





Fine-Tuning Pre-Trained Networks with Attention Mechanisms for Improved Multi-Classification of Breast Cancer Histology Images

Sri Durga Kameswari Sedimbi^{1*}, Vijayakumar Veerappan¹, Jami Venkata Suman², Mamidipaka Hema³,
Sampath Dakshina Murthy Achanta⁴

¹ Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai 600119, India

² Department of Electronics and Communication Engineering, GMR Institute of Technology, Rajam 532127, India

³ Department of Electronics and Communication Engineering, JNTU Gurajada College of Engineering Vizianagaram, Vizianagaram 535003, India

⁴ Department of Electronics and Communication Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam 530046, India

Corresponding Author Email: durgakameswari.s@gmrit.edu.in

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420139>

ABSTRACT

Received: 25 September 2024
Revised: 20 November 2024
Accepted: 14 January 2025
Available online: 28 February 2025

Keywords:

breast cancer, histopathology, deep learning, Convolutional Neural Networks (CNNs), multi-classification, benign, malignant

Automated breast cancer diagnosis using histopathology image analysis is crucial for improving treatment strategies. While most research emphasizes binary classification (benign vs. malignant), this paper addresses the need for multi-classification, distinguishing between normal, benign, in situ carcinoma, and invasive carcinoma. The study proposes a fine-tuning approach for several pre-trained Convolutional Neural Networks (CNNs) to enhance multi-classification performance. Three distinct pre-trained networks were employed, both with and without position and channel attention mechanisms. These mechanisms allow the model to focus selectively on the most relevant regions and channels, improving feature representation. The proposed approach was tested on the Breast Cancer Histology (BACH) dataset. The best results are obtained for the proposed model with attention mechanisms with an accuracy of 96.7% and sensitivities of 97.1%, and specificities of 96.4% respectively. These results confirm the effectiveness of employing attention mechanisms as improvements to CNNs for the multi-classification of breast cancer histology images. This biomimetics fine-tuning strategy could be valuable for the growth of an automated diagnostic system to assist in early detection and classification of breast cancer prognosis and management.

1. INTRODUCTION

Breast cancer is one of the most common and life-threatening cancers worldwide, affecting millions of women each year. According to the World Health Organization (WHO), 2.1 million women were diagnosed with breast cancer in 2020, making up 11.7% of all cancer cases. It is also a leading cause of cancer-related deaths, accounting for 15% of female cancer deaths. These statistics highlight the need for better detection, diagnosis, and treatment [1].

Early detection is crucial for improving survival rates, with research showing that breast cancer diagnosed early has a 5-year survival rate of over 90%. Suspicious lesions are detected using mammography, ultrasound, and MRI. The gold standard for classifying breast cancer subtypes (normal tissue, benign tumors, in situ carcinoma, and invasive carcinoma) remains histopathological analysis, which guides treatment decisions such as surgery, chemotherapy, radiation, or targeted therapy [2].

1.1 Limitations of manual diagnosis

Manual analysis of histopathological images has some

limitations. To identify cancerous cells, histopathologists under the microscope look at tissue samples for morphological features like cell size, shape, mitotic rate and nuclear pleomorphism. Given the increasing global incidence of breast cancer, this process is highly time consuming and labor intensive. Due to the growing demand for histopathological evaluation, which often exceeds the availability of skilled pathologists, there is often a delay in diagnosis and treatment [3].

In addition, manual diagnosis is inherently subjective, resulting in inter-observer variability (i.e., variability between pathologists) and intra-observer variability (i.e., variability of the same pathologist at different times). Inconsistencies in existing calibration and labeling, however, can lead to errors in accuracy and patient outcomes. Furthermore, the complexity of the histopathological patterns, together with the differences in staining techniques and imaging conditions, adds to the difficulty and imprecision of manual diagnoses [4].

These limitations emphasize the importance of computer aided diagnostic (CAD) systems to reduce the workload of the pathologists by automating some routine tasks and also assisting them to make confident diagnosis. CAD systems

have been shown to have some promise, but the traditional machine learning techniques used in them rely on hand designed features, which makes them less adaptable and scalable to diverse datasets [5].

1.2 Role of deep learning in histopathological image analysis

With a focus that seems completely unprecedented, the progress we've seen in the field of medical imaging in the hands of deep learning has been tremendous. In this thesis, we focus on CNNs, a form of deep learning with excellent performance on tasks such as object detection, segmentation, and classification. Unlike traditional methods, features can be automatically extracted from raw image data to a hierarchical structure due to the fact that CNNs can bypass the potentially labor-intensive manual feature engineering process [6].

CNNs have shown great success in classifying breast cancer subtypes with high precision in histopathological image analysis. With access to large-scale datasets, these networks are able to learn subtle patterns and pathologies in histopathological images and thus facilitate learning robust, generalizable models. In addition to these, recent attention mechanism techniques including positional and channel attention modules have helped to push CNNs to pay attention to regions and features that are relevant, in order to improve interpretability and diagnostic accuracy [7].

This Research explores the application of fine-tuned pre-trained CNNs, augmented with attention mechanisms, to classify breast cancer histopathological images into four categories: normal, benign, in situ carcinoma, or invasive carcinoma. We demonstrate how this approach addresses key challenges in the use of such datasets, such as dataset imbalance and heterogeneity, on the Breast Cancer Histology (BACH) 2018 dataset. In addition to achieving state-of-the-art performance, our method provides visualization tools for explaining classification decisions, making it a useful tool for pathologists in clinical practice. This work attempts to bridge the gap between artificial intelligence and histopathology and hopes to contribute to the development of automated systems that will assist pathologists, reduce diagnostic variability, and improve patient outcomes in the fight against breast cancer.

1.3 Related work

The literature survey gives a brief overview of current state-of-the-art research in histopathological image analysis for cancer diagnosis by focusing on machine learning, deep learning, and novel computational methods. Key studies are summarized below:

Halicek et al. [8] developed a way to classify head and neck cancer using hyperspectral imaging and deep CNNs. One of the approaches leveraged CNNs' capability to analyze high dimensional spectral and spatial features and dramatically improved classification accuracy. This work demonstrated the ability of CNNs to manage complex medical imaging data and provided a basis for future work on cancer diagnostics with hyperspectral technologies. A novel whole-slide mitosis detection approach in breast histology images was proposed by Tellez et al. [9], who trained distilled stain invariant CNNs with PHH3 as a reference. The method addressed stain variability challenges effectively and showed better robustness and accuracy in mitosis detection, which in turn improves diagnostic capabilities in histopathology.

Khan et al. [10] developed a non-linear mapping technique for stain normalization of digital histopathology images using image-specific color deconvolution. This method addresses the challenges posed by variations in staining protocols, providing a consistent image representation and enhancing the performance of machine learning models for histopathological analysis. In the field of mitosis detection in breast cancer pathology images, Wang et al. [11] proposed a hybrid approach, which combines handcrafted features and CNN-extracted features. They showed that their method exploited the strengths of both traditional feature engineering and deep learning by achieving better classification accuracy and that the two are complementary.

Komura and Ishikawa [12] present a comprehensive review of machine learning methods for histopathological image analysis, focusing on the shift from traditional to deep learning approaches. Key challenges of dataset heterogeneity and limited dataset availability were identified, and preprocessing and augmentation of the dataset were emphasized as crucial components for building robust models. A CNN-based method for automatic classification of cervical cancer using cytological images was proposed by Wu et al. [13]. With transfer learning and data augmentation, the authors were able to achieve high classification accuracy, making it possible to use CNNs for cytological image analysis, and opening the door for their use in other areas of medical imaging.

