Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 10, No. 4, November, Pp. 487-498

Classification of breast cancer histopathology images using an attention-based multiple instance learning method

Safira Hasna Setiyani*[1], Edi Noersasongko[1], Affandy[1]
Department of Magister Teknik Informatika, Universitas Dian Nuswantoro Semarang, Indonesia[1]

Article Info
Keywords:
Breast Cancer, Classification, Image, Machine Learning

Article history:
Received: May 06, 2025
Accepted: July 25, 2025
Published: November 01, 2025

Cite:
S. H. Setiyani, E. Noersasongko, and Affandy, "Classification of Breast Cancer Histopathology Images Using an Attention-Based Multiple Instance Learning Method", *KINETIK*, vol. 10, no. 4, Nov. 2025.
https://doi.org/10.22219/kinetik.v10i4.2310

*Corresponding author.
Safira Hasna Setiyani
E-mail address:
shafira1995@gmail.com

Abstract
Breast cancer is one of the deadliest types of cancer among women worldwide. Early detection plays a crucial role in increasing the chances of successful treatment and reducing the risk of death. Various efforts have been made by both the general public and medical professionals to raise awareness, promote early screening, and ensure timely medical intervention. With advances in technology, the use of computer-based systems, particularly in the field of medical image analysis, has become increasingly important. One such application is histopathological image analysis to support the diagnostic process in breast cancer cases. Histopathological image classification has gained significant attention from researchers in recent years, and various machine learning and deep learning techniques have been applied to improve its accuracy. Convolutional Neural Networks (CNNs), as part of the deep learning framework, have shown promising results in identifying tissue patterns in histopathological images. However, despite their high accuracy, CNNs are often less interpretable, making it difficult to understand the reasoning behind their predictions—especially when dealing with subtle features such as small spots, dots, or fine lines that may be overlooked. This study addresses these limitations by proposing a method that not only classifies histopathological images with high accuracy but also enhances readability through localization techniques. The goal is to make the classification process more transparent and clinically useful. Using widely recognized datasets like BreakHIS, the proposed method achieves a classification accuracy of up to 97.50%, demonstrating its potential as a reliable tool in medical diagnostics and breast cancer research.

1. Introduction

Cancer is a complex disease characterized by abnormal and uncontrolled cell growth, in which these cells have the ability to invade surrounding tissues and spread to other organs through a process known as metastasis [1]. Cancer cells can develop in almost any part of the human body, from the skin surface to internal organs such as the lungs, liver, or brain [2]. According to WHO data, cancer is the second leading cause of death worldwide, after cardiovascular diseases such as stroke and heart attack, with the number of cases continuing to rise each year. The causes of cancer are diverse and complex, involving a combination of genetic factors, lifestyle choices (such as diet, alcohol consumption, smoking, and lack of physical activity), and exposure to environmental carcinogens such as radiation or air pollution.

Although cancer is often equated with tumors, it is important to understand the fundamental difference between the two. A tumor is an abnormal mass of tissue that can be benign (non-spreading) or malignant (spreading and damaging other tissues), while cancer always refers to malignant tumors [3]. Among the various types of cancer, breast cancer is one of the most common in women and remains the leading cause of death among women globally [4]. Statistics show a significant increase in the number of new cases each year, particularly in developing countries where access to early screening remains limited.

Breast cancer diagnosis is traditionally performed by pathologists through histopathological slide examination using a microscope [5]. This process is manual, subjective, and time-consuming [6]. Pathologists must observe important features such as nuclear morphology, cell distribution and size, and tissue structure at the microscopic level [7]. The accuracy of diagnosis depends heavily on the skill and experience of the physician, which opens the possibility of variation among individuals in assessing the severity of the disease. Therefore, early detection is the key to improving patient survival rates and the effectiveness of therapy.

Early detection efforts are generally conducted through screening mammography [8], a non-invasive imaging technique that can identify tumors before they progress to advanced stages. However, this technique is not without challenges, such as limited image resolution, interpretation errors due to overlapping breast tissue, and reduced

sensitivity in women with dense breast tissue. In some cases, mammography produces false positives or false negatives, which may lead to misdiagnosis and inappropriate therapy.

In response to these limitations, alternative approaches such as analyzing hematological parameters from patient blood samples have begun to be developed [9]. Some studies suggest that changes in certain blood parameters can serve as early indicators of cancer, including breast cancer. This technique is more accessible, non-invasive, and has the potential to provide rapid results. On the other hand, histopathology-based imaging approaches remain crucial as they provide detailed information about cellular and tissue structure, which is essential for cancer classification and staging.

Histopathology itself is a branch of anatomical pathology that focuses on the microscopic analysis of body tissues to detect morphological changes in response to disease [9]. Histopathological examination not only identifies the presence of cancer cells but also provides important insights into the biological characteristics of tumors, such as differentiation level, mitotic rate, and tissue invasion patterns [10]. All this information is vital in establishing prognosis and determining the most appropriate therapeutic strategy.

Recent developments in breast cancer diagnosis include the integration of artificial intelligence (AI), deep learning, and computer-aided diagnosis (CAD) in histopathological image analysis. This approach aims to enhance the accuracy, consistency, and efficiency of diagnosis while reducing the workload of pathologists. Machine learning algorithms can identify subtle patterns in digital images that are difficult for human observers to detect and can perform automatic classification based on cellular morphological features. Additionally, research is also moving toward personalized treatment based on tumor molecular profiles and patient responses, known as precision medicine.

