An interpretable integration model improving disease-free survival prediction for gastric cancer based on CT images and clinical parameters

Xiaoping Cen[1,3,4], Can Hu[2], Li Yuan[2], Huanming Yang[1,3,4], Xiangdong Cheng[2], Wei Dong[1,3,4], Yahan Tong[2,*]

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[2]Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, China
[3]HIM-BGI Joint Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, China
[4]BGI-Shenzhen, Shenzhen, China

*Correspondence should be addressed to Yahan Tong: tongyh2590@zjcc.org.cn.

Key Points:

- An integration model combining deep features, radiomics features and clinical parameters improved disease-free-survival prediction of gastric cancer by 3%-5% C-index.
- Embedding radiomics and clinical features into deep learning model through concatenation and loss design improved feature extraction ability of deep network.
- The model revealed disease progression trajectory and tumor heterogeneity.

Abstracts

Preoperative prediction of disease-free survival of gastric cancer is significantly important in clinical practice. Existing studies showed the potentials of CT images in identifying predicting the disease-free survival of gastric cancer. However, no studies to date have combined deep features with radiomics features and clinical features. In this study, we proposed a model which embedded radiomics features and clinical features into deep learning model for improving the prediction performance. Our models showed a 3%-5% C-index improvement and 10% AUC improvement in predicting DFS and disease event. Interpretation analysis including T-SNE visualization and Grad-CAM visualization revealed that the model extract biologically meaning features, which are potentially useful in predicting disease trajectory and reveal tumor heterogeneity. The emedding of radiomics features and clinical features into deep learning model could guide the deep learning to learn biologically meaningful information and further improve the performance on the DFS prediction of gastric cancer. The proposed model would be extendable to related problems, at least in few-shot medical image learning.

Keywords: Deep learning; Radiomics; Interpretable analysis; Integration model; Disease-free survival

Introduction

Gastric cancer (GC) is the fifth most common malignant cancer worldwide and the fourth leading cause of global cancer deaths [1]. GC patients are usually diagnosed at an advanced stage with poor prognoses. The 5-year overall survival rate of GC is 30−40% after curative resection, whereas metastasis-related or relapse-related death remain a challenge for curative treatment [2]. Preoperative prediction of the disease-free survival (DFS) of gastric cancer is remarkably important for precision treatment of patients.

Efforts have been made on predicting the DFS of gastric cancer based on clinical reports and CT image-based radiomics. Discovered risk factors include age, tumor size, tumor stage, tumor location, and Lauren type. Besides, CT images have also shown effective in predicting the DFS of gastric cancer. Radiomics features extracted from CT images showed considerable performance in predicting DFS of gastric cancer with stage II/III [3]. Multi-detector CT-based radiomics nomogram indicated high C-index and better performance in 2-year survival prediction [4].

However, due to the high tumor heterogeneity, radiomics features are not enough to extract complex information from images. Deep learning has been widely applied in medical image analysis, showing considerable performance in many cancer-related tasks [5], including the DFS prediction of gastric cancer in a multi-institution study [6] . Nevertheless, deep learning has some weaknesses in medical imaging analysis, including the lack of interpretation and the overfitting issue. Some studies have integrated deep learning features with radiomics features using linear regression models. While the method showed good performances in large dataset, directly integrating fixed deep features with radiomics features or clinical features relies largely on the feature interaction, which is inflexible and lacks of generalizability. Wei et al proposed a strategy by first mapping radiomics features, clinical features into common feature space using variational autoencoder, and then integrate them with deep features, which showed improvement of performance in predicting the overall survival of hepatocellular carcinoma [7]. Zhou et al applied multi-modal integration scheme to predict the microvascular invasion using multi-phase contrast-enhanced MRI images [8]. Based on these studies, we hypothesized that the integration of clinical parameters, deep learning features, and radiomics features in a deep learning framework could guide deep learning to learn biologically meaningful features and extract the high-level tumor heterogeneity, which would promote the DFS prediction of gastric cancer.

