**University of Dundee**

**Joint Prediction and Classification of Brain Image Evolution Trajectories from Baseline Brain Image with Application to Early Dementia**

Gafurolu, Can; Rekik, Islem; Alzheimer's Disease Neuroimaging Initiative

# Joint Prediction and Classification of Brain Image Evolution Trajectories from Baseline Brain Image with Application to Early Dementia

Can Gafuroğlu, Islem Rekik*, and for the Alzheimer's Disease Neuroimaging Initiative

BASIRA lab, CVIP group, School of Science and Engineering, Computing, University of Dundee, UK

**Abstract.** Despite the large body of existing neuroimaging-based studies on brain dementia, in particular mild cognitive impairment (MCI), *modeling and predicting* the early dynamics of dementia onset and development in healthy brains is somewhat overlooked in the literature. The majority of computer-aided diagnosis tools developed for classifying healthy and demented brains mainly rely on either using single timepoint or longitudinal neuroimaging data. Longitudinal brain imaging data offer a larger time window to better capture subtle brain changes in early MCI development, and its utilization has been shown to improve classification and prediction results. However, typical longitudinal studies are challenged by a limited number of acquisition timepoints and the absence of inter-subject matching between timepoints. To address this limitation, we propose a novel framework that learns how to *predict* the developmental trajectory of a brain image from a single acquisition timepoint (i.e., baseline), while *classifying* the predicted trajectory as 'healthy' or 'demented'. To do so, we first rigidly align all training images, then extract 'landmark patches' from training images. Next, to predict the patch-wise trajectory evolution from baseline patch, we propose two novel strategies. The first strategy learns in *a supervised manner* to select a few training atlas patches that best boost the classification accuracy of the target testing patch. The second strategy learns in *an unsupervised manner* to select the set of most similar training atlas patches to the target testing patch using multi-kernel patch manifold learning. Finally, we train a linear classifier for each predicted patch trajectory. To identify the final label of the target subject, we use majority voting to aggregate the labels assigned by our model to all landmark patches' trajectories. Our image prediction model boosted the classification performance by 14% without further leveraging any enhancing methods such as feature selection.

## 1 Introduction

The early detection of Alzheimer's Disease (AD) allows patients to undergo preventive treatment to minimize the impact of the disease on their lives. In particular, distinguishing early Mild Cognitive Impairment (eMCI), which is the earliest form of AD, from

---

* Corresponding author: irekik@dundee.ac.uk, www.basira-lab.com

normal controls (NC) is a challenging task, due to the subtlety of the anatomical differences between NC and eMCI subjects. However, this classification problem has great significance due to AD being the most common cause of dementia [1, 2]. Presently, AD is diagnosed using neuroimaging along with measures of cognitive performance [2]. This process places a burden on the experts as they must examine structural Magnetic Resonance Imaging (MRI). The atrophy patterns in early MCI patients in particular are subtle and these patients show no signs of cognitive impairment aside from minor memory concerns. Applying machine learning to automate the diagnosis of eMCI can alleviate the burden on these experts and provide consistent interpretations of the MRI data. As such, there have been a vast number of studies that aim to apply machine learning methods on neuroimaging data to make predictions and diagnoses regarding dementia, some of which use data from a single time point [3–5] and some of which utilize data from multiple time points [6–9].

[6] used similarity maps between aligned baseline and follow-up images to predict progression from MCI to AD by classifying stable MCI patients against MCI converters. [7] used incomplete longitudinal data from MCI subjects in conjunction with sparse learning algorithms and found that the longitudinal data improves their classification accuracy for identifying converters. [8] proposed a temporally structured support vector machine (SVM) classifier, which is designed specifically to work with longitudinal MRI data, with no limit on the number of follow-ups that can be used and achieved state-of-the-art performance in classifying MCI converters. [9] used data from multiple modalities measured at multiple time points in order to predict progression from MCI to AD using an SVM classifier. This work nicely modeled the relationship between data acquired at different time points, where the estimations of derived features between time points was introduced in the learning model, such as cortical thinning speed. By incorporating multiple time points into their framework, all of these works leveraged additional relevant information to improve classification accuracy. However, these methods are limited by the requirement of more than a single acquisition timepoint; hence, early MCI diagnosis or MCI conversion to AD prediction from baseline data is impeded. As preventive treatment is more likely to succeed the earlier the disease is detected, requiring subjects to wait for multiple measurements at different time points may hinder their treatment and recovery processes.

