Combining Data Sources Nonlinearly for Cell Nucleus Classification of Renal Cell Carcinoma

Mehmet Gönen[1] *Aydın Ulaş*[2] *Peter Schüffler*[3]
Umberto Castellani[2] *Vittorio Murino*[2,4]

[1] Aalto University School of Science, Department of Information and Computer Science
Helsinki Institute for Information Technology (HIIT), Espoo, Finland
[2] University of Verona, Department of Computer Science, Verona, Italy
[3] ETH Zurich, Department of Computer Science, Zürich, Switzerland
[4] Istituto Italiano di Technologia (IIT), Genova, Italy

Outline

- Introduction
- Methodology
- Data Set
- Experiments
- Conclusion

Introduction

- Kernel function = similarity measure
- Main factor of empirical performance
- Cross-validation to pick the best kernel
- *Multiple kernel learning* (MKL) to learn a better similarity measure

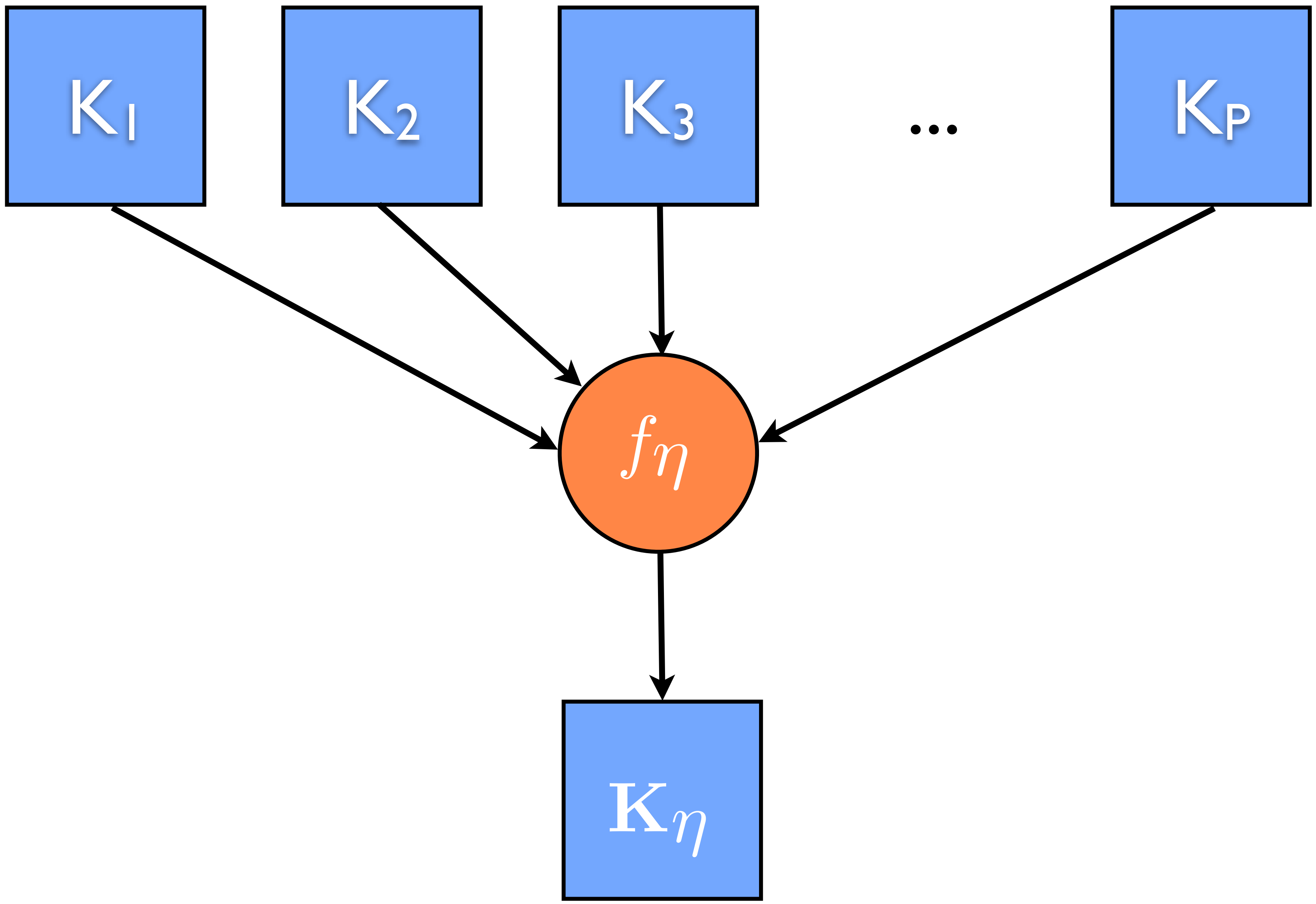
Our Contribution

- Formulate a nonlinear MKL variant
- Test it on cell nucleus classification of *renal cell carcinoma* (RCC)
- Combine different feature representations from *Tissue microarray* (TMA) images
- Compare our variant with single-kernel SVMs and linear MKL algorithms

