JCBI
Journal of Computing & Biomedical Informatics Volume 11 Issue 01
ISSN: 2710 - 1606 2026

Research Article
https://doi.org/10.56979/1101/2026/1286

Fast Vision Transformer with Hierarchical Attention for Multi-Class Breast Cancer Classification in Histopathology Images

Aruna Shankar[1*], M. M. Poornima[2], Heena Farheen Ansari[3], Senthilnathan S[4], Maheshwari S. Divate[5], and P. Annan Naidu[6]

[1]Department of Computer Information Science, Higher College of Technology, Abu Dhabi, United Arab Emirates.
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andra Pradesh, India.
[3]Department of Computer Science and Engineering (Cyber Security), St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India.
[4]Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam, India.
[5]Department of Information Technology, Dr D Y Patil Institute of Technology, Pimpri, India.
[6]Department of Information Technology, GMR Institute of Technology(GMRIT)- Deemed to be University, Rajam, Vizianagaram, Andhra Pradesh, India.
[*]Corresponding Author: Aruna Shankar Email: ashankar@hct.ac.ae

Received: January 18, 2026 **Accepted:** May 27, 2026

Abstract: Breast cancer (BC) is a leading cause of female mortality across the globe. The histopathology analysis of breast tissues is a more crucial tool for diagnosing and staging BC. In recent times, there has been a considerable increase in research examining the usage of deep-learning (DL) models for BC detection from histopathology images (HIs). DL is developed as a precious tool for BC detection by automatically learning discriminative features from medical images. Its capability to model complex patterns has significantly enhanced diagnostic accuracy and maintained earlier and more reliable cancer identification. Consequently, explainable artificial intelligence (XAI) methods have been integrated to show how DL models arrive at their conclusions. Therefore, this study presents a Transformer with Hierarchical Attention for Explainable Multi-Class Breast Cancer Detection (THAXMCBCD) model. The primary intention of this paper is to develop an intelligent system capable of accurately detecting and classifying BC from histopathology images. The proposed THA-XMCBCD model begins with a robust preprocessing stage, including color normalization, image enhancement, and pixel normalization to prepare the input image data for analysis. Following that, a fast vision transformer with hierarchical attention is utilized for feature extraction, which captures local cellular patterns and global tissue context. The extracted features are further passed into a heterogeneous graph attention network for BC classification, including Adenosis, Ductal Carcinoma, Fibroadenoma, Lobular Carcinoma, Mucinous Carcinoma, Papillary Carcinoma, Phyllodes Tumor, and Tubular Adenoma. Furthermore, the model is trained using the AdamW optimizer to ensure stable and efficient convergence. To provide interpretability, ScoreCAM is employed to visualize the region's most influential in the classification decision. The THA-XMCBCD technique is evaluated using the benchmark BreakHis dataset under magnification factors of 100X and 400X. An extensive simulation analysis is performed, and the obtained results demonstrate the betterment of the THA-XMCBCD technique over the recent methods.

Keywords: Breast Cancer; Histopathology Images; Fast Vision Transformer with Hierarchical Attention; Explainable Artificial Intelligence; Deep Learning; Hyperparameter Tuning

1. Introduction

A tumor is an abnormal development of cells, whereas cells separate without stopping the production of new abnormal cells known as harmless or malignant cancers. Therefore, usual cells decrease when the

abnormal (cancerous) cells keep growing, expelling usual cells. Breast cancer (BC) is the 2nd most prevalent tumor in female adults [1]. Initially, to identify this type of cancer, it is essential to get suitable treatment to cure and decrease the number of deaths. BC starts with harmless cancer cells, which develop into malignant cells when there is no early detection [2]. There are various types of BC based on the type of cells and the range that impacts the breast. The main breast formation parts are the connective tissue, lobules, and ducts. The lobule glands secrete milk, which is transmitted to the nipple through ducts. Conjunctive tissue keeps the lobules and creates the natural breast [3]. Generally, BC cells start to propagate from lobules and ducts, and they develop over time and enter the breast tissue and then extend into lymph nodes or other organs. Therefore, early detection of BC is essential to a patient's life [4].

Many models for diagnosing BC in medical practice comprise thermal imaging, mammography, and histopathological images (biopsy). The biopsy is a more effective detection technique that can determine whether a part is tumorous. Initial detection is key for disease treatment and a Safe prediction [5]. Non-invasive BC showing the process that comprises medical breast valuation and tomography validations, like ultrasound, MRI, and mammography [6]. Moreover, BC is identified by pathological analysis, where H&E-stained tissue sections from a biopsy are inspected microscopically to determine malignancy. There are multiple analytical approaches deployed for BC identification. Because of the critical scale of its practical utilization, computer-aided diagnosis (CAD) techniques have become indispensable for the detection and classification of BC concerns [7]. A CAD system is created to assist clinicians in detecting and diagnosing tumors, along with minimizing mortality, through remotely validating clinical images [8].

Recently, the rapid growth of artificial intelligence (AI) models has attained remarkable outcomes in multiple domains. Various models, like rule-driven and machine-learning (ML) techniques, are utilized to examine BC by digital pathology images [9]. In recent years, it was shown that deep-learning (DL)-based methods, which systematize the entire procedures, performed well as a traditional ML method in various image-valuation tasks [10]. Successful utilizations of convolutional neural networks (CNNs) in clinical imaging have enabled for the timely detection of several critical illnesses. Initial DL-driven functions in conventional microscopy image processing have validated their capability to be efficient in the detection of BC [11]. ML has progressively worked a significant part in BC diagnoses over the last few periods. A few statistical, probabilistic, and optimization statistics can be utilized in the ML model to obtain a classification technique from a dataset. DL approaches are able to autonomously remove features, retrieve details from data, and learning innovative abstract clarifications of data. DL methods are very robust. They could solve a standard feature selection concern and have been adopted across a range of segments, comprising biomedicine and computer optics [12].

1.1. Research Contribution and Novelty

To enhance the accuracy and interpretability of histopathology image-based breast cancer diagnosis, this paper presents a Transformer with Hierarchical Attention for Explainable Multi-Class Breast Cancer Detection (THA-XMCBCD) model. The proposed THA-XMCBCD approach aims to effectively detect and classify different BC types using deep learning models. The proposed THA-XMCBCD model consists of five primary stages: comprehensive image preprocessing to standardize and improve the input image data, hierarchical transformer-based feature extraction, graph-based DL classification, hyperparameter tuning, and explainability analysis. The performance assessment of the THA-XMCBCD model is assessed using a benchmark dataset under diverse measures. The significant essential contribution of this paper is as follows:

- Introduce a fast vision transformer with hierarchical attention (FasterViT) to efficiently extract both local cellular details and global tissue context from histopathology (HI) images.
- Integrate a heterogeneous graph attention network (H-GAT) for modeling complex feature relationships and precisely classifying eight BC categories.
- Incorporate ScoreCAM to generate visual explanations that highlight diagnostically relevant regions and enhance model transparency.
- Comprehensive experiments are carried out on the benchmark BreakHis dataset at 100X and 400X magnifications, demonstrating the supremacy of the THA-XMCBCD model compared to other methods.

Novelty: The novelty of this paper lies in the design of an explainable multi-class BC detection system that integrates fast vision transformers with hierarchical attention and heterogeneous graph attention networks for HI image analysis. Unlike traditional convolution-based techniques, the proposed model efficiently extracts both local cellular-level features and global tissue-level contextual information through hierarchical transformer-driven feature extraction. Moreover, the use of a heterogeneous graph attention network allows the modeling of complex inter-feature relationships, resulting in better classification among different BC subtypes. The integration of ScoreCAM offers visual interpretability by emphasizing diagnostically significant regions, improving transparency, and clinical trust.

2. Related Work on Breast Cancer Detection

This section provides a brief synthesis of the literature concerning BC detection using HI images. Alzoubi et al. [13] examined an automated and more precise application for categorizing histological images of BC utilizing modern AI methods. The main objective is to improve detection speed and dependability when decreasing the assignment on physicians, thus facilitating improved patient care and results. The presented method uses the DL framework for dimensionality reduction from histological images. Deka et al. [14] introduced a computing model that assists in diagnosing BC more precisely. Initially, it cleans the mammogram images by filtering to eliminate unwanted noise. And then, it divides the significant portion of the image utilizing a specific approach called Thresholding-Based Level Set. And, it only selects pertinent information by the image using a novel hybrid technique named improved grey wolf optimizer (IGWO) with seagull optimizer algorithm (SOA). Aldakhil et al. [15] introduced the difficulties associated with multi-class classification for BC detection, focusing on enhancing patient care, effectiveness, and timeliness. In this study, the authors concentrate on utilizing HI and evaluate the present layer of BC classification, especially with AI, particularly DL and CNNs. This study exposes many difficulties that interrupt the efficiency of present detection techniques.

Ponraj et al. [16] introduced a new model, MPa-DCAE, which applies a multi-patch-driven deep convolutional autoencoder (DCAE) architecture integrated with VGG19 to identify and classify BC in HI. Through utilizing a multi-patch technique, regions of interest (ROI) are removed from pathology images to simplify localized feature learning, improving the technique's classification power. The AE module allows unsupervised feature learning, enhancing flexibility and resilience to differentiate in image characteristics. Pradeepa et al. [17] proposed a higher-performance DL architecture for reliable classification of BC in HI, representing the restrictions in feature representation and spatial relationship learning. A hybrid DL method is present, enhancing EfficientNetV2 for multi-class feature extraction with GRU developed through an attention mechanism to capture consecutive dependencies. The attention-based mechanism, which is applied in the introduced work, then fine-tunes the method by ignoring higher weights to the most diagnostically related features of the image. Thatha et al. [18] developed a modern DL architecture for HI classification, combining AlexNet and GRU applications, optimized utilizing the hippopotamus optimizer algorithm (HOA). In the initial stages, DenseNet-41 removes complex spatial attributes from HI imagery. These characteristics are processed using AlexNet-GRU method. HOA is utilized to fine-tune hyperparameters, confirming optimal performance of the technique.

