Outcome Prediction in Cancer Therapy based on Dempster-Shafer Theory and PET Imaging

Chunfeng Lian[1,2], Su Ruan[2], Thierry Denœux[1], Pierre Vera[3,2]

[1]Sorbonne Universités, Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, Compiègne, France

[2]Université de Rouen, QuantIF – EA 4108 LITIS, Rouen, France

[3]Centre Henri-Becquerel, Department of Nuclear Medicine, Rouen, France

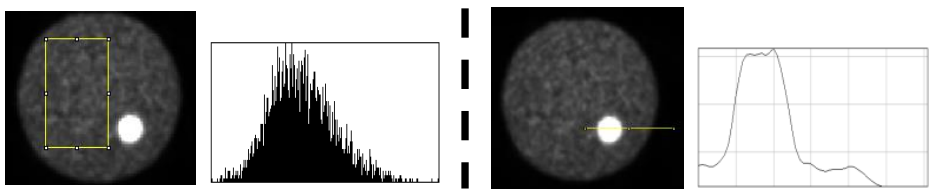
Contents

Outcome Prediction in Cancer Therapy

- Outcome prediction prior to or even during the cancer therapy
  $\Longrightarrow$ tailoring and adapting a treatment planning.
- To this end, there are diverse information sources :
  - Patient's Demography :
    patient gender, patient age, country ...
  - Clinical Characteristics :
    tumor stage, tumor location, histology, genomic data ...
  - Medical Imaging :
    anatomical images and functional images $\Rightarrow$ tumor volume, intensity, texture features ...

Difficulties We Have

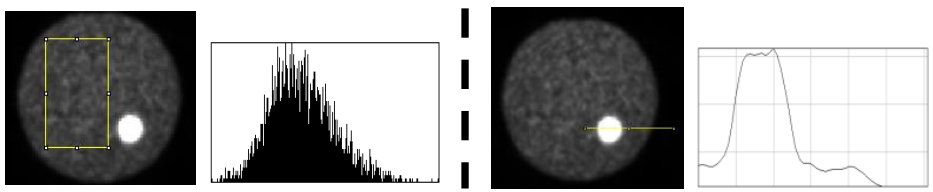
1. Information sources are imprecise :
   - Positron Emission Tomography (PET) is blurring and noisy.



   - Clinical characteristics offered by clinicians are in some sense subjective and inaccuracy.

Difficulties We Have

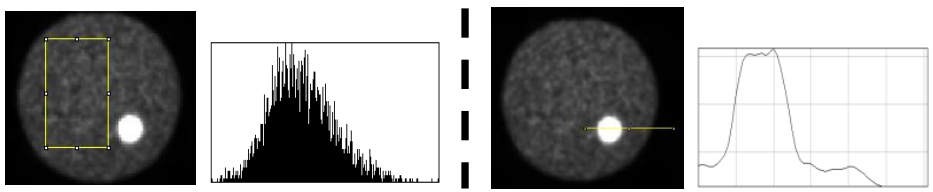
1. Information sources are imprecise :
   - Positron Emission Tomography (PET) is blurring and noisy.



   - Clinical characteristics offered by clinicians are in some sense subjective and inaccuracy.
2. No consensus to determine the most predictive features :
   - Dozens of PET imaging features, texture features and clinical features. Maybe redundant, irrelevant or even interference.

Difficulties We Have

1. Information sources are imprecise :
   - Positron Emission Tomography (PET) is blurring and noisy.

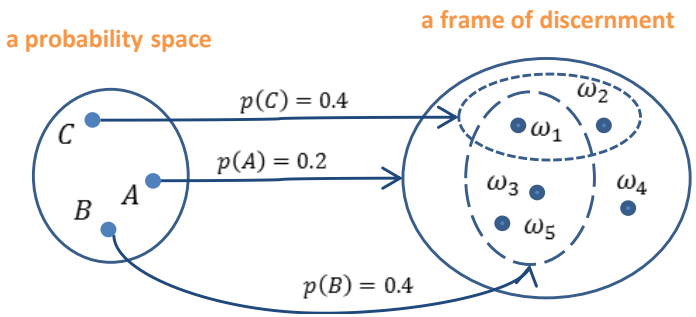


   - Clinical characteristics offered by clinicians are in some sense subjective and inaccuracy.
2. No consensus to determine the most predictive features :
   - Dozens of PET imaging features, texture features and clinical features. Maybe redundant, irrelevant or even interference.

