

Prediction of Preoperative Lymph Node Metastasis of Esophageal Cancer by Self-coding Fusion Model Based on Multi-scale Features

Yixiang Hepeng¹, Xu Wu^{1,*}, Zhi Xu², Kailin Qu¹

¹Department of Thoracic Surgery, Southern Medical University Nanfang Hospital, Guangzhou, 510515, China

²Medical Center for Overseas Patients, Southern Medical University Nanfang Hospital, Guangzhou, 510515, China

*Corresponding author: wuxu_southhospital@163.com

Keywords: Esophageal cancer; Multi-scale features; Self-coding fusion model; State prediction

Abstract: Esophageal cancer is one of the eight most common malignant tumors, with high incidence and mortality rates. Most patients are already in the advanced or late stage at the time of diagnosis, missing the best treatment period and resulting in extremely poor prognosis. Factors affecting the prognosis of esophageal cancer include clinical staging, lymph node status, and pathological type. Early diagnosis and personalized treatment plans, as well as controlling the risk factors for esophageal cancer, can effectively improve the prognosis of patients. Therefore, early diagnosis and treatment of esophageal cancer have become extremely important. This paper proposes a model that comprehensively predicts lymph node metastasis status through multi-level image features. It utilizes sparse self-encoded feature fusion networks to process high-dimensional features from different levels, including machine vision features, imaging genomics features, and perceptual features. The model is constructed using statistical methods and experimentally verified for its discriminative ability, identification capability, and clinical practicality.

1. Introduction

Currently, malignant tumors have become one of the major public health threats to human health, posing a serious threat to life and health. According to the 2018 Global Cancer Report published by the International Agency for Research on Cancer of the World Health Organization in the authoritative journal "CA Cancer J Clin" [1], in 2018, there were a total of 18.1 million new cases of cancer globally, with 9.6 million new deaths. Esophageal cancer ranked ninth, further increasing the global burden of cancer. The incidence of cancer in China remains high. According to the 2019 National Cancer Report released by the National Cancer Center [2], in 2015, there were 3.929 million new cases nationwide, with 2.338 million deaths, and esophageal cancer ranked fifth. Therefore, early diagnosis and treatment of esophageal cancer are particularly important. A large number of clinical studies have confirmed that controlling the risk factors for cancer reasonably is

an effective way to reduce the incidence and mortality rates [3]. Standard and accurate diagnosis, personalized treatment plans, can effectively improve patients' survival time and quality of life. Compared with the five-year survival rate of patients with lymph node metastasis, the five-year survival rate of patients without lymph node metastasis is longer. Among patients with lymph node metastasis, the number of lymph node metastases is closely related to the survival rate. Generally, the more lymph node metastases, the shorter the survival time [4]. At the same time, the status and number of lymph node metastases directly affect the choice of surgical methods. Therefore, preoperative accurate diagnosis of lymph node metastasis status helps to choose a reasonable treatment method and predict prognosis. This article addresses the clinical problem of the difficulty in determining whether lymph node metastasis has occurred before esophageal cancer surgery, and proposes a multi-level imaging feature comprehensive prediction model for lymph node metastasis. It constructs a sparse autoencoder feature fusion network to process high-dimensional features from different aspects such as machine vision features, imaging genomics features, and perceptual features, further selects a feature set related to lymph node metastasis status, constructs a model using statistical methods, and comprehensively evaluates the model's discriminative ability, identification ability, and clinical utility.

2. Deep learning features

The features of radiomics and machine vision are limited, only reflecting some characteristics of the tumor area, only reflecting the superficial characteristics of the tumor, such as the radiomics features, only reflecting the part that can be described by a formula, and ignoring the part of the features that cannot be described by a formula; while the machine vision features, to a certain extent, compensate for the limited number of features defined in radiomics, but they are still superficial features. Therefore, to obtain more comprehensive information about the tumor area, it is necessary to explore its deep features and explore the characteristics of tumor heterogeneity at a deeper level.

