We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 7

Computer-Aided Diagnosis in Neuroimaging

Francisco J. Martínez-Murcia,
Juan Manuel Górriz and Javier Ramírez

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/64980

Abstract

This chapter is intended to provide an overview to the most used methods for computer-aided diagnosis in neuroimaging and its application to neurodegenerative diseases. The fundamental preprocessing steps, and how they are applied to different image modalities, will be thoroughly presented. We introduce a number of widely used neuroimaging analysis algorithms, together with a wide overview on the recent advances in brain imaging processing. Finally, we provide a general conclusion on the state of the art in brain imaging processing and possible future developments.

Keywords: neuroimaging, VBM, feature extraction, CT, MRI, PET, SPECT, machine learning, classification

1. Introduction

Neuroimaging has meant a major breakthrough for the diagnosis and treatment of neurodegenerative diseases. Not so long ago, biomedical signal processing was limited to filtering, modelling or spectral analysis, prior to visual inspection. In the past decades, a number of powerful mathematical and statistical tools have been developed and evolved together with an increasing development and use of neuroimaging. Structural modalities such as computed tomography (CT) or the widely known magnetic resonance imaging (MRI), and later functional imaging techniques such as positron emission tomography (PET) and single photon emission computed tomography (SPECT) provide unprecedented insight in the internals of the brain, allowing the study of the structural and functional changes that can be linked to neurodegenerative diseases. This means a huge amount of data where automatic tools can help to identify patterns, reduce noise and enhance our knowledge of the brain functioning.

INTECH
open science | open minds

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

Computer-aided diagnosis (CAD) systems in neuroimaging include a variety of methods that range from preprocessing of the images (just after acquisition) to advanced machine-learning algorithms to identify disease-related patterns. Algorithms used in the reconstruction of medical imaging, such as the tomographic reconstruction (TR) or the filtered back-projection (FBP) lay outside the scope of this chapter, focused on the application of CAD systems to neuroimaging.

This chapter starts with an exposition of the preprocessing methods used in different neuroimaging modalities, including registration, normalization and segmentation. We provide references on the algorithms behind well-known pieces of software such as statistical parametric mapping (SPM) [1], FreeSurfer [2] or the FMRIB Software Library (FSL) [3]. Later, the most used computer-aided diagnosis systems in psychiatry, psychology and neurology are described. These include SPM [1] and voxel-based morphometry (VBM) [4], voxels as features (VAFs) [5] and how the computation of regions of interest (ROIs) work in semiquantitative analysis. In the next section, new advances in neuroimaging analysis are presented, starting with the basis of machine learning and classification, including support vector machines (SVMs) [5–8], but also logistic regression [9, 10] or classifier ensembles [11, 12]. Given the characteristics of neuroimaging data, where we study large, possibly correlated, data, the extraction of higher-level features is essential. Therefore, in the last section, we provide an introduction to commonly used image decomposition algorithms such as principal component analysis (PCA) [8, 13–18] and independent component analysis (ICA) [19–22]. Finally, other recent feature extraction algorithms including spatial and statistical methods such as texture analysis [23–31], morphological tools [31–33] or artificial neural networks [34–40] are presented.

2. Preprocessing

Preprocessing of the neurological images is a fundamental step in CAD systems as it ensures that all the images, either structural or functional, are comparable. We consider a preprocessing step, an algorithm that, applied after the acquisition and reconstruction of the images–usually a machine-dependent procedure–is intended to produce directly comparable images that represent a certain magnitude. The number and type of procedures to follow in preprocessing differs from one modality to another, although normalization and smoothing are used throughout all of them (see **Figure 1**).

2.1. Spatial normalization or registration

The anatomy of every subject's brain is slightly different in shape and size. In order to compare images of different subjects, we need to eliminate these particularities and transform the images so that the subsequent group analysis or comparison can be performed. To do so, the individual images are mapped from their individual subject space (current anatomy) to a reference space, a common anatomical reference that allows the comparison. This procedure is known as *spatial normalization* or *registration*.

There are a number of algorithms used in image registration, but the procedure usually involves the computation of a series of parameters to map the source images to a template that works as a common anatomical reference (see Section 2.1.2 for an overview on registration algorithms). The most widely used template is the Montreal Neurological Institute (MNI) template.