We need a stable prediction rule and feature selection.

Dempster-Shafer Theory

- Also known as the Theory of Belief Functions and Evidence Theory.
- An extension of Probability theory and Set-Membership Approach.



- A powerful framework for reasoning and making decision with partial (uncertain, imprecise) knowledge.

Contents

Modified EK-NN Classification Rule

Given a query instance X_t and training sample $(X_j, Y_j = \omega_q)$, evidence regarding X_t's label can be quantified as [Denoeux, 1995] :

$$\begin{cases} m_{t,j}(\omega_q) & = \alpha \cdot exp(-\gamma_q \cdot d_{t,j}^2) \\ m_{t,j}(\Omega) & = 1 - m_{t,j}(\omega_q) \end{cases}$$

Modified EK-NN Classification Rule

Given a query instance X_t and training sample $(X_j, Y_j = \omega_q)$, evidence regarding X_t's label can be quantified as [Denoeux, 1995] :

$$\begin{cases} m_{t,j}(\omega_q) & = \alpha \cdot exp(-\gamma_q \cdot d_{t,j}^2) \\ m_{t,j}(\Omega) & = 1 - m_{t,j}(\omega_q) \end{cases}$$

A mixed combination rule to fuse K-NNs' evidence

1. NNs with the same label : Dempster's rule $\Rightarrow m^{\Gamma_q}(\omega_q) + m^{\Gamma_q}(\Omega) = 1$;
2. Between different m^{Γ_q}, where $q = 1, \ldots, c$:
   - Discounting according to each group's cardinality $|\Gamma_q|$:

$$\begin{cases} dm_t^{\Gamma_q}(\{\omega_q\}) & = (|\Gamma_q|/|\Gamma_{max}|)^\eta \times m_t^{\Gamma_q}(\omega_q) \\ dm_t^{\Gamma_q}(\Omega) & = 1 - (|\Gamma_q|/|\Gamma_{max}|)^\eta \times m_t^{\Gamma_q}(\omega_q) \end{cases}$$

   - Global fusion via Yager's rule :

$$\begin{cases} m_t(\{\omega_q\}) & = dm_t^{\Gamma_q}(\{\omega_q\}) \times \prod_{h \in \{1,\ldots c\} \setminus q} dm_t^{\Gamma_h}(\Omega), \quad \forall q = 1 \ldots c \\ m_t(\Omega) & = 1 - \sum_{q=1}^{c} m_t(\{\omega_q\}) \end{cases}$$

Modified EK-NN

Examples of mixed combination

Comparing fusion results between different combination rules

• Assume ♯1 ∼ ♯4 are the training samples with the same distance to a query instance :

CASE 1 : A NORMAL CONFLICTING SITUATION

	♯1	♯2	♯3	Dempster's rule	Yager's rule	mixed rule
$m(\{\omega_1\})$	0.8	0.8	0.0	0.8276	0.1920	**0.7680**
$m(\{\omega_2\})$	0.0	0.0	0.8	0.1379	0.0320	**0.0080**
$m(\Omega)$	0.2	0.2	0.2	0.0345	0.7760	**0.2240**

CASE 2 : A HIGH CONFLICTING SITUATION

	♯1	♯2	♯3	♯4	Dempster's rule	Yager's rule	mixed rule
$m(\{\omega_1\})$	0.8	0.8	0.0	0.0	0.4898	**0.0384**	**0.0384**
$m(\{\omega_2\})$	0.0	0.0	0.8	0.8	0.4898	**0.0384**	**0.0384**
$m(\Omega)$	0.2	0.2	0.2	0.2	0.0204	**0.9232**	**0.9232**

Modified EK-NN
two-step Classification

Assumption

- **Using belief functions** :

  precise objects $\simeq$ additional evidence for imprecise objects.

Approach

1. Mass functions constructed via the proposed mixed combination rule ;
2. Making decision for easy to classify objects
   $\Rightarrow$ bigger group of "training pairs" ;
3. Calculating prototypes (i.e., class centers)
   $\Rightarrow$ Making decision for imprecise objects.