On the other hand, several works focused on using a single acquisition timepoint for dementia diagnosis and classification, which avoids the limitation of requiring patients to wait for multiple scans. For instance, [3] used SVMs to classify AD against NC from MRI images. Similarly, [4] used SVM classifiers with data acquired using varying scanning equipments. [5] compared ten previously explored methods of classification on a single dataset with three different AD-related classification problems: classifying AD against NC, MCI against NC, as well as classifying MCI converters against stable MCI patients. However, these works do not integrate longitudinal data into their frameworks, thereby foregoing the potential improvement in classification accuracy that could be achieved with information from additional timepoints.

To solve this issue, we propose to diagnose a patient in an early stage from their baseline image; however, more reliably by leveraging longitudinal information *predicted* at later timepoints. Specifically, we unprecedentedly propose a joint image evo-
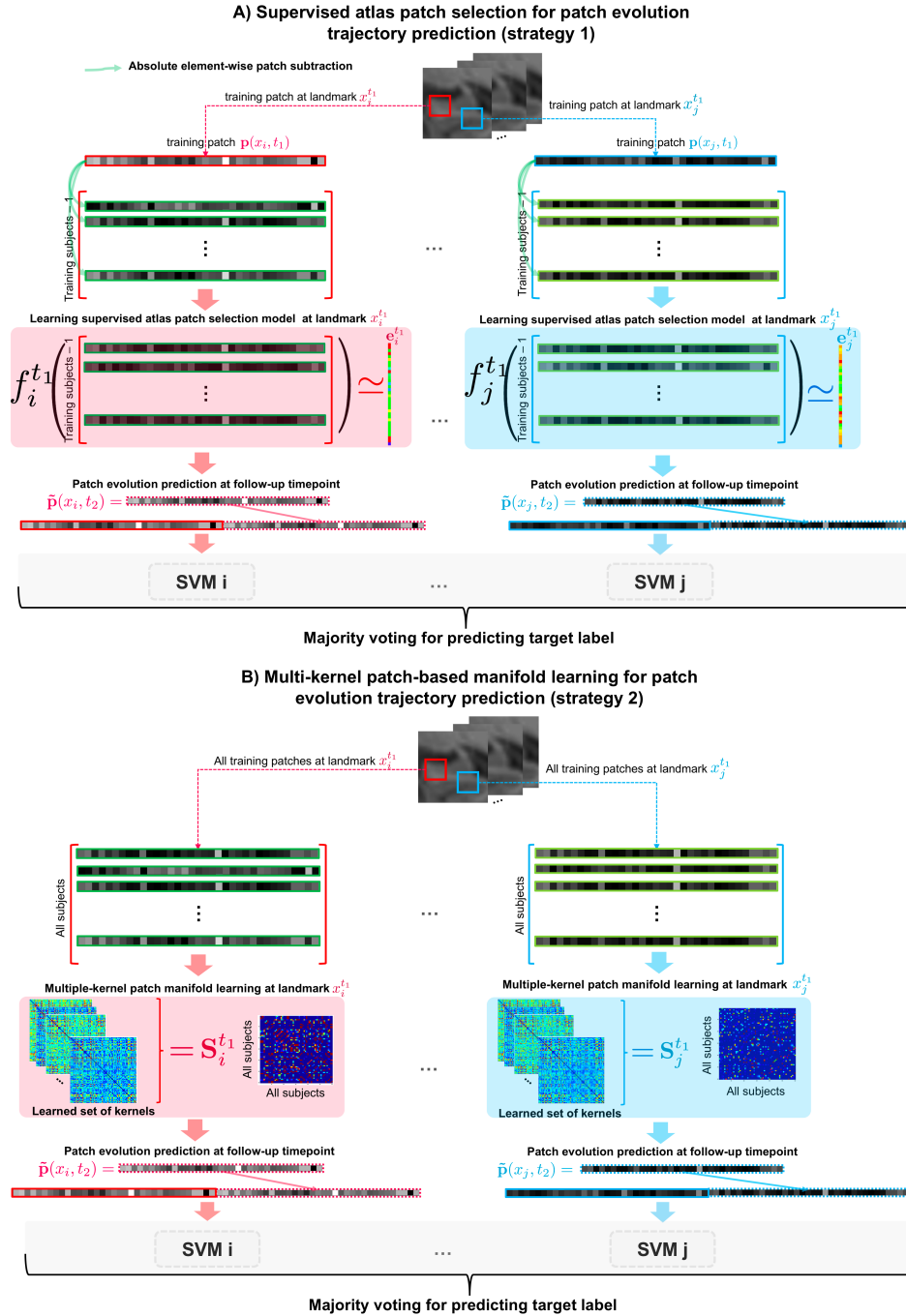
lution trajectory prediction and classification framework from a single acquisition time-point. Our framework comprises four steps. First, we detect key landmarks in the target anatomical region of interest (ROI) across all training subjects at baseline $t_1$. Second, for each baseline patch centered at a specific landmark, we learn how to predict its evolution trajectory at follow-up timepoints. To do so, we propose two different novel strategies. The first strategy learns in a supervised way how to select baseline 'atlas patches' that are expected to yield the smallest patch prediction error at a follow-up timepoint $t_2$ for an input testing patch. The second strategy learns in an unsupervised way how to nest both training and testing patches in a high-dimensional manifold using multiple-kernel learning. Next, we select the closest baseline training patches (or atlas patches) on the manifold to the testing patch. Following the selection of the best training atlas patches at baseline $t_1$, we simply average their corresponding atlas patches at $t_2$ to predict the testing patch at $t_2$. Ultimately, we train an ensemble classifier to predict the label for each learned patch-wise evolution trajectory, which are then aggregated using majority voting to classify the target subject.

