ISSN 2618-1584 Vol. 4, Issue 4: 64-69, DOI: 10.25236/FMSR.2022.040410

Identification of Molecular Subtypes of Breast Cancer Based on Multimodal Deep Learning

Hongmei Tang[1], Luhang Dai[2]

[1]*College of Life Sciences and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China*
[2]*College of Life Science, Shaanxi Normal University, Xi'an, Shaanxi, 710119, China*

Abstract: *The identification of breast cancer subtypes plays a key role in the prognosis of breast cancer. In recent years, deep learning (DL) has shown good performance in intelligent identification of breast cancer subtypes. However, most of the traditional DL models use single-mode data, and the extracted features are limited, so the association between patient characteristics and breast cancer subtypes cannot be established stably. In order to improve the effect of recognition, this study proposes a multimodal fusion deep learning (MFDL) model. This model combined with the breast cancer gene modal data and image modal data established a multilayer perceptron network and the depth of the convolution neural network for feature extraction, and then based on the idea of weighted aggregation on the output of the two characteristics of the network integration. Finally, the fusion features were used to identify breast cancer subtypes. The experimental results show that compared with other models in AUC value, accuracy and other indicators, the MFDL model proposed in this study is more accurate and efficient in the identification of breast cancer subtypes.*

Keywords: *Breast Cancer Subtype, Multimodal, Deep Learning, Feature Fusion, In-Telligent Recognition*

1. Introduction

Breast cancer is a highly heterogeneous cancer composed of a variety of prognostic molecular subtypes, among which inter-tumor or intra-tumor heterogeneity is the key to drug resistance and treatment failure [1]. Therefore, in order to provide precise treatment, it is necessary to further identify the patient's breast cancer subtype. With the rapid development of medical science, immunohisto-chemical markers (IHC) have enabled the classification of breast cancer molecular subtypes: LuminalA, LuminalB, Her2-enriched, and basal-like [2]. Although IHC has obvious advantages in the identification accuracy of breast cancer molecular subtypes, its identification cycle and high cost are a drawback that cannot be ignored.

In recent years, machine learning and deep learning technologies have been used for intelligent diagnosis of molecular subtypes of breast cancer. Ha et. Al. [3] proposed a customized 14-layer convolutional neural network (CNN) for the identification of molecular subtypes of breast cancer. They adopted MRI data set of 216 breast cancer patients and identified them according to medical subtype classification, and finally achieved 70% accuracy in the classification of four subtypes. Coutureet.al.[4] adopted a weighting strategy to identify molecular subtypes of breast cancer using improved VGG16. They used pathological image data of 859 patients, and finally obtained a recognition accuracy of 77%. However, due to factors such as low sensitivity, low positive predictive value, high false positive rate and limited dimension, it is difficult to accurately identify molecular subtypes with single-mode features [5]. Therefore, this paper proposes a multi-mode fusion model based on the concept of multi-mode deep learning based on the gene modal data and image modal data of breast cancer patients, which improves the identification accuracy of molecular subtypes of breast cancer.

2. Data Acquisition and Data Preprocessing

2.1. Introduction to Data Sets

The TCGA-BRCA public dataset was used as a sample dataset to identify the molecular subtypes of breast cancer, which contains gene expression data and copy number variation (CNVs) data from 1098

ISSN 2618-1584 Vol. 4, Issue 4: 64-69, DOI: 10.25236/FMSR.2022.040410

breast cancer patients with desensitized information and 1-10 unequal full-size histopathological images from each breast cancer patient sample. The gene expression data and CNVs data were one-dimensional data, while the pathology data were color images. Therefore, the data set is divided into two types of modal data, namely gene modal data and image modal data. Some samples lack data or labels. In this paper, the interference samples were screened and the gene expression data were preprocessed by Log2($y = \log_2 x$), which is a commom method for gene data preprocessin[6].The specific data expression finally obtained is shown in Table 1.

Table 1: A detailed description of TCGA multimodal dataset.

The data type	Indicators	The file fomat	Original processing mode
Gene expression	20530	txt	Log2 process
CNVs	24776	txt	None
Pathological image	1-10(page)	svs	None

In this paper, the samples were shuffled and divided into training set, validation set and test set in the ratio of 8:1:1 according to the total amount of data samples. At the same time, considering the difference in sample quantity distribution of the four molecular subtypes of breast cancer, the same proportion of stratified sampling method was used to obtain the sample sets of the four molecular subtypes.