The extraction of deep learning features in this paper is achieved through deep convolutional neural networks. Generally, there are two sources of deep convolutional neural network models. For data with sufficient training samples and corresponding labels, a network model can be trained. For data with a small amount of data or insufficient annotated data, such as training a network model that is prone to overfitting, transfer learning or data augmentation can be used. Since it is difficult to obtain preoperative CT image data of esophageal cancer studied in this paper, the sample size is limited and not sufficient for training network models, so the transfer learning method is adopted for the extraction of deep features. The network model structure and parameters of this paper are shown in Figure 1.

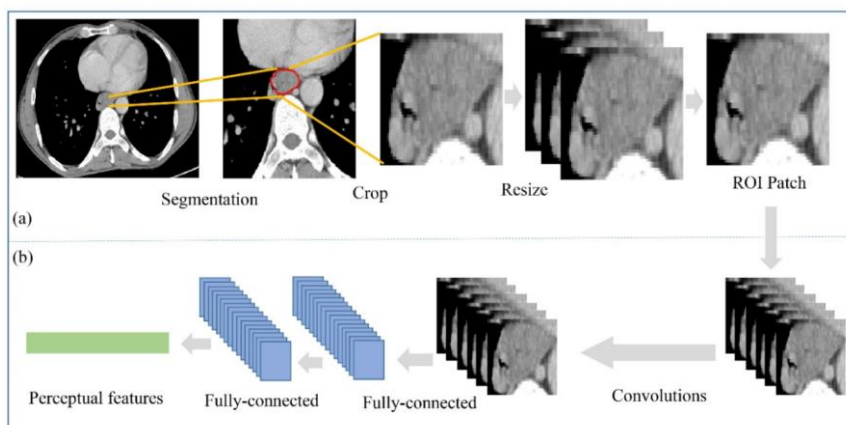


Figure 1: Flowchart of deep learning feature extraction

This text uses a pre-trained network model CNN-Fb [5] as a deep feature extractor. CNN-F consists of five convolutional layers (Conv1-5) and three fully connected layers (fully-connected1-3), with pooling layers following conv1, conv2, and conv5. CNN-F is a network model trained on the ILSVRC-2012 dataset, with training parameters including: momentum: 0.9; weight decay: 5×10^{-4} ; initial learning rate: 0.01 and reduced by 10 times when validation error stops decreasing; dropout: 0.5; during training, stochastic gradient descent (SGD) is used to adjust network parameters. The trained model is transferred to the preoperative CT lymph node metastasis dataset in this paper, used as a feature extractor for deep feature extraction.

The process of deep feature extraction is as follows: the input image of CNN-F is a three-channel 224×224 picture, while medical images encoded in DICOM format are single-channel grayscale images with a wider grayscale range (16 bits). In order to align with the input channels of the pre-trained model, the following operations are performed on the CT images. First, the slice with the largest tumor area is selected from each patient's CT slice, and the tumor is manually segmented along the tumor boundary (excluding air in the tumor). Then, the segmented tumor area is cropped, with the cropping border encompassing the entire tumor area, and the cropped area is resized back to 224×224 using bicubic interpolation. Finally, the adjusted single-channel image is encoded as a three-channel image. At this point, it meets the input requirements of the model. As a feature extractor, the last output layer of the CNN-F model (fully connected layer 8) is removed, and deep features can only be computed through forward propagation and extracted from fully connected layer 7.

3. Design of sparse autoencoder network

The sparse autoencoder network introduced in this article is used for the overall framework of feature fusion, as shown in Figure 2.

It mainly includes three steps: (1) Data cleaning. For the extracted imaging omics features, deep features, and visual features mentioned above, inter-group consistency detection is first performed to remove features with low repeatability; then the remaining features are normalized. (2) Building the autoencoder network structure. For the standardized features after data cleaning, a high- to low-dimensional mapping autoencoder network structure is constructed. (3) Extracting fused features. The low-dimensional features from step two are extracted for model construction, further evaluating the model's performance in predicting the preoperative lymph node metastasis status.

