Predicting histopathological findings of gastric cancer via deep generalized multi-instance learning

Mengjie Fang, Wenjuan Zhang, Di Dong, Junlin Zhou, Jie Tian

Mengjie Fang, Wenjuan Zhang, Di Dong, Junlin Zhou, Jie Tian, "Predicting histopathological findings of gastric cancer via deep generalized multi-instance learning," Proc. SPIE 10949, Medical Imaging 2019: Image Processing, 109491Q (15 March 2019); doi: 10.1117/12.2512435

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Predicting Histopathological Findings of Gastric Cancer via Deep Generalized Multi-instance Learning

Mengjie Fang[a,b], Wenjuan Zhang[c], Di Dong[a,b], Junlin Zhou[c], Jie Tian[a,b,*]
[a] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Department of Radiology, Lanzhou University Second Hospital, Lanzhou 730000, Gansu, China

ABSTRACT

In this paper, we investigate the problem of predicting the histopathological findings of gastric cancer (GC) from preoperative CT image. Unlike most existing classification systems assess the global imaging phenotype of tissues directly, we formulate the problem as a generalized multi-instance learning (GMIL) task and design a deep GMIL framework to address it. Specifically, the proposed framework aims at training a powerful convolutional neural network (CNN) which is able to discriminate the informative patches from the neighbor confusing patches and yield accurate patient-level classification. To achieve this, we firstly train a CNN for coarse patch-level classification in a GMIL manner to develop several groups which contain the informative patches for each histopathological category, the intra-tumor ambiguous patches, and the extra-tumor irrelative patches respectively. Then we modify the fully-connected layer to introduce the latter two classes of patches and retrain the CNN model. In the inference stage, patient-level classification is implemented based on the group of candidate informative patches automatically recognized by the model. To evaluate the performance and generalizability of our approach, we successively apply it to predict two kinds of histopathological findings (differentiation degree [two categories] and Lauren classification [three categories]) on a dataset including 433 GC patients with venous phase contrast-enhanced CT scans. Experimental results reveal that our deep GMIL model has a powerful predictive ability with accuracies of 0.815 and 0.731 in the two applications respectively, and it significantly outperforms the standard CNN model and the traditional texture-based model (more than 14% and 17% accuracy increase).

Keywords: Multi-instance learning, Convolutional neural network, Classification, Gastric cancer

1. INTRODUCTION

Gastric cancer (GC) is the fourth most common human malignant disease and the second leading cause of cancer-related death worldwide.[1] The histopathological findings of GC, including differentiation degree and Lauren classification, is closely related to the patients' prognosis. Accurate preoperative histopathological assessment can facilitate selection of the optimal treatment strategy.[2-3] Non-invasive contrast-enhanced CT scan is recommended by the National Comprehensive Cancer Network (NCCN) guidelines as a routine preoperative imaging examination for GC in clinical practice. Several previous studies have been performed on the CT-based inter-tumor characteristics measurement and

* Corresponding author: Jie Tian, email: tian@ieee.org

Medical Imaging 2019: Image Processing, edited by Elsa D. Angelini, Bennett A. Landman, Proc. of SPIE
Vol. 10949, 109491Q · © 2019 SPIE · CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2512435

their correlation to GC histopathological findings.[4-5] However, due to the relative poor soft-tissue contrast of CT images and the complex context around the GC, CT-based prediction of the histopathological findings is a challenge problem, and there is still lack of an efficient classification approach.

Deep learning methods, especially convolutional neural network (CNN) models, have shown significant ability of learning predictive feature representations from medical images.[6] But in the applications where the training set size is limited and the image textural phenotype is poor, such as the diagnosis of GC and colorectal cancer, this kind of data-driven methods haven't achieved satisfied performances.[7] The imaging phenotype in a fixed patch depends on not only the lesion's own local attribute but its posture, location and the neighbor context in the patch. We hypothesis that the extracted patches hold various amount of information about the classification object, and predicting based on the most informative patches would be helpful. While only patient-level annotations are provided, a weakly supervised learning method should be developed.

