Vision Transformer for the Categorization of Breast Cancer from H&E Histopathology Images

Md Shakhawat Hossain [1,2,3,*], Ashifur Rahman [2], Munim Ahmed [2], Kaniz Fatema [1,2], MM Mahbubul Syeed [1,2], Mohammad Anowar Hussen [4], and Mohammad Faisal Uddin [1,2]

[1] Department of Computer Science and Engineering, Independent University Bangladesh, Dhaka, Bangladesh
[2] RIoT Center, Independent University Bangladesh, Dhaka, Bangladesh
[3] School of Informatics, Kochi University of Technology, Kami, Kochi 782-8502, Japan
[4] Department of Histopathology, Armed Forces Institute of Pathology, Dhaka, Bangladesh
Email: shakhawat@iub.edu.bd (M.S.H.); ashifurrahman.bubt@gmail.com (A.R.); munim@iub.edu.bd (M.A.); kaniz.fatema@iub.edu.bd (K.F.); mahbubul.syeed@iub.edu.bd (M.M.S.); anwarhussen9625@gmail.com (M.A.H.); faisal@iub.edu.bd (M.F.U.)
*Corresponding author

***Abstract*—Breast Cancer (BC) is the most frequent form of cancer, accounting for 24.5% of all cancer cases worldwide, with projections estimating 364,000 cases by 2040. Accurate diagnosis and effective categorization of BC are essential for proper treatment planning, patient management, and improved survival. Traditionally, pathologists examine histopathology specimens manually using a microscope to categorize the BC, which is labor-intensive, time-consuming, prone to subjectivity and constrained by experts' availability. An automated approach can address these limitations; however, previous methods, particularly those based on Convolutional Neural Networks (CNNs), often struggle with data imbalance, poor accuracy and poor generalizability across datasets, especially in multiclass BC categorization. This study presents an automated BC categorization method leveraging whole slide histopathology images and a transformer-based deep learning model. The proposed method uses a cascade of transformers to classify BC using 40× histopathology images, following the taxonomy defined by the BRACS dataset, distinguishing between benign, atypical, and malignant cases. First, it classifies BC into three primary categories—benign, atypical and malignant—and subsequently determines the specific sub-types within each category. The proposed method was validated using two widely recognized datasets: BRACS and BreakHis. On BRACS, it achieved 95.6% accuracy in classifying BC into benign, atypical, and malignant categories, with sub-type accuracies of 94.7% for benign, 98.6% for atypical, and 99.1% for malignant cases. On the BreakHis dataset, the model achieved 93% accuracy for binary benign-malignant classification, with sub-type accuracies of 94% and 91% for benign and malignant cases, respectively. The proposed method outperformed existing methods in accuracy and robustness, making it a promising tool for automated BC diagnosis and classification.**

***Keywords*—breast cancer, whole slide image, vision transformer, histopathology image, cancer classification**

I. Introduction

Breast Cancer (BC) is one of the major causes of cancer related deaths among women worldwide. In 2022, approximately 2.3 million women were diagnosed with BC and 670,000 among them died worldwide [1]. It is estimated that 310,720 women in the United States alone will be diagnosed with Invasive Breast Cancer (IVC) in 2024, while 56,500 women will be diagnosed with Ductal Carcinoma in Situ (DCIS). Approximately 42,250 of them are predicted to die from this cancer [2]. According to the American Cancer Society, one in every eight women in the United States is diagnosed with BC at some point in their lifetime [2]. Accurate BC diagnosis and categorization play a crucial role in reducing cancer deaths by enabling personalized treatment and effective patient management. Several countries reported that precise diagnosis and efficient categorization improves treatment planning and patient management, which could achieve an annual breast cancer mortality reduction of 2–4% per year [3–5].

Traditionally, the diagnosis and categorization of BC are performed based on the Hematoxylin and Eosin (H&E) histopathology images. At first, the biopsy is performed to collect tissue from the suspected region, which is then processed in multiple steps, which include embedding, sectioning, mounting and staining to prepare the specimen for microscopic observations. H&E is considered the gold standard staining for histopathology examination, which involves sequential tissue immersion in hematoxylin and eosin solutions. Hematoxylin is a basic dye that binds to the tissue's deoxyribonucleic acid and stains cell nuclei blue-purple. On the other hand, eosin is an acidic dye which binds to the cytoplasmic and stromal tissue proteins, giving them a pink stain. This highlights cellular and tissue structures, allowing for visualization under a microscope. Pathologists manually examine the H&E-stained specimen to observe the tissue structure, cellular morphology, nuclei size and shape, stromal changes, mitotic activity and other pathological changes and abnormalities for the diagnosis and categorization of cancer. However, such manual examination requires experience and much time and effort. Moreover, such diagnosis often suffers from inter-

Manuscript received February 12, 2025; revised April 21, 2025; accepted May 8, 2025; published August 7, 2025.

observer variability.

With the recent advancements in Artificial Intelligence (AI) and computational techniques, along with the digitization of histopathology specimens, it is now possible to automatically analyze histopathology images for accurate cancer categorization and diagnosis without demanding pathologists' direct supervision. The WSI scanner converts the entire histopathology specimen into high-resolution digital images, which can be observed on a computer screen and processed using computer algorithms, particularly AI methods. Several AI-based methods were proposed for the diagnosis and categorization of BC patients. These methods include traditional machine learning models, deep learning models and the combination of these models. Most of these methods were developed based on the publicly available BC dataset BreakHis [6] and BRACS [7]. Both datasets were prepared by breast pathologists and widely used for developing BC categorization.

Traditional machine learning model, such as the Support Vector Machine (SVM) with empirically selected features, was also utilized for BC classification [8]. Although this type of network is suitable for binary classification with the limited dataset, they failed to achieve good accuracy for multi-class BC categorization. Convolutional Neural Networks (CNN) based architecture, a sub-group of deep learning that can learn spatial relationships in images, was found very effective for BC histopathology image analysis [6, 9]. Hybrid networks combining multiple deep learning and traditional machine learning models were also applied for BC classification [10, 11].

Although CNNs have been widely used in histopathological image analysis, they exhibit several limitations that hinder their performance in categorizing BC. CNNs rely on fixed-size receptive fields, which restrict their ability to capture long-range spatial dependencies and global context which is crucial for interpreting complex tissue structures. Their performance depends heavily on extensive data augmentation and large, well-annotated datasets to generalize effectively. However, publicly available digital pathology datasets frequently suffer from class imbalance, particularly among rare BC sub-types, which impairs the learning of minority classes and reduces overall classification accuracy. Moreover, CNN-based models often demonstrate poor generalization across datasets obtained from different institutions or under varying staining conditions, limiting their robustness in real-world clinical settings. These challenges highlight the need for alternative deep learning architectures that are more context-aware and capable of modeling local and global features [12]. Transformers, particularly Vision Transformers (ViTs), address these limitations through self-attention mechanisms that enable the modeling of complex patterns and long-range dependencies [13]. ViTs have shown promising results in recent medical imaging studies but remain underexplored in BC categorization using Hematoxylin and Eosin (H&E) stained histopathology images. This capability is especially valuable for processing high-resolution Whole Slide Images (WSIs), which can reach dimensions of up to 100,000×100,000 pixels [14].

This paper presents a BC categorization method utilizing a ViT-based model. The proposed method first divided the BC patients into benign, atypical and malignant classes. Then, it determined the sub-categories of each class of patients. Benign BC was subdivided into normal, Pathological Benign (PB) and Usual Ductal Hyperplasia (UDH). Atypical cancers were divided into Flat Epithelial Atypia (FEA) and Atypical Ductal Hyperplasia (ADH). Finally, the malignant cancers were sub-typed as DCIS and IVC. Such a BC taxonomy is provided by the BRACS dataset, which is helpful in differentiating the atypical cancers from benign and malignant cases for appropriate clinical treatment and surgical planning [15, 16]. Identifying atypical cancers is essential as they pose a high possibility of developing future DCIS or IVC. More importantly, these lesions remain undetected in mammography and other imaging techniques, even if they cannot be identified in a physical breast examination. Therefore, in this study, we designed the method to categorize BC according to the taxonomy given by BRACS and trained the transformers using BRACS dataset. Further, we have demonstrated the efficacy of the proposed method for the BreakHis dataset, which classifies the BC into eight classes. According to the BreakHis, BC patients are divided into benign and malignant categories. Further, the benign patients are subdivided into adenosis, fibroadenoma, phyllodes tumor and tubular adenoma classes. The malignant patients are subdivided into ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma. The proposed transformer-based BC categorization method achieved high accuracy in the BRACS and BreakHis datasets, outperforming the previously proposed methods. This ensures the robustness of the method. Moreover, this method utilized the 40× magnification WSI, which allows faster diagnosis than the methods that relied on 100×, 200× or 400× WSI. We used 40× magnification histopathology images for training and evaluation. Although prior studies often rely on higher magnifications such as 100×, 200×, or 400× to capture detailed tissue structures, we chose 40× to balance computational efficiency and diagnostic precision. High-resolution images are essential for capturing cellular morphology, gland formation, nuclear features, and other histological patterns critical for cancer classification. In contrast, lower-resolution images (e.g., 5× or 10×) generally lack such fine details and are unsuitable for diagnostic purposes; instead, they serve to identify artifacts and pen marks or assess overall tissue coverage. Using 40× images allows our model to retain sufficient diagnostic information while ensuring practical computational requirements.