In this study, we aim to integrate clinical parameters, deep learning features, and radiomics features in a deep learning framework to improve gastric cancer DFS prediction. Besides, interpretation analyses are performed to reveal the model's capability in characterizing tumor heterogeneity.

Material and Methods

Dataset and Preprocessing

Study Population

The study obtained approval from the local institutional review board, and the requirement for patients' informed consent was waived. All patients underwent surgical treatment within one month of contrast-enhanced abdominal CT examination. There were primarily 1001 patients diagnosed as GC. We applied the inclusion criteria followed to identify required patient samples: (1) postoperative pathologically confirmed gastric adenocarcinoma; (2) patients who underwent contrast-enhanced CT of the whole abdomen or the upper abdomen within one month before surgery; (3) no distant metastases prior to surgery. The exclusion criteria were as follows: (1) incomplete clinical or pathological information; (2) treatment was performed before surgery; (3) poor CT image quality or unrecognizable lesion; and (4) other concurrent malignancies.

CT image acquisition

All patients underwent contrast-enhanced abdominal CT using the multidetector row CT systems: BrightSpeed, Optima CT680 Series (GE Medical Systems); Siemens Somatom definition AS 64, Perspective (Siemens Medical Systems). The acquisition parameters were the following: detector configuration 128 × 0.6 mm; tube voltage, 120-130 kV; tube current, 150-300 mAs; reconstructed axial-section thickness 5 mm, slice interval 5 mm, pitch 0.6. The contrast agents were Ultravist (Bayer Schering Pharma, Berlin, Germany), Optiray (Liebel-Flarsheim Canada Inc., Kirkland, Quebec, Canada), and Iohexol (Beijing North Road Pharmaceutical Co. Ltd., Beijing, China). A contrast medium of 1.5 ml/kg was injected through the antecubital vein. The portal venous phase were performed at 50-60 s after injection of the contrast medium.

Tumor segmentation

The CT images of portal vein phase were exported in digital imaging and communications in medicine (DICOM) format from image storage and communication systems. Lesions were segmented using ITK-SNAP (version 3.8.0, http://www.itksnap.org). Two radiologists with over five years of experience in diagnosing abdominal diseases observed each CT image of the patient and performed delineation on the largest layer of the tumor.

Radiomics feature extraction

A total of 944 radiomics features were extracted from the delineated ROI regions, including 2D shape, firstorder, glcm, glrlm, glszm, gldm, ngtdm features with original image or images after log and wavelet filtering. Lasso cox regression model with 10-fold cross validation was applied to determine useful features for DFS prediction. The Pyradiomics package in Python 3.7.3 and glmnet package in R were utilized for radiomics feature extraction and feature selection, respectively.

The Proposed Method

The Overview of the Proposed Framework

The proposed framework was summarized in **Fig.1**.

Descriptions and the Implementation

Images were resized into 28*28 as input of convolutional neural network(CNN). A ResNet50 model with pre-trained ImageNet weights for all convolutional layers was utilized to extract deep features. The ResNet50 model consists of 49 convolutional layers, followed by a fully connected layer. We added a fully connected layer into two layers to reduce the dimension of features into 256 dimension, and the output 1000 dimension was changed into one dimension referring to the risk score predicted by the model. In order to embed the radiomics features and clinical parameters into the model, we first normalized the two sets of features, and then passed them into the deep network. A FCN layer was then utilized to embed the features into 250 and 6 dimensions, respectively. The three kinds of features, were then concatenated and formed 512 dimension. Then, another FCN layer was then applied to predict risk score based on the features. We applied two kinds of strategies for calculating the loss value of the network: 1) simply using the loss value of the final 512 dimension features, named CON model; 2) combine the individual loss with the concatenated loss based on the design of deeply supervised nets [9], named DSN model. The DeepSurv [10] was used as loss function for cox regression. The DSN loss was defined as follows:

$$L_{dsn} = L_{concat} + w_a L_{deep} + w_b L_{radio} + w_c L_{clini} \quad (1)$$

All experiments were implemented using the PyTorch machine learning library. The learning rate was set as 0.0001 and Adam optimizer was used to finetune all the model parameters. To alleviate the effect of overfitting problem to the model performance, we applied early stopping strategy to advancedly stop the iterations before the overfitting occurred [11].