Experiments have been conducted based on machine learning techniques ranging from the lowest to the highest level, with quite varied accuracy results [11], and many success rates are close to excellent, with only some noise that needs to be improved. Therefore, the development of this machine learning method continues in an effort to achieve the most effective approach with optimal results. Deep machine learning algorithms are capable of producing performance equivalent to that of an expert in object detection and image classification tasks [12].

The application of Machine Learning in the health sector is highly impactful [13] because it can enhance clinical decision support, detect diseases, and help personalize treatment approaches for patients in need. The limited number of skilled doctors is a factor that affects the optimization of the healthcare organization's services. The most widely used machine learning technique to describe objects in complex images is Convolutional Neural Network (CNN). ResNet18 and VGG16, which are part of the Convolutional Neural Network architecture, are capable of generating large image network data sets.

The application of Convolutional Neural Networks (CNNs) in breast cancer detection has advanced significantly. A study by Alanazi et al. (2021) demonstrated that a five-layer CNN model could achieve an accuracy of 87%, outperforming traditional machine learning approaches [14]. In 2022, Pradeep Kumar et al. applied a three-layer CNN to the DDSM dataset and achieved a high accuracy of 97.2%, with a precision rate of 99% [15]. More recent work by Murtaza et al. (2024) utilized various architectures—including InceptionV3, DenseNet121, and VGG16—combined with data augmentation and U-Net-based segmentation, reaching an accuracy of up to 98.87% on datasets like CBIS-DDSM and Inbreast [16]. Another study in 2024 introduced a modified ResNet architecture for patch-based analysis of Whole Slide Images (WSI), reporting an accuracy of 89% [17]. A systematic review by Yusoff et al. (2023) in *Diagnostics* highlighted CNNs as the most dominant and effective deep learning approach across nearly 100 studies on breast cancer diagnosis. Overall, these recent studies underscore the effectiveness of CNN-based models—particularly modified and hybrid versions—in improving the accuracy, efficiency, and real-world applicability of breast cancer detection.

Multiple Instance Learning (MIL) is a learning framework that is often used in supervised learning [15]. Unlike supervised learning, the data sample for MIL is a *bag* consisting of a set of instances. In the MIL setting [16], the bag is labeled, while the individual instances within the bag remain unlabeled. Researchers have studied various paradigms for MIL, including—but not limited to—multiple instance classification [18], multiple instance regression, multiple instance clustering, unbalanced multi-instance learning, and multiple multi-instance labeling.

Attention-based Multiple Instance Learning (A-MIL) is a machine learning method used when training labels are available only for groups of data (bags) rather than for individual instances within those groups. In A-MIL, each bag consists of several instances, and the label assigned to the bag applies to the entire bag, not to each individual instance [20].

In medical images, especially histopathology images, there is great variation in tissue structure as well as pattern and texture. In addition, the cancer lesions are very small and the patterns are scattered, while for part of the image is normal tissue. Therefore, multiple instance learning techniques are very effective in classifying a set of instances [21], which allows the system to ignore irrelevant structures (such as image areas with normal tissue) and focus on detecting relevant areas (cancer cells). Multiple instance learning can also detect very small areas, allowing cancerous lesions to be identified even when the most of the image appears to show normal tissue. Especially in cancer datasets, the labeling

process at the instance level often takes a long time due to the large amount of data used. Multiple instance learning allows to use datasets with more general labels (in bags), making the process faster and more efficient.

In this research, the attention-based multiple instance learning method [18] is used to produce more complicated localization of regions in breast cancer histopathology images, particularly in handling noise—an area where other classification methods often struggle, such as spots in images [22]. This is achieved by giving a bag [23] of several examples of image spots using the attention-based multiple instance learning (A-MIL) method to classify them [23]. The research provides new innovations to breast cancer classification by using different testing methods and datasets from several studies that have been done before. Increasing the amount of training data will help evaluate the capability of multiple instance learning in improving classification accuracy, while the addition of texture-based feature extraction aims to reduce noise from previous studies.

Therefore, by utilizing bag-level labels and capturing complex patterns within the bag [19], MIL models can support early detection, accurate classification, risk prediction, and personalized treatment planning. In addition, incorporating domain knowledge and expert annotations into the MIL framework can further improve the performance and interpretability of cancer diagnostic models.

In contrast to conventional CNNs or fully supervised approaches, A-MIL can capture complex patterns across whole-slide images, making it suitable for real-world clinical data. This research introduces innovations by evaluating A-MIL with various datasets and expanding the amount of training data. It also incorporates texture-based feature extraction to reduce the impact of noise, thereby improving classification accuracy and model robustness.

A major contribution of this work is the development of an interpretable and reliable breast cancer classification system that not only predicts overall labels but also provides visual localization maps highlighting diagnostically important areas. The use of public datasets such as BreakHIS ensures that the findings are reproducible and comparable with existing methods.

Beyond its technical merits, this research offers significant social value. The AI-based diagnostic system can assist medical practitioners, particularly in regions with limited access to pathologists. It enables more efficient and accurate early detection of breast cancer, potentially lowering mortality rates, accelerating the diagnostic process, and reducing the workload of healthcare professionals in underserved areas.

2. Research Method

2.1 Dataset

The dataset used in this research consists of breast cancer histopathology images obtained from the publicly available BreakHis dataset (Breast Cancer Histopathological Image Classification). BreakHis is a widely recognized dataset in the field of medical image analysis and has been used extensively in various studies related to the classification of breast cancer. The dataset comprises a total of 7,909 histopathology images, which are categorized into two main classes, namely Benign (non-cancerous) and Malignant (cancerous). These images were collected from breast tumor biopsies of 82 patients and represent a broad range of tissue types, providing diverse and rich data for training and evaluating machine learning models.