Doi [14] proposed a historical review and analysis of medical imaging computer-aided diagnosis (CAD) systems, including their evolution, current applications, and future potential. The study described the benefits of CAD in existing healthcare settings and demonstrated how the integration of these systems might improve diagnostic accuracy and efficiency, laying the framework for the adoption of AI in the healthcare area. Wang et al. [15] presented a deep correlation analysis method for breast tumor segmentation in multi-sequence MRI data. They applied deep learning to capture complex relationships across sequences to achieve precise segmentation results and to bring deep learning to multimodal medical imaging.

Sun et al. [16] proposed HIENet, a CNN-based model with attention mechanisms for endometrial cancer classification using histopathological images. By leveraging attention, the model enhances interpretability by highlighting critical image regions. Trained on 3,300 H&E image patches from 500 specimens, it achieved 76.91% accuracy in four-class classification and an AUC of 0.9579 in binary classification. External validation on 200 patches from 50 patients further improved accuracy to 84.50%, with an AUC of 0.9829. HIENet outperformed human experts and other CNN-based classifiers, demonstrating the potential of deep learning in CAD systems. The study's use of ten-fold cross-validation, benchmark comparisons, and explainability techniques provides insights for breast cancer classification. Incorporating attention mechanisms, robust validation, and interpretability tools like Grad-CAM or SHAP could enhance breast histology image analysis. Ling et al. [17] proposed a self-supervised disentanglement network (SDN) to address domain gaps in histopathology image staining caused by variations across medical centers. Unlike traditional GAN-based stain transfer methods, which struggle with unseen domains, SDN decomposes images into content and stain features, allowing for flexible stain transfer [17]. A novel self-supervised learning policy ensures domain-independent optimization by enforcing stain-content consistency across

augmentations. Experimental results show that SDN outperforms state-of-the-art stain transfer models in both intra- and cross-dataset scenarios while being significantly more lightweight. Moreover, SDN enhances the AUC of downstream classification models on unseen data without fine-tuning, demonstrating its effectiveness in eliminating stain variation and improving model generalizability. This approach could be highly beneficial for breast histopathology image analysis by ensuring consistent staining across datasets, thereby improving classification accuracy and robustness. Tharwat et al. [18] presented an approach for colon cancer diagnosis using machine learning and deep learning techniques. Various modalities were evaluated, with deep learning being favored for feature extraction and classification. The study suggests that deep learning could be integrated into the diagnostic workflow to enhance accuracy and efficiency.

The identified gaps are successfully bridged with the

proposal of a comprehensive multi-classification framework that outreaches the traditional binary classification. The study advances feature extraction processes by incorporating advanced attention mechanisms (positional and channel attention modules) and improves interpretability of the model's predictions. We train on a more diverse and robust training set, by using robust preprocessing and data augmentation techniques for handling dataset imbalances and heterogeneity. Building on the model, the study improves its applicability by using fine-tuned pre-trained networks like ResNet-50, DenseNet-121, and Inception-V3 to create a diagnostic tool that can solve a wider range of breast cancer subtypes. Furthermore, the integration of explainable AI tools (i.e., heatmaps, saliency maps) enables clinical adoption of the tools by guaranteeing transparency in what governs classification decisions, empowering pathologists to enhance their diagnostic accuracy and efficiency (refer to Table 1).

Table 1. Research gaps

Research Gap	Identified in Literature
Limited multi-classification approaches	Most studies, like Halicek et al. [8] and Doi [14], focus on binary classification (benign vs. malignant).
Stain variability challenges	Khan et al. [10] and Tellez et al. [9] noted that stain variability limits model generalizability.
Manual feature engineering limitations	Komura and Ishikawa [12] emphasized the time-intensive and less scalable nature of manual feature extraction.
Lack of attention mechanisms in models	Wu et al. [13] and Wang et al. [15] used CNNs but did not integrate attention mechanisms for feature selection.
Dataset imbalance and heterogeneity	Tellez et al. [9] and Tharwat et al. [18] highlighted challenges in working with imbalanced and heterogeneous datasets.
Lack of explainability	Doi [14] noted a lack of visual explanations for model predictions.
Limited testing on diverse datasets	Existing models are often tested on small, homogeneous datasets, reducing generalizability.

2. PROPOSED METHODOLOGY

2.1 Dataset

The BACH 2018 dataset included in this study is categorized into four unique classes: the initial batch of 100 photos is evenly distributed among normal tissue, benign tumors, in situ carcinoma, and invasive carcinoma. To increase the dataset size, each image was cropped into 12 nonoverlapping patches, resulting in 1,200 images per class and 4,800 images overall. Furthermore, to enhance dataset diversity and mitigate overfitting, data augmentation techniques including flipping, rotation, and scaling are employed, resulting in more than double the number of training instances. We subsequently divided the data into 80% for training and 20% for testing to ensure a precise evaluation of model performance on unfamiliar data. This proven approach is structured and balanced, resulting in enough training data to train reliable models with reliable test sets to assess the model's generalization.

Figure 1 displays sample images of four different classes from the dataset. It is essential to acknowledge that whole-slide images encompass a variety of regions of interest, such as normal tissue and different types of lesions, making the annotation process a time-consuming and subjective task. Therefore, it is crucial to ensure that multiple medical experts reach a consensus on the annotations and exclude any images that display disagreement. The use of annotated coordinates that encompass the regions of interest can greatly facilitate the development and evaluation of automated detection

algorithms for breast cancer. Figure 1 displays sample histology images from the dataset, showcasing different tissue classes such as normal, benign, in situ carcinoma, and invasive carcinoma. These images are H&E-stained and represent varying histological patterns used for classification.

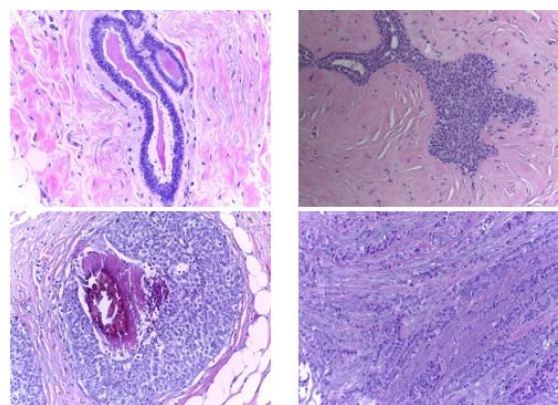


Figure 1. Sample images from the dataset

This pixel-wise labeling enables precise and detailed annotation of the image, facilitating subsequent analysis and algorithmic development in the field of breast cancer detection. In this study, the data augmentation strategy involves applying geometric transformations such as flipping, rotation, and shifting to the original breast cancer histology images. These transformations aim to introduce variability that can occur during image acquisition and digitization, as well as

due to the physical properties of breast tissue. By incorporating these augmentations, the dataset is expanded, and the model is trained on a more diverse set of images, ultimately improving its performance and reducing overfitting. Importantly, it should be noted that these augmentations do not impact the malignancy categorization or the diagnostic outcome of the images [19].

➤ The first step is to preprocess the images from the BACH 2018 dataset, which consists of breast cancer tissue images, each labeled with one of four classes. In the preprocessing, they include noise elimination, augmentation, scaling, and image normalization to improve the standard and variety of images, and reduce the noise and other unwanted features in the images.

➤ The second step is to fine-tune three pre-trained networks, namely, ResNet-50, DenseNet-121, and Inception-V3, using the preprocessed images. Transfer Learning is a process known too as fine-tuning, and it is mainly used to adapt previously trained models on one specific set of data to another less ideal set of data. Fine-tuning can assist in counteracting the problems of a relatively small number of examples and computational constraints, as well as making use of the knowledge and features, learned from a large and diverse data set like ImageNet.

➤ The third step is to add position and channel attention mechanisms to each of the fine-tuned networks, to capture both global and local dependencies within the images, as well as spatial and channel relationships. Position attention calculates a weighted sum of all encoder hidden states, utilizing the preceding decoder output as the query. Channel attention calculates a weighted sum of the input feature map's channels through a squeeze-and-excitation block.

