Multi Scale and Multi-Source Fusion Framework for Breast Cancer Detection Using YOLOv8

Laavanya Mohan[1], Shashi Kant Gupta[2]

[1]Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia; [2]Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

laavanvijay@gmail.com

Abstract: Breast cancer detection from histopathological images requires accurate analysis of both fine-grained cellular structures and broader tissue level context. Most approaches rely on microscopic images which capture detailed cellular features but often lack global structural information while whole slide images (WSIs) provide comprehensive tissue context but are computationally expensive to process directly. Therefore, this work proposes a multi-sourced learning framework that integrate microscopic BreakHis images with contextual information derived from WSIs. In addition, a patch-based strategy is employed to extract informative regions from WSIs using a sliding window approach; then an intensity-based filtering is used to remove background regions and texture-based filtering is used to eliminate non-informative patches. Furthermore, a novel "on the fly" fusion mechanism is introduced where each BreakHis image will be dynamically combined with a randomly selected WSI patch using weighted blending thus enhancing contextual representation while preserving label integrity and avoiding additional memory overhead. The fused dataset will be constructed in a yolo compatible format allowing for efficient training without storing intermediate data in memory. A pre-trained YOLOv8 nano model will be fine-tuned on the generated dataset for tumor detection and classification. Experimental results demonstrate improved performance in terms of precision, recall, and mAP@50 alongside stable convergence behaviour. As such the proposed framework effectively captures multi-scale features while maintaining computational efficiency making it suitable for real-time and large-scale medical imaging applications.

Keywords: Breast Cancer Detection; Histopathological Image Analysis; Multi-Source Learning; Whole Slide Images; YOLOv8 Object Detection.

Introduction

Breast cancer is one of the leading reasons for deaths among females around the globe, and hence early and accurate diagnosis plays a vital role in raising the survival rate among them. The analysis of histopathological images has been considered a golden standard for breast cancer diagnosis, as it helps obtain in-depth knowledge about structures present at the cellular and tissue levels. However, manual analysis of such images has been a time-consuming and subjective process, and hence there has been a growing concern for developing automated approaches for image analysis.

In recent times, deep learning has shown tremendous improvements in the automated analysis of histopathological images, and hence accuracy and efficiency in cancer detection can be enhanced. The conventional deep learning-based approaches rely on microscopic images, such as those obtained from the BreakHis dataset, which contain fine-grained features present in breast cancer tissues. However, such approaches often do not consider global tissue level features, which play a vital role in accurate diagnosis. On the other hand, whole slide images contain features present at the tissue level, and hence they are computationally expensive, as they contain a large amount of resolution information.

Recent studies have also addressed the problem through the use of multi-magnification learning, attention-based models, and patch-based analysis for WSI to improve performance. Even though these approaches improve feature representation, these methods are mostly limited to classification tasks and are not able to provide an end-to-end approach for detection that includes localization and classification. Moreover, the use of heterogeneous sources, such as microscopic images and WSI, is a problem that is still to be addressed, especially because of computational and memory limitations.

To improve the performance and address the limitations, the proposed approach utilizes both multi-scale and multi-source approaches for breast cancer detection. The approach utilizes an on-the-fly approach that is able to dynamically combine local cell features and global tissue features without the need for large amounts of data to be stored in memory. Additionally, a YOLOv8 approach is used for classification and localization, utilizing pseudo bounding boxes for the images. The dataset is also constructed using a disk-based approach.

Background

Deep learning methods have shown promise for histopathology analysis in breast cancer image analysis in recent times, and the use of data augmentation and multi-magnification learning has been crucial for improving performance, where features are learned at tissue and cellular levels, as shown in various studies that use multi-resolution networks, VGG16 with XGBoost and CBAM, and channel-spatial attention networks; however, such methods are found to be ineffective for dealing with inter-magnification variations and cross-scale feature extraction [1-3]. Attention mechanisms have shown promise for improving feature interpretability and selectivity, and BreNet, cascaded attention networks for WSIs, and SE Conformer have shown remarkable accuracy; however, such methods are found to be ineffective for dealing with object detection scenarios [4-6]. The use of patch-based and WSIs has shown promise for bridging the gap between image analysis and high-resolution images, and coarse annotation ensembles and hybrid CNN-transformer networks have shown promise for improving object detection accuracy; however, such methods are found to be ineffective for dealing with object detection and segmentation scenarios [7-8]. The use of YOLO object detection networks and various hybrids has shown promise for object detection scenarios, and this highlights the need for improving object detection and its interpretability; however, a unified framework that incorporates object detection, segmentation, and multi-scale fusion and attention mechanisms is yet to be developed, as shown in Table 1 [9-12].