The dataset was divided into three subsets: training, validation, and testing data. Specifically, 80% of the total data (6,327 images) were allocated for training the model to learn and recognize the relevant features of each class. Meanwhile, 10% of the data (791 images) were used for validation during training to fine-tune model parameters and prevent overfitting. The remaining 10% (791 images) were used for final testing to evaluate the generalization performance of the trained model.

One of the unique characteristics of the BreakHis dataset is that the histopathology images are captured at four different magnification levels, namely 40x, 100x, 200x, and 400x. This allows the model to learn features at varying scales, simulating the real-world scenario where pathologists often examine tissue samples under different zoom levels to observe both macro and micro characteristics. The availability of images at multiple magnifications enhances the robustness of the classification model and provides a more comprehensive understanding of breast cancer tissue morphology.

2.2 Division of Bag

Each image in the BreakHis dataset has a resolution of 700×460 pixels and is available at four different magnification levels. Each whole slide image is divided into smaller patch sizes, resulting in 60 new patches per image. At this stage, the process begins with preparing the breast cancer histopathology images, followed by segmenting each whole image into small pieces (patches) of a fixed size, which then become instances.

All patches from one image or suspicious region are grouped into one bag based on certain criteria, such as image origin or tumor segmentation results. The labeling process assigns a positive label to any bag that contains at least one patch indicating cancer, while the bag without any cancer-indicative patches is labeled negative.

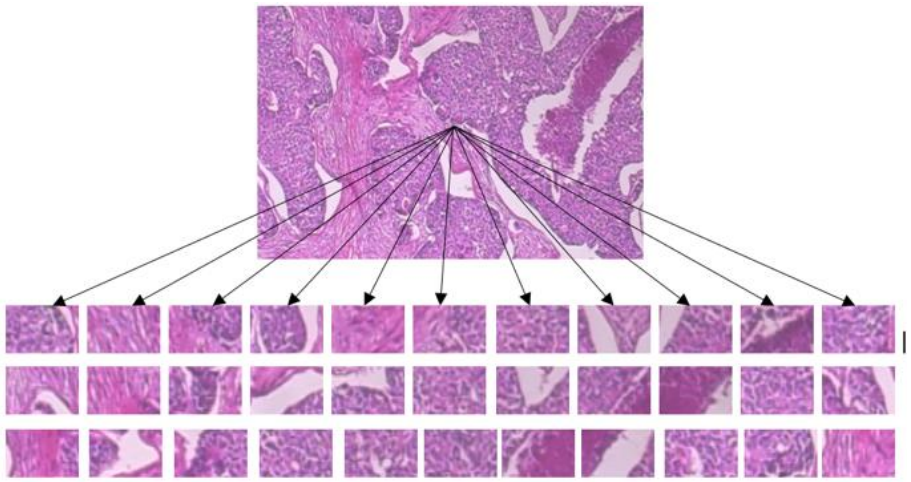

Figure 1. Patch Division

In the Attention Multiple Instance Learning (AMIL) method, the patch division process is a very important initial stage for handling large-scale histopathology images, such as whole slide images (WSI). Since the size of WSI is very large and complex, the image cannot be directly analyzed as a whole by the model. Therefore, as illustrated in Figure 1, each image is first divided into small pieces known as patches. Each image is treated as one bag, while the patch pieces in it are treated as instances.

Patch division is typically performed using a grid-based cropping method, where the image is cut into fixed-sized squares, such as 224×224 pixels. The patches may or may not overlap depending on the system design requirements and the desired level of detail. After division, patches that contain only background or irrelevant information are usually filtered out and ignored, so they do not affect the performance of the model. The extraction results in representation vectors, which are then combined to form one representation bag for the entire image.

2.3 Prepocessing

In the first stage, the image is reprocessed using data augmentation techniques, including vertical and horizontal flips as well as rotations of 90°, 180°, and 270°, with the aim of improving the performance of the model to be more effective. This augmentation technique produces a new dataset of 47,454 images.

After performing the data augmentation technique, the next step is to verify the dimensions of the augmented data to ensure that all images have the same size and that no NULL or improperly sized data remain.

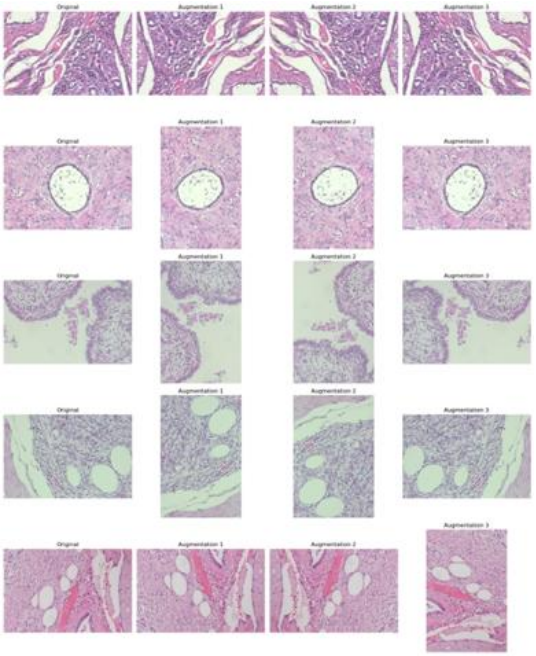

Figure 2. Augmentation Data

In Figure 2, after the data augmentation process was completed, testing was conducted on the VGG16 network initialized with weights that had been trained using the ImageNet dataset. For VGG16 training, random patches of size 224 x 224 were extracted from the training images. The following are the input and output images generated from the training process using VGG16.