## 2 Joint Prediction and Classification of Brain Image Evolution Trajectories from Baseline

In this section, we present the supervised and unsupervised learning strategies proposed for jointly predicting and labeling brain image evolution trajectories from a single acquisition timepoint. We denote matrices by boldface capital letters, e.g., $\mathbf{X}$, and vectors are denoted by bold lowercase letters, e.g., $\mathbf{x}$. **Fig.** 1 illustrates the pipeline for each of the proposed strategies based on a supervised or unsupervised selection of atlas patches to predict the evolution trajectory of a baseline testing patch. In the training stage, both strategies share a fundamental step, which consists of detecting key landmarks across all training images to learn a landmark-wise prediction and classification model.

**Proposed landmark detection method.** Since brain disorders affect the morphology of a particular anatomical brain region, we expect that the regions at the boundary (or edge) of the target ROIs capture the most discriminative information. Hence, for each training subject, we apply a Sobel filter to the training label map of the target ROI to detect its edge. Next, we average all training edges to generate an 'edge density' at each voxel. Ultimately, by thresholding the edge density map, we generate the key landmarks that represent the centers of our training and testing patches for learning each of our supervised and unsupervised image prediction models.

**Supervised atlas patch selection for patch evolution trajectory prediction (strategy 1).** Inspired by the work of [10] on atlas selection learning for brain image segmentation, we propose a patch-based atlas selection strategy that learns in a supervised manner how to minimize the prediction error for a testing patch evolution trajectory. For each target landmark $x_i$, we learn how to select the best training atlas patches at baseline timepoint $t_1$ that minimize the prediction error for a target testing patch at a follow-up timepoint $t_k$, $k \geq 2$. Our learning model assumes that two similar intensity patches at baseline will evolve similarly at follow-up timepoints. Specifically, given $n-1$ training subjects, we first build an intensity disparity matrix of size $(n-2)^2$ at $t_1$, where each row denotes an element-wise absolute difference between two training

**A) Supervised atlas patch selection for patch evolution trajectory prediction (strategy 1)**

Absolute element-wise patch subtraction

training patch at landmark $x_i^{t_1}$     training patch at landmark $x_j^{t_1}$

training patch $\mathbf{p}(x_i, t_1)$     training patch $\mathbf{p}(x_j, t_1)$

Training subjects

Learning supervised atlas patch selection model at landmark $x_i^{t_1}$   $\mathbf{e}_i^{t_1}$

Learning supervised atlas patch selection model at landmark $x_j^{t_1}$   $\mathbf{e}_j^{t_1}$

$$f_i^{t_1}\left(\ \right) \simeq$$

$$f_j^{t_1}\left(\ \right) \simeq$$

Patch evolution prediction at follow-up timepoint
$\tilde{\mathbf{p}}(x_i, t_2) =$