Table 1. Comparison of related works with proposed method

Ref	Multi Scale Learning	WSI Utilization	Fusion Strategy	Detection Capability
[1]	Yes	No	No	No
[2]	No	No	No	No

[3]	No	No	No	No
[4]	Yes	No	No	No
[5]	Yes	Yes	No	No
[6]	Yes	No	No	No
[7]	Yes	No	Yes	Yes
[8]	Yes	No	No	No
[9]	Yes	No	Yes	Yes
[10]	Yes	No	Yes	Yes
[11]	No	No	No	No
[12]	Yes	Yes	Yes	No
This Work	Yes	Yes	Yes	Yes

Key Contribution

The major contributions of this work are summarized as follows:

- **Leverages both multi scale and multi-source data for improved breast cancer detection**: A unified pipeline integrating BreakHis microscopic images with WSI derived contextual patches for multi scale feature representation.
- **On the Fly Fusion Strategy**: A novel memory efficient fusion mechanism that dynamically combines WSI patches with labeled images, avoiding label ambiguity and reducing memory overhead.
- **Efficient WSI Patch Filtering**: Implementation of intensity and texture-based filtering to extract only diagnostically relevant regions from WSIs.
- **Adaptation of YOLOv8 for Histopathology**: Transformation of a classification dataset into a detection framework using pseudo bounding boxes and fine tuning of a lightweight YOLOv8 model.

Method

Data Acquisition and Multi-Source Integration: The framework utilizes two types of breast histopathology images to effectively capture multi-scale features. The first image is derived from the BreakHis dataset, containing labeled microscopic images with fine-grained cellular details. The second image is derived from the whole slide image, used to extract multi-scale features at the tissue level. This integration is necessary to enable the framework to learn both local and global features, which are vital for accurate breast cancer detection.

Whole Slide Image Patch Extraction and Filtering: Due to the high-resolution nature of whole slide images, processing is computationally expensive. Therefore, the framework is designed to use a patch-based method to segment the whole slide image into smaller regions using the sliding window method. To filter the image and retain only relevant information, some patches are discarded. Patches with low average intensity are discarded to filter out background information. Similarly, patches with low variance are discarded to filter out non-informative textures. The remaining patches are resized to fit the dimensions of the input image used in the detection model.

BreakHis Data Preprocessing and Augmentation: The BreakHis dataset is preprocessed to resize each image to a uniform size of 640 x 640 pixels. Data augmentation techniques are also used to prevent overfitting. Each image is assigned a binary label to represent the two types of breast cancer. This is used to generate the ground truth for training the detection model.

On the Fly Multi-Source Fusion Strategy: To enable the integration of contextual information with high efficiency, an on-flight fusion strategy is proposed. Instead of using preprocessed fusion images, each image in the BreakHis dataset is fused with a randomly selected whole slide image patch. The fusion is done using weighted blending:

$$I_{fused} = \alpha I_{BreakHis} + (1 - \alpha) I_{WSI}$$

where α=0.7. This approach enriches the input data with contextual information while preserving the original labels and avoiding additional memory usage.

Memory Efficient Dataset Construction: The construction of the dataset is performed sequentially, and the images are saved on the disk as soon as they are constructed, ensuring a YOLO-compatible structure while eliminating the need to load the entire dataset into memory for training–validation splitting, thereby significantly reducing memory usage. The detection framework is based on YOLOv8, a state-of-the-art model that includes a backbone for feature extraction, a neck for feature aggregation, and a head for object detection; since the BreakHis dataset does not include bounding box annotations, pseudo bounding boxes are generated using normalized coordinates, and a YOLOv8 nano model is fine-tuned on the prepared dataset. The model is trained using an input resolution of 640 × 640 pixels with a batch size of 16 over 15 epochs, incorporating early stopping to prevent overfitting, while optimization is performed using a composite loss function that combines localization, classification, and confidence losses to effectively learn both spatial and semantic features.

