Integration of Multimodal Data for Breast Cancer Classification Using a Hybrid Deep Learning Method

Rui Yan[1,2], Fei Ren[2], Xiaosong Rao[5], Baorong Shi[3], Tiange Xiang[4], Lingling Zhang[5], Yudong Liu[2], Jun Liang[5※], Chunhou Zheng[1※], Fa Zhang[2※]

[1]College of Computer Science and Technology, Anhui University, Hefei, China
[2]High Performance Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[3]School of Computer Science, Wuhan University, Wuhan, China
[4]School of Computer Science, The University of Sydney, Sydney, Australia
[5]Department of Pathology, Peking University International Hospital, Beijing, China
liangjun1959@aliyun.com, zhengch99@126.com, zhangfa@ict.ac.cn

Abstract. Although the application of deep learning has greatly improved the performance of benign and malignant breast cancer classification algorithm, the accuracy of classification using only the pathological image has been unable to meet the requirements of clinical practice. Inspired by the real scene when the pathologist read the pathological image for diagnosis, in this paper, we propose a new hybrid deep learning method for benign and malignant breast cancer classification. From the perspective of multimodal data fusion, our method combines pathological image and structured data in the clinical electronic medical record (EMR) to further improve the accuracy of breast cancer classification. Thus, the proposed method can be useful for breast cancer diagnosis in real clinical practice. Experimental results based on our datasets show that the proposed method significantly outperforms the state-of-the-art methods in terms of overall classification accuracy.

Keywords: Breast Cancer Classification, Pathological Image, Electronic Medical Record, Deep Learning, Multimodal Data Fusion.

※Corresponding author.

1 Introduction

Nowadays, even with the rapid advances in medical sciences, the analysis of pathological image remains the most widely used method for breast cancer diagnosis. However, the complexity of histopathological images and the dramatic increase in workload make this task time consuming, and the results may be subject to pathologist subjectivity. Facing this problem, the development of automatic and precise diagnosis methods is challenging but also essential for the field [1].

Recently, deep learning methods have made great progress and achieved remarkable performance in the field of computer vision and image processing [2]. This has also inspired many scholars to apply this technique to pathological image analysis. In spite of this, the accuracy of the benign and malignant classification of breast cancer using only the pathological image data of single mode cannot be improved to meet the requirements of clinical practice [3].

Although it is not possible to obtain high accuracy only by using pathological image, pathological image provides a rich environment to integrate data from EMR, making novel information accessible and quantifiable. In particular, the raw pathological images are highly dimensional information. It requires less human labor to obtain but it contains a large amount of potentially undiscovered information. The clinical information extracted by clinicians from EMR have fewer feature dimensions, but they usually provide more instructional information for diagnosis.

Therefore, we proposed a fusion method to mimic diagnosis tasks in clinical practice. From the perspective of multimodal data fusion, we try to combine pathological image and structured data in EMR to further improve the accuracy of breast cancer diagnosis. This is also consistent with the pathologist's actual scenario of reading pathological images for diagnosis. When reading pathological images, pathologists will repeatedly refer to the relevant information in patients' EMR as a priori until the final diagnosis is made.

There is almost no literature that classifies breast cancer using multimodal data, but the approach of multimodal fusion in other areas of medicine (text, images, genomics) has yielded good results. Although their fusion method has achieved good results than traditional methods, it still has some problems, such as the feature representation of image is not rich enough, information fusion is insufficient, especially the loss of high-dimensional information before data fusion is not addressed.

In this paper, we proposed a hybrid deep neural network to integrate multimodal

data for breast cancer classification. The main contributions of our work are as follows:
(1) To the best of our knowledge, this is by far the first time integrate multimodal data to diagnose breast cancer, and the multimodal network significantly outperforms methods using any single source of information alone.
(2) In order to make pathological image can be integrated more sufficient with structured data in EMR, we proposed a method to extract richer feature representation of the pathological image from multiple convolutional layers.
(3) In order not to lose the information of each mode before data fusion, we use the method of low-dimensional data amplification instead of reducing the high-dimensional data to the low-dimensional data before data fusion.

2 Related work

Multimodal data fusion: Recently, deep learning has demonstrated excellent performance in the medical imaging field such as pathological image classification. Bayramoglu et al. [4] proposed a magnification-independent deep learning method for the breast cancer pathological image classification task, with a classification accuracy of approximately 83%. However, such classification accuracy is not enough to be used in clinical practice. Inspired by the actual situation of the pathologist in diagnosis, the method of multimodal data fusion provides a good opportunity. Moreover, many research results show that the performance of multimodal fusion is better than that of single modal. Although the fusion of multiple modes has achieved good results, each modality of multi-modal objects has different characters with each other, leading to the complexity of heterogeneous data. Therefore, heterogeneous data poses another challenge in multimodal deep learning methods.