The authors [19] established an innovative attention-guided deep network CNN (ADBNet). The ADBNet has a developed convolutional block focusing unit that concentrates on particular features while controlling unrelated ones. It even has a deep network block that improves the broad's flexibility to multiple magnification scales. In addition, it uses a reproductive oppositional system integrated with spreading the dataset by expertly verified images. This improves the dataset for categorizing similar cancer types. Singh et al. [20] developed a method to categorize BC using HI properly through the use of segmentation methods. The author presents an integrated classification and segmentation method for classifying BC using HI to solve these concerns. Chan-Vese structure is utilized for segmentation to describe ROI along with the HI precisely, succeed by the presented SegEIR-Net for detection. The presented technique applies three important networks, ResNet, EfficientNet, and InceptionNet, and connects the results from every block, followed by Dropout and Dense layers. Table 1 presents the comparative analysis of existing BC detection models.

Previous histopathological image classification techniques are primarily depending on CNN-based or transformer-based methods, which concentrating on separating features at either image-level or patch-

level representations. Moreover, they frequently fail to explicitly acquire inter-patch relationships and structural dependencies within tissue areas. This limitation is mainly important in multiclass breast cancer recognition, where subtle morphological variations need both global context-based understanding and relational reasoning among spatial regions.

Table 1. Related works on the breast cancer detection

Authors	Aim	Methods	Datasets	Results	Limitations
Alzoubi et al. [13]	To build a fast and accurate system for detecting BC from tissue images to assist doctors.	DT, KNN, SVM and RF models.	BreakHis dataset.	Accuracy of 96.45 %.	The system is tested on only one dataset, which may limit its use in real clinical settings.
Deka et al. [14]	To develop a computer-based method that helps detect BC early and accurately from images.	CatBoost algorithms	MAMMSIT dataset.	Accuracy of 99.2 %.	Its effectiveness on images from other hospitals or sources is not yet proven.
Aldakhil et al. [15]	To advance BC detection via AI and tissue images for faster and more accurate results.	GAN and TL techniques.	BreakHis dataset.	Achieved better cancer detection	Existing methods are limited by small, unbalanced datasets and may not work well on new data.
Ponraj et al. [16]	To improve efficiency in BC detection using a multi-patch auto-encoder framework.	DCAE method	CBIS-DDSM and MIAS datasets.	Accuracy of 98.36% and 98.95%.	The model is validated on limited datasets, which may restrict real-world applicability.
Pradeepa et al. [17]	To create a deep learning system that accurately identifies BC from tissue images by analyzing spatial and contextual patterns.	GRU model	BreakHis and Camelyon17 datasets.	Precision of 98.15%.	Tested on only two datasets, so performance on other clinical images is uncertain.
Thatha et al. [18]	To enhance early BC recognition with a hybrid AI model optimized for high accuracy.	HOA and GRU methods.	BreakHis and BACH datasets.	Accuracy of 99.60%.	Performance on images from different hospitals or scanners is not yet proven.
Rahman et al. [19]	To design an automatic tool for precise BC detection with high accuracy across multiple magnifications.	CNN model	BreakHis, LC25000, and Cifar-10 datasets.	Accuracy of 99.33%.	Limited dataset testing may reduce reliability in diverse clinical settings.
Singh et al. [20]	To create a model combining segmentation and classification for accurate BC detection in tissue images.	SegEIR-Net methods.	breakHis, BACH and UCSB datasets.	Accuracy of 98.66%, 93.33% and 96.55%.	Generalizability is limited due to testing on selected datasets only.

3. Proposed THA-XMCBCD Framework

The proposed THA-XMCBCD model follows a systematic multi-stage pipeline for accurate and explainable multi-class BC classification from HI images. Fig. 1 showcases the end-to-end architecture of the THA-XMCBCD model. The presented architecture introduces a hybrid framework integrating Faster Vision Transformers (FasterViT) with a Heterogeneous Graph Attention Network (H-GAT). Dissimilar from traditional model, which treat transformer outputs autonomously, this method converts transformer-

derived features into graph nodes and uses heterogeneous attention mechanisms to model complex inter-region dependencies. The key innovation lies in this structured transformation of deep visual features into a graph-based representation for enhanced relational learning. As depicted in Fig. 1, the proposed model consists of several sequential stages, as described below.

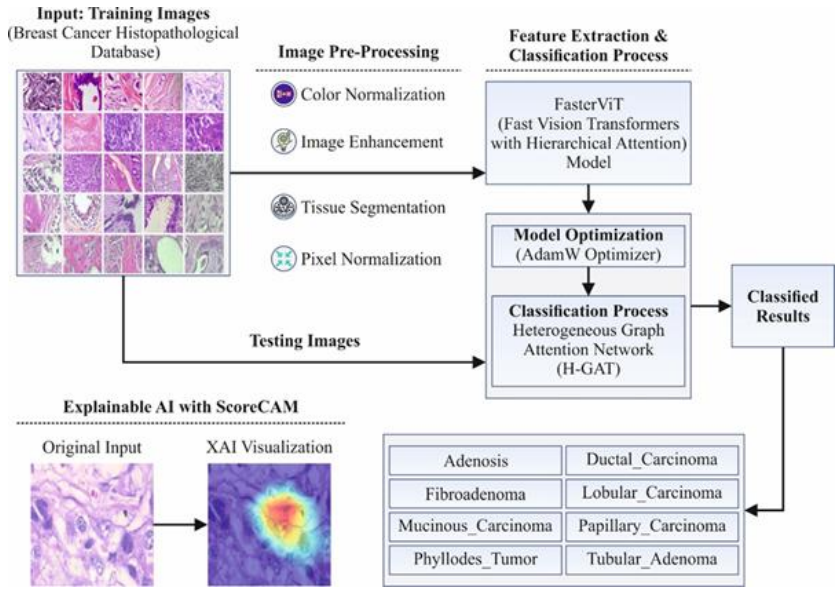

Figure 1. End-to-end architecture of the THA-XMCBCD model

- The raw HI images are first pre-processed to improve image quality and standardize input data. This step comprises color normalization to reduce staining variations, image enhancement to improve contrast and clarity, tissue segmentation for isolating diagnostically relevant regions, and pixel normalization for stabilizing model training.
- The preprocessed images are then passed into a FasterViT with hierarchical attention for feature extraction. This stage extracts discriminative representations by learning local cellular patterns and global tissue contextual information, which enables effective classification of complex histopathological structures.
- The extracted features are then fed into an H-GAT for performing multi-class BC classification. The graph-based model efficiently learns relationships among features and allows for precise classification of eight BC categories.
- To guarantee stable and effective convergence during training, the proposed THA-XMCBCD system is optimized using the AdamW optimizer, which balances adaptive learning rates with robust regularization.
- Finally, ScoreCAM is incorporated into the proposed model to offer visual explanations by emphasizing the most influential regions contributing to the classification decision, thus enhancing model transparency and clinical interpretability.

3.1. Histopathology Image Preprocessing

The HIs exhibit significant variances in texture, contrast, and color intensity because of differences in staining techniques, imaging conditions, and microscope settings [21]. The histopathology images are initial subjected to color normalization to decrease staining dissimilarities and assure constant color distribution over instance. This is succeeded by image improvement to enhance contrast and make cellular and tissue structures more different. Tissue segmentation is then used to eliminate background regions and hold only relevant tissue areas for analysis. At last, pixel normalization is executed to scale intensity values into a uniform range, assuring constant and effective model training over the dataset. The preprocessing pipeline stages are followed by:

Color Normalization: Differences in H&E (Hematoxylin and Eosin) staining over the slides is presented variations. Initiated with the Optical Density (OD) text conversion with the mathematical - log10((RGB + ε)/255). Later, by employing the Macenko or Reinhard technique to achieve stain vector

estimation, these dual approaches utilize SVD (Singular Value Decomposition) under OD space. In this normalization stage, it can match the target stain matrix and normalize the incorporated levels. In conclusion, the RGB (Red, Green, Blue) reconstruction is converted back to RGB space.

Image Enhancement: For increasing the image quality, it is essential to improve these parameters as, contrast, sharpness, and brightness level. Co-operatively, these parameters are boosted by positioning the histogram images. The HE methods have been extensively utilized for improving the handled digital grayscale images. Generally, the above-mentioned method is unsophisticated to use the comparatively lower computation expenses, but it demonstrates greater effectiveness. The core of this technique is to modify the stages of the halftone image on the basis of the probability distribution function of a specified image (2) and, thus, the improvement in the dynamic ranges of brightness distributions. It provides the improvement meant of visual implications, like clarity, sharpness, and brightness contrast.

$$P(i) = \frac{n_i}{n}, i = 0..255 \quad (1)$$

$$H(j) = 255 \sum_{i=0}^{j} p(i) \quad (2)$$

Here, the parameter $P(i)$ refers to the possibility of the arrival of the pixels with brightness i, the normalization function of the histogram of the original images, j means the pixel systematizes the image processing, $H(j)$ represents the converted image.

Tissue Segmentation: It can be used for separating the center regions that include appropriate HIs, while eliminating the background, fat, and alternative non-informative regions. It supports the model concentrated on diagnostically important tissue regions. Segmentation is accomplished by employing contour-based segmentation that can identify the cellular boundaries and tissue structures, followed by morphological operations to refine the segmented regions and remove small artifacts. The subsequent tissue masks have been utilized for extracting the expressive features for classification.