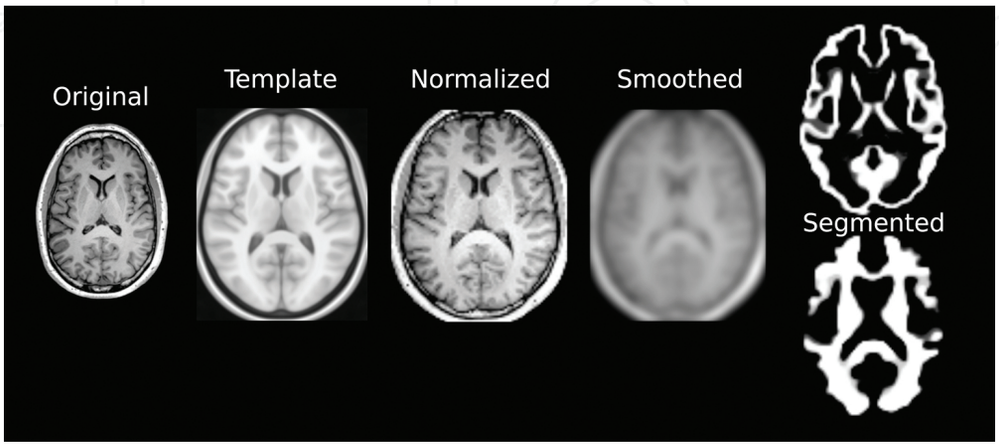


Figure 1. Example of the pipeline followed in structural MRI preprocessing, comprising spatial normalization, smoothing and segmentation.

2.1.1. The MNI space and template

The MNI space is the most widely used space for brain registration and was recently adopted by the International Consortium for Brain Mapping (ICBM) as its standard template. It defines a standard three-dimensional (3D) coordinate system (also known as 'atlas'), which is used to map the location of brain structures independently of the size and shape of each subject's brain.

The MNI space was intended to replace the Talairach space, a system based on a dissected and photographed brain for the Talairach and Tournoux atlas. In contrast to this, the MNI created a new template that was approximately matched to the Talairach brain but using a set of normal MRI scans. The current standard MNI template is the ICBM152 [41], which is the average of 152 normal MRI scans that have been matched to an older version of the MNI template using a nine-parameter affine transform.

2.1.2. Registration algorithms

Algorithms used in registration can be categorized in *linear* transformations (being the *affine* transform the most complex) and *non-rigid* or elastic transformations. Affine transformations are applied as a matrix multiplication and include terms for translation, scale, rotation, shear, squeeze and so on. These lineal transformations are applied globally to the image and therefore do not account for local geometric differences. Most neuroimaging software includes some kind of affine registration, including FreeSurfer [2], FSL (via package FLIRT) [3] or Elastix.

The estimation of the parameters is performed via the optimization of a given cost function, the minimum-squared difference between the source image and the template being the most basic. Modern software include more refined functions, for example, Tukey's biweight function (in mri_robust_template of FreeSurfer) [11], or the mutual information (in FLIRT) [42], that operate under a high-complexity schema involving local and global multiresolution optimization. When working with images of the same modality, the preferred cost function is the minimum-squared difference between the source image and the template, whereas in the case of multimodal registration, the maximization of the mutual information is preferred.

Non-rigid transformations can apply local transformation to align the target image to the template. Many of these non-rigid transformations are applied as a local fine-tuning after a previous affine transformation, although some of them use higher-complexity models that do not need this previous step. Some non-rigid transformations include radial basis functions (RBFs) (thin plate or surface splines, multiquadrics and compactly supported transformations), physical continuum models and large deformation models (diffeomorphisms). Of these, the most popular are diffeomorphic transformations, which feature the estimation and application of a warp field, and they are used in SPM (default) [1], FreeSurfer [2], FSL (FNIRT) [3], Elastix or ANTs.

2.1.3. Co-registration

Sometimes, we use different modalities, usually a functional and a structural image, for the same subject. It would be therefore very useful that these two (or even more) different images spatially match each other, so that any processing can be fit to any of them and applied to all. Since functional images have low resolution, the procedure for performing spatial normalization frequently involves co-registration to the structural image, which has a higher level of detail. Then the spatial normalization (or registration) parameters are estimated on the structural image and applied to all the co-registered maps. In the context of low-resolution functional imaging, affine co-registration to the structural image is preferred.