Interpretability Analysis

To assess whether the proposed models captured more information, we performed T-SNE analysis on both feature-level and sample-level using the Rtsne package. We utilized Grad-CAM visualization method to visualize the area which the model has focused on.

Statistical Analysis

The data set was randomly split into the training set (160 GCs) and independent test set (54 GCs). For each deep learning model, we performed five times training and testing to assess the stability of the models, and the final performances were calculated using mean±sd. Once the predictions were made, Kaplan-Meier curve was plotted using the median value of the predicted risk score as cutoffs.

Results

Statistical description of clinical parameters

Table 1 presented the statistical description of clinical parameters among training set and validation set. Among the 214 GCs, there are 163 GCs obtaining relapse or metastasis after surgery. The DFS ranges from 7 to 177 months, with a medium value

of 26 months.

Model performance

We systematically compared the proposed framework with baseline models in Table 2. Radiomics model showed slightly better performance among the three baseline models. Lasso-based integration models did not improve the performance, while the proposed integration models showed improved C-index performance of 3%-5% in predicting DFS. Using the predicted score and the event label, we found that the DSN and CON model improved AUC of 10% in predicting event occurrence.

In order to identify the time-dependent property of the models, we performed time-dependent ROC analysis for each model. According to Fig.2 and Table3, the proposed CON or DSN models present higher AUC among different time points for relapse prediction, and especially when the time is 21 months (1.75 year), the two models could perform significantly better than radiomics model. Besides, we found DSN model could predict patient relapse at eight month, which indicates DSN model could identify particularly severe gastric cancer.

Kaplan-Meier curves **(Fig.3)** also proved that the proposed DSN and CON models showed good differentiation ability for high-risk and low-risk patients, while the radiomics and CNN model showed no statistical significance in differentiating high-risk and low-risk patients.

Interpretation analysis

Figure 4 present the TSNE visualization plot from feature-level (a-b) and patient-level (c-f). Compared with CNN-based deep features (Fig.4a), CON-based deep features (Fig.4b) distribute more densely and show higher accordance with radiomics feature space. While radiomics features distributed in a continuous way, deeply-learned features present different clusters far from each other, indicating that deep network potentially extract different patterns from tumor images. Fig.4c showed that the radiomics model split the patients into two groups, while the two groups showed no distinction between DFS time. Different from radiomics model, CNN, CON and DSN model showed significant trends from high DFS to low DFS (Fig.4d-f). However, the patient distributions are very dense in CNN and CON model while the DSN model explicitly separate different patients.

We performed visualization on the last layer of deep networks to reveal what properties have been learned. Grad-CAM visualization on one case with 23 months of DFS showed that, CNN model focused more on the background, while the CON-2m model and DSN-2m model focused more on the tumor lesion (Fig.5). Besides, DSN-2m focused a large part of the tumor while CON-2m focused only a small part of the tumor. This indicated that radiomics-based knowledge guided deep learning to

focus more information on tumor lesion, and combining radiomics loss with deep learning loss could further update the model weights to extract more meaningful deep features.

Discussion

In this study, we developed an integration model which intuitively embedded radiomics features and clinical features into deep learning model to predict the DFS of gastric cancer. We found our proposed model obtained an improvement of C-index of 3%-5%. Interpretation analysis revealed the addition of radiomics features helps deep learning extract useful information to predict disease progression. Our proposed models fully leverage the features extracted from clinical reports and single phase CT image, which potentially revealed inter-tumor heterogeneity of gastric cancer patients.

Our studies identified new radiomics biomarkers for preoperative DFS prediction. From Supplementary Table S1, tumor size and age are previously reported a biomarker for patient prognosis, and in our clinical model, the features were selected though with no statistical significance. In the model combining radiomics features with clinical features, only one clinical variable—age was selected. Nine radiomics features were selected in radiomics model, whereas the “ratio” feature was involved. Though the “ratio” feature was roughly calculated by dividing the two dimension image size, the feature potentially represent the morphology and structure of the tumor. As reported in previous studies, the morphology of cancer is related to the tumor growth patterns and is related to patient prognosis [12, 13]. Our study preliminarily implies the morphology indicated by CT image would be an important factor to DFS prediction of GC patients. However, when integrating deep features with radiomics features, the ‘ratio’ feature has not been selected. This may be due to the rough definition of the morphology. Further investigation is required in the future works.

