Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction

Chunfeng L., Su R., Thierry D., Fabrice J., Pierre V.

Haojie LU, Haifei ZHANG, Zinan ZHOU

Université de Technologie de Compiègne

4 November 2019

Outline

Outline

Introduction

In this paper, it proposes a prediction system primarily using radiomic features extracted from FDG-PET images.

- The solid application is still hampered by some practical difficulties:
  - Abounding features
  - The number of observations is less than input features
  - Input features are irrelevant with the outcome label
- Conventional feature selection
  - Seriously imbalanced training set
  - High overlapping or noisy training set
- Evidential Feature Selection (EFS)
- Improved Evidential Feature Selection (iEFS)

Outline

Main protocol

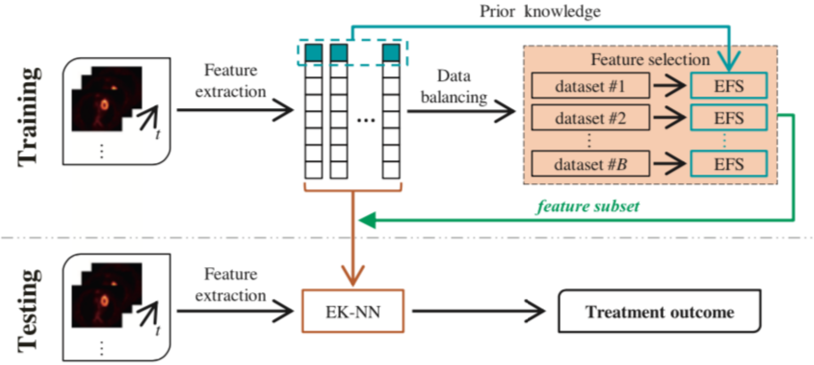

Fig. 1. Protocol of the prediction system.

Outline

Feauture extraction

Three types of PET imaging features are quantified

- SUV-based features :
  SUVmin, SUVmax, SUVpeak, MTV and TLG
- Texture features :
  Gray Level Size Zone Matrix(GLSZM)
- Temporal changes of image features :
  Relative difference between the baseline and the follow-up PET acquisitions as additional features, $\Delta f = (f_t - f_0)/f_0$

Dempster-Shafer theory

- Let ω be a variable taking one and only one value in a finite set $\Omega = \{\omega_1, \ldots, \omega_c\}$, called frame of discernment
- Evidence (uncertain information) about ω can be represented by a mass function $m : 2^{\Omega} \rightarrow [0, 1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1$$

- Belief and plausibility functions

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

$$Pl(A) = \sum_{B \bigcap A \neq \emptyset} m(B)$$

Dempster's rule

- Two mass functions are fused via the the Transferable Belief Model (TBM) conjunctive rule

$$m_{1\cap 2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C)$$

- If $m_{1 \bigcap 2}(\emptyset) < 1$, the new mass function obtained by Dempster's rule can be represented as

$$m_{1\oplus 2}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{m_{1\cap 2}(A)}{1 - m_{1\cap 2}(\emptyset)} & \text{otherwise} \end{cases}$$

Evidential feature selection

Let $\{(X_i, Y_i) \,|\, i = 1, \ldots, N\}$ be a collection of N training pairs, where $X_i = \left[x_{i,1}, \ldots, x_{i,V}\right]^T$ is the ith training instance with V features, and $Y_i \in \{\omega_1, \ldots, \omega_c\}$ is the corresponding class label.

- Weighted Euclidian distance (dissimilarity)

$$d_{i,j} = \sqrt{\sum_{p=1}^{V} \lambda_p d_{ij,p}^2}$$

where $d_{ij,p} = \left|x_{i,p} - x_{j,p}\right|$ and features are selected via changing the value of the ***binary vector*** $\Lambda = [\lambda_1, \ldots, \lambda_V]^T$

Evidential feature selection

- For X_i, the evidence offered by the training sample $\left(X_j, Y_j = \omega_q\right)$ is represented by

$$\begin{cases} m_{i,j}\left(\{\omega_q\}\right) & = e^{-\gamma_q d_{i,j}^2} \\ m_{i,j}(\Omega) & = 1 - e^{-\gamma_q d_{i,j}^2} \end{cases}$$

- Global mass function : Dempster+Yager rule
- Loss function

$$\arg\min_{\Lambda} \frac{1}{N} \sum_{i=1}^{N} \sum_{q=1}^{c} \left\{ Pl_i\left(\{\omega_q\}\right) - t_{i,q} \right\}^2 + \frac{1}{N} \sum_{i=1}^{N} m_i(\Omega) + \beta \|\Lambda\|_0$$

Improved EFS

What happens if $d_{i,j}$ is too large?