2.2. Smoothing

Despite the spatial normalization applied to the images, differences between subjects are still a problem that can reduce the signal-to-noise ratio (SNR) of the neuroimaging data. This problem increases as the number of subjects increases. To increase the SNR, it is recommended to filter out the highest frequencies, that is, applying a *smoothing*. Smoothing removes the smallest scale changes between voxels, making the detection of large-scale changes easier.

On the other hand, smoothing the images lowers its resolution. Smoothing is usually applied using a 3D Gaussian kernel to the image, determined by its full-width at half maximum (FWHM) parameter, which is the diameter of the smoothing kernel at half of its height. The choice of the size of the kernel is therefore of fundamental importance, and it depends on the signal to be detected. A small kernel will make our further processing confound noise with activation signal, but a larger kernel can parse out significant signal. As a general rule, the size of the kernel should be smaller than the activation to be detected.

http://dx.doi.org/10.5772/64980

2.3. Functional MRI-specific steps

The acquisition of functional MRI (fMRI) involves a more complex preprocessing of the images, given its dynamic features. fMRI studies acquire a long sequence of low-resolution MRI images that contain a magnitude known as blood-oxygen-level-dependent (BOLD) contrast. This sequence of three-dimensional volumes is combined into a four-dimensional (4D) volume that conceptually works similar to a video. In this context, the outcome can be much more affected by the subject motion. Therefore, procedures such as slice-timing correction and motion correction are mandatory in fMRI.

2.3.1. Slice-timing correction

In fMRI, the scanner acquires each slice within a single brain volume at different times. Different methods are used to acquire the slices: descending order (top to down), ascending order (bottom to up) and interleaved (the slices are acquired in a certain sequence). The time interval between one slice and another is usually small, but after acquiring a whole brain volume, there might be a difference of several seconds between the first and the last acquisition.

To compensate for the time differences between slice acquisition within a single volume, a *slice-timing correction* is performed by temporally interpolating the slices so that the resulting volume is approximately equivalent to acquiring the whole brain at the same time point. These time differences are especially important in event-driven experimental design, where timing is very relevant. Linear, quadratic or spline functions are used to interpolate the slices.

2.3.2. Motion correction

Motion correction, also known as *realignment*, accounts for the undesirable effect of head movement during the acquisition of the data. It is almost impossible for a subject to lie perfectly still during the acquisition of an fMRI sequence, and even the small movements can lead to variation in the BOLD values that, if uncorrected, can lead to activation of false positives.

The correction of this problem usually uses a rigid-body transformation similar to those used in registration. In this case, a model characterized by six parameters that account for translation and rotation is frequently used. The parameter estimation is performed by minimizing a cost function, such as correlation or mutual information, between the volumes and a reference time volume, usually the mean image of all time points.

Sometimes, the movement of the head is so fast that motion correction cannot correct its effects. In that case, the most used approach is to eliminate the images acquired during that fast movement, using an artefact detection algorithm that identifies large variations between images at adjacent time points.

2.4. Intensity normalization

Most functional neuroimaging modalities, in contrast to unitless structural MRI images, are the representation of the distribution of a certain contrast over the brain. There exist a larger number of sources of variability that can affect the final values: contrast uptake, radiotracer

decay time, metabolism, and so on. In order to establish comparisons between subjects, an *intensity normalization* procedure is required.

Intensity normalization methods are to be linear in nature, since it is essential to maintain the intensity ratio between brain regions, acting on the whole brain. In its simplest form, it consists of a division by a constant. This parameter is often estimated [6, 7] as the average value of the 95th bin of the histogram of the image, that is, the average of the 5% higher-intensity values, in what is known as the *normalization to the maximum*. Another approach, called *integral normalization*, estimates this parameter as the sum of all values of the image. A more complex approach requires *a priori* knowledge of the distribution of intensity in a normal subject. This is designed so that the whole image is divided by the binding potential (BP) [43], a specific-to-non-specific ratio between the intensities in areas where the tracer should be concentred and the non-specific areas.