Modified EK-NN

A example of two-step Classification

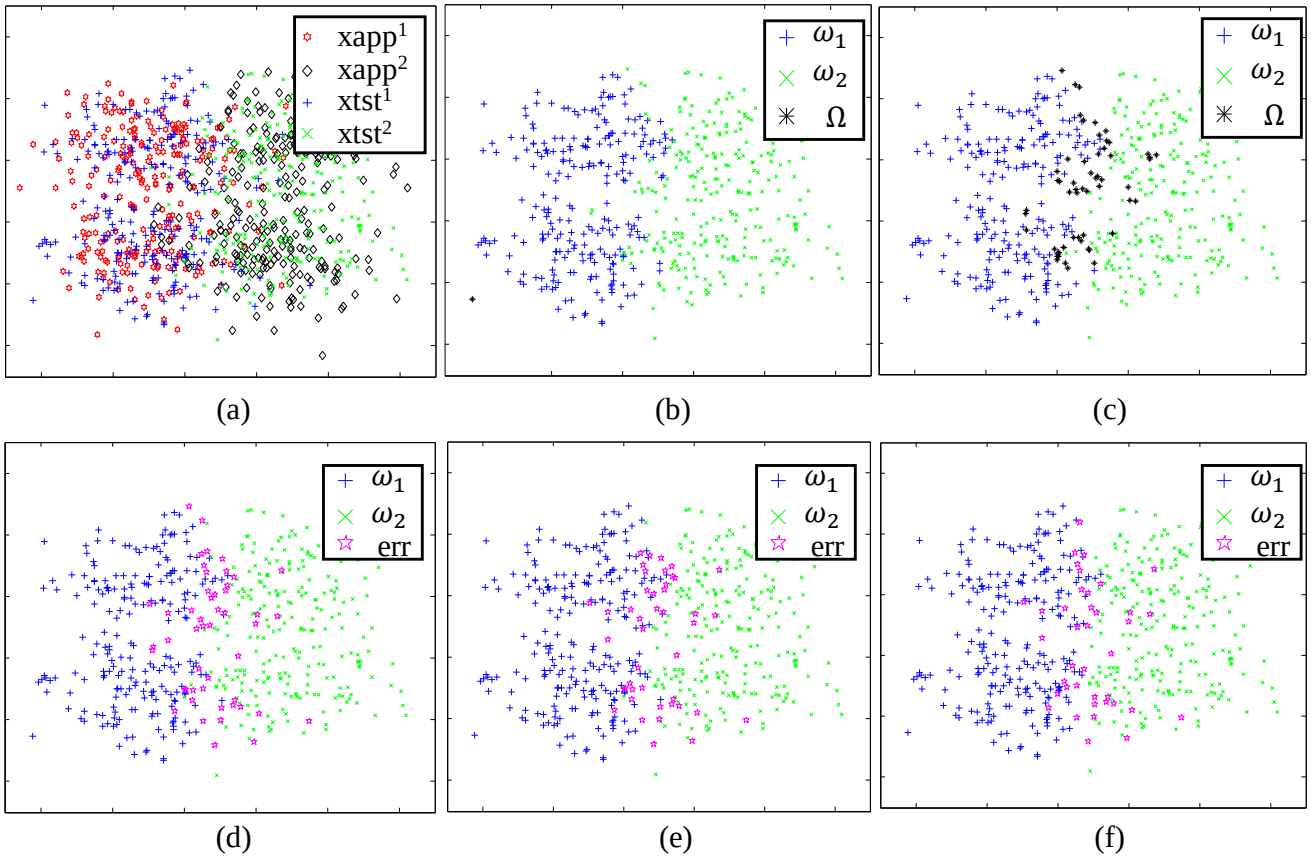


FIGURE: (b) and (c) are credal partition results for EK-NN and mEK-NN ; (d)-(f) are decision making results. The error rates are respectively **9.80%**, **8.80%** and **7.80%**.

Evidential Feature Selection
Main idea

Requirements

A good feature subset should satisfy three requirements :

- High classification accuracy ;
- Low imprecision and uncertainty (small overlaps between different classes) ;
- Sparsity to reduce the risk of over-fitting.

Evidential Feature Selection
Main idea

Requirements

A good feature subset should satisfy three requirements :

- High classification accuracy ;
- Low imprecision and uncertainty (small overlaps between different classes) ;
- Sparsity to reduce the risk of over-fitting.

According to these requirements, we developed an Evidential Feature Selection (EFS) method based on **DST** and **mEK-NN**.