The deep learning-based integration framework proposed presented several advantages. Firstly, the model can improve the feature extraction ability of deep learning and allieviate the few-shot learning problem. The Grad-CAM visualization plot and T-SNE plot at feature level indicated that the knowledge-guided deep learning model improved the feature extraction ability. Secondly, the model can potentially identify tumor heterogeneity and infer disease progression trajectory. T-SNE plot at sample level showed that the capability of deep learning based model in identifying disease progression trajectory. While the DSN model also presented disease progression trajectory, the deep features could also extracted the difference between different GCs, showing its potentials in identifying inter-tumor heterogeneity. Thirdly, the model is flexible and can integrate different knowledge. The proposed model is an extendable and flexible model which could be fitted to many different scenerios. To our best knowledge, the study is the first work to show that an

integration model integrating radiomics features, clinical features in a deep learning framework could guide deep learning to extract biologically meaning information for patient prognosis.

Beyond the discoveries made, our study has some limitations. Firstly, the samples included were limited, which greatly affect the generalization of deep learning model. Secondly, we have not adopted the data augmentation strategy to further alieviate the overfitting problem. Although we finetuned the models on pretrained ResNet model, our model still suffers from overfitting problem. Data augmentation would help further improve the model generalization and allieviate overfitting. Finally, although we performed interpretation on the meaning of the deep features learned by proposed model, we have not consider more in-depth investigation on how these deep features could be related to biological processes. Our future works will focus on interpreting these imaging-based deep features using paired RNA-seq data with images.

In conclusion, our study suggests that the guidance of radiomics features and clinical features improved the performance of deep learning model for the DFS prediction of gastric cancer. Our model would be extendable to related problems and would be an inspiration in dealing with few-shot medical image learning.

Acknowledgements

The authors are very grateful to the Zhejiang Cancer Hospital for providing CT images and clinical histopathology reports of gastric cancer for this study.

References

1. Thrift, A.P., T.N. Wenker, and H.B. El-Serag, *Global burden of gastric cancer: epidemiological trends, risk factors, screening and prevention.* Nature Reviews Clinical Oncology, 2023. **20**(5): p. 338-349.
2. Machlowska, J., et al., *Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies.* Int J Mol Sci, 2020. **21**(11).
3. Jiang, Y., et al., *Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer.* EBioMedicine, 2018. **36**: p. 171-182.
4. Wang, S., et al., *Preoperative computed tomography-guided disease-free survival prediction in gastric cancer: a multicenter radiomics study.* Med Phys, 2020. **47**(10): p. 4862-4871.
5. Hu, C., et al., *Deep learning radio-clinical signatures for predicting neoadjuvant chemotherapy response and prognosis from pretreatment CT images of locally advanced gastric cancer patients.* Int J Surg, 2023. **109**(7): p. 1980-1992.
6. Jiang, Y., et al., *Predicting peritoneal recurrence and disease-free survival from CT images in gastric cancer with multitask deep learning: a retrospective study.* The Lancet Digital Health, 2022. **4**(5): p. e340-e350.
7. Wei, L., et al., *A deep survival interpretable radiomics model of hepatocellular carcinoma patients.* Physica Medica, 2021. **82**: p. 295-305.
8. Wang, G., et al., *Prediction of Microvascular Invasion of Hepatocellular Carcinoma Based on Preoperative Diffusion-Weighted MR Using Deep Learning.* Acad Radiol, 2021. **28 Suppl 1**: p.