Thus, the main contributions of the paper are 1) the development of an accurate BC classification method based on the H&E histopathology images, 2) the investigation of the performance of the transformer model for BC classification, 3) the demonstration of the method using two widely used BC classification dataset to ensure its robustness and generalized performance.

II. Literature Review

Accurate diagnosis of BC and categorizing them into the appropriate sub-types enable to control the development of tumor cells into malignant cancers. However, the manual examination-based analysis of histopathology images for categorizing the BC into multiple classes and sub-classes is challenging due to the inter-observer variability, lack of experienced pathologists and tedious and lengthy process. Several automated methods were proposed to analyze histopathology images, mainly utilizing AI. These methods are capable of fixing the issues related to the manual examination. However, these methods require high-magnification images, lack accuracy and fail when implemented for images prepared in a different laboratory.

AI-based automated BC categorization methods can be divided into three groups: traditional machine learning based methods, deep learning-based methods, and hybrid methods. Traditional machine learning-based methods include SVM, logistic regression, Linear Discriminant Analyzer (LDA), K-Nearest Neighbor (KNN), and random forest models. These models are suitable for training binary classifiers using comparatively smaller datasets; however, considering the complexity of a multiclass categorization problem, these methods are not suitable for producing high accuracy when trained with a limited dataset.

In 2015, Spanhol *et al.* [6] published the BreakHis dataset. They investigated the performance of Random Forest (RF), KNN, LDA and SVM classifiers for eight class multiclass BC categorization based on the handcrafted features. The SVM produced the highest accuracy of 85.2 using 200× images. Singh *et al.* [17] proposed an SVM-based binary classifier for classifying BC patients into benign and malignant classes, which achieved 92.3%. Belsare *et al.* [18] proposed a LDA based binary classifier which utilized Gray Level Co-occurrence Matrix (GLCM) and Graph Run Length Matrix (GRLM) based features to detect the malignant cancer with 80% accuracy. Aswathy *et al.* [19] investigated the performance of SVM, KNN and RF methods for classifying the BC histopathology images as benign and malignant based on the handcrafted features. This method found SVM the most accurate, which yielded 89.1% accuracy for both BreakHis and the University of California Santa Barbara (UCSB) dataset. In a separate study, Chan *et al.* [8] applied SVM to classify the BreakHis dataset into eight classes, which produced only 55.5% accuracy. These studies express that SVM is mostly used among the traditional classifiers for BC categorization. The studies also indicate the limitation of traditional classifiers for multiclass BC classification.

Deep learning-based methods, particularly CNN, have been deployed by many researchers and reported to achieve higher accuracy than traditional classifiers. The publisher of the BreakHis dataset also investigated the performance of CNN models on their dataset [20]. They implemented a pre-trained CNN model, CaffeNet, on 200× images to improve the classification accuracy to 88.7%. In a separate study, they implemented a pretrained AlexNet model, which yielded 82.7% accuracy for 200× images [21]. Han *et al.* [22] proposed a class structure-based CNN classifier, a modified version of GoogLeNet architecture. They tested the performance of this model for binary and multiclass BC categorization using the BreakHis dataset. The accuracy was 96.9 and 93.9 using 100× images for binary and multiclass classification, respectively. Motlagh *et al.* [23] proposed a ResNetbased CNN model using 40× images, which achieved 98.7% accuracy in benign and malignant binary classification. This method achieved 94.8% and 96.4% accuracy for further classifying benign and malignant cases into sub-classes. Golatkar *et al.* [24] also proposed a multiclass BC categorization method utilizing the inception-based CNN model. This method was trained using 20× images of Breast Cancer Histology images Challenge dataset (BACH) [25]. The model produced 93% accuracy for benign-malignant binary classification. This is significant considering the magnification of the image. However, the accuracy of the model dropped to 85% for normal, benign, DCIS and IVC multiclass classification. Jiang *et al.* [9] proposed another ResNet-based model for the BreakHis dataset. They integrated squeeze and excitation block with the ResNet architecture, unlike the [24]. However, it produced little improvement in accuracy. Gour *et al.* [26] also utilized residual block but increased the number of layers by 152. The accuracy of this ResNet model was lower (84.3%) compared to the squeeze and excitation-based ResNet model for benign-malignant binary classification (98.8%). Another ResNet-based model was proposed by Zewdie *et al.* [27]. Their model achieved approximately 96% accuracy in benign malignant binary classification and in sub-typing the benign and malignant classes.

Brancati *et al.* [7] developed a custom CNN-based model, which was trained and tested using the BRACS dataset. They categorized the BC patients in two steps: firstly, the patients were divided into benign, atypical and malignant classes with 70.3% accuracy and then, each class was categorized into appropriate sub-classes with a maximum accuracy of 69.6%. Although this method utilized two-stage classification, which is suitable for achieving higher accuracy in this type of multi-class problem, they failed to achieve sufficient accuracy. This method used 40× images. Another ResNet-based method was proposed by Fahad *et al.* [28] utilizing the BRACS dataset. Unlike Brancati's method, they classified the BC into seven classes rather than implementing a two-stage classification. However, the accuracy was 96.2%, which is significantly better. This method also used 40× images. Chu *et al.* [29] proposed a multiple instance learning based method for the same purpose with achieved 87.6% accuracy. These studies demonstrate the superiority of deep learning-based methods, particularly CNN models, over traditional machine learning models for BC categorization. Although the CNN-based methods achieved higher accuracy, they required more extensive training data and computational power. Alternatively, the traditional models can be trained using features derived from limited data and computational resources, but they

lack accuracy.

Therefore, some researchers proposed hybrid networks combining multiple CNN models or traditional classifiers with CNN models. Deniz *et al.* [30] combined the feature extraction capability of CNN models with low computation classification of traditional models. In their work, they utilized AlexNet and VGG16 models for extracting features from histopathology images, which were then used by an SVM classifier to predict the class of BC. This approach produced 89.0% accuracy for the binary classification of the BreakHis dataset using 100× images. Another issue of training the CNN models was over-fitting. Bardou *et al.* [10] combined multiple CNN models to create an ensemble of models suitable for handling over-fitting problems. This ensemble model achieved a significantly high accuracy of 97.8% for the binary classification of the BreakHis dataset. Other hybrid networks were proposed by Nahid *et al.* [11] and Tasleem *et al.* [31]. They also did not achieve significant improvement compared to the CNN-based models.

III. Materials and Methodology

A. Ethics Statement

In this study, we used anonymized human biopsy data. We obtained Institutional Review (IRB) approval from the Independent University Bangladesh Research Ethics Committee and conducted the research following the IRB (IRB-2022-SETS-06) guidelines.

B. Dataset

In this study, we have utilized the two most commonly used and publicly available histopathology image datasets of breast cancer, which are BRACS [7] and BreakHis [6].

In this study, we followed the BC taxonomy provided by the BRACS dataset and primarily trained and validated the proposed method using the BRACS dataset. Then, the method was tested on the BreakHis dataset to evaluate the robustness and generalization ability. We also trained and tested the model using BreakHis to ensure the efficacy of the transformer-based model regardless of the dataset.

The BRACS dataset contains 4539 labelled image patches extracted from 547 whole slide images belonging to 189 breast cancer patients. The WSIs were scanned using an Aperio AT2 scanner at 40× magnification to provide a 0.25 μm/pixel resolution. The 4539 images included 1837 benign, 1263 atypical and 1439 malignant images. The benign class included 484 healthy or normal, 836 PB and 517 UDH images. The atypical class included 756 FEA and 507 ADH images. The malignant class included 790 DCIS and 649 IVC images. The distribution of images in each class and the distribution of images for training, validating and testing each classifier is given in Fig. 1.

