Learning and combining image neighborhoods using random forests for neonatal brain disease classification

Veronika A. Zimmer[a,*], Ben Glocker[b], Nadine Hahner[c,d], Elisenda Eixarch[c,d], Gerard Sanroma[a], Eduard Gratacós[c,d], Daniel Rueckert[b], Miguel Ángel González Ballester[a,e], Gemma Piella[a]

[a]*SIMBioSys, Universitat Pompeu Fabra, Barcelona, Spain*
[b]*BioMedIA Group, Imperial College London, London, UK*
[c]*Fetal i+D Fetal Medicine Research Center, BCNatal - Barcelona Center for Maternal-Fetal and Neonatal Medicine (Hospital Clínic and Hospital Sant Joan de Déu), Institut Clínic de Ginecologia, Obstetricia i Neonatologia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Universitat de Barcelona.*
[d]*Centre for Biomedical Research on Rare Diseases (CIBER-ER), Barcelona, Spain.*
[e]*ICREA, Barcelona, Spain*

Abstract

It is challenging to characterize and classify normal and abnormal brain development during early childhood. To reduce the complexity of heterogeneous data population, manifold learning techniques are increasingly applied, which find a low-dimensional representation of the data, while preserving all relevant information. The neighborhood definition used for constructing manifold representations of the population is crucial for preserving the similarity structure and it is highly application dependent. The recently proposed neighborhood approximation forests learn a neighborhood structure in a dataset based on a user-defined distance. We propose a framework to learn multiple pairwise distances in a population of brain images and to combine them in an unsupervised manner optimally in a manifold learning step. Unlike other methods that only use a univariate distance measure, our method allows for a natural combination of multiple distances from heterogeneous sources. As a result, it yields a representation of the population that preserves the multiple distances. Furthermore, our method also selects the most predictive features associated with the distances. We evaluate our method in neonatal magnetic resonance images of three groups (term controls, patients affected by intrauterine growth restriction and mild isolated ventriculomegaly). We show that combining multiple distances related to the condition improves the overall characterization and classification of the three clinical groups compared to the use of single distances and classical unsupervised manifold learning.

Keywords: Random forest, neighborhood approximation forest, manifold learning, similarity measure, brain development

1. Introduction

During early childhood, the brain undergoes complex structural changes, which makes it challenging to characterize and quantify normal and abnormal brain development. Depending on the condition occurring during the pregnancy, brain structure could have overt lesions or more subtle and general structural changes that could make it difficult to quantify these changes. The diagnostic and subsequent therapy, however, often relies only on one dimensional measurements, such as the width of the ventricles in a specific plane manually determined by the experts. Therefore, there is a growing need for identifying brain imaging biomarkers to improve the characterization of diseases, their diagnosis and therapy. Medical images, such as magnetic resonance (MR) images, provide structural information of the brain which can be analyzed to find changes and disease-specific differences in a population.

*Corresponding author
Email address: veronika.zimmer@upf.edu (Veronika A. Zimmer)

Techniques of manifold learning have been used successfully in medical imaging for various applications (Aljabar et al., 2012). Their aim is to reduce the complexity (dimensionality) of the data while preserving all relevant information (the intrinsic structure of the data). The obtained new representation of the data can facilitate further analysis such as classification, visualization and compression. For an overview of manifold learning techniques see, e.g., the review of van der Maaten et al. (2009). Popular dimensionality reduction and manifold learning methods, both linear (e.g., PCA (Pearson, 1901), MDS (Cox and Cox, 2001)) and non-linear (e.g., kernel PCA (Schölkopf et al., 1998), Isomap (Tenenbaum et al., 2000) and Laplacian Eigenmaps (Belkin and Niyogi, 2003)), require some notion of distance between data samples to capture global and local structures in the population. Often, they optimize an objective function based on the image's neighborhoods to learn the underlying manifold structure.

The neighborhood definition is crucial for the quality of the resulting new representation. This is highly application dependent and a field of on-going research. A typical choice is to calculate the Euclidean distance between data samples to approximate distances on the manifold (in the context of brain imaging the Euclidean distance between the voxels of the image or a region of interest). However, the Euclidean distance may not capture all the differences in a population or be representative of anatomical characteristics. The works by Aljabar et al. (2011), Gray et al. (2013), Wolz et al. (2012) and Zimmer et al. (2015b) have shown that more sophisticated and application-dependent distance definitions improve the overall performance of the methods.

At the same time, there has been much interest in identifying and combining different kinds of heterogeneous information in the definition of image neighborhoods to improve the new representation resulting from manifold learning. The manifold structure of brain images has been estimated by Gerber et al. (2010) based on non-rigid transformations, whereas Aljabar et al. (2008) derived similarities from overlaps of their structural segmentations. Aljabar et al. (2011) combined shape and appearance information in a joint representation for an improved characterization of brain development and Wolz et al. (2012) incorporated clinical information into the manifold learning step. Regional approaches were presented by Bhatia et al. (2014) and Ye et al. (2014) that learned a manifold for image regions rather than for the entire image. To define such pairwise similarities, additional information, e.g., structural segmentations, shape information or non-rigid deformations between images, have to be extracted or estimated.

However, the extraction of additional information, such as structural segmentations and shape information, is computationally expensive. Moreover, it is in general not clear how to combine and weight multiple heterogeneous neighborhood definitions. In multiple kernel learning, different kernels, representing different notions of similarity between information coming from different sources, are combined by learning an optimal linear or non-linear weighting of the kernels (Gönen and Alpaydın, 2011; Sanchez-Martinez et al., 2017; Zimmer et al., 2015b). Wolz et al. (2012) combined non-imaging information with appearance-based image features by including additional edges in the neighborhood graph used for Laplacian Eigenmaps. Aljabar et al. (2011) combined new embedding coordinates of multiple manifold representations, obtained using different information, for a joint representation.

Another powerful method that can be used to approximate distance measurements is a random forest (Breiman, 2001). A random forest is an ensemble of decision trees, where at each node of the trees the data samples are partitioned into two sets, according to a test, whose parameters are learned a priori from training data. A notion of similarity between the data samples can be defined based on the co-occurrences in leaf nodes (Shi and Horvath, 2006) which has been applied successfully to several clustering and classification tasks (Gray et al., 2013; Nowak and Jurie, 2007; Shi et al., 2005). In particular, Gray et al. (2013) used random forests for deriving the pairwise distances from different imaging modalities for manifold learning. Additionally, the most important features for the classification problem were extracted. One of the advantages of using random forests is the feature selection which is done simultaneously to the classification, and can give additional insight to the application at hand.

In this work, we focus on how to define and combine multiple neighborhood structures to obtain a new data representation of a complex population of MR brain images, which best captures all relevant information. Our contributions are threefold. First, we propose to learn the neighborhood structure in brain MR images based on application-specific properties of the population. We employ neighborhood approximation forests (NAFs) (Konukoglu et al., 2013), which are a type of random forests specifically designed to

approximate neighborhoods. Second, we use several definition of similarity, based on multiple user-defined criteria. The corresponding information comes from heterogeneous sources of supervised information (i.e., user-defined distances relevant to the pathology) to take into account the complex nature of the data. The proposed method allows a seamless combination of those multiple neighborhoods. The distances between images are expressed in the same units (co-occurrences on leaf nodes) across heterogeneous data information. Another advantage of using random forests to define image similarity is the ability to define discriminative features for interpretation purposes. To the best of our knowledge, the only work using random forests to approximate distances for manifold learning in medical image analysis is Gray et al. (2013). They take advantage of multiple modalities of the data, whereas we extract multiple neighborhood structures from a single modality. Third, we propose to optimally combine multiple neighborhood structures, based on the quality of the resulting data representation (e.g., by optimizing an embedding quality measure Zimmer et al. (2015b)).

We use NAFs to approximate different definitions of similarities based on the appearance information in MRI and obtain approximated neighborhoods. The pairwise affinities are based on the co-occurrence counts in the leaf nodes of the NAFs. This allows seamlessly combining several neighborhoods induced by arbitrary criteria since they are all represented in the same units (co-occurrence counts) and therefore can be readily combined. The combined affinity matrices are used in a manifold learning step to extract a vector representation of the images that preserves the combined image neighborhoods and encodes different heterogeneous information simultaneously. Thus, the proposed method enables the combination of heterogeneous sources of supervised information and obtain a new representation useful for analyzing the pathology. Moreover, our method automatically extracts relevant features (without the need to engineer them a priori).

We apply the proposed method to a population of neonates. Two conditions are present in this population: intrauterine growth restriction (IUGR) and mild isolated ventriculomegaly (VM). IUGR is due to placental insufficiency and is related to changes in brain size and more subtle changes in several brain structures (Batalle et al., 2012; Esteban et al., 2010; Padilla et al., 2011), which makes prior analysis and feature extraction difficult. VM, on the other hand, is defined by an enlargement of the ventricular system and it has been described as being associated with changes in cortical development and changes in white and gray matter volumes (Gilmore et al., 2008; Lyall et al., 2012; Wyldes and Watkinson, 2004).

