Leveraging Attention Mechanisms for Breast Cancer Diagnosis

Akumalla Brahma Reddy[*], Bach-Tung Pham[*], Tung-Yu Zhuang [*], Bima Paristao [‡], Pao-Chi Chang[†], Jia-Ching Wang[*]
[*]Department of Computer Science and Information Engineering, National Central University, Taiwan
[†]Department of Communication Engineering, National Central University, Taiwan
[†]Department of Informatics, Institut Teknologi Kalimantan, Indonesia
Email: jcw@csie.ncu.edu.tw

***Abstract*—Recent advancements in deep learning have significantly impacted breast cancer diagnosis, enhancing the accuracy and efficiency of detection and classification. In this study, we applied various attention mechanisms in different configurations to evaluate their effects on our model's performance, specifically in terms of F1 score and accuracy. We utilized ConvNeXt for initial feature extraction, followed by the implementation of attention mechanisms such as SRM (Spatial Relation Module), GCT (Global Contextual Transformer), LCT (Local Contextual Transformer), and Triplet Attention. These attention mechanisms were arranged both in parallel and series configurations to explore their impact on feature extraction. Our experiments demonstrated that the choice and arrangement of attention mechanisms significantly influenced the model's performance, with the Triplet_LCT in the parallel configuration achieving the highest accuracy and F1 score.**

I. Introduction

Breast cancer is a disease where cells in the breast grow uncontrollably. It is one of the most common cancers affecting women worldwide, though it can also occur in men. Breast cancer can manifest in various forms, primarily categorized into benign (non-cancerous) and malignant (cancerous) masses and calcifications.Benign masses are non-cancerous growths in the breast. They typically have smooth, well-defined edges and are often oval or circular in shape. Common types of benign breast masses include fibroadenomas and cysts. While benign masses are not life-threatening, they may require monitoring or removal if they cause discomfort or other issues.Malignant masses, on the other hand, are cancerous and can invade surrounding tissues. These masses often have irregular shapes and poorly defined edges. Malignant tumors can spread (metastasize) to other parts of the body, making early detection and treatment crucial. The most common types of malignant breast tumors include invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC)[1].

Breast calcifications are small deposits of calcium that appear as white spots on a mammogram. They are common, especially in women over 50, and can be classified into two types: microcalcifications and macrocalcifications.Microcalcifications are tiny calcium deposits that may indicate the presence of breast cancer, particularly ductal carcinoma in situ (DCIS), which is an early form of breast cancer. These calcifications often appear in clusters and may require further investigation through a biopsy to determine if they are benign or malignant .Macrocalcifications are larger, coarser calcium deposits that are usually benign. They are often associated with aging, past injuries, or benign breast conditions such as fibroadenomas or cysts. Macrocalcifications typically do not require further testing unless they appear in a suspicious pattern.

Machine learning (ML) has become a pivotal tool in the fight against breast cancer, offering significant advancements in early detection, diagnosis, and treatment planning. Here are some key areas where machine learning is making an impact:

Machine learning algorithms, particularly deep learning models like convolutional neural networks (CNNs), are being used to analyze mammograms, ultrasounds, and MRI scans to detect breast cancer at an early stage. These models can identify patterns and anomalies that may be indicative of cancer, often with higher accuracy than traditional methods. For instance, models like Inception-V3 and ResNet-50 have shown superior performance in identifying various breast tissue types and detecting abnormalities[2].

ML techniques are also employed to classify breast lesions as benign or malignant. This involves analyzing features from medical images to determine the nature of the lesion. Advanced models, such as those incorporating transfer learning and ensemble methods, have demonstrated high accuracy in distinguishing between benign and malignant masses. For example, a study using a deep ensemble-based model achieved 99% accuracy on the MIAS datase[3]. Predictive models using machine learning can assess the risk of breast cancer recurrence. By analyzing patient data, including tumor characteristics and treatment history, these models can provide valuable insights into the likelihood of cancer returning. Algorithms like AdaBoost have been used to develop robust prediction models, helping clinicians make informed decisions about follow-up care and treatment[4].