- Choose only the first K nearest neighbors of each query pattern X_i
- Resulting mass function $m_i^{\Theta_q}$ can be represented as when $\Theta_q \neq \emptyset$

$$\begin{cases} m_i^{\Theta_q}(\{\omega_q\}) = 1 - \prod_{\chi_{i_p \in \Theta_q}^{p=1,\ldots,K}}^{p=1,\ldots,K} \left(1 - e^{-\gamma_q d_{i_1 j_p}^2}\right) \\ m_i^{\Theta_q}(\Omega) = \prod_{\chi_{i_p} \in \Theta_q}^{p=1,\ldots,K} \left(1 - e^{-\gamma_q d_{i_p p}^2}\right) \end{cases}$$

when $\Theta_q = \emptyset$, $m_i^{\Theta_q}(\Omega) = 1$

Improved EFS

- By TBM conjunctive rule, the global mass function is represented as

$$\begin{cases} M_i(\{\omega_1\}) & = m_i^{\Theta_1}(\{\omega_1\}) \cdot m_i^{\Theta_2}(\Omega) \\ M_i(\{\omega_2\}) & = m_i^{\Theta_2}(\{\omega_2\}) \cdot m_i^{\Theta_1}(\Omega) \\ M_i(\Omega) & = m_i^{\Theta_1}(\Omega) \cdot m_i^{\Theta_2}(\Omega) \\ M_i(\Phi) & = m_i^{\Theta_1}(\{\omega_1\}) \cdot m_i^{\Theta_2}(\{\omega_2\}) \end{cases}$$

- Loss function

$$\begin{aligned} L(\Lambda) = & \frac{1}{N}\sum_{i=1}^{N}\sum_{q=1}^{2}\left\{M_i(\{\omega_q\}) - t_{i,q}\right\}^2 \\ & + \frac{1}{N}\sum_{i=1}^{N}\left\{M_i(\Phi)^2 + M_i(\Omega)^2\right\} \\ & + \beta\|\Lambda\|_0 \end{aligned}$$

Prior knowledge

- We incorporate a most important SUV-based feature into EFS as a fixed element of the optimal feature subset
- RELIEF : A feature ranking method

$$S(\tilde{f}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{k}\sum_{j=1}^{k}\mathrm{diff}\left(\tilde{f}, X_i, \mathrm{mis}_j^i\right) - \frac{1}{k}\sum_{j=1}^{k}\mathrm{diff}\left(\tilde{f}, X_i, \mathrm{li}\ i_j^i\right)\right)$$

where hit_j^i and $\mathrm{mis}\,s_j^i, j \in \{1, \ldots, k\}$, are the nearest neighbors of X_i from the same class and the opposite class.

- The SUV-based feature with the largest score will be included in EFS.

Data Balancing

Problem : A small training sample size and class imbalance

Table 1
Description of the three clinical datasets.

Dataset	Sample size	Feature size	Imbalance ratio
Lung tumor	25	52	0.24
Esoph. tumor	36	29	0.36
Lymph tumor	45	27	0.13

Solution : Adaptive Synthetic Sampling (ADASYN)

ADASYN Algorithm

Algorithm 1: ADASYN-based balancing for feature selection (He et al., 2009)

input : imbalanced dataset $\{(X_i, Y_i)|i = 1, \ldots, N\}$, where $X_i = [x_{i,1}, \ldots, x_{i,V}]^T$ and $Y_i \in \{\omega_1, \omega_2\}$. Assume ω_1 and ω_2 represent the minority class and the majority class, respectively. Let n_{maj} and n_{min} be the number of majority class instances and the number of minority class instances, respectively.

```
Set the number of synthetic minority class instances as n_syn = n_maj − n_min.
for each sample X_j with Y_j = ω_1 do
    Find k nearest neighbors of X_j in the training pool.
    Calculate the parameter r_j for X_j as r_j = Δ_j/k where Δ_j
    is the number of nearest neighbors of X_j that belong to
    the majority class.
end
for each sample X_j with Y_j = ω_1 do
    Define the level of difficulty in learning for X_j as
    r̄_j = r_j / Σ_{j=1}^{n_min} r_j.
    Determine the number of synthetic instances for X_j as
    n_j = r̄_j × n_syn.
    for i = 1, 2, ..., n_j do
        Randomly select a minority class instance, X_r, from the
        neighbors of X_j.
        Randomly generate a scalar δ ∈ [0, 1].
        Generate a minority synthetic instance as
        S_i^j = X_j + δ × (X_r − X_j).
    end
end
```

Data Balancing

- Problem : due to the random nature of ADASYN and a limited trainning samples, the balanced training dataset CAN NOT ALWAYS be more representive than the original dataset.
- Solution : Execute ADASYN B times to provide B>1 balanced training datasets.

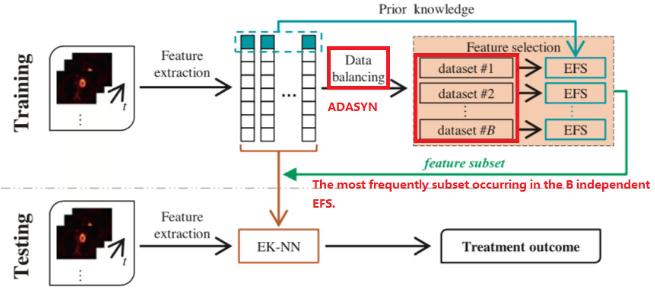

Fig. 1. Protocol of the prediction system.