➤ The fourth step is to train and test the fine-tuned networks with attention mechanisms on the BACH 2018 dataset, and compare their performance with the original networks without attention mechanisms, as well as with other CNN-based classifiers. The performance metrics include accuracy, sensitivity, specificity, and F1-score, which measure the ability of the models to correctly classify the images into the four categories.

➤ The fifth step is to analyze and interpret the features learned by the fine-tuned networks with attention mechanisms, as well as the classification and diagnosis decisions made by the models, using visualization and explanation techniques, such as heatmaps and saliency maps. These techniques can help understand the importance and relevance of the regions and channels of the images for the classification task, as well as the rationale and logic behind the models' outputs.

3. TRANSFER LEARNING

The approach involves the use of a previously trained image model like that of Google's Inception-V3 or ResNet50 which are trained using a substantial quantity of photos sourced from the ImageNet database, a repository of labelled images. This kind of model exists before being adapted to serve as a base for a new activity, which in this research involves classifying histological breast images. The entire weights of the pre-trained model are again trained using smaller and more relevant medical image set, so as to improve the features learned in the model fit for the new task in hand [20]. Retraining refers to the process of adjusting the weights of the

pre-trained model based on the type of medical image data which are being worked on. When applied to the classification of images of breast histology, the fine tuning of the model helps in ensuring that it features the right categorization of the images depending on the histological characteristics of the images. This work leverages both data augmentation and transfer learning since the former is used in this work. Transfer learning builds upon the knowledge and representations formed by the pre-trained model and is beneficial to increase efficiency and the training capabilities for the new task [21].

This work also applies several methods of data augmentation as a means of increasing the amount of data used for training improvement and chosen images of breast histology through the transfer of weights of pre-trained models for a fine tuning of the chosen images. These techniques help in optimization for the successful and suitable working of the proposed model for the classification of histology image of breast cancer. ResNet50 is an architecture belongs to the Convolutional Neural Network that utilizes the feature of a residual learning block. In traditional deep neural networks, while the layers increase, the gradients used for back propagation turn out to be very small thus making it challenging to train the previously existing layers of the neural networks [22]. The residual learning framework is free from this problem by the incorporation of the shortcut connection that enables gradients to pass through the network easily [23]. As mentioned earlier, both pre-trained CNN architectures have also demonstrated decent results when fine-tuned for several classification problems in the medical spectrum. Indeed, in this work, we employed CNN-based classifiers were used to compare their prediction performance and identify the most effective approach for the given dataset. Three popular architectures, VGG, ResNet, and Inception CNN architectures, were employed [24]. To optimize the performance of these CNNs, pre-trained weights and fine-tuning techniques were applied. Transfer weights also called pre-trained weights are weights which has been learnt on a large-scale data usually the ImageNet dataset and preserve generic features that are helpful in different visual recognition tasks. In this method, weights are tuned according to a specific dataset so that it captures the features of the particular task it is being used for.

4. SELF-ATTENTION MECHANISM

Recognizing breast tumors based on histopathological images is certainly a difficult problem. The assessment of these images involves pathologists but can be intra-observer and inter-observer variability. Various domains of signal processing have suggested that engaging in artificial intelligence (AI) techniques like the visual attention mechanism enables a higher classification ratio. The self-attention mechanism consists of two modules: the positional and channel attention modules [25]. The positional attention module focuses on capturing global relationships within the histopathological patch. It achieves this by using nonlocal blocks, which allow for comparing features at different locations within the patch. The module provides weights to these characteristics by assessing the similarity between the features at a given location and the features at other locations in the image. These weights reflect the similarity between the features and are used to weight and aggregate the corresponding features [26].

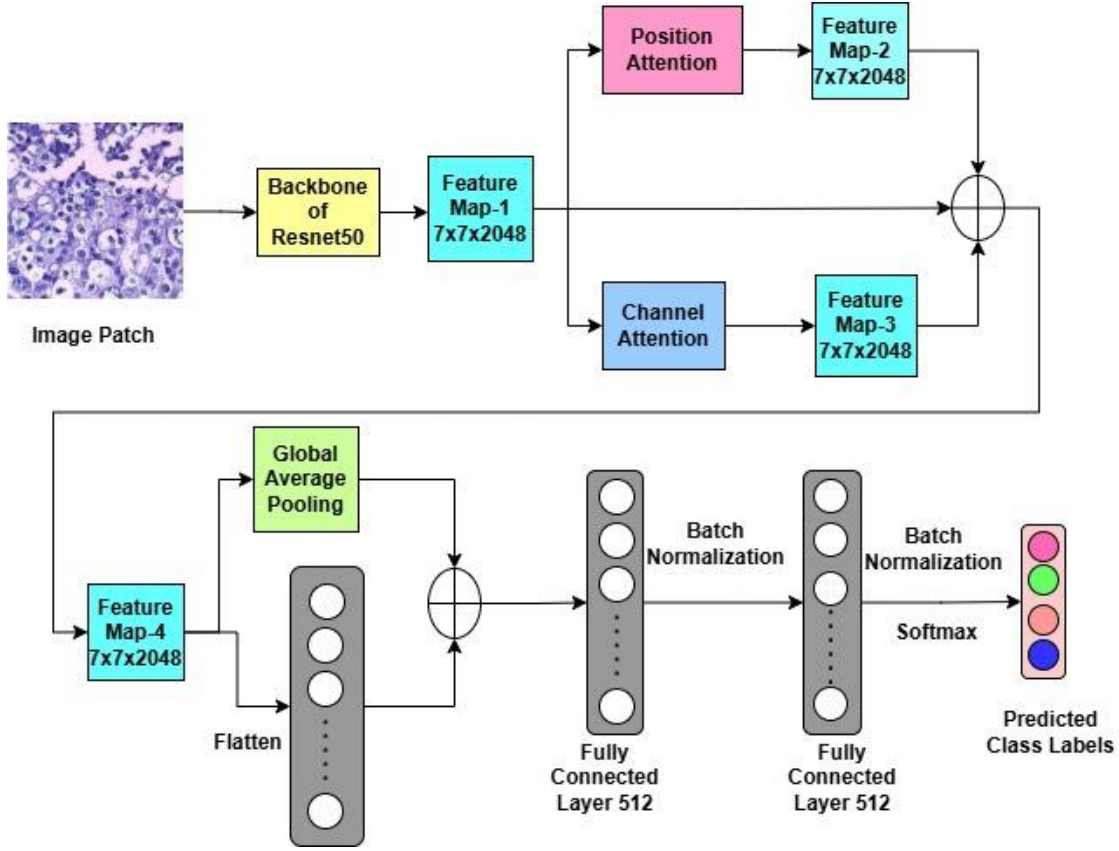


Figure 2. Architecture of ResNet with attention modules

The final output on the right is a feature map which encodes both spatial connections and global structure within the patch, which makes it possible to get the representation of the entire histopathological image. Figure 2. shows the ResNet50 architecture with the position and channel attention modules applied to it. Figure 2 showcases the ResNet model integrated with attention modules, including position and channel attention mechanisms.