S118-s127.
9. Lee, C.-Y., et al., *Deeply-Supervised Nets*, in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, L. Guy and S.V.N. Vishwanathan, Editors. 2015, PMLR: Proceedings of Machine Learning Research. p. 562--570.
10. Katzman, J.L., et al., *DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network.* BMC Med Res Methodol, 2018. **18**(1): p. 24.
11. Yasaka, K., et al., *Deep learning with convolutional neural network in radiology.* Jpn J Radiol, 2018. **36**(4): p. 257-272.
12. Lee, J., et al., *Interfacial geometry dictates cancer cell tumorigenicity.* Nature Materials, 2016. **15**(8): p. 856-862.
13. Almagro, J., et al., *Tissue architecture in tumor initiation and progression.* Trends Cancer, 2022. **8**(6): p. 494-505.

Figure Legends

Figure 1. The proposed framework for integrating deep features, radiomics features and clinical features in a deep learning network

Figure 2. Time dependent ROC curves for different models

Figure 3. The comparion of KM curves for different models. (a) KM curve for radiomics model; (b) KM curves for baseline CNN model; (c) KM curves for CON-2m model; (d) KM curves for DSN-2m model.

Figure 4. TSNE visualization comparing the features extracted from the models. (a) TSNE plot comparing deep features with radiomics features; (b) TSNE plot comparing CON features with radiomics features; (c) TSNE for radiomics model; (d) TSNE for CNN model; (e) TSNE for CON model; (f) TSNE for DSN model

Figure 5. Grad-CAM visualization of the last layer features learned by the deep networks. (a) Grad- CAM plot of baseline CNN; (b) Grad-CAM plot of CON model; (c) Grad-CAM plot of DSN model.

Table Legends

Table 1: Statistical description of clinical parameters in training and validation set*