In this paper, we propose a deep generalized multi-instance learning (deep GMIL) framework to select the informative patches and leverage them to achieve better classification performance. The experimental results reveal that our approach can estimate the probability of each extracted patch being informative, ambiguous or irrelative (patch-level), and yield an accurate prediction of histopathological findings (patient-level).

2. METHODS

2.1 Deep Generalized Multi-instance Learning

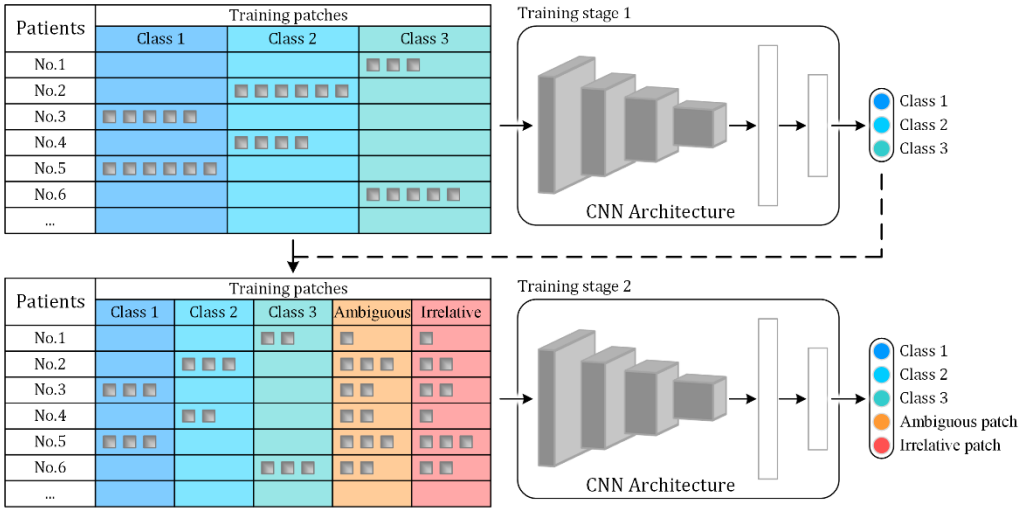

Figure 1. Illustration of the training stages in deep GMIL framework for three-category classification

In this study, GMIL is intended to identify the label of each patch (i.e. the instance) extracted from the CT scan (i.e. the bag) and achieve an accurate final classification based on the informative patches. Please note that the standard multi-instance learning considers there is only one discriminative instance in a positive bag[8-9], which is not suitable for our issue. Since there is large intra-class variation and inter-class ambiguity in the CT imaging phenotype of GC, a group of informative patches should be taken into account to improve the performance. The proposed deep GMIL framework consists of two stages of model training (as shown in Figure 1) and one inference stage.

For the training stages, we extract 32×32 patches from each CT slice as much as possible with 16-pixel step size and ensure that central points of these patches are in the lesion, which are manually segmented by an experienced radiologist.

In the training stage 1, we define the loss function of CNN model as a combined cross-entropy, which counts only the instances with the highest or lowest probability to be correctly classified within each bag. It's formulated as:

$$L_1 = -(1/B) \times \sum_{b=1}^{B} \sum_{c=1}^{C_1} g_c^b \left(\log\left[\max_{i \in I_b}\{p_c^{b,i}\}\right] + \log\left[\min_{i \in I_b}\{1 - 1/C_1 + p_c^{b,i}, 1\}\right]\right) \quad (1)$$

where B is the number of the bags in a training batch, I_b is the instances in bth bag, C_1 is the number of classification categories, $g_c^b \in [0,1]$ and $p_c^{b,i} \in (0,1)$ are respectively the ground truth and the predicted result. We use this design to eliminate the imbalance influence from the various lesion size and to enhance the feature representation of the informative patches. Then we develop several groups containing the informative patches for each category (i.e. with probability $> \alpha$ to be correctly classified, α is set to $1/C_1$ in this study) and the intra-tumor ambiguous patches (the rest patches). We also develop a group containing the extra-tumor irrelative patches by extracting several patches adjacent to the lesion to guide the model to recognize other tissues.