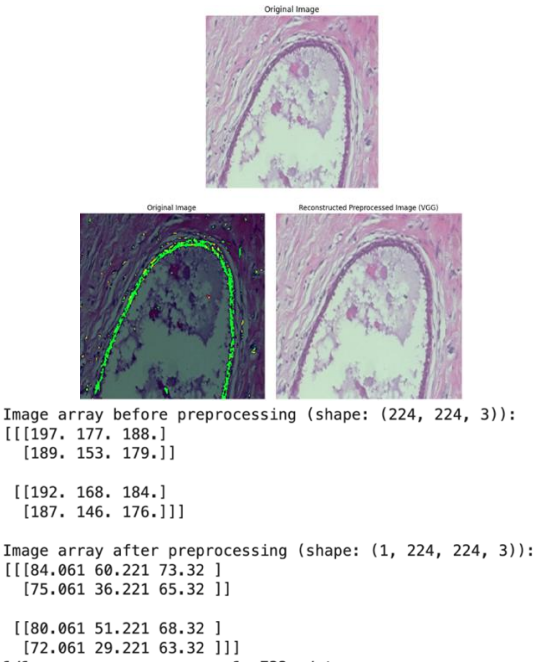

Figure 3. VGG16 Feature Extraction Result

Figure 3 shows the preprocessing pipeline applied to the histopathological images before feeding them into the deep learning model. The top panel presents the original histopathological image, while the bottom panels compare (left) the original image with overlaid feature activations and (right) the reconstructed preprocessed image produced by the VGG-based preprocessing step. The overlaid activations highlight the regions of interest extracted from the raw image, whereas the reconstructed image on the right demonstrates the normalized and resized version of the tissue sample used as input for the network.

The accompanying array outputs further confirm the effect of preprocessing. The image array before preprocessing shows raw pixel intensity values in the original RGB scale, whereas the image array after preprocessing reflects the normalized and scaled values with altered dynamic ranges, consistent with the VGG preprocessing routine. This transformation standardizes the input data to improve model stability and convergence during training while retaining diagnostically important structures.

These visual and numerical results collectively illustrate how the preprocessing stage adapts raw histopathological images to the feature space expected by the convolutional neural network, ensuring that subsequent layers focus on relevant patterns rather than variations in raw pixel intensity.

Finally, the extracted feature dimensions for both the training and testing data are displayed to ensure that the extraction process has been executed correctly. This step prepares the data for use in the subsequent learning models or for further analysis. Since both VGG16 networks have many adjustable parameters, a large data augmentation method is used to ensure effective training. This data augmentation includes various transformations such as horizontal and vertical flips, rotation at various angles, and changes to image scale and brightness.

Table 1. Preprocessing Architecture Detail

Dimension Input	Layer
[3 x 128 x 128]	Conv1
[64 x 128 x 128]	Maxpool
[64 x 64 x 64]	Conv2
[64 x 64 x 64]	Maxpool
[64 x 32 x 32]	Conv3

[128 x 32 x 32]	Maxpool
[128 x 16 x 16]	Conv4
[128 x 16 x 16]	Maxpool
[3 x 8 x 8]	Conv5
[124 x4 x 4]	Flatten
2048	FC1
1024	FC2

The architecture described in the Table 1 represents a series of feature extraction processes from histopathology images using a Convolutional Neural Network (CNN) approach. The process begins with an input image of size [3 x 128 x 128], where 3 denotes the RGB color channel and 128 x 128 corresponds to the spatial dimension of the image. This image is first processed through the initial convolution layer (Conv1), which produces 64 feature channels while maintaining the spatial dimension of [64 x 128 x 128]. A max-pooling layer is applied, reducing the spatial size to [64 x 64 x 64] to lower the computational complexity and minimize redundant features.

A similar process is repeated in subsequent layers. After Conv2 and the second pooling, the dimension becomes [64 x 32 x 32]. Conv3 increases the number of channels to 128, producing feature maps of size [128 x 32 x 32], followed by another pooling step that reduces the size to [128 x 16 x 16]. The Conv4 layer maintains this size before further pooling. Although the table reads the output after the fourth pooling as [3 x 8 x 8], this appears to be a typographical error, as the number of channels should remain at 128. The Conv5 layer then changes the output to [124 x 4 x 4], indicating the final features of the convolution process.

The final output of this process is then flattened into a one-dimensional vector of size 2048, which is then passed to two fully connected (FC) layers. The FC1 layer changes the dimension from 2048 to 1024, followed by FC2, which serves as an advanced stage of classification. This entire sequence serves as an important feature extraction stage before class prediction is performed, ensuring that the image is optimally represented for distinguishing between benign and malignant tissues.

2.4 Attention Mechanism

Each patch within the bag is processed in a packet by extracting features to fulfill the instance-level features. The dense layer extracts 500 features from each instance. The attention score block calculates the attention weight using 500 features from each instance. These attention weights are used during attention-based aggregation to obtain bag-level characteristics. This accumulation calculation makes the bag highly informative for classifiers at the bag level. The A-MIL framework is trained using a single batch size.

The table below illustrates the sequence of the feature extraction process in the A-MIL model that uses several layers in a convolutional neural network (CNN). The process begins with an input image of size 3 x 28 x 28 pixels (RGB). In the first step, the initial convolution layer (Conv1) is applied, resulting in features measuring 24 x 24 x 24. This is followed by a maxpooling layer, which reduces the spatial dimensions to 20 x 12 x 12. The second convolution layer (Conv2) produces features with dimensions of 50 x 8 x 8, which are reprocessed with maxpooling to produce a size of 50 x 4 x 4. The output is then flattened into a one-dimensional vector with a dimension of 800, which is then processed in the fully connected layer that produces an output with a dimension of 50.

