

Attention-Based Multimodal Fusion Model for Breast Cancer Diagnostics

Anni Zheng¹, Wei Qi Yan²

¹School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, 1010, New Zealand. Email: byc1902@autuni.ac.nz

²School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, 1010, New Zealand. Email: weiqi.yan@aut.ac.nz

Summary

Computer-assisted breast cancer diagnosis has emerged as a promising approach to enhance the accuracy and efficiency of breast cancer classification. However, the question of how to effectively utilise both magnetic resonance imaging (MRI) and electronic health record (EHR) data in the model to enhance prediction accuracy remains unanswered. In this paper, we present a new attention-based multimodal model for breast cancer label classification. In the proposed model, inspired by attention blocks in transformer architecture, we innovatively adapt EHR-guided attention at mid and late stage of modality fusion highlighting the complementary strength of different modality data. We compare performance on two breast cancer datasets with other fusion methods and multimodal models. The experimental result shows that our model achieved better accuracy on both datasets and has potential to assist real-world clinical decisions.

Background

Breast cancer is a devastating disease with a profound global impact. In clinical practices, doctors made diagnosis decisions based on risk factors combined with mammography results such as Magnetic Resonance Imaging (MRI). With the development of digital image processing methods and deep learning, there're a lot of promising improvements in diagnosis accuracy [1,2]. When applying deep learning models to these tasks, past work typically only includes single modality data or late-stage fusion of modalities [3,4]. The potential of cross-modal information is not efficiently leveraged in these models. Building upon previous research in medical image processing and fusion methods, this paper focuses on a multimodal fusion model that enhances diagnostic accuracy on breast cancer datasets.

Methodology

In this research project, we present a new multimodal model for breast cancer risk prediction. To learn the representation for each MRI image in the encoder layer, we use self-supervised contrastive learning methods (SimCLR) [5]. Multi-Instance Contrastive Learning (MICLe) [6] is applied in the embedding layer to maximize the mutual information between MRI images from the same patient. In the fusion layer, we make use of Multimodal Adaptation Gate (MAG) [7] to learn the attention weights for each modality. In the middle fusion stage, the EHR-guided Transformer using architecture of Attention Bottleneck [8]. For the architecture details, Figure 1 shows the pseudocode of the proposed model of this paper. For our experiments, we chose the two open dataset. Duke-Breast dataset [9] contains 922 samples of clinical data for retrospective breast cancers including MRI images and non-image EHR data. The MRI data in this dataset has both CC and MLO view for each patient. The non-image data includes

demographic, clinical, pathology, genomic, treatment clinical records. VinDr Mammo Dataset [10] includes 5,000 mammography exams and 20,000 samples of mammography results. This dataset also compasses Breast Imaging Reporting and Data System (BIRADS) data for each patient. In the model training, we use 10-fold validation to get average results as an evaluation matrix. The performance metric is the area under the receiver operating characteristic curve (AUROC).