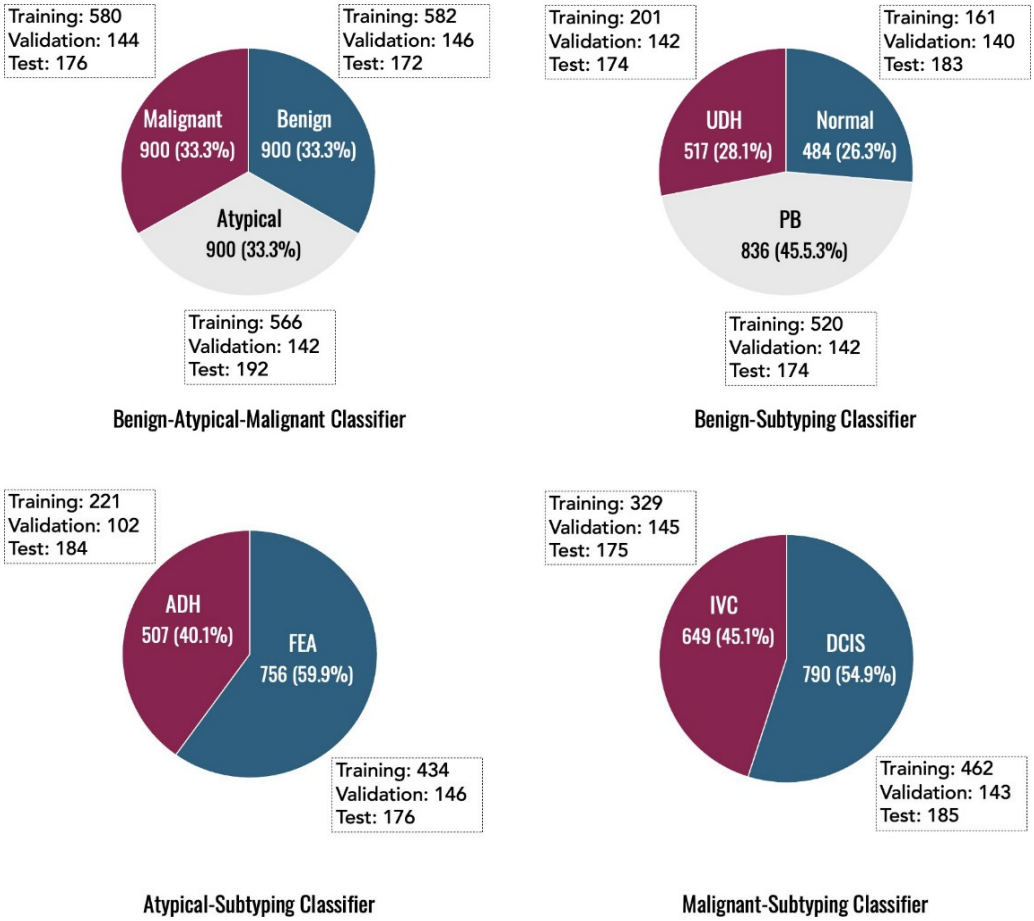

Fig. 1. Distribution of BRACS images for training, validating and testing the classifiers.

The BRACS-trained transformer model was tested on 545 images of the BreakHis dataset, which included 176 benign and 369 malignant images. Later, the model was trained, validated, and tested using 2113 images of the BreakHis dataset. Among these images, 1148 images were used for training, validating, and testing the benign-malignant classifier, 468 images for the benign sub-type classifier, and 500 images for the malignant sub-type classifier. These images were selected randomly from the BreakHis dataset.

C. Architecture of the Proposed System

The proposed system trained four ViT models independently to accomplish benign-atypical-malignant classification, benign sub-type, atypical sub-type and malignant sub-type classification. Firstly, we plotted the BRACS dataset using t-distributed Stochastic Neighbor Embedding, an unsupervised non-linear dimensionality reduction technique for visualizing high-dimensional data. This was done separately for seven classes and three classes (benign, atypical and malignant), as shown in Fig. 2 using the same parameters. Fig. 2 shows the two-dimensional projection of seven-class and three-class data distribution. The three-class projection shows better separable clusters than the seven-class one. This indicates that classifying the BRACS dataset into benign, atypical and malignant classes is less complex than directly applying seven class classifications. Therefore, we implemented the classifiers in two stages to create a cascade of transformers, as illustrated in Fig. 3. The proposed system starts by evaluating the sharpness of the image. WSI images often suffer from out-of-focus problems due to the stage alignment problem or other related hardware issues [32]. Therefore, the sharpness of the images was estimated to determine the focus quality of the image. The sharpness was calculated based on the average width of edges in the image according to Eq. (1). Sharp edges have abrupt intensity changes over a small spatial region, resulting in lower edge width. In constant, blurred edges have gradual intensity transitions over a larger spatial region, leading to higher edge width. Therefore, a sharp image results in a lower average edge width according to Eq. (1).

$$\text{Average edge width} = \frac{1}{N}\sum_{i=1}^{N} w(i) \quad (1)$$

In Eq. (1) N is the number of detected edges in the image, $w(i)$ is the width of edge i.

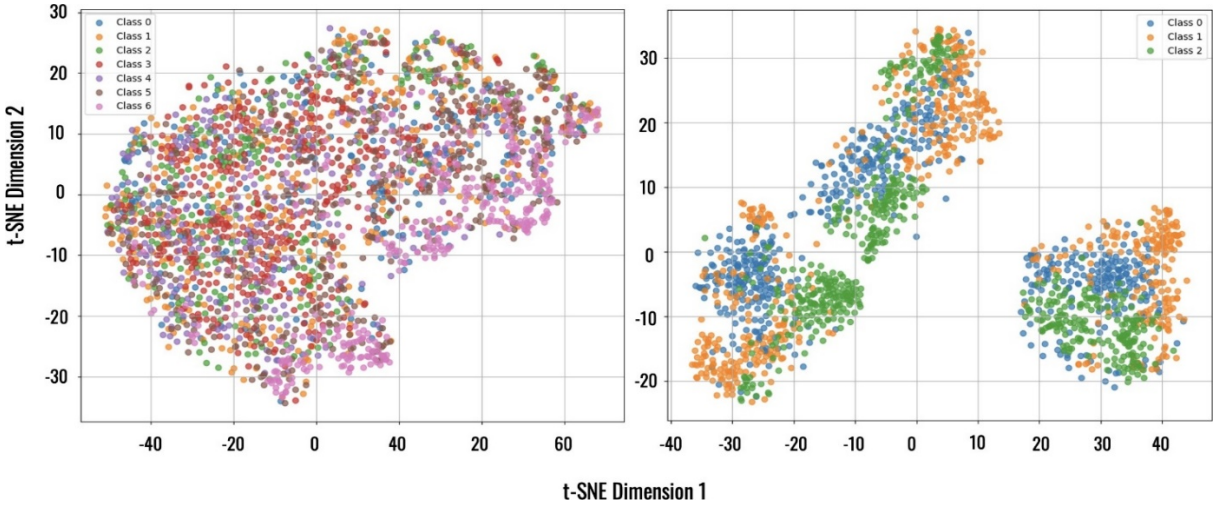

Fig. 2. Visualization of BRACS dataset by t-distributed Stochastic Neighbor Embedding: seven class (left) and three class data distribution which included benign, atypical and malignant classes (right).

Our experiment found that an image having an out of-focus issue typically results in an average width higher than 6. Consequently, such images were rejected for analysis. After sharpness evaluation, the proposed method normalized the RGB color values of the image.

Color variation is a common issue in pathology image analysis. It often impacts the effectiveness of AI models, particularly when the models are used on images scanned with a different WSI scanner than the one they were trained on. Therefore, this method transforms the RGB color values to the sRGB color space to compensate for the color variation caused by different WSI scanners. Sharpness evaluation and color normalization comprise the system's pre-processing steps, which allow the method to handle images scanned by various scanners and prepared in various laboratories. These steps are related to the system's generalization ability.

After color normalization, the stage one classifier classifies the image into one of three categories: benign, atypical and malignant. If the image is classified as benign in the first stage, then it is processed using only a benign sub-typing classifier in the second stage. Similarly, in the second stage, atypical sub-typing and malignant sub-typing classifiers are selected in the input image and classified as atypical and malignant cancer by the stage one classifier. The selection of a second stage classifier for an image is dependent on the result of the stage one classifier. Thus, the stage one classifier plays a crucial role in the proposed system.