Patch evolution prediction at follow-up timepoint
$\tilde{\mathbf{p}}(x_j, t_2) =$

**SVM i**    ...    **SVM j**

**Majority voting for predicting target label**

**B) Multi-kernel patch-based manifold learning for patch evolution trajectory prediction (strategy 2)**

All training patches at landmark $x_i^{t_1}$     All training patches at landmark $x_j^{t_1}$

All subjects

Multiple-kernel patch manifold learning at landmark $x_i^{t_1}$

Multiple-kernel patch manifold learning at landmark $x_j^{t_1}$

$$= \mathbf{S}_i^{t_1}$$     All subjects

$$= \mathbf{S}_j^{t_1}$$     All subjects

**Learned set of kernels**    All subjects

Patch evolution prediction at follow-up timepoint
$\tilde{\mathbf{p}}(x_i, t_2) =$

Patch evolution prediction at follow-up timepoint
$\tilde{\mathbf{p}}(x_j, t_2) =$

**SVM i**    ...    **SVM j**

**Majority voting for predicting target label**

**Fig. 1:** *Pipeline of the proposed joint prediction and classification framework of brain image evolution trajectories from baseline image using two strategies (A) and (B).* A) *Supervised atlas patch selection strategy.* A regression function $f_i^{t_1}$ is learned at each landmark $x_i$ to map absolute distance vectors between two patches to an error prediction vector. This allows to *learn* how to select the best atlas patches at $t_1$ for our patch evolution trajectory prediction at $t_2$. B) *Unsupervised atlas patch selection strategy.* A patch-based multi-kernel manifold learning is used to select the most similar atlas patches to the target testing patch at $t_1$. The predicted follow-up patches at $t_2$ are then concatenated with the baseline patch at $t_1$ and classified by majority voting using an ensemble of SVM classifiers in both approaches.

patches $\mathbf{p}_{i,s}^{t_1}$ and $\mathbf{p}_{i,s'}^{t_1}$. Next, we build a corresponding error vector $\mathbf{e}_i^{t_1}$ that quantifies the prediction error of using $\mathbf{p}_{i,s}^{t_1}$ for subject $s$ at $t_1$ as an atlas patch to predict $\tilde{\mathbf{p}}_{i,s'}^{t_2}$ for subject $s'$ at $t_2$. Specifically, we define $\tilde{\mathbf{p}}_{i,s'}^{t_2} = \alpha \times p_{i,s'}^{t_2}$, where $\alpha = \mathbf{p}_{i,s'}^{t_1}/\mathbf{p}_{i,s}^{t_1}$. If $\mathbf{p}_{i,s}^{t_1}$ has a zero-element, then the corresponding element in vector $\alpha$ is set to a high value. As for the prediction error, we define it as the average absolute difference between the ground truth patch $\mathbf{p}_{i,s'}^{t_2}$ and the predicted patch $\tilde{\mathbf{p}}_{i,s'}^{t_2}$ for the 'testing' subject $s'$. Next, in the *training stage*, we learn a support vector regressor function $f_i^{t_1}$ that maps the intensity disparity matrix at each landmark $x_i$ onto the corresponding prediction error vector (**Fig.** 1–A). In the *testing stage*, we compute the pairwise distance between each training patch $\{\mathbf{p}_{i,s}^{t_1}\}_{s=1}^{n-1}$ and the testing patch $\mathbf{p}_{i,tst}^{t_1}$. Then, by testing the learned regression function for each pairwise absolute distance $f_i^{t_1}(|\mathbf{p}_{i,s}^{t_1} - \mathbf{p}_{i,tst}^{t_1}|)$, we predict the error of using subject $s$ to predict $\mathbf{p}_{i,tst}^{t_2}$. Ultimately, we select the top $K$ atlas patches at $t_1$ with the lowest prediction errors, then average their corresponding patches at $t_2$ to output $\tilde{\mathbf{p}}_{i,tst}^{t_2}$.