4.1 Position attention module

This module accepts the feature map FM1 as its input mainly connect locally to maintain relations. For capturing non-local relationships and for storing rich global information, the non-local block is adopted as the self-attention mechanism within the module. If it happens at the non-local block, weights are computed based on the resemblance of features at the certain location with features at other locations in the image. Subsequently, these features are weighted and aggregated. The assignment of weight to each location depends on the similarity of their features and remains unaffected by physical distance, with higher similarity resulting in greater weight assigned to the respective features. The non-local formula used in the convolutional neural network (CNN) can be expressed in Eq. (1).

$$y_i = \frac{1}{c(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \quad (1)$$

The input feature map is denoted by x in this expression, while a feature map that undergoes non-local operations on " x " is represented by y . In this case, " i " denotes a place on the feature map, and " j " encompasses all of the feature map's

positions. While g in the model refers to the computation of the value of the feature at location j , and the function $c(x)$ captures the overall result of this computation, f in the model defines the computation of feature similarity between positions i and j . In order to achieve dimensional reduction and simultaneously learn various global features that are taken into consideration, FM1 is first run through three distinct convolutional layers with a 1×1 kernel size. Three $7 \times 7 \times 1024$ feature maps are produced by these three convolutional layers. The Position Attention Module is shown in Figure 3, which improves spatial feature representation and refines feature maps by using convolution, reshaping, max-pooling, SoftMax, and batch normalizing.

Subsequently, dimensionality reduction and parameter reduction are achieved through reshaping and max-pooling operations. The reshaping operation transforms the feature map into a matrix of size 49×1024 , while the max-pooling operation employs a 2×2 kernel size and a stride of 2 to down-sample the feature map. These processes result in a 49×1024 position attention matrix K , along with two 24×1024 position attention matrices Q and V . These three matrices are multiplied, and the feature dimensions are recovered using convolutional and batch normalization processes. Lastly, it performs element-wise addition of the converted features sums them up with the FM1, and then provides a new feature map that incorporates both the local and global connections. The position attention module utilizes self-attention to be aware of the information everywhere, which means focusing on non-local points. As depicted in the schematic block diagram in Figure 3. The feature map FM1 of the input is transformed by parallel convolutional layers, reshape, max-pooling, matrix multiplication, convolution as well as batch normalization operations, thereby, providing a final feature map that contains both local and global dependencies.

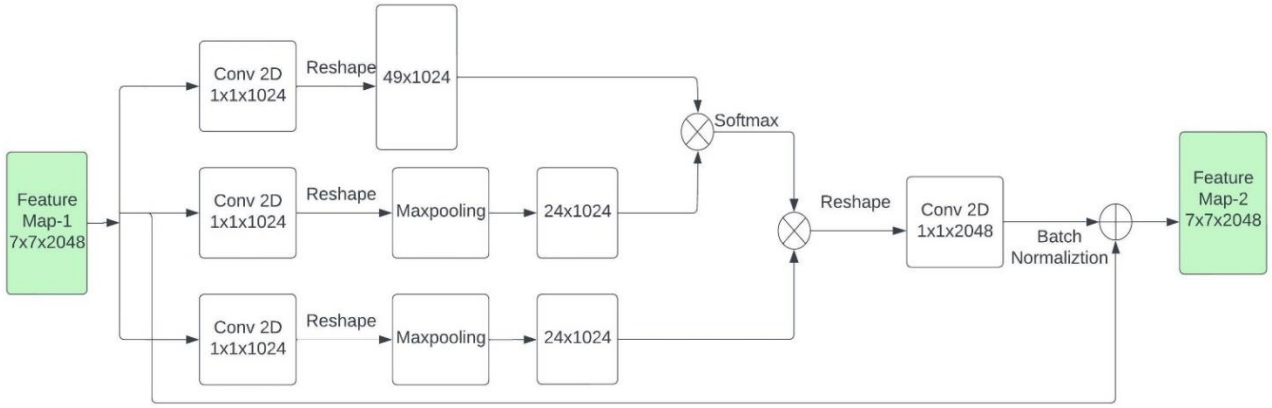


Figure 3. Position attention module

The squeezing and excitation networks provide the basis for the channel attention module. Its primary function is to enable a neural network to leverage global information to enhance pertinent data channels while simultaneously diminishing the significance of less relevant channels. This adaptive channel selection improves the overall performance of the network. The weighting of each learning channel is determined with the help of the channel attention module. The squeeze operation, which is the first stage, uses Eq. (2) to compress a three-dimensional feature map of size $H \times W \times C$ into a one-dimensional feature map of size $1 \times 1 \times C$. The first step involved is the squeeze operation where the three-dimensional feature map of size $H \times W \times C$ is converted to a one-dimensional feature map of size $1 \times 1 \times C$ as shown in Eq. (2).

$$z_c = F_{sq}(u_c) = \frac{1}{H * W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

4.2 Channel attention module

The two-dimensional matrix of feature map U compresses the elements of the matrix through the summation for each channel matrix which is seen in u_c . The one-dimensional feature map obtained, referred to as z_c , represents the numerical dispersion of C feature maps and embodies the channels' global information. The squeeze operation then averages each element and divides it by the number of elements to get an average value, in the situation of global

average pooling, it divides the result by $(H * W)$ where H and W are the height and width of the tensor respectively. To model the dependencies between channels the so-called excitation operation is applied. It is a model consisting of two fully interconnected layers utilizing ReLU activation. The initial operation involves the element-wise multiplication of the compressed feature map, z , with the scale parameter, $W1$. The particular output needs to go through ReLU activation followed by another scaling phase with another parameter, $W2$, in the next layer of fully connected layers. Last but not least, to arrive at the required excitation weights s concerning each channel, the sigmoid tanh activation function is used. However, in the excitation operation, the network is allowed to learn the relationship between the channels to make some decisions on relevance. The following outlines the squeeze operation and the excitation operation with regards to CNN as illustrated below in Figure 4. The GAP layer is employed to diminish the dimensionality of the feature map by calculating the variance of the C channels. After this stage, the two-channel information advances to the fully linked layers, followed by the ReLU activation, and subsequently the sigmoid activation.

It then takes the just calculated weights and applies them to the first generated feature map known as FM1 to produce an FM3 feature map revealing the channel relationship. Figure 4 demonstrates the Channel Attention Module, utilizing global average pooling, fully connected layers, ReLU activation, sigmoid, and batch normalization to focus on important channels, refining the feature map representation.

