CADDementia based on structural MRI using Supervised Kernel-based Representations

D. Cárdenas-Peña, A. Álvarez-Meza, and G. Castellanos-Dominguez

Signal Processing and Recognition Group
Universidad Nacional de Colombia
Campus La Nubia, km 7 via al Magdalena, Manizales-Colombia
{dcardenasp, amalvarezme, cgcastellanosd}@unal.edu.co

Abstract. Magnetic Resonance Image (MRI) analysis allows to find patterns on physiological and pathological processes. Nevertheless, extracting MRI information posses a challenge due to high-dimensional voxel-wise feature spaces. As an alternative, a new supervised kernel-based representation approach for MRI discrimination is proposed. In this sense, the shape features of brain structures is highlighted by the inherent inter-slice similarities of a 3D MRI volume. For testing the proposed approach, an SVM classifier is trained using the ADNI dataset and tested on the CADDementia images for dementia diagnosis. Attained classification results (86% average accuracy on ADNI and 83% on CADDementia) prove that our fully automatic methodology is able to discriminate dementia patients.

1 Introduction

The use of Brain Magnetic Resonance Images (MRI) allows analyzing the influence of physiological and pathological processes on structural or functional properties of brain regions [1]. Specifically, for Alzheimer's Disease (AD), which is the leading form of dementia, it is well-known that structures as grey matter and hippocampus tend to be reduced [6]. Such facts have raised the interest on the development automatic MRI discrimination tools for supporting the AD diagnosis.

Some approaches in MRI discrimination are the following: In [2], an introduced mean shift algorithm is employed to perform atlas stratification to determine whether each population is best represented the considered multi-modal distribution. In [1], a ranked atlas selection is performed by computing image similarities among subject images based on measures like sums of squared differences (SSD), cross-correlation, or mutual information. In any case, those estimators do not guarantee convergence due to the involved highly-dimensional spaces.

Here, to improve MRI discrimination, we propose a new kernel-based representation based on the computed inherent Inter-Slice Kernel (ISK) relationship that makes prominent brain structure distributions. Specifically, we compare three different types of ISK-based feature representation to estimate pairwise

MRI similarities using generalized Euclidean metrics. We tune all needed metric parameters by means of a centered alignment approach, so that the obtained kernels resemble the most prior demographic information [4,3]. The proposed approach is tested on MRI data discrimination using dementia categories (namely, Normal Control (NC), Late Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD)).

2 Proposed Algorithm Description

2.1 MRI Representation based on Inter-Slice Similarities

A 3D Magnetic Resonance Image (MRI) volume comprises a spatially structured set of intensity voxels $\Psi=\{x_{\boldsymbol{r}}\in\mathbb{R} : \boldsymbol{r}=(i,j,k)\}$, where $x_{\boldsymbol{r}}$ is the magnetic field intensity measured at location $\boldsymbol{r}\in\boldsymbol{R}\subset\mathbb{N}^3$. Provided this spatial structure, we describe the MRI volume as an ordered set of 2D slices along the axis views as:

$$\Psi = \{\boldsymbol{X}_i^v \in \mathbb{R}^{L_{v'}\times L_{v''}} : i \in \{1,\cdots,L_v\}\} \tag{1}$$

where v is each one of the considered axes, noted as: *axial*, a, *sagittal* $-\ s$, and *coronal* $-\ c$, i indexes the slices, L_v corresponds to the volume size in the considered axis and $L_{v'}$, $L_{v''}$ are volume sizes in the remaining axes. The arrangement in Eq (1) provides a useful way to analyze MRIs by medical specialists since to read information on the whole 3D volume is harder than on a single 2D slice. So, we take advantage of this introduced slice view and propose the use of the Inter-Slice Kernel (ISK) to encode pairwise similarities of the image slice set in the vector, $\boldsymbol{s}^v\in\mathbb{R}^{P_v}$, with elements described as: $s_{ij}^v=\kappa_{\boldsymbol{X}}\left\{\mathrm{d_X}\left(\boldsymbol{X}_i^v,\boldsymbol{X}_j^v\right) : \forall i<j\right\}$, where $\mathrm{d_X}:\mathbb{R}^{L_v'\times L_v''}\times\mathbb{R}^{L_v'\times L_v''}\to\mathbb{R}$ is a used distance operator for implementing the positive definite kernel function $\kappa_{\boldsymbol{X}}\{\cdot\}$, and $P_v=L_v(L_v-1)/2$, so $\boldsymbol{s}^v$ becomes the ISK representation of the image Ψ along each axis v. It is worth noting that the ISK representation becomes much smaller than the original image space, i.e., $L_v(L_v-1)/2\ll L_vL_{v'}L_{v''}$.