Attention mechanisms have become a crucial component in enhancing the performance of deep learning models for breast cancer detection and classification. These mechanisms help models focus on the most relevant parts of the input data, improving accuracy and interpretability. Here are some key applications of attention mechanisms in breast cancer research: Attention mechanisms can refine the feature extraction process by emphasizing important regions in medical images. For instance, the Efficient Channel Spatial Attention Network (ECSAnet) integrates a convolutional block attention module (CBAM) to improve the classification of histopathological images. This approach has shown superior performance compared to traditional models like AlexNet and ResNet50[5].Incorporating attention mechanisms into deep learning models can significantly boost classification accuracy. The Residual Attention Neural Network for Breast Cancer Classification (RANN-BCC) utilizes residual neural networks with attention layers to classify invasive ductal carcinoma (IDC) and non-IDC. This model achieved high accuracy, recall, precision, and F1 score, outperforming other models such as CNN and AlexNet[6].

II. PREPROCESSING

A. Data

The DDSM (Digital Database for Screening Mammography) and the CDD (Curated Breast Imaging Subset of DDSM) datasets are essential resources in breast cancer research and diagnosis.

The DDSM dataset comprises 2,620 scanned film mammography studies, including normal, benign, and malignant cases with verified pathology information. Its extensive scale and ground truth validation make it invaluable for developing and testing decision support systems. However, historical limitations such as non-standard compression files and imprecise lesion annotations have hindered direct comparisons and replication of results.

In contrast, the CDD dataset is an updated and standardized version of DDSM, curated by trained mammographers. It includes decompressed images converted to DICOM format, improved ROI segmentation, and precise pathologic diagnoses for training data. By addressing the limitations of DDSM, CDD provides a consistent evaluation platform for future CADx (computer-aided diagnosis) and CADe (computer-aided detection) research in mammography.

By integrating the diverse breast tissue variations from DDSM with the structured clinical information provided by CDD, researchers can develop sophisticated machine learning models. These models can accurately differentiate between benign and malignant lesions, leading to earlier detection and improved treatment strategies. Ultimately, this combination enhances breast cancer diagnosis and patient care.

B. Preprocessing

The original datasets contained small regions extracted from entire breast images, including some artifacts and irrelevant portions. To improve data quality, we meticulously removed these non-useful regions, ensuring that the dataset only contained relevant information for training and evaluation.

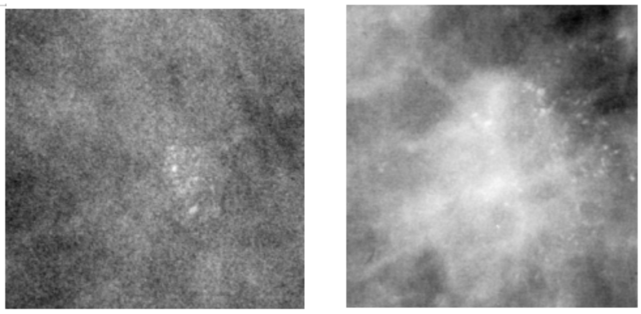

Fig. 1. Calcification images

We carefully divided the data into training, validation, and test subsets, selecting samples from both mass (benign/malignant) and calcification (benign/malignant) cases. This balanced approach allowed the model to learn from a diverse set of scenarios, ensuring robustness. Additionally, we performed augmentations such as flipping, rotation, and brightness adjustments to enhance the dataset. Images in Figure 1 depict calcification, while images in Figure 2 show masses. To enhance model generalization, various augmentations were applied to the training data. Images were horizontally or vertically flipped to create additional variations. Random rotations were introduced to simulate different orientations, and brightness adjustments were made to account for varying lighting conditions. These augmentations help the model learn from a more diverse set of scenarios, improving its robustness and performance.

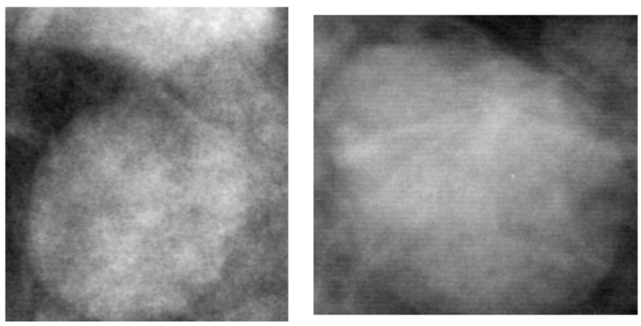

Fig. 2. Mass images

C. Augmentations

To enhance model generalization, various augmentations were applied to the training data. Images were horizontally or vertically flipped to create additional variations. Random rotations were introduced to simulate different orientations, and brightness adjustments were made to account for varying lighting conditions. These augmentations help the model learn from a more diverse set of scenarios, improving its robustness and performance.