Finally, important features are extracted with a final dimension of 50, which are used for advanced models or for classification.

Table 2. A-MIL Feature Extraction

Dimension Input	**Layer**
3 x 28 x 28	Conv1
24 x 24 x 24	Maxpool
20 x 12 x 12	Conv2
50 x 8 x 8	Maxpool
50 x 4 x 4	Flatten
800	Fully Connected
50	Extrac Feature

The feature extraction stage of the Attention-Multiple Instance Learning (A-MIL) method in Table 2 shows the convolutional network architecture designed to extract feature representations from each small image patch of 3 x 28 x 28, where 3 denotes the color channel (RGB) and 28 x 28 refers to the resolution of the patch being analyzed. The process begins with the first convolution layer (Conv1), which converts the input into a 24 x 24 x 24dimensional output,

indicating that 24 feature maps are successfully generated with a slight reduction in spatial size due to the convolution process.

Afterward, the output from Conv1 goes through a max-pooling process, which reduces the spatial dimensions while retaining the important features. This produces an output with dimensions of 20 x 12 x 12, showing 20 feature channels with a smaller spatial size. Next, the second convolution layer (Conv2) increases the number of channels to 50 while reducing the spatial size to 8 x 8, indicating a more complex feature deepening process. A second pooling layer further reduces the spatial size to 4 x 4, while maintaining the 50 channels.

The next stage is flattening, where the three-dimensional features [50 x 4 x 4] are flattened into a single 800-dimensional vector (the result of 50 × 4 × 4). This vector is then passed into a fully connected layer, producing a 50-dimensional feature representation, which serves as the final output of the feature extraction process. This 50-dimensional vector is subsequently used in the next stage of A-MIL, the attention mechanism, which determines the importance of each patch in the final classification process. Overall, this design ensures that each patch is represented efficiently, concisely, and informatively for multiple instance learning.

The A-MIL algorithm allows different samples within each bag to be assigned varying weights. This weighted aggregation ensures that the resulting bag representation is highly informative for bag-level classification. The A-MIL framework was trained using a single batch size, a learning rate of 0.001, and binary cross-entropy as the loss function.

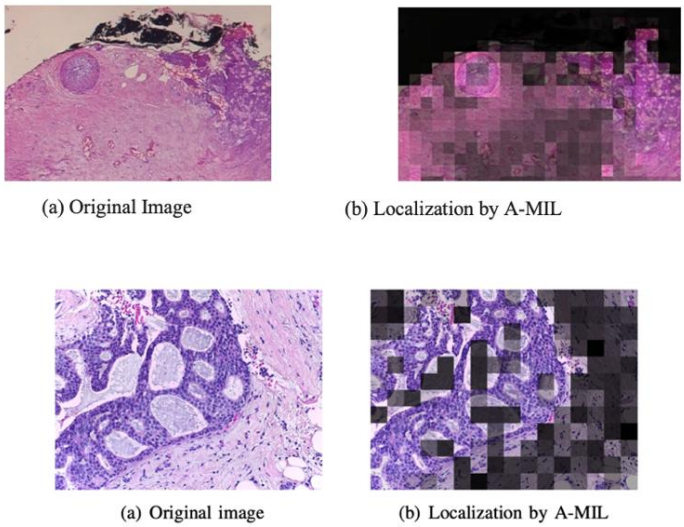

Figure 4. A-MIL Prediction

Figure 4 illustrates the visual interpretability of the proposed Attention-based Multiple Instance Learning (A-MIL) framework on histopathological images. Panels (a) present the original input images at different magnifications, whereas panels (b) depict the corresponding localization maps generated by A-MIL. The highlighted regions in the localization maps represent tissue areas that the model identified as most informative for classification. These attention maps clearly demonstrate that A-MIL is capable of focusing on diagnostically relevant structures—such as epithelial clusters or abnormal glandular formations—while down-weighting less informative regions. This visualization provides evidence that the model not only achieves high classification performance but also offers transparent and interpretable decision-making, which is crucial for building trust among pathologists and ensuring clinical applicability.

2.5 Multiple Instance Learning

MIL is useful in situations where obtaining precise label at the instance level is difficult, while labeling groups of instances is easier or more practical [22]. MIL offers a more effective and flexible approach for various cases in machine learning, where labeling individual instances is difficult or impractical. The purpose of MIL is to build a more complex machine learning model to predict the label of a new bag using each information from each labeled instance [23].

The most important part of Multiple Instance Learning (MIL) [22] is the instance combination process. The purpose of this model-level aggregation is to collect features from each instance and combine them to form a bag-level representation. Two commonly used aggregation methods in MIL are mean pooling and max pooling.

In mean pooling, the average of all instances in a bag is used to predict the bag label, meaning each instance contributes equally to the final outcome. In contrast, max pooling assigns the bag label based on the instance with the highest activation. This method focuses on very strong instances and is considered more suitable for determining bag labels.

However, both methods have limitations. Max pooling may rely on outliers—high-activation instances that do not truly represent the whole bag—resulting in the model failing to capture important information from the majority of the instances. On the other hand, mean pooling treats all instances equally, causing rare or minority instances to be overshadowed in the pooling process. This may reduce the performance of the model when working with data with fewer classes.

