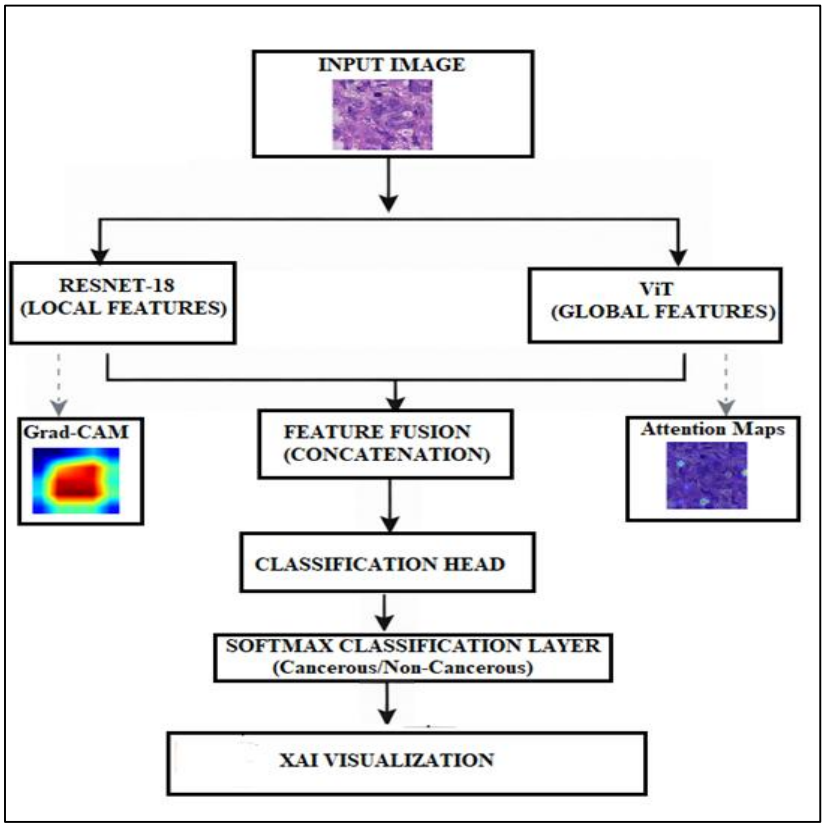


Figure 1. Architecture of Proposed Model

The histopathological image inputs standardized, modified and improved using data modifications like rotation to enhance the model stability and generalization. The hybrid model is trained with cross-entropy and evaluated from freely accessible dataset like Histopathologic Cancer diagnosis (~220,000 labeled patches) [16] to examine recall and reduce false negative for early cancer diagnosis. The proposed system improves CNN and transformer models for accuracy and accessibility to create high synchronization of attention-based interpretations with histopathological structures. The system provides a reliable and clinically accessible method of identifying breast cancer includes both local and global representation using XAI for major limitation of existing automated techniques.

4. Experimental Results and Analysis of Hybrid Approach

The hybrid CNN-Transformer model performs better than traditional CNN and Transformer models in the evaluation process. It has an accuracy of 97%, precision of 96%, recall of 97%, F1-score of 97% and an AUC of 0.9955 represented in fig 2.

```
Device: cuda
Loaded ResNet checkpoint: resnet18_oscc.pth
model.safetensors: 100% 22.9M/22.9M [00:01<00:00, 14.8MB/s]
Loaded ViT checkpoint: vit_oscc.pth
Epoch 1/3 | Train Loss: 0.2018 | Train AUC: 0.9742 | Val Loss: 0.1171 | Val AUC: 0.9911
  -> New best model saved (val AUC: 0.9911)
Epoch 2/3 | Train Loss: 0.1382 | Train AUC: 0.9872 | Val Loss: 0.1060 | Val AUC: 0.9924
  -> New best model saved (val AUC: 0.9924)
Epoch 3/3 | Train Loss: 0.1170 | Train AUC: 0.9907 | Val Loss: 0.0938 | Val AUC: 0.9944
  -> New best model saved (val AUC: 0.9944)
```

	Model	Accuracy	Precision	Recall	F1-score	AUC
0	ResNet18	0.8362	0.8920	0.7651	0.8237	0.9184
1	Vision Transformer (ViT)	0.8989	0.8492	0.9702	0.9057	0.9797
2	Hybrid	0.9719	0.9685	0.9755	0.9720	0.9955

Figure 2. Results of Hybrid Approach

The balance in recall and precision shows the efficient reduction of false positive and false negative is important for early detection of cancer. Figure 3 represented the confusion matrix. This model demonstrates better and balanced results with large number of accurate non-cancerous (22,815) and cancerous (22,986) samples. It has reduced false positives (748) and false negatives (577) ensures with high specificity and reliable method.

