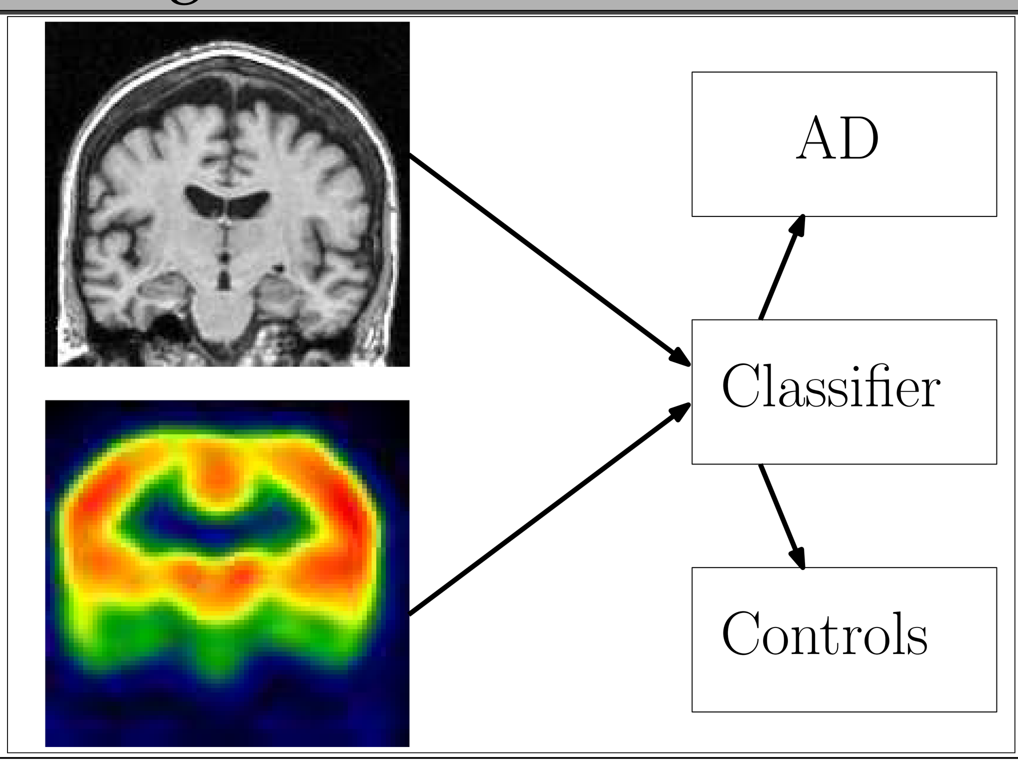
MKL for Robust Multi-modality AD Classification

Chris Hinrichs[1,2], Vikas Singh[2,1], Guofan Xu[3], Sterling C. Johnson[3]

[1]Computer Sciences	[2]Biostatistics & Medical Informatics	[3]Medicine
hinrichs@cs.wisc.edu	vsingh@biostat.wisc.edu	{gxu,scj}@medicine.wisc.edu

University of Wisconsin – Madison

Our Aim:
To Combine Multiple Scanning Modalities into a single machine learning framework.