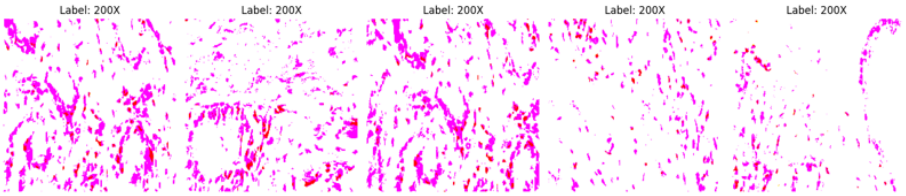

Figure 5. Data Labels in One Class

Based on Figure 5 above, a histopathology image is divided into many small patches, which are then grouped into a bag. If the image contains malignant cancerous tissue, the entire bag is labeled as malignant even though not all patches necessarily show malignant features. Some patches may contain only normal or irrelevant tissue. This illustrates the uneven distribution of label data within a class in the MIL method.

The Attention-based Multiple Instance Learning (A-MIL) model was developed with three main components: a feature extractor, an attention mechanism, and a classifier. The feature extractor transforms the input data into a latent representation, while the attention mechanism assigns importance weights to each instance within a bag. The classifier then uses the information from the attention mechanism to generate the final prediction.

The training process involves feeding the model with training data incrementally in batches. The model calculates the loss based on the difference between the prediction and the true label, then updates the weights using gradient-based optimization. After each training epoch, the model is evaluated using validation data to assess accuracy and loss, providing insight into its performance on unseen data.

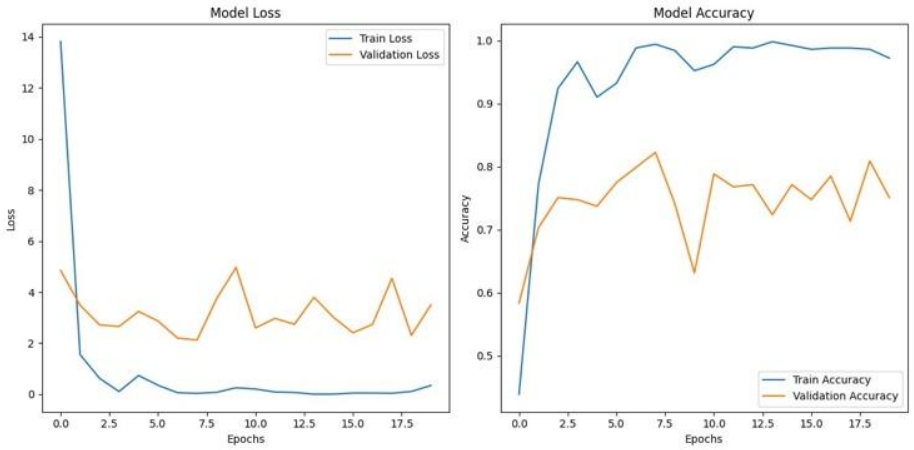

Figure 6. Validation and Loss

In Figure 6, once the model is trained, the testing process is performed. The model is used to make predictions on new data, including both individual images and image sets. For individual images, features are extracted using a pre-trained model (such as VGG16), and then organized into a dataset (bag) that can be processed by the A-MIL model.

Predictions are generated in the form of class probabilities, and the class with the highest probability is selected as the final output. The model's performance is evaluated using metrics such as classification report, which provides insight into accuracy, precision, and recall. This process ensures that the model is not only learns patterns from the training data but is also capable of generalizing to new data.

3. Results and Discussion

At the A-MIL testing stage, which used 80% training data, 10% testing data, and 10% validation data from the BreakHIS dataset with four different magnification levels, the accuracy results obtained are as follows:

Table 3. A-MIL Classification Accuracy Result

Training Data	Testing Data	Accuracy			
		40x	100x	200x	400x
80% 6.327 image	20% 1582 image	89.90%	95.50%	97.50%	90.43%

Based on Table 3 above, the highest accuracy was achieved in the class trial with an image magnification level of 200x, reaching 97.50%. Figure 7 shows an image of a sample of the cancer classification visualization results using a dataset with a magnification level of 100x.

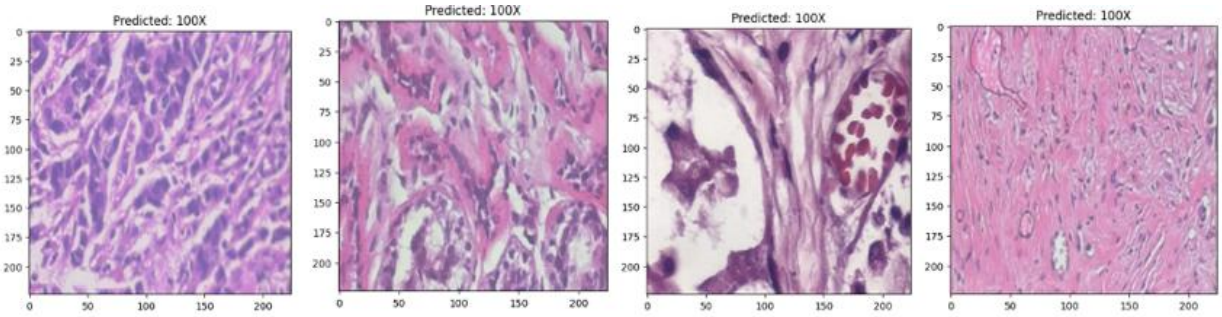

Figure 7. Visualization of Cancer Classification Output

The initial prediction result is a probability score for each class, which is then converted into a final label by selecting the class with the highest probability using the argmax function. Next, the numerical prediction labels are converted back to the original labels using the label encoder previously used during training. The same process is performed on the original labels of the test data to ensure comparability.

Afterward, a classification report is generated using the classification_report function, presenting metrics such as accuracy, precision, recall, and F1-score for each class. Additionally, a confusion matrix is calculated to provide a detailed overview of the distribution of correct and incorrect predictions. This confusion matrix is then visualized as a heatmap using Seaborn to facilitate analysis of the test results, as shown below.