Richer feature representation: The precondition of multi-mode fusion is feature extraction of single-mode data. The current deep learning method enables to learn very good feature representation from some unstructured data, such as pathological images. This also makes the deep learning method achieve good results in tasks such as classification, detection and segmentation.

However, different tasks have different characteristics, which makes it necessary to adjust the required feature representation according to specific tasks. For example, in the field of semantic segmentation and edge detection, more multi-level features have a better impact on the final result. Xie et al. [5] proposed a method, holistically-nested edge detection, to automatically learns rich hierarchical representations that are important in order to resolve the challenging ambiguity in object boundary detection.

For the multimodal fusion task, the key factor is whether the fusion between different modes is sufficient or not. Therefore, it is necessary for each mode to learn rich enough feature representation before fusion, so as to provide a fertile environment for full multi-mode fusion. However, the current classification method using original deep learning (such as CNN) does not enable each mode (such as pathological image) to learn enough rich feature representation.

High-dimensional and low-dimensional data fusion: High-dimensional data (unstructured data) generally have high dimensions of feature representation. In contrast, the dimension of structured data is inherently low. How to solve the problem of fusing high-dimensional data and low-dimensional data will have a significant impact on the final result of fusion.

According to the level of fusion, information fusion technology can be divided into three categories: data-level, feature-level and decision-level fusion [6]. Generally speaking, the larger the amount of information, the smaller the information loss, the more sufficient the information fusion, and the better the final fusion result. From the existing studies, the accuracy of this viewpoint is also demonstrated, especially that the feature-level fusion achieves better results than the decision-level fusion. This indicates that the information of each mode, especially the high-dimensional information, should be kept as complete as possible before fusion, and then reduced to the required level after fusion. So, they are enough effective to capture the complex correlations over different modalities for heterogeneous data.

At present, most data fusion is image and image or image and text, and these data are all of high dimension, so fusion is relatively simple. Zhang et al. [7] introduce the semantic knowledge of medical images from diagnostic reports to provide an inspirational network training and an interpretable prediction mechanism with their proposed novel multimodal neural network, namely TandemNet.

For the fusion of low-dimensional structured data and high-dimensional unstructured data, there was little early work involved. Xu et al. [8] first reduced the high-dimensional image data to low-dimensional, and then merged with the low-dimensional structured data. This approach has yielded good results. However, in this way, a lot of information is lost before the fusion, making the fusion insufficient.

3 Method

In this section, we describe our proposed method for breast cancer classification. For simplicity, we first present an overview of our method framework. And then introduce our innovation points from two aspects: richer feature representation and high-dimensional and low-dimensional data fusion, respectively.

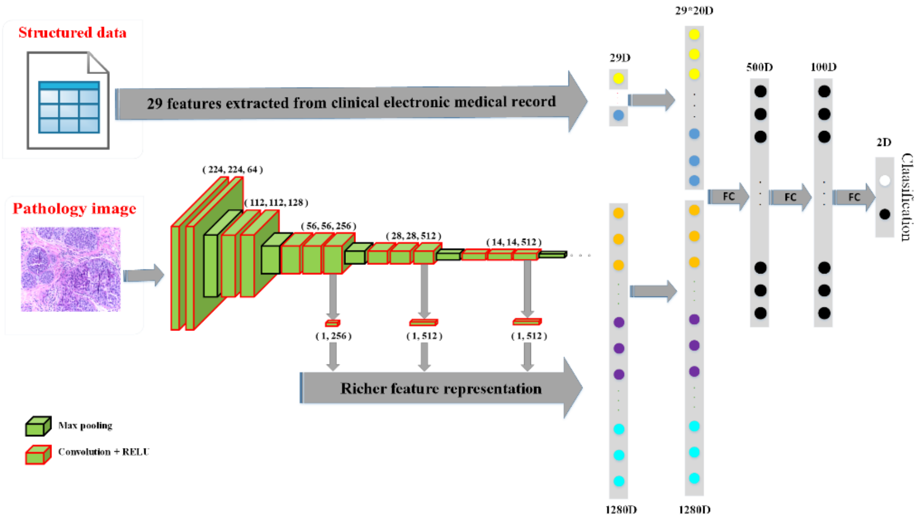

Fig. 1. The framework of our proposed method. (1) In terms of structured data, 29 representative features were extracted from EMR, which are closely related to the diagnosis of breast cancer. We copy this 29-dimensional vector 20 times on average (29D*20); (2) In terms of pathological image, the third, fourth and fifth convolution layers were extracted from VGG16 network (1280-dimensional) as richer feature representation; (3) Finally, the vector of 29D*20 dimensions extracted from the structured data was concatenated with the vector of 1280D dimensions extracted from the pathological images to form a vector of 1860D. This vector then goes through the next three full connective layers to get a classification result.