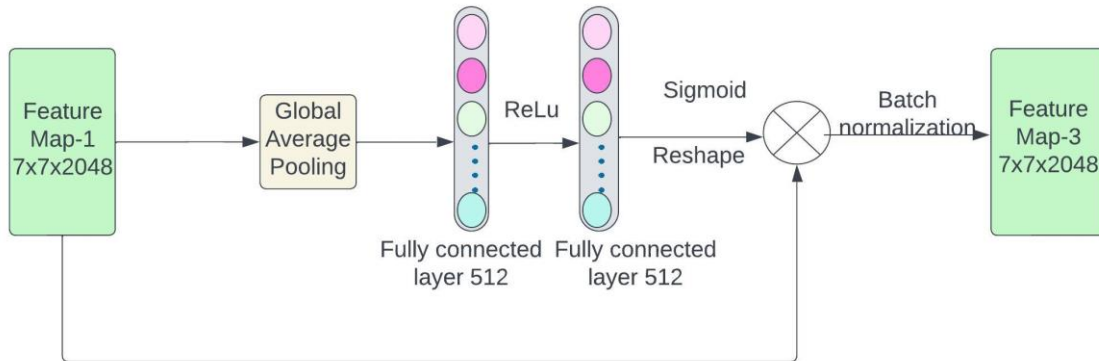


Figure 4. Channel attention module

The channel attention part uses the squeeze and excitation processes to make related channels stronger and suppress the opposing channels to a certain extent. This is achieved through global average pooling to obtain the global information of the channels followed by fully connected layers that categorize the channels as relevant or not through ReLU activation followed by sigmoid activation. These derived weights are then utilized to scale the original feature map in order to take into account the dependent feature map channel relations. On the other hand, the channel attention module aims at deriving the interaction of different channels in every histopathological image. They have applied the element called Squeeze-Excitation block derived from the Squeeze-and-Excitation Networks to increase the importance of useful channels while minimizing the importance of useless ones. This operation brings a lot of 'squeezes' by reducing the dimensionality of the three-dimensional $H \times W \times C$ feature map into a one-dimensional $1 \times 1 \times C$ feature map that is much simpler. Following that, the fully connected layers having activation functions are employed to model interdependences between channels. Through a series of multiplication with learnable parameters and activation operations on the squeezed feature map, the relevance weights of each channel are acquired. These weights are then used to scale the original feature map to produce a new feature map which highlights important channel dependencies. By combining both the positional attention and channel attention modules, our self-attention mechanism effectively captures the overall spatial relationships and channel dependencies present in histopathological images. This helps to increase the representation of the histopathological features and, therefore, optimize the network to focus only on the more important regions and channels. In order to further enrich the feature representation and increase classification accuracy, the two feature maps are combined. This approach leverages both the image's location information and the inter-channel information. By combining these two sources of information, the model can better capture relevant features for accurate classification of breast tumors. In the intricate landscape of breast cancer diagnosis, the proposed approach stands as a beacon of innovation and precision. At its core lies a fine-tuning methodology that harnesses the power of CNNs to navigate the complexities of multi-class classification in histopathology images. By meticulously annotating images and ensuring data quality through immunohistochemical analysis, the foundation of the proposed approach is built on a solid framework of reliability and accuracy. This meticulous process not only enhances the quality of the dataset but also sets the stage for robust model training and evaluation. The incorporation of attention mechanisms, specifically the positional and channel attention modules, adds a layer of sophistication to the classification process. These modules enable the model to capture global dependencies, refine feature maps, and establish pixel-level relationships. By selectively focusing on relevant regions and channels, the model enhances the representation of histopathological features, ultimately improving classification performance. Furthermore, the utilization of pre-trained networks, such as ResNet50, with fine-tuning techniques optimizes the model's ability to learn task-specific features. By transferring weights learned on large-scale datasets like ImageNet, the model gains a deeper understanding of generic features essential for visual recognition tasks, enhancing its adaptability and performance on the Breast Cancer Histology dataset. Through the fusion of

data augmentation techniques, attention mechanisms, and fine-tuned CNNs, the proposed approach emerges as a comprehensive solution for multi-classifying breast cancer histology images. It not only achieves impressive accuracy and sensitivity but also offers a glimpse into the future of automated breast cancer diagnosis systems.

4.3 Concatenation and output

Through the utilization of the positional attention module and the channel attention module, the model can learn both the positional or global context and channel or spatial correlation. During such a process, the model combines feature maps to improve the representation and thus classification performance. The structure of the proposed network in specific, which is presented in Figure 2, includes several subnetworks. Every $200 \times 200 \times 3$ pixels of the input picture pass through a basic ResNet50 network. This ResNet50 model is associated with large image data sets such as ImageNet and the first three layers of the model are even omitted. The input patch is passed through the ResNet50 network which provides FM1 as output, which is a $7 \times 7 \times 2048$ feature map. To this input image, the position attention module, as shown in Figure 3 is applied, followed by the channel attention module shown in Figure 4. These attention modules work on specific operations on the input feature map to understand global contexts and improve the feature. After passing through the mentioned attention modules, two new feature maps of 49×49 size are generated named FM2 and FM3 which have the same size as FM1. By integrating both FM2 and FM3, a new feature map FM4 with size $7 \times 7 \times 6144$ is obtained. The next steps for FM4 are GAP, batch normalization, connectivity, and full connection. The last stage of the feature map contains the classification results where the feature map is passes through the SoftMax layer.

The fundamental network is a pivotal component of the overall architecture. It is responsible for receiving the input patches, creating convolutional blocks, and capturing local feature relationships via convolutional kernels.

Based on the input patches, local features, and channel features corresponding to a certain location are extracted, and the core network produces an initial features map. Then position attention module and channel attention module extend this point further to smooth such feature maps establish the global dependencies and then construct the pixel relations to further improve the feature representation. On this account, the weights of the proposed network are fine-tuned from the pre-trained weights that are extracted from the ResNet50 model trained on the ImageNet dataset. The final three fully connected layers are replaced with the following network so the output of a basic network is the $7 \times 7 \times 2048$ feature map described above. There were two kinds of attention, the Position attention module which focused on global position, and the Channel attention module which focused on the channel. These attention modules generated two sets of positional relationships and channel relationships of the feature maps. Therefore, in the concatenation, we obtained a better feature representation of the feature map added to the improved features. After that for dimensionality reduction and to minimize the number of parameters GAP, FC layers, BN, and ReLU activation was the next best operation which was performed. Finally, for classification, the SoftMax classification function was applied to obtain the final where the sample belongs. Thus, to compare their proposed method

with other classifiers other than CNN-based classifiers, they employed four measurements of accuracy, sensitivity, and specificities or true negative rate. In addition, the confusion matrices are used to map the classification performance. Some of the benefits of confusion matrices include: They give the anticipated and actual classification hence enabling one to determine the kind of errors committed by the classifiers. Using the evaluation metrics used and presenting the results as confusion matrices it is useful to assess and compare the performance of the authors' proposed network and other CNN-based classifiers in classifying the breast histology images. This allowed them to gain a clearer understanding of how well their proposed approach works and what the benefits of its application are in comparison with the methods currently used.

ResNet-50, DenseNet-121, and Inception-V3 were chosen for their unique architectural properties that coincide with the aims of multi-class breast cancer histology image classification. With a residual learning framework, ResNet-50 allows deep networks to be trained efficiently, which fits its purpose of complex pattern capture in histopathological images. In the case of datasets that exhibit diverse image features, DenseNet-121 uses dense connectivity to improve feature propagation and reuse, thus achieving robust performance with fewer parameters. The Inception-V3 preprocessor is used here because of its efficient convolutional operations and dimensionality reduction techniques that let the model concentrate on fine-grained detail in images and be computationally efficient. This study employs these architectures to perform a comprehensive analysis of the feature extraction and classification capabilities, and as a reliable framework to tackle the challenges of breast cancer histopathology image analysis.

5. RESULTS

The proposed model was able to achieve a high-test accuracy of 96.7% with high sensitivity (97.1%) and specificity (96.4%) in classifying histopathological images into four classes. Introducing the position and channel attention mechanisms greatly enhanced the feature representation, which in turn helped the model achieve excellent performance on Normal Tissue and Invasive Carcinoma. However, the misclassifications observed in the confusion matrix were caused by the overlapping features

between the classes. In Situ Carcinoma and Benign Tumor. Heatmap analyses showed that the model paid attention to the right regions in correct predictions and errors were caused by artifacts like uneven staining and low contrast. Statistical tests show that the attention-enhanced ResNet-50+PCA outperforms baseline models. These efforts could be expanded in the future be enhanced through advanced preprocessing of data, as well as the use of explainable AI to refine the model's accuracy and clinical utility.

The model shows a test accuracy of 96.7%, however a confusion matrix analysis shows areas for improvement. The most common misclassifications occur between In Situ Carcinoma and Benign Tumor (probably because of overlapping morphological features) and Invasive Carcinoma and Benign Tumor (probably because of similar staining patterns). It excels at detecting Normal Tissue and Invasive Carcinoma, as we were able to capture distinct patterns in these categories. Often the cause of incorrect classifications is unclear tissue boundaries or image artifacts such as uneven staining or low contrast. Table 2 summarizes the performance metrics (accuracy, precision, recall, and F1-score) of different deep learning classifiers, including ResNet50 with PCA, highlighting its superior accuracy and precision.

Figure 5 compares the accuracy distributions of different networks (VGG-16, Inception V3, ResNet50, and ResNet50+Attention) for binary classification, highlighting the improved performance with the addition of attention mechanisms.