Evidential Feature Selection

- The dissimilarity between two training instances X_i and X_j is measured by a weighted euclidian distance :

$$d_{j,i} = \sqrt{\sum_{p=1}^{m} \lambda_p \cdot \left(d_{j,i}^p\right)^2}$$

$\Longrightarrow$ λ_p is the binary coefficient for feature selection.

Evidential Feature Selection

- The dissimilarity between two training instances X_i and X_j is measured by a weighted euclidian distance :

$$d_{j,i} = \sqrt{\sum_{p=1}^{m} \lambda_p \cdot \left(d_{j,i}^p\right)^2}$$

$\Longrightarrow$ λ_p is the binary coefficient for feature selection.

- Based on mEK-NN, features are selected through

$$\arg_\lambda \min \frac{1}{n} \sum_{i=1}^{n} \sum_{q=1}^{c} \left(Pl_i(\{\omega_q\}) - t_{i,q}\right)^2 + \frac{\rho}{n} \sum_{i=1}^{n} m_i(\Omega) + \delta \times l_0$$

  - m_i and Pl_i are the mass and plausibility function ;
  - label indicator $t_{i,q} = 1$ iff $Y_i = \omega_q$;
  - ρ and δ are two hyper-parameters.

Evidential Feature Selection

Specified loss function

$$\arg \min_{\lambda_1,\ldots,\lambda_m} \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{c}(1 - t_{i,q} - \sum_{h\neq q} B_h^i)^2 + \rho \times \frac{1}{n}\sum_{i=1}^{n}(1 - \sum_{q=1}^{c} B_q^i) + \delta \sum_{p=1}^{m}[1 - \exp(-5\lambda_p)] \quad (1)$$

with

$$B_q^i = A_q^i \prod_{s\in\{1,\ldots c\}\setminus q}(1 - A_s^i)$$

and

$$A_q^i = \left(\frac{|\Gamma_q^i|}{|\Gamma_{max}^i|}\right)^{\eta}\left(1 - \prod_{j\in\Gamma_q^i}\left[1 - \alpha \exp(-\gamma_q \cdot d_{i,j}^2)\right]\right).$$

$\Longrightarrow$ Solved via integer genetic algorithm [Deep et al., 2009].

Evidential Feature Selection

A test on synthetic data

- Data generation [Perkins et al., 2003] :
  There were n_r relevant, n_c redundant and n_i irrelevant features uniformly distributed between [-1,1]. Class label was determined only by relevant features :

$$y = \begin{cases} \omega_1 & \text{if } \max_i(x_i) > 2^{1-\frac{1}{n_r}} - 1, \\ \omega_2 & \text{otherwise.} \end{cases}$$

  where x_i $(1 \leq i \leq n_r)$ is the ith relevant feature.

Evidential Feature Selection

A test on synthetic data

- Data generation [Perkins et al., 2003] :
  There were n_r relevant, n_c redundant and n_i irrelevant features uniformly distributed between [-1,1]. Class label was determined only by relevant features :

$$y = \begin{cases} \omega_1 & \text{if } \max_i(x_i) > 2^{1-\frac{1}{n_r}} - 1, \\ \omega_2 & \text{otherwise.} \end{cases}$$

  where x_i $(1 \leq i \leq n_r)$ is the ith relevant feature.

- Obtained results :

n_r	n_c	n_i	subset size	EK-NN	mEK-NN	
					no fs	with fs
2	2	6	**2**	14.67	12.67	**2.67**
2	2	16	**2**	17.33	12.00	**1.33**
2	2	26	**2**	23.33	18.67	**4.00**
2	2	36	**2**	28.67	26.67	**5.33**
2	2	46	**2**	29.33	23.33	**4.67**

Contents

- Background
  - Outcome Prediction in Cancer Therapy
  - Difficulties in Outcome Prediction
  - Dempster-Shafer Theory
- Our Approach
  - Modified EK-NN Classification Rule
  - Evidential Feature Selection
- Experimental Results
  - on UCI data sets
  - on Clinical Data Sets
- Conclusion

UCI Data Sets

feature selection

TABLE: comparing EFS with classical wrapper methods using 10-fold cross validation. The robustness is evaluated via [Somol and Novovicova, 2010].