Table 4. Classification Matrix

	Precision	Recall	F1 - Score
***Benign* - 400x**	0.88	0.81	0.84
***Benign* - 200x**	1.00	0.82	0.90
***Benign* - 100x**	0.92	0.69	0.87
***Benign* - 40x**	0.90	0.78	0.86
***Malignant* - 400x**	0.87	0.62	0.90
***Malignant* - 200x**	0.89	0.88	0.89
***Malignant* - 100x**	0.82	0.82	0.89
***Malignant* - 40x**	0.84	0.83	0.87

Based on Table 4, the model's performance in classifying benign tissues shows consistently high precision across all magnifications, with the highest value at 200× (1.00) and the lowest at 100× (0.92). The recall for the benign class varies between 0.69 and 0.82, indicating that although the model is highly precise at most magnifications, its ability to correctly retrieve all benign samples is less optimal, particularly at 100×. Consequently, the F1-scores range from 0.84 to 0.90, reflecting moderate to high overall performance.

For the malignant tissue classification, the model exhibits relatively stable precision (0.82–0.89), with recall being lower at 400× (0.62) but improving at lower magnifications (0.82–0.88). The F1-scores for the malignant class remain high (0.87–0.90), indicating a balanced trade-off between precision and recall. These findings suggest that image magnification influences the model's performance, with intermediate magnifications (e.g., 200×) providing the most consistent and optimal results for both benign and malignant tissues.

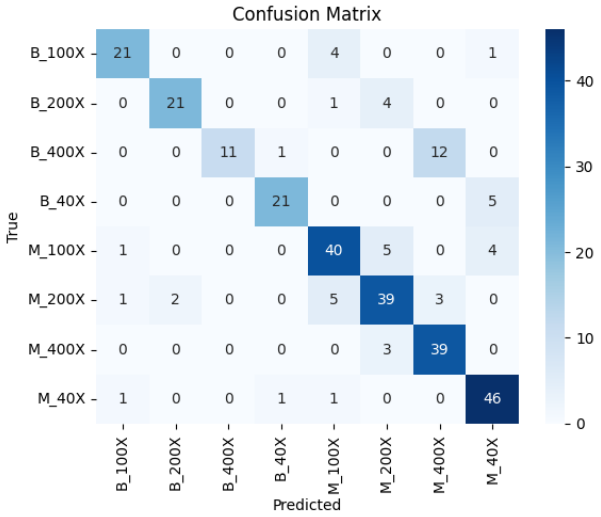

Figure 8. Classification Matrix Graph

Based on Figure 8, the classification matrix evaluates the performance of the breast histopathology image classification model across two main categories, namely Benign and Malignant, each further classified based on microscope magnification levels of 40X, 100X, 200X, and 400X. This matrix compares the actual labels (vertical axis) with the model's predicted labels (horizontal axis). Each cell represents the number of images predicted for a particular category. Correct predictions appear along the main diagonal from the top left to the bottom right, highlighted in dark blue to represent a high value.

Overall, the model shows good classification performance, with most predictions concentrated along the main diagonal. For example, in the B_100X class, 21 images were correctly classified, and in the M_40X class, 46 images were also accurately identified. However, several notable misclassifications occurred. For example, in the B_400X class, 12 images were incorrectly predicted as M_400X, indicating that the model had difficulty distinguishing benign and malignant tissues at that level of magnification. Similarly, in the M_100X class, some images were misclassified into the M_200X and M_400X classes, indicating that visual features at multiple resolutions tend to be similar and difficult for the model to distinguish.

These findings show that, although the model has performed adequately, further improvement is needed, especially in distinguishing high-resolution images that have similar features between classes. Improvements can be made through approaches such as data augmentation, class balancing, or the use of more complex architectures to capture finer visual patterns. By doing so, it is hoped that classification accuracy can be improved across all image categories.

4. Conclusion

Based on the results of the research conducted, it can be concluded that the Multiple Instance Learning method can be used to classify breast cancer using histopathology images, as reflected in the accuracy obtained. This study has shown better classification performance through the use of the attention method, which enhances accuracy in noise-prone areas by providing more focused exposure related to histopathology images to spot patterns. The highest accuracy was achieved using images with a 200x magnification level, reaching 97.50%.

To support the continuation and further development of this research, several suggestions are proposed. These include increasing the number of datasets used for the classification process for varied data such as increasing the number of classes or adding features related to data preprocessing, in the hope of improving accuracy results, as well as adding or developing the methods used in order for comparisons, such as adding feature extraction or using two classification methods as a performance comparison.

Acknowledgement

Thank you to everyone who contributed to this research, especially Mr. Edi and Mr. Affandy. I also extend my gratitude to previous researchers whose work provided valuable references that supported the completion of this study.

References

[1] J. S. Brown, S. R. Amend, R. H. Austin, R. A. Gatenby, E. U. Hammarlund, and K. J. Pienta, “Updating the Definition of Cancer,” *Molecular Cancer Research*, vol. 21, no. 11, pp. 1142–1147, 2023. https://doi.org/10.1158/1541-7786.MCR-23-0411

[2] T. Agustin, “Potensi Metabolit Aktif Dalam Sayuran Cruciferous Untuk Menghambat Pertumbuhan Sel Kanker,” 2020.