As an example of application, the new representation is used to classify different diagnostic groups (CN: normal controls, IUGR and VM). That is, we evaluate how well the resulting low-dimensional data representation separates (using simple kNN classification) the different diagnostic groups. We compare classification accuracy with several methods using single affinities and classical unsupervised manifold learning. The new vector representations obtained by our method are able to separate the three groups (normal controls (NC), IUGR and VM) with a high accuracy. The features used for the neighborhood definition are extracted automatically and reflect the disease's structural changes.

This work extends our previous conference paper (Zimmer et al., 2015a) with a more detailed description of the method, a more sophisticated combination method for the approximated neighborhoods and a more extensive evaluation on new data.

The remainder of the paper is structured as follows: Section 2 presents the proposed pipeline and explains the main parts in detail, including the construction of NAFs, feature selection, and combination in a manifold learning step. Section 3 reports experimental results on a neonatal population, followed by further discussion and conclusions in Section 4.

2. Methodology

The pipeline of the proposed method is shown in Fig. 1. For a given population of images, NAFs are constructed based on multiple distances. These distances are chosen such that they represent different properties of subpopulations. Affinity matrices are derived from the NAFs as neighborhood representations and used to obtain a new representation that best characterizes the original data. In the following, the different parts of the pipeline are explained in detail, namely, the computation of affinities (Fig. 1 (a)-(d)) and the manifold learning (Fig. 1 (e)).

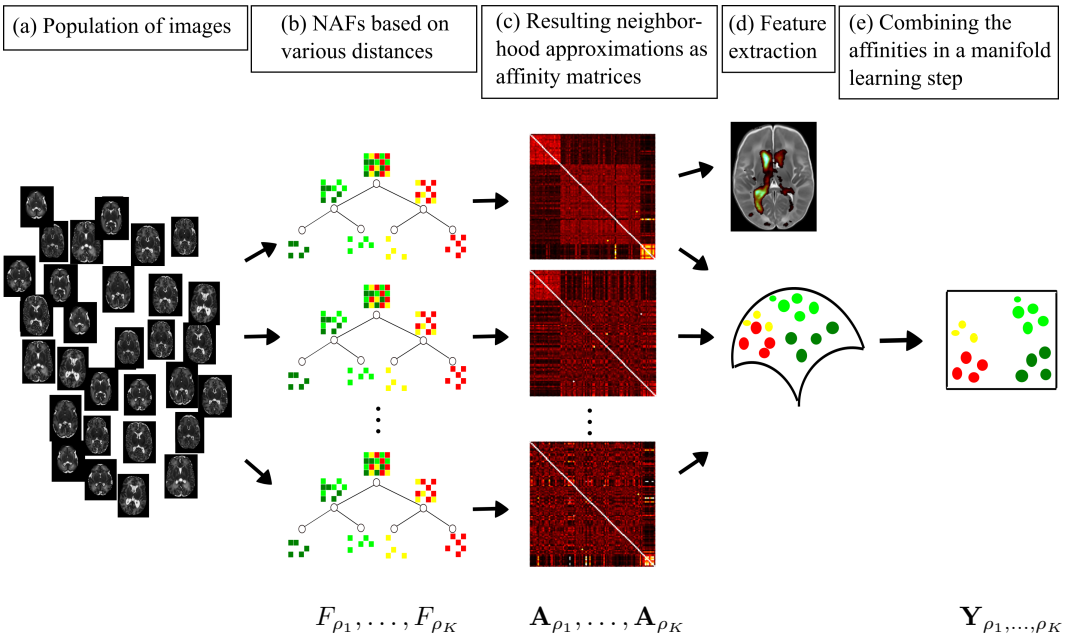

Figure 1: Pipeline of the proposed method. For a given population, NAFs are constructed using various distances to obtain neighborhoods within the population and for feature extraction. The combined information is used for classification.

2.1. Pairwise Image Similarities using Random Forests

In this work, we choose to learn the neighborhood structure of images using NAFs. A NAF learns a neighborhood structure of a given dataset in a supervised manner induced by an arbitrary notion of similarity between images. In the training step, the algorithm learns to cluster the images based on appearance features according to the distance function. For testing, the learned features are used to predict the closest neighbors of a test image in the training database. Given a population $\mathcal{I}$ of images, a subset $\mathbf{I} = \{I_p\}_{p=1}^{P} \in \mathcal{I}$ is used for training and each I_p is represented by a high-dimensional intensity-based feature vector $\mathbf{f}(I_p) \in \mathbb{R}^Q$. The population $\mathcal{I}$ is equipped with a user-defined distance function $\rho : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$ which allows the definition of pairwise distances $\rho(I_m, I_n)$ between the images.

Training phase. In the training phase, T individual trees are constructed. For each tree t, a random subset of features $\mathbf{f}^{(t)} \subset \mathbf{f}$ is selected with $\mathbf{f}^{(t)} \in \mathbb{R}^q$, $q < Q$. At each node of tree t, the algorithm divides the data samples present in the current node into two sets. This branching of the set of images $\mathbf{I}_s^{(t)}$ in node s of tree t is based on a binary test: for $I_n \in \mathbf{I}_s^{(t)}$, $I_n \in \mathbf{I}_{s_R}^{(t)}$ if $\mathbf{f}_m^{(t)}(I_n) > \tau$, and $I_n \in \mathbf{I}_{s_L}^{(t)}$ otherwise. Here, $\mathbf{f}_m^{(t)}(I_n)$ is the mth feature in image I_n in tree t, $\tau \in \mathbb{R}$ is a threshold, and $\mathbf{I}_{s_R}^{(t)}$ and $\mathbf{I}_{s_L}^{(t)}$ are the sets of images splitted to the right and left child node, respectively. For each node in each tree, the binary test is optimized with respect to the parameters m and τ such that the data samples are clustered according to a user-defined distance function ρ by maximizing a compactness criterion (Konukoglu et al., 2013).

Testing phase. Given a test set $\hat{\mathbf{I}} = \{\hat{I}_r\}_{r=1}^{R} \in \mathcal{I}$, a test image $\hat{I}_r$ is passed down each tree in the forest. At a node, the binary test with the parameters learned in the training phase is applied. According to this test, the image is sent to the left or to the right child of the current node. This is repeated until the image arrives at a leaf node. If the leaf node contains the training image I_p, their affinity $a_\rho^{(pr)}$ is increased by one. This procedure is repeated for each tree and yields an affinity matrix $\mathbf{A}_\rho = \{a_\rho^{(pr)}\}_{\substack{p=1,\ldots,P \\ r=1\ldots,R}}$ between the samples of the training and testing set.

Feature Selection. During the training phase of NAFs, the parameters m and τ of each binary test of tree t are optimized to obtain an optimal partitioning of the training data samples. The parameter m denotes the component of the feature vector $\mathbf{f}^{(t)}$ which is tested at the current node. There exist several ways of determining the importance of individual features for the growing of the decision trees. Konukoglu et al. (2013) considered a feature important if it is selected in the first three levels of the trees. A more sophisticated approach was used by Gray et al. (2013), where the decrease in the Gini impurity criterion (Breiman, 2001) was measured for the individual features in each node. In this work, we adopt the former and simpler approach. The frequency of the selected features in the first three levels of the trees is recorded, and the values are normalized by the number of nodes in the tree level.

Pairwise Similarities based on leaf co-occurrences. For each distance function ρ, a NAF F_ρ is trained using the training set $\mathbf{I}$ and a pairwise affinity matrix $\mathbf{A}_\rho \in \mathbb{R}^{P\times P} = \{a_\rho^{(ij)}\}_{i,j=1,\ldots,P}$ is computed, where $a_\rho^{(ij)}$ reports how often images $I_j \in \mathbf{I}$ and $I_i \in \mathbf{I}$ finish in the same node across each tree in the forest. The corresponding distance matrix $\mathbf{D}_\rho = \{d_\rho^{(ij)}\}_{i,j=1,\ldots,P}$ is constructed with $d_\rho^{(ij)} = 1 - a_\rho^{(ij)}/T$, where T is the number of trees in the forest. The matrix $\mathbf{D}_\rho$ can now be interpreted as pairwise distances of the image set $\mathbf{I}$ and can be used for constructing a manifold representation of the training set.

2.2. Manifold learning using multiple approximated neighborhoods

Combination of affinity matrices. To consider different features simultaneously, the affinity matrices from NAFs based on different user-defined distances ρ_k, $k = 1,\ldots,K$, can be combined, producing a joint embedding of various neighborhood approximations. Assuming that the NAFs F_{ρ_k} contain the same number of trees T, the affinity matrices can be additively combined by

$$\mathbf{A}_{\rho_1,\ldots,\rho_K} = \sum_{k=1}^{K} \mu_k \mathbf{A}_{\rho_k}. \tag{1}$$

The weights $\mu_k \in \mathbb{R}^+$, $\sum_{k=1}^{K} \mu_k = 1$ control the influence of each A_{ρ_k} in the combination (and will be computed following a weighting optimization scheme described later). The components of the joint distance matrix $\mathbf{A}_{\rho_1,\ldots,\rho_K}$ are $d_{\rho_1,\ldots,\rho_K}^{(ij)} = 1 - a_{\rho_1,\ldots,\rho_K}^{(ij)}/T$.