	Iris			**Seeds**		
	Error(%)	Robustness(%)	Subset Size	Error(%)	Robustness(%)	Subset Size
All	2.67	n/a	4	7.62	n/a	7
SFS	4.67	54.55	1	11.90	57.97	2
SBS	5.33	21.05	2	10.95	23.88	3
SFFS	5.33	21.62	3	5.24	54.93	2
EFS*	***2.00***	***100***	3	***4.76***	***81.18***	3
	Wine			**Yeast**		
	Error(%)	Robustness(%)	Subset Size	Error(%)	Robustness(%)	Subset Size
All	13.04	n/a	13	38.87	n/a	8
SFS	30.50	75	1	61.99	***100***	1
SBS	6.24	42.47	5	48.35	***100***	1
SFFS	7.29	57.58	4	36.21	40	5
EFS*	***5.13***	***91.89***	3	***32.51***	***100***	2
	WDBC			**Parkinsons**		
	Error(%)	Robustness(%)	Subset Size	Error(%)	Robustness(%)	Subset Size
All	7.20	n/a	30	13.37	n/a	22
SFS	14.44	80	1	15.82	33.33	1
SBS	19.67	22.22	2	19.03	23.91	2
SFFS	9.87	25	4	13.79	43.65	3
EFS*	***5.80***	***92.37***	3	***8.63***	***100***	3

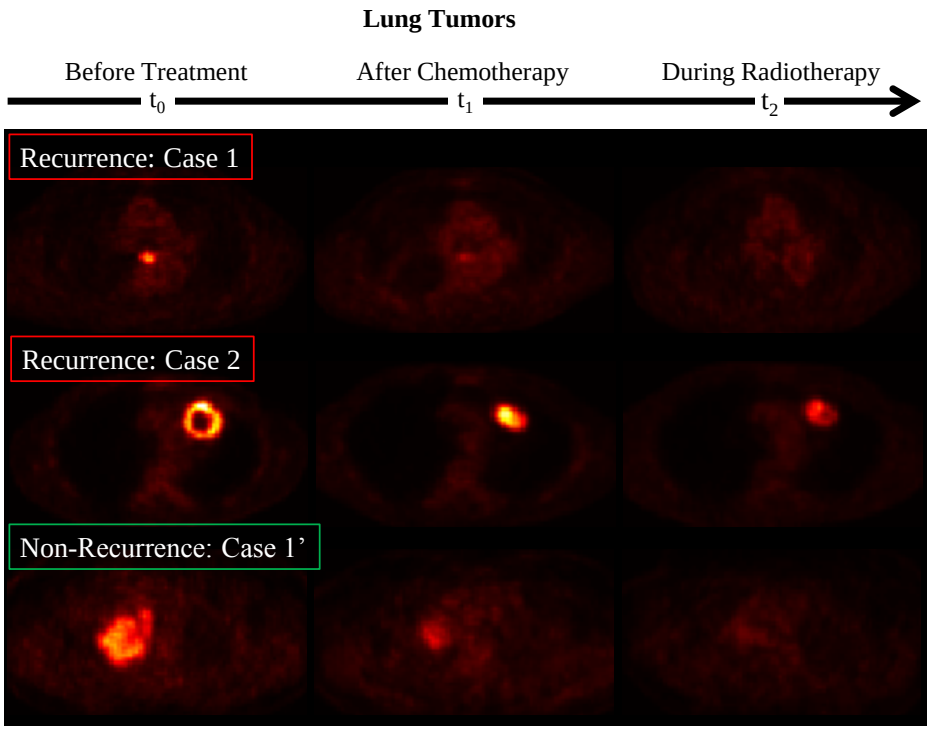
UCI Data Sets

classification

TABLE: comparing the classification performance. SFFS was used to select features for other methods. For BK-NN [Liu et al., 2013] and CCR [Liu et al., 2014], R_e and R_i represent, respectively, the error rate and imprecision rate.

		Iris	Seeds	Wine	Yeast	WDBC	Parkinson
SFFS +	ANN	8.00	7.62	9.64	32.57	9.15	9.63
	CART	8.00	7.14	9.09	37.55	10.04	11.21
	SVM	6.00	7.14	6.83	36.14	8.28	13.26
	EK-NN	5.33	6.67	6.18	35.07	9.70	16.39
	BK-NN (R_e, R_i)	4.00 4.67	2.38 11.90	6.74 5.13	16.31 40.84	7.22 8.44	9.18 13.37
	CCR (R_e, R_i)	4.00 4.67	3.81 18.57	3.99 15.33	19.53 36.11	5.99 15.83	16.42 12.26
our method		**2.00**	**4.76**	**5.13**	**32.51**	**5.80**	**8.63**