3.1 Richer feature representation

Since objects in pathological images possess various scales and high complexity, learning the rich hierarchical representations is very critical for fusion of multimodal data. CNN has been proved to be effective for this task. In addition, the convolutional features in CNN gradually become coarser with the increase of the convolutional layers. Inspired by these observations, we attempt to use richer convolutional features in such a challenging fusion task. Richer convolutional features provide richer representations compared to features just extracted from the final fully connected layer. Because multi-level convolutional layer retained complementary information such as local textures and fine details lost by higher levels.

We extract the third, fourth and fifth feature map of VGG16 network, and then use average pooling to compress the original 56*56*256,28*28*512 and 14*14*512 into 1*256,1*512 and 1*512. Then concatenate the three vectors into a 1280 (512+512+256) dimension vector, which was used as the richer feature representation of the pathological image. The specific fusion process is shown in Fig. 1.

3.2 High-dimensional and low-dimensional data fusion

After extracting the richer feature representation, we can combine the data of different modes. Compared with the 1280-dimension feature representation of the pathological image, there are only 29 representative features extracted from the EMR, namely a vector of 29 dimensions. If we integrated them directly, the vector of 29 dimensions would be completely overwhelmed by the vector of 1280 dimensions. The previous method is to represent the features of high-dimensional image data as dimensionality reduction. However, in this way, a large amount of information has been lost before the fusion of different modes, making the information fusion insufficient.

Instead, we copy the lower dimensional vector by a certain ratio, so that it is on the same order of magnitude as the higher dimensional data. In particular, by experimenting without duplicating and duplicating 10,15,20,25, and 30 times respectively, we found that the best results were obtained by making 20 copies of a vector with 29 dimensions. Then it was concatenated with the vector of 1280 dimensions extracted from the pathological images to form a vector of 1860. This vector then goes through the next three full connective layers to get a classification result.

4 Dataset

In this work, we collected a new dataset with pathological images and pairwise multiple types of features extracted from EMR for breast cancer classification.

Pathological image: We collected the medical records of 185 breast cancer patients (82 benign, 103 malignant), and for each patient we selectively cropped 2-10 representative image areas from WSI (whole slide image). In the end, we collected a total of 3764 high resolution (2048×1536 pixels) H&E stained pathological images (1332 benign, 2432 malignant). Each image is labeled as benign or malignant according to the main cancer type in each image. Figure 2 shows the example of pathological images in the dataset and summarizes the image distribution.

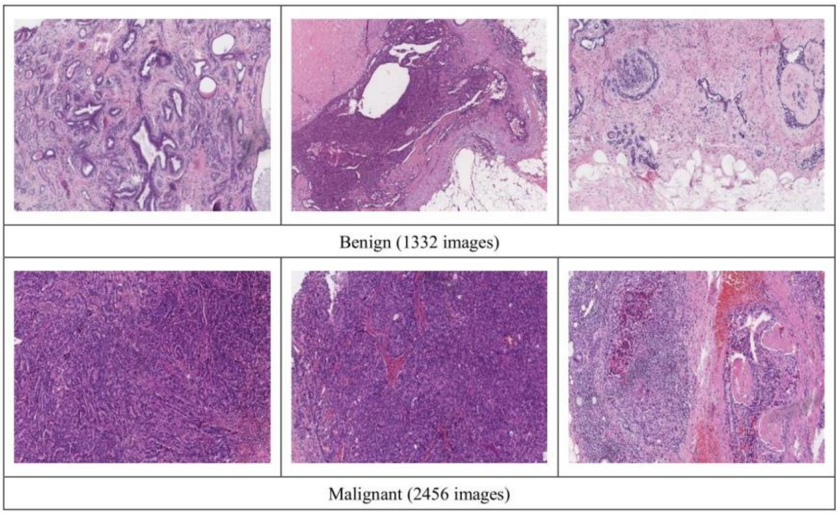

Fig. 2. Examples of pathological images in our collected dataset.