Methodology

- Instead of picking a single kernel using cross-validation
- Combine *P* different kernels
  - similarity measures (i.e., different kernel functions)
  - feature representations (i.e., coming from different data sources or modalities)

Methodology

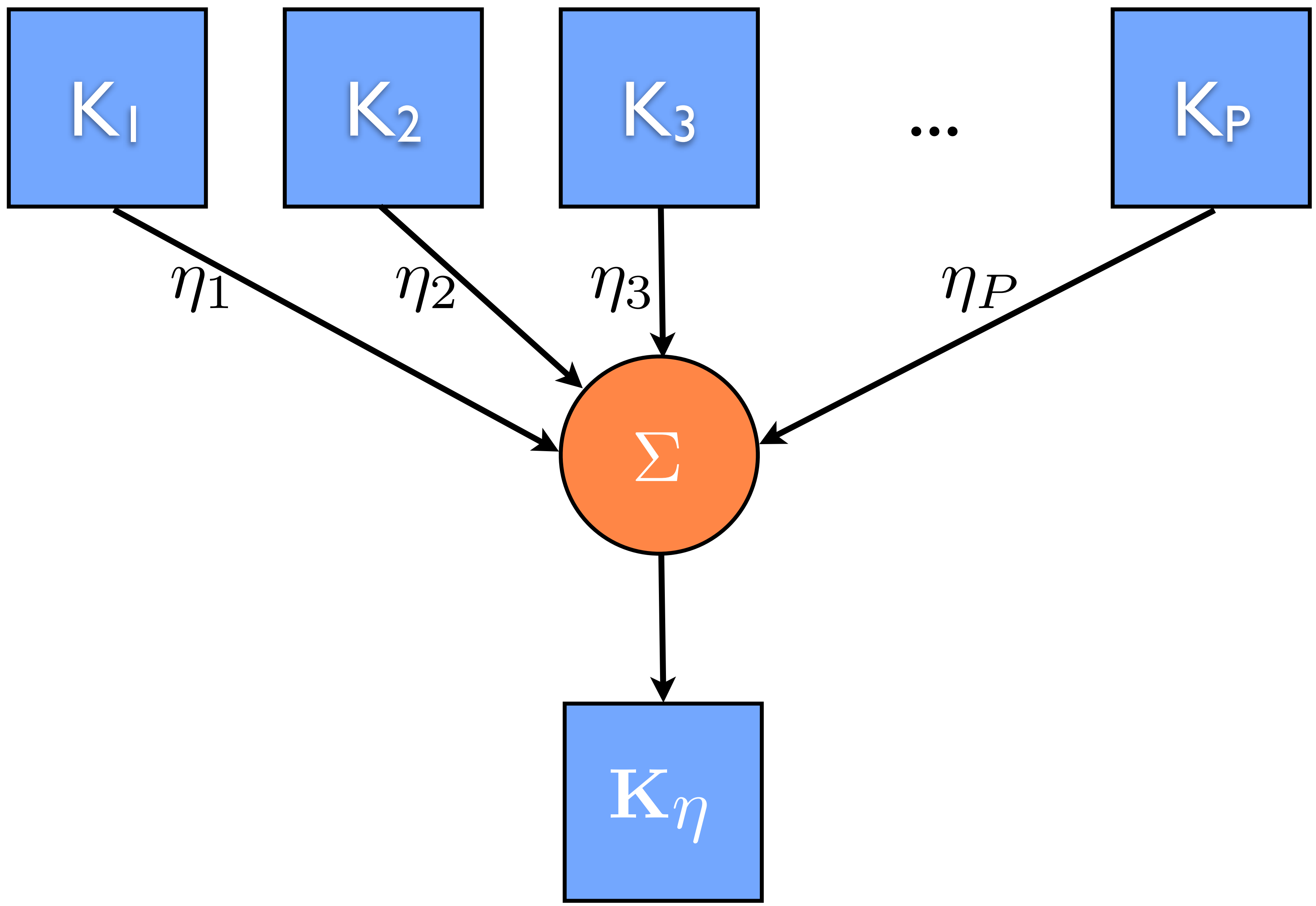


$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta}) = f_\eta(\{k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)_{m=1}^P\}; \boldsymbol{\eta})$$