Since the affinity matrices are expressed in the same units (i.e., leaf counts), they can be readily combined without the need of additional normalization or scaling, as opposed to other similarity combination methods.

Manifold learning method. Laplacian Eigenmaps (LE) (Belkin and Niyogi, 2003) is used for learning the manifold. It captures the intrinsic low-dimensional structure of a manifold by finding an optimal embedding which preserves local neighborhoods. This can be posed as the minimization problem

$$\min_{\substack{\mathbf{Y} \\ \mathbf{Y}^T\mathbf{D}\mathbf{Y}=\mathbf{Id}}} \sum_{ij} \parallel y_i - y_j \parallel^2 w^{(ij)},$$

where $\mathbf{Y} = [y_1,\ldots,y_P] \in \mathbb{R}^{P\times P}$ are the embedding coordinates, $\mathbf{Id}$ the $P \times P$ identity matrix and $w^{(ij)}$ the similarity between input data points. By using the above objective function, neighboring points in the original space ($w^{(ij)}$ large) receive a penalty if they are mapped far apart in the embedding space ($y_i - y_j$ large). Thus, by minimizing the objective function, local neighborhoods are preserved. It can be shown that solving this minimization problem is equivalent to the generalized eigenvalue problem $\mathbf{L}\mathbf{v} = \lambda\mathbf{M}\mathbf{v}$ of the corresponding graph Laplacian $\mathbf{L} = \mathbf{M}-\mathbf{W}$ (with degree matrix $\mathbf{M}$ and weight matrix $\mathbf{W} = (w^{(ij)})_{i,j=1,\ldots,P}$), whose eigenvectors $\mathbf{v}_i$ give the optimal embedding. Typically, only the eigenvectors corresponding to the lowest (nonzero) n eigenvalues λ_i, $i = 1\ldots,n$ are kept for the final embedding. We estimate the best value of n by a goodness-of-fit parameter G (Mardia et al., 1980):

$$G = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{P} \lambda_i},$$

describing how well these n eigenvalues represent the full matrix of eigenvalues.

Here, the similarity coefficients $w^{(ij)}$ are functions of the distances based on leaf co-occurences between images I_i and I_j, i.e., functions of the elements $d^{(ij)}_{\rho_1,\ldots,\rho_K}$. We choose

$$w^{(i,j)}_{\rho_1,\ldots,\rho_K}(d^{(ij)}_{\rho_1,\ldots,\rho_K}) = \exp\left(- \frac{\left(d^{(ij)}_{\rho_1,\ldots,\rho_K}\right)^2}{2\sigma^2_{\rho_1,\ldots,\rho_K}} \right),$$

where $\sigma \in \mathbb{R}^+$ plays the role of a scale parameter and is estimated as $\sigma = \frac{1}{P^2}\sum_{i,j=1}^{P}\sqrt{d^{(ij)}_{\rho_1,\ldots,\rho_K}}$.

We denote the new embedding coordinates as $\mathbf{Y}_{\rho_1,\ldots,\rho_K}$ using affinity matrices obtained from NAFs trained on the user-defined distance functions ρ_k, $k = 1,\ldots,K$.

Weight optimization. The weights μ_k for the affinity combination in Eq. (1) can be chosen in different ways. For $\mu_k = \frac{1}{K}$, $k = 1,\ldots,K$, each approximated neighborhood has the same influence (uniform weighting). Alternatively, we can choose the weights μ_k such that the resulting joint embedding has certain properties. In this work, we propose two alternative different criteria: (i) minimizing the intra-class variance and (ii) preserving the neighborhoods. For the former, the variance of the new coordinates of the training set within each class is minimized (supervised) and for the latter a measure of quality assessing the new neighborhoods in the embedding is optimized (unsupervised). In the literature, there exist several measures that assess the quality of embeddings by comparing the k nearest neighbors of samples in the original space and the embedding space (Chen, 2006; Lee and Verleysen, 2009). We adopt a measure called Neighborhood Preservation Measure (NPM) (Zimmer et al., 2015b), which quantifies neighborhood intrusions in the new embedding space. Let the matrix $\mathbf{D}$ be the $P \times P$ matrix of pairwise distances between data samples in the original space, $\mathbf{Y}$ the $P \times d'$ matrix of data samples in the embedding space with new dimension $d' < d$ and P the number of data samples. Note that we use here the approximated pairwise distances using NAFs, whereas in (Zimmer et al., 2015b) Euclidean distances were calculated. Let $\mathcal{S}^{\mathbf{D}}_k$ be the set of indices of the k-nearest neighbors of sample i in the original neighborhood. To measure neighborhood intrusions, we search for the hypersphere in the new representation space that contains all the neighboring points in $\mathcal{S}^{\mathbf{D}}_k$. The radius of this hypersphere is calculated as $r^{\mathbf{D},\mathbf{Y}}_k(i) = \max_{j \in \mathcal{S}^{\mathbf{D}}_k} \parallel \mathbf{y}_i - \mathbf{y}_j \parallel^2$. Subsequently, the set of indices of all points that are located inside this hypersphere are derived as $\mathcal{S}^{\mathbf{D},\mathbf{Y}}_k(i) = \{p_j | \parallel \mathbf{y}_i - \mathbf{y}_j \parallel^2 \leq r^{\mathbf{D},\mathbf{Y}}_k(i), i \neq j\}$. This set will evidently contain all original k-nearest neighbors, but depending on the quality of the embedding, points that were initially outside this neighborhood could erroneously be mapped inside it. NPM is defined by summing all point intrusions into the embedding as follows:

$$\mathrm{NPM}(\mathbf{D},\mathbf{Y},k) = 1 - \frac{\sum_{i=1}^{P} |\mathcal{S}^{\mathbf{D},\mathbf{Y}}_k(i)| - k}{P(P-1-k)}. \tag{2}$$

An ideal embedding preserving all neighborhoods without any local or global distortion would result in a NPM value of 1. The worst possible mapping, i.e., to a single point in the new space, would result in a value of 0. Therefore, we choose the weights μ_k by maximizing the NPM of the resulting embedding using the Nelder-Mead Simplex algorithm in Matlab.

Projection of new data points. To be able to relate unseen data points with the new manifold representation, the so-called out-of-sample extensions are applied to project new data points to the low-dimensional space. For non-linear techniques such as LE, there is no straightforward extension to out-of-sample data points and the projection has to be estimated. Here, we use an out-of-sample extension based on the Nyströms approximation (Bengio et al., 2004; van der Maaten et al., 2009). According to it, the eigenvectors of a large $p \times p$ matrix can be approximated by the eigendecomposition of a $q \times q$ submatrix with $q < p$. Therefore, we can approximate the coordinates in the new representation of an unseen data sample I_r as

$$\mathbf{y}_r = \frac{1}{\kappa} \sum_{i \in \mathcal{N}^{\kappa}_A(i)} \mathbf{y}_i a^{(r,i)}_{\rho}, \tag{3}$$

where $\mathbf{y}_i$ are the new embedding coordinates of data sample $I_i \in \mathbf{I}$, where $i \in \mathcal{N}_A^{\kappa}(r)$, the κ nearest neighbors of new sample $I_r \in \hat{\mathbf{I}}$ in training set $\mathbf{I}$ according to the neighborhood approximated by affinity matrix $\mathbf{A}_\rho$. The affinity $a_\rho^{(r,i)}$ is based on co-occurrences in leaf nodes of the sample I_r to the sample I_i (which was used for training the NAF F_ρ), when I_r is passed down each tree of the forest F_ρ.

3. Experiments and results

3.1. Dataset

We applied the proposed pipeline to explore the best combination of user-defined distances for a population of neonates that were scanned at the Hospital Clínic de Barcelona. The population consists of T2 weighted MR images, scanned on a 3 T system (TrioTrim, Siemens Healthcare) in axial orientation. The study protocol and recruitment and scanning procedures were approved by the Institutional Ethics Committee, and written informed consent was obtained from the parents of each child to participate in the research studies. Two conditions are present in this population: intrauterine growth restriction (IUGR) and isolated mild ventriculomegaly (VM). IUGR due to placental insufficiency, which includes lack of oxygen and nutrition, affects 5-10 % of all pregnancies and is associated with neurostructural and neurodevelopmental anomalies (Batalle et al., 2012; Esteban et al., 2010; Padilla et al., 2011). A newborn is diagnosed with IUGR if its birth weight percentile is less than ten. A percentile of 50 for a newborn means that out of 100 newborns, 50 are bigger and 50 are smaller. The second condition, isolated mild VM, occurs in 0.15-0.7 % of all pregnancies and is defined as an atrial diameter between 10 mm and 12 mm of the lateral ventricles at $18 - 22$ weeks gestational age (GA) (Lyall et al., 2012; Wyldes and Watkinson, 2004). VM can affect both lateral ventricles (bilateral) or only one ventricle (unilateral) and is associated with neurodevelopmental disorders (Lyall et al., 2012; Wyldes and Watkinson, 2004).