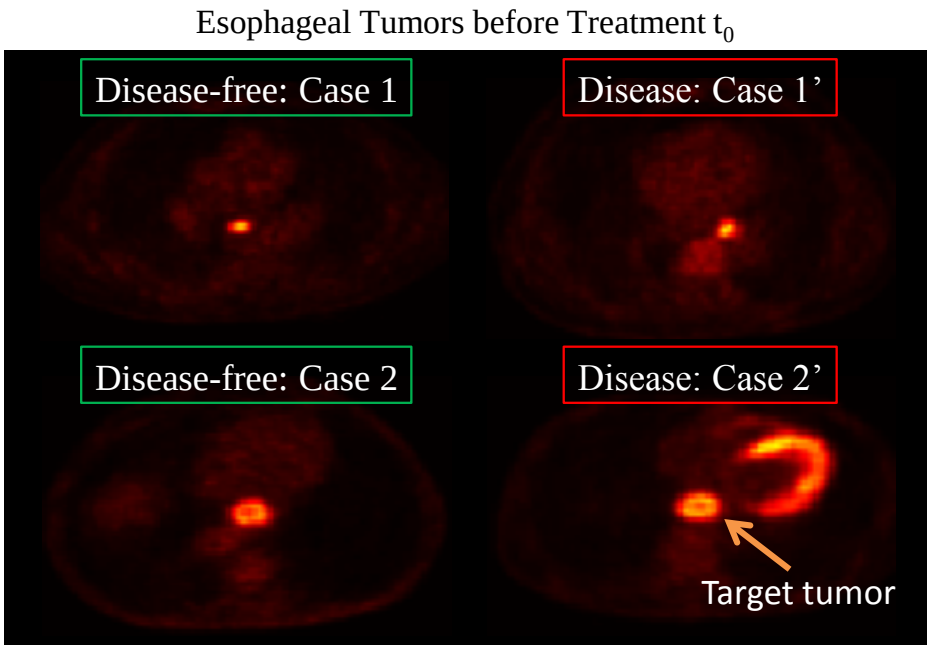
Clinical Data Sets
Lung Tumor Data



FDG-PET uptakes for lung tumor

- twenty-five patients with stage II-III non small cell lung cancer were treated with curative intent chemo-radiotherapy.
- FDG-PETs : before treatment, after chemotherapy and during radiotherapy.
- definition of recurrence after one year : local/distant recurrence (19 patients) and no recurrence (6 patients).

Clinical Data Sets
Esophageal Tumor Data

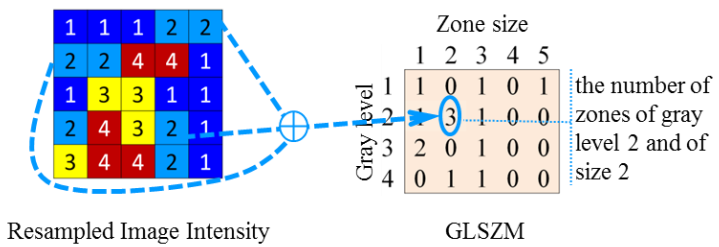


FDG-PET uptakes for esoph. tumor

- thirty-six patients with esophageal squamous cell carcinomas were treated with chemo-radiotherapy, and followed up in a long term up to five years.
- FDG-PETs : only before treatment is available.
- neither loco regional nor distant recurrence (13 patients) and disease-positive (23 patients).

Clinical Data Sets
Feature Extraction

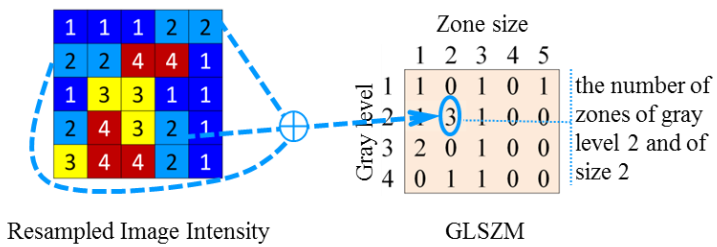
- Three types of PET imaging features.
  - SUV-based features : SUV_{max}, SUV_{mean}, SUV_{peak}, metabolic tumor volume (MTV) and total lesion glycolysis (TLG) ;
  - texture features : gray level size zone matrix [Tixier et al., 2012] ;



  - longitudinal change : relative difference between baseline features and follow-up features.