In the training stage 2, we use the patch groups developed above to retrain the CNN model. As there are two additional groups in this stage, the number of classification categories C_2 is two larger than C_1. The standard cross-entropy is used as the loss function:

$$L_2 = -(1/N) \times \sum_{n=1}^{N} \sum_{c=1}^{C_2} g_c^n \log(p_c^n) \quad (2)$$

where N is the number of the patches in a training batch.

For the inference stage, considering the accessibility and robustness needed in the clinical practice, we ask the radiologist to mark patches' central points with 16-pixel interval at least in the lesion, and ensure the extracted patches completely cover the lesion. The predicted category of the instance i in the bth bag is $P_{b,i} = \text{argmax}_{c \in \{1,2,\cdots,C_2\}} p_c^{b,i}$. An informative patch group I_b^* is then developed using the patches with $P_{b,i^*} \in \{1,2,\cdots,C_1\}$. We combine the class probabilities of the informative patches to derive the final bag-level (i.e. patient-level) category: $P_b = \text{argmax}_{c \in \{1,2,\cdots,C_1\}} \sum_{i \in I_b^*} p_c^{b,i}$.

2.2 Dataset

To investigate the performance of our deep GMIL framework for predicting the histopathological findings of GC, we evaluate it on two kinds of clinical classification systems: 1) differentiation degree (two categories); 2) Lauren classification (three categories). A dataset consisting of 433 GC patients was collected from Lanzhou University Second Hospital. The venous phase contrast-enhanced CT volumes were acquired on one of the two multi-detector row CT scanners (Discovery CT 750 HD, GE Healthcare or Sensation 64, Siemens). In our experiments, the CT slices are resampled to 0.5 mm/pixel in the two axes.

In our experiment, 60% and 10% of the patients are randomly selected and set aside as the training set and validation set respectively. The remaining 30% are used as the test set. The numbers of patients with different histopathological findings in each set are listed in Table 1. For the training set, we select three slices with the largest lesion regions from each CT volume and augment them 2 times with random translations and rotations. We treat each single slice as a sample and extract the patches using the method for the training stage. Around 36,000 training patches are extracted. For the validation set and test set, one slice with largest lesion region is selected from each CT volume and

we extract patches from these slices using the method for the inference stage.

Table 1. Numbers of patients in each set

Characteristics		Training set	Validation set	Test set
Differentiation degree (application 1)	Poorly	127	21	64
	Well/Moderately	133	22	66
Lauren type (application 2)	Intestinal	103	16	46
	Diffuse	86	15	43
	Mixed	71	12	41

2.3 Implementation Details

In our experiment, the CNN inserted into the proposed deep GMIL framework contains three convolutional layer groups, of which input patches have size 32×32, 16×16 and 8×8 respectively, and two fully-connected layers, the former of which has 32 nodes. Each convolutional layer group is composed of two convolutional layers and a 2×2 max pooling layer. Each convolutional layer has 16 3×3 filters with a stride of 2 pixels. Batch normalization is used after convolution. The rectified linear units is used as nonlinearity activation function. We initialize all parameters with random Gaussian distributions and train the network with mini-batch size of 32 bags using Adam Optimizer with a learning rate of 0.001. We stop training when the network performance do not significantly improve on the validation set for 10 epochs.

We compare the proposed method with a standard CNN model and a texture-based model. The standard CNN model has the same structure with our deep GMIL model except that its final fully-connected layer has two less nodes. We train it using the method for the training stage 2 descried above. Then it inferences the test set based on the patches with the highest predicted probability.