Algorithm 1: Proposed Model Algorithm

```

Input: A sequence of MRI image matrix  $\{\mathbf{L}_{cc}, \mathbf{L}_{mlo}, \mathbf{R}_{cc}, \mathbf{R}_{mlo}\}$ 
        EHR data as non-image matrix  $\mathbf{I}$ , A label matrix  $\mathbf{Y}$ , Maximum epoch
        of image encoder training  $MaxEpoch_1$ , Maximum epoch of multimodal model
        training  $MaxEpoch_2$ , Image embedding  $e$ , multimodal model  $f$ .
1 for  $I_{image} \in \{\mathbf{L}_{cc}, \mathbf{L}_{mlo}, \mathbf{R}_{cc}, \mathbf{R}_{mlo}\}$  do
2   while not  $MaxEpoch_1$  do
3      $\mathcal{L}^{Image} \leftarrow \text{run}_{image}(\mathbf{I}_{image})$ 
4     Update  $e$  by backpropagating  $\mathcal{L}^{Image}$ 
5   end
6 end
7 Function  $\text{run}_{image}(\mathbf{I}_{image})$ :
8    $\mathbf{M} \leftarrow e(\mathbf{I}_{image})$ 
9    $\mathcal{L}^{Image} \leftarrow \sum_{i=1}^{|V|} \mathcal{L}_i^{MICLe}$ 
10  return  $\mathcal{L}^{Image}$ 
11 end
12  $\mathbf{M}_l \leftarrow \text{Concatenate}(e(\mathbf{L}_{cc}), e(\mathbf{L}_{mlo}))$ 
13  $\mathbf{M}_r \leftarrow \text{Concatenate}(e(\mathbf{R}_{cc}), e(\mathbf{R}_{mlo}))$ 
14  $\mathbf{E} \leftarrow \text{ReLU}(\text{Linear}(\mathbf{I}))$ 
    // EHR-guided Attention Bottleneck
15  $\mathbf{M}_l \leftarrow \text{EHRRGuidedAttention}(\mathbf{M}_l, \mathbf{E})$ 
16  $\mathbf{M}_r \leftarrow \text{EHRRGuidedAttention}(\mathbf{M}_r, \mathbf{E})$ 
    // Multimodal Fusion with Attention Gate
17 while not  $MaxEpoch_f$  do
18    $\mathcal{L}^{\text{Softmax}} \leftarrow \text{run}_{\text{AttentionGate}}(\mathbf{M}_l, \mathbf{M}_r, \mathbf{E})$ 
19   Update  $f$  by backpropagating  $\mathcal{L}^{\text{Softmax}}$ 
20 end
21 Function  $\text{run}_{\text{AttentionGate}}(\mathbf{M}_l, \mathbf{M}_r, \mathbf{E})$ :
22    $g_1 \leftarrow \text{ReLU}(\text{Linear}([\mathbf{E}; \mathbf{M}_l]))$ 
23    $g_2 \leftarrow \text{ReLU}(\text{Linear}([\mathbf{E}; \mathbf{M}_r]))$ 
24    $\mathbf{H} \leftarrow \text{Linear}([g_1 \mathbf{M}_l; g_2 \mathbf{M}_r])$ 
25    $\mathbf{R} \leftarrow \mathbf{E} + \min(\frac{\|\mathbf{E}\|_2}{\|\mathbf{H}\|_2} \theta, 1) \mathbf{H}$ 
26    $\hat{\mathbf{Y}} \leftarrow \text{softmax}(\text{Linear}(\mathbf{M}))$ 
27    $\mathcal{L}^{\text{Softmax}} \leftarrow -\frac{1}{B} \sum_{i=1}^B \mathbf{Y}_i \log(\hat{\mathbf{Y}}_i) + (1 - \mathbf{Y}_i) \log(1 - \hat{\mathbf{Y}}_i)$ 
28   return  $\mathcal{L}^{\text{Softmax}}$ 
29 end
30 Function  $\text{EHRRGuidedAttention}(\mathbf{M}, \mathbf{E})$ :
31    $\mathbf{Q} \leftarrow \text{Linear}(\mathbf{E})$ 
32    $\mathbf{K} \leftarrow \text{Linear}(\mathbf{M})$ 
33    $\mathbf{V} \leftarrow \mathbf{M}$ 
34    $\mathbf{A} \leftarrow \text{softmax}(\frac{\mathbf{QK}^T}{\sqrt{d_k}})$ 
35    $\mathbf{M}_{out} \leftarrow \mathbf{AV}$ 
36   return  $\mathbf{M}_{out}$ 
37 end
    
```

Fig. 1. The proposed model algorithm

Results and Observations

In the experiments, we compare the model performance on Duke-Breast dataset and VinDr Mammo Dataset. The baseline models include the image-only model, the early fusion method, the late fusion approach and two recent LLM multimodal classification models which are pretrained on medical dataset. As result shows in Table 1, our model is more effective than other early and late fusion methods in multimodal breast cancer diagnosis.