The dataset consists of 111 neonates with an age range of $38.63 - 48.40$ (mean of 43.70 ± 2.24) weeks GA at scan. The subjects were prenatally diagnosed either as healthy term born controls (NC, 70 subjects), affected by IUGR (27 subjects) or VM (14 subjects). The study protocol and the recruitment and scanning procedures were approved by the Institutional Ethics Committee, and written informed consent was obtained from the parents of each child to participate in the research studies (IUGR cohort: HCB/2012/7715; VM cohort: HCB/2014/0484).

Information for each group regarding GA, birth weight percentile and type of VM is provided in Table 1. The subjects of the three groups have a similar mean GA at scan. The birth weight percentiles differ between the groups. While the IUGR subjects have a percentile around three, the other two groups have larger percentiles as expected. From the VM subjects, the majority have a unilateral VM affecting the right ventricle.

Image noise was removed (Coupé et al., 2012). The images were skull-stripped using BET (Smith, 2002), corrected for bias using N4 (Tustison et al., 2010) and intensity normalized using histogram matching (Nyúl et al., 2000). A group template was created using ANTs (Avants et al., 2008) for the same population as used in the experiments. All subjects were non-rigidly aligned using a coarse control-point spacing of 20 mm to the group template to account for size differences. We performed automatic segmentation (Sanromà et al., 2016) into four regions (gray matter, white matter, deep gray matter and ventricular structure). The aligned images in atlas space were of size $166 \times 245 \times 62$ with a voxel size of $0.625 \times 0.625 \times 2$ mm. The images were smoothed using a Gaussian filter with physical size of 4 mm in each dimension. Figure 2 shows subjects for the three subpopulations and the group template.

3.2. Experimental setup

We selected distance functions ρ specific to the properties of the conditions (as described below) and applied the pipeline shown in Fig. 1 to the affinity matrices obtained by the NAFs as described in Section 2. For evaluation, we employed a stratified 10-fold cross-validation strategy, taking randomly 10% of the data as test data in each fold and used the rest for training the NAFs and construction of the low-dimensional embedding. In all folds, all three diagnostic groups are represented with a similar distribution as in the whole set for both training and testing sets. Each sample is exactly once used for testing and in all other

Table 1: For each group of the neonatal population (NC: normal controls; IUGR: affected by intrauterine growth restriction; VM: affected by isolated mild ventriculomegaly), information on the total number of subjects (N), weeks gestational age (GA) at birth and at scan, birth weight percentile and on the type of ventriculomegaly (bilateral or unilateral) are provided.

	N	GA at birth	GA at scan	birth weight percentile	type of VM		
					bi	left	right
NC	70	39.6 ± 1.3	44.0 ± 2.3	47.4 ± 29.0	—	—	—
IUGR	27	37.3 ± 2.1	42.7 ± 2.4	3.3 ± 3.9	—	—	—
VM	14	40.1 ± 0.8	43.9 ± 1.6	77.5 ± 28.4	5	2	7

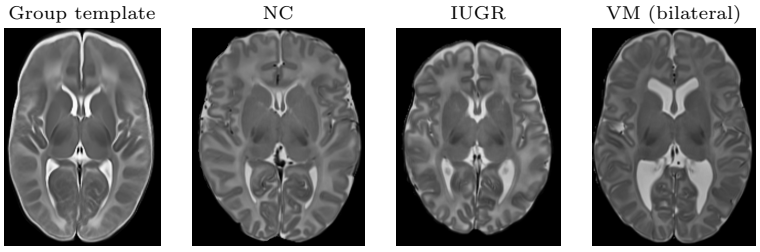

Figure 2: From left to right: group template, NC subject (normal controls), IUGR subject (affected by intrauterine growth restriction), VM subject (affected by isolated mild ventriculomegaly).

folds for training. The feature vector for each image was composed of the intensities of randomly chosen voxels inside the brain mask of the group template. We chose a feature vector of length $Q = 100{,}000$ (13.6% of all voxels in the brain mask). NAFs F_ρ were trained for three different definitions of the distance function ρ, namely, (i) the difference in diagnosis (-1: IUGR; 0: NC; 1: VM), F_D, (ii) the difference in the birth weight percentile, F_P and (iii) the difference in volume of the right lateral ventricle, F_V. Combinations of those are denoted as $F_{D,P}$, $F_{P,V}$ and $F_{D,P,V}$.

3.3. Evaluation strategy

The aim of this work was to find a low-dimensional representation of the data, which captures its heterogeneity and all relevant information. The evaluation of such representation is not straightforward. One could expect, however, that it preserves local neighborhoods and distances. This can be measured using Euclidean distances. Thus, we assessed the quality of the new data representations by how well unknown data samples are classified according to population-specific characteristics using its k nearest neighborhood (kNN) in the projected space. We performed classification into three diagnostic groups (IUGR/NC-VM, VM/NC-IUGR and NC/IUGR/VM) using a simple kNN classifier.

We compared our method to classical LE with distance $d^{(ij)}$ between images I_i and I_j being the Euclidean distance of the feature vector $\mathbf{f}$ of the images, to the random forest classifier (NAF, using directly the output of the NAFs as classification results) and to a logistic regression classifier applied to the original features (LR) and to the selected features for each NAF (LR-F).

We evaluate the classification results using three metrics: the balanced accuracy, the sensitivity and the specificity. The balanced accuracy takes into account the unbalanced nature of our dataset and is computed as the average accuracy across classes such that each class has the same contribution in the evaluation regardless of the number of samples

When the projection captures the heterogeneity of the data, it should have a high classification performance for low values of k. With increasing k, the accuracy should decrease slowly. On the other hand, a data representation of low quality should have a low classification accuracy irrespective of the value of k.

To illustrate the influence of k in the classification, Fig. 3, shows the accuracy of new data representations obtained using different methods (LE, NAFs F_D and F_P, and ML-NAFs $\mathbf{Y}_{D,P}^{\mathrm{NPM}}$ optimized using NPM) for

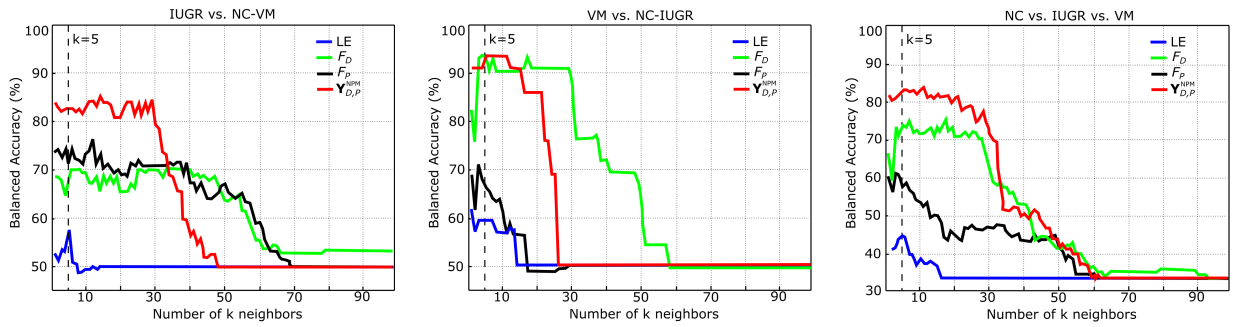

Figure 3: Influence of the number of nearest neighbors for the classification performance of LE (Laplacian Eigenmaps), F_D and F_P (random forest classifier NAF, trained on the diagnosis and birth weight percentile, respectively), and $\mathbf{Y}^{\mathrm{NPM}}_{D,P}$ (proposed method combining F_D and F_P using NPM).

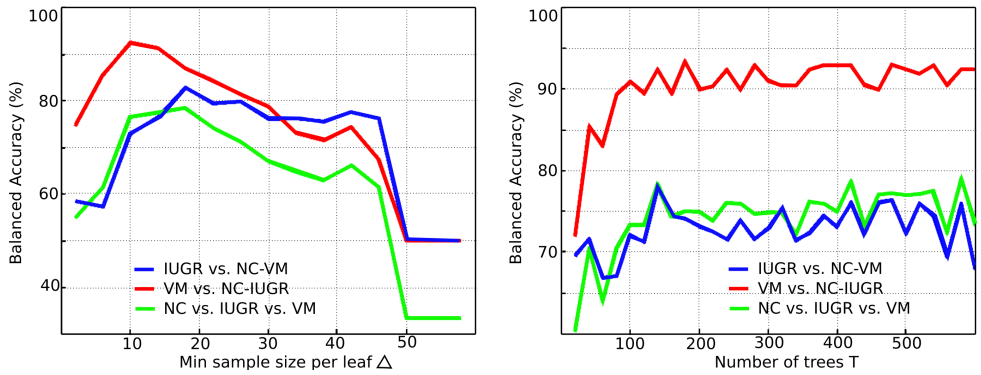

Figure 4: Influence of the parameters Δ (left) and T (right) on the classification performance.

those three classification tasks for varying k. LE yields low accuracies for all three tasks and all values of k and gives therefore a low quality data projection. The embedding $\mathbf{Y}^{\mathrm{NPM}}_{D,P}$ has the highest accuracies for all classification tasks for low values of k, which decrease with increasing k. For the classification of IUGR and VM, the decrease in accuracy starts for $k \approx 30$ and $k \approx 15$, respectively, which is related to the size of those subpopulations ($N = 27$ for IUGR and $N = 14$ for VM, see Table 1). For F_D and F_P, the accuracies are lower. As expected, the results of F_P are good for the classification task of IUGR and F_D performs reasonable for all classifications, because all diagnostic labels are used for training the forest. The method $\mathbf{Y}^{\mathrm{NPM}}_{D,P}$, combining those information in a low-dimensional space outperforms all other methods.