To build the texture-based model, we extract 90 hand-crafted texture features defined in our previous study.[10-11] Support vector machine with radial basis function kernel is adopted, and its hyper-parameters are set via grid search while the input features are selected using the forward stepwise selection strategy.

3. RESULTS

Example of the extracted patches used in the training step 2 are shown in Figure 2, we could find that the informative patches usually contain more textural tumor tissues while there are more edges in the ambiguous patches. The adjacent organs, blood vessels, lymph nodes, fat tissue and gastric contents etc. are extracted as the extra-tumor irrelative patches. Table 2 lists the results of our approach and the compared models in the experiments. The proposed deep GMIL model achieves promising performance on both applications, all outperforming the standard CNN model and the traditional texture-based model. The predictive accuracy of our approach outperforms that of the standard CNN model by 15.11% and 14.58% on the two applications respectively. The superiority suggests that the deep GMIL framework is able to extract the discriminative information, while reducing irrelevant noise and redundant information.

When compared with the traditional texture-based model, 17.78% and 31.95% accuracy increase reveals the power of significant feature representation of our deep GMIL model. Specifically, as the CT imaging phenotype of Lauren classification is relatively obscure, there is only indirect association between this kind of clinical categories and the common hand-crafted features making it a difficult task to build a useful traditional model. Meanwhile, leveraging the

informative image patches and class-specific quantification of image characteristics, the proposed method has greater ability of efficiently modeling for such challenging problem.

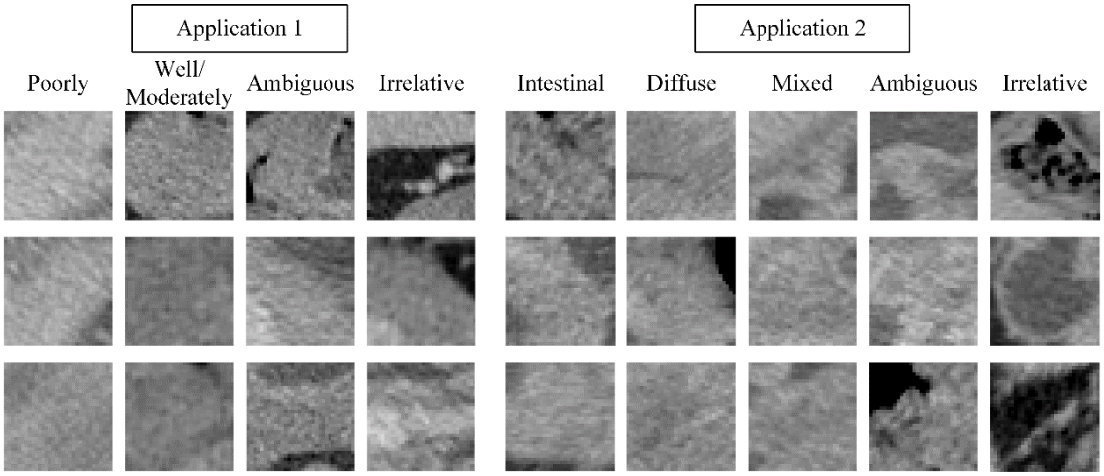

Figure 2. Extracted patches of different groups

Table 2. Predictive results of each model on the test set

Models	Application 1		Application 2	
	Predictive result[a]	Accuracy	Predictive result[b]	Accuracy
Texture-based model	43/47	0.692	29/20/23	0.554
Standard CNN model	47/45	0.708	26/28/29	0.638
Deep GMIL model	54/52	0.815	31/35/29	0.731

Note. a: Numbers of patients whose Differentiation degree are predicted as “Poorly” or “Well/Moderately”; b: Numbers of patients whose Lauren type are predicted as “Intestinal”, “Diffuse” or “Mixed”.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

There is lack of research on building image-based model for predicting the histopathological findings of GC. Our formulation of this approach using deep GMIL is novel, and outperforms the standard CNN model and the traditional texture-based model. Numerical results on two classification tasks not only indicate the powerful discriminative ability of our approach, but also imply its potential for being applied to other diagnostic systems.