Classification

- The classification rule : EK-NN
- Use the original taining dataset with selected features to train classification rule

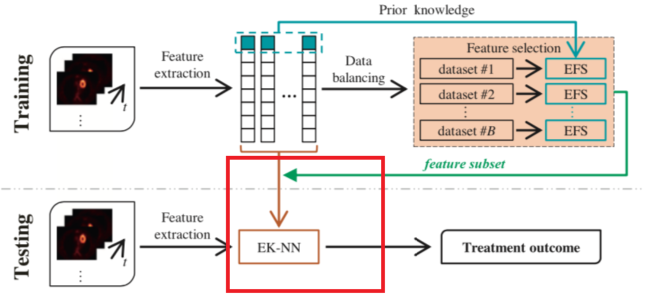

Fig. 1. Protocol of the prediction system.

Outline

1. Introduction
2. Main protocol
3. Method
4. Experimental result
   - Feature selection performance
   - Prediction performance
5. Conclusion

Feature selection performance

- LOOCV and .632+ Bootstrapping
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
- AUC : the Area Under the ROC Curve
- Robustness : the relative weighted consistency

Feature selection performance evaluated by the LOOCV. EFS represents our previous work (Lian et al., 2015a), while *i*EFS denotes the improved EFS that proposed in this paper. "All" represents the results for all the input features (without selection).

	All	RELIEF	FAST	SFS	SFFS	SVMRFE	KCS	HFS	EFS	*i*EFS
Lung tumor data										
Robustness	–	0.64	0.65	0.85	0.32	0.56	0.50	**1.00**	0.94	**1.00**
Accuracy	0.76	0.72	0.76	0.88	0.80	0.76	0.84	**1.00**	**1.00**	**1.00**
AUC	0.50	0.60	0.35	0.95	0.61	0.74	0.81	**1.00**	**1.00**	**1.00**
Subset size	52	10	14	2	5	5	3	3	4	4
Esophageal tumor data										
Robustness	–	0.94	**1.00**	0.26	0.23	0.80	0.94	0.53	0.92	**1.00**
Accuracy	0.64	0.56	0.64	0.64	0.58	0.72	0.69	0.72	0.83	**0.89**
AUC	0.12	0.54	0.12	0.50	0.55	**0.76**	0.57	0.67	0.69	**0.77**
Subset size	29	2	27	5	5	5	2	5	3	3
Lymph tumor data										
Robustness	–	**1.00**	0.85	0.72	0.34	0.64	**1.00**	0.90	0.57	**0.95**
Accuracy	0.87	**0.96**	0.82	0.89	0.87	0.89	**0.96**	0.87	0.89	**0.93**
AUC	0.50	0.68	0.26	0.65	0.29	0.83	0.68	0.36	**0.92**	**0.95**
Subset size	27	1	5	2	5	5	1	4	4	4

Prediction performance

- The selected feature subsets led to the higher AUC than the input feature for all the three examples

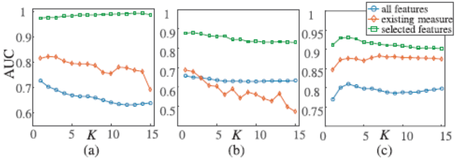

Fig. 3. Prediction performance of the EK-NN classifier with respect to different K : (a) lung tumor dataset, (b) esophageal tumor dataset, and (c) lymph tumor dataset. "all features", "selected features", and "existing measure" denote the results obtained by the input features, the selected feature subset and the predictor that has been clinically proven, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Outline

Conclusion

- An improved EFS with prior knowledge and data balancing has been proposed.
- After feature selection, the EK-NN classifier has been trained to predict the outcome.
- More radiological features extracted from other image modes (CT, MRI and multi-trace PET) to further improve the reliability of the prediction system

Reference

[1] Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., et al., 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature Commun. 5.

[2] Barwick, T.D., Taylor, A., Rockall, A., 2013. Functional imaging to predict tumor response in locally advanced cervical cancer. Current Oncol. Rep. 15 (6), 549–558. Bloch, I., 1996. Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. Pattern Recognit. Lett. 17 (8), 905–919.

[3] Brooks, F.J., Grigsby, P.W., 2014. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J. Nuclear Med. 55 (1), 37–42.

[4] Calais, J., Thureau, S., Dubray, B., Modzelewski, R., Vera, P., 2015. Areas of high 18F-FDG uptake on preradiotherapy PET/CT identify pref- erential sites of local relapse after chemoradiotherapy for non–small cell lung cancer. J. Nuclear Med. 56 (2), 196–203.

[5] Denœux, T., 1995. A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans. Syst. Man Cybern. 25 (5), 804–813.

[6] Denœux, T., Smets, P., 2006. Classification using belief functions: relationship between case-based and model-based approaches. IEEE Trans. Syst. Man Cybern.