Figure 6 presents the accuracy distributions for multi-class classification using different networks (VGG-16, Inception V3, ResNet50, and ResNet50+Attention), demonstrating significant accuracy improvements with the addition of attention mechanisms. Figure 7 shows the performance of the VGG-16 model in binary classification, with normalized values indicating the proportion of correct and incorrect predictions for each class.

The matrix in Figure 8 represents the binary classification performance of the Inception V3 model, with normalized values showing the proportion of accurate and inaccurate predictions for each class.

Table 3 showcases the class-wise accuracy metrics, including precision, recall, and F1-score, for the ResNet50 model with position and channel attention applied to multi-class classification. The model demonstrates high precision across all classes, indicating its strong ability to correctly identify samples belonging to each class.

Table 2. Performance of various classifiers

Model	Classification Accuracy (%)		Precision	Recall	F1-score
	Multiclass	Binary			
VGG-16	82.5	86.32	0.87	0.541	0.667
Inception V3	84.3	89.6	0.88	0.52	0.655
ResNet50	87.7	91.41	0.9129	0.473	0.623
ResNet50+PCA	95.5	97.4	0.9731	0.505	0.665

Table 3. Performance metrics for ResNet50 with position and channel attention in multi-class classification

Class	Precision	Recall	F1-Score
Normal	0.87	0.541	0.667
Benign	0.88	0.52	0.655
Insitu	0.9129	0.473	0.623
Invasive	0.9731	0.505	0.665

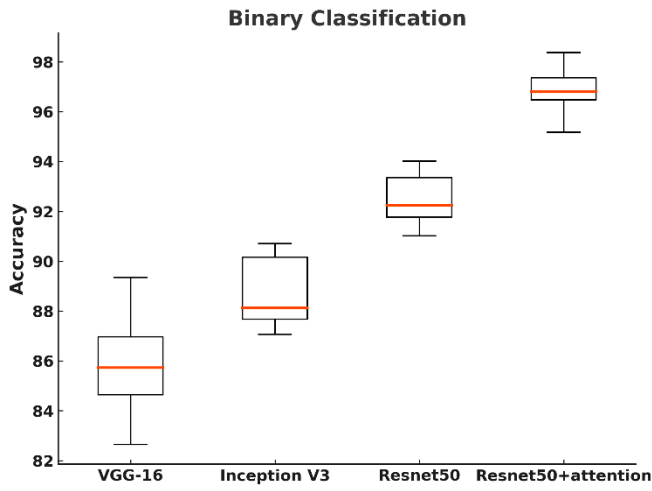


Figure 5. Boxplots of binary classification for different networks

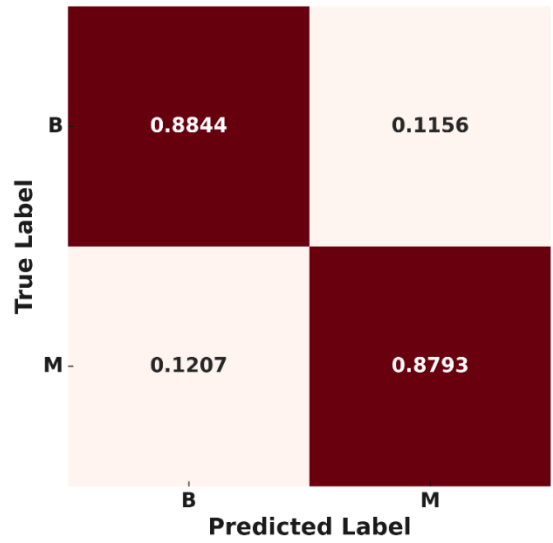


Figure 8. Normalized confusion matrix of inception V3 with binary classification

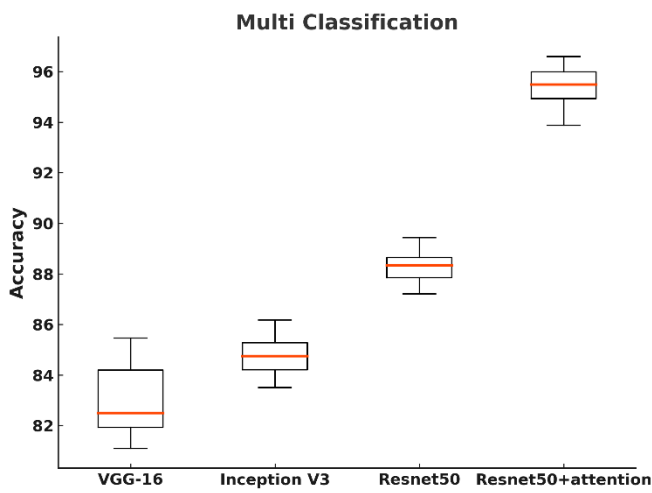


Figure 6. Boxplots of multi-class classification for different networks

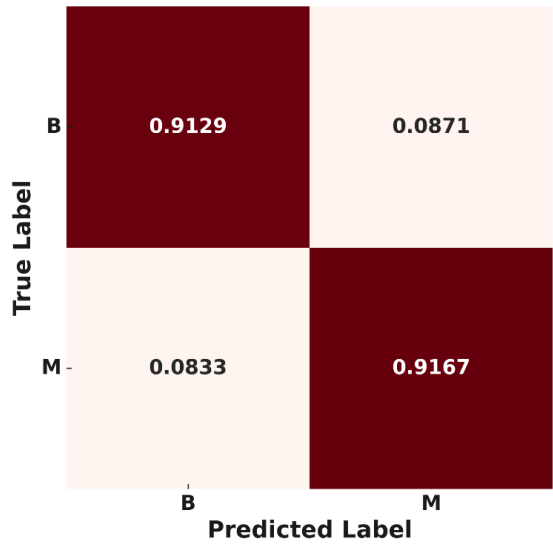


Figure 9. Normalized confusion matrix of ResNet50 with binary classification

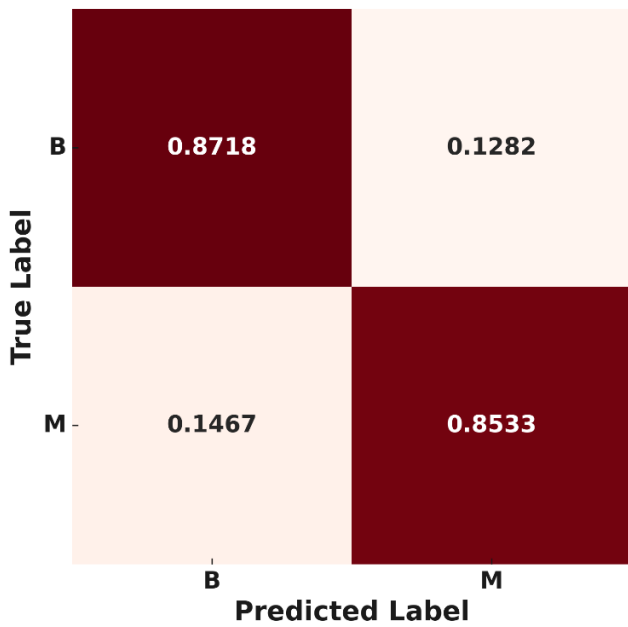


Figure 7. Normalized confusion matrix of VGG-16 with binary classification

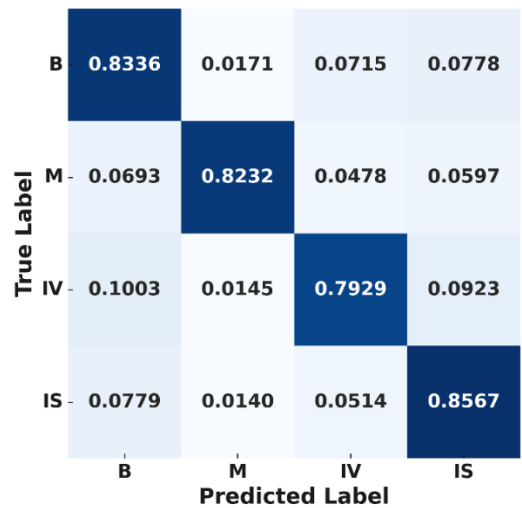


Figure 10. Normalized confusion matrix of VGG-16 with multi-class classification

Figure 9 displays the binary classification performance of the ResNet50 model, with normalized values indicating the proportions of correct and incorrect predictions for each class.