III. METHODOLOGY

For our methodology, we began by utilizing ConvNeXt for feature extraction. ConvNeXt, known for its robust perfor-

TABLE I
DATA DISTRIBUTION

Partition	Calcification (Benign)	Mass (Benign)	Calcification (Malignant)	Mass (Malignant)	Pathology (Malignant/Benign)	Total
Train	1007	768	541	655	1775/1196	2971
Test	110	99	75	87	209/162	371
Validation	110	99	74	85	209/162	371

mance, allowed us to effectively capture and represent the essential features of our data. This initial step was crucial in ensuring that our subsequent processes were built on a strong foundation. Following the feature extraction, we employed

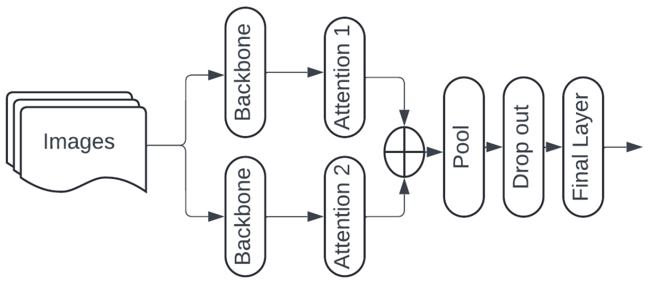

Fig. 3. Parrel Configaration

attention mechanisms to further refine and extract features in various forms. Attention mechanisms have the ability to focus on different parts of the input data, making them highly effective for our purposes. We explored different configurations by arranging two attention mechanisms both in parallel and in series. This approach enabled us to investigate how different arrangements could impact the feature extraction process.

In our experiments, we tested several types of attention mechanisms, each offering unique advantages. The first attention mechanism we tested was the Spatial Relation Module (SRM). SRM focuses on capturing spatial relationships within the data, which can be particularly useful for tasks that require an understanding of spatial context. Next, we experimented

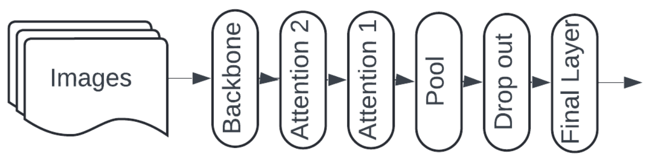

Fig. 4. Series Configaration

with the Global Contextual Transformer (GCT). GCT is designed to capture global contextual information, allowing the model to understand the broader context of the data. This can be beneficial for tasks that require a comprehensive understanding of the entire input.

We also tested the Local Contextual Transformer (LCT), which focuses on capturing local contextual information. LCT is particularly useful for tasks that require a detailed understanding of specific parts of the input data. By combining LCT with other attention mechanisms, we aimed to achieve a balance between local and global context.

Lastly, we experimented with Triplet Attention, which is designed to capture relationships between triplets of data points. This unique approach allowed us to explore how interactions between multiple data points could enhance the feature extraction process. By systematically testing these different attention mechanisms, we were able to identify the most effective configurations for our specific use case, ultimately enhancing the performance of our model.

IV. EXPERIMENT SETTING

Our current experiment was conducted on a Linux operating system, utilizing a GTX 2090 GPU to ensure efficient processing and model training. We employed Python and PyTorch as our primary programming and deep learning frameworks, respectively. These tools provided the flexibility and robustness needed to implement and test various attention mechanisms in our model.

For the optimization process, we used the Adam optimizer combined with a cosine annealing learning rate schedule. This approach helped in achieving a smooth convergence during training. The model was trained for 100 epochs with a batch size of 8, allowing us to effectively balance computational efficiency and model performance. This setup enabled us to thoroughly evaluate the impact of different attention mechanisms on our model's accuracy and F1 score.