In the remaining experiments, we fixed the numbers of neighbors for the kNN classifier as $k = 5$.

3.4. Parameter selection for NAFs

The most influential parameters on the performance of NAFs are the number of trees for each forest, T, and the minimum sample size at each leaf, Δ, (Konukoglu et al., 2013). Low values of Δ yield high classification errors due to overfitting and high values to a decrease in the discriminating power. We performed similar experiments as in (Konukoglu et al., 2013) to analyze the influence of the parameters on our experiments. Figure 4 shows the influence of the parameters Delta and T on the classification performance. Each plot was obtained by varying the corresponding parameter while keeping all others constant. We observe that for $\Delta \in [10, 20]$ good results are obtained for all three diagnostic groups and that for $T > 200$, no significant increase in accuracy is observed. Based on this, for the following experiments we chose $\Delta = 10$ and $T = 500$. In each tree, $q = \mathrm{round}(\sqrt{Q}) = 316$ features were evaluated.

3.5. Pairwise affinities and feature extraction

The affinity matrices extracted from NAFs F_D, F_P and F_V are shown in Fig. 5 and illustrate the expected properties of the new approximated neighborhoods. For F_D, three blocks associated with the three classes

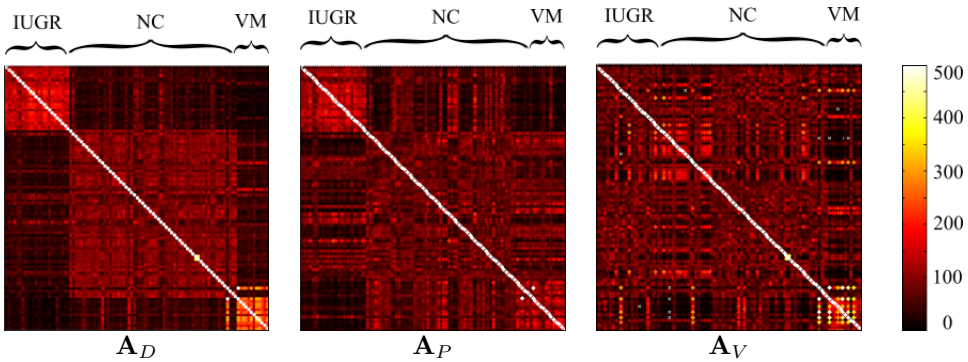

Figure 5: Affinity matrices extracted from the NAFs F_D (left), F_P (middle) and F_V (right) showing the similarity structure in the approximated neighborhoods.

are visible, showing that subjects in the approximated neighborhood are more similar to subjects of their own class than to subjects of other classes. For F_P and F_V, the block structure is degenerated. For F_P, only the block of IUGR subjects is clearly visible, indicating that the birth weight percentile is a discriminating factor for IUGR but not for VM or NC. When trained on the right lateral ventricle volume, the VM subjects show higher similarities to other VM subjects than to subjects belonging to other classes.

As described in Section 2.1, the most discriminative features of the NAFs F_D, F_P and F_V are automatically selected. The frequencies of the features selected in the first three levels of the trees, normalized with the total number of nodes in these levels, are shown in Fig. 6 after convolution with a Gaussian kernel with standard deviation $\sigma = 2$.

For F_V, the NAF trained on the right ventricle volume, the most discriminative features are selected around the right ventricle (right column in Fig. 6). The majority of the VM subjects have a dilated right lateral ventricle (12 out of 14 subjects, see Table 1) and the left ventricle is only affected in half of the cases. This explains that mainly features from the right ventricle are selected in the first levels of the trees, even though the lateral ventricles form normally a symmetric structure. For F_P (Fig. 6, middle column), the selected features are located mainly at the cortex, suggesting that changes in the cortex are connected with a low birth weight percentile. This is in line with other studies (Dubois et al., 2008; Egaña-Ugrinovic et al., 2013; Tolsa et al., 2004) that reported changes in the cortex for subjects affected by IUGR. For F_D, features of both locations (right ventricle and cortex) are selected. When training based on the diagnosis (Fig. 6, left column), many features are selected around the ventricles. This indicates that the dilation of the ventricles is a distinguishing structural difference between the two conditions and the controls (see Fig. 2). The image-based classification of IUGR is difficult, because the structural changes in the brain are more subtle than those in VM subjects. However, as we can see in Fig. 6, left column, some features from the cortex are also selected.

3.6. Data representation

Figure 7 shows examples of new data representations using the first and second embedding coordinates with different color coding. The columns show different data representations (from left to right: $\mathbf{Y_f}$, $\mathbf{Y}_D$, $\mathbf{Y}_P$ and $\mathbf{Y}_{D,P}$), and the rows different color codings (from top to bottom: diagnostic labels, birth weight percentile, right lateral ventricle volume). It can be observed that the representation $\mathbf{Y_f}$ (left column, obtained by applying LE using the Euclidean distance between the original feature vectors $\mathbf{f}$) is not able to cluster the samples according to the diagnostic labels (in particular IUGR subjects) or the birth weight percentile (first and second row). It shows a separation according to the right lateral ventricle volume, which is related to more drastic structural changes than, e.g., the birth weight percentile, and can therefore be captured by the simple Euclidean distance. With $\mathbf{Y}_D$, $\mathbf{Y}_P$ and $\mathbf{Y}_{D,P}$, the three diagnostic groups are well clustered. However, when combining NAFs trained on multiple distances as in $\mathbf{Y}_{D,P}$, the embedding better reflects the different characteristics of the population, which is confirmed by the correlation values (r-scores) for the birth weight percentile and the volume of the right lateral ventricle. Note that when using image

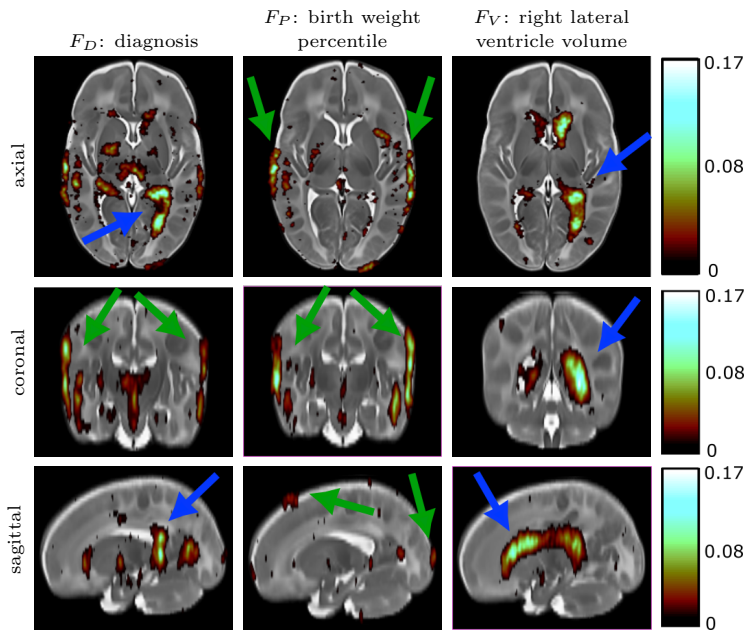

Figure 6: The features frequency of the NAFs F_D, F_P, F_V, selected in the first three levels of the trees. For F_P (middle column) and F_V (right column), the most discriminative features are located at the cortex (green arrows) and around the right lateral ventricle (blue arrows), respectively. For F_D (left column), features of both locations are selected.

neighborhoods based on the diagnostic labels, similar data samples are mapped very close to each other in the embedding space (see $\mathbf{Y}_D$ in Fig. 7, top row). When multiple image neighborhoods are combined, the data samples are wider spread in the embedding space.

3.7. Classification results using single NAFs

Table 2 shows the classification results (balanced accuracy, sensitivity and specificity) of the three diagnostic groups using the different methods without combination. When applying LE, the sensitivity for classifying IUGR and VM subjects is poor. However, when using logistic regression on the original features, the classification metrics increase. This shows that the classical Euclidean distance cannot capture the differences between the features of the three diagnostic groups.

