Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2019 April ; 2019: 1303–1306. doi:10.1109/ISBI.2019.8759512.

COLLABORATIVE CLUSTERING OF SUBJECTS AND RADIOMIC FEATURES FOR PREDICTING CLINICAL OUTCOMES OF RECTAL CANCER PATIENTS

Hangfan Liu¹, Hongming Li¹, Pamela Boimel², James Janopaul-Naylor², Haoyu Zhong², Ying Xiao², Edgar Ben-Josef², Yong Fan¹

¹Center for Biomedical Image Computing and Analysis, Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

²Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

Abstract

Most machine learning approaches in radiomics studies ignore the underlying difference of radiomic features computed from heterogeneous groups of patients, and intrinsic correlations of the features are not fully exploited yet. In order to better predict clinical outcomes of cancer patients, we adopt an unsupervised machine learning method to simultaneously stratify cancer patients into distinct risk groups based on their radiomic features and learn low-dimensional representations of the radiomic features for robust prediction of their clinical outcomes. Based on nonnegative matrix tri-factorization techniques, the proposed method applies collaborative clustering to radiomic features of cancer patients to obtain clusters of both the patients and their radiomic features so that patients with distinct imaging patterns are stratified into different risk groups and highly correlated radiomic features are grouped in the same radiomic feature clusters. Experiments on a FDG-PET/CT dataset of rectal cancer patients have demonstrated that the proposed method facilitates better stratification of patients with distinct survival patterns and learning of more effective low-dimensional feature representations that ultimately leads to accurate prediction of patient survival, outperforming conventional methods under comparison.

Keywords

Collaborative clustering; unsupervised learning; nonnegative matrix tri-factorization; radiomics; patient stratification; rectal cancer

1. INTRODUCTION

Rectal cancer is a major cause of tumor-related deaths in the US. One of the most widely-used treatments for rectal cancer is chemoradiation therapy (CRT). However, only about 15~20% tumors treated by CRT attain pathologic complete response (pCR), and most of the others achieve different degrees of partial response [1]. Patients with pCR expect optimistic outcomes, whereas those without pCR have a high risk of developing local recurrences and distant metastasis. For better treatment planning, it is desired to early predict treatment response of individual patients.

Recent radiomic studies on rectal cancer have shown promising results on cancer prognosis by building prediction models on imaging features computed from radiological imaging data, providing an alternative non-invasive means to conventional cancer prognosis based on clinical factors [2–5]. The general radiomic framework consists of feature extraction, feature selection or dimension reduction, and prediction modeling [6, 7]. Since radiomic studies often have limited samples that are characterized by high dimensional radiomic features, feature selection and dimensionality reduction are typically adopted to build robust prediction models for improving the prediction accuracy [8, 9].

Typical feature selection methods identify discriminative features via optimizing prediction models using validation datasets. Obviously, when the sample size is small, such supervised learning techniques may produce models that overfit the training data. Feature dimensionality reduction techniques, such as principal component analysis (PCA), learn new feature presentations in a low dimension space. However, these unsupervised learning approaches lack relevant guidance in the feature learning, hence the generated low-dimensional representations might not be sufficiently informative for building prediction models.

Furthermore, conventional feature selection and feature dimensionality reduction techniques for prediction of treatment response ignore the potentially huge gap in radiomic features of heterogeneous groups of patients. Since it is highly likely radiomic features computed from patients with distinctive clinical outcomes are different and high dimensional features contain redundant correlated information, simultaneously clustering patients and their high dimensional features may improve both the patient stratification and dimensionality reduction. In fact, joint clustering of patients and radiomic features has demonstrated promising performance in radiomic studies of lung cancer [10].

Grounded on the discussions above and inspired by matrix factorization based clustering techniques [11], we propose an unsupervised collaborative clustering analysis approach for predicting overall survival of rectal cancer patients based on their radiomic features computed from radiological imaging data collected before treatment. Rather than clustering all radiomic features obtained from different subjects as a whole, we consider the underlying difference between patients and their corresponding features, and simultaneously cluster the radiomic features and the patients via orthogonal nonnegative matrix tri-factorization, so as to achieve improved low-dimensional representation of radiomic features as well as patient stratification.