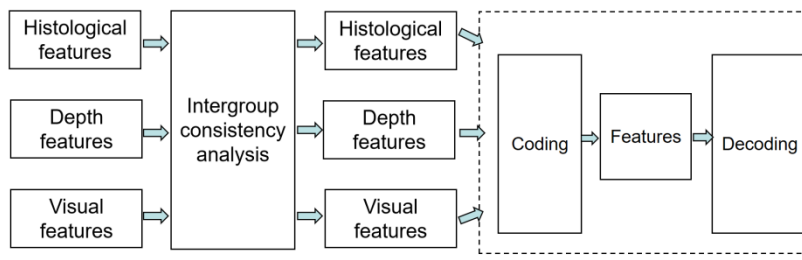


Figure 2: Flowchart of feature fusion

3.1 Design of sparse autoencoder network model structure

The structure of the sparse autoencoder network in this article is shown in Figure 3. The network structure consists of an input layer, three hidden layers, and an output layer. After data cleaning, 4004 features are used as the input to the autoencoder network. In order to represent the nonlinear relationships between the data and improve the model's expressive power, three hidden layers are designed with the numbers of neurons in each hidden layer being {2048, 1024, 256}. The output

consists of 64-dimensional features after dimensionality reduction.

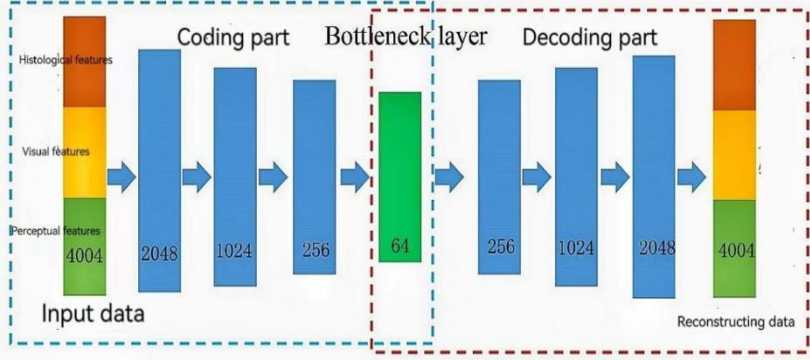


Figure 3: The structure of the sparse autoencoder network

Autoencoders realize high-dimensional feature-to-low-dimensional mapping, including: (1) Network model pre-training: When there are many hidden layers in the model, the network training process may encounter challenges such as local optimal solutions and overfitting. In this chapter, we use a layer-by-layer greedy algorithm to individually train each layer of the network, ultimately obtaining the parameters for each layer. After cleaning and standardizing the data, we input it into the autoencoder network. In the first layer, we obtain the network weights $w^{(1)} = [w^{(1,1)}, w^{(1,2)} \dots]$ and bias $b^{(1)} = [b^{(1,1)}, b^{(1,2)} \dots]$. We then save the network weights and biases of the first layer, and through the activation function, activate the output of the first layer to serve as the input for the second encoder. We then train the second encoder separately, obtaining its weights $w^{(2)} = [w^{(2,1)}, w^{(2,2)} \dots]$ and bias $b^{(2)} = [b^{(2,1)}, b^{(2,2)} \dots]$, and save these weights and biases. This process continues until all layers are trained and all parameters are obtained. For the weights and biases of each layer, they can be represented as follows:

$$w^{(1)} = [w^{(1,1)}, w^{(1,2)} \dots] \quad (1)$$

$$b^{(1)} = [b^{(1,1)}, b^{(1,2)} \dots] \quad (2)$$

Where l denotes the number of layers.

Taking into account the trained parameters, the weights and biases of the entire network are as follows:

$$W = [w^{(1)}, w^{(2)}, \dots, w^{(l)}] \quad (3)$$

$$b = [b^{(1)}, b^{(2)}, \dots, b^{(l)}] \quad (4)$$

(2) Parameter fine-tuning: According to the principle of minimizing the error between input data and network reproduction data, fine-tune the entire network for parameters W and b using error backpropagation algorithm.

1) Forward propagation, calculating the output values $h_k^{(1)} = \{h_k^{(1)}, h_k^{(2)}, \dots, h_k^{(l)}\}$ of each hidden layer and the output layer \widehat{X}_j .

