

Bridging Knowledge Gaps in Multimodal Breast Cancer Diagnosis Using CTGA-Net

Himanish Shekhar Das¹, Subrata Chowdhury^{2,3}

¹ Postdoctoral Researcher, Lincoln University College, Petaling Jaya 47301, Selangor, Malaysia

² Lincoln University College, Petaling Jaya 47301, Selangor, Malaysia

³ Department of Computer Science Engineering, SVCET College, Chittoor, Andhra Pradesh 517127, India

Lincoln University College, Petaling Jaya 47301, Selangor, Malaysia

pdf.himanishshekhar@lincoln.edu.my , pdfsv.subrata@lincoln.edu.my

Abstract: Breast cancer is the most common cancer diagnosed in women worldwide and sentinel node biopsy for early detection has significantly improved survival. While mammography, ultrasound and histopathology have auxiliary diagnostic methodologies, current artificial intelligence (AI) proposed solutions are mostly unimodal and do not have standardized multimodal fusion strategies. In this study, we proposed a novel Multimodal CTGA-Net for comprehensive breast cancer diagnosis based on multimodality of mammogram, ultrasound, and histopathology images. It incorporates multiple components (CNNs for local feature extraction, Transformers for global dependencies modeling, BiGRU to learn context information and attention mechanisms to enhance discriminative features). We present a cross-modal attention fusion module that deals with the distribution mismatch and domain shift discovered across heterogeneous modalities. It does this by addressing several knowledge gaps, including a limitation of unimodal data and absence of a unified segmentation, classification pipelines as well as limited interpretability. We present the experimental design, architecture formulation, and potential clinical implications. The proposed method is a scalable and interpretable, AI-based decision-support system designed to be applicable in low-resource healthcare environments.

Keywords: Breast Cancer, Multimodal Learning, CNN, Transformer, BiGRU, Attention, Medical Imaging, Deep Learning

Introduction

Breast cancer is still one of the most common and serious diseases among women in the world today. Recent global health data show it contributes to quarter of all cancer in women, and continues to pose a public health problem. An early diagnosis is fundamental for better chances of survival, with rates above 90% when the disease occurs. But delayed diagnosis limits treatment options and increases the risk of mortality. Therefore, effective clinical management and better patient outcomes depend upon comprehensive and timely detection methods. Medical imaging techniques (mammography, ultrasound and histopathology) are the integral techniques in breast cancer screening and diagnosis. Mammography is the most commonly used method for large-scale screening and is capable of detecting microcalcifications and various structural abnormalities. Ultrasound imaging offers additional insights which are especially useful in the context of dense breast tissues where mammography may display limited sensitivity. The gold standard for malignancy confirmation at the cellular level is histopathological examination of biopsy samples, mainly hematoxylin and eosin (H&E) stained slides. Although these

modalities individually can provide diagnostic information, there are limitations associated with each including variance in interpretation, limited sensitivity and reliance on expert interpretations by radiology or pathology. Data-driven approaches are gaining ground in the analysis of medical images, revealing new opportunities for automating diagnostics and assisting clinical decision-making of human experts. Deep architectures such as convolutional neural networks have demonstrated strong performance on accurately identifying breast cancer patterns from individual imaging modalities. However, most available methods treat unimodal data, which restricts the opportunity to exploit complementary information among multimodal imaging sources. The use of unimodal representation together with semi-supervision limits the generalization of models and reduces diagnostic robustness. To tackle these challenges, the study presents a multimodal artificial intelligence framework to process mammogram, ultrasound and histopathology images in the same architecture. As this study was able to combine state-of-the-art deep learning paradigm with cross-modal feature fusion, it the authors expect that such an approach is likely to improve diagnostic accuracy and generalization making it a robust AI-assisted decision-support system for breast cancer diagnosis.