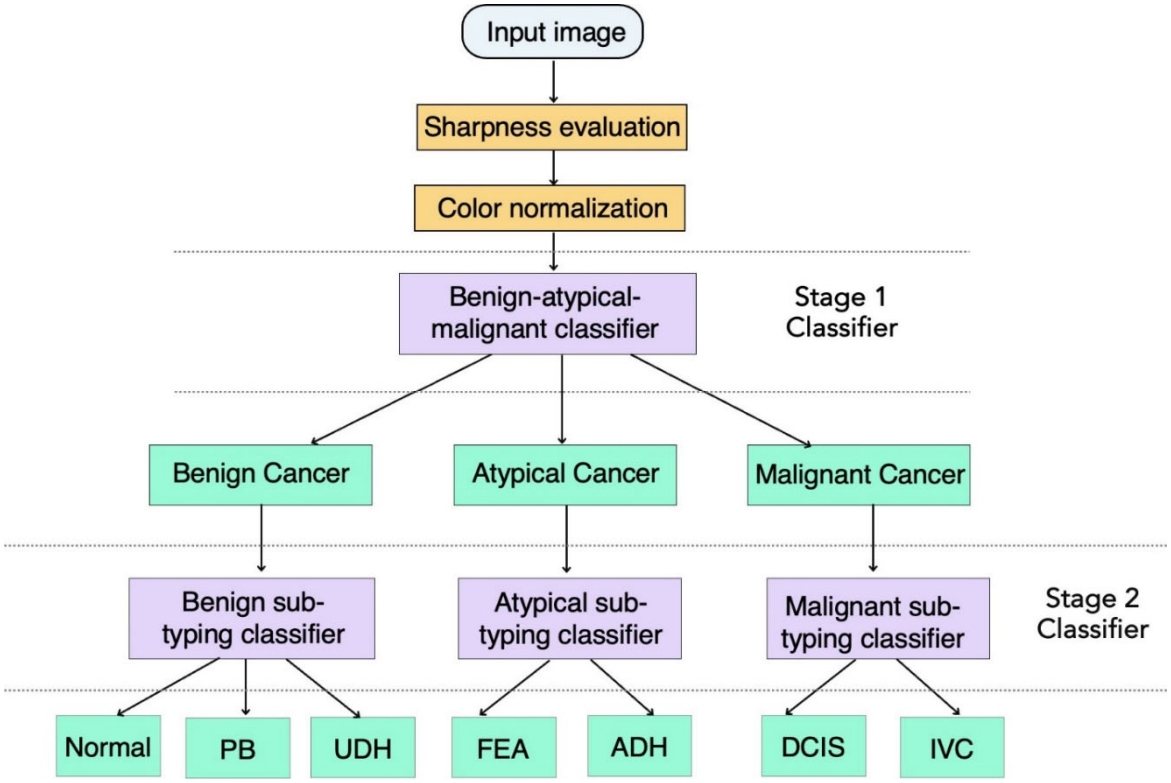

Fig. 3. Flowchart of the proposed method for BC categorization.

D. Model Training and Evaluation

The proposed method was designed to categorize BC according to the BRACS taxonomy. Therefore, the method was mainly trained, validated, and tested using the BRACS dataset. Then, the BRACS-trained model was also tested on the BreakHis dataset. Although the classes of the BreakHis and BRACS datasets do not align completely, we applied the BRACS-trained stage one benign atypical-malignant classifier on the benign and malignant class images of the BreakHis dataset. This evaluated the robustness of the proposed transformer-based method. After that, we also trained and tested the transformers using the BreakHis dataset to show the effectiveness of the proposed transformer-based method. The proposed method independently trained four transformer models using the BRACS dataset for benign atypical-malignant, benign sub-type, atypical sub-type and malignant sub-type classification. We utilized the ViT models and then fine-tuned them by optimizing their hyperparameters to select the best transformer network for each classifier.

Transformer architecture is popular in processing natural languages; recently, it has been structured to process images. ViT is one of the very first transformer architectures applied on the images [14]. The core concept of transformer architecture is to apply self-attention to understand the relationships between different parts or patches of the input sequence. In the case of image-based transformers, the input image is divided into a sequence of patches, and then the attention is applied to the patches to capture the significance of each patch of the image for predicting the output. This patch-wise attention enables the capture of both local and global spatial information from the image. In ViT, the input image is divided into fixed-size patches, which are then flattened and linearly embedded to obtain token representations. The transformer encoder processes these token representations. The encoder consists of multiple Multi-Head Self-Attention (MHSA) layers and Feedforward Neural Networks (FFN). FFN layers have two times more weight than the MSHA layers. Fine-tuning the FFN layers is, therefore, time-consuming. Limiting the fine-tuning to the MHSA layer allows modifying only a relatively small number of parameters. Thus, the training time becomes short. Fine-tuning FFN is time consuming and is recommended when achieving high accuracy, which is challenging. Finally, a linear classifier is used at the top layer to predict the class labels of the input. The classification head is basically a two-layer multi-layer perception network.

The ViT model comes in three variants: base model (ViT-B), large model (ViT-L) and huge (ViT-H) model. The base model has 12 layers, 12 heads and 86M parameters. The large and huge models have more parameters and require much training time. This study aims to evaluate the suitability of transformer models for histopathology image-based BC categorization; therefore, in this study, we utilized the base model, finetuned only the MSHA layers of the model and experimented with the commonly used hyperparameters to keep the network simple. This allowed us to achieve adequate accuracy without long training and high computational resources. Table I shows the hyperparameter optimization space of the transformer models. This network optimization approach was followed to develop each type of classifier by training the transformers independently. Then, the

network that produced the highest test accuracy for each type of classification was selected. After that, we plotted the Receiver Operating Characteristic (ROC) curves for the classifiers and performed a 5-fold cross-validation experiment. Finally, the BC categorization results produced by the selected transformer-based classifier were compared with the previously proposed methods.

TABLE I. LIST OF HYPERPARAMETERS AND THEIR VALUES EXPLORED TO FINE-TUNE THE CNN AND TRANSFORMER NETWORKS

Criteria	Search space
Models	[ViT-B/16, ViT-B/32]
Pre-training datasets	[ImageNet-1K]
Epochs	[50, 75, 100]
Batch sizes	[16, 32]
Patch sizes	[16, 32]
Optimizers	[AdamW, SGD]
Loss functions	[Categorical Cross Entropy]
Learning rates	[0.03, 0.01, 0.001]
Weight decay	[0.001, 0.0001]
Transformer layers	[8, 16, 32, 64]
MLP head units	[1024, 2048]

IV. RESULTS

A. Suitability of Transformers

Our study aimed to develop a comprehensive set of classifiers using the ViT-based image transformer for BC categorization. We experimented with various hyperparameters, creating multiple versions of the base model. These networks were then trained on the BRACS dataset, which aligns with the BC taxonomy we adopted. This resulted in four distinct classifiers: benign-atypical-malignant, benign sub-type, atypical sub-type, and malignant subtype classifiers, each serving a unique purpose in the proposed system. The best transformer networks were selected based on their test accuracy, 5-fold cross-validation accuracy, and Area Under the Curve (AUC) value for each classifier.

At first, the appropriate images for training specific classifiers were pre-processed. Then, the networks were trained, validated and tested using these images to select the best-fine-tuned network through the holdout validation. For example, in the case of a holdout validation experiment for selecting the benign-atypical malignant classifier, 1728, 432 and 540 pre-processed images were used for training, validating and testing the networks, shown in Fig. 1. The 1728 training images included 582 benign, 580 malignant and 566 atypical images. The 432 validation images include 146 benign, 144 malignant and 142 atypical images. Lastly, the 540 test images included 172 benign, 176 malignant and 192 atypical images. The test images were unseen to the networks during training and validation. This holdout validation was done independently to select the benign-atypical-malignant, benign sub-type, atypical sub-type, and malignant sub-type classifier.

The dataset distribution for the holdout validation experiment is given in Fig. 1. Table II shows the performance evaluation of the four ViT-based classifiers. It shows that the benign-atypical-malignant, stage-one classifier achieved approximately 95% test accuracy. The accuracy of the stage one classifier is significant as it is related to the performance of all the stage two classifiers. The test accuracy of the second-stage benign sub-type classifier (94.7%) was slightly lower than the stage one classifier. However, the test accuracy of other second-stage classifiers was significantly better, higher than 98%. The loss values were low in the training and validation, which indicated that the classifiers did not fit the criteria. Table II also indicates that benign subtype classification is more challenging than other BC categorizations.