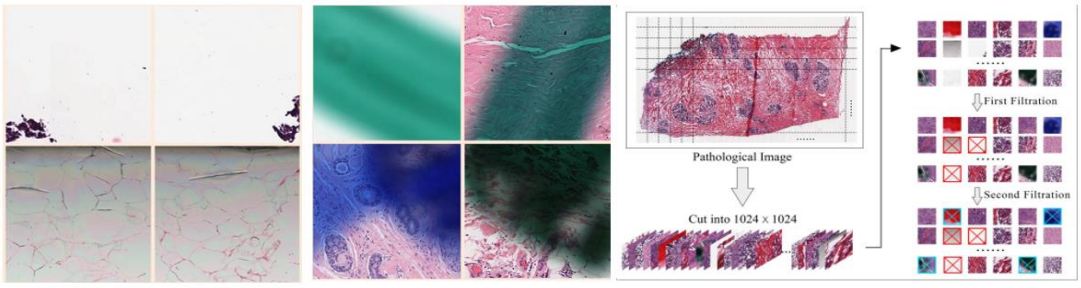
2.2. Preprocessing of Gene Modal Data

The gene expression data after standardized processing was integrated with CNVs to obtain gene modal data, which contained 45,306 data indicators. Using high-dimensional data for network training will lead to more network parameters and longer training time. More importantly, the training samples of the data set in this paper are limited, and the final identification results are often inaccurate [7]. In this paper, PCA method was used to reduce dimension of gene modal data and the principal component whose eigenvalues of the contribution rate reaches 90% was selected as a dimension reduction result. Finally, 398 data indexes were obtained as the input of gene modal feature extraction network.

2.3. Image Modal Data Preprocessing

The number of pixels contained in the full-size pathological images of each sample is tens of millions. In this paper, a size of 1024×1024 was selected to cut full-size pathological images, and about 100 submaps could be cut from each full-size pathological image. As some local areas of full-size pathological images tend to be colorless or white, they contain very little characteristic information, as shown in Figure 1 (a). This "white" noise image will adversely affect network convergence and lead to performance degradation. Therefore, all pathological subgraphs were filtered for the first time in this paper to improve the robustness of the data set. Top50 selection method was adopted to select 50 pathological submaps with the most information as the initial image modal data set of a sample according to the average gray value.

In addition, the residual staining fluid after hematoxylin-eosin staining of tissue sections is the main cause of these noises in Figure 1 (b). The existence of such dyeing noise information will deteriorate the performance of DCNN and lead to the network convergence difficult. A second filter is used to further enhance the effective features of the image set. Due to the fact that an overstained subgraph will usually have long strips in place. Therefore, the Hough transform method was used in this paper to filter and eliminate the pathological subgraph with excessive staining.



(a) (b) (c)

Figure 1: (a) The"white" noise image (b) The overdyed noise image (c) Schematic diagram of the basic flow of cutting and secondary filtering of full-dimensional pathological images.

ISSN 2618-1584 Vol. 4, Issue 4: 64-69, DOI: 10.25236/FMSR.2022.040410

Finally, random affine transformation is mainly used in this paper to enhance and standardized the pathological subgraph, to break the limitation of small sample data to a certain extent, improve the generalization ability of DCNN model and the performance of the model on the test set, and accelerate the convergence rate of the network.

3. Construction of Multi-Mode Fusion Deep Learning Model

3.1. Multilayer Perceptron Model Based on Gene Modes

MLP model can well extract features from one-dimensional data, and can describe rich internal information of one-dimensional data to obtain feature output. In order to extract the abstract features of gene modes more fully, an MLP network structure was innovatively defined in this paper with the design of "inverted pyramid" and co- mbining multiple activation functions, which referred to the layer design of Lenet-5 [8]. The specific MLP structure and activation functions of each layer are shown in Figure 2 (a). L2 regularization model [9], exponential weighted moving average model (EWMA) and simple and computation-efficient Adam optimizer are also adopted in this paper to prevent the over-fitting phenomenon.