2.2 Learning MRI Similarities by Kernel Centered Alignment

We establish an MRI Similarity Kernel (MSK), $\boldsymbol{K}^v\in\mathbb{R}^{N\times N}$ from a set of MRI volumes $\{\Psi_n : n\in\{1,\ldots,N\}\}$, that is the ISK matrix version representing high-dimensional image information along the axes, where N is the number of considered MRIs. Specifically, we perform MRI similarities, for every axis v, by computing each pairwise relationship, $k_{nm}^v\in\boldsymbol{K}^v$, between the ISK-based features as:

$$k_{nm}^v = \kappa_{\boldsymbol{s}}\left\{\mathrm{d_{SA}}\left(\boldsymbol{s}_n^v,\boldsymbol{s}_m^v\right)\right\} : n,m\in\{1,\ldots,N\} \tag{2}$$

where $\mathrm{d_{SA}}:\mathbb{R}^{P_v}\times\mathbb{R}^{P_v}\to\mathbb{R}$ is a certain a distance operator implementing the positive definite kernel function $\kappa_{\boldsymbol{s}}\{\cdot\}$. In order to reveal the main ISK relationships for learning MRI similarities, we rely on the Mahalanobis distance defined in P_v-dimensional space with inverse covariance matrix $\boldsymbol{A}^v\boldsymbol{A}^{v\top}$ as:

$$\mathrm{d_{SA}}^2(\boldsymbol{s}_n^v,\boldsymbol{s}_m^v) = \left(\boldsymbol{s}_n^v-\boldsymbol{s}_m^v\right)\boldsymbol{A}^v\boldsymbol{A}^{v\top}\left(\boldsymbol{s}_n^v-\boldsymbol{s}_m^v\right)^\top. \tag{3}$$

where matrix $\boldsymbol{A}^v \in \mathbb{R}^{P_v \times D_v}$ holds the linear projection $\boldsymbol{\nu}_n^v = \boldsymbol{s}_n^v \boldsymbol{A}^v$, with $\boldsymbol{\nu}_n^v \in \mathbb{R}^{D_v}$ and $D_v \leq P_v$. Moreover, we propose to learn the matrix $\boldsymbol{A}^v$ based on the already estimated ISK-based feature similarities and by adding prior subject diagnosis information enclosed in the matrix $\boldsymbol{B} \in \mathbb{R}^{N \times N}$. Thus, we measure the dependence between both matrices $\boldsymbol{K}^v$ and $\boldsymbol{B}$ through the following kernel target centered alignment function [3,4]:

$$\rho\left(\boldsymbol{K}^v, \boldsymbol{B}\right) = \frac{\langle \boldsymbol{H}\boldsymbol{K}^v\boldsymbol{H}, \boldsymbol{H}\boldsymbol{B}\boldsymbol{H} \rangle_F}{\|\boldsymbol{H}\boldsymbol{K}^v\boldsymbol{H}\|_F \, \|\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}\|_F}, \; \rho \in [0, 1] \tag{4}$$

where $\boldsymbol{H} = \boldsymbol{I} - N^{-1}\mathbf{1}\mathbf{1}^\top$, with $\boldsymbol{H} \in \mathbb{R}^{N \times N}$, is a centering matrix, $\mathbf{1} \in \mathbb{R}^N$ is an all-ones vector, and notations $\langle \cdot, \cdot \rangle_F$ and $\|\cdot, \cdot\|_F$ stand for the Frobenius inner product and norm, respectively. Generally, the centered version of the alignment coefficient in Eq. (4) gets better correlation estimates than its uncentered version [4,3].