Clinical Data Sets
Feature Extraction

- Three types of PET imaging features.
  - SUV-based features : SUV_{max}, SUV_{mean}, SUV_{peak}, metabolic tumor volume (MTV) and total lesion glycolysis (TLG) ;
  - texture features : gray level size zone matrix [Tixier et al., 2012] ;



  - longitudinal change : relative difference between baseline features and follow-up features.
- Patients' clinical characteristics for esophageal tumor data.
  - gender, tumor stage, dysphagia grade, WHO performance status, weight loss, tumor location.

Clinical Data Sets
results

TABLE: Comparing feature selection performance using leave-one-out cross-validation. **EFS*** denotes the proposed method.

Method	Lung Tumor Data			Esophageal Tumor Data		
	Accuracy	Robustness	subset size	Accuracy	Robustness	subset size
All features	76±44	n/a	52	64±49	n/a	29
SFS	84±37	60	3	53±44	51	3
SFFS	72±46	54	4	**81±40**	53	3
SVMRFE	92±28	57	5	75±44	80	5
EFS*	**100±0**	**94**	4	**81±40**	**92**	3

TABLE: Comparing classification performance. **mEK-NN*** denotes the proposed classification method.

Classifier	Lung Tumor Data		Esophageal Tumor Data	
	without EFS	with EFS	without EFS	with EFS
ANN	68±48	92±28	67±48	83±38
SVM	76±44	**100±0**	64±49	81±40
EK-NN	68±48	96±20	64±49	83±38
mEK-NN*	56±51	**100±0**	53±44	**89±32**

Contents

- Background
  - Outcome Prediction in Cancer Therapy
  - Difficulties in Outcome Prediction
  - Dempster-Shafer Theory
- Our Approach
  - Modified EK-NN Classification Rule
  - Evidential Feature Selection
- Experimental Results
  - on UCI data sets
  - on Clinical Data Sets
- Conclusion

Conclusions

- More information about this presentation :
  - C.Lian, S.Ruan and T.Denœux, "An evidential classifier based on feature selection and two-step classification", *Pattern Recognition*, Vol. 48, pages 2318-2327, 2015.
  - C.Lian, S.Ruan, T.Denœux and P.Verra, "Outcome prediction in tumor therapy based on Dempster-Shafer Theory", *IEEE-ISBI*, New York, USA, pages 63-66, April 2015.

Conclusions

- More information about this presentation :
  - C.Lian, S.Ruan and T.Denœux, "An evidential classifier based on feature selection and two-step classification", *Pattern Recognition*, Vol. 48, pages 2318-2327, 2015.
  - C.Lian, S.Ruan, T.Denœux and P.Verra, "Outcome prediction in tumor therapy based on Dempster-Shafer Theory", *IEEE-ISBI*, New York, USA, pages 63-66, April 2015.
- Future work :
  - Tackling imbalanced learning problem and small sample size effect, so as to improve the performance ;
  - Evaluating the proposed method on larger clinical data sets.

Thanks for Your Attention.

chunfeng.lian@utc.fr

References I

Deep, K., Singh, K. P., Kansal, M., and Mohan, C. (2009).
A real coded genetic algorithm for solving integer and mixed integer optimization problems.
Applied Mathematics and Computation, 212(2) :505–518.

Denoeux, T. (1995).
A k-nearest neighbor classification rule based on dempster-shafer theory.
Systems, Man and Cybernetics, IEEE Transactions on, 25(5) :804–813.

Liu, Z.-g., Pan, Q., and Dezert, J. (2013).
A new belief-based k-nearest neighbor classification method.
Pattern Recognition, 46(3) :834–844.

Liu, Z.-g., Pan, Q., Dezert, J., and Mercier, G. (2014).
Credal classification rule for uncertain data based on belief functions.
Pattern Recognition, 47(7) :2532–2541.

Perkins, S., Lacker, K., and Theiler, J. (2003).
Grafting : Fast, incremental feature selection by gradient descent in function space.
The Journal of Machine Learning Research, 3 :1333–1356.

References II

Somol, P. and Novovicova, J. (2010).
Evaluating stability and comparing output of feature selectors that optimize feature subset cardinality.
Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(11) :1921–1939.

Tixier, F., Hatt, M., Le Rest, C., Le Pogam, A., Corcos, L., and Visvikis, D. (2012).
Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18f-fdg pet.
Journal of Nuclear Medicine, 53(5) :693–700.