Table 2: Performance comparison between different models

Table 3: Comparison between time dependent AUC for different models

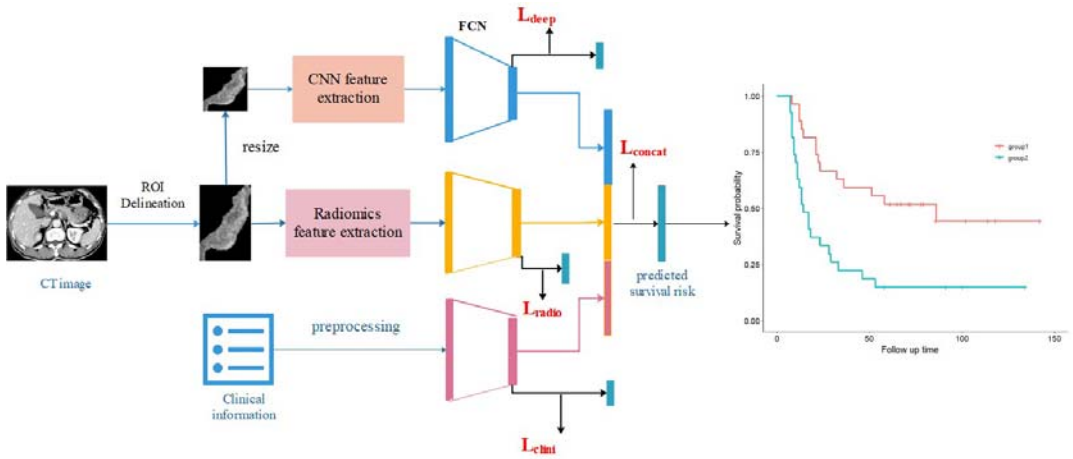

Figure 1

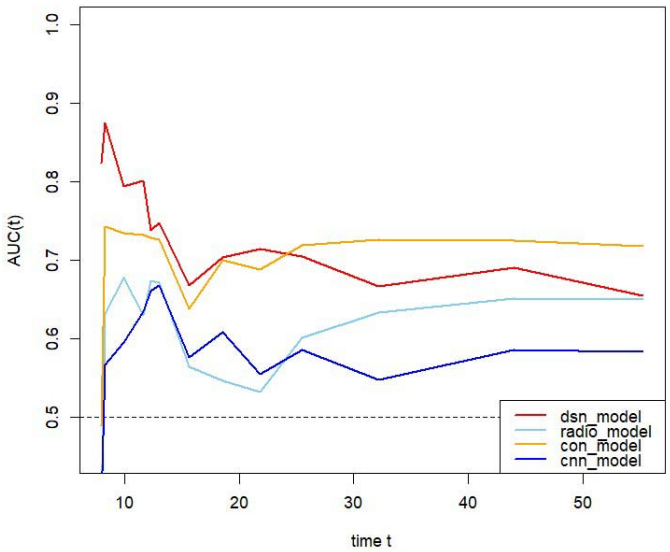

Figure 2

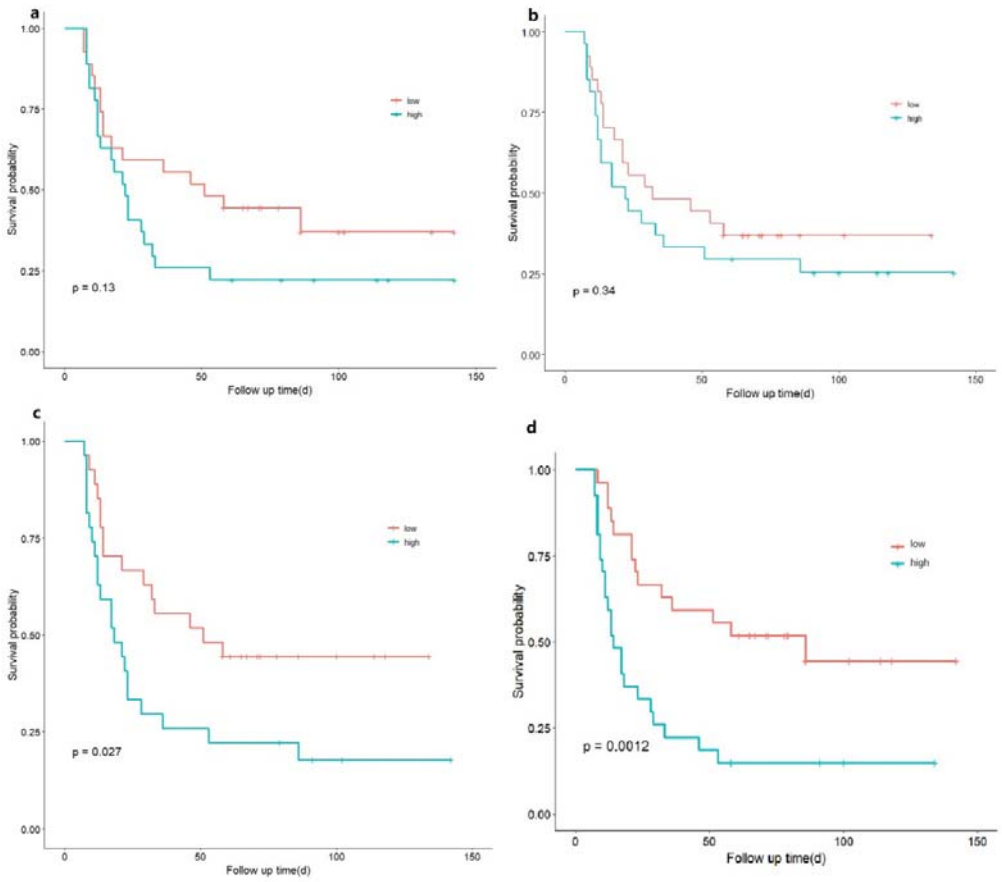
a
b
c
d
Survival probability
Follow up time(d)
low
high
p = 0.13
p = 0.34
p = 0.027
p = 0.0012
1.00
0.75
0.50
0.25
0.00
0
50
100
150