Therefore, we propose to learn MRI similarities from ISK-based features taking advantage of the Kernel Center Alignment (KCA) cost function described in Eq. (4). In this sense, prior patient information, e.g, demographic data as age and gender, can be employed to reveal MRIs dependencies by learning the matrix $\boldsymbol{A}^v$ that parameterizes a Mahalanobis distance between pairwise images (see Eq. (3)). Thereby, given a demographic-based similarity matrix $\boldsymbol{B}$, a KCA-based function can be formulated to compute the projection matrix $\boldsymbol{A}^v$ in Eq. (2) as:

$$\boldsymbol{A}^{v*} = \underset{\boldsymbol{A}^v}{\operatorname{argmax}} \, \rho\left(\boldsymbol{K}^v_{\boldsymbol{A}^v}, \boldsymbol{B}\right), \tag{5}$$

where $\boldsymbol{K}^v_{\boldsymbol{A}^v}$ is the resulting MSK matrix for a provided $\boldsymbol{A}^v$ projection as given in Eq. (2). Consequently, we term each $\boldsymbol{K}^v_{\boldsymbol{A}^{v*}}$ as a Learned MRI Similarity Matrix (LMSK).

2.3 MRI Discrimination using LMSK

The proposed LMSK is learned from an MRI Mahalanobis distance as in Eq. (4). With regard to the needed kernel functions, because of its universal approximating capability [5], we choose the well-known Gaussian kernel noted as follows:

$$g\left\{\mathrm{d}_{\mathrm{z}}\left(z, z'\right); \sigma\right\} \triangleq \exp\left(-\,\mathrm{d}_{\mathrm{z}}\left(z, z'\right)^2 / (2\sigma^2)\right),$$

where $\sigma \in \mathbb{R}^+$ is the kernel bandwidth; $z, z' \in \mathcal{Z}$ is a sample pair in a given feature space $\mathcal{Z}$, and $\mathrm{d}_{\mathrm{z}} : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ is a distance operator in $\mathcal{Z}$. In this sense, we calculate each ISK-based feature vector $\boldsymbol{s}^v$ from MRI using the Frobenious norm:

$$s_{ij}^v = g\left(||\boldsymbol{X}_i^v - \boldsymbol{X}_j^v||_F; \sigma_{s_v}\right). \tag{6}$$

Afterwards, we calculate each $\boldsymbol{K}^v$ matrix encoding pairwise MRI relationship as in Eq. (2), yielding:

$$k_{nm}^v = g\left(\mathrm{d}_{\mathrm{SA}}(\boldsymbol{s}_n^v, \boldsymbol{s}_m^v); \sigma_{SA_v}\right). \tag{7}$$

We optimize the KCA-based cost function in Eq. (5) to learn $\boldsymbol{A}^v$ by a gradient descent solver, where the initial feasible solution is calculated by the Principal Component Analysis algorithm. In addition, the elements of the label kernel $\boldsymbol{B}$ are set as: $b_{nm}{=}\delta\left(c_n - c_m\right)$, being δ the delta function and where $c_n{\in}\{1, 2, \ldots, C\}$ is the label of Ψ_n. Namely, three classes are considered from the `CADDementia Challenge` dataset: Alzheimer's disease (AD), late mild cognitive impairment (MCI) and healthy controls (NC). It is worth noting that every kernel bandwidth in Eqns. (6) and (7) must be properly tuned. Since the variability of the Gaussian kernel $g(\cdot;\sigma)$ tends to zero whenever the kernel bandwidth tends to either zero or infinity to get an appropriate σ value spanning widely all similarity values, we propose to adjust the Gaussian kernel bandwidth employing the following criterion (Notation $\mathrm{var}(\cdot)$ stands for the variance operator):

$$\sigma^* = \arg\max_{\sigma} \left\{\mathrm{var}(g(\cdot;\sigma)\right\} \tag{8}$$

3 Dataset and Preprocessing

For training the proposed MRI discrimination approach, the ADNI dataset was employed. Specifically, a subset of 451 3T MRI volumes are considered from subjects aged from 55 to 90 years (148 NC, 205 MCI, and 98 AD). Provided images are bias filtered using the well-known N3 algorithm. As a further preprocessing stage, each image is registered to the MNI305 template by an affine transform to reference the whole dataset to the Talairach space.