Then, we have general linear transformations, defined as $Y = aX + b$. These procedures use estimates of the probability density function (PDF) of the source images and then estimate the parameters a and b, which transform their original PDF to an expected range. Methods to estimate the PDF parameters range from the simplest, non-parametric histogram [44] to more advanced estimates such as an analysis of covariance (ANCOVA) approach used in SPM [44, 45], or parametric estimates involving the Gaussian or the more general alpha-stable distribution [46], which has been recently tested with great success.

Structural modalities also suffer from some sources of intensity variability, for example, magnetic field inhomogeneity, noise, evolution of the scanners, and so on. Field inhomogeneity causes distortions in both geometry and intensity of the MR images [47], usually addressed via increasing the strength of the gradient magnetic field or preprocessing. Intensity variability is especially noticeable in multicentre-imaging studies, where images should share certain characteristics. To improve the homogeneity of a set of structural images acquired at different locations, the use of quantitative MRI images has been recently proposed [48]. In contrast to typical unitless $T1$-weighted images, quantitative imaging can provide neuroimaging biomarkers for myelination, water and iron levels that are absolute measures comparable across imaging sites and time points.

2.5. Segmentation

Segmentation, mostly of structural MRI images, involves a series of algorithms aimed at constructing maps of the distribution of different tissues. The general approach is to separate the image in three different maps containing grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF), although some software can also output data for bone, soft tissue or very detailed functional regions and subregions [49–51]. The procedure is applied after all aforementioned steps, including field inhomogeneity correction, which is essential for a correct segmentation. Here, we provide insight on the most used algorithms for segmentation

2.5.1. Statistical parametric mapping

In SPM, the segmentation procedure uses an expectation-maximization (EM) algorithm to obtain the parameters corresponding to a mixture of Gaussians that represents the tissue classes. Afterwards, an affine transformation is applied using tissue probability maps that are in the ICBM/MNI space. It currently takes normalized MRI images and extracts up to six maps: GM, WM, CSF, bone, soft tissue and air/background.

2.5.2. FMRIB software library

Two algorithms are used in FSL to perform segmentation: the FMRIB Automated Segmentation Tool (FAST) and the FMRIB Integrated Registration and Segmentation Tool (FIRST).

FAST is based on a hidden Markov random field model optimized through the EM algorithm. It firstly registers the brain volume to the MNI space and then segments the volume into three tissue classes (GM, WM and CSF). Skull-stripped versions of the anatomical image are needed.

On the other hand, FIRST is intended to extract subcortical structures of the brain characterized by parameters of surface meshes and point distribution models located in a database, built using manually segmented data. The source images are matched to the database, and the most probable structure is extracted based on the shape of the image.

2.5.3. FreeSurfer

FreeSurfer uses surfaces to perform posterior analysis, such as cortical thickness estimation. Therefore, the main aim of the command reckon_all, which performs most preprocessing steps, is not to obtain new image maps containing tissues, but surfaces that identify the different areas in the brain.

After registration to the MNI305 space, voxels are classified into white mater and other based on their location, intensity and local neighbourhood intensities. Afterwards, an estimation of the bias field is performed using these selected voxels, to correct the image. Next, the skull is stripped using a deformable template model [49]. The hemispheres are separated using cutting planes based on the expected MNI305 location of the corpus callosum and pons, and several algorithms that detect these shapes in the source images. The surfaces for each hemisphere are estimated first using the outside of the white matter mass, and then refined to follow the intensity gradients between WM and GM, and GM and CSF, called the pial surface, which will allow the estimation of cortical thickness [50].

FreeSurfer also implements a volume-based stream designed to obtain cortical and subcortical tissue volumes and label them. This stream performs a different, pathology-insensitive affine registration to the MNI305 space, followed by an initial volumetric labelling. Then, the intensity bias is corrected, and a new non-linear alignment to a MNI305 atlas is performed. The labelling process is a very complex one and is more thoroughly explained in Ref. [51].

3. Basic analyses

After performing a preprocessing of the source images, many procedures can be applied to extract the information required for clinical practice. In this section, we focus on those analyses that are more extended in clinical practice. These are currently preferred by medical staff, since they are easily interpretable and require little knowledge of computer science. Nevertheless, they are computer-aided systems that yield significant information to assist in the procedure of diagnosis. In later sections, we develop the application of more advanced systems that make use of machine learning to help in the same procedure.