2. METHODS

The proposed collaborative clustering method is based on tri-factorization of the radiomic feature matrix for patient stratification and computing meta-features that are used to build prediction models for predicting clinical outcomes of patients. Fig. 1 illustrates the overall framework of the present study.

2.1 Collaborative Clustering

In order to achieve collaborative clustering of radiomic features and patients, this study adopts a nonnegative matrix tri-factorization approach [11]. The radiomic features of all the patients are contained in matrix $Y \in \mathbb{R}^{p \times f}$, where p is the number of patients and f denotes the number of radiomic features. The tri-factorization procedure decomposes nonnegative matrix Y into three matrices Φ , X , and Θ via minimizing:

$$\begin{aligned} \min_{\Phi \geq 0, X \geq 0, \Theta \geq 0} & \|Y - \Phi X \Theta\|_F^2 \\ \text{s.t. } & \Phi^T \Phi = I, \Theta^T \Theta = I, \end{aligned} \quad (1)$$

where I is an identity matrix, $\Phi \in \mathbb{R}_+^{p \times k_p}$ encodes the relationship of the patients and the corresponding k_p patient groups, $\Theta \in \mathbb{R}_+^{k_F \times f}$ encodes the mapping between the features and the corresponding k_F feature clusters, while $X \in \mathbb{R}_+^{k_p \times k_F}$ reflects the magnitude of mappings upon different dimensions and the interactions between Φ and Θ . In fact, Φ and Θ can be seen as two sets of orthonormal basis used to decompose the feature matrix, and X contains the corresponding low dimensional representation. The combination of Φ and X could serve as a low dimensional representation of Y , referred to as meta-features hereafter. The optimal number of feature clusters k_F can be evaluated by gap criterion [12], which aims to find solutions with the largest local or global gap value within a tolerance range. The present study uses the correlation between features as the distance metric to compute the gap statistic.

The objective function (1) can be optimized efficiently by an alternative optimization method as elaborated in [11]. After the matrix tri-factorization is performed, it is easy to calculate the meta-feature matrix $M \in \mathbb{R}_+^{p \times k_F}$ as

$$M = \Phi X. \quad (2)$$

Such meta features are essentially weighted combinations of the radiomic features in the same clusters. In this way, we reduce the feature dimensionality with discriminative information conveyed in the original features, and improve the robustness against feature noise. The meta features are utilized in prediction modeling to predict clinical outcomes.

In such a collaborative approach, both the patient stratification and feature dimensionality reduction are jointly improved. Particularly, the feature dimensionality reduction helps reduce redundant information of the original features to improve the patient grouping/stratification, while the patient grouping/stratification helps the feature dimensionality reduction to compute discriminative meta-features for predicting clinical outcomes.

2.2. Patient Stratification and Prediction of Clinical Outcomes

Based on the collaborative clustering results, we can stratify the patients and predict their overall based on the meta-features. The present study adopts Kaplan-Meier estimator [13] to estimate survival functions of different groups of patients in terms of survival, and employs the log-rank test [14] to compare the survival functions of different patient groups for evaluating the patient stratification performance.

To predict clinical outcomes for individual patients, we build prediction models on their meta features using three survival modeling methods, including Cox proportional hazard regression (Cox) [15], Cox with LASSO (CoxL) [16], and random survival forests (RSF) [17]. We train and test the prediction models using the same 5-fold cross-validation, and the concordance index (C-index) [18] is used to measure the performance of the prediction models. We repeat the cross-validation procedure 100 times and obtain average performance scores.

