Peer-Reviewed Structured Abstract

Attention-Based Multimodal Fusion Model For Breast Cancer Diagnostics

Anni Zheng, Wei Qi Yan
Auckland University of Technology, Auckland, 1010, New Zealand.

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 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 and deep learning methods, there're a lot of promising improvements in diagnosis accuracy [1,2]. While 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 work in medical image processing and fusion methods, the focus of this paper is on a multimodal fusion model that enhances diagnostic accuracy on breast cancer datasets.

Methodology

In this research paper, we present a new multimodal model for breast cancer risk prediction. To learn the representation for each MRI image in the encoder layer, we make use of 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 take advantage of Multimodal Adaptation Gate (MAG) [7] to learn the attention weights for each modality. In the middle fusion stage, the EHR-guided Transformer was conducted by using architecture of Attention Bottleneck [8]. With regard to the architecture details, Figure 1 shows the pseudocode of the proposed model of this paper. Regarding our experiments, we chose two open datasets. 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] capsulates 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 are use of 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.

for $I_{image} in \{\mathbf{L}_{cc}, \mathbf{L}_{mlo}, \mathbf{R}_{cc}, \mathbf{R}_{mlo}\}$ **do**
  while *not* $MaxEpoch_1$ **do**
    $\mathcal{L}^{\text{Image}} \leftarrow \mathtt{run}_{image}(\mathbf{I}_{image})$
    Update e by backpropagating $\mathcal{L}^{Image}$
  end
end
Function $\mathtt{run}_{image}(\mathbf{I}_{image})$**:**
  $\mathbf{M} \leftarrow e(\mathbf{I}_{image})$
  $\mathcal{L}^{\text{Image}} \leftarrow \sum_{i=1}^{|V|} \mathcal{L}_i^{\text{MICLe}}$
  return $\mathcal{L}^{\text{Image}}$
end
$\mathbf{M}_l \leftarrow Concatenate(e(\mathbf{L}_{cc}), e(\mathbf{L}_{mlo}))$
$\mathbf{M}_r \leftarrow Concatenate(e(\mathbf{R}_{cc}), e(\mathbf{R}_{mlo}))$
$\mathbf{E} \leftarrow \mathtt{ReLU}(\mathtt{Linear}(\mathbf{I}))$
// EHR-guided Attention Bottleneck
$\mathbf{M}_l \leftarrow EHRGuidedAttention(\mathbf{M}_l, \mathbf{E})$
$\mathbf{M}_r \leftarrow EHRGuidedAttention(\mathbf{M}_r, \mathbf{E})$
// Multimodal Fusion with Attention Gate
while *not* $MaxEpoch_f$ **do**
  $\mathcal{L}^{\text{Softmax}} \leftarrow \mathtt{run}_{AttentionGate}(\mathbf{M}_l, \mathbf{M}_r, \mathbf{E})$
  Update f by backpropagating $\mathcal{L}^{\text{Softmax}}$
end
Function $\mathtt{run}_{AttentionGate}(\mathbf{M}_l, \mathbf{M}_r, \mathbf{E})$**:**
  $g_1 \leftarrow \mathtt{ReLU}(\mathtt{Linear}([\mathbf{E}; \mathbf{M}_l]))$
  $g_2 \leftarrow \mathtt{ReLU}(\mathtt{Linear}([\mathbf{E}; \mathbf{M}_r]))$
  $\mathbf{H} \leftarrow \mathtt{Linear}([g_1\mathbf{M}_l; g_2\mathbf{M}_r])$
  $\mathbf{R} \leftarrow \mathbf{E} + \mathtt{min}(\frac{||\mathbf{E}||_2}{||\mathbf{H}||_2}\theta, 1)\mathbf{H}$
  $\hat{\mathbf{Y}} \leftarrow \mathtt{softmax}(\mathtt{Linear}(\mathbf{M}))$
  $\mathcal{L}^{\text{Softmax}} \leftarrow -\frac{1}{B}\sum_{i=1}^{B} \mathbf{Y}_i \mathtt{log}(\hat{\mathbf{Y}}_i) + (1 - \mathbf{Y}_i)\mathtt{log}(1 - \hat{\mathbf{Y}}_i)$
  return $\mathcal{L}^{\text{Softmax}}$
end
Function $EHRGuidedAttention(\mathbf{M}, \mathbf{E})$**:**
  $\mathbf{Q} \leftarrow \mathtt{Linear}(\mathbf{E})$
  $\mathbf{K} \leftarrow \mathtt{Linear}(\mathbf{M})$
  $\mathbf{V} \leftarrow \mathbf{M}$
  $\mathbf{A} \leftarrow \mathtt{softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}})$
  $\mathbf{M}_{out} \leftarrow \mathbf{A}\mathbf{V}$
  return $\mathbf{M}_{out}$
end

Figure 1 The proposed algorithm

Results

In our experiments, we compare the model performance on Duke-Breast dataset and VinDr Mammo Dataset. The baseline includes 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 the result shows in Table 1, our model is more effective than others in multimodal breast cancer diagnosis.

Table 1 Our experimental results

Model	DukeDataset [9]	VinDrDataset [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%
Our 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 the existing clinical risk models.

References

1. Kwon, B.C., Choi, M.J., et al, 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.
2. Zubair, M., Wang, S. and Ali, N., 2021. Advanced approaches to breast cancer classification and diagnosis. Frontiers in Pharmacology, 11, p.632079.
3. Morais, M., Calisto, F.M., et al. 2023, Classification of breast cancer in Mri with multimodal fusion. In IEEE International Symposium on Biomedical Imaging (ISBI) (pp. 1-4). IEEE.
4. Kline, A. et al. (2022) Multimodal machine learning in precision health: A scoping review. NPJ Digital Medicine, 5(1), p. 171.
5. Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In International Conference on Machine Learning (pp. 1597-1607). PMLR.
6. Azizi, S., Mustafa, B., et al, 2021. Big self-supervised models advance medical image classification. In Proceedings of the IEEE/CVF ICCV (pp. 3478-3488).

7. Rahman, W., Hasan, M.K., et al. 2020. Integrating multimodal information in large pretrained transformers. In Proc Conf Assoc Comput Linguist Meet. (Vol. 2020, p. 2359).
8. Nagrani, A., Yang, S., et al. 2021. Attention bottlenecks for multimodal fusion. Advances in Neural Information Processing Systems, 34, pp.14200-14213.
9. Saha, A., Harowicz, M.R., et al. 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.
10. Nguyen, H.T., Nguyen, H.Q., et al 2023. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. Scientific Data, 10(1), p.277.
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.
12. Tu, T., Azizi, S., et al, 2024. Towards generalist biomedical AI. NEJM AI, 1(3), p.AIoa2300138.
13. Yan, W. 2023. Computational Methods for Deep Learning: Theory, Algorithms, and Implementations (2nd ed.), Springer.