3.2. Image Modal-Based Deep Convolutional Neural Network

3.2.1. Improved Concrete Structure of Deep Convolutional Neural Network

Based on the famous VGG16 model [10], this paper independently designed a DCNN model to extract high-dimensional abstract features from pathological images. The DCNN model consists of 28 layers, including input layer, convolutional layer, Inception layer, pooling layer and output layer, which can fully extract the key abstract features of the three-channel pathological submap. Inception layer refers to the penultimate layer in InceptionV3 model [11], which is introduced to further break the bottleneck of feature extraction from pathological images. The specific CNN model structure is shown in Figure 2 (b). In order to prevent overfitting, Dropout technology is introduced in training networks [12].

3.2.2. Transfer Learning Method

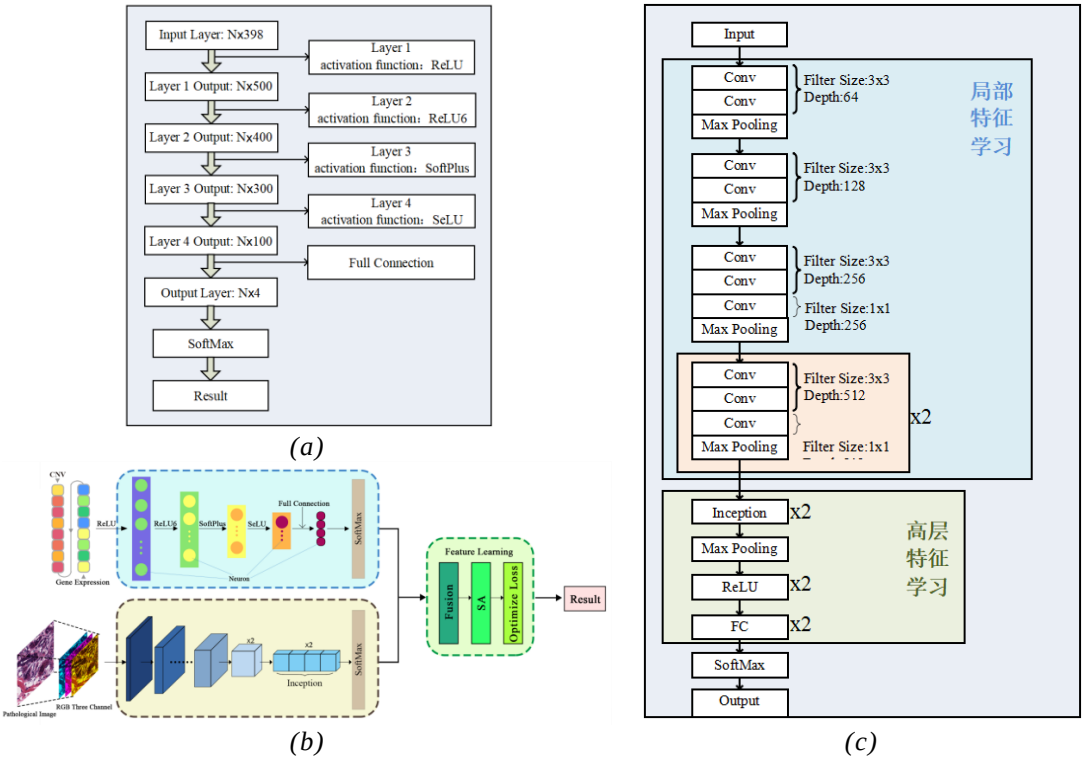


Figure 2: (a) Specific structure of each MLP model designed in this paper (b) Specific description of the parameters of each layer of the DCNN model (c) The whole process of identification of molecular subtypes of breast cancer by a multi-modal fusion architecture.

ISSN 2618-1584 Vol. 4, Issue 4: 64-69, DOI: 10.25236/FMSR.2022.040410

Due to the limited number of effective image training samples, transfer learning is also introduced to accelerate network convergence. Firstly, the DCNN model in this paper was pre-trained on ImageNet data set to extract reuse features, and then the pre-trained model was converted to pathological image data for fine-tuning training. In fine-tuning training, the Inception model trained by Google on ImageNet dataset is directly used as the migration module in the DCNN model in this paper. In addition, the "local feature learning" part after pre-training is frozen, and only the structural parameters of "high level feature learning" part are updated, which including inception layer, max pooling layer, Relu layer and FC layer, so as to achieve the purpose of fast matching pathology image data set.