The proposed architecture in Figure 1 introduces a multi-scale, multi-source fusion framework for breast cancer detection by combining BreakHis microscopic images with contextual information from whole slide images using YOLOv8. BreakHis images are preprocessed and augmented, while WSI patches are extracted and filtered to retain meaningful regions. An on-the-fly fusion module dynamically blends both sources, enabling learning of local and global features without high memory cost. The fused data are stored in a YOLO-compatible format with pseudo bounding boxes for efficient training. The YOLOv8 model then performs feature extraction, aggregation, and detection. Finally, the system outputs benign or malignant predictions with corresponding bounding boxes for accurate analysis.

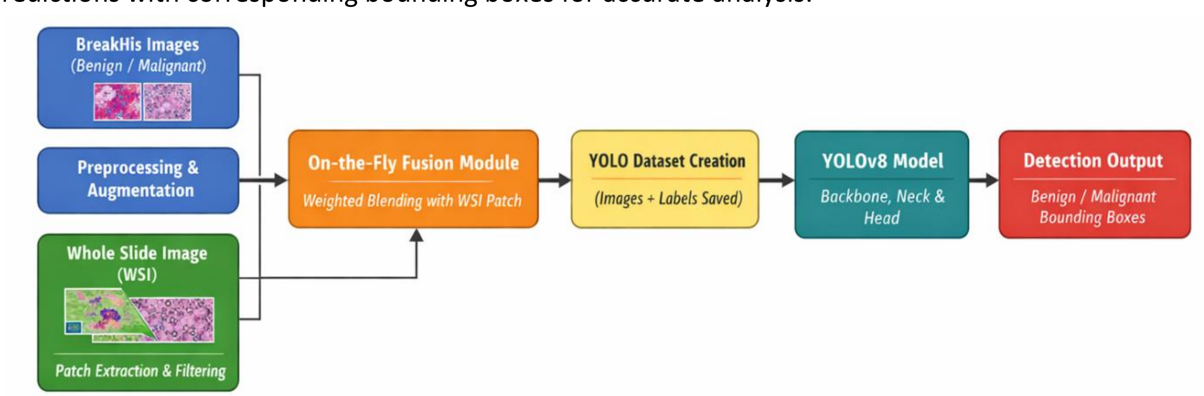

Figure 1. The proposed model.

Experiments and Results

The performance of the proposed multi-source fusion framework is evaluated using standard metrics such as loss curves, precision, recall, and mean average precision (mAP@50), demonstrating strong convergence and high detection accuracy. The loss curves (Figure 1) show that both bounding box and classification losses start high and gradually decrease, stabilizing at low values in later epochs, which indicates effective learning and stable optimization without divergence. Similarly, the precision–recall curves (Figure 2) initially fluctuate but quickly improve, reaching values close to 1.0, reflecting high correctness and completeness of predictions due to effective feature fusion. The mAP@50 curve (Figure 2) shows rapid improvement followed by saturation near its peak, with the model achieving an excellent score of 0.995, confirming its ability to learn discriminative features efficiently. Quantitatively, the model achieves mAP@50 of 0.995, precision of 0.9996, and recall of 1.0, indicating near-perfect performance in breast cancer detection and classification. Visual results (Figure 2) further confirm that the model accurately localizes tumor regions and consistently classifies them as malignant across varying tissue structures, correctly identifying key pathological features such as dense cells, irregular nuclei, and abnormal tissue patterns. The confidence scores, ranging from approximately 0.26 to 0.45, suggest moderate certainty, indicating room for improvement in confidence calibration. Additionally, overlapping bounding boxes are observed in some cases, showing sensitivity to subtle feature variations but also indicating redundancy due to non-optimal non-maximum suppression (NMS) and the use of pseudo bounding box annotations; nevertheless, the consistent malignant classification across samples highlights the robustness of the learned representations.