**Multi-kernel patch-based manifold learning for patch evolution trajectory prediction (strategy 2).** In this strategy, we use the multi-kernel manifold learning (MKML) method proposed in [11], to identify the baseline atlas patches whose follow-up images best represent the testing baseline patch by learning patch-to-patch similarities (**Fig.** 1–B). In this section, we briefly present the MKML framework introduced in [11] and how we extended it to our aim. MKML learns a distance metric in an unsupervised manner by combining multiple Gaussian kernels, and uses it to output a learned pairwise similarity matrix of size $n \times n$ from an input matrix of size $n \times d$ where $n$ is the number of subjects and $d$ is the size of their vectorized patches. For each landmark $x_i$, to predict $\tilde{\mathbf{p}}_{i,tst}^{t_2}$ for a testing subject, we first learn a baseline similarity matrix $\mathbf{S}_i^{t_1}$ for all training and testing samples. This allows to learn an intensity patch manifold where all patch vectors $\{\mathbf{p}_{i,1}^{t_1}, \ldots, \mathbf{p}_{i,n}^{t_1}\}$ are nested. Instead of using one predefined distance metric which may fail to capture the nonlinear relationship in the patch data, we use multiple Gaussian kernels with learned weights to better explore in depth the similarity patterns among patches centered at a fixed landmark across a set of subjects. In other words, adopting multiple kernels allows to better fit the true underlying statistical distribution of the input matrix of intensity patch features. Additionally, constraints are imposed on kernel weights to avoid a single kernel selection [11].

The Gaussian kernel is expressed as follows: $\mathbf{K}(\mathbf{p}_{i,s}^{t_1}, \mathbf{p}_{i,s'}^{t_1}) = \frac{1}{\epsilon_{s,s'}\sqrt{2\pi}}e^{(-\frac{|\mathbf{p}_{i,s}^{t_1} - \mathbf{p}_{i,s'}^{t_1}|^2}{2\epsilon_{s,s'}^2})}$, where $\epsilon_{s,s'}$ is defined as: $\epsilon_{s,s'} = \sigma(\mu_s + \mu_{s'})/2$, where $\sigma$ is a tuning parameter and $\mu_s = \frac{\sum_{l \in KNN(\mathbf{p}_{i,s}^{t_1})} |\mathbf{p}_{i,s}^{t_1} - \mathbf{p}_{i,s'}^{t_1}|}{k}$, where $KNN(\mathbf{p}_{i,s}^{t_1})$ represents the top $k$ neighboring subjects of subject $s$. The computed kernels are then averaged to further learn the similarity matrix $\mathbf{S}_i^{t_1}$ at landmark $x_i$ and baseline timepoint $t_1$ through an optimization framework formulated as follows:

$\min_{\mathbf{S}_i^{t_1},\mathbf{L},\mathbf{w}} \sum_{k,j} -w_l\mathbf{K}_l(\mathbf{h}^k,\mathbf{h}^j)\mathbf{S}_i^{t_1}(k,j) + \beta||\mathbf{S}_i^{t_1}||_F^2 + \gamma\mathbf{tr}(\mathbf{L}^T(\mathbf{I}_n - \mathbf{S}_i^{t_1})\mathbf{L}) + \rho\sum_l w_l log w_l$

Subject to: $\sum_l w_l = 1$, $w_l \geq 0$, $\mathbf{L}^T\mathbf{L} = \mathbf{I}_c$, $\sum_j \mathbf{S}_i^{t_1}(k,j) = 1$, and $\mathbf{S}_i^{t_1}(k,j) \geq 0$ for all $(k,j)$, where:

- $\sum_{k,j} -w_l \mathbf{K}_l(\mathbf{h}^k, \mathbf{h}^j) \mathbf{S}_i^{t_1}(k,j)$ refers to the relation between the similarity and the kernel distance with weights $w_l$ between two subject-specific patches. The learned similarity should be small if the distance between a pair of patches is large.

- $\beta \|\mathbf{S}_i^{t_1}\|_F^2$ denotes a regularization term that avoids over-fitting the model to the data.

- $\gamma \mathbf{tr}(\mathbf{L}^T (\mathbf{I}_n - \mathbf{S}_i^{t_1}) \mathbf{L})$: $\mathbf{L}$ is the latent matrix of size $n \times c$ where $n$ is the number of subjects and $c$ is the number of clusters. The matrix $(\mathbf{I}_n - \mathbf{S}_i^{t_1})$ denotes the graph Laplacian.