Pixel Normalization: Pixel intensities are normalized to stabilize the input data for the NNs. It can be ensured that entire images are similarly offered during training to prevent computational variability. Also, the normalization increases the model's generalization and convergence speed. In general, pixel values can be measured for a constant range or normalized to include the unit variance and zero mean.

3.2. Hierarchical Vision Transformer-Based Feature Extraction

Fast ViT is well-suited for BC classification from HI images by proficiently extracting both local tissue patterns and global structural context [22]. The hierarchical attention mechanism processes high-resolution cellular details in early layers and progressively aggregates discriminative features such as nuclei arrangement and gland morphology. This design reduces memory overhead while allowing precise learning of multi-scale pathological features. The following are the components of fasterViT.

Stem: A source image defined by its height, width, and three-color channels is broken down into overlapping segments. These segments are analyzed by two sequential 3×3 convolutional filters, each applied with a stride of 2. The resulting features are afterward mapped into a D-dimensional vector space (embedding). The resulting embedded features (tokens) are further advanced using batch normalization (BN) and the ReLU activation function after every filtering stage.

Downsampler Blocks: FasterViT uses hierarchical architecture: the spatial resolution is decreased with 2 among phases through a downsampling block. By applying the 2D layer normalization on spatial attributes, the convolutional layer with a stride of 2, Conv Blocks, and a kernel of $3x3$ is presented.

Phase 1 and 2 contain residual convolutional blocks that can be described as

$$\hat{x} = GELU\left(BN\left(Conv_{3\times3}(x)\right)\right), \quad x = BN(Conv_{3\times3}(\hat{x})) + x, \quad (3)$$

Hierarchical Attention: A new development of windowed attention. Initially, local windows presented in swin transformer. Subsequently, the concept of carrier tokens (CTs) that perform the brief part of complete local windows. The primary attention block has been used to CTs for summarizing and passing the global features. Similarly, the local window tokens and CTs can be integrated in such a manner that each local window takes the accessibility only to its individual set of CTs, with the help of executing self-attention with combined tokens. It can enable the global and local data exchange at decreased expense. With the purpose of changeable local (windowed) and sub-global (CTs) self-attention, it expresses a concept of hierarchical attention.

Consider the specified input feature maps $x \in \mathbb{R}^{H\times W\times d}$ and , H, and W mean the number of feature maps, height, and width, respectively, and assume the set $H = W$ for simplicity. It first partitions input feature maps into $n \times n$ local windows with $n = \frac{H^2}{k^2}$, where k is the window size,

$$\hat{x}_1 = Split_{k\times k}(x) \tag{4}$$

This method centers on developing CTs that facilitate broader attention spans beyond a local window, while reducing computational overhead. Initially, these CTs are created by pooling $L = 2^c$tokens within a specified window.

$$\begin{aligned} \hat{x}_c &= Conv_{3\times 3}(x) \\ \hat{x}_{ct} &= AvgPool_{H^2\to n^2 L}(\hat{x}_c) \end{aligned} \tag{5}$$

Here, the term $Conv_{3\times 3}$ signifies effectual position encoding. $\hat{x}_{ct}$ and AvgPool represent the CTs and feature pooling operation, correspondingly; c is set to 1, then it is altered to control latency. The existing method with conv+pooling offers an image size with more flexibility. In each HAT block, CTs experience the attention process:

$$\begin{aligned} \hat{x}_{ct} &= \hat{x}_{ct} + \gamma_1 \cdot MHSA\big(LN(\hat{x}_{ct})\big) \\ \hat{x}_{ct} &= \hat{x}_{ct} + \gamma_2 \cdot MLP_{d\to 4d\to d}\big(LN(\hat{x}_{ct})\big) \end{aligned} \tag{6}$$

Here, the LN characterizes layer normalization, MHA signifies multi-head self-attention, γ is a learned per-channel scale multiplier, $MLP_{d\to 4d\to d}$ refers to a 2-layer MLP with GeLU activation function.

For modeling short- and long-range spatial data, it can be used for computing the interface between the local and CTs, the terms $\hat{x}_1$ and $\hat{x}_{ct,1}$, correspondingly. Primarily, the local features and CTs have been combined into one. Every local window is only accessible for its respective CTs:

$$\hat{x}_w = Concat\big(\hat{x}_1, \hat{x}_{ct,1}\big) \tag{7}$$

These tokens undertake an alternative set of attention methods:

$$\begin{aligned} \hat{x}_w &= \hat{x}_w + \gamma_1 \cdot MHSA\big(LN(\hat{x}_w)\big), \\ \hat{x}_w &= \hat{x}_w + \gamma_2 \cdot MLP_{d\to 4d\to d}\big(LN(\hat{x}_w)\big). \end{aligned} \tag{8}$$

In conclusion, tokens can be more separated and utilized in the following hierarchical attention layers:

$$\hat{x}_1, \hat{x}_{ct,1} = Split(\hat{x}_w), \tag{9}$$

Measures are determined in Eqs. (6) to (9) have been iteratively used for the numbers of layers in the phase. In order to enable long and short-range interrelations, this can be performed by the global data propagations, similar to those at the end of the phase. Lastly, the output of this phase can be calculated as given below,

$$x = Upsample_{n^2 L\to H^2}\big(\hat{x}_{ct,1}\big) + Merge_{n^2 k^2\to H^2}(\hat{x}_1) \tag{10}$$

MHAs are in Eqs. (6) and (8) are token position invariant, but the position of attributes in the spatial dimension can be evidently informative. In order to overcome this complexity, initially, multiply the total position bias straightly to CTs as well as local window tokens. It is motivated by SwinV2 and utilized as the dual layers MLP for embedding complete 2Ds token position into feature sizes. In order to enable image-related locality inductive bias, it improves the attention with log space comparative position bias through SwinV2 (dual layer MLP). This can ensure that the comparative location of tokens offers collective attention patterns. This method generates flexibility with respect to image size, as the positional encoding is incorporated by the MLP. An attention map comparative evaluation among effectual global-local self-attention is achieved. Fig. 2 model diagram of FasterViT with hierarchical attention.

3.3. Heterogeneous Graph-Based Breast Cancer Classification

The H-GATs are DL methods that have received considerable attention in different domains for their capability to model complex graphs with numerous edges and node forms [23]. With the help of combining the attention mechanisms, the H-GATs are learned correlative specifics significance weights that gain various architectures and semantic connections. It can allow efficient representation learning in H-GAT. Accordingly, the H-GATs are effectively used through an extensive of real-time applications. In the graph construction process where each image is first processed through hierarchical convolutional and attention-based stages to generate patch-level feature maps, and these extracted patch tokens are treated as graph nodes. The number of nodes per image corresponds to the fixed grid of extracted patches after downsampling stages. Each node is represented using a 128-dimensional feature vector obtained through a learnable linear projection applied on the transformer/feature encoder outputs. Edges in the graph are constructed based on a combination of spatial adjacency and feature similarity (k-nearest neighbors),

ensuring both local and contextual relationships between tissue regions. Heterogeneous relations correspond in practice to spatial proximity and semantic similarity between patches, enabling the model to capture diverse inter-region dependencies. The H-GAT is employed for, standard multi-head GAT layers, where hierarchical structure is introduced by stacked attention layers and graph pooling, when the AM itself tracks the original GAT formulation without modification. H-GAT-Concat achieves the efficient classification result that could be benefited from well-maintained relation-driven data, where Het-GAT-Attn offers reasonable results with adaptive relation weighting.

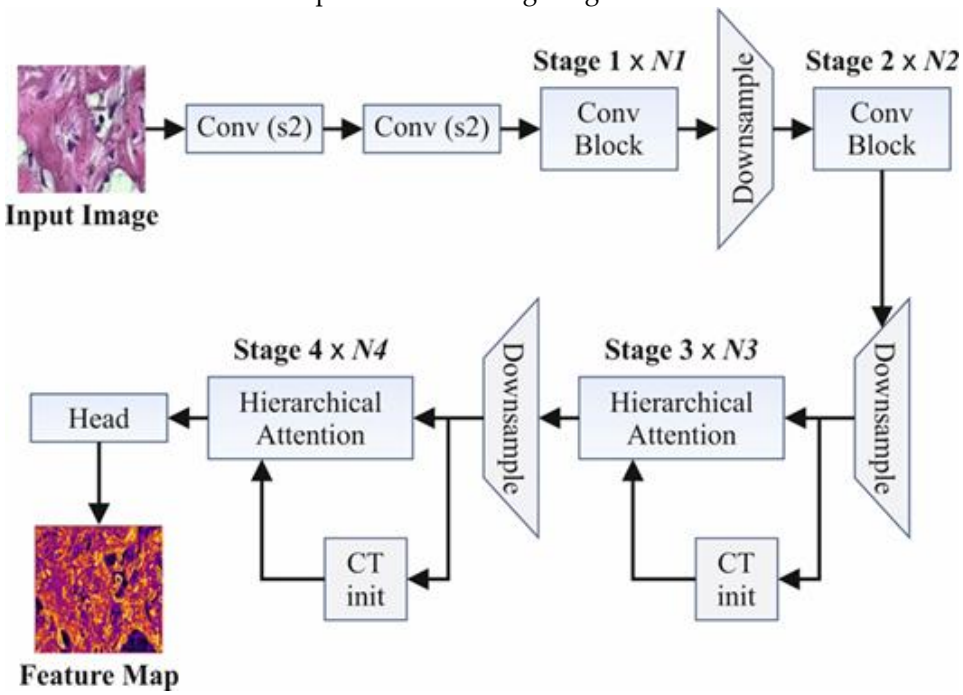

Figure 2. Model diagram of FasterViT with hierarchical attention

Heterogeneous graph attention layer: The H-GAT is processed every individual edge class graph $L_\tau(G)$ self-reliantly utilizing the context-specific Graph Attention Networks (GATs) to obtain the diverse spatial and structural correlations that is crucial for BC histopathological classification. Assume that the term $\mathcal{N}^r(v)$ signifies the neighbors of node u with relation $r \in \{ss, tt, st, pp\}$. In every individual layer l and relation category r, the model is used to calculate the attention weights α_{vu}^r among nodes u and v. Afterwards, it can be implemented context-specific weight matrices W_r for learning the particular representations at all the edge types.