3.3. Multimodal Fusion

In order to fuse the features of the two modes, the following fusion methods are defined in this paper:

$$\begin{cases} result_{fusion} = \alpha \cdot result_{mlp} + \beta \cdot result_{dcnn} \\ \alpha + \beta = 1 \\ 0 \le \alpha \le 1, 0 \le \beta \le 0 \end{cases}$$

Where $result_{mlp}$、$result_{dcnn}$ are respectively the MLP model feature extraction results obtained by Softmax classifier from gene modal data and image modal data, and $result_{fusion}$ represents the result after feature fusion.

In this paper, simulated annealing algorithm (SA) is used to obtain the best combination of (α, β) is (0.88185, 0.11815) on the verification set. The realization process of the entire multi-mode fusion architecture is shown in Figure 2 (c).

4. Experimental Results and Discussion

4.1. Multimodal Fusion Results

The identification accuracy of MLP model, DCNN model and MFDL model is 86.21%, 70.11% and 88.51%, respectively. The identification accuracy of MFDL model is 2.7% higher than that of the optimal single mode model. In addition, the loss value of each model on the loss function is also counted in this paper, among which the loss value of MLP model is 0.26036, DCNN model is 0.48633, and MFDL model is only 0.17954.

4.2. Cross Validation and AUC Test

In order to improve the reliability of model identification results, this paper carried out ten cross-validation of MLP model, DCNN model and MFDL model. The results of ten fold cross validation are shown in Figure 3 (a). The average accuracy of MFDL model was 88.07%, while MLP model was 85.06%, DCNN model was 72.77%.

In order to evaluate the performance of MFDL model for identification of a certain molecular subtype of breast cancer, ROC curves were made for each molecular subtype, as shown in Figure 3(b)-(e), and AUC values of the model for identification of each molecular subtype were calculated, as shown in Table 2. Since ROC curve drawing and AUC value calculation are based on dichotomies, this paper takes the method of classifying other kinds of molecular subtypes into one class when evaluating the identification performance of a certain molecular subtype.

Table 2: The AUC values obtained from the MFDL model, MLP model, and DCNN models on the subtype identification work in this pager.

Molecular subtype category	MFDL model	MLP model	DCNN model
Basal-like	0.9331	0.8548	0.6364
Her2-enriched	0.9732	0.8707	0.6902
Luminal A	0.9316	0.8558	0.6751
Luminal B	0.9328	0.8425	0.6709
Average	0.9427	0.8561	0.6682

ISSN 2618-1584 Vol. 4, Issue 4: 64-69, DOI: 10.25236/FMSR.2022.040410

The results show that MFDL model is superior to the other two models based on single mode in the prediction of various molecular subtypes from the size of AUC value, and in the AUC level rating, MFDL model has opened a gap with MLP model to some extent. In the ten-fold cross validation, the accuracy of MFDL model is 3.53% higher than that of MLP model, and the difference between them is not particularly large. At the AUC level, the average AUC value of MFDL model is 10.12% higher than that of DNN model, which also indicates that MFDL model is more robust and has stronger prediction ability than MLP model.

5. Conclusions

This paper mainly studies the intelligent recognition of molecular subtypes of breast cancer. A multi-modal fusion deep learning (MFDL) model is proposed to identify the molecular subtypes of breast cancer, and extract the deep features of gene modal data and image modal data. Finally, the two modal data are fused to identify the molecular subtypes of breast cancer intelligently. The results of 10 fold cross validation and AUC test show that the multimodal MFDL model proposed in this paper is superior to the traditional single mode model, and may become potential choice for intelligent identification of molecular subtypes of breast cancer in the future. However, the number of samples in the data set used in this paper is limited and the interference information in the image source is not completely removed, so this paper still has some room for improvement. As a new feature recognition technology, multimodal fusion technology shows great advantages in the field of recognition and can be extended to more research fields.