- $\rho \sum_l w_l log w_l$ imposes constraints on the kernel weights to avoid selection of a single kernel.
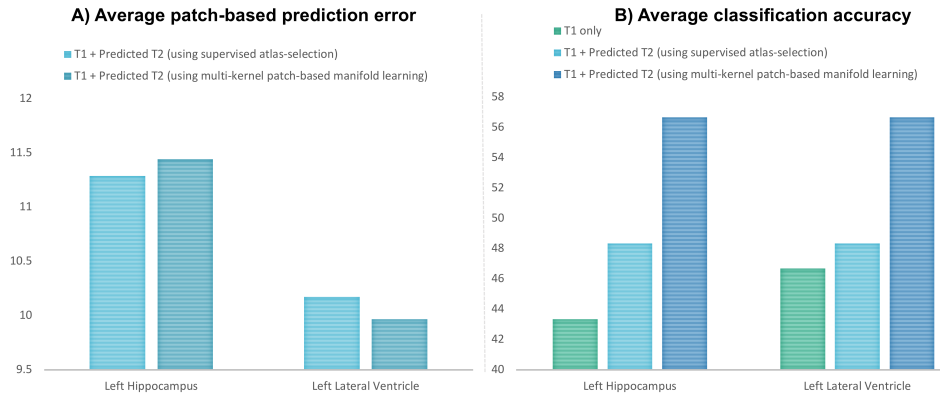
An alternating convex optimization is adopted where each variable is optimized while fixing the other variables until convergence [11]. Finally, based on the landmark-specific learned matrix $\mathbf{S}_i^{t_1}$, we select the top $K$ training patches (or $K$ atlas patches) that are most similar to the target testing patch at baseline. Finally, we predict $\tilde{\mathbf{p}}_{i,tst}^{t_2}$ as a weighted average of corresponding $K$ atlas patches at follow-up $t_2$.

**Predicted trajectories' labeling using an ensemble classifier**. The last shared step in both proposed strategies is to *label* patch evolution trajectory for a testing subejct at each landmark $x_i$ as 'healthy' or 'disordered'. To do so, for each landmark $x_i$, we train a support vector machine (SVM) using the concatenation of baseline training patches $\{\mathbf{p}_{i,s}^{t_1}\}_{s=1}^{n-1}$ and their corresponding predicted patches $\{\tilde{\mathbf{p}}_{i,s}^{t_2}\}_{s=1}^{n-1}$ using strategies 1 or 2. The left out testing subject is then classified using majority voting on predicted labels outputted by all SVMs (i.e, across all landmarks).

## 3   Results and Discussion

**Data and model parameters.** We used data from 30 NC (Normal Control) and 30 eMCI subjects acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (adni.loni.usc.edu). All training baseline and follow-up MR images along with their AAL-atlas based label maps are rigidly aligned to a common space. For evaluation, we tested our methods on landmark patches extracted from the left hippocampus and the left lateral ventricle. We selected these preliminary ROIs due to: (1) their prevalence in the dementia literature [6, 5], and (2) the fact that atrophy rates were found to be faster in the left hemisphere compared to the right hemisphere [12]. We fixed the patch size to $7 \times 7 \times 7$ across all methods. The number of neighbours (the best atlas patches) used for predicting the follow-up trajectory is set to 2 for both strategies. For MKML parameters, we set the number of clusters to $c = 2$, and $m = 2$ kernels.

**Evaluation.** We used leave-one-out cross-validation to evaluate the classification accuracy of the proposed methods. **Fig.** 2–A displays the average patch prediction error computed using the mean absolute error (MAE) between ground-truth and predicted patches at the follow-up timepoint. We note that for both of the selected ROIs, both methods led to similar results, where unsupervised prediction slightly outperformed supervised prediction in the hippocampus as opposed to the ventricle. As for the eMCI/NC classification results (**Fig.** 2–B), integrating predicted follow-up timepoints significantly increased by ∼5% (resp. ∼2%) when using supervised prediction method and by ∼%14 (resp. ∼10%) when using unsupervised prediction strategy for the hippocampus (resp.

**A) Average patch-based prediction error**

- T1 + Predicted T2 (using supervised atlas-selection)
- T1 + Predicted T2 (using multi-kernel patch-based manifold learning)

**B) Average classification accuracy**

- T1 only
- T1 + Predicted T2 (using supervised atlas-selection)
- T1 + Predicted T2 (using multi-kernel patch-based manifold learning)

**Fig. 2:** (A) Prediction accuracy for each ROI using proposed strategies for predicting the follow-up image evolution trajectory from baseline. (B) Classification accuracy by our proposed methods.

the ventricle). Clearly, multi-kernel patch-based manifold learning for patch evolution prediction consistently produced the best classification results. We would like to highlight that the main contribution of this work is to highlight the great potential of predicting follow-up data from baseline data in boosting classification and disorder diagnosis. Both strategies proposed in **Fig.** 1 can be boosted by using enhancing methods such as feature selection or deep-learning approaches.