The complete H-GAT embedding layer aggregates the mathematical representation of multi-relational features as given below,

$$x_v^{(l+1)} = MLP^{(l)}\left(\oplus_{r\in\{ss,tt,st,pp\}}\left(\sum_{u\in\mathcal{N}^r(v)} \alpha_{vu}^r W_r x_u^{(l)}\right)\right) \quad (11)$$

Here, the α_{vu}^r represents attention weight among nodes v and u with respect to relation r, the parameter W_r is learned relation weight matrices, and $MLP^{(l)}$ denotes the layer-related MLP. The aggregative operator $\oplus$ denotes 3 various methods.

H-GAT-Concat: The relation-based features to maintain different histopathological semantic contributions with every edge type, thus avoiding the combination of spatial and structural indications:

$$\oplus = Concat: x_v^{(l+1)} = MLP^{(l)}\left(\left[h_v^{ss} \| h_v^{tt} \| h_v^{st} \| h_v^{pp}\right]\right) \quad (12)$$

H-GAT-Sum: Summation of relation-based features for an incorporated representation with decreased dimensionality:

$$\oplus = Sum: x_v^{(l+1)} = MLP^{(l)}\left(h_v^{ss} + h_v^{tt} + h_v^{st} + h_v^{pp}\right) \quad (13)$$

H-GAT-Attn: It is used to learn the attention weights, β^r is dynamically weights the significance of every relation type,

$$\oplus = Attention: x_v^{(l+1)} = MLP^{(l)}\left(\sum_r \beta^r h_v^r\right) \quad (14)$$

Now, the term $h_v^r = \sum_{u\in\mathcal{N}^r(v)} \alpha_{vu}^r W_r x_u^{(l)}$ denotes the combined features with relation r.

This multi-relational attention enables BC classification by using these steps

(i) spatial dependency model where st hierarchical tissue continuity $(i \to j \to k)$, whereas ss/tt model separating or merging structures,

(ii) relation-based feature learning with st edges are identified as satisfactory path extensions, and pp edges are morphologically recognized as the same areas,

(iii) Avoiding feature concatenation maintains different semantic contributions, unlike homogeneous methods that combine neighborhoods from conflated heterogeneous connections.

3.4. Optimization Strategy

AdamW is an enhanced version of the Adam optimizer that dissociates weight decay with the gradient update, thus increasing the generalization in DL methods [24]. With the help of separating regularization from the adaptive learning rate method, it avoids the massive gradient parameters and overcomes a main error in Adam. AdamW is modified the update rules to dissociate the updating weight decay with the original errors of the loss function. Thus, the mathematical form is,

$$\begin{aligned} v_t &= \beta_1 v_{t-1} + (1-\beta_1)\nabla f(x_t) \\ c_t &= \beta_2 c_{t-1} + (1-\beta_2)\nabla f(x_t) \otimes \nabla f(x_t) \\ x_{t+1} &= x_t - \varepsilon_t v_t \otimes \frac{1}{\sqrt{c_t}} - \lambda x_t \end{aligned} \tag{15}$$

Where, the symbol λ is L2 regularization parameter. By employing this update rule, it must avoid the L2 regularization portion in the loss function. It requires to decay λ due to the λ cannot be multiplied by the learning rate. This decay λ is the same as the learning rate decay method. In the case of no decoupling process, original Adam decay the weights that can be reduced due to the 2nd momentum division. Therefore, the concept completely creates the intelligence and the optimistic outcomes. Likewise, the learning rate and regularization hyperparameters could not be related to any further extent. Therefore, this method can be used the grid search rather than random search under learning rate and regularization to determine the better integration of hyperparameters.

Table 2 represented the optimized hyperparameter specification utilized for the presented H-GAT-driven technique. The graph input is denoted utilizing 128-dimensional node dimensionality to extract adequate structural and characteristic data. The H-GAT model contains of two states with 128 hidden units and eight-headed attention mechanism, executes efficient learning of challenging graph relations, when focus and feature dropout rates of 0.3 are used to decrease the overfitting. The classifier head comprises a fully connected state with 64 units and ReLU activation to improve nonlinear discriminative of feature learning. Model optimization process is executed utilizing the AdamW optimizer with a learning rate of 1e-3, weight decay of 0.01, and momentum parameters β_1 = 0.9 and β_2 = 0.999 to ensure a constant integration. Training is conducted with a batch size of 32 for 100 epochs, permitting the method to learn strong representations and attains dependable performance.

Table 2. Parameter setting

Component	Hyperparameter	Optimal Value
Graph Input	Node Feature Dimension	128
H-GAT Layers	Number of Layers	2
	Hidden Units	128
	Attention Heads	8
	Attention Dropout	0.3
	Feature Dropout	0.3
Classifier Head	Fully Connected Units	64
	Activation Function	ReLU
Optimization (AdamW)	Learning Rate	1e-3
	Weight Decay	0.01
	β_1	0.9
	β_2	0.999
Training Setup	Batch Size	32
	Epochs	300

3.5. Explainable AI Analysis with ScoreCAM

Explainability techniques such as ScoreCAM are vital for interpreting, trusting, and debugging complex "black-box" models. They play a critical role in fostering trust, ensuring fairness by identifying bias, and enabling debugging. The ScoreCAM removes the gradient dependency by calculating the modification in class confidence once the feature maps disturb the input images, which provides a more accessible and possibly more consistent method for explaining the outcomes [25]. The ScoreCAM technique includes evaluating the improvement of confidence for a class while particular portions of the input images have been highlighted. Specifically, the ScoreCAM utilizes every feature map in the final convolutional layer to produce a soft spatial mask for the input images. Every mask is primarily normalized values in the ranges of $[0,1]$ followed by upsampled to the resolution of the input image for matching the sizes of the original images. In order to utilize the masks, they can be element-wise multiplied with the original to make a form that only underlines the areas of the feature map.

The addressed regions of ScoreCAM-generated attention maps are examined rely on their correspondence with histopathologically relevant tissue structures namely glandular formations, density variations, and abnormal morphological patterns. The visual clarifications are qualitatively assessed by comparing them with established pathological features included in the literature, assuring the activated parts align with diagnostically important areas. Though no direct quantitative clinical scoring is executed, the steadiness of attention localization over various instance aids the reliability of the approach's decision-making process. It supports further, the transparency of the architecture by delivering interpretable visual evidence that links methods forecasting with significant histological features, thus enhancing trust in the classification results.

The ScoreCAM passed the agitated image by using the CLIP to gain a new class confidence score. Subsequently, the variance among this score of the original image and new score has been determined as the significance weight for this feature map. It can be represented about the significant feature map is used in the target label or class c. The-above mentioned method is constant for whole k feature maps in final convolutional layers of CLIP for calculating the set of significance weights α_k^c. These can be used to linearly integrate the feature maps into an individual heatmap, which must be overlapped at the original images,

$$L_{ScoreCAM}^c = ReLU(\sum_k \alpha_k^c A_k) \quad (16)$$

The ScoreCAM technique takes dual benefits over gradient-based techniques such as the GradCAM. This cannot be used for noisy gradients and have a tendency to create clear images to be more class-specific; but it can be computation complex because of recurrent forward enters in the CLIP.

The explainability is executed on the feature representations attained from FasterViT prior to graph construction. Particularly, the output tokens from the final encoder block of FasterViT are redesigned into a 2D spatial grid to form a surrogate feature map $F \in \mathbb{R}^{H \times W \times C}$, allowing spatial localization. This transformation is significant as transformer-based techniques do not inherently deliver traditional feature maps. According to this representation, class activation mapping methods like ScoreCAM, LayerCAM, and EigenCAM are utilized to generate visual explanations. The activation maps are compared with respect to the final class prediction, thus addressing the most discriminative parts contributing to the model's decision. This formulation assures that the clarifications are directly related with the FasterViT feature extraction process when maintaining coherence with the subsequent graph-based classification by the H-GAT module.

4. Experimental Findings and Interpretation

In this part, the experimental evaluation of the presented model is provided in detail with experimental dataset and performance metrics. The assessment shows the feasibility and applicability of the proposed model in the BC classification task.

4.1. Description of the Dataset Used

The experimental analysis of the THA-XMCBCD technique is verified below the BreakHis dataset [26]. This dataset has been developed to be utilized in ML operations for both binary and multiclass classification issues. It comprises of 3901 microscopic images of breast tumor tissue, acquired utilizing various magnification factors (100X and 400X). The 100× and 400× magnification levels scales are providing the most diagnostically relevant histopathological details while ensuring consistency with the dataset acquisition protocol. The images in this dataset have been resized to 224x224 pixels and structured based

on binary and multiclass classification activities. Table 3 shows the complete distribution of the dataset for the 100× and 400× magnification instances. Fig. 3 represents the sample histopathology images from the BreakHis dataset. Fig. 4 demonstrates the sample images of pre-processed and feature map representations.