Table 1. Experiment result

Model	Duke Dataset [9]	VinDr Dataset [10]
Image-only: SimCLR	61.87%	69.31%
Early fusion	61.21%	68.66%
Late fusion	65.57%	69.24%
Med-BERT [11]	65.50%	70.91%
Med-PaLM [12]	67.29%	71.76%
Proposed Model	69.71%	75.10%

Conclusion

In this paper, we proposed a novel model for breast cancer prediction and evaluated its performance against several baseline models. The results demonstrated that our model significantly outperforms the baseline models on both the Duke Breast Cancer and Vindr Mammo dataset. Our analysis also highlighted the importance of modality fusion methods. The proposed model is a promising tool for breast cancer prediction, with potential for further improvements through enhanced training data, fusion methods, and integration with existing clinical risk models.

References

1. Kwon, B.C., Choi, M.J., Kim, J.T., Choi, E., Kim, Y.B., Kwon, S., Sun, J. and Choo, J., 2018. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE transactions on visualization and computer graphics*, 25(1), pp.299-309. Available at: <https://dl.acm.org/doi/10.1109/TVCG.2018.2865027>
2. Zubair, M., Wang, S. and Ali, N., 2021. Advanced approaches to breast cancer classification and diagnosis. *Frontiers in Pharmacology*, 11, p.632079. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7952319/>
3. Morais, M., Calisto, F.M., Santiago, C., Aleluia, C. and Nascimento, J.C., 2023, April. Classification of breast cancer in Mri with multimodal fusion. In 2023 IEEE 20th international symposium on biomedical imaging (ISBI) (pp. 1-4). IEEE. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10230686>
4. Kline, A. et al. (2022) "Multimodal machine learning in precision health: A scoping review," *NPJ digital medicine*, 5(1), p. 171. Available at: <https://doi.org/10.1038/s41746-022-00712-8>
5. Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020, November. A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR. Available at: <https://dl.acm.org/doi/abs/10.5555/3524938.3525087>
6. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T. and Natarajan, V., 2021. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3478-3488). Available at: https://openaccess.thecvf.com/content/ICCV2021/papers/Azizi_Big_Self-Supervised_Models_Advance_Medical_Image_Classification_ICCV_2021_paper.pdf
7. Rahman, W., Hasan, M.K., Lee, S., Zadeh, A., Mao, C., Morency, L.P. and Hoque, E., 2020, July. Integrating multimodal information in large pretrained transformers. In *Proceedings of the*

- conference. Association for Computational Linguistics. Meeting (Vol. 2020, p. 2359). Available at: <https://aclanthology.org/2020.acl-main.214/>
8. Nagrani, A., Yang, S., Arnab, A., Jansen, A., Schmid, C. and Sun, C., 2021. Attention bottlenecks for multimodal fusion. *Advances in neural information processing systems*, 34, pp.14200-14213. Available at: <https://proceedings.neurips.cc/paper/2021/hash/76ba9f564ebbc35b1014ac498fafadd0-Abstract.html>
 9. Saha, A., Harowicz, M.R., Grimm, L.J., Kim, C.E., Ghate, S.V., Walsh, R. and Mazurowski, M.A., 2018. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. *British journal of cancer*, 119(4), pp.508-516. Available at: <https://pubmed.ncbi.nlm.nih.gov/30033447/>
 10. Nguyen, H.T., Nguyen, H.Q., Pham, H.H., Lam, K., Le, L.T., Dao, M. and Vu, V., 2023. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Scientific Data*, 10(1), p.277. Available at: <https://www.nature.com/articles/s41597-023-02100-7>
 11. Rasmy, L., Xiang, Y., Xie, Z., Tao, C. and Zhi, D., 2021. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, 4(1), p.86. Available at: <https://www.nature.com/articles/s41746-021-00455-y>
 12. Tu, T., Azizi, S., Driess, D., Schaeckermann, M., Amin, M., Chang, P.C., Carroll, A., Lau, C., Tanno, R., Ktena, I. and Palepu, A., 2024. Towards generalist biomedical AI. *NEJM AI*, 1(3), p.Aloa2300138. Available at: <https://ai.nejm.org/doi/full/10.1056/Aloa2300138>