2) For the output layer, calculate its error term $\delta^{l+2} = -(\nabla_{h^{l+2}} J) f'(z_i^{(2)})$.

3) Hidden layer: $\delta^l = ((w^l)^T \delta^{l+1}) f'(z_i^{(l)})$.

4) Backward differentiation: $\nabla_{w^l} J(W, b) = \delta^{l+1} (h^l)^T$, $\nabla_{b^l} J(W, b) = \delta^{l+1}$

3.2 Clinical Model Construction and Evaluation Indicators

(1) Model Construction

Before constructing the model, the features were extracted from the bottleneck layer first and the Maximum Relevance Minimum Redundancy (MRMR) algorithm was used to rank the features from high to low.

$$\max(D - R) = \max \left\{ \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) - \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \right\} \quad (5)$$

Where S represents the subset of target features, and I represents mutual information. By maximizing the correlation between features and labels while minimizing the correlation between features, the optimal features are selected. The number of key features used to build the model is determined by 10-fold cross-validation C-statistics. The selected key features are used to build the model using logistic regression $f(x) = 1/(\exp(-x))$, and are evaluated on internal and external validation sets.

(3) Evaluation Indicators

In order to comprehensively evaluate the performance of the constructed model, a comprehensive evaluation is conducted from the overall performance of the model, the model's ability to distinguish between the lymph node metastasis status of esophageal cancer, and the clinical practicality of the model.

1) Overall performance evaluation of the model

The Brier score is used to evaluate the overall performance of the model. The Brier score is used to measure the consistency between the observed event outcome and the predicted probability of the outcome. In this paper, it refers to the consistency between the predicted probabilities of lymph node metastasis being positive or negative and the actual outcomes.

$$Brier_{score} = \frac{1}{N} \sum_{i=1}^N (y_i - prob_i)^2 \quad (6)$$

Where the variable y_i represents the actual occurrence of sample i , where a transition represents occurrence with a probability of 1, and non-occurrence with 0. The variable $prob$ represents the model's predicted probability for sample i .

2) Model Discrimination and Calibration

Discrimination measures the model's ability to distinguish between lymph node metastasis and non-metastasis. This is typically evaluated using the area under the receiver operating characteristic curve (AU-ROC) or the C-index. The AU-ROC (or C-index) ranges from 0.50 to 1.00, with a higher value indicating stronger discrimination. Calibration measures the model's predictive ability on unknown data, often assessed using the Hosmer-Lemeshow (H-L) test, which examines the degree to which the observed lymph node metastasis ratio matches the predicted ratio relative to the predicted risk percentage.

3) Clinical Utility

This evaluates the practical benefits the model can bring to clinical practice or patients. It primarily uses methods such as the net benefit curve and decision curve analysis.

4. Clinical Model Evaluation

(1) Overall Performance of the Clinical Model

The overall performance of the lymph node metastasis prediction model is shown in Table 1. The constructed model can satisfactorily predict the lymph node metastasis status. The model's ability to distinguish between lymph node metastasis and non-metastasis reached 0.856 (training set), 0.838 (internal validation), and 0.814 (external validation), and the Hosmer-Lemeshow test confirmed that

the model has predictive significance.

Table 1: Indicators for the overall evaluation of the model

| | Training set | Internal validation | External validation |
|------------------------|--------------|---------------------|---------------------|
| Overall Brier | 0.151 | 0.172 | 0.186 |
| Discrimination C-index | 0.856 | 0.838 | 0.814 |
| Calibration H-L test | 0.538 | 0.522 | 0.058 |