Table 3. Details of dataset

Classes	Magnification Image Samples	
	100X	400X
Adenosis	113	106
Ductal Carcinoma	903	788
Fibroadenoma	260	237
Lobular Carcinoma	170	137
Mucinous Carcinoma	222	169
Papillary Carcinoma	142	138
Phyllodes Tumor	121	115
Tubular Adenoma	150	130
Total Images	2081	1820

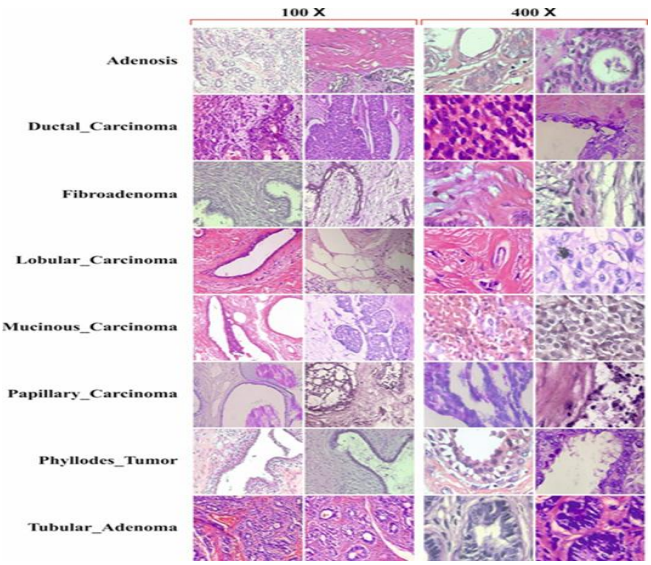

Figure 3. Sample histopathology images from the BreakHis dataset at 100X and 400X

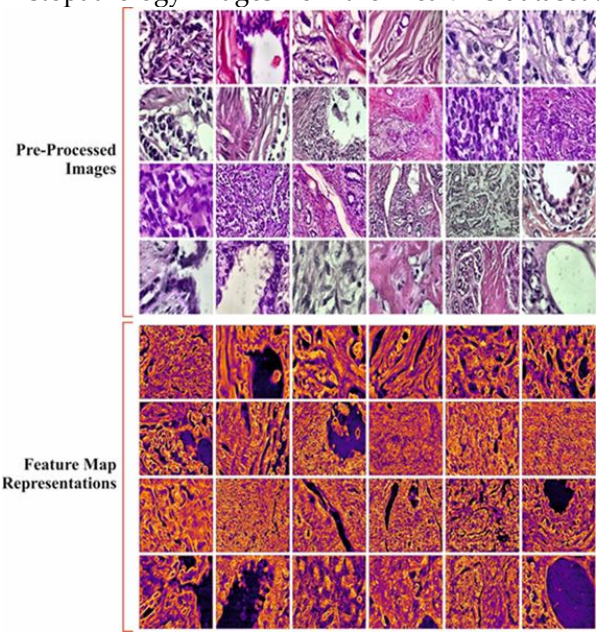

Figure 4. Sample images of preprocessed and feature map representations

4.2. Performance Evaluation Metrics

To assess the model performance, this study employs multiple evaluation metrics, which are detailed below.

Accuracy: It signifies the total proportion of correctly classified histopathological samples among the total test samples. In the proposed eight-class tissue classification task, it represents the effectual classification of all tissue classes.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (17)$$

Recall: It denotes the percentage of positive image patches that are appropriately identified.

$$Recall = \frac{TP}{(TP+PN)} \quad (18)$$

Precision: It measures reflects the percentage of positive forecasts that are accurately classified. This is computed by applying this expression:

$$Precision = \frac{TP}{(TP+FP)} \quad (19)$$

F1-Score: It is a metric that integrates the effect of precision as well as recall employing the harmonic mean, providing equal penalties to maximum values. This is usually computed by applying this expression:

$$F1 - Score = \frac{(2\times recall\times precision)}{(recall+precision)} \quad (20)$$

AUC: It determines the capability of distinguishing among classes. The highest AUC value suggests excellent discrimination among the instances.

True positives (TP): Samples where the forecasted class and original class are both positive. This implies that the identifiers correctly categorized the sample with a positive class.

False positives (FP): Samples where the forecasted class is positive; however, the original class is negative. This indicates that the categorizes inaccurately identified the sample with a positive class.

True negatives (TN): Samples where the forecasted class and original class are both negative. This signifies that the classifier properly categorized the sample with a negative class.

False negatives (FN): Samples where the forecasted class is negative but the original class is positive. This indicates that the identifies inaccurately categorized the sample with a negative class.

4.3. Confusion Matrices at 100X and 400X Magnifications (70:30 Split)

Fig. 5 implements the confusion matrices developed by the THA-XMCBCD method with 5a-5b 100X and 5c-5d 400X using 70:30 of TRAP/TESP. The results indicate that the THA-XMCBCD model has effective recognition and detection of all classes accurately.

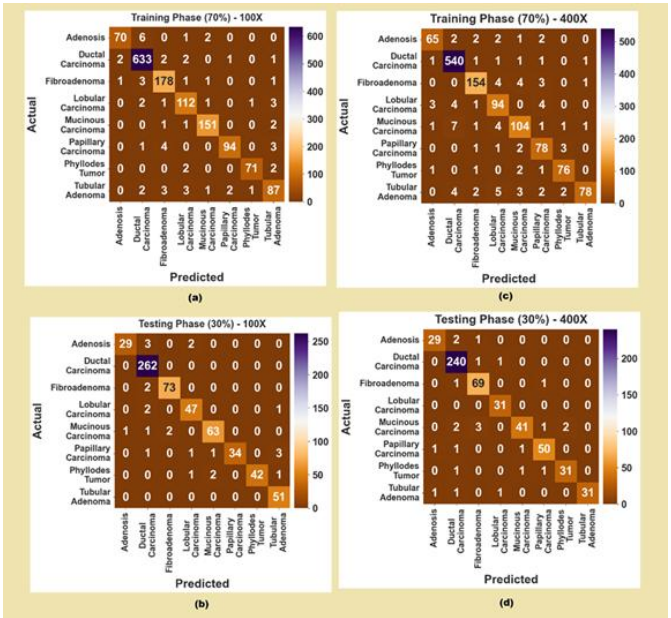

Figure 5. Confusion matrices of THA-XMCBCD system with (a-b) 100X and (c-d) 400X using 70:30

4.4. Performance Analysis at 100X Magnification

Table 4 show the BC classification results of the THA-XMCBCD model utilizing 100X on 70:30. The outcomes mentioned that the THA-XMCBCD model precisely classified the samples. On 70% TRAP, the THA-XMCBCD techniques deliver average value with $accur_y$ of 98.97%, $preci_n$ of 94.82%, $recal_l$ of 93.63%, $F1_{score}$ of 94.19%, and AUC_{Score} of 96.49%, respectively. Meanwhile, on 30% TESP, the THA-XMCBCD technique delivers average value with $accur_y$ of 99.04%, $preci_n$ of 96.17%, $recal_l$ of 93.37%, $F1_{score}$ of 94.59%, and AUC_{Score} of 96.37%, respectively.

Table 4. Breast cancer classifier result of THA-XMCBCD approach using 100X

Class Labels	$Accur_y$	$Preci_n$	$Recal_l$	$F1_{Score}$	AUC_{Score}
Training Phase (70%)					
Adenosis	99.18	95.89	88.61	92.11	94.19
Ductal Carcinoma	98.49	97.84	98.75	98.29	98.52
Fibroadenoma	98.76	94.18	96.22	95.19	97.68
Lobular Carcinoma	98.76	91.80	93.33	92.56	96.29
Mucinous Carcinoma	99.38	96.79	97.42	97.11	98.52
Papillary Carcinoma	99.24	96.91	92.16	94.47	95.97
Phyllodes Tumor	99.59	97.26	94.67	95.95	97.26
Tubular Adenoma	98.35	87.88	87.88	87.88	93.50
Average	98.97	94.82	93.63	94.19	96.49
Testing Phase (30%)					
Adenosis	99.04	96.67	85.29	90.62	92.56
Ductal Carcinoma	98.56	96.68	100.00	98.31	98.76
Fibroadenoma	99.36	97.33	97.33	97.33	98.48
Lobular Carcinoma	98.88	92.16	94.00	93.07	96.65
Mucinous Carcinoma	98.88	95.45	94.03	94.74	96.75
Papillary Carcinoma	99.04	100.00	85.00	91.89	92.50
Phyllodes Tumor	99.36	100.00	91.30	95.45	95.65
Tubular Adenoma	99.20	91.07	100.00	95.33	99.56
Average	99.04	96.17	93.37	94.59	96.37

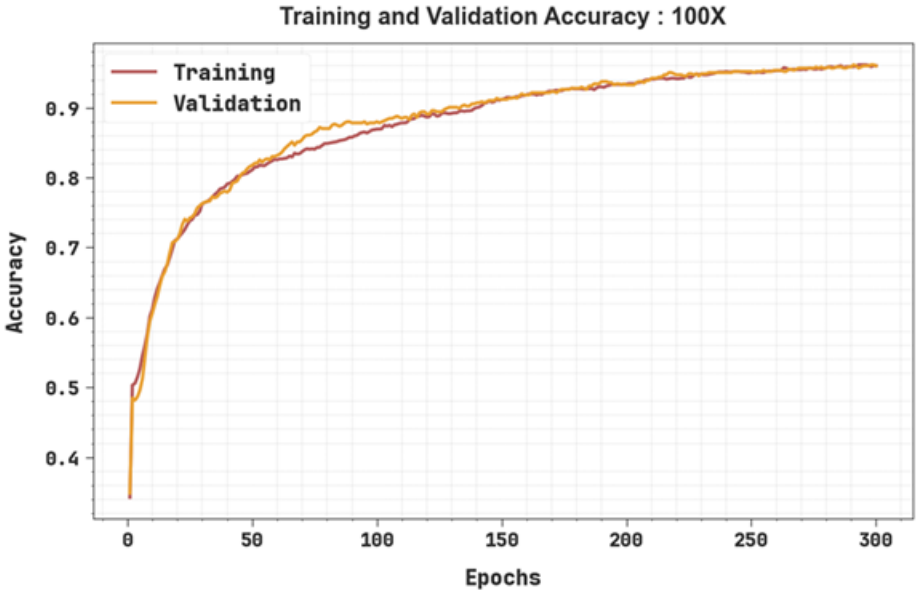

Figure 6. $Accur_y$ curve of THA-XMCBCD approach with 100X

Fig. 6 displays the training (TRAIN) and validation (VALID) accuracy curves of the THA-XMCBCD system with 100X over 300 epochs. TRAIN accuracy rises constantly and reaches a higher-level, specifying efficient learning and integration. The VALID accuracy closely following the training trend with a lesser and constant gap, validating good generalization and least overfitting.