3.1. Analysis of regions of interest (ROIs)

The analysis of *regions of interest (ROIs)* is the most basic computationally aided analysis of neuroimaging. It involves either purely manual or computer-assisted delimitation of regions in both structural and functional imaging and a posterior analysis of the delimited volumes.

A number of analyses can be performed on these regions, depending on the image modality. In the case of structural MRI, a frequent approach is the estimation of the volume of cortical and subcortical structures, called *morphometry*. The delimitation of ROIs is often performed automatically and then manually refined and allows the quantification of diseases that alter the normal distribution and size of GM and WM. This is the case of brain atrophy, a common issue in dementia, or brain tumours.

This ROI analysis is hardly used in fMRI, but very extended in nuclear imaging (PET or SPECT). Since the maps obtained with these techniques quantify the uptake of certain drugs, the total uptake can be obtained as the sum of intensities inside the drawn volume. However, the most used measure in these modalities is a ratio between the intensities in specific and non-specific areas, especially with drugs that bind to specific targets such as dopamine, amyloid plaques, and so on.

3.1.1. Cortical thickness

A specific case of a fully computer-assisted ROI analysis is the *cortical thickness* measure performed in FreeSurfer [50]. As aforementioned, once the GM-WM and GM-CSF surfaces have been estimated, the amount of GM in a direction perpendicular to the surface can be estimated. Combined with the subcortical segmentation algorithms, this allows to a very powerful estimation of the average cortical thickness by anatomical region. Thus, per region GM differences such as atrophy or hypertrophy can be characterized, making it perhaps the most widely used method in neuroscience.

3.2. Voxel-wise analyses

To overcome the time-consuming procedure of the traditional analysis of ROIs, several algorithms that act at the voxel level have been proposed. These include the statistical parametric mapping (SPM), a voxel-based morphometry (VBM) or the first machine-learning approach in this chapter, called voxel as features (VAF) (see **Figure 2**).

http://dx.doi.org/10.5772/64980

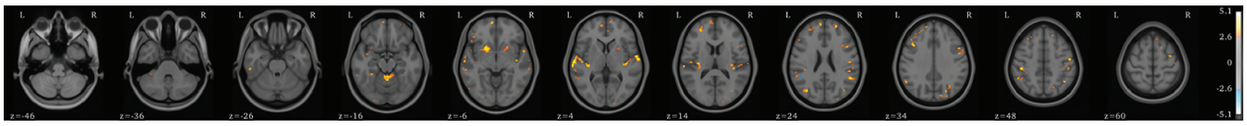


Figure 2. Example of a voxel-based morphometry output. Significant areas ($p < 0.05, |t| > 2.56$) in a comparison of autism-affected patients and healthy controls are highlighted.

3.2.1. Statistical parametric mapping (SPM)

Statistical parametric mapping (SPM) was proposed by Friston et al. [45] to automatically examine differences in brain activity during functional-imaging studies involving fMRI or PET. The technique can be applied to both single images (PET) and time series (fMRI), and the idea behind it is construct a general linear model (GLM) to describe the variability in the data in terms of experimental and confounding effects.

The level of activation is assessed at a voxel level using univariate statistics, and featuring a correction for type I errors (false positives), due to the problem of multiple comparisons. In the case of time series analysis, a linear convolution of the voxel signal with an estimation of the hemodynamic response is performed, and then the activation is tested against the analysed task.

The representation of the activation is frequently presented as an overlay of the Z-scores obtained for each voxel after the multiple comparisons correction on a structural image. The *Z*-score–or standard score–is the signed number of standard deviations an observation is above the mean. The resulting maps allow a visual inspection of the active brain areas, which can later be related to a certain disease or task.

3.2.2. Voxel-based morphometry (VBM)

Voxel-based morphometry (VBM) is the application of SPM to structural MRI images [4]. The principle behind VBM is the registration to a template, and then a smoothing of the structural image so that the smaller anatomical differences between subjects are reduced. Finally, a GLM is applied voxel-wise to all the images, in order to obtain a *Z*-score map that highlights the areas where the differences are greater.