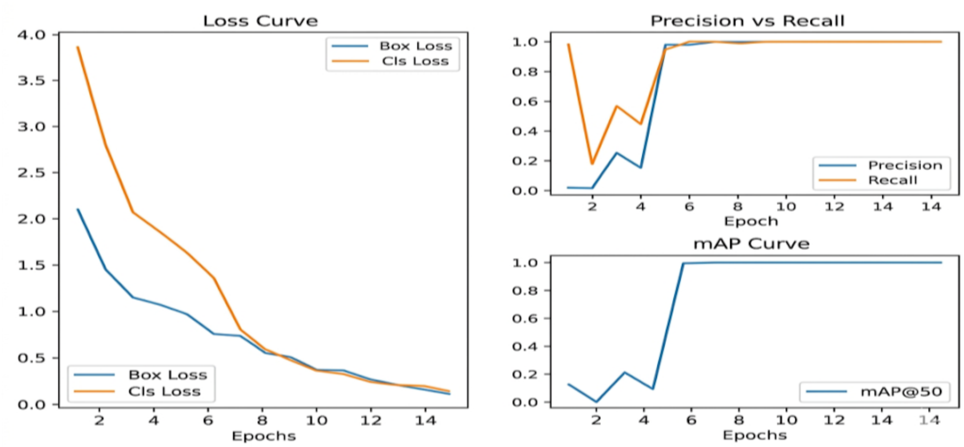

Figure 2. Evaluation of model training performance metrics over epochs.

Furthermore, the model performs well across different texture distributions and magnification characteristics, validating the effectiveness of the proposed multi-source fusion strategy in capturing both local cellular details and global tissue context. However, the relatively moderate confidence values and coarse localization suggest that incorporating fully annotated datasets and improving post-processing techniques could further enhance detection precision. Overall, the results confirm that the proposed framework achieves reliable and consistent malignant region detection while maintaining computational efficiency.

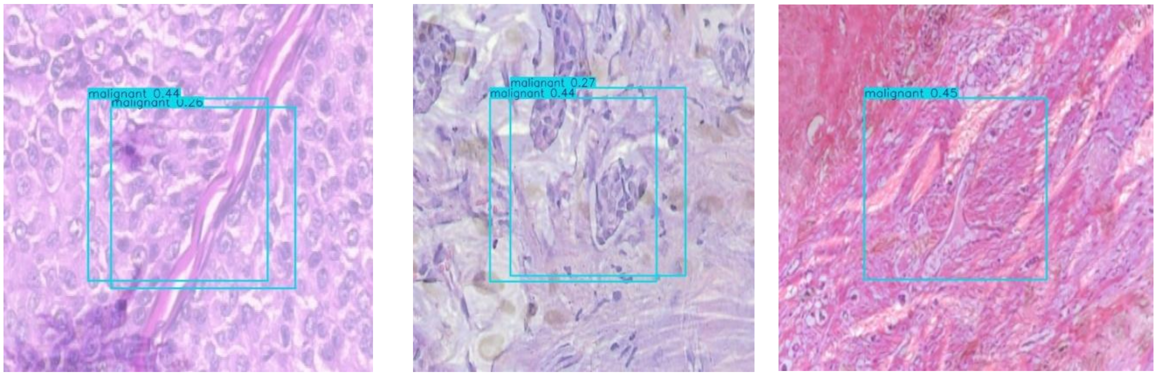

Figure 3. Localization of tumor regions

Discussions

The superior performance can be attributed to the proposed multi scale and multi-source fusion strategy, which combines microscopic cellular features from BreakHis images with contextual information from WSI patches. This fusion enhances the model's ability to capture both local and global patterns. Additionally, the on-the-fly data generation pipeline ensures efficient training without memory constraints, enabling the use of diverse training samples.

The high precision value indicates that the model produces very few false positives, while the perfect recall suggests that almost all relevant instances are correctly detected. The stable loss curves further confirm the robustness of the training process.

Overall, the results demonstrate that the proposed framework achieves highly accurate and stable performance, validating the effectiveness of multi-source fusion learning for breast cancer detection.