Literature Review

Advancements in artificial intelligence and deep learning methods have led to breakthroughs in the automated detection of breast cancer via various medical imaging modalities. Convolutional Neural Networks (CNNs) have been extensively used for mammogram-based classification as they can learn spatial features and texture patterns [8] that characterize a malignant lesion. For instance, Jabeen et al. proposed an advanced deep learning-based framework which included feature selection strategies to improve classification precision on mammographic images [1]. Similarly, Chugh et al. performed a comparative analysis that proves the transfer learning models achieved superior diagnostic accuracy than conventional machine learning methods in breast carcinoma diagnosis [3]. Deep learning architectures have been investigated with respect to combating speckle noise in ultrasound imaging and variabilities associated with image acquisition [3]. Ayana et al. proposed a multi-stage transfer learning method resulted improved ultrasound-based breast tumor classification [4]. Furthermore, to diversify training data and improve robustness of models, various methods that augment using GANs have been proposed [2]. So has deep learning-based frameworks for histopathological image analysis. Deep learning-based classification with CNN to predict malignant tissue pattern from H&E-stained biopsy image has shown good performance [5]. Hybrid architectures such as CNN-LSTM have also improved classification accuracy in histopathology images, by capturing spatial-temporal dependencies [6]. These approaches show high performance for each modality, while most of the existing studies are still unimodal. To this end, more recent work emphasises the need for multimodal frameworks, as well as explainable AI methods to improve diagnosis reliability across different imaging sources [7], [8].

Identified Knowledge Gaps

Although there have been great strides in AI-based breast cancer detection, several limitations exist in current research. Identically, existing models are limited as they consider single imaging modality and unable to leverage complementary diagnostic information available from mammography, ultrasound and histopathology images. These unimodal designs commonly lead to low generalisability and lack of robustness when used with multi-modal clinical datasets. Moreover, several methods typically consider

segmentation and classification as two separate tasks causing diagnostic pipelines to be existed alongside. A further relevant limitation rests in the lack of standardized multimodal fusion strategies that can efficiently incorporate heterogeneous imaging features. Additionally, explainability and interpretability in clinical practice is often ignored leading to limited adoption by physicians in everyday medical decision-making.

Proposed Methodology

In this study, we propose a Multimodal CNN-Transformer-BiGRU with Attention Network (CTGA-Net) method for breast cancer diagnosis that fuses a multimodality set of images (mammography, ultrasound, and histopathology images). The goal is to acquire complementary diagnostic information from heterogeneous image sources, and bridge these imaging sources through a unified multi-input framework allowing joint learning of local, global and contextual features. The proposed architecture is represented in Figure 1.

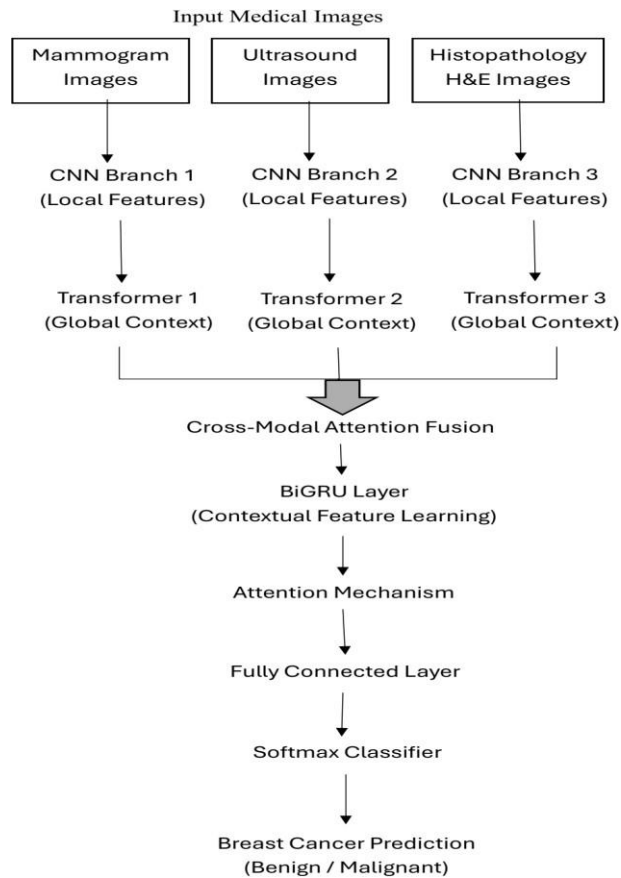


Fig. 1. Proposed Multimodal Breast Cancer Detection Architecture

1. Data Preprocessing

The input data includes images from three modalities: namely mammogram, ultrasound and hematoxylin and eosin (H&E) stained histopathology images. Owing to the differences in resolution, channel characteristics and feature distributions between these datasets a preprocessing pipeline is being used. These include image processing steps like resizing, normalization, and noise reduction; as well as

augmentation processes like rotating, flipping or scaling the images to help with generalization in the models. Images of mammograms are converted into a canonical form to eliminate modality mismatch.