Constructing Kernels

- scaling a kernel with a positive number
  - $ak_1(\boldsymbol{x}_i^1, \boldsymbol{x}_j^1)$
- summing up two kernels
  - $k_1(\boldsymbol{x}_i^1, \boldsymbol{x}_j^1) + k_2(\boldsymbol{x}_i^2, \boldsymbol{x}_j^2)$
- multiplying two kernels
  - $k_1(\boldsymbol{x}_i^1, \boldsymbol{x}_j^1)k_2(\boldsymbol{x}_i^2, \boldsymbol{x}_j^2)$

Linear MKL Algorithms



$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta}) = \sum_{m=1}^{P} \eta_m k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)$$

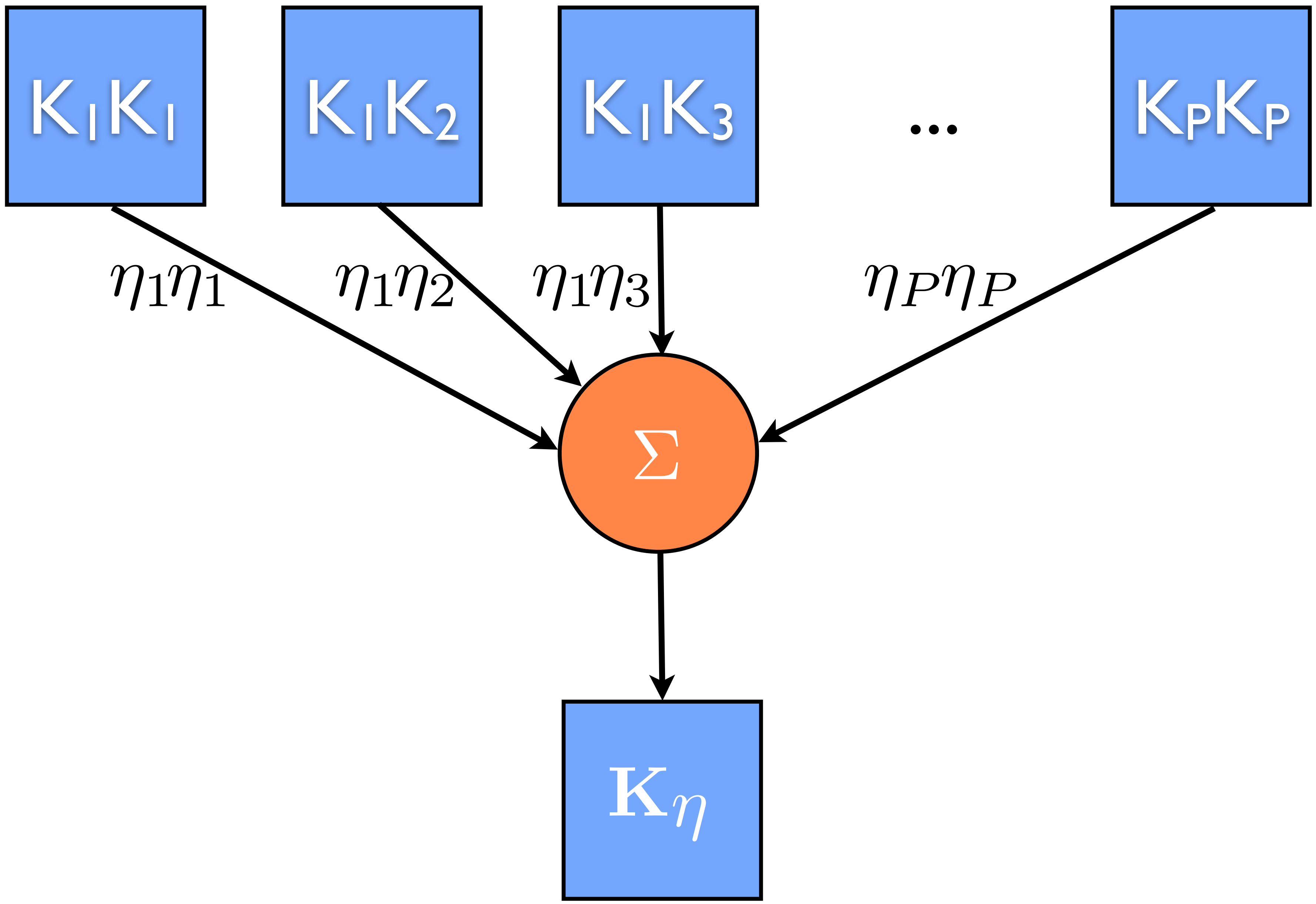
Linear MKL Algorithms

- Linear combination $\{\boldsymbol{\eta} \colon \boldsymbol{\eta} \in \mathbb{R}^P\}$
  - arbitrary kernel weights
- Conic combination $\{\boldsymbol{\eta} \colon \boldsymbol{\eta} \in \mathbb{R}^P_+\}$
  - positive kernel weights
- Convex combination $\{\boldsymbol{\eta} \colon \boldsymbol{\eta} \in \mathbb{R}^P_+, \;\; \mathbf{1}^\top \boldsymbol{\eta} = 1\}$
  - kernel weights on a simplex

Our Nonlinear Variant