Structured data in EMR: After consulting with pathologists, 29 representative features were extracted from EMR, which are closely related to the diagnosis of breast cancer, and used as structured data to represent the clinical condition description of a patient. Specifically, these 29 features include Age, Gender, Disease Course Type, Pectoral Muscle Adhesion, Personal Tumor History, Family Tumor History, Prophase Treatment, Neoadjuvant Chemotherapy, Dimple Sign, Orange Peel Appearance, Redness And Swelling Of Skin, Skin Ulcers, Tumor, Breast Deformation, Nipple Change, Nipple Discharge, Axillary Lymphadenectasis, Swelling of Lymph Nodes, Tumor Position, Tumor Number, Tumor Size, Tumor Texture, Tumor Border, Smooth Surface, Tumor Morphology, Activity, Capsules, Tenderness and Skin Adhesion.

5 Experiments

In this section, we present the performance of our proposed algorithm on our collected dataset. The 80% of the datasets are randomly selected to train the model and the remaining 20% to test. All experiments in this paper are finished on an NVIDIA Tesla K40 GPU using the TensorFlow framework [9].

5.1 Accuracy comparison with previous methods

The performance of our proposed method is shown in Table 1. For the 2-class classifications, our method achieved 90.6% average accuracy. The structured data in

EMR only plays an auxiliary role, and the pathological images are the gold standard for the final diagnosis in clinical practice. Therefore, there is currently almost no paper that classifies breast cancer using only clinical EMR, we only compare it with the method only use single-mode histopathological images. Because some of the previously published papers reported 2-class classification and others reported 4-class classification. For the comprehensive comparison, we compared the accuracy with all the methods.

Table 1. Comparation of accuracy with previous methods.

Methods	Accuracy
Bayramoglu et al. (two-class) [4]	83%
Fabio A. Spanhol et al. (two-class) [10]	85%
Araujo et al. (two-class) [11]	83.3%
Rakhlin et al. (four-class) [12]	87.2%
Yeeleng S. Vang et al. (four-class) [13]	87.5%
Aditya Golatkar et al. (four-class) [14]	85%
Aresta, Guilherme et al. (BACH contest) [15]	87%
Our proposed	90.6%

5.2 Accuracy comparison using different dimensional fusion

For our proposed methods with different strategies to integrate low-dimensional structured data and high-dimensional unstructured data, we compare their overall performance via average classification accuracy in Table 2. When only structured data from EMR were used, the classification accuracy was not very high, only 81.5% on the test set. This is a reasonable result. Because the structured data in EMR only plays an auxiliary role, the pathological image is the gold standard for the final diagnosis in clinical practice. In addition, Due to the small amount of structured data in EMR and the large amount of pathological image data, especially after the enhancement of pathological image data, the use of structured data alone is the only case in which the phenomenon of overfitting exists.

When we used VGG16 to classify pathological images, we got a relatively high accuracy of 83.6%. Although the accuracy of using only structured data is not high, the leverage of structured data can improve the accuracy of pathological image classification. We got 87.9% accuracy when integrated 29-dimensional structured data and 29-dimensional feature representation of pathological images. We added a full connectivity layer with 29 nodes at the end of the VGG network, thus obtaining the 29-

dimensional feature representation of pathological images.

Further, we compared two different fusion methods of high-dimensional pathological images and low-dimensional structured data. The experimental results show that it is better to first reduce the 4096-dimensional vector extracted from the last full connected (FC) layer of VGG16 to 29 dimensions, and then fuse it with the 29-dimensional structured data. This strategy is better than directly integrated 29-dimensional structured data with 4096-dimensional feature representation of pathological image. Because the dimension of the 29-dimensional is too low compared to the 4096-dimensional. In this fusion, the higher-dimensional vectors completely overwhelm the lower-dimensional ones.

Finally, after trying different multiples of the amplification, a 20-fold amplification of structured data yields best results. When the amplification of structured data is 10 times, the overall accuracy is barely changed. The reason for this is that the amplification is insufficient. However, if amplification too much on low-dimensional structured data, the results will go down.

Table 2. Comparation of accuracy using different dimensional fusion of structured data and pathological image.

Method	**Accuracy (train)**	**Accuracy (test)**
Structured data only (29D)	89.3%	81.5%
Pathological image only (VGG16)	84.5%	83.6%
Structured data (29D) + Pathological image (29D)	88.2%	87.9%
Structured data (29D) + Pathological image (4096D)	85.1%	84.2%
Structured data (29D*10) + Pathological image (4096D)	86.2%	84.8%
Structured data (29D*20) + Pathological image (4096D)	92.3%	90.1%
Structured data (29D*30) + Pathological image (4096D)	85.5%	85.2%

5.3 ROC comparison using richer feature representation

We can further improve our model using richer convolutional features extracted from different convolutional layers of VGG16. After trying fusion different number of convolutional layers, we get our best model by integrating 29-dimensional structured data and the third, fourth and fifth convolution layers extracted from VGG16 network (1860-dimensional). And Fig. 3 shows the Area Under Curve (AUC) using different fusion method based on Receiver Operating Characteristic (ROC) analysis.