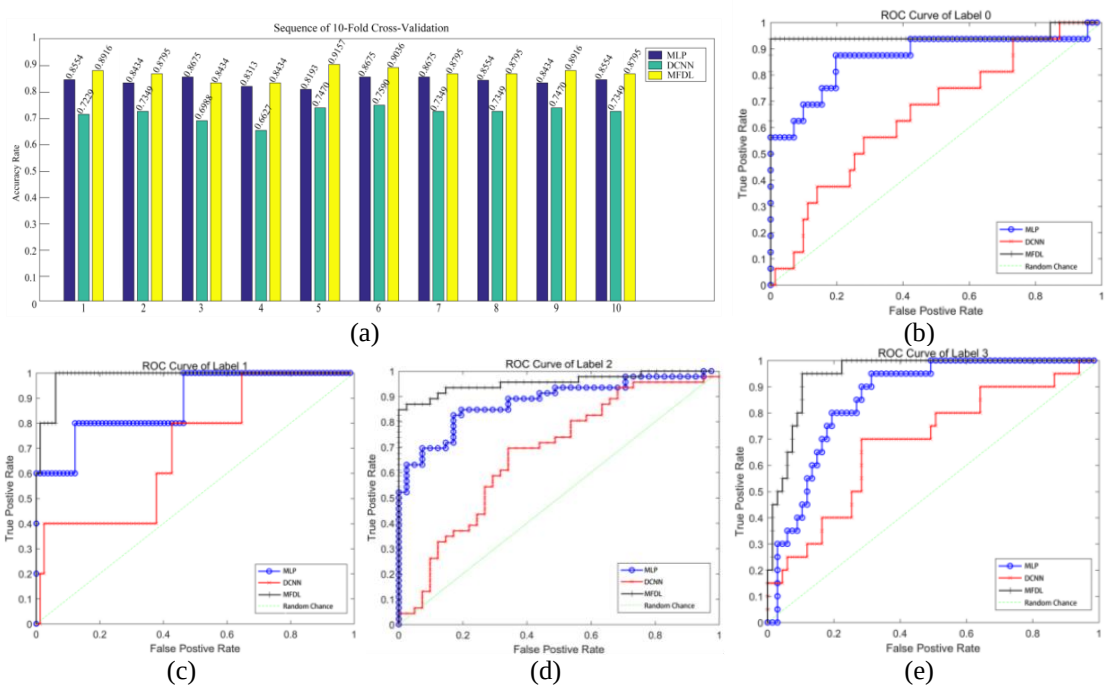


Figure 3: (a) Specific outcome plots of ten times of ten-fold cross-validation (b), (c), (d) Plot of the representation of the ROC curves made by the different molecular subtypes.

References

[1] McGranahan N, Swanton C, Clonal heterogeneity and tumor evolution: past, present, and the future [J]. Cell, 2017, 168(4): 613-628.

[2] Guiu S, Michiels S, Andr éF, et al., Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement [J]. Annals of oncology, 2012, 23(12): 2997-3006.

[3] Ha R, Mutasa S, Karcich J, et al., Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm [J]. Journal of digital imaging, 2019, 32(2): 276-282.

[4] Couture H D, Williams L A, Geradts J, et al., Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype [J]. NPJ breast cancer, 2018, 4 (1): 1-8.

ISSN 2618-1584 Vol. 4, Issue 4: 64-69, DOI: 10.25236/FMSR.2022.040410

[5] Lahat D, Adali T, Jutten C, Multimodal data fusion: an overview of methods, challenges, and prospects [J]. Proceedings of the IEEE, 2015, 103(9): 1449-1477.
[6] Quackenbush J, Microarray data normalization and transformation [J]. Nat Genet, 2002, 32: 496–501.
[7] Fu X, Wang L, Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2003, 33(3): 399-409.
[8] LeCun Y, Bottou L, Bengio Y, et al., Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[9] Murugan P, Durairaj S, Regularization and optimization strategies in deep convolutional neural network [J]. arXiv preprint arXiv: 1712.04711, 2017.
[10] Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
[11] Szegedy C, Vanhoucke V, Ioffe S, et al., Rethinking the inception architecture for computer vision [C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
[12] Srivastava N, Hinton G, Krizhevsky A, et al., Dropout: a simple way to prevent neural networks from overfitting [J]. The journal of machine learning research, 2014, 15(1): 1929-1958.