Fig. 7 demonstrates the TRAIN and VALID loss curves of the THA-XMCBCD technique with 100X over 300 epochs. Both losses reduce sharply when the early epochs, specifying fast learning, and then progressively decline in a stable and smooth manner, indicating efficient integration. The close alignment among training and validation loss, with only a least stable gap, recommends small overfitting and robust generalization ability.

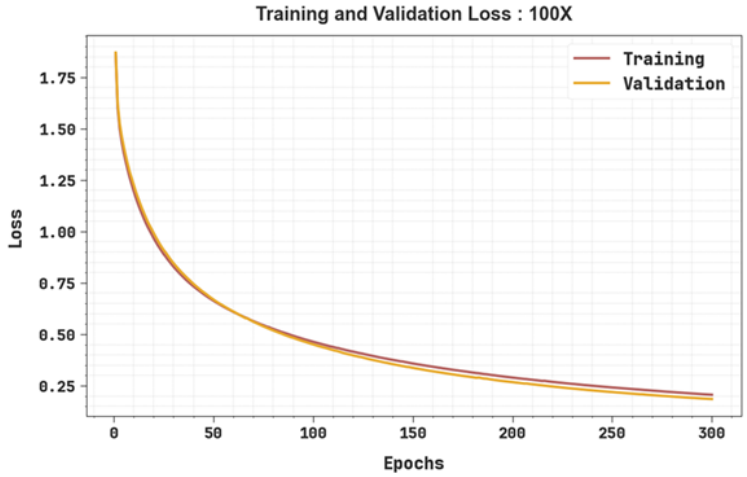

Figure 7. Loss curve of THA-XMCBCD technique with 100X

4.5. Performance Analysis at 400X Magnification

Fig. 8 illustrates the BC classification results of the THA-XMCBCD model utilizing 400X on 70:30. The outcomes mentioned that the THA-XMCBCD technique accurately classified the examples. On 70% TRAP, the THA-XMCBCD approach provide average value with $accur_y$ of 98.33%, $preci_n$ of 91.29%, $recal_l$ of 90.21%, $F1_{score}$ of 90.64%, and AUC_{Score} of 94.60%, respectively. Meanwhile, on 30% TESP, the THA-XMCBCD method deliver average value with $accur_y$ of 98.90%, $preci_n$ of 95.14%, $recal_l$ of 93.42%, $F1_{score}$ of 94.18%, and AUC_{Score} of 96.34%, respectively.

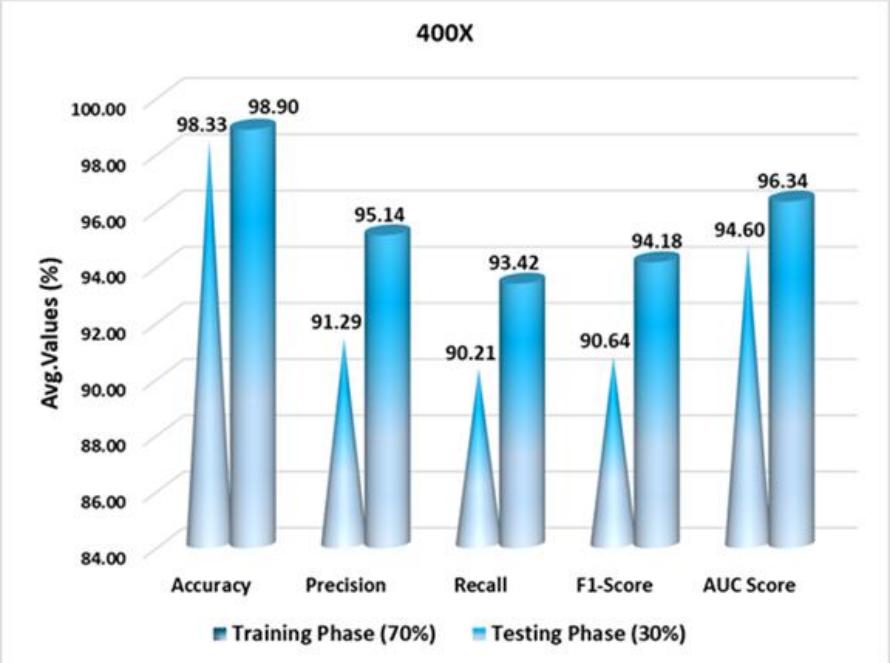

Figure 8. Average of THA-XMCBCD technique using 400X with 70:30 split

Fig. 9 demonstrates the TRAIN and VALID accuracy curves of the THA-XMCBCD approach with 400X over 300 epochs. TRAIN accuracy improves consistency and reaches a higher-level, signifying efficient learning and integration. The VALID accuracy sharply follows the training trend with a least and constant gap, validating good generalization and least overfitting.

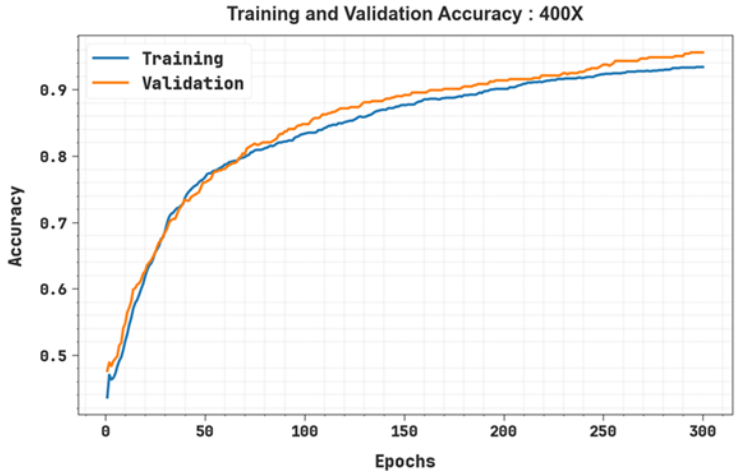

Figure 9. $Accur_y$ curve of THA-XMCBCD method with 400X

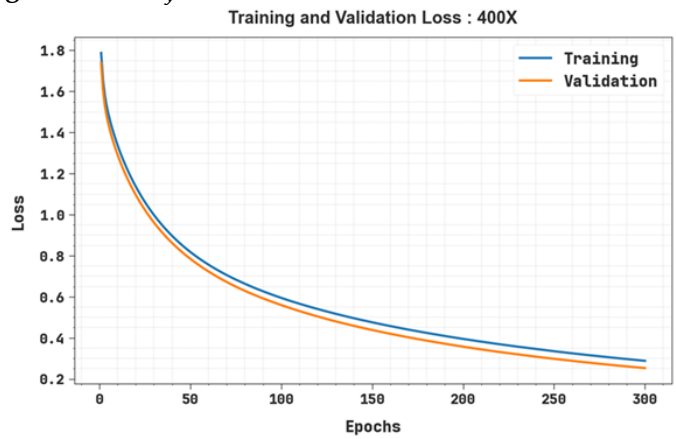

Figure 10. Loss curve of THA-XMCBCD method with 400X

Fig. 10 represents the TRAIN and VALID loss curves of the THA-XMCBCD method with 400X over 300 epochs. Both losses reduces sharply when the early epochs, specifying quick learning, and then progressively deny in a consistent and smooth manner, showing efficient integration. The close alignment among training and validation loss, with only a minimal consistent gap, recommends least overfitting and strong generalization ability.

Fig. 11 considered the ROC curve of the THA-XMCBCD approach with 100X and 400X is examined. The outcomes mention that the THA-XMCBCD method reaches improved ROC result over each class, exposing important ability in class discrimination. This dependable trend of enhanced ROC values over multiple classes indicate the capable performance of the THA-XMCBCD approach on estimating classes, underlining the strong nature under classification procedures.