Motivation

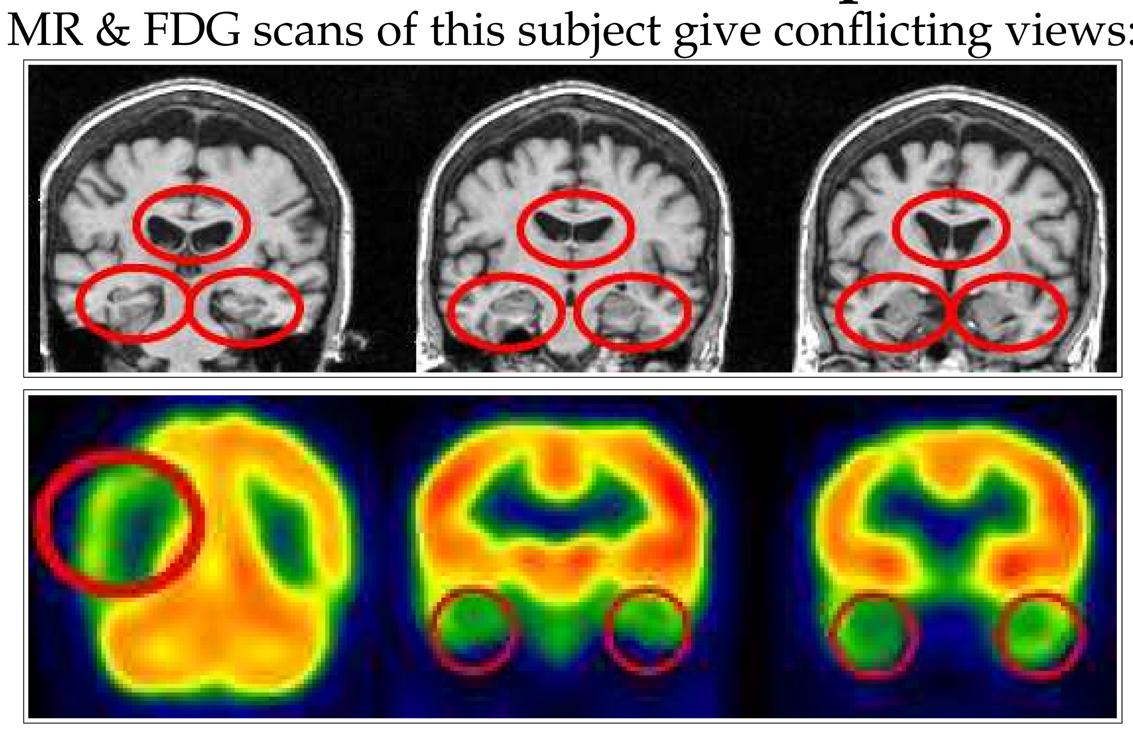
1. Most techniques utilize only one view of the data.
2. Different scanning modalities highlight different aspects of pathology.
3. Their combination is likely to be more accurate.
4. We would also like to detect outliers and reduce their influence in training.

Contributions

1. An efficient *Multi-modality* framework for classifying clinical groups from medical images with outliers
2. Experimental results on the ADNI data set

An illustrative example

MR & FDG scans of this subject give conflicting views:



- If the 2 views disagree, then there must be some error in single view predictions.
- The weights given to different modalities should not all be the same.

Classical SVM formulation

(primal)

$$\min_{\mathbf{w},\xi} \frac{||\mathbf{w}||}{2} + C\sum_i \xi_i$$
$$\text{s.t. } y_i \mathbf{w}^T x_i + \xi_i \geq 1 \; \forall i$$
$$\xi_i \geq 0 \; \forall i$$

(dual)

$$\max_{\alpha} \sum_i \alpha_i - \sum_{i,j} \alpha_i \alpha_j y_i y_j \underbrace{x_i^T x_j}_{\text{kernel}}$$
$$\text{s.t. } 0 \leq \alpha_i \leq C \; \forall i$$
$$\sum_i y_i \alpha_i = 0 \; \forall i$$

- Note the **kernel** function in the dual problem.
- We extend this formulation to use multiple kernels and suppress outliers.
- Robustness to outliers can be added by replacing the ξ hinge-loss function with a robust ηhinge-loss:

$$\eta\text{-hinge}(\mathbf{w}, \mathbf{x}, y) = \eta(1 - y\mathbf{w}^\mathbf{T}\mathbf{x})_+ + (1 - \eta)$$

Related works

1. *Sonnenburg, Raetsch* introduced an efficient MKL implementation
2. *Xu, Schuurmans* introducted Robust SVMs which detect outliers and reduce their influence in training