2. Modality-Specific Feature Extraction (CNN)

The input data for each modality is a separate CNN branch to extract low/mid-level morphological features (including tumor boundary, texture, structural patterns). Convolution, batch normalization and pooling operations are applied in order to learn features from spatial domains. Let X_m represent the input image from modality m . The CNN extracts local features as:

$$F_m = CNN_m(X_m)$$

3. Global Feature Modeling (Transformer Encoder)

These CNN features are then fed to a Transformer encoder, which encodes long-range dependencies and global contextual relationships across the image. Using self-attention, transformers are able to determine relationships between spatial locations in a manner that is especially relevant for detecting tumors with complex morphologies.

$$G_m = Transformer(F_m)$$

4. Cross-Modal Feature Fusion

Once the modality-specific global features are extracted, the modalities are integrated through a cross-modal attention fusion module. This module can adaptively weight for each modality and relieve domain shifts owing to the differences in image format as well as statistical distributions.

$$F_{fusion} = Attention(G_1, G_2, G_3)$$

5. Contextual Learning (BiGRU)

Feature representation of pairs (after fusion) are fed through a Bidirectional Gated Recurrent Unit network with stacked cell layers, to capture contextual dependencies and relationships between sequential features. BiGRU utilizes information in both forward and backward directions, which allows the model to extract more informative and rich feature interactions.

6. Attention-Based Classification

Lastly, an attention layer focuses on features pertinent to the disease and subsequently feeds the representation into a fully connected with Softmax classifier for benign or malignant prediction.

Conclusion

This paper has provided an integrated CTGA-Net for accurate breast cancer diagnosis. The framework bridges significant research gaps in multimodal medical AI by incorporating local morphological features with global contextual dependencies and sequential relationships across mammogram, ultrasound, and histopathology images.

References

1. Jabeen K et al (2023) BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection. *Diagnostics*. <https://doi.org/10.3390/diagnostics13071238>
2. Jiménez-Gaona Y et al (2024) Gan-based data augmentation to improve breast ultrasound and mammography mass classification. *Biomed Signal Process Control* 94:106255
3. Melekoodappattu JG, Dhas AS, Kandathil BK, Adarsh KS (2023) Breast cancer detection in mammogram: combining modified CNN and texture feature-based approach. *J Ambient Intell Humaniz Comput* 14(9):11397–11406. <https://doi.org/10.1007/s12652-022-03713-3>
4. Chugh G, Kumar S, Singh N (2024) TransNet: a comparative study on breast carcinoma diagnosis with classical machine learning and transfer learning paradigm. *Multimed Tools Appl*. 83(11):33855–33877. <https://doi.org/10.1007/s11042-023-16938-x>
5. Ayana G, Park J, Jeong JW, Choe SW (2022) A novel multistage transfer learning for ultrasound breast cancer image classification. *Diagnostics* 12(1):135. <https://doi.org/10.3390/diagnostics12010135>
6. Chen S-H, Wu Y-L, Pan C-Y, Lian L-Y, Su Q-C (Sep.2023) Breast ultrasound image classification and physiological assessment based on GoogLeNet. *J Radiat Res Appl Sci*. 16(3). <https://doi.org/10.1016/j.jrras.2023.100628>
7. Alruily M et al (2023) Breast ultrasound images augmentation and segmentation using GAN with identity block and modified U-Net 3+. *Sensors* 23(20):8599
8. Wu L et al (2024) Ultrasound-based deep learning radiomics nomogram for differentiating mass mastitis from invasive breast cancer. *BMC Med Imaging* 24(1):189
9. Yamlome P, Akwaboah AD, Marz A, Deo M, Convolutional Neural Network Based Breast Cancer Histopathology Image Classification, in (2020) 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, Jul. 2020, pp. 1144–1147. <https://doi.org/10.1109/EMBC44109.2020.9176594>