Figure 3

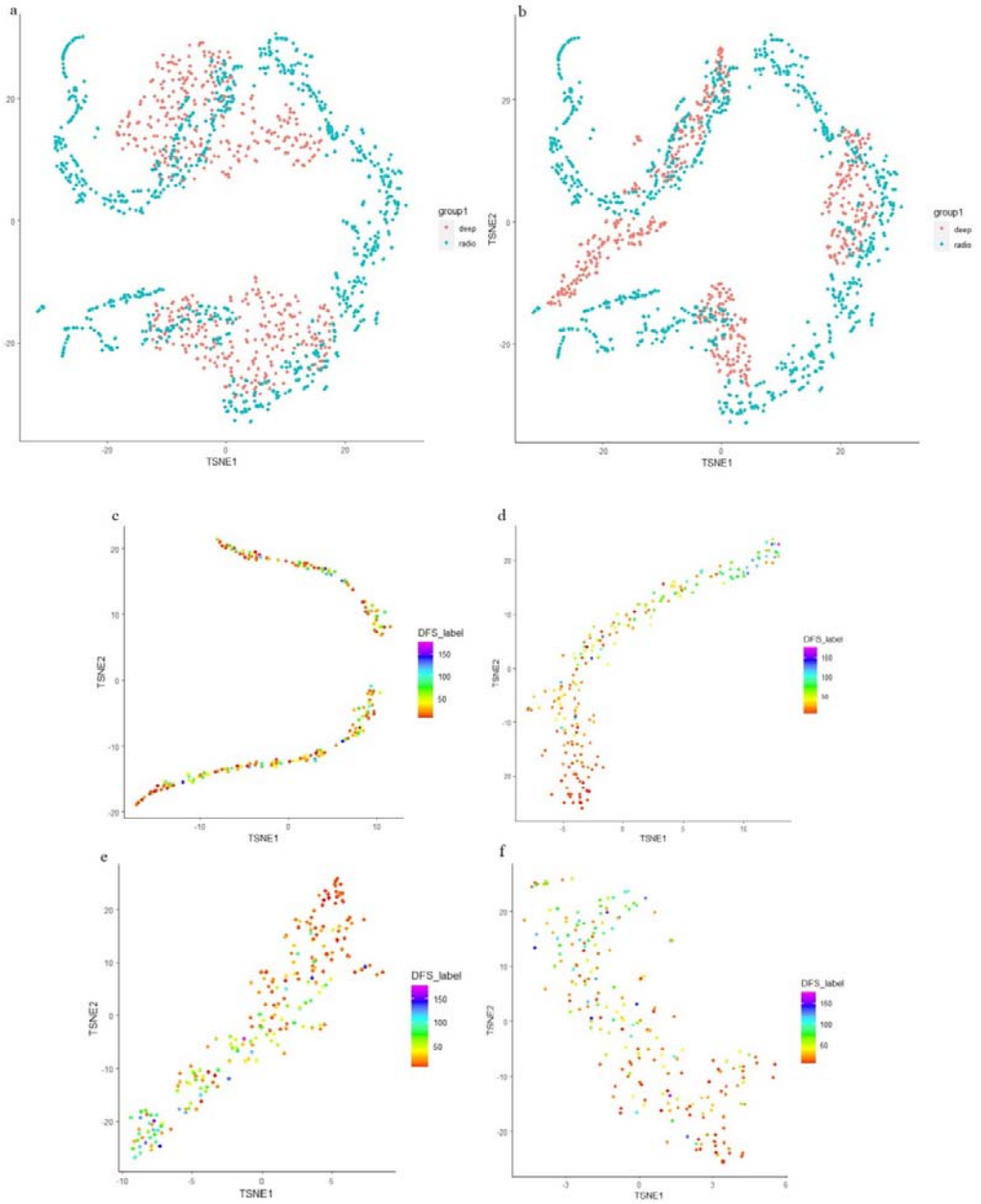

Figure 4

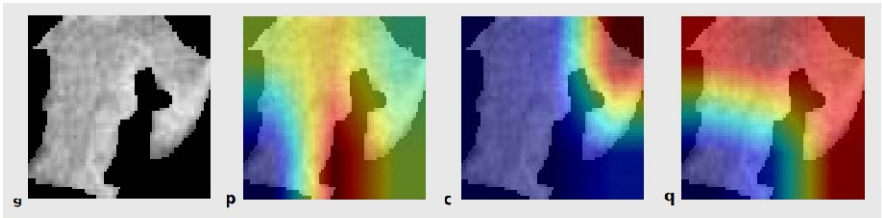

Figure 5