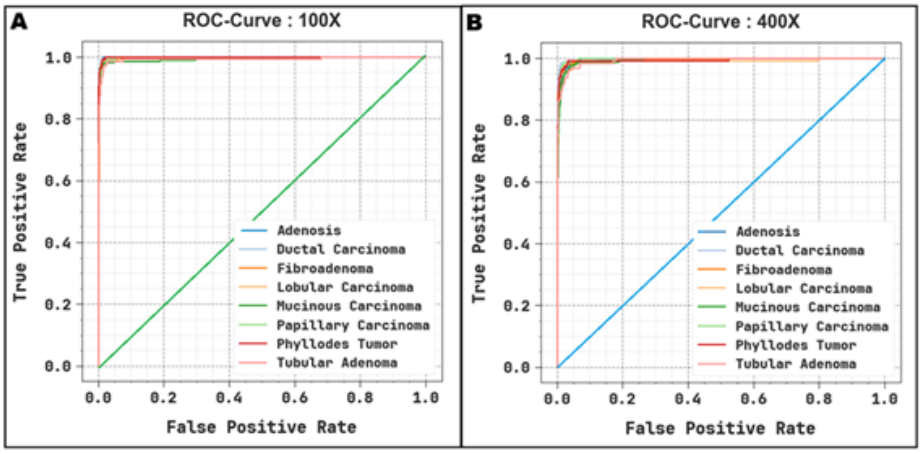

Figure 11. ROC-curve of THA-XMCBCD system with 100X and 400X

4.6. Discussion of Results

The comparative examination of the THA-XMCBCD model THA-XMCBCD method with 100X and 400X utilizing multiple measures is shows in Table 5 [27, 28]. According to 100X, the P4M-DenseNet and Multiple CNNs technique have described poor performance. Additionally, DRDA-Net models have attained kind of higher-performance. In the meantime, 3-CNN fusion technique, Dense-Att Network, ResNeXt+KD, and ST-Double-Net methods have attained reasonable and closer outcomes. Therfore, the THA-XMCBCD method described improved performance with higher $accur_y$ of 99.04%, $preci_n$ of 96.17%, $recal_l$ of 93.37%, and $F1_{score}$ of 94.59%, respectively. Meanwhile, according to 400X, the Multiple CNNs technique have got minimal value with $accur_y$ of 80.10%, $preci_n$ of 90.23%, $recal_l$ of 92.31%, and $F1_{score}$ of 92.78%. Also, the 3-CNN integrating model and Dense-Att Network approaches have archives may higher results. Additionally, the P4M-DenseNet, DRDA-Net, ResNeXt+KD, ST-Double-Net techniques have reached closer and strong outcomes. Lastly, the THA-XMCBCD method have attained higher value with $accur_y$ of 98.90%, $preci_n$ of 95.14%, $recal_l$ of 93.42%, and $F1_{score}$ of 94.18%, respectively.

Table 5. Comparative analysis of THA-XMCBCD method with 100X and 400X

Approaches	$Accur_y$		$Preci_n$		$Recal_l$		$F1_{Score}$	
	100X	400X	100X	400X	100X	400X	100X	400X
P4M-DenseNet	89.80	93.14	85.14	85.53	89.91	89.17	89.64	91.18
3-CNN fusion model	98.00	90.05	90.61	91.43	89.11	87.37	86.59	89.32
Dense-Att Network	97.56	88.77	94.05	90.35	87.75	91.71	89.80	87.40
Multiple CNNs	84.20	80.10	88.53	90.23	92.18	92.31	85.66	92.78
DRDA-Net	94.41	96.84	89.84	91.44	90.20	89.67	85.74	87.23
ResNeXt+KD	96.92	96.64	94.99	88.27	90.47	92.95	85.76	93.29
ST-Double-Net	96.86	95.05	92.91	94.22	91.62	89.89	88.55	88.92
THA-XMCBCD	99.04	98.90	96.17	95.14	93.37	93.42	94.59	94.18

Table 6 denotes the outcome of 10-fold cross-validation performed at two magnification levels (100X and 400X), assessing the model using $Accur_y$, $Preci_n$, $Recal_l$, and $F1_{Score}$. Across all folds, the model demonstrates consistently high performance, with $Accur_y$ values generally exceeding 92% for both magnifications. $Preci_n$ and $Recal_l$ values indicate a good balance between correctly identified positive instances and minimal false negatives, although slight variations are observed between folds due to data distribution differences. The $F1_{Score}$ further confirms the robustness of the model by maintaining stable values across folds, reflecting an effective trade-off between precision and recall. Overall, the results highlight strong model reliability, minimal variance across folds, and good generalization capability for different magnification levels.

Table 6. K-fold cross validation of THA-XMCBCD method with 100X and 400X

K-Fold Cross Validation	$Accur_y$		$Preci_n$		$Recal_l$		$F1_{Score}$	
Fold-1	100X	400X	100X	400X	100X	400X	100X	400X
Fold-2	96.13	96.38	94.11	91.64	85.03	88.01	85.51	89.18
Fold-3	95.89	93.51	92.23	94.33	88.56	91.91	92.01	89.85
Fold-4	97.76	92.79	92.44	92.57	91.09	87.03	92.51	88.55
Fold-5	94.35	97.03	94.49	93.86	87.95	90.87	88.51	85.61
Fold-6	97.97	92.73	94.16	89.56	88.25	85.49	91.46	86.08
Fold-7	92.44	96.53	92.03	94.56	92.37	86.93	85.43	86.71
Fold-8	93.51	95.92	93.79	91.27	91.92	91.44	87.37	85.23
Fold-9	95.68	96.70	95.28	91.22	92.11	92.69	85.76	90.90
Fold-10	95.60	94.67	91.06	91.10	87.54	91.46	88.33	91.52

4.7. Visual Interpretation Results

Fig.12 illustrates the visual explanation of the ScoreCAM with other XAI methods under 8 classes. The figure specifies that multiple CAM approach is used for offering optical explanation for abnormality identification. These models deliver a heatmap denotes the regions on the input image important for decision making through the applied technique. The figure demonstrates an optical comparison of multiple class activation mapping (CAM) models used to illustrative histopathological BC instance. The first column displays the original microscopic images, when the following columns show the associated attention maps produced by LayerCAM, EigenCAM, and ScoreCAM. Over instance 1–6, the heatmaps underline regions that includes most to the technique's predictions, with warmer colors (red/yellow) signifying higher-importance. LayerCAM and EigenCAM commonly deliver smoother and broader attention regions, acquiring learning patterns within cancer areas. In contrast, ScoreCAM deliver more localized and sharper activations, concentrating on the most discriminative cellular areas.

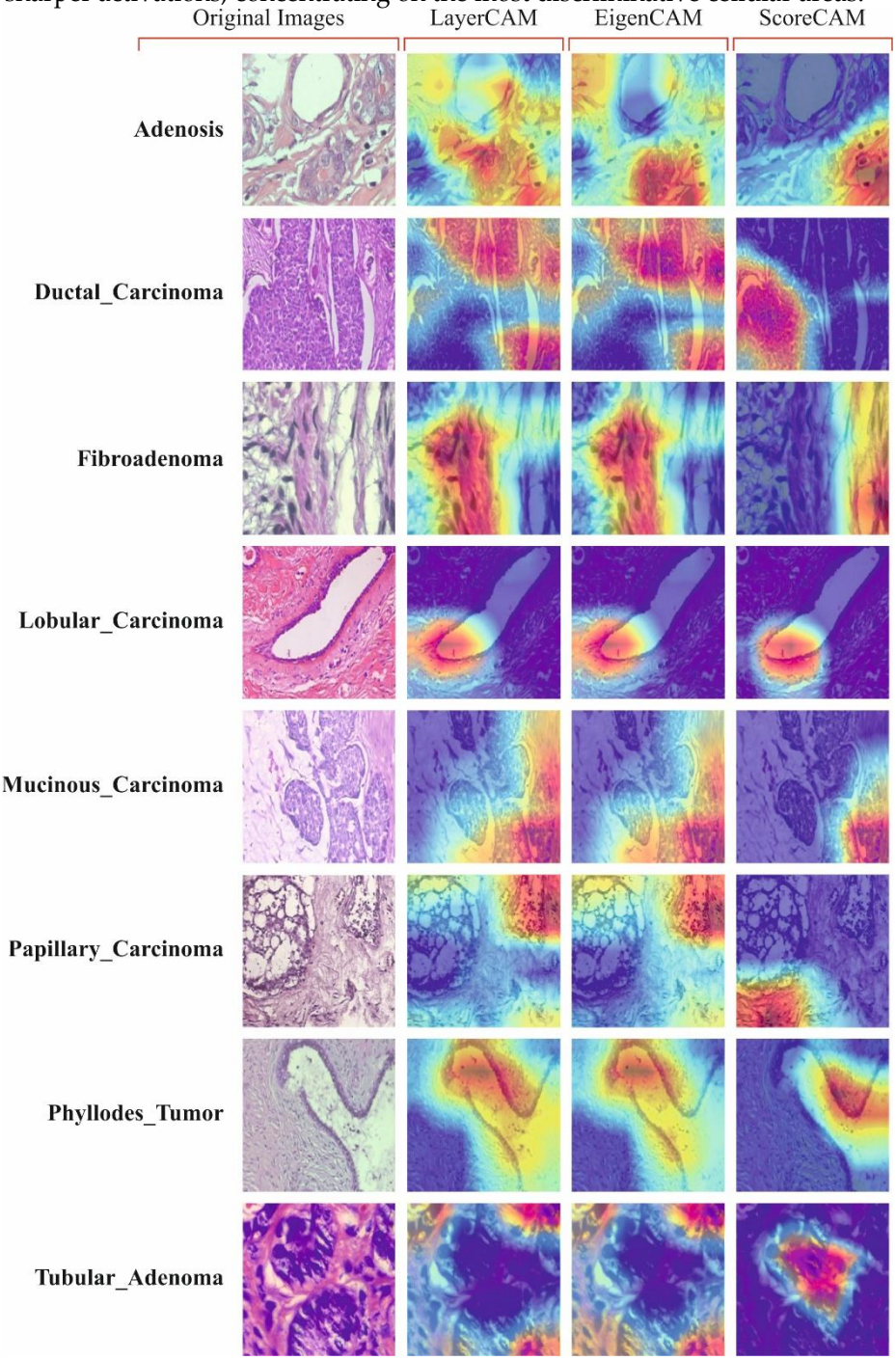

Figure 12. Comparative visual interpretation outcomes of XAI method under eight classes

5. Conclusion

In this article, the THA-XMCBCD approach has been presented to precisely detect and classify BC from HI images. Initially, the THA-XMCBCD model preprocessed the input images using color normalization, image enhancement, tissue segmentation, and pixel normalization. Subsequently, a FasterViT with hierarchical attention is employed for feature representation, capturing local cellular patterns and global tissue context. The extracted features are further fed into an H-GAT for BC classification. Moreover, the model is trained with the AdamW optimizer, which ensures steady and robust convergence. By incorporating ScoreCAM into the proposed model, the model's predictions become more explainable, supporting pathologists in confirming AI-assisted diagnoses and raising confidence in automated BC systems. To guarantee the improved performance of the THA-XMCBCD approach, a comprehensive simulation analysis is carried out, and the results prove the improvement of the THA-XMCBCD model with an accuracy of 99.04 and 98.90 under 100X and 400X magnifications. compared to other models under diverse measures for multi-class breast cancer classification.