$$
\begin{array}{ccccc}
 & \eta_1 & \eta_2 & & \eta_P \\
\eta_1 & K_1K_1 & K_1K_2 & \cdots & K_1K_P \\
\eta_2 & K_2K_1 & K_2K_2 & \cdots & K_2K_P \\
 & \vdots & \vdots & \ddots & \vdots \\
\eta_P & K_PK_1 & K_PK_2 & \cdots & K_PK_P
\end{array}
$$

Our Nonlinear Variant



$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{m=1}^{P} \sum_{h=1}^{P} \eta_m \eta_h k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m) k_h(\boldsymbol{x}_i^h, \boldsymbol{x}_j^h)$$

Our Nonlinear Variant

- Modified optimization problem

$$\underset{\boldsymbol{\eta}\in\mathcal{M}}{\text{minimize}} \;\; J_\eta = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\text{maximize}} \;\; \mathbf{1}^\top\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^\top((\boldsymbol{y}\boldsymbol{y}^\top)\odot\mathbf{K}_\eta)\boldsymbol{\alpha}$$

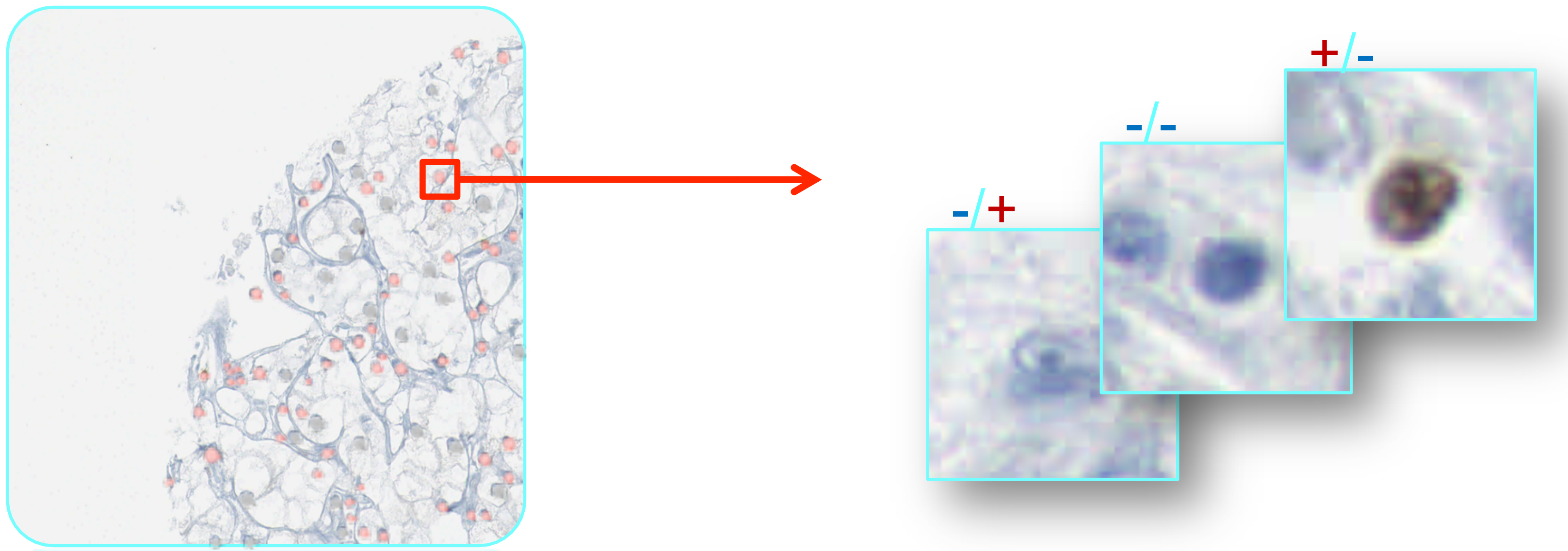
$$\mathcal{M} = \{\boldsymbol{\eta}\colon \boldsymbol{\eta}\in\mathbb{R}_+^P,\;\; \mathbf{1}^\top\boldsymbol{\eta} = 1\}$$