It should be noted that our supervised atlas selection strategy relies on the assumption that the multiplication of an intensity patch of subject $s$ at baseline by a weighting vector $\alpha$ can produce the intensity patch of a different subject $s'$. Through propagating this rule to follow-up timepoints, we can learn how to predict patch evolution trajectory. However, there is no theoretical proof that this assumption holds aside from the fact that the follow-up data predicted in this way improved our classification accuracy when compared to classifying using the baseline images only. We also note that our framework leverages information from only brain MR images, however brain dementia also atrophies the cortical surface [13]. Hence, based on the seminal shape evolution learning models for predicting infant cortical development from a single timepoint [14] and inspired from the joint shape-image regression model proposed in [15], we aim to build a unified model which simultaneously predicts cortical shape and brain image evolution trajectories for a more accurate early diagnosis from baseline data.

## 4 Conclusion

In this work, we proposed a novel joint image prediction and classification framework for the early diagnosis of MCI. Our initial results show that there is a possibility of improvement in classification accuracy when predicted longitudinal data is leveraged for classification even without additional use of enhancing approaches such as feature selection. Our future work will involve improving the framework by investigating the impact of various feature selection and learning methods, including deep learning architectures. Additionally, we wish to extend our framework by investigating whether timepoint-to-timepoint similarity maps such as those proposed in [6] can further im-

prove our classification accuracy when the follow-up images are not measured, but *predicted* instead.

## References

1. Apostolova, L.G., Thompson, P.M.: Mapping progressive brain structural changes in early alzheimer's disease and mild cognitive impairment. Neuropsychologia **46** (2008) 1597 – 1612
2. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol **6** (2010) 67–77
3. Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S., Benali, H.: Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology **51** (2009) 73–83
4. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.: Automatic classification of MR scans in Alzheimer's disease. Brain **131** (2008) 681–689
5. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. NeuroImage **56** (2011) 766 – 781
6. Sanroma, G., Andrea, V., Benkarim, O.M., Manjón, J.V., Coupé, P., Camara, O., Piella, G., González Ballester, M.A.: Early prediction of Alzheimer's Disease with non-local patch-based longitudinal descriptors. Patch-Based Techniques in Medical Imaging (2017) 74–81
7. Thung, K.H., Wee, C.Y., Yap, P.T., Shen, D.: Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans. Brain Structure and Function **221** (2016) 3979–3995
8. Zhu, Y., Zhu, X., Kim, M., Shen, D., Wu, G.: Early diagnosis of Alzheimer's Disease by joint feature selection and classification on temporally structured support vector machine. Medical Image Computing and Computer-Assisted Intervention (2016) 264–272
9. Zhang, D., Shen, D., Initiative, A.D.N.: Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. PLOS ONE **7** (2012) 1–15
10. Sanroma, G., Wu, G., Gao, Y., Shen, D.: Learning-based atlas selection for multiple-atlas segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014) 3111–3117
11. Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., Batzoglou, S.: Visualization and analysis of single-cell rna-seq data by kernel-based similarity learning. Nature Methods **14** (2017) 414 EP –
12. Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W.: Dynamics of gray matter loss in Alzheimer's disease. J. Neurosci. **23** (2003) 994–1005
13. Möller, C., Pijnenburg, Y.A., van der Flier, W.M., Versteeg, A., Tijms, B., de Munck, J.C., Hafkemeijer, A., Rombouts, S.A., van der Grond, J., van Swieten, J., et al.: Alzheimer disease and behavioral variant frontotemporal dementia: automatic classification based on cortical atrophy for single-subject diagnosis. Radiology **279** (2015) 838–848
14. Rekik, I., Li, G., Yap, P.T., Chen, G., Lin, W., Shen, D.: Joint prediction of longitudinal development of cortical surfaces and white matter fibers from neonatal MRI. NeuroImage **152** (2017) 411–424
15. Fishbaugh, J., Prastawa, M., Gerig, G., Durrleman, S.: Geodesic regression of image and shape data for improved modeling of 4D trajectories. Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on (2014) 385–388