TABLE II. PERFORMANCE EVALUATION OF THE VIT MODEL FOR BC CLASSIFICATION AND SUBTYPING USING BRACS DATASET

Metrics	Benign-Atypical-Malignant Classifier	Benign-Subtyping Classifier	Atypical-Subtyping Classifier	Malignant-Subtyping Classifier
Training accuracy	0.967	0.948	0.980	0.991
Training loss	0.093	0.164	0.057	0.028
Validation accuracy	0.951	0.936	0.975	0.989
Validation loss	0.153	0.194	0.099	0.029
Average test accuracy	0.956	0.947	0.986	0.991
Average test precision	0.955	0.947	0.977	0.989
Average test recall	0.955	0.946	0.994	0.994
Average 5-fold cross validation accuracy	0.928	0.923	0.951	0.961
Macro average AUC	0.990	0.980	1.00	1.00
Micro average AUC	0.990	0.980	1.00	1.00

After that, we performed the 5-fold cross-validation experiment for each classifier. For this purpose, we divided the images into five different groups. For example, in the 5-fold cross-validation of the benign-atypical malignant classifier, we split the 2700 images into five groups. Each group contained 540 images, which included 180 images for each class. The images were assigned randomly to each group to contain 180 benign, 180 atypical and 180 malignant images. Then, the classifier was trained using four groups of images and tested using the images of the remaining group, termed as a fold. This was done to ensure that each group is used once for testing the network, resulting in a 5-fold training and test of the classifier. We calculated the accuracy for each fold and derived the average accuracy, which is crucial to evaluating the classifiers' effectiveness. Table III shows the 5-fold cross-validation accuracies for the classifiers. The accuracies were higher than 92% with a minimal standard deviation, indicating the classifiers' stability. In 5-fold cross-validation, the malignant sub-type classifier achieved the highest accuracy and the benign sub-type classifier the lowest, similar to the holdout validation.

TABLE III. FIVE-FOLD CROSS VALIDATION ACCURACY FOR BRACS DATASET

Metrics	Benign-Atypical-Malignant Classifier	Benign-Subtyping Classifier	Atypical-Subtyping Classifier	Malignant-Subtyping Classifier
Fold 1	0.924	0.920	0.938	0.963
Fold 2	0.946	0.937	0.950	0.916
Fold 3	0.916	0.905	0.983	0.994
Fold 4	0.901	0.922	0.963	0.966
Fold 5	0.951	0.930	0.922	0.963
Average ± standard deviation	0.928 ± 0.018	0.923 ± 0.010	0.951 ± 0.020	0.961 ± 0.025

Fig. 4 shows the classifiers' accuracy and loss curves during holdout training validation. Fig. 4 indicates that the accuracies increased and losses decreased with the epochs for all. Meaning the networks were well-trained without over-fitting. Figs. 5 and 6 show the confusion matrices and ROC curves of the classifiers. The AUC values were higher than 0.98 for all the classifiers. The confusion matrices reveal that the number of false positives and false negatives was very low for the classifiers, particularly atypical and malignant sub-type classifiers. Consequently, the precision and recall values were significantly high. The precision was 97% and 99% for atypical and malignant sub-type classifiers. The recall was 99% for both. The findings from our holdout and 5-fold cross-validation experiment are presented in Tables II and III, Figs. 4–6 which clearly indicate that the transformer-based networks achieve sufficient accuracy for BC categorization when trained using BRACS dataset which is one of the most widely used and accepted BC categorization dataset.

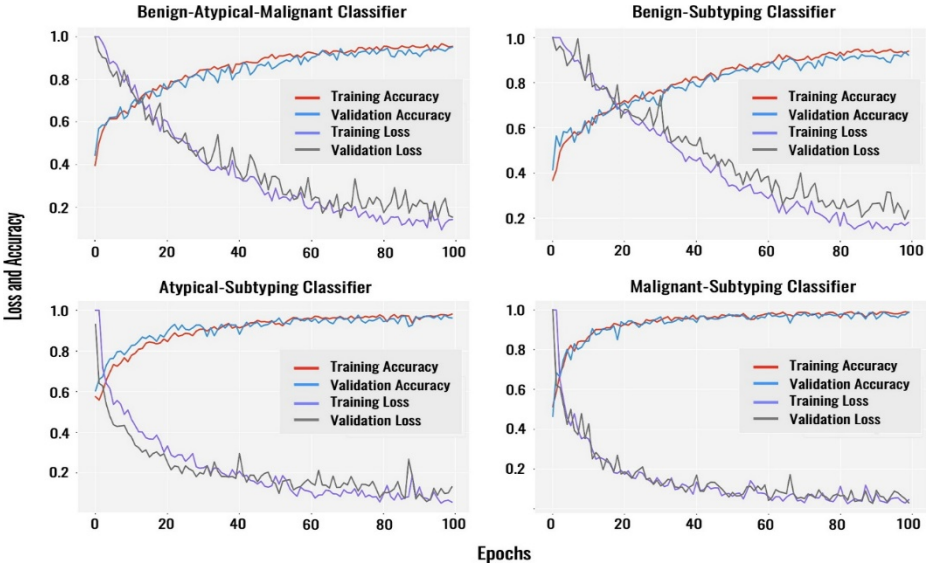

Fig. 4. Training and validation curves of holdout validation using BRACS dataset.

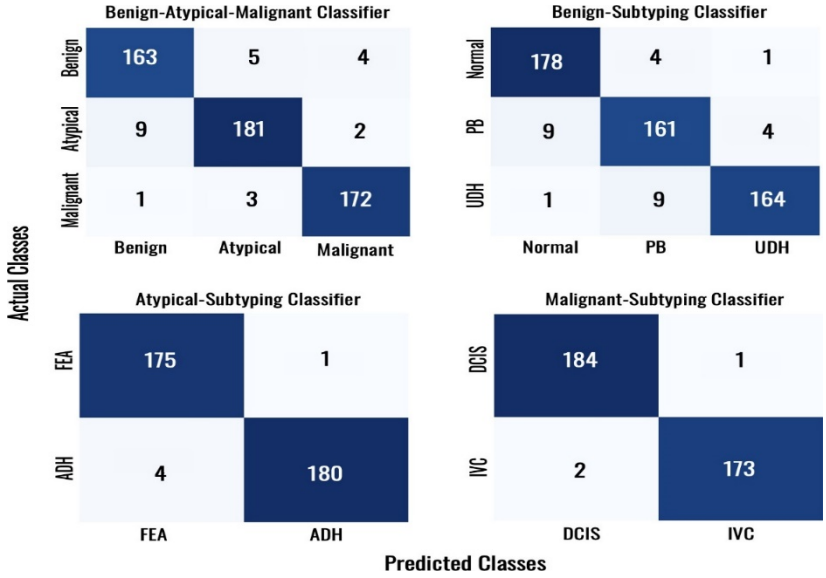

Fig. 5. Confusion matrices for the test images of BRACS dataset.

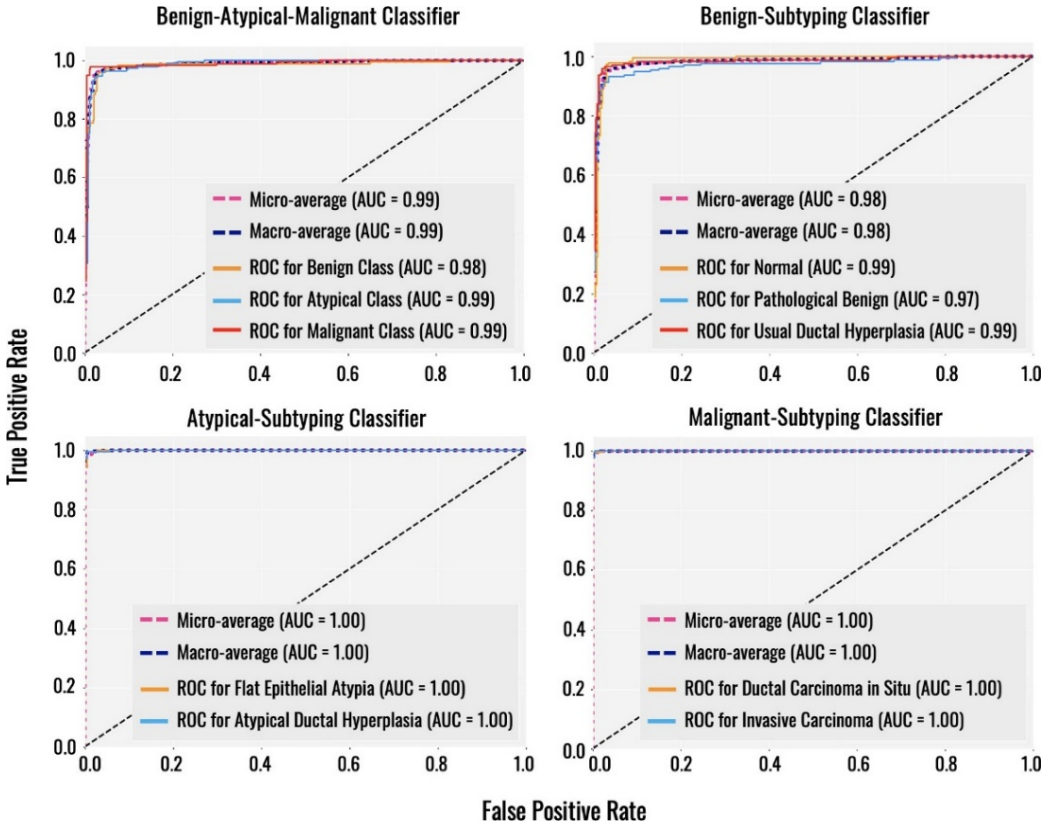

Fig. 6. ROC curves of the proposed method for BRACS dataset.