Standard MKL formulation

Based on *Sonnenburg, et. al.*

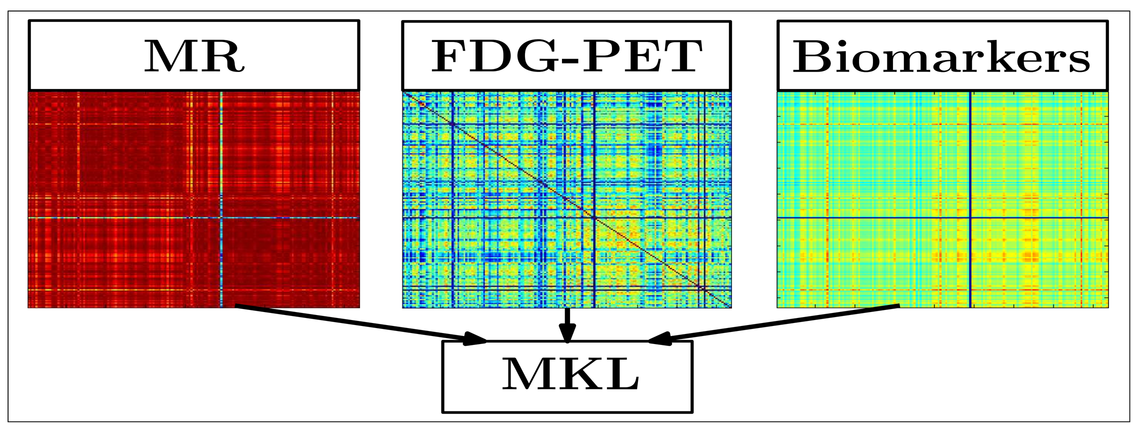
(primal)

$$\min_{\mathbf{w}_{1:K},\xi,b} \frac{1}{2}\left(\sum_{k=1}^{K} ||\mathbf{w_k}||_2\right)^2 + C\sum_{i=1}^{N}\xi_i$$
$$\text{s.t. } \xi_i \geq 0$$
$$y_i\left(\sum_{k=1}^{K} \mathbf{w_k}^\mathbf{T}\phi_\mathbf{k}(\mathbf{x_i}) + b\right) \geq 1 - \xi_i, \forall i$$

(dual)

$$\min_{\gamma,\alpha} \gamma - \sum_{i=1}^{N}\alpha_i$$
$$\text{s.t. } 0 \leq \alpha \leq C$$
$$\frac{1}{2}\sum_{i,j=1}^{N} \alpha_i\alpha_j y_i y_j \mathbf{k_k}(\mathbf{x_i}, \mathbf{x_j}) \geq \gamma, \; \forall k$$

- The 1-norm applied to separate blocks of weights encourages sparsity. The 2-norm applied to weights within each kernel does not.
- For Multi-modality, we compute kernels (e.g., linear Gaussian, etc.) for all modalities.



Robust MKL formulation

Our idealized formulation incorporates both of these ideas:

$$\min_{\eta} \min_{\mathbf{w},\xi} \sum_k ||w_k||_2 + C\sum_i \xi_i + D\sum_{i,k}\eta_{i,k}$$
$$\text{s.t. } y_i\left(\sum_{k=1}^{K} \mathbf{w}^\mathbf{T}\phi_\mathbf{k}(\mathbf{x_i})\right) \geq 1 - \xi_i \; \forall i = 1 \dots N$$
$$0 \leq \eta \leq 1$$
$$\xi \geq 0$$

- We have extended η so that it now discounts examples separately in each view.
- The problem is *not convex*; we require an approximation in which η is treated as constant.