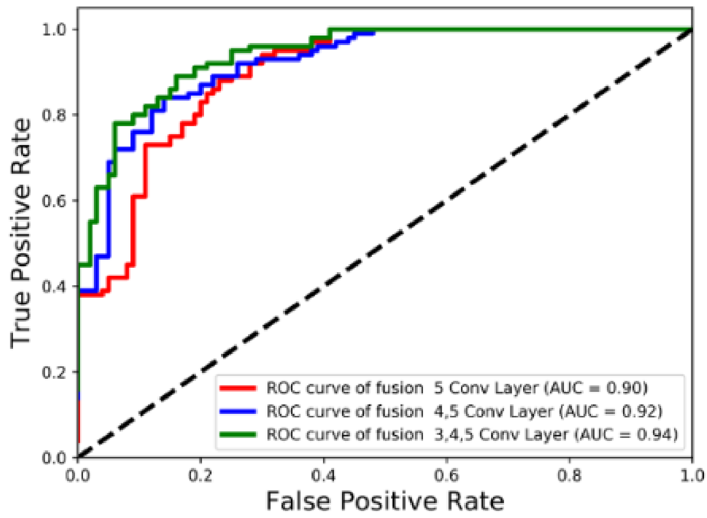

Fig. 3. ROC curves were compared integrating different convolutional layers of VGG16.

6 Conclusions

In this paper, we utilized a new method to integrate highly heterogeneous data to leverage structured data of EMR to improve pathological images classification accuracy. Therefore, the application of automatic classification algorithm in clinical practice becomes possible. Due to the generality of the proposed fusion workflow, it can be straightforwardly extended to other fusion of structured data and unstructured data.

Acknowledgement

This research is supported by the National Key Research and Development Program of China (No. 2017YFE0103900 and 2017YFA0504702), the NSFC projects Grant (No. U1611263, U1611261, 61502455 and 61672493), Peking University International Hospital Research Grant (No. YN2018ZD05), Beijing Municipal Natural Science Foundation Grant (No. L182053) and Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase).

References

1. Holzinger, A., et al., *Towards the Augmented Pathologist: Challenges of Explainable-AI in Digital Pathology.* 2017.
2. Lecun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436.

3. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Medical Image Analysis, 2017. **42**(9): p. 60-88.
4. Bayramoglu, N., J. Kannala, and J. Heikkilä. *Deep learning for magnification independent breast cancer histopathology image classification*. in *International Conference on Pattern Recognition*. 2017.
5. Xie, S. and Z. Tu, *Holistically-Nested Edge Detection.* International Journal of Computer Vision, 2017. **125**(1-3): p. 3-18.
6. Ramachandram, D. and G.W. Taylor, *Deep Multimodal Learning: A Survey on Recent Advances and Trends.* IEEE Signal Processing Magazine, 2017. **34**(6): p. 96-108.
7. Zhang, Z., et al. *TandemNet: Distilling Knowledge from Medical Images Using Diagnostic Reports as Optional Semantic References*. in *International Conference on Medical Image Computing & Computer-assisted Intervention*. 2017.
8. Tao, X., et al. *Multimodal Deep Learning for Cervical Dysplasia Diagnosis*. in *International Conference on Medical Image Computing & Computer-assisted Intervention*. 2016.
9. Abadi, M., et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.* 2016.
10. Spanhol, F.A., et al. *Breast cancer histopathological image classification using Convolutional Neural Networks*. in *International Joint Conference on Neural Networks*. 2016.
11. Araújo, T., et al., *Classification of breast cancer histology images using Convolutional Neural Networks.* Plos One, 2017. **12**(6): p. e0177544.
12. Rakhlin, A., et al. *Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis*. in *International Conference Image Analysis and Recognition*. 2018.
13. Vang, Y.S., Z. Chen, and X. Xie, *Deep Learning Framework for Multi-class Breast Cancer Histology Image Classification.* 2018: p. 914-922.
14. Golatkar, A., D. Anand, and A. Sethi, *Classification of Breast Cancer Histology using Deep Learning.* 2018.
15. Aresta, G., et al., *BACH: Grand Challenge on Breast Cancer Histology Images.* 2018.