When the NAFs are directly used as a non-linear classifier, the classification accuracy of the three diagnostic groups improves notably. The forests F_D and F_V yield balanced accuracies of 92% and 87%, respectively, to classify VM subjects, which indicates that the size of the right ventricle volume is a discriminative factor in the population. For the classification of IUGR subjects, where the structural changes are more subtle, the balanced accuracy is 72% for F_D and 71% for F_P. For both diagnostic groups, the forests obtaining the highest classification accuracies were trained on distance functions which are associated with the respective diseases (diagnosis and birth weight percentile for IUGR and diagnosis and right lateral ventricle volume for VM). When trained on less suitable distance functions (birth weight percentile for VM and right lateral ventricle volume for IUGR), the performance in classification decreases. This is also seen in the multi-class classification problem for the third diagnostic group. The highest accuracy is obtained with F_D, where information about both diseases is encoded in the forest. The balanced accuracy decreases with F_P and F_V.

The most discriminative features selected in each forest reflect the relation between the appearance of the images in the population and the distance function used for training the NAFs. When only those features are used for classification, the performance is comparable or increases in almost all cases compared to the classification results directly obtained by the NAFs, as shown in Table 2. This highlights the strong relation between the regions from where the features were selected and the respective diseases.

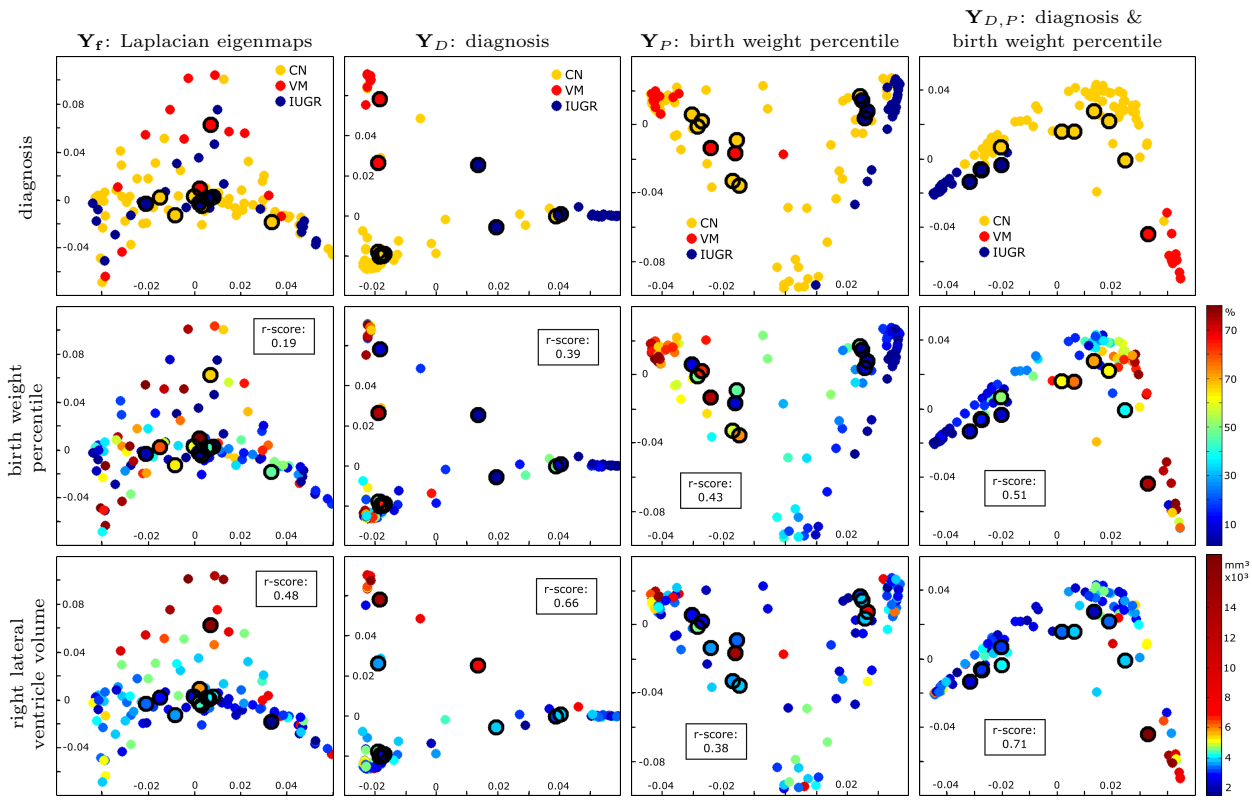

Figure 7: First and second coordinates of the new representation when applying the proposed pipeline to single and combined NAFs ($\mathbf{Y}_D$: trained on the diagnosis; $\mathbf{Y}_P$: trained on the birth weight percentile; $\mathbf{Y}_{D,P}$, $\mathbf{Y}_{P,V}$: combination when optimizing NPM of $\mathbf{Y}_D$ and $\mathbf{Y}_P$; $\mathbf{Y_f}$: using classical Laplacian eigenmaps. The color coding is with respect to the diagnosis (top row; controls in yellow, IUGR in blue and VM in red). The samples with a black circle are the test samples projected into the new space, with the color of their true class labels.

When applying the proposed method (Fig. 1) to the affinity matrices extracted from F_D, F_P, F_V and performing classification in the new representations $\mathbf{Y}_D$, $\mathbf{Y}_P$, $\mathbf{Y}_V$, the overall best results with balanced accuracy, sensitivity and specificity for single NAFs are obtained in the new space $\mathbf{Y}_D$, where already information of both IUGR and VM was encoded (highlighted in bold in Table 2).

3.8. Classification results using combined NAFs

The affinity matrices extracted from the NAFs approximate image neighborhoods resulting from the corresponding user-defined distance function. To incorporate several neighborhood definitions, the matrices can be combined as detailed in Section 2.2 to form a new manifold representation. The classification results when using several combination methods are shown in Table 3.

It can be observed that the classification performance for all three diagnostic groups increases when combining image neighborhoods based on different distance functions. For the proposed method, we used three different weighted combination methods, as described in Section 2.2. The overall best performance for all groups is obtained when optimizing the weights μ_k in Eq. (1) according to the embedding assessment measure NPM (highlighted in bold in Table 3).

4. Discussion and Conclusions

We have proposed a method to learn and combine pairwise image similarities induced by application specific distance functions for finding a new, low-dimensional data representation through manifold learning.

Table 2: Classification results of the three diagnostic groups (Bacc: balanced accuracy; Sens: sensitivity; Spec: specificity) using different methods without combination. LR: logistic regression classifier on the original features; LE: classical Laplacian Eigenmaps using the Euclidean distance between original features $\mathbf{f}$; NAF: neighborhood approximation forest classifier of forests F_D, F_P, F_V; LR-F: logistic regression classifier on the most discriminative features $\mathbf{f}_D^*$, $\mathbf{f}_P^*$, $\mathbf{f}_V^*$, selected from F_D, F_P, F_V, respectively; ML-NAF: proposed method (Fig. 1) applied to the affinity matrices $\mathbf{A}_D$, $\mathbf{A}_P$, $\mathbf{A}_V$ extracted from F_D, F_P, F_V to obtain new representations $\mathbf{Y}_D$, $\mathbf{Y}_P$, $\mathbf{Y}_V$.

Method		IUGR			VM			NC-IUGR-VM		
		Bacc.	Sens.	Spec.	Bacc.	Sens.	Spec.	Bacc.	Sens.	Spec.
		in %			in %			in %		
LR	$\mathbf{f}$	75	52	98	66	35	98	61	77	81
LE	$\mathbf{Y_f}$	58	18	98	59	2	99	45	68	72
NAF	F_D	72	57	88	92	90	95	77	79	85
	F_P	71	58	84	66	35	98	57	69	78
	F_V	57	18	96	87	80	94	63	73	78
LR-F	$\mathbf{f}_D^*$	74	55	93	81	65	98	68	77	83
	$\mathbf{f}_P^*$	75	55	94	76	55	98	61	76	80
	$\mathbf{f}_V^*$	69	43	94	87	75	99	68	77	82
ML-NAF	$\mathbf{Y}_D$	**78**	**72**	**85**	**93**	**90**	**96**	**80**	**79**	**88**
	$\mathbf{Y}_P$	77	82	73	61	25	97	57	63	77
	$\mathbf{Y}_V$	47	0	94	87	80	94	56	64	73

Table 3: Classification results of the three diagnostic groups (Bacc: balanced accuracy; Sens: sensitivity; Spec: specificity) using different methods with combination. LR-F: logistic regression classifier on the concatenation of the most discriminative features $\mathbf{f}_{D,P}^*$, $\mathbf{f}_{P,V}^*$, $\mathbf{f}_{D,P,V}^*$, selected from F_D, F_P, F_V, respectively; ML-NAF: proposed method (Fig. 1) applied to a combination of affinity matrices $\mathbf{A}_D$, $\mathbf{A}_P$, $\mathbf{A}_V$ extracted from F_D, F_P, F_V to obtain new representations $\mathbf{Y}_{D,P}$, $\mathbf{Y}_{P,V}$, $\mathbf{Y}_{D,P,V}$.