The proposed THA-XMCBCD model may display biased learning because of the unequal distribution of instance over various BC classes, leads to minimal sensitivity and normalization performance for minority classes despite high accuracy. Future work will focus on addressing class imbalance by incorporating advanced strategies like class-balanced loss functions, synthetic data augmentation, and adaptive sampling methods to enhance minority class representation and assure more strong and clinically reliable classification performance.

Data Availability Statement: The data that support the findings of this study are openly available at https://data.mendeley.com/datasets/jxwvdwhpc2/1, reference number [26].

Declarations:

Acknowledgements: Not Applicable

Funding Information:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical Statement:

This study did not involve any human or animal subjects and, therefore, did not require ethical approval.

Conflict of Interest:

The authors declare no conflicts of interest related to this study.

Supplementary Materials:

No supplementary materials are associated with this article

References

1. Attallah, O.; Pacal, I. Impact of Magnification on Deep Learning Approaches through Comprehensive Comparative Study of Histopathological Breast Cancer Classification. Biomedical Signal Processing and Control 2025, 113, 108973, doi:10.1016/j.bspc.2025.108973.
2. Satheeshkumar, N.; Senthilkumar, S.; Rajalakshmi, S. Optimized Multi-Dimensional Attention Spiking Neural Network Based Breast Cancer Detection from Histopathology Images. *Biomedical Signal Processing and Control* **2025**, *113*, 109176, doi:10.1016/j.bspc.2025.109176.
3. Jaikumar, S.S.; Praveena, S.M. Breast Cancer Diagnosis from Histopathological Images and Molecular Signatures by Fusing Features with an Explainable AI-Based Residual Tabular Network Model. *Journal of Computer-Aided Molecular Design* **2025**, *40*, 3, doi:10.1007/s10822-025-00709-x.
4. Christobel, J.S.; Rani, K.S.S. A Novel Vision-Efficient Grad-CAM Network for Early Breast Cancer Detection Using Multi-Scale Histopathological Image Analysis. *Biomedical Signal Processing and Control* **2025**, *112*, 108939, doi:10.1016/j.bspc.2025.108939.
5. V, P.; Gunasundari, S. A Methodological Review on Breast Cancer Detection Using Deep Learning in Histopathological Images. *MDPI* **2025**, 1119–1129, doi:10.1109/icscsa66339.2025.11171246.
6. Shree, T.B.; Selvan, M.P. Deep Learning-Enabled Breast Cancer Detection from Histopathological Images in IoT Environments. Proceedings of the 3rd International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2025; pp. 1–8. https://doi.org/10.1109/ICDSNS65743.2025.11168650.
7. Shahid, M.S.; Imran, A. Breast Cancer Detection Using Deep Learning Techniques: Challenges and Future Directions. *Multimedia Tools and Applications* **2025**, *84*, 3257–3304, doi:10.1007/s11042-025-20606-7.
8. Karthiga, R.; Narasimhan, K.; V, T.; M, H.; Amirtharajan, R. Review of AI & XAI-Based Breast Cancer Diagnosis Methods Using Various Imaging Modalities. *Multimedia Tools and Applications* **2024**, *84*, 2209–2260, doi:10.1007/s11042-024-20271-2.
9. Ukwuoma, C.C.; Cai, D.; Eziefuna, E.O.; Oluwasanmi, A.; Abdi, S.F.; Muoka, G.W.; Thomas, D.; Sarpong, K. Enhancing Histopathological Medical Image Classification for Early Cancer Diagnosis Using Deep Learning and Explainable AI – LIME & SHAP. *Biomedical Signal Processing and Control* **2024**, *100*, 107014, doi:10.1016/j.bspc.2024.107014.
10. Shanmugarajeshwari, V.; Ilayaraja, M. IoT-Based Multiclass Decision Support System of Chronic Kidney Disease Using Optimal DNN. *Advances in Complex Systems* **2023**, *15*, doi:10.1142/s1793962324410022.
11. Chitta, S.; Yandrapalli, V.K.; Sharma, S. Advancing Histopathological Image Analysis: A Combined EfficientNETB7 and VIT-S16 Model for Precise Breast Cancer Detection. *MDPI* **2024**, 1–6, doi:10.1109/otcon60325.2024.10687939.
12. Saha, M.; Chakraborty, M.; Maiti, S.; Das, D. Breast-NET: A Lightweight DCNN Model for Breast Cancer Detection and Grading Using Histological Samples. *Neural Computing and Applications* **2024**, *36*, 20067–20087, doi:10.1007/s00521-024-10298-9.
13. Alzoubi, H.; Amin, M.; Aljehani, A.M.; Alhatamleh, S.; Madain, R.; Madae'en, S.; Serhan, H.A.; Alomani, A.A. Enhancing Breast Cancer Diagnosis with ResNet50 and Salp Swarm-Based Feature Reduction on BreakHis Dataset. *Biomedical Signal Processing and Control* **2025**, *114*, 109319, doi:10.1016/j.bspc.2025.109319.
14. Deka, A.; Misra, D.D.; Barbhuyan, M.F.K.; Khan, M.; Hassan, M.Md.M.; Hussain, M.A. An Improved Hybrid Feature Selection and Classification Framework for Breast Cancer Detection Using Mammography Images. *Ain Shams Engineering Journal* **2025**, *17*, 103897, doi:10.1016/j.asej.2025.103897.
15. Aldakhil, L.A.; Alhasson, H.F.; Alharbi, S.S.; Khan, R.U.; Qamar, A.M. Image-Based Breast Cancer Histopathology Classification and Diagnosis Using Deep Learning Approaches. *Applied Computational Intelligence and Soft Computing* **2025**, *2025*, doi:10.1155/acis/7011984.
16. Ponraj, A.; Nagaraj, P.; Balakrishnan, D.; Srinivasu, P.N.; Shafi, J.; Kim, W.; Ijaz, M.F. A Multi-Patch-Based Deep Learning Model with VGG19 for Breast Cancer Classifications in the Pathology Images. *Digital Health* **2025**, *11*, 20552076241313161, doi:10.1177/20552076241313161.
17. Pradeepa, M.; Sharmila, B.; Nirmala, M. A Hybrid Deep Learning Model EfficientNet with GRU for Breast Cancer Detection from Histopathology Images. *Scientific Reports* **2025**, *15*, 24633, doi:10.1038/s41598-025-00930-6.
18. Thatha, V.N.; Karthik, M.G.; Gaddam, V.G.; Krishna, D.P.; Venkataramana, S.; Lella, K.K.; Pamula, U. Histopathological Image Based Breast Cancer Diagnosis Using Deep Learning and Bio Inspired Optimization. *Scientific Reports* **2025**, *15*, 19034, doi:10.1038/s41598-025-04136-8.

19. Rahman, M.; Deb, K.; Dhar, P.K.; Shimamura, A.T. ADBNET: An Attention-Guided Deep Broad Convolutional Neural Network for the Classification of Breast Cancer Histopathology Images. *IEEE Access* **2024**, *12*, 133784–133809, doi:10.1109/access.2024.3419004.
20. Singh, P.; Kumar, R.; Gupta, M.; Al-Turjman, F. SEGEIR-NET: A Robust Histopathology Image Analysis Framework for AccurateBreast Cancer Classification. *Current Medical Imaging Formerly Current Medical Imaging Reviews* **2024**, *20*, e15734056278974, doi:10.2174/0115734056278974231211102917.
21. Yunianta, A. MSB-UNET: A Multi-Scale Bifurcation U-Net Architecture for Precise Segmentation of Breast Cancer in Histopathology Images. *Preprints.org* **2026**, doi:10.20944/preprints202601.0713.v1.
22. Hatamizadeh, A.; Heinrich, G.; Yin, H.; Tao, A.; Alvarez, J.M.; Kautz, J.; Molchanov, P. FasterViT: Fast Vision Transformers with Hierarchical Attention. *arXiv (Cornell University)* **2023**, doi:10.48550/arxiv.2306.06189.
23. Guettala, W.; Holló-Szabó, Á.L.; Gulyás, L.; Botzheim, J. Heterogeneous Graph Neural Networks for Scalable Asymmetric Traveling Salesman Problem Optimization. *Neurocomputing* **2026**, *671*, 132718, doi:10.1016/j.neucom.2026.132718.
24. Norouzi, S.; Ebrahimi, M. *A Survey on Proposed Methods to Address Adam Optimizer Deficiencies*. Technical Report, University of Toronto, Toronto, ON, Canada, 2019.
25. Zou, A. *Explaining CLIP Predictions with SingularCAM: A Class Activation Mapping Method*. Master's Thesis, University of California, San Diego, La Jolla, CA, USA, 2025.
26. Mendeley, Accessed 2025, https://data.mendeley.com/datasets/jxwvdwhpc2/1.
27. Rafiq, A.; Chursin, A.; Alrefaei, W.A.; Alsenani, T.R.; Aldehim, G.; Samee, N.A.; Menzli, L.J. Detection and Classification of Histopathological Breast Images Using a Fusion of CNN Frameworks. *Diagnostics* **2023**, *13*, 1700, doi:10.3390/diagnostics13101700.
28. Zhao, M.; Hou, C.; Cao, L.; Zhang, J. Breast Cancer Histopathological Image Classification Based on High-Order Modeling and Multi-Branch Receptive Fields. *Applied Sciences* **2025**, *15*, 6085, doi:10.3390/app15116085.