$$\mathcal{A} = \{\boldsymbol{\alpha}\colon \boldsymbol{\alpha}\in\mathbb{R}_+^P,\;\; \boldsymbol{y}^\top\boldsymbol{\alpha} = 0,\;\; \boldsymbol{\alpha}\le C\}$$

- A projection-based gradient-descent algorithm

$$\frac{\partial J_\eta}{\partial \eta_m} = -\frac{1}{2}\sum_{h=1}^{P}\eta_h\boldsymbol{\alpha}^\top((\boldsymbol{y}\boldsymbol{y}^\top)\odot\mathbf{K}_h\odot\mathbf{K}_m)\boldsymbol{\alpha}$$

Data Set



8 TMA images from 8 patients

Nuclei extraction by two pathologists

- 1633 patches in total
- Pathologists agreed on labels of 1273 patches (891 benign and 382 malignant)

Data Set

Name	Feature Description
ALL	Patch Intensity
FG	Foreground Intensity
BG	Background Intensity
LBP	Local Binary Patterns
COL	Color Feature
FCC	Freeman Chain Code
SIG	1D-Signature
PHOG	Pyramid Histograms of Oriented Gradients

Experiments

- 10-fold stratified cross-validation on 1273 nuclei samples (from 8 patients)
- 8 feature representations (`ALL`, `FG`, `BG`, `LBP`, `COL`, `FCC`, `SIG`, and `PHOG`)
- 3 basic kernel functions (`LIN`, `POL`, and `GAU`)

Experiments

- `SVM`: each feature representation separately
- `RBMKL`: using the mean of the kernels
- `SimpleMKL`: benchmark linear MKL
- `GLMKL`: group Lasso-based MKL
- `NLMKL`: our nonlinear MKL variant

SVM Results

	LIN	POL	GAU
ALL	70.0±0.2	71.9±2.9	68.7±2.9
FG	70.0±0.2	71.2±3.7	65.9±4.3
BG	70.2±0.6	72.7±3.8	69.6±3.1
LBP	70.0±0.2	63.6±2.7	68.4±6.3
COL	70.2±3.0	62.9±3.5	67.2±3.4
FCC	70.0±0.2	69.8±0.7	62.9±5.5
SIG	70.0±0.2	69.6±3.4	66.0±3.0
PHOG	76.0±3.4	70.5±3.3	**76.9**±**2.7**

MKL Results

	LIN	POL	GAU	LIN+POL+GAU
SVM	76.0±3.4	72.7±3.8	76.9±2.7	NA
RBMKL	77.3±4.0	77.2±2.4	82.7±3.6	81.8±3.8
SimpleMKL	77.1±3.3	77.3±2.3	81.8±3.8	81.6±3.9
GLMKL	77.1±3.5	76.5±3.2	81.8±4.3	81.8±3.8
NLMKL	77.9±3.9	79.2±3.8	**83.3±3.6**	83.1±3.5

Training Times

	LIN	POL	GAU	LIN+POL+GAU
SVM	4.45	5.81	3.52	NA
RBMKL	1.56	0.87	1.35	2.57
SimpleMKL	35.55	11.07	11.71	32.81
GLMKL	11.11	4.61	5.20	14.27
NLMKL	45.25	39.21	44.28	323.83

Conclusion

- Our NLMKL is better than single-kernel SVMs and linear MKL methods
- Better results may be possible
  - using more complex combination schemes
  - adding new modalities

Some Notes

- Many MKL algorithms in the literature
- See Gönen & Alpaydın (2011) for a recent survey
- MKL Matlab Toolbox is available at http://users.ics.tkk.fi/gonen/mkl