Approximation Scheme

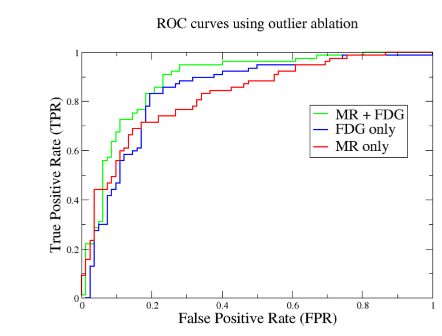
We approximate the optimal setting for η as follows:

$$\eta_{i,k} = \frac{\left(\sum_j \alpha_i\alpha_j y_i y_j K_k(\mathbf{x}_i, \mathbf{x}_j)\right)_-}{\left|\left|\sum_{i'}\left(\sum_{j'} \alpha_{i'}\alpha_{j'} y_{i'} y_{j'} K_k(\mathbf{x}_{i'}, \mathbf{x}_{j'})\right)_-\right|\right|}$$

- The Numerator is the extent to which example i is misclassified in kernel k.
- The Denominator is the total amount of incorrectness among training examples in kernel k. This has the effect of normalizing each kernel separately.
- We implement this by multiplying each kernel matrix, $\mathbf{\Phi}(\mathbf{X})^\mathbf{T}\mathbf{\Phi}(\mathbf{X})$, by $\eta\eta^\mathbf{T}$.

Experimental Results

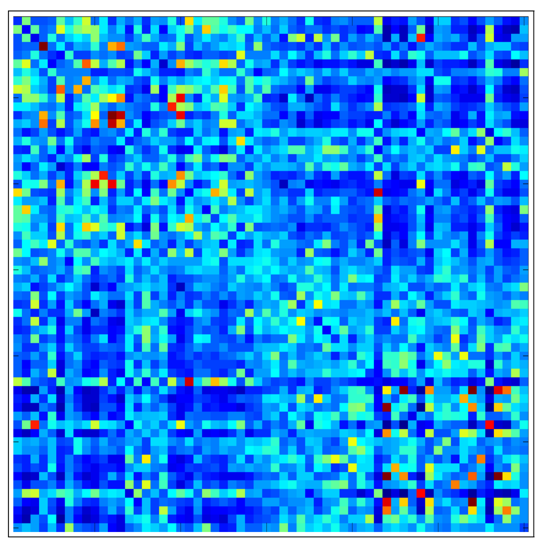
Our Robust Multi-modality framework showed an improvement over single-modality methods on ADNI data.



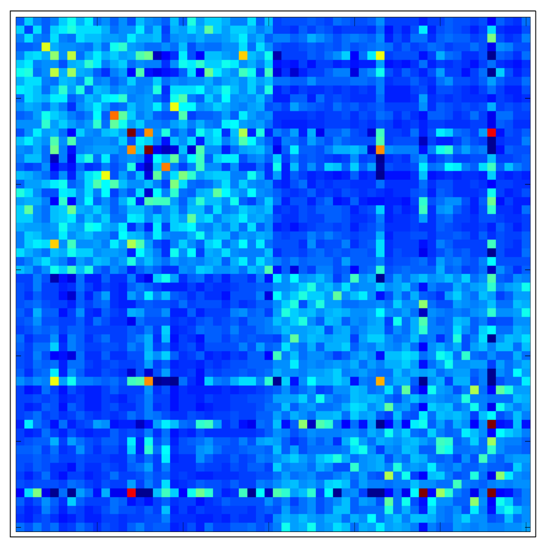
Method	Acc.	Sens.	Spec.	AUC
robust-MKL MR	75.3%	63.1%	81.9%	0.83
robust-MKL FDG	79.4%	78.6%	78.9%	0.84
robust-MKL (multimod	**81.0%**	**78.5%**	**81.8%**	**0.89**