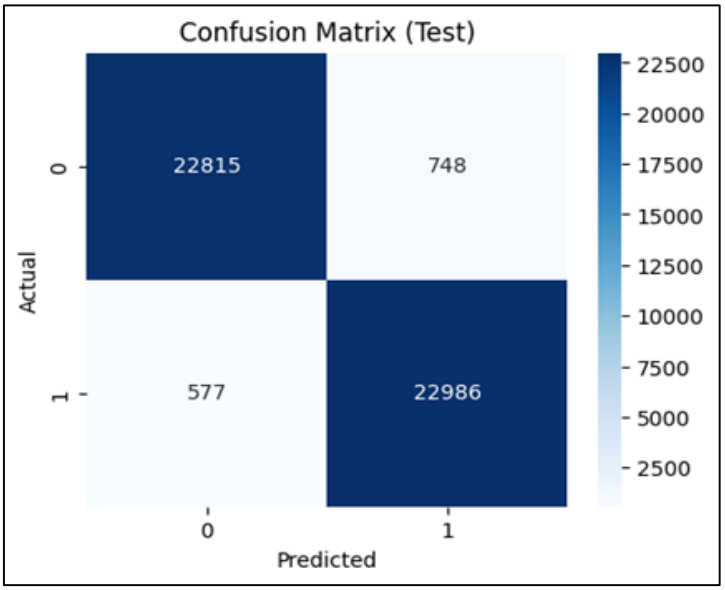


Figure 3. Confusion Matrix

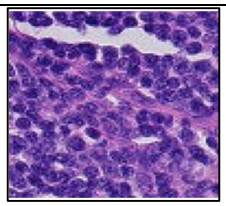
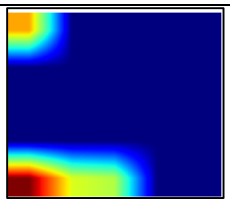
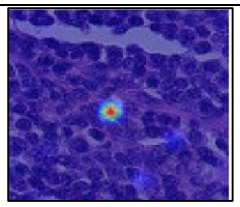
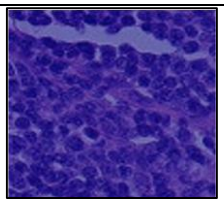
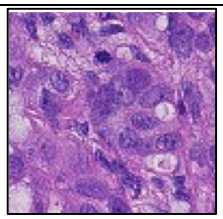
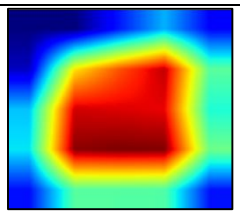
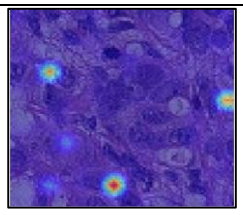
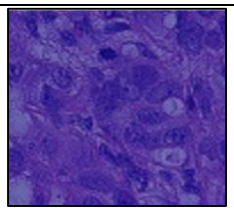
The balanced distribution of errors is effective in combining the local feature extraction in ResNet-18 with global contextual modelling in ViT. This hybrid model reduces

the misclassification for both classes and useful in clinical decisions compared to the standard CNN or transformer-based models. This model has an average inference latency about 10ms per image on GPU hardware and delivers real-time performance.

5. Explainability (XAI) Analysis

The quantitative analysis demonstrates the proposed hybrid CNN-Transformer model required for clinical reliability and acceptability involving automated histopathological analysis. Explainable Artificial Intelligence (XAI) methods are used to present and explain the regions for model predictions. CNN ResNet-18 based model using Grad-CAM presents the discriminative local features and the attention maps are produced using Vision Transformer (ViT) to determine global contextual dependencies. In this hybrid model, the two mechanisms of explanation combined to provide a comprehensive and detailed graphical representation. Table 1 illustrates the result analysis of XAI.

Table 1. XAI Analysis

Original Image	ResNet-18 (Grad-CAM)	ViT (Attention Map)	Hybrid (Grad-CAM + Attention)
			
			

The XAI analysis of ResNet-18 indicates the model is localized, connected with cellular morphology and has texture patterns include nuclear edges and tissue microstructures. This indicates the local features are effective, distributed spatially represents the rate of false negative when using quantitative evaluation. The ViT attention maps have a significant tissue region shows extended contextual effects are modelled accurately. The global attention

directed at non-discriminative domains as per the increased false positive rate with the transformer based models.

The proposed model produces the explanations are informative and interpretable by clinician and it uses the combination of accurate local activation per image basis using Grad-CAM models and widely consistent patterns of attention on global scale using the transformer branch. The identified areas have increased diagnosis used histopathological structures and provide both contextual and spatial awareness. This model also achieves better classification accuracy, transparency and confidence confirms applicability for reliable clinical decision support. The quantitative analysis of XAI was conducted with the help of disturbance-based accuracy measurements (inserting and deleting of AUC) access the degree of variation of predictions in salient regions are systematically changed. When the regions removed, the decreased large probability shows the explanations are the fundamental cause of decision-making.