[3] E. Marfianti, "Peningkatan Pengetahuan Kanker Payudara dan Ketrampilan Periksa Payudara Sendiri (SADARI) untuk Deteksi Dini Kanker Payudara di Semutan Jatimulyo Dlingo," 2021. https://doi.org/10.20885/jamali.vol3.iss1.art4
[4] J. Zhou, Z. Wu, D. Aili, L. Wang, and T. Liu, "Exploration of the carcinogenetic and immune role of CHK1 in human cancer," *J Cancer*, vol. 15, no. 18, pp. 5927–5941, 2024. https://doi.org/10.7150/jca.93930
[5] W. Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, "Klasifikasi Penyakit Kangker Payudara Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2022, no. 12, pp. 445–452, 2022.
[6] Z. Zhang *et al.*, "A new clinical prognosis model for breast cancer with ADSS as the hub gene," *J Cancer*, vol. 15, no. 18, pp. 5910–5926, 2024. https://doi.org/10.7150/jca.95589
[7] A. Indah Sari, "Skrining Mamografi dan Mortalitas Kanker Payudara," vol. 7, no. 7, p. 11, 2022.
[8] R. Resmiati and T. Arifin, "SISTEMASI: Jurnal Sistem Informasi Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination," 2021. https://doi.org/10.32520/stmsi.v10i2.1238
[9] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021. https://doi.org/10.3322/caac.21660
[10] B. Zhang, H. Shi, and H. Wang, "Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach," 2023, *Dove Medical Press Ltd*. https://doi.org/10.2147/JMDH.S410301
[11] N. Aidossov *et al.*, "An Integrated Intelligent System for Breast Cancer Detection at Early Stages Using IR Images and Machine Learning Methods with Explainability," *SN Comput Sci*, vol. 4, no. 2, Mar. 2023. https://doi.org/10.1007/s42979-022-01536-9
[12] I. Idawati, D. P. Rini, A. Primanita, and T. Saputra, "Klasifikasi Kanker Payudara Menggunakan Metode Convolutional Neural Network (CNN) dengan Arsitektur VGG-16," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 3, p. 529, Apr. 2024. https://doi.org/10.30865/json.v5i3.7553
[13] X. Xu, Q. Guo, Z. Li, and D. Li, "Uncertainty Ordinal Multi-Instance Learning for Breast Cancer Diagnosis," *Healthcare (Switzerland)*, vol. 10, no. 11, Nov. 2022. https://doi.org/10.3390/healthcare10112300
[14] S. A. Alanazi *et al.*, "Boosting Breast Cancer Detection Using Convolutional Neural Network," *J Healthc Eng*, vol. 2021, 2021. https://doi.org/10.1155/2021/5528622
[15] P. Kumar, S. Srivastava, R. K. Mishra, and Y. P. Sai, "End-to-end improved convolutional neural network model for breast cancer detection using mammographic data," *Journal of Defense Modeling and Simulation*, vol. 19, no. 3, pp. 375–384, Jul. 2022. https://doi.org/10.1177/1548512920973268
[16] D. Kumar Saha, T. Hossain, M. Safran, S. Alfarhood, M. F. Mridha, and D. Che, "Segmentation for mammography classification utilizing deep convolutional neural network," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024. https://doi.org/10.1186/s12880-024-01510-2
[17] J. Nouri Pour, M. A. Pourmina, and M. N. Moghaddasi, "Improving Breast Cancer Detection with Convolutional Neural Networks and Modified ResNet Architecture," *Curr Med Imaging Rev*, vol. 20, Apr. 2024. https://doi.org/10.2174/0115734056290499240402102301
[18] N. Yudistira, M. S. Kavitha, J. Rajan, and T. Kurita, "Attention-effective multiple instance learning on weakly stem cell colony segmentation," *Intelligent Systems with Applications*, vol. 17, Feb. 2023. https://doi.org/10.1016/j.iswa.2023.200187
[19] T. P. Theodore Armand, S. Bhattacharjee, and H. C. Kim, "Overview of the Potentials of Multiple Instance Learning in Cancer Diagnosis: Applications, Challenges, and Future Directions," in *International Conference on Advanced Communication Technology, ICACT*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 419–425. https://doi.org/10.23919/ICACT60172.2024.10471995
[20] Z. Shao *et al.*, "TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification," Jun. 2021. https://doi.org/10.48550/arXiv.2106.00908
[21] Y. Kim, T. Wang, D. Xiong, X. Wang, and S. Park, "Multiple instance neural networks based on sparse attention for cancer detection using T-cell receptor sequences," *BMC Bioinformatics*, vol. 23, no. 1, Dec. 2022. https://doi.org/10.1186/s12859-022-05012-2
[22] M. A. Carbonneau, E. Granger, and G. Gagnon, "Bag-Level Aggregation for Multiple-Instance Active Learning in Instance Classification Problems," *IEEE Trans Neural Netw Learn Syst*, vol. 30, no. 5, pp. 1441–1451, May 2019. https://doi.org/10.1109/TNNLS.2018.2869164
[23] L. Cai, S. Huang, Y. Zhang, J. Lu, and Y. Zhang, "Rethinking Attention-Based Multiple Instance Learning for Whole-Slide Pathological Image Classification: An Instance Attribute Viewpoint," Mar. 2024. https://doi.org/10.48550/arXiv.2404.00351
[24] H. Xiang, J. Shen, Q. Yan, M. Xu, X. Shi, and X. Zhu, "Multi-scale representation attention based deep multiple instance learning for gigapixel whole slide image analysis," *Med Image Anal*, vol. 89, p. 102890, 2023. https://doi.org/10.1016/j.media.2023.102890
[25] S. Fatima, S. Ali, and H. C. Kim, "A Comprehensive Review on Multiple Instance Learning," Oct. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. https://doi.org/10.3390/electronics12204323