5. CONCLUSIONS

In this paper, we propose a deep GMIL framework to preoperatively predict the histopathological findings of GC by building a powerful CNN model using discriminative features extracted from the automatically recognized informative patches. We evaluate the method on two kinds of clinical classification systems. The experimental results show that our method is effective, accurate and universal for the CT-based prediction, and performs significantly better than the two compared methods. Future work will focus on optimizing the framework parameters and applying it on more applications.

This work has not been submitted for publication or presentation elsewhere.

6. ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2017YFA0205200, 2017YFC1308700, and 2017YFC1309100), National Natural Science Foundation of China (81771924, 81501616 and 81527805), the Beijing Natural Science Foundation (L182061), the Beijing Municipal Science and Technology Commission (Z171100000117023 and Z161100002616022), the Bureau of International Cooperation of Chinese Academy of Sciences (173211KYSB20160053), and the Youth Innovation Promotion Association CAS (2017175).

REFERENCES

[1] Fitzmaurice, C., Dicker, D., Pain, A., Hamavid, H., Moradi-Lakeh, M., MacIntyre, M.F., Allen, C., Hansen, G., Woodbrook, R., Wolfe, C. and Hamadeh, R.R., "The global burden of cancer 2013," JAMA Oncology 1(4), 505-527 (2015).

[2] Zu, H., Wang, H., Li, C. and Xue, Y., "Clinicopathologic characteristics and prognostic value of various histological types in advanced gastric cancer," International Journal of Clinical and Experimental Pathology 7(9), 5692 (2014).

[3] Qiu, M.Z., Cai, M.Y., Zhang, D.S., Wang, Z.Q., Wang, D.S., Li, Y.H. and Xu, R.H., "Clinicopathological characteristics and prognostic analysis of Lauren classification in gastric adenocarcinoma in China," Journal of Translational Medicine 11(1), 58 (2013).

[4] Liu, S., Liu, S., Ji, C., Zheng, H., Pan, X., Zhang, Y., Guan, W., Chen, L., Guan, Y., Li, W. and He, J., "Application of CT texture analysis in predicting histopathological characteristics of gastric cancers," European Radiology 27(12), 4951-4959 (2017).

[5] Tsurumaru, D., Miyasaka, M., Muraki, T., Nishie, A., Asayama, Y., Oki, E. and Honda, H., "Histopathologic diversity of gastric cancers: Relationship between enhancement pattern on dynamic contrast-enhanced CT and histological type," European Journal of Radiology 97, 90-95 (2017).

[6] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H. and Aerts, H.J., "Artificial intelligence in radiology," Nature Reviews Cancer 18, 500–510 (2018).

[7] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B. and Sánchez, C.I., "A survey on deep learning in medical image analysis," Medical Image Analysis 42, 60-88 (2017).

[8] Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Zhang, S., Metaxas, D.N. and Zhou, X.S., "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," IEEE transactions on Medical Imaging 35(5), 1332-1343 (2016).

[9] Hering, J., Kybic, J. and Lambert, L., "Detecting multiple myeloma via generalized multiple-instance learning," in Proceedings of SPIE Medical Imaging 2018: Image Processing 105740N (2018).

[10] Cheng, S., Fang, M., Cui, C., Chen, X., Yin, G., Prasad, S.K., Dong, D., Tian, J. and Zhao, S., "LGE-CMR-derived texture features reflect poor prognosis in hypertrophic cardiomyopathy patients with systolic dysfunction: preliminary results," European Radiology 28(11), 4615-4624 (2018).

[11] Fan, L., Fang, M., Li, Z., Tu, W., Wang, S., Chen, W., Tian, J., Dong, D. and Liu, S., "Radiomics signature: A biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule," European Radiology 29(2), 889-897 (2019).