Finally, we compared the proposed transformer-based BC categorization method with the existing methods proposed previously for this purpose. Table IV shows the comparison. This Table shows that the transformer-based method outperformed the existing methods even when trained using a comparatively low magnification images of 40×. Moreover, the proposed method was tested using heterogeneous dataset to confirm it robustness and generalized performance.

TABLE IV. COMPARISON BETWEEN THE EXISTING METHODS AND PROPOSED METHOD

Methods	Technology	Data description	Classification accuracy %
Spanhol *et al.* [6] (2015)	SVM	Dataset: BreakHis [6]; Image Mag: 200×	Benign-Malignant eight sub-classes: 85.2
Chan *et al.* [8] (2016)	SVM	Dataset: BreakHis [6]; Image Mag: 40×	Benign-Malignant eight sub-classes: 55.6
Singh *et al.* [16] (2020)	SVM	Dataset: BreakHis [6]; Image Mag: 40×	Benign/Malignant: 92.3
Aswathy *et al.* [18] (2021)	SVM	Dataset: BreakHis [6]; Image Mag: 40×	Benign/Malignant: 89.1
Spanhol *et al.* [19] (2017)	CaffeNet CNN	Dataset: BreakHis [6]; Image Mag: 200×	Benign-Malignant eight sub-classes: 88.7
Spanhol *et al.* [20] (2017)	AlexNet CNN	Dataset: BreakHis [6]; Image Mag: 200×	Benign-Malignant eight sub-classes: 82.7
Han *et al.* [21] (2017)	Class structure-based deep CNN	Dataset: BreakHis [6]; Image Mag: 100×	Benign/Malignant: 96.9, Benign-Malignant eight sub-classes: 93.9
Motlagh *et al.* [22] (2018)	ResNet-based CNN	Dataset: BreakHis [6]; Image Mag: 40×	Benign/Malignant: 98.7, Benign sub-classes: 94.8, Malignant sub-classes: 96.4
Golatkar *et al.* [23] (2018)	Inception-based CNN	Dataset: BACH [24]; Image Mag: 20×	Benign/Malignant: 93, Normal/Benign/In-situ Carcinoma/Invasive Carcinoma: 85
Jiang *et al.* [9] (2019)	Squeeze-Excitation block-based ResNet CNN	Dataset: BreakHis [6]; Image Mag: 40×	Benign/Malignant: 98.8, Benign-Malignant eight sub-classes: 94.4

Gour *et al.* [25] (2020)	Residual learning-based 152-layered CNN	Dataset: BreakHis [6]; Image Mag: 40×	Benign/Malignant: 84.3
Zewdie *et al.* [26] (2021)	ResNet50 CNN	Dataset: BreakHis [6] and Private dataset; Image Mag: 40×	Benign/Malignant: 96.7, Benign sub-classes: 96.7, Malignant sub-classes: 95.7
Brancati *et al.* [7] (2022)	Custom CNN with resid- ual structure	Dataset: BRACS [7]; Image Mag: 40×	Benign/Atypical/ Malignant: 70.3, Benign sub-classes: 50.9, Atypical sub-classes: 44.0, Malignant sub-classes: 69.6
Fahad *et al.* [27] (2023)	ResNet50-based CNN	Dataset: BRACS [7]; Image Mag: 40×	Benign-Atypical-Malignant seven sub-classes: 96.2
Chu *et al.* [28] (2024)	Retentive Multiple In- stance Learning	Dataset: BRACS [7]; Image Mag: 40×	Benign-Atypical-Malignant seven sub-classes: 87.6
Deniz *et al.* [29] (2018)	AlexNet-VGG16 features with SVM	Dataset: BreakHis [6]; Image Mag: 100×	Benign/Malignant: 86.0
Bardou *et al.* [10] (2018)	Ensemble CNN Model	Dataset: BreakHis [6]; Image Mag: 40×, 100×, 200× and 400×	Benign/Malignant: 97.8, Benign-Malignant eight sub-classes: 88.2
Nahid *et al.* [11] (2018)	Combination of Kmeans, CNN and LSTM Models	Dataset: BreakHis [6]; Image Mag: 40×	Benign/Malignant: 94.4, Benign-Malignant eight sub-classes: 91.0
Tasleem *et al.* [30] (2023)	CNN with Wavelet trans- formation	Dataset: BRACS [7]; Image Mag: 40×	Benign-Atypical-Malignant seven sub-classes: 72.2
Proposed method	**ViT**	**Dataset: BRACS [7] and BreakHis[6]; Image Mag: 40×**	**Benign/Atypical/ Malignant: 95.6, Benign sub-classes: 94.7, Atypical sub-classes: 98.6, Malignant sub-classes: 99.1**

B. Robustness and Generalized Performance

Evaluating the robustness and generalized performance of machine learning models is important to estimate its performance under various conditions to ensure that it maintains its accuracy and reliability across different scenarios. Therefore, we evaluated the robustness of the method to ensure two things: 1) the method works efficiently when the histopathology image profile changes and 2) when the BC taxonomy changes. The profile of histopathology images changes when the tissue specimens are prepared in a different lab and when the specimens are scanned by a different WSI scanner. Therefore, in this study, we tested the robustness of the method firstly, by testing the BRACS-trained classifier using BreakHis images. Secondly, we trained the same classifiers using the BreakHis images instead of BRACS and then tested them using BreakHis images. We applied the stage one benign-atypical-malignant classifier on the BreakHis dataset to evaluate it effectiveness on the histopathology image prepared in different lab and scanned using a different WSI scanner. The images were scanned at 40× magnifications; however, they were labeled as either benign or malignant. The benign-atypical-malignant classifier was applied on randomly selected 545 images of BreakHis which included 176 benign and 369 malignant labelled images. The BRACS-trained benign-atypical-malignant classifier correctly identified 147 benign and 257 malignant images, shown in Fig. 7. It wrongly classified 30 malignant images as benign (30/369) and 7 benign images as malignant (7/176) which is not so high. The benign atypical-malignant classifier also identified 22 benign and 82 malignant images as atypical. It was difficult to consider them false negatives as some of these images may include atypical images. BreakHis dataset did not consider atypical class and labeled them either benign or malignant. Therefore, we ignored them in the true positive rate estimation. The true positive rates were 95.5% and 89.5% for detecting the benign and malignant cancers. This indicates the robustness of the transformer-based classifier. For the stage two sub-type classifiers such robustness evaluation using heterogeneous dataset was not performed as the sub-type class labels of BRACS and BreakHis dataset do not matches. Then, we trained the networks selected as the benign-atypical-malignant, benign sub-type and malignant subtype classifiers using the BreakHis dataset to perform benign-malignant, benign sub-type and malignant subtype classification respectively according to the BreakHis BC taxonomy. We selected a subset of 2113 images from the BreakHis dataset randomly for this experiment. The benign-malignant classifier was trained, validated and tested using 688, 230 and 230 image respectively. The benign sub-type classifier was trained, validated and tested using 284, 92 and 92 images respectively. The malignant sub-type classifier was trained, validated and tested using 300, 100 and 100 images respectively. The distribution of image per class was uniform in the training, validation and test datasets for each classifier. For example, 284 training images of malignant sub-type classifier included 71 adenosis, 71 fibroadenoma, phyllodes tumor and 71 tubular adenoma class images. Table V shows the performance of BreakHis trained classifiers' performance. It shows that the classifiers yielded test accuracies higher than 90% indicating the adaptability of transformer models when the

BC taxonomy is changed. The BreakHis-trained benign-malignant classifier had lower test accuracy (92.5%) than the BRACS trained benign-atypical-malignant classifier (95.5%). The benign sub-type classification test accuracy was 94% for both cases. The malignant sub-type classifier achieved only 91% test accuracy using BreakHis; however, it was 99% for BRACS. The accuracy of transformed based classifiers dropped slightly for BreakHis; however, it could have been caused by the small training data. Fig. 8 shows the accuracy and loss curves for training and validation. Although the losses for the classifiers were higher than the BRACS trained classifiers but they indicated no over-fitting.