Conclusion

This work describes a multi scale and multi-source fusion framework for breast cancer detection, where microscopic BreakHis images are fused with contextual information extracted from whole slide images. The proposed method utilizes an on-the-fly fusion technique, facilitating the effective combination of local cellular information and global tissue information without high memory overhead. Additionally, a YOLOv8-based detection model is fine-tuned with pseudo bounding box annotations for simultaneous classification and detection of breast cancer regions. The experimental outcomes show that the proposed method achieves superior performance with a mAP@50 of 0.995, precision of 0.9996, and recall of 1.0, along with stable convergence during training. These outcomes confirm the efficacy of the proposed fusion strategy in improving feature representation and detection accuracy. Moreover, disk-based direct construction of datasets facilitates scalability and computational efficiency for real-world applications. Even though satisfactory outcomes have been reported for this work, it is still subject to some limitations, such as the utilization of pseudo annotations and a small number of datasets. In future, this work can be extended to explore more annotated datasets, attention mechanisms, and a segmentation-based analysis approach for more accurate detection of breast cancer regions. In conclusion, this work can be considered an efficient and effective approach for multi scale breast cancer detection and can be extended for more medical image analysis applications.

References

1. D. J. Ho et al., “Deep Multi Magnification Networks for multi class breast cancer image segmentation,” Comput. Med. Imaging Graph., vol. 88, p. 101866, 2021. DOI: 10.1016/j.compmedimag.2021.101866
2. V. A. V. Winnarasi and V. Vaishnavi, “Enhanced Breast Cancer Diagnosis Using Customized Transfer Learning and Attention Mechanisms,” Proc. ICCPCT, 2024. DOI: 10.1109/ICCPCT61902.2024.10673240
3. L. A. Aldakhil et al., “Leveraging Attention Based Deep Learning in Binary Classification for Early Stage Breast Cancer Diagnosis,” Diagnostics, vol. 15, no. 6, p. 718, 2025. DOI: 10.3390/diagnostics15060718
4. A. Alshehri, “BreNet: A Multi Scale CNN Ensemble with Attention Mechanisms for Breast Cancer Classification,” IEEE Access, vol. 13, pp. 359845–359856, 2025. DOI: 10.1109/ACCESS.2025.3598450
5. D. Liu and B. Hu, “A Novel Framework for Whole Slide Pathological Image Classification Based on the Cascaded Attention Mechanism,” Sensors, vol. 25, no. 3, p. 726, 2025. DOI: 10.3390/s25030726
6. L. S. Nair et al., “Advancing Breast Cancer Detection: SE Conformer Framework,” IEEE Access, vol. 13, pp. 357682–357690, 2025. DOI: 10.1109/ACCESS.2025.3576825
7. M. Naouar et al., “Robust Tumor Detection from Coarse Annotations via Multi Magnification Ensembles,” arXiv:2303.16533, 2023. DOI: 10.48550/arXiv.2303.16533
8. R. Monjezi et al., “RI ViT: Residual Hybrid Vision Transformer for Breast Cancer Classification,” IEEE Access, vol. 12, pp. 335868–335879, 2024. DOI: 10.1109/ACCESS.2024.3358686
9. A. D. Mohammed and D. Ekmekci, “Breast Cancer Diagnosis Using YOLO Based Multiscale Parallel CNN,” Appl. Sci., vol. 14, no. 7, p. 2680, 2024. DOI: 10.3390/app14072680
10. A. Anas et al., “Hybrid YOLOv5 and Mask R CNN Framework for Improved Breast Lesion Detection,” IEEE Access, vol. 12, pp. 351432–351441, 2024. DOI: 10.1109/ACCESS.2024.3514322
11. A. Maleki et al., “Breast Cancer Diagnosis from Histopathology Images Using DenseNet201 and XGBoost,” Biomed. Signal Process. Control, vol. 86, p. 105152, 2023. DOI: 10.1016/j.bspc.2023.105152
12. F. Abdullakutty et al., “Explainable Multi Modal Approaches for Breast Cancer Diagnosis,” Front. Med., vol. 11, Art. 1450103, 2024. DOI: 10.3389/fmed.2024.1450103