Kernel Matrices

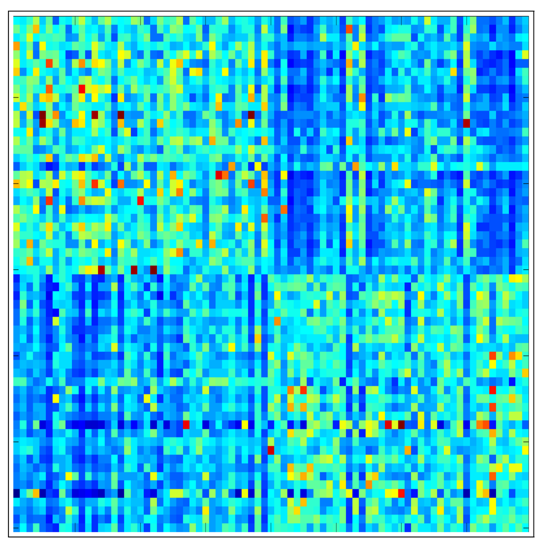
- We can visually inspect the kernel matrices after applying outlier ablation.
- A kernel matrix consists of 3 regions:
  - similarities between positive examples (positive region, top left),
  - similarities between negative examples (negative region, bottom right),
  - similarities between the 2 classes, (interclass region, top right & bottom left).
- Outliers will appear as vertical and horizontal bars in the interclass region.



Sum of unablated kernels. Note the presence of many outliers here.



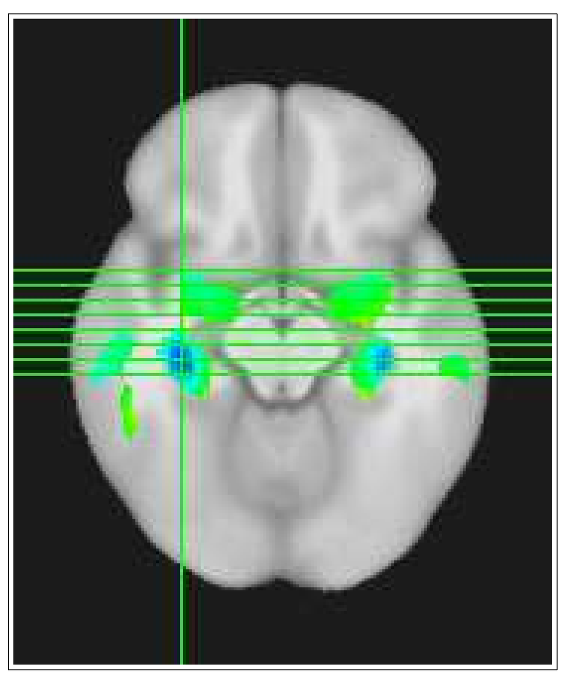
Sum of ablated kernels. The outliers have been visibly ablated here.



Test set outliers appear as vertical lines because they have not been ablated. Relative to test subjects, training outliers' influence has been reduced, (no horizontal lines).

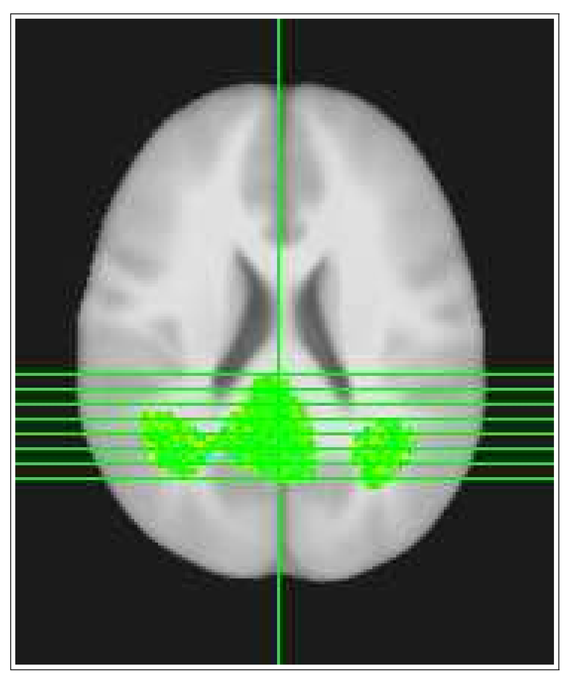
Resulting Classifier

- We can evaluate the derived classifier in terms of the degree of emphasis it puts on relevant brain regions.



MR

In MR the hippocampus and parahippocampal gyri are highlighted.



FDG-PET

In FDG-PET we see the posterior cingulate cortex and lateral parietal lobules bilaterally.

THE UNIVERSITY of WISCONSIN MADISON