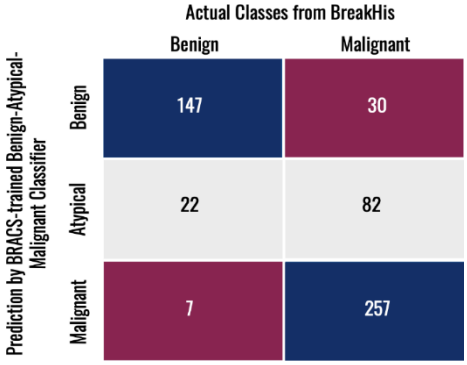

Fig. 7. BRACS-image trained Benign-Atypical-Malignant Classifier tested on BreakHis images.

TABLE V. PERFORMANCE EVALUATION OF THE VIT MODEL FOR BC CLASSIFICATION AND SUBTYPING USING BREAKHIS DATASET

Metrics	Benign-Malignant Classifier	Benign-Subtyping Classifier	Malignant-Subtyping Classifier
Training accuracy	0.951	0.958	0.988
Training loss	0.129	0.1034	0.034
Validation accuracy	0.956	0.973	0.975
Validation loss	0.183	0.161	0.206
Average test accuracy	0.930	0.940	0.910
Average test precision	0.925	0.945	0.907
Average test recall	0.905	0.945	0.907
Average 5-fold cross validation accuracy	0.896	0.901	0.902
Macro average AUC	0.960	0.990	0.990
Micro average AUC	0.960	0.990	0.990

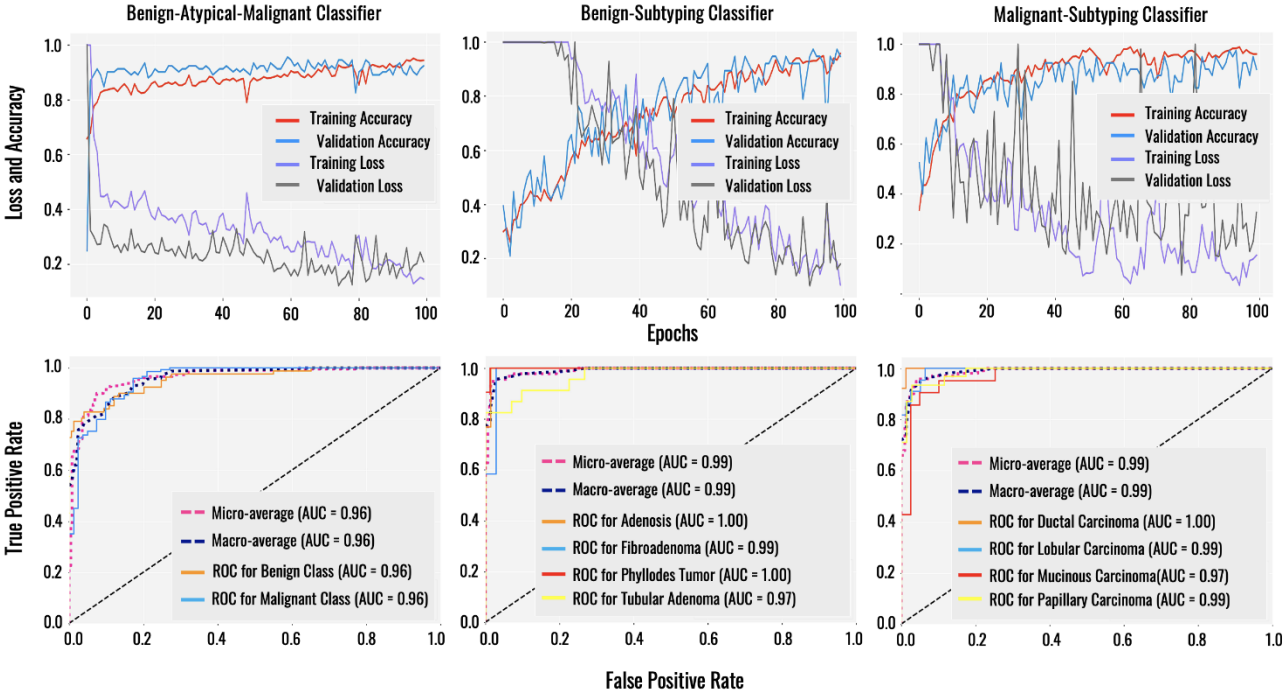

Fig. 8. Training, validation and ROC curves of the ViT model for BreakHis dataset.

After that, 5-fold cross validation experiments were also conducted for the classifiers using the BreakHis dataset. Table VI shows the 5-fold cross validation accuracy of the networks trained using BreakHis dataset. It indicated the 5-fold cross validation accuracies was also lower for the BreakHis trained classifiers. Further, we estimated the AUC values for the classifiers, shown in Fig. 8. Finally, we compared the accuracies of the classifier for both datasets. Fig. 9 shows the comparison of the classifiers. It reveals that the transformer-based networks tend to achieve better accuracy when trained using BRACS dataset. However, the test accuracies, cross validation accuracies and average AUC values indicate that the transformer model yielded comparatively good performances considering the number of images used for trained the models. These tests, expressed the transformers

adaptability when the data domain is drifted and output classes are shifted. It also identified that transformer models can lack in accuracy when trained using a limited dataset, specially when the number of classes are high.

TABLE VI. FIVE-FOLD CROSS VALIDATION ACCURACY FOR BREAKHIS DATASET

Metrics	Benign-Malignant Classifier	Benign-Subtyping Classifier	Malignant-Subtyping Classifier
Fold 1	0.886	0.905	0.872
Fold 2	0.901	0.881	0.905
Fold 3	0.873	0.884	0.939
Fold 4	0.899	0.929	0.910
Fold 5	0.921	0.906	0.883
Average ± standard deviation	0.896 ± 0.016	0.901 ± 0.017	0.902 ± 0.023

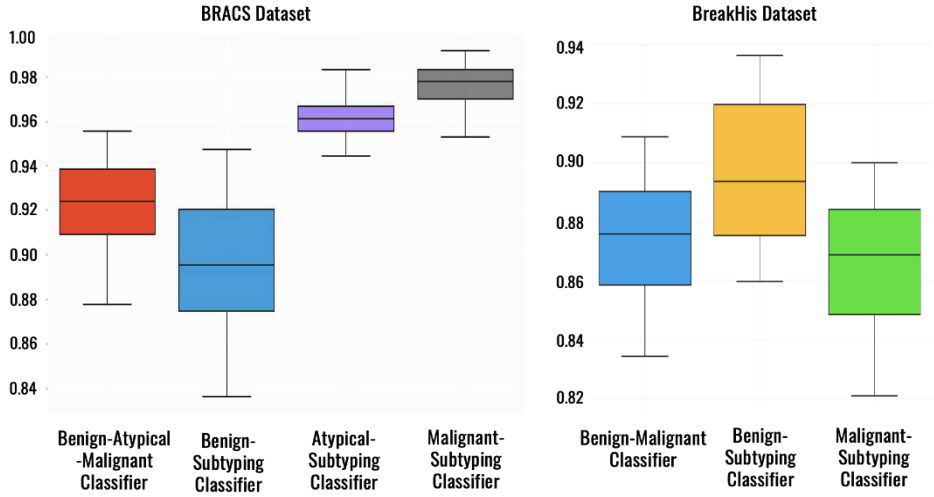

Fig. 9. Boxplot of accuracy of the classifiers for both BRACS and BreakHis datasets.

V. DISCUSSION

In this study, we proposed an automated BC categorization method utilizing transformer based deep learning technique. This method adopted the BC taxonomy provided by the BRACS dataset, as differentiating atypical cancers from the benign and malignant classes improves the BC treatment planning and patient management. We designed a two-stage classification in which four transformer-based classifiers were organized to firstly categorize the BC patients into benign, atypical and malignant classes using the stage one classifier. Then in the second stage another three transformer-based classifiers were used to sub-categorize the benign, atypical or malignant cases, depending on the result of the first stage classifier. This architecture reduced the complexity for the multi-class classification of BC.

The proposed method was primarily trained and tested using the BRACS dataset in which it achieved 95.6% test accuracy for benign-atypical-malignant classification, 94.7% accuracy for benign sub-type classification, 98.6% accuracy for atypical sub-type and 99.1% for malignant sub-type classification, outperforming the previously proposed methods. This ensured the effectiveness of transformer models for BC classification. After that, this method was tested on BreakHis dataset without being trained on BreakHis to assess its robustness in which it achieved a true positive rate of 95.5% for benign class and 89.5% for malignant class. This ensured the robustness of the method. Further, to evaluate how the well this method adapts to concept drift when the BC taxonomy changes, we trained the method following the BreakHis taxonomy and tested it using BreahHis dataset. The proposed method achieved 92.5%, 94.0% and 91.0% accuracy respectively for benign malignant, benign sub-type and malignant sub-type classification. This ensured the generalization capability of the method.