V. METRIC

Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome. It is calculated as the ratio of correctly predicted instances (both true positives and true negatives) to the total number of instances. The formula for accuracy is:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

In binary classification, it can also be expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP: True Positives
- TN: True Negatives
- FP: False Positives
- FN : False Negatives

The **F1 score** is a metric that combines precision and recall into a single value, providing a balance between the two. It

Model	type	Accuracy	F1 Score	Loss	F1 Score (Test)
Base model	-	0.78	0.7655	0.557	0.74
Parrel	SRM_GCT	0.787	0.7736	0.5017	0.74
	tripple_gct	0.787	0.7736	**0.4812**	0.72
	tripple_lct	**0.7924**	**0.7818**	0.5583	0.75
Series	tripple_gct	0.787	0.7669	0.6313	0.75
	tripple_lct	0.7843	0.7575	0.6793	0.752
	SRM_GCT	0.7789	0.7696	0.4913	**0.758**

TABLE II
MODEL PERFORMANCE COMPARISON

is particularly useful when you need to account for both false positives and false negatives. The F1 score is the harmonic mean of precision and recall, calculated as:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

- **Precision** is the ratio of true positive predictions to the total number of positive predictions (true positives + false positives).
- **Recall** (or sensitivity) is the ratio of true positive predictions to the total number of actual positive instances (true positives + false negatives).

VI. RESULTS

The Base model achieved an accuracy of 0.78, an F1 Score of 0.7655, a loss of 0.557, and an F1 Score (Test) of 0.74. This serves as a benchmark for evaluating the performance of the other models.

In the Parallel configuration, the Triplet_LCT model stood out with the highest accuracy of 0.7924 and an F1 Score of 0.7818. However, it also had a relatively high loss of 0.5583. The SRM_GCT model in this configuration had a slightly lower accuracy of 0.787 and an F1 Score of 0.7736, but it achieved the lowest loss of 0.5017 among the parallel models. The Triplet_ACT model also performed well with an accuracy of 0.787, an F1 Score of 0.7669, and a loss of 0.6313.

In the Series configuration, the SRM_GCT model achieved an accuracy of 0.7789, an F1 Score of 0.7696, and the lowest loss of 0.4913 among all models. The Triplet_LCT model in this configuration had a higher accuracy of 0.7843 and an F1 Score of 0.7575, but it also had the highest loss of 0.6793. Overall, the parallel configuration models generally performed better in terms of accuracy and F1 Score compared to the series configuration models. The choice of attention mechanism and its arrangement significantly impacted the performance metrics, highlighting the importance of selecting the right configuration for the specific use case.

VII. CONCLUSION

In our experiments, the addition of attention mechanisms to the models demonstrated noticeable improvements in performance metrics such as accuracy and F1 score. By incorporating attention mechanisms like SRM, GCT, and LCT, we were able to enhance the model's ability to focus on relevant features, leading to better overall performance. These attention mechanisms allowed the model to capture both local and global contextual information, which is crucial for accurate feature extraction in complex datasets.

Among the various configurations tested, the combinations involving Triplet Attention yielded the most significant performance gains. The Triplet Attention mechanism, when combined with other attention mechanisms like GCT and LCT, provided a more comprehensive understanding of the relationships between data points. This resulted in higher accuracy and F1 scores compared to other configurations. The parallel configuration of Triplet_LCT, in particular, achieved the highest accuracy and F1 score, highlighting the effectiveness of this combination in improving model performance.Overall, the integration of attention mechanisms, especially in combination with Triplet Attention, proved to be a valuable approach in enhancing the performance of our models.

REFERENCES

[1] J. Hao, R. Jin, Y. Yi, *et al.*, "Development of a humanized anti-fabp4 monoclonal antibody for potential treatment of breast cancer," *Breast Cancer Research*, vol. 26, no. 1, pp. 1–15, 2024.

[2] D. Zuo, L. Yang, Y. Jin, H. Qi, Y. Liu, and L. Ren, "Machine learning-based models for the prediction of breast cancer recurrence risk," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 276, 2023.

[3] A. Khalid, A. Mehmood, A. Alabrah, *et al.*, "Breast cancer detection and prevention using machine learning," *Diagnostics*, vol. 13, no. 19, p. 3113, 2023.

[4] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis," *Designs*, vol. 2, no. 2, p. 13, 2018.

[5] L. A. Aldakhil, H. F. Alhasson, and S. S. Alharbi, "Attention-based deep learning approach for breast cancer histopathological image multi-classification," *Diagnostics*, vol. 14, no. 13, p. 1402, 2024.

[6] C. K. Toa, M. Elsayed, and K. S. Sim, "Deep residual learning with attention mechanism for breast cancer classification," *Soft Computing*, pp. 1–11, 2023.