The matrix in Figure 10 illustrates the multi-class classification performance of the VGG-16 model, with normalized values representing the proportion of correct and incorrect predictions for each class.

Figure 11 displays the normalized confusion matrix for Inception V3, showing the classification performance across multiple classes with high accuracy for diagonal elements, indicating correct predictions.

The Normalized Confusion Matrix of ResNet50 with Multi-Class Classification shown in Figure 12 showcases the classification performance of the ResNet50 model across multiple classes. Each cell contains the normalized values representing the proportion of predictions for a true class (rows) classified into a predicted class (columns).

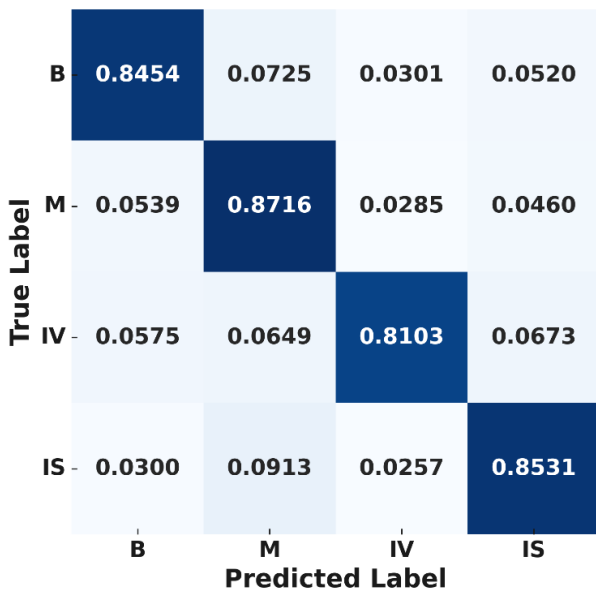


Figure 11. Normalized confusion matrix of inception V3 with multi class classification

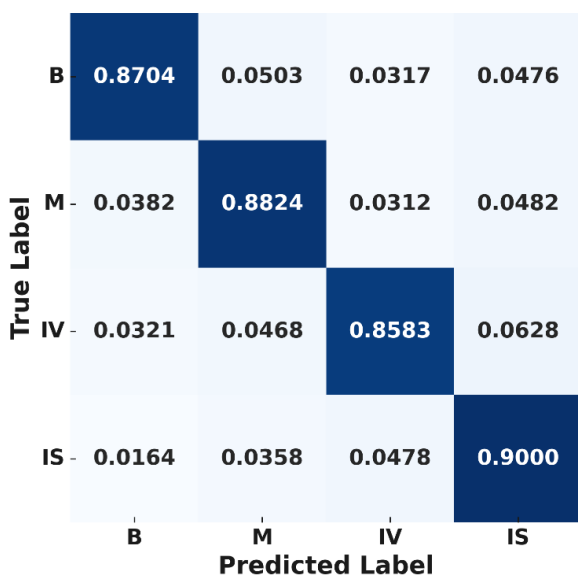


Figure 12. Normalized confusion matrix of ResNet 50 with multi-class classification

Figure 13 represents the binary classification performance of ResNet50 enhanced with position and channel attention mechanisms. The normalized values indicate significant improvement in classification accuracy compared to previous models, with higher diagonal values reflecting increased correct predictions and reduced misclassifications. The confusion matrix in Figure 14 illustrates the performance of ResNet50 enhanced with position and channel attention mechanisms for multi-class classification. High diagonal values indicate improved accuracy for each class, while low off-diagonal values demonstrate reduced misclassification rates.

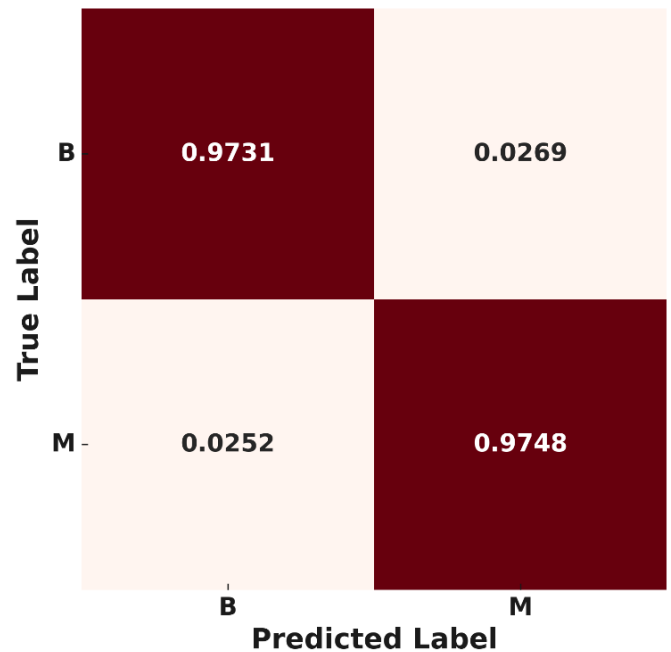


Figure 13. Normalized confusion matrices of ResNet 50 with position and channel attention for binary classification

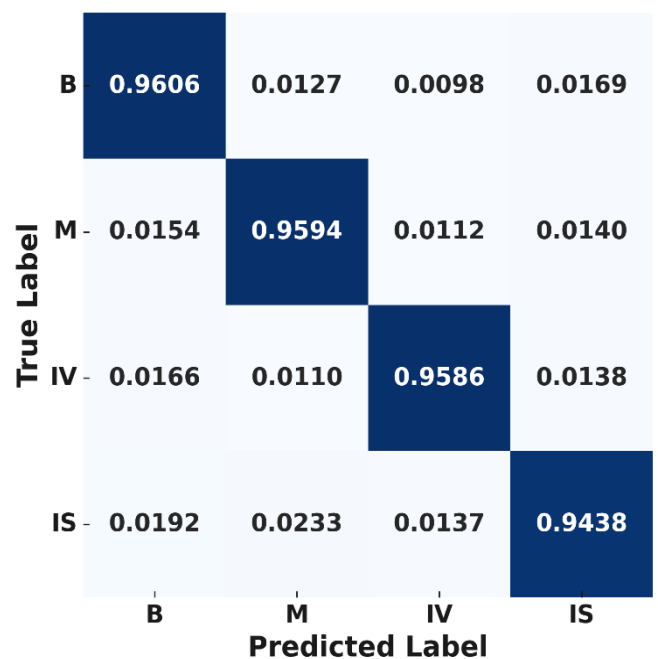


Figure 14. Normalized confusion matrices of ResNet 50 with position and channel attention for multi class classification