4 Results

Figs. 1(a) to 1(c) show a concrete MRI example illustrating all three views. As seen in Figs. 1(d) to 1(e) displaying their corresponding estimated ISK representations, the red corner patches keep the MRI edges with no content, i.e., the background. Moreover, as the Sagittal ISK (see Fig. 1(e)) exhibits symmetry respect to the anti-diagonal, it is clear that such representation is able to keep the head sagittal symmetry. Therefore, due to the kernel shape varies accordingly to the brain structure distribution, we infer that proposed ISK suitably characterizes head shapes.

Using the above proposed feature extraction stage, a new MRI similarity matrix, comparing each pair of subjects, is learned using the LMSK approach. The prior diagnosis information matrix $\boldsymbol{B}$ and the resulting similarity matrix $\boldsymbol{K^v}$ are depicted in Fig. 2(a) and Fig. 2(b) using the coronal axis view ($v = c$), respectively. From both Figures, it can be seen how the resulting kernel resembles the most the prior label matrix. For the sake of visualization, resulting LMSK is decomposed using the well-known PCA and the first three eigen-components are shown in Fig. 2(c), where the three considered classes can be clearly identified.

For the sake of evaluating the proposed approach a Support-Vector-Machine-based classifier is trained over the LMSK representation using the whole ADNI dataset and tested on the CADDementia MRIs. Obtained class-wise Receiver

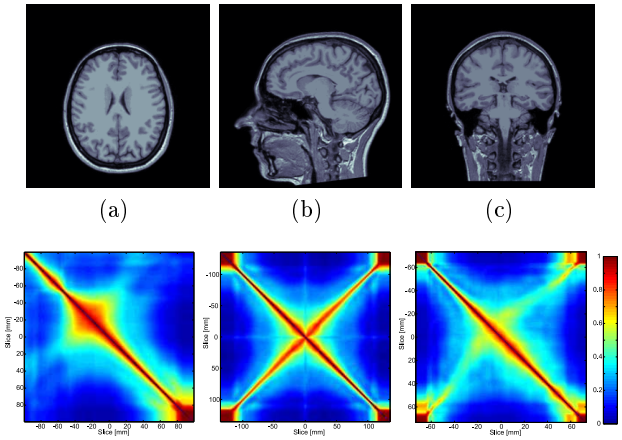


Fig. 1: Database subject and their estimated ISK representation per view

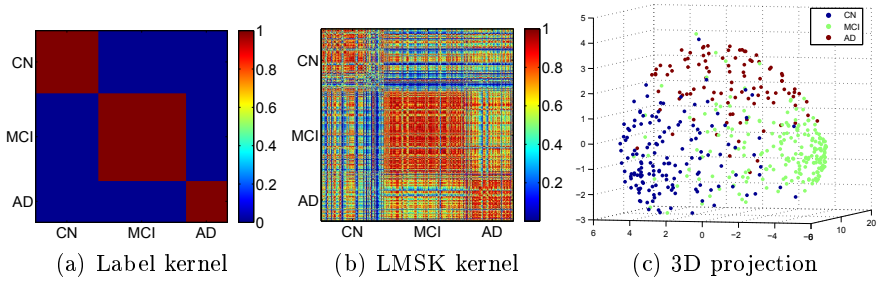
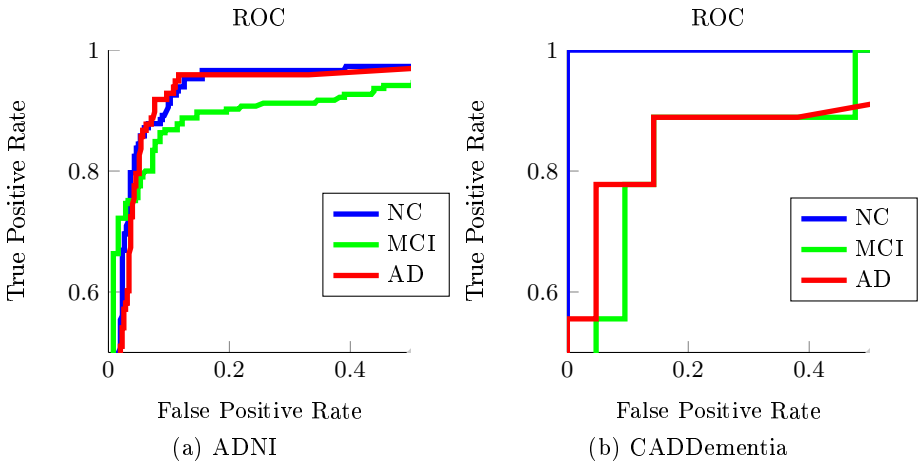