(2) Clinical practicability evaluation

In clinical research, traditional diagnostic test indicators such as accuracy, sensitivity, specificity, and the area under the ROC curve only measure the diagnostic accuracy of the predictive model and do not consider the clinical utility of the model constructed. Decision curve analysis (DCA) serves as a simple and intuitive way to evaluate clinical predictive models, diagnostic tests, molecular biomarkers, etc., effectively integrating patient or decision maker preferences. As shown in Figure 4, the horizontal axis represents the threshold probability. In the assessment of lymph node metastasis risk, if a patient is diagnosed with a probability of lymph node metastasis denoted as P_i , when P_i reaches a certain threshold P_t , they are diagnosed as positive (i.e., metastasis), and corresponding treatment measures are taken clinically. At this point, some patients will benefit from the treatment, while others will suffer losses due to the treatment, i.e., they will not benefit. The vertical axis represents the net benefit after weighing the pros and cons. The red curve in the figure represents the clinical diagnostic model constructed in this study, and the remaining two curves represent two different extreme scenarios: the horizontal line represents the net benefit level of not treating all patients when all patients are negative for lymph node metastasis, while the slanted line represents the benefit situation when all patients are positive for lymph node metastasis and receive corresponding treatment.

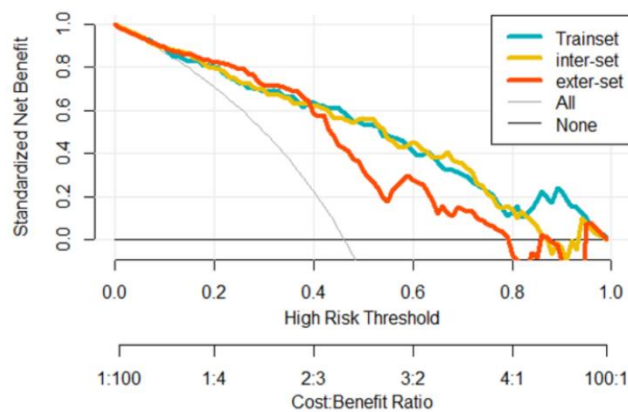


Figure 4: Assessment of Clinical Utility - Net Benefit Curve

From the figure, it can be seen that the constructed model is far from the two extreme scenarios, indicating that the model has practical value. Specifically, to interpret the marked points on the graph: assuming 60% of the predicted probability is determined as lymph node metastasis and corresponding treatment is administered, then among 100 patients using the constructed model, 40 people will benefit from the treatment measures, while no patients will benefit if all patients receive corresponding treatment (gray slanted line) or if no measures are taken (gray horizontal line). Therefore, it can be seen that the model constructed in this study can to some extent benefit patients from the corresponding treatment measures.

5. Conclusion

The main challenge in the diagnosis and treatment of esophageal cancer is the inaccurate clinical staging and the difficulty in determining the status of lymph node metastasis, which affects the decision-making for treatment. The status of lymph node metastasis is crucial for treatment planning and prognosis prediction in resectable esophageal cancer. Accurate lymph node dissection can alleviate patient suffering, reduce local recurrence, and improve survival rates. This study addresses the clinical problem of predicting lymph node metastasis in esophageal cancer before surgery. We used autoencoder networks to integrate high-dimensional feature data and selected a highly correlated feature set related to lymph node metastasis. Statistical methods were used for modeling, and in addition to conventional evaluation metrics, we also explored the model's clinical utility. This research primarily aims to address the difficulty in predicting lymph node metastasis in clinical practice.

Acknowledgement

Funding: The Start-up Plan of Southern Medical University Clinical Study (LC2016ZD004).

References

- [1] Bray F, Ferlay J, Soerjomataram I, et al. *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA Cancer J Clin, 2018, 68(6): 394-424.*
- [2] Zheng R S, Sun KX, Zhang Z, et al. *Prevalence of malignant tumors in China in 2015 [J]. Chinese Journal of Cancer, 2019, 41(1): 19-28.*
- [3] Chen S-B, Weng H-R, Wang G, et al. *Prognostic Factors and Outcome for Patients with Esophageal Squamous Cell Carcinoma Underwent Surgical Resection Alone: Evaluation of the Seventh Edition of the American Joint Committee on Cancer Staging System for Esophageal Squamous Cell Carcinoma [J]. Journal of Thoracic Oncology, 2013, 8(4): 495-501.*
- [4] Chen M-F, Chen P-T, Lu M-S, et al. *Survival benefit of surgery to patients with esophageal squamous cell carcinoma [J]. Scientific Reports, 2017, 7(1): 46139.*
- [5] Jiang H, Learned-Miller E. *Face Detection with the Faster R-CNN[C]. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017: 650-657.*