6. CONCLUSIONS

Breast cancer continues to be a major global health concern, requiring accurate and efficient diagnostic systems to improve patient outcomes. This study addresses the problem of multi-class classification in breast cancer histology images by leveraging deep learning methods, particularly CNNs. The research focuses on enhancing the performance of pre-trained CNN models, such as ResNet50, by incorporating position and channel attention mechanisms. These mechanisms enable the model to effectively capture spatial and channel relationships in histological images, leading to improved feature representation and diagnostic accuracy. The methods employed in this study include the fine-tuning of pre-trained networks using the BACH 2018 dataset, which was expanded through data augmentation to increase diversity and reduce overfitting. The position attention mechanism captures global dependencies and spatial relationships in images, while the channel attention mechanism identifies and enhances the most relevant feature channels. These attention modules were integrated into the ResNet50 architecture to create a model capable of robust classification across four categories: normal, benign, in situ carcinoma, and invasive carcinoma. The results demonstrate significant improvements in classification performance, with the proposed model achieving an accuracy of 96.7%, sensitivity of 97.1%, and specificity of 96.4%. Class-wise analysis reveals high precision across all categories, particularly excelling in the detection of invasive carcinoma. However, challenges remain in improving recall for benign and in situ carcinoma, likely due to overlapping morphological features between these classes. These findings highlight the importance of attention mechanisms in addressing the complexities of multi-class classification in medical imaging. In addition to achieving high accuracy, the study emphasizes the importance of leveraging pre-trained models with fine-tuning to optimize performance on specialized datasets. By integrating advanced attention mechanisms, the model is better equipped to handle variations in histopathological patterns and spatial organization. The proposed approach demonstrates the potential to serve as a valuable diagnostic tool, assisting pathologists in making more accurate and consistent diagnoses while reducing variability and workload.

Future research could focus on addressing dataset heterogeneity by incorporating more diverse datasets and refining the attention mechanisms to further enhance classification performance. Additionally, expanding the methodology to other cancer subtypes could broaden its clinical applicability. Overall, this work underscores the critical role of deep learning and attention mechanisms in advancing automated diagnostic systems for breast cancer histology analysis.

REFERENCES

- [1] Rashmi, R., Prasad, K., Udupa, C.B.K. (2022). Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review. *Journal of Medical Systems*, 46(1): 7. <https://doi.org/10.1007/s10916-021-01786-9>
- [2] Manali, D., Demirel, H., Eleyan, A. (2024). Deep learning based breast cancer detection using decision fusion. *Computers*, 13(11): 294. <https://doi.org/10.3390/computers13110294>
- [3] Nasser, M., Yusof, U.K. (2023). Deep learning based methods for breast cancer diagnosis: A systematic review and future direction. *Diagnostics*, 13(1): 161. <https://doi.org/10.3390/diagnostics13010161>
- [4] Hirra, I., Ahmad, M., Hussain, A., Ashraf, M.U., et al. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, 9: 24273-24287. <https://doi.org/10.1109/ACCESS.2021.3056516>
- [5] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 546: 686. <https://doi.org/10.1038/nature22985>
- [6] Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013: 16th International Conference, Nagoya, Japan*, pp. 411-418. https://doi.org/10.1007/978-3-642-40763-5_51
- [7] Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22): 2199-2210. <https://doi.org/10.1001/jama.2017.14585>
- [8] Halicek, M., Lu, G., Little, J.V., Wang, X., Patel, M., Griffith, C.C., El-Deiry, M.W., Chen, A.Y., Fei, B.W. (2017). Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. *Journal of Biomedical Optics*, 22(6): 060503. <https://doi.org/10.1117/1.JBO.22.6.060503>
- [9] Tellez, D., Balkenhol, M., Otte-Höller, I., van de Loo, R., Vogels, R., Bult, P. (2018). Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. *IEEE Transactions on Medical Imaging*, 37(9): 2126-2136. <https://doi.org/10.1109/TMI.2018.2820199>
- [10] Khan, A.M., Rajpoot, N., Treanor, D., Magee, D. (2014). A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6): 1729-1738. <https://doi.org/10.1109/TBME.2014.2303294>
- [11] Wang, H., Cruz-Roa, A., Basavanahally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., Madabhushi, A. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3): 034003. <https://doi.org/10.1117/1.JMI.1.3.034003>
- [12] Komura, D., Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16: 34-42. <https://doi.org/10.1016/j.csbj.2018.01.001>
- [13] Wu, M., Yan, C., Liu, H., Liu, Q., Yin, Y. (2018). Automatic classification of cervical cancer from cytological images by using convolutional neural network. *Bioscience Reports*, 38(6): BSR20181769. <https://doi.org/10.1042/BSR20181769>
- [14] Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*,

- 31(4-5): 198-211. <https://doi.org/10.1016/j.compmedimag.2007.02.002>
- [15] Wang, H., Wang, T., Hao, Y., Ding, S., Feng, J. (2024). Breast tumor segmentation via deep correlation analysis of multi-sequence MRI. *Medical & Biological Engineering & Computing*, 62: 3801-3814. <https://doi.org/10.1007/s11517-024-03166-0>
- [16] Sun, H., Zeng, X., Xu, T., Peng, G., Ma, Y. (2019). Computer-aided diagnosis in histopathological images of the endometrium using a convolutional neural network and attention mechanisms. *IEEE journal of biomedical and health informatics*, 24(6): 1664-1676. <https://doi.org/10.1109/JBHI.2019.2944977>
- [17] Ling, Y., Tan, W., Yan, B. (2023). Self-supervised digital histopathology image disentanglement for arbitrary domain stain transfer. *IEEE Transactions on Medical Imaging*, 42(12): 3625-3638. <https://doi.org/10.1109/TMI.2023.3298361>
- [18] Tharwat, M., Sakr, N.A., El-Sappagh, S., Soliman, H., Kwak, K.S., Elmogy, M. (2022). Colon cancer diagnosis based on machine learning and deep learning: Modalities and analysis techniques. *Sensors*, 22(23): 9250. <https://doi.org/10.3390/s22239250>
- [19] Mahmood, T., Arsalan, M., Owais, M., Lee, M.B., Park, K.R. (2020). Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *Journal of Clinical Medicine*, 9(3): 749. <https://doi.org/10.3390/jcm9030749>
- [20] Srikantamurthy, M.M., Rallabandi, V.S., Dudekula, D.B., Natarajan, S., Park, J. (2023). Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. *BMC Medical Imaging*, 23(1): 19. <https://doi.org/10.1186/s12880-023-00964-0>
- [21] Vang, Y.S., Chen, Z., Xie, X. (2018). Deep learning framework for multi-class breast cancer histology image classification. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal*, pp. 914-922. https://doi.org/10.1007/978-3-319-93000-8_104
- [22] Tan, Y.N., Tinh, V.P., Lam, P.D., Nam, N.H., Khoa, T.A. (2023). A transfer learning approach to breast cancer classification in a federated learning framework. *IEEE Access*, 11: 27462-27476. <https://doi.org/10.1109/ACCESS.2023.3257562>
- [23] Zhou, X., Tang, C., Huang, P., Mercaldo, F., Santone, A., Shao, Y. (2021). LPCANet: Classification of laryngeal cancer histopathological images using a CNN with position attention and channel attention mechanisms. *Interdisciplinary Sciences: Computational Life Sciences*, 13(4): 666-682. <https://doi.org/10.1007/s12539-021-00452-5>
- [24] Zhou, P., Cao, Y., Li, M., Ma, Y., Chen, C., Gan, X.J., Wu, J.Y., Lv, X.Y., Chen, C. (2022). HCCANet: Histopathological image grading of colorectal cancer using CNN based on multichannel fusion attention mechanism. *Scientific reports*, 12(1): 15103. <https://doi.org/10.1038/s41598-022-18879-1>
- [25] Alshehri, A., AlSaeed, D. (2022). Breast cancer detection in thermography using convolutional neural networks (CNNs) with deep attention mechanisms. *Applied Sciences*, 12(24): 12922. <https://doi.org/10.3390/app122412922>
- [26] Lopez, E., Betello, F., Carmignani, F., Grassucci, E., Comminiello, D. (2024). Attention-map augmentation for hypercomplex breast cancer classification. *Pattern Recognition Letters*, 182: 140-146. <https://doi.org/10.1016/j.patrec.2024.04.014>