2.3. Implementation of the Proposed Method

We evaluated the proposed method based on a dataset of 83 rectal cancer patients treated by CRT for locally advanced rectal cancer. All the patients had pre-treatment FDGPET/CT scans, and 8 patients deceased within a median follow-up of 3 years. The tumors were manually contoured by experienced radiologists. Standardized uptake values (SUV) were computed for the PET scans. We computed shape and texture features of tumors from the CT scans and SUV maps respectively using a radiomic feature extraction method [19]. Different extraction parameters including wavelet band-pass filtering and quantization of gray levels were adopted. The number of CT radiomic features was 1249, and the number of PET radiomic features was 1254. All these features were pooled together for patient stratification and prediction modeling.

The number of feature clusters k_F was set to 11 according to feature correlation based gap criterion [12], and the number of patient groups k_P was empirically set to 2 and 3, with the setting of 2 groups for stratifying patients into a low-risk group and a high-risk group respectively in terms of mortality, while the setting of 3 groups for stratifying patients into low-risk, medium-risk, and high-risk groups respectively. We display the clustering result with 3 patient groups and 11 feature clusters in Fig. 2, demonstrating that highly-correlated radiomic features were grouped into the same clusters, while patients with similar radiomic features were also grouped together. The blocks along the diagonals of the Pearson correlation matrices shows relatively higher correlation coefficients within the same groups/clusters than those between groups/clusters.

The R packages *survival*, *glmnet* and *randomForestSRC* were employed to build the prediction models. The sparsity parameter in CoxL was tuned by a nested 3-fold cross validation, and 500 decision trees were used in the RSF model with the minimum leaf size of trees set to 5.

3. RESULTS

We evaluated the performance of the proposed method by comparing it with PCA based feature dimensionality reduction for building prediction models.

3.1. Patient Stratification

To evaluate the patient stratification performance, we obtained survival functions of patient groups obtained by the proposed collaborative clustering (CC) method in terms of overall survival. Fig. 3 shows survival functions of patient groups obtained at different settings of the number of patient groups, and Table 1 summarizes p -values (log-rank test) between different groups stratified by the proposed method. When $k_p = 2$, the difference in the survival functions between the two groups obtained by the proposed method was statistically significant with a p -value of 0.035. As shown in Fig. 3 (top), an evident gap between the red line and the green line indicates a clear separation of the high-risk group from the low-risk group. When $k_p = 3$ (bottom of Fig. 3), the survival functions of Group 2 and Group 3 were significantly different with a p -value of 0.0032 although the differences between other groups were not statistically significant.

3.2. Prediction of Overall Survival

To evaluate the prediction performance of low-dimensional meta features obtained by the proposed collaborative clustering method from the original radiomic features, prediction models were built on the meta features to predict overall survival of individual patients. For comparison, we also built prediction models based on low dimensional meta features learned using PCA from the radiomic features. Table 2 summarizes prediction performance of different prediction models measured by C-index. For the PCA based prediction, the best results were obtained by tuning the number of meta features within a range of 5~13. For the CC based prediction, the results were obtained with $k_p = 3$ and $k_F = 11$. The results summarized in Table 2 indicated that prediction models built on meta features obtained by the proposed method outperformed those built upon the PCA features. The best performance was achieved by the Cox prediction model built on meta-features obtained by the proposed approach, with a gain of ~7% over the corresponding PCA based counterparts. The results demonstrated that PCA based feature dimensionality reduction did not improve the prediction models built on the original radiomic features without feature dimensionality reduction, indicating that unsupervised learning approaches like PCA often lack relevant guidance in learning meta features, and the produced low-dimensional representations might not be sufficiently informative for the prediction models.

4. CONCLUSIONS

This paper presents a collaborative clustering approach for stratifying patients based on radiomic features and learning informative low dimensional representations of the radiomic features for predicting clinical outcomes of rectal cancer patients. Different from conventional feature dimensionality reduction methods, such as PCA, the collaborative clustering method considers the potential gap in radiomic features of heterogeneous patient groups, and utilizes an orthogonal nonnegative matrix tri-factorization technique to cluster

highly-correlated features and patients simultaneously. Quantitative evaluation measures of both patient stratification and prediction of overall survival have demonstrated the proposed approach could achieve promising performance in radiomic studies.