Method		Combination	IUGR			VM			NC-IUGR-VM		
			Bacc.	Sens.	Spec.	Bacc.	Sens.	Spec.	Bacc.	Sens.	Spec.
			in %			in %			in %		
NAF	$F_{D,P}$	uniform	79	72	87	91	85	97	79	79	87
	$F_{P,V}$		67	50	84	87	80	95	70	73	81
	$F_{D,P,V}$		76	65	88	89	85	94	77	78	86
LR-F	$\mathbf{f}_{D,P}^*$	concat.	75	55	95	81	65	98	67	77	82
	$\mathbf{f}_{P,V}^*$		72	48	96	87	75	99	69	79	83
	$\mathbf{f}_{D,P,V}^*$		75	55	95	81	65	98	67	77	82
ML-NAF	$\mathbf{Y}_{D,P}$	uniform	82	80	84	**93**	**90**	**97**	82	80	88
	$\mathbf{Y}_{P,V}$		72	62	81	87	80	94	71	70	82
	$\mathbf{Y}_{D,P,V}$		78	72	85	92	90	94	80	78	88
	$\mathbf{Y}_{D,P}$	intra-class variance	81	77	85	**93**	**90**	**97**	82	80	88
	$\mathbf{Y}_{P,V}$		77	82	73	77	60	94	68	67	83
	$\mathbf{Y}_{D,P,V}$		82	77	87	92	90	95	82	81	90
	$\mathbf{Y}_{D,P}$	NPM	**85**	**90**	**80**	**93**	**90**	**97**	**84**	**79**	**89**
	$\mathbf{Y}_{P,V}$		78	78	77	87	80	95	75	72	85
	$\mathbf{Y}_{D,P,V}$		75	65	85	92	90	94	77	77	86

We used NAFs to approximate image neighborhoods, which are learned on user-defined distances and based on the appearance of images. Thus, the proposed method allows the combination of heterogeneous sources of information (i.e., user-defined distances relevant to a pathology) and extracts automatically relevant features without the need to engineer them a priori (e.g., shape-based features). However, the neurobiological meaning of those predictive features remains unclear and has to be investigated further (Shmueli, 2010). We proposed to combine those multiple neighborhoods in an unsupervised manner through manifold learning by assessing the quality of the resulting embedding through NPM. We applied the method to a population of neonatal brain MR images and evaluated it using classification in the resulting manifold representation space. Results showed an improved classification performance compared to using single affinities and classical unsupervised manifold learning regarding structural changes related to neonatal diseases.

A key motivation for using random forests, here in particular NAFs, is that they provide a natural way for combining the similarities learned from multiple distances. It may not be possible to capture the complex brain changes and characteristics, which come along with diseases in the developing brain, by single, pre-selected features. Subsequently, an important part of our framework is the combination of approximated neighborhoods. We have shown that an optimal combination can improve the performance of the new representation with respect to classification of structural and clinical information.

The proposed framework is flexible in the sense that the different methods, in particular for learning and combining the neighborhoods and constructing the manifold representation, are interchangeable. We chose to learn the pairwise affinities using NAFs but other implementations of random forests can be used as well. As a manifold learning method, we employed LE, but other non-linear methods are possible. In our previous work (Zimmer et al., 2015a), we applied Isomap to the approximated neighborhoods obtained from the NAFs, because it had the best performance on the given dataset. In Section 2.2, we presented two methods to optimize the weights in Eq. (1). First, a supervised method which optimizes the intra-class variance in the new manifold representation such that the classes are well separated in the optimal low-dimensional space. Second, as an alternative, we proposed an unsupervised method measuring the quality of the resulting manifold representation in terms of intrusions in the k nearest neighborhood of samples in the new space (Zimmer et al., 2015b). Other methods to find the optimal combination, see, e.g., the works of Gray et al. (2013) and Wolz et al. (2012) or to combine the different distances after the manifold learning step (Aljabar et al., 2011) are possible.

The evaluation of such manifold embeddings is not straightforward. Since one could expect that a good low-dimensional representation capturing the data's heterogeneity preserves local neighborhoods and distances, we evaluate our method through the classification of the conditions present in the data population. Classification methods for brain disorders have been studied extensively and although classification was not the main objective of this work, we use it to assess the quality of the low-dimensional data representations. One major difference is the type of features they use to analyze the data. Many works use hand-crafted features, such as shape information (Tang et al., 2014) and deformations (Baloch and Davatzikos, 2009) to classify brain disorders, others combine simple features such as image intensities with other information, e.g., other measurements such as volume, thickness and shape (Aljabar et al., 2011; de Vos et al., 2016), or meta-information, such as age, gender and cognitive scores (Wolz et al., 2012; Moradi et al., 2015). Our method goes in line with the latter approaches and we avoid the computation of complex features. With this, we follow an increasing trend in computer vision to substitute hand-crafted features (through non-straightforward procedures, such as extraction of shape descriptors) by features learned using data-driven approaches (representation learning) (LeCun et al., 2015). The idea of those approaches is to make methods and algorithms less dependent on labor intensive feature engineering (Bengio et al., 2013). However, the proposed framework allows an easy incorporation of hand-crafted features. One option would be to define distance functions for training the NAFs based on the difference of those features, e.g., as done in Aljabar et al. (2011) based on shape features. A second option would be to train the NAFs on more sophisticated features. We performed experiments incorporating information about pairwise image deformation by learning the approximated neighborhoods on the deformation Jacobians instead of the image intensities. However, using those supposedly more informative features did not significantly improve the results.

We have applied our framework to neonatal MR images of three different groups (term controls and subjects affected by one out of two conditions: IUGR and VM). We learned the pairwise image distances

using clinical (diagnosis and birth weight percentile) and structural information (lateral ventricle volume) to construct manifold representations which are able to separate between the three groups. The manifold representations constructed from individual neighborhoods were specific to the criterion their neighborhoods were trained on. The combination of pairwise distances through associated affinity matrices obtained from the NAFs improves the overall performance of the joint representation regarding the full characteristics of the population.

Diseases affecting the brain structure during brain development are highly complex and might affect not only single, well identified brain regions, but influence overall brain development. By learning image distances through NAFs based on, e.g., clinical information, we couple the neighborhood approximation with the feature extraction step and do not have to rely on feature extraction methods which require prior knowledge. Interesting to highlight are the approximated neighborhoods and most discriminative features trained on the diagnosis or the birth weight percentile to detect IUGR (see Fig. 6). The structural changes which go in hand with IUGR are very subtle and hard to identify with MR imaging. However, by learning the distances based on clinical information, we are able to detect the differences in the cortex. As future work, it would be of great interest to relate the new embedding coordinates with the neurodevelopmental outcome of patients affected by VM and IUGR. This would go one step further towards its prediction and biomarker identification.

A limitation of this study is the relatively small population size of 111 subjects. The lack of neonatal data due to challenges during acquisition and data accessibility, is a well known issue. Nonetheless, the sampling of the manifold by the original data is an important aspect for the quality and performance of the new representation obtained by manifold learning (Aljabar et al., 2012). We had only 14 VM subjects, and therefore we cannot assume that they represent the whole variability of the disease. The same holds for the 27 IUGR subjects. For further clinical studies on the specific diseases, including biomarker extraction, a larger amount of data is required.

Under the hypothesis that the space of brain images can be represented by a low-dimensional non-linear manifold, many manifold learning techniques have been proposed to discriminate features over the brain. However, it remains an open question whether one can consider brains affected by different pathologies to lie on or close to the same manifold. Typically, only a single pathology is studied, together with normal controls. We hypothesize that neonatal brain images affected by different conditions, as IUGR and VM, have manifold representations approximately in the same space. Another approach would be to construct new representations for each disease and compare them.

A related issue is the projection of new data samples to an already constructed manifold representation. In non-linear manifold learning, determining the manifold coordinates for new subjects is not straightforward. In this work, we chose a projection based on Nyströms formula (Bengio et al., 2004) as detailed in Section 2.2. The new coordinates are a weighted linear combination of the coordinates of the nearest training samples in the original space. By using such an approximation, the assumption is made that the mapping from the high- to the low-dimensional space is an isometry. Although LE does not preserve distances on the manifold (Belkin and Niyogi, 2003), it does preserve locality. Here, we use only the k=10 nearest neighbors (in the original space determined by the affinity matrices derived from the NAFs) to obtain the embedding coordinates of an unseen data sample. Other approaches are possible (He and Niyogi, 2004).

To conclude, we have presented a framework encoding simultaneously heterogeneous information (clinical and image-based) in new manifold representations and have shown on a population of neonatal brain MR images that this approach improves the overall characterization of the population and corresponding diseases. We believe that incorporating different kind of information in the analysis may help in studying abnormal brain development, which is characterized by changes in multiple biomarkers.

Acknowledgements. The authors would like to thank E. Konukoglu for providing the implementation of the NAFs (publicly available at `http://www.nmr.mgh.harvard.edu/~enderk/software.html`).