Fig. 2: LMSK results

Operating Characteristic (ROC) curve for both datasets is depicted in Fig. 3, while the confusion matrices are shown in Table 1. The Fig. 3 shows a lower area under the curve for the second class, as the Table 1 shows the lowest accuracy for that class. Both facts imply that MCI subjects are the most difficult to classify, which can be due to the wide spread class distribution (see green subjects in Fig. 2(c)). From a morphological perspective, the low accuracy in MCI subjects can be related to nature of such class. Since MCI is an intermediate class between Healthy and Alzheimer's Disease classes, those subjects tend to be more misdiagnosed than the ones belonging to NC and AD.

5 Conclusion

A new supervised kernel-based image representation is introduced for automatic MRI-based discrimination of dementia. The proposed approach encodes inter-slice similarities, which are related to the shape features of brain structures.



(a) ADNI (b) CADDementia

Fig. 3: Obtained ROC curve for training MRIs in the ADNI and CADDementia datasets.

Table 1: Dementia classification accuracy [%]

(a) ADNI 86%

Class	CN	MCI	AD
CN	87.8	9.8	4.1
MCI	7.4	81.5	7.1
AD	4.8	8.7	88.8

(b) CADDementia 83.3%

Class	CN	MCI	AD
CN	100	0	0
MCI	0	0.66	0.34
AD	0	22.2	77.8

Table 2: Computation time per subject of all methodology stages

Stage	Time [s]
Registration (Rigid)	21.60 ± 7.74
Feature extraction (LMSK)	0.05 ± 0.01
Classification (SVM)	0.65 ± 0.01
Total	22.3

Furthermore, an SVM is trained using the LMSK representation for classifying three dementia categories (NC, MCI and AD). Taking into account the obtained results over the ADNI and CADDementia datasets, our proposed representation proves to find the natural inherent distributions of MRI. The methodology achieves 86% average classification accuracy on the ADNI training set and 83% on the CADDementia, using the coronal axis view. It is important to highlight that all stages on the current methodology are fully automatic. As future work, we plan to evaluate axis view combination strategies and the inclusion of patient demographic data for enhancing the representation and class separability.

References

1. Aljabar, P., Heckemann, R.a., Hammers, a., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage **46**(3) (July 2009) 726–38
2. Blezek, D.J., Miller, J.V.: Atlas stratification. Medical image analysis **11**(5) (October 2007) 443–57
3. Brockmeier, A., Choi, J., Kriminger, E., Francis, J., Principe, J.: Neural decoding with kernel-based metric learning. Neural Computation **26** (2014) –
4. Cortes, C., Mohri, M., Rostamizadeh, A.: Algorithms for learning kernels based on centered alignment. The Journal of Machine Learning Research **13** (2012) 795–828
5. Liu, W., Principe, J.C., Haykin, S.: Kernel Adaptive Filtering: A Comprehensive Introduction. Volume 57. John Wiley & Sons (2011)
6. Rao, A., Lee, Y., Gass, A., Monsch, A.: Classification of Alzheimer ' s Disease from Structural MRI using Sparse Logistic Regression with Optional Spatial Regularization. Evaluation (3) (2011) 4499–4502