6. Conclusion

In this proposed work, the hybrid CNN-Transformer model detects breast cancer from histopathology images automatically reduces the major limitations of previous system developed by deep learning provides reliable and interpretable results. The proposed method achieves improved capturing of complex histopathological patterns and reduces false negatives due to integrating CNN-based local feature extraction with transformer-based global contextual modelling. The explainable AI (XAI) features such as Grad-CAM and transformer attention maps integration provide clinical visual explanations related to diagnosing important tissue regions. In this experimental study, the hybrid model achieves high performance in different evaluation measures like recall and AUC used in early cancer detection. This model also provides sensitive, interpretative and clinically reliable solution for AI-based histopathological analysis.

7. Future Work

In future, the proposed model can be implemented in a real-time decision making situation. It develops the model into flexible or web-based diagnostic platform to access it on online, scalability and processing large amount of histopathological images effectively. Quantization, sorting and hardware-aware augmentation process are some of the model

optimization methods evaluated to decrease the complexity of computations and time taken for inference process implemented on resource-constrained clinical systems. The further work aims to implement system validation with the help of predictive clinical data and real-time feedback provided by user pathologists to examine robustness, usability and clinical impact

References

[1]. Alom, Md Romzan, Fahmid Al Farid, Muhammad Aminur Rahaman, Anichur Rahman, Tanoy Debnath, Abu Saleh Musa Miah, and Sarina Mansor. "An explainable AI-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images." Scientific Reports 15, no. 1 (2025): 17531.

[2]. Filiot, Alexandre, Ridouane Ghermi, Antoine Olivier, Paul Jacob, Lucas Fidon, Axel Camara, Alice Mac Kain, Charlie Saillard, and Jean-Baptiste Schiratti. "Scaling self-supervised learning for histopathology with masked image modeling." MedRxiv (2023): 2023-07.

[3]. Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[4]. Mir, Aqib Nazir, Danish Raza Rizvi, and Md Rizwan Ahmad. "Enhancing histopathological image analysis: an explainable vision transformer approach with comprehensive interpretation methods and evaluation of explanation quality." Engineering Applications of Artificial Intelligence 149 (2025): 110519.

[5]. Wu, Yawen, Michael Cheng, Shuo Huang, Zongxiang Pei, Yingli Zuo, Jianxin Liu, Kai Yang et al. "Recent advances of deep learning for computational histopathology: principles and applications." Cancers 14, no. 5 (2022): 1199.

[6]. Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: visual explanations from deep networks via gradient-based localization." International journal of computer vision 128, no. 2 (2020): 336-359.

[7]. Siden, Hagia Sofia, and I. Gusti Ngurah Lanang Wijayakusuma. "Implementation of Convolutional Neural Networks (CNN) for Breast Cancer Detection Using ResNet18 Architecture." Journal of Applied Informatics and Computing 9, no. 4 (2025): 1423-1430.

[8]. An, Jianpeng, Yong Wang, Qing Cai, Gang Zhao, Stephan Dooper, Geert Litjens, and Zhongke Gao. "Transformer-based weakly supervised learning for whole slide lung cancer image classification." IEEE Journal of Biomedical and Health Informatics (2024).

[9]. Wang, Xiyue, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Junzhou Huang, Wei Yang, and Xiao Han. "Transpath: Transformer-based self-supervised learning for histopathological image classification." In International Conference on Medical Image Computing and Computer-Assisted Intervention, Cham: Springer International Publishing, (2021): 186-195.

[10]. Kumar, Tajinder, Manoj Arora, Vikram Verma, Sachin Lalar, and Shashi Bhushan. "Pre-Examination of Breast Cancer Dataset Using Exploratory Data Analysis (EDA) Approach." In 2024 International Conference on Computational Intelligence and Computing Applications (ICCICA), vol. 1, IEEE, (2024): 1-7.

[11]. Anandhi, K., and K. Karuppasamy. "Histopathological Cancer Detection Using Deep Convolutional Neural Networks: A Step Toward Transformer-Based Explainable AI Models."

[12]. Anandhi, K., and K. Karuppasamy. "A Comparative Study of CNN, Vision Transformer, and Hybrid CNN–Transformer Models for Histopathology-Based Cancer Detection."

[13]. Campanella, Gabriele, Matthew G. Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J. Busam, Edi Brogi, Victor E. Reuter, David S. Klimstra, and Thomas J. Fuchs. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images." Nature medicine 25, no. 8 (2019): 1301-1309.

[14]. Stebbing, Richard V., and J. Alison Noble. "Delineating anatomical boundaries using the boundary fragment model." Medical image analysis 17, no. 8 (2013): 1123-1136.

[15]. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, (2016): 770-778.

[16]. Kaggle. (2018). 'Histopathologic cancer detection dataset.' Kaggle. https://www.kaggle.com/competitions/histopathologic-cancer-detection.