V. A. Zimmer is supported by the grant FI-DGR 2013 (2013 FI B00159) from the Generalitat de Catalunya. This research was partially funded by the Spanish Ministry of Economy and Competitiveness (TIN2012-35874). This study was also supported by Instituto de Salud Carlos III (PI16/00861), integrated in the Plan Nacional de I+D+I and co-financed by ISCIII-Subdirección General de Evaluaciión and Fondo

Europeo de Desarrollo Regional (FEDER) “Una manera de hacer Europa”; additionally, the research leading to these results has received funding from “la Caixa” Foundation.

References

Aljabar, P., Rueckert, D., Crum, W.R., 2008. Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. NeuroImage 43, 225–235.

Aljabar, P., Wolz, R., Rueckert, D., 2012. Manifold learning for medical image registration, segmentation, and classification, in: Suzuki, K. (Ed.), Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis. IGI Global, Hershey, PA, USA, pp. 351–372.

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S.J., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2011. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. IEEE Transactions on Medical Imaging 30, 2072–2086.

Avants, B., Epstein, C., Grossman, M., Gee, J., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12, 26–41.

Baloch, S., Davatzikos, C., 2009. Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis. NeuroImage 45, S73 – S85. Mathematics in Brain Imaging.

Batalle, D., Eixarch, E., Figueras, F., Muñoz-Moreno, E., Bargalló, N., Illa, M., Acosta-Rojas, R., Amat-Roldan, I., Gratacós, E., 2012. Altered small-world topology of structural brain networks in infants with intrauterine growth restriction and its association with later neurodevelopmental outcome. NeuroImage 60, 1352 – 1366.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15, 1373–1396.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1798–1828.

Bengio, Y., Paiement, J.F., Vincent, P., Delalleau, O., Le Roux, N., Ouimet, M., 2004. Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. Advances in Neural Information Processing Systems 16, 177–184.

Bhatia, K.K., Rao, A., Price, A.N., Wolz, R., Hajnal, J.V., Rueckert, D., 2014. Hierarchical manifold learning for regional image analysis. IEEE Transactions on Medical Imaging 33, 444–461.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Chen, L., 2006. Local multidimensional scaling for nonlinear dimension reduction, graph layout and proximity analysis. Ph.D. thesis. University of Pennsylvania.

Coupé, P., Manjon, J.V., Robles, M., Collins, D.L., 2012. Adaptive multiresolution non-local means filter for three-dimensional magnetic resonance image denoising. IET Image Processing 6, 558–568.

Cox, T.F., Cox, M.A.A., 2001. Multidimensional scaling. Chapman & Hall/CRC.

Dubois, J., Benders, M., Borradori-Tolsa, C., Cachia, A., Lazeyras, F., Ha-Vinh Leuchter, R., Sizonenko, S.V., Warfield, S.K., Mangin, J.F., Hüppi, P.S., 2008. Primary cortical folding in the human newborn: an early marker of later functional development. Brain 131, 2028–2041.

Egaña-Ugrinovic, G., Sanz-Cortés, M., Figueras, F., Bargalló, N., Gratacós, E., 2013. Differences in cortical development assessed by fetal MRI in late-onset intrauterine growth restriction. American Journal of Obstetrics and Gynecology 209, 126–e1.

Esteban, F.J., Padilla, N., Sanz-Cortés, M., de Miras, J.R., Bargalló, N., Villoslada, P., Gratacós, E., 2010. Fractal-dimension analysis detects cerebral changes in preterm infants with and without intrauterine growth restriction. NeuroImage 53, 1225 – 1232.

Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R., The Alzheimer’s Disease Neuroimaging Initiative, 2010. Manifold modeling for brain population analysis. Medical Image Analysis 14, 643–653.

Gilmore, J.H., Smith, L.C., Wolfe, H.M., Hertzberg, B.S., Smith, J.K., Chescheir, N.C., Evans, D.D., Kang, C., Hamer, R.M., Lin, W., Gerig, G., 2008. Prenatal mild ventriculomegaly predicts abnormal development of the neonatal brain. Biological Psychiatry 64, 1069–1076.

Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. Journal of Machine Learning Research 12, 2211–2268.

Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., The Alzheimer’s Disease Neuroimaging Initiative, 2013. Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease. NeuroImage 65, 167–175.

He, X., Niyogi, P., 2004. Locality preserving projections, in: Neural information processing systems, MIT. pp. 153–160.

Konukoglu, E., Glocker, B., Zikic, D., Criminisi, A., 2013. Neighbourhood approximation using randomized forests. Medical Image Analysis 17, 790 – 804.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lee, J.A., Verleysen, M., 2009. Quality assessment of dimensionality reduction: Rank-based criteria. Neurocomputing 72, 1431–1443.

Lyall, A.E., Woolson, S., Wolfe, H.M., Goldman, B.D., Reznick, J.S., Hamer, R.M., Lin, W., Styner, M., Gerig, G., Gilmore, J.H., 2012. Prenatal isolated mild ventriculomegaly is associated with persistent ventricle enlargement at ages 1 and 2. Early Human Development 88, 691–698.

van der Maaten, L.J.P., Postma, E.O., van den Herik, H.J., 2009. Dimensionality reduction: A comparative review. Technical Report. Tilburg, Netherlands: Tilburg Centre for Creative Computing, Tilburg University, Technical Report: 2009-005.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1980. Multivariate analysis (probability and mathematical statistics). Academic Press London.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015. Machine learning framework for early mri-based alzheimer's conversion prediction in {MCI} subjects. NeuroImage 104, 398 – 412.

Nowak, E., Jurie, F., 2007. Learning visual similarity measures for comparing never seen objects, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.

Nyúl, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging 19, 143–150.

Padilla, N., Falcón, C., Sanz-Cortés, M., Figueras, F., Bargalló, N., Crispi, F., Eixarch, E., Arranz, A., Botet, F., Gratacós, E., 2011. Differential effects of intrauterine growth restriction on brain structure and development in preterm infants: A magnetic resonance imaging study. Brain Research 1382, 98 – 108.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6 2, 559–572.

Sanchez-Martinez, S., Duchateau, N., Erdei, T., Fraser, A.G., Bijnens, B.H., Piella, G., 2017. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. Medical Image Analysis 35, 70–82.

Sanromà, G., Benkarim, O.M., Piella, G., Ballester, M.Á.G., 2016. Building an Ensemble of Complementary Segmentation Methods by Exploiting Probabilistic Estimates. Springer International Publishing. pp. 27–35.

Schölkopf, B., Smola, A., Müller, K.R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10, 1299 – 1319.

Shi, T., Horvath, S., 2006. Unsupervised learning with random forest predictors. Journal of Computational and Graphical Statistics 15, 118–138.

Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A., Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Modern Pathology 18, 547–557.

Shmueli, G., 2010. To explain or to predict? Statistical Science 25, 289–310.

Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17, 143–155.

Tang, X., Holland, D., Dale, A.M., Younes, L., Miller, M.I., for the Alzheimer's Disease Neuroimaging Initiative, 2014. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and alzheimer's disease: Detecting, quantifying, and predicting. Human Brain Mapping 35, 3701–3725.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323.

Tolsa, C.B., Zimine, S., Warfield, S.K., Freschi, M., Rossignol, A.S., Lazeyras, F., Hanquinet, S., Pfizenmaier, M., Hüppi, P.S., 2004. Early alteration of structural and functional brain development in premature infants born with intrauterine growth restriction. Pediatric Research 56, 132–138.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging 29, 1310–1320.

de Vos, F., Schouten, T.M., Hafkemeijer, A., Dopper, E.G.P., van Swieten, J.C., de Rooij, M., van der Grond, J., Rombouts, S.A.R.B., 2016. Combining multiple anatomical mri measures improves alzheimer's disease classification. Human Brain Mapping 37, 1920–1929.

Wolz, R., Aljabar, P., Hajnal, J.V., Lötjönen, J., Rueckert, D., The Alzheimer's Disease Neuroimaging Initiative, 2012. Nonlinear dimensionality reduction combining MR imaging with non-imaging information. Medical Image Analysis 16, 819–830.

Wyldes, M., Watkinson, M., 2004. Isolated mild fetal ventriculomegaly. Archives of Disease in Childhood - Fetal and Neonatal Edition 89, F9–F13.

Ye, D.H., Desjardins, B., Hamm, J., Litt, H., Pohl, K.M., 2014. Regional manifold learning for disease classification. IEEE Transactions on Medical Imaging 33, 1236–1247.

Zimmer, V.A., Glocker, B., Aljabar, P., Counsell, S.J., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., González Ballester, M.A., Rueckert, D., Piella, G., 2015a. Learning and combining image similarities for neonatal brain population studies, in: Zhou, L., Wang, L., Wang, Q., Shi, Y. (Eds.), Machine Learning in Medical Imaging. Springer International Publishing. volume 9352 of *LNCS*, pp. 110–117.

Zimmer, V.A., Lekadir, K., Hoogendoorn, C., Frangi, A.F., Piella, G., 2015b. A framework for optimal kernel-based manifold embedding of medical image data. Computerized Medical Imaging and Graphics 41, 93 – 107.