ACKNOWLEDGEMENTS

This study was supported in part by National Institutes of Health grants [CA223358, EB022573].

5. REFERENCES

- [1]. Maas M, et al., Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncology*, 2010 11(9): p. 835–844. [PubMed: 20692872]
- [2]. Joye I, et al., Can clinical factors be used as a selection tool for an organ-preserving strategy in rectal cancer? *Acta Oncologica*, 2016 55(8): p. 1047–1052. [PubMed: 27142064]
- [3]. Nie K, et al., Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clinical Cancer Research*, 2016 22(21): p. 52565264.
- [4]. Dinapoli N, et al., Radiomics for rectal cancer. *Translational Cancer Research*, 2016 5(4): p. 424–431.
- [5]. Li H, et al. Deep Convolutional Neural Networks for Imaging Based Survival Analysis of Rectal Cancer Patients. *International Journal of Radiation Oncology• Biology• Physics*. 2017 99(2): p. S183.
- [6]. Gillies RJ, et al., Images are more than pictures, they are data. *Radiology*. 2015 278(2): p. 563–77. [PubMed: 26579733]
- [7]. Kumar V, et al., Radiomics: the process and the challenges. *Magnetic Resonance Imaging*, 2012 30(9), p.1234–1248. [PubMed: 22898692]
- [8]. Peng H, et al. Feature selection by optimizing a lower bound of conditional mutual information. *Information Sciences*, 2017 418: p. 652–667. [PubMed: 30283157]
- [9]. Peng H, et al. A general framework for sparsity regularized feature selection via iteratively reweighted least square minimization. In *AAAI Conference on Artificial Intelligence*, 2017 p. 2471–2477.
- [10]. Li H, et al. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol*. 2018 129 (2), p. 218–226. [PubMed: 30473058]
- [11]. Ding C, et al., Orthogonal nonnegative matrix tri-factorizations for clustering. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006 p. 126135.
- [12]. Tibshirani R, et al., Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 2001 63(2): p. 411–423.
- [13]. Kaplan EL et al., Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 1958, 53(282): p. 457–481.
- [14]. Mantel N, Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 1966, 50: p. 163–70. [PubMed: 5910392]
- [15]. Fox J, Cox proportional-hazards regression for survival data. *An R and S-PLUS Companion to Applied Regression*, 2002.
- [16]. Tibshirani R The lasso method for variable selection in the Cox model. *Statistics in Medicine*, 1997 16(4): 385–95. [PubMed: 9044528]
- [17]. Ishwaran H, et al., Random survival forests. *The Annals of Applied Statistics*, 2008 2(3): p.841–860.
- [18]. Harrell FE, et al., Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, 1984 3(2): p. 143–152. [PubMed: 6463451]
- [19]. Vallieres M, et al., A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology*, 2015 60(14): p. 5471–96. [PubMed: 26119045]

Fig. 1.
Overall framework of the present radiomic study.

Fig. 2. Visualization of the collaborative clustering results. From top to bottom and left to right: Clustering results overlaid on the feature matrix of all patients; the Pearson correlation matrix between features (left); the Pearson correlation matrix between patients (right). Different sub-clusters are separated by red lines.

Fig. 3. Kaplan-Meier plots of patient groups stratified by the proposed approach regarding overall survival.

Table 1.

The p -values between different groups stratified by the proposed method.

	$k_p=2$	$k_p=3$
Group 1 vs. Group 2	0.035	0.0711
Group 3 vs. Group 2	--	0.0032
Group 1 vs. Group 3	--	0.8624

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Prediction performance (C-index) comparison between the proposed collaborative clustering (CC) approach and the PCA approach.

methods	C-index
Cox	0.6322
CoxL	0.6086
RSF	0.5378
<i>PCA</i> + Cox	0.6291
<i>PCA</i> + CoxL	0.5827
<i>PCA</i> + RSF	0.4978
CC + Cox	0.6731
<i>CC</i> + CoxL	0.6663
<i>CC</i> + RSF	0.5486