In this research, we have utilized base ViT model with pre-trained weights derived from ImageNet-1k dataset. In the future, it necessary to evaluate the performance of the model for advanced ViT models or ViT models pre-trained on ImageNet-21K dataset. Recently some customized models of ViT were proposed such as Data efficient image transformer (DeiT), Pyramid Vision Transformer (PVT) for medical image classification with limited dataset. Efficacy of such models can be evaluated for BC classification. Another limitation of this study is that we only experimented with 40× magnification images to balance computational efficiency and classification accuracy, in contrast to prior methods that used higher magnifications, such as 100× or more. In future work, we plan to investigate how model performance varies across different magnification levels, ranging from 20× to 200×, to optimize diagnostic accuracy and computational feasibility for real-world deployment.

In addition to its high classification accuracy, the proposed transformer-based model offers practical computational advantages, particularly due to its use of 40× resolution images. Compared to existing methods that

rely on higher magnifications, such as 100× or above, using lower-resolution images reduces data size and computational burden, resulting in faster inference times. For practical implementation, the system will be implemented using standard GPU (e.g., NVIDIA RTX 2080 Ti), enabling real-time or near real-time diagnosis workflows. Furthermore, the architecture's modular nature allows future enhancements such as pruning, quantization, or knowledge distillation to improve deployment performance on resource-constrained systems. These characteristics make the method accurate and feasible for real-world integration into clinical workflows. However, this study does not include a detailed time complexity or computational cost analysis. We plan to comprehensively evaluate the model's inference time and resource requirements in future work to better understand its scalability and feasibility for clinical use. Another potential future work is using nature-inspired algorithms for feature optimization and efficient implementation, as they have shown effectiveness in enhancing deep learning models in healthcare [33].

VI. Conclusion

This study aims to evaluate the performance of image-based transformer models for BC categorization, which remained unexplored in the previous studies. Therefore, in this study, we utilized the base ViT model which contained the minimum number of the parameters and then it by modifying the values of standard hyperparameters to keep the model simple. This is helpful to understand the performance of transformer-based networks for BC categorization. However, in the future, we intend to explore the different customization approaches of ViT to further improve the accuracy of BC categorization. The results of this study showed that ViT, a transformer-based network achieved better accuracy and robustness than the existing deep learning and traditional machine learning models.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Md Shakhawat Hossain: Supervision, Investigation, Formal analysis, Writing—review & editing Conceptualization. Md Ashifur Rahman: Writing—original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. Munim Ahmed: Writing—review & editing, Writing—original draft. Kaniz Fatema: Validation, Software, Methodology, Investigation, Formal analysis, Data curation. Mohammad Anowar Hussen: Investigation, Formal analysis, Data curation. MM Mahbubul Syeed: Project administration, Funding acquisition. Mohammad Faisal Uddin: Supervision, Project administration, Investigation, Funding acquisition. All authors had approved the final version.

Funding

This work was partially funded by the Bangladesh Bureau of Educational Information and Statistics' Grant for Advanced Research in Education (GARE) (Grant Reference Number: ET20222056) and Sponsored Research Grant (2022-SETS-06) from Independent University, Bangladesh.

Acknowledgment

The authors thank Dr. Suresh Advani, Dr. Shingla Shivam and Dr. Sahria Bakar for contributing to preparing the image dataset and evaluating the results of our experiments.

References

[1] World Health Organization. (Mar. 2019). Breast cancer. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] American Cancer Society. Key statistics for breast cancer. [Online]. Available: https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html

[3] World Health Organization. Global impact in reducing breast cancer mortality. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[4] A. Woeckel *et al.*, "The screening, diagnosis, treatment, and follow-up of breast cancer," *Deutsches Ärzteblatt International*, vol. 115, no. 18, pp. 316–323, 2018.

[5] A. Giuliano, S. Edge, and G. Hortobagyi, "Breast cancer staging manual: AJCC cancer staging manual," *Ann. Surg. Oncol.*, vol. 25, no. 7, pp. 1783–1785, 2018.

[6] F. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2015.

[7] N. Brancati *et al.*, "BRACS: A dataset for breast carcinoma subtyping in H&E histology images," *Database*, vol. 2022, p. baac093, 2022.

[8] A. Chan and J. Tuszynski, "Automatic prediction of tumour malignancy in breast cancer with fractal dimension," *R. Soc. Open Sci.*, vol. 3, no. 8, 160558, 2016.

[9] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PLoS ONE*, vol. 14, no. 4, e0214587, 2019.

[10] D. Bardou, K. Zhang, and S. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.

[11] A. Nahid and Y. Kong, "Histopathological breast-image classification using local and frequency domains by convolutional neural network," *Information*, vol. 9, no. 1, p. 19, 2018.

[12] H. Luo, J. Li, L. Cai, and M. Wu, "STrans-YOLOX: Fusing swin transformer and YOLOX for automatic pavement crack detection," *Appl. Sci.*, vol. 13, no. 3, p. 1999, 2023.

[13] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," arXiv Prepr., arXiv:2010.11929, 2020.

[14] M. Hossain *et al.*, "Region of Interest (ROI) selection using vision transformer for automatic analysis using whole slide images," *Sci. Rep.*, vol. 13, no. 1, 11314, 2023.

[15] T. Menes *et al.*, "Subsequent breast cancer risk following diagnosis of atypical ductal hyperplasia on needle biopsy," *JAMA Oncol.*, vol. 3, no. 1, pp. 36–41, 2017.

[16] N. Houssami, L. Irwig, and O. Ung, "Review of complex breast cysts: Implications for cancer detection and clinical practice," *ANZ J. Surg.*, vol. 75, no. 12, pp. 1080–1085, 2005.

[17] S. Singh and R. Kumar, "Histopathological image analysis for breast cancer detection using cubic SVM," in *Proc. 2020 7th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, 2020, pp. 498–503.

[18] A. Belsare, M. Mushrif, M. Pangarkar, and N. Meshram, "Classification of breast cancer histopathology images using texture

feature analysis," in *Proc. TENCON 2015-2015 IEEE Region 10 Conf.*, 2015, pp. 1–5.
[19] M. Aswathy and M. Jagannath, "An SVM approach towards breast cancer classification from H&E-stained histopathology images based on integrated features," *Med. Biol. Eng. Comput.*, vol. 59, no. 9, pp. 1773–1783, 2021.
[20] F. Spanhol, L. Oliveira, P. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *Proc. 2017 IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2017, pp. 1868–1873.
[21] F. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Proc. 2016 Int. Jt. Conf. Neural Netw. (IJCNN)*, 2016, pp. 2560–2567.
[22] Z. Han *et al.*, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.*, vol. 7, no. 1, p. 4172, 2017.
[23] M. Motlagh *et al.*, "Breast cancer histopathological image classification: A deep learning approach," *bioRxiv*, p. 242818, 2018.
[24] A. Golatkar, D. Anand, and A. Sethi, "Classification of breast cancer histology using deep learning," in *Image Analysis and Recognition*, vol. 10882, Springer, 2018, pp. 837–844.
[25] G. Aresta *et al.*, "BACH: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, 2019.
[26] M. Gour, S. Jain, and T. S. Kumar, "Residual learning based CNN for breast cancer histopathological image classification," *Int. J. Imaging Syst. Technol.*, vol. 30, no. 3, pp. 621–635, 2020.
[27] E. Zewdie, A. Tessema, and G. Simegn, "Classification of breast cancer types, sub-types and grade from histopathological images using deep learning technique," *Health Technol.*, vol. 11, no. 6, pp. 1277–1290, 2021.
[28] F. Ahmed *et al.*, "Improved breast cancer diagnosis through transfer learning on hematoxylin and eosin stained histology images," arXiv Prepr., arXiv:2309.08745, 2023.
[29] H. Chu *et al.*, "RetMIL: Retentive multiple instance learning for histopathological whole slide image classification," arXiv Prepr., arXiv:2403.10858, 2024.
[30] E. Deniz *et al.*, "Transfer learning based histopathologic image classification for breast cancer detection," *Health Inf. Sci. Syst.*, vol. 6, no. 1, pp. 1–7, 2018.
[31] T. Kausar, Y. Lu, and A. Kausar, "Breast cancer diagnosis using lightweight deep convolution neural network model," *IEEE Access*, vol. 11, pp. 124869–124886, 2023.
[32] M. Hossain *et al.*, "Practical image quality evaluation for whole slide imaging scanner," in *Biomedical Imaging and Sensing Conf.*, vol. 10711, 2018, pp. 203–206.
[33] Z. Amiri *et al.*, "The applications of nature-inspired algorithms in Internet of Things-based healthcare service: A systematic literature review," *Trans. Emerg. Telecommun. Technol.*, vol. 35, no. 1, p. e4969, 2024.