As commented in Section 2.2, the size of the smoothing kernel is an important parameter. A small kernel will lead to artefacts in the Z-maps, including misalignment of brain structures, differences in folding patterns or misclassification of tissue types. On the other hand, a larger kernel will not be able to detect smaller regions.

Newer algorithms expand the idea behind VBM using multivariate approaches, to reveal different patterns. These algorithms include an independent component analysis (ICA) decomposition of the dataset and conversion to *Z*-scores, called source-based morphometry [52] and a multidimensional tensor-based morphometry [53].

3.2.3. Voxels as features (VAF)

Voxel-as-features (VAFs) is another voxel-wise approach proposed in Ref. [5] for the diagnosis of Alzheimer's disease using functional imaging. It can be considered the first machine-learning approach in this chapter, since it features a linear support vector machine (SVM) classifier whose input features are the intensities of all voxels in the images. It has been used in many works [6, 13, 25, 32, 33] as a baseline and an estimation of the performance achieved by expert physicians by means of visual analysis.

Additionally, some improvements can be done over the raw VAF, for example, using statistical hypothesis testing to obtain the most significant voxels, thus reducing the computational load and increasing the accuracy. The weight vector of the linear SVM can be inversely transformed to the dimension of the original images, and therefore provide a visual map that reflects the most influential voxels, in a similar way to the Z-maps of SPM and VBM.

4. Advances in brain image analysis

The application of new machine-learning techniques in CAD systems is a current trend. Works on this topic have increased exponentially in the past 10 years, and it is expected to grow even more. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data, and therefore it is very useful in neuroimaging.

Two approaches exist in machine learning. *Supervised learning* explores the patterns that lead to a certain outcome, for example, the brain activation patterns that are related to a certain disease. On the other hand, *unsupervised learning* explores the underlying structure of the data. Machine learning in CAD is mostly based on supervised learning, since it is focused on the prediction and analysis of patterns related to a certain disease.

In its simplest form, a machine-learning pipeline for neuroimaging consists of a single *classifier*, just as the VAF approach that we mentioned. However, classifiers can improve their detection power if higher-level features are extracted from the data, for example, features that represent the distribution of the voxel intensities, the texture of the images or the sources of variance of the maps. This is known as *feature extraction*, and the most common technique is *image decomposition*.

4.1. Classification in neuroimaging

We mentioned that the simplest approach to machine learning is classification. Classification basically is induction, that is, using a set of samples extracted from the real world from which we know their class (also known as label or category), and build a model that can identify the class of new samples that were never seen.

Statistical classification is a fertile field, and many classification algorithms are being developed at the moment. A wide range of strategies exist, among them are the following: *instance-based methods*, where the new samples are classified by a measure of similarity with known samples;

hyperplane-based methods, where a multidimensional function that weights the input features produces a score that identifies the class; *decision trees*, where the information is encoded in hierarchical nodes that test one feature at a time; *classifier ensembles*, *artificial neural networks* and many others.

Hyperplane-based methods are perhaps the most extended in neuroimaging and among them, by far, *support vector machines (SVM)* [6–8]. They offer high accuracy in high-dimensional spaces and can be expanded with the use of kernels to tackle non-linearly separable data. Furthermore, they are very robust to overfitting (unlike decision trees) and therefore more generalizable. Other frequent methods include the hyperplane-based *logistic regression* [9, 10]—and its multiclass version, the *Softmax classifier*–decision trees and *random forests* [54–56], and ensembles of neural networks [34] or ensembles of SVM [11, 12].

4.1.1. Support vector machines (SVM)

Support vector machines (SVM) have been widely used in neuroimaging [5–8] and other high-dimensional problems due to their overfitting robustness. In its linear version, training the classifier is equivalent to finding the coefficient vector $\vec{w}$ that defines the separation hyperplane:

$$(\vec{\boldsymbol{w}}\vec{\boldsymbol{x}}_i - b) = 0 \tag{1}$$

so that the distance between the hyperplane and the nearest point $\vec{x}_i$ is maximized. That way, the problem reduces to minimize the loss function (based on the Hinge loss):

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max[0, y_i(\vec{\boldsymbol{w}}\vec{\boldsymbol{x}}_i - b)]\right] + \lambda \parallel \vec{\boldsymbol{w}} \parallel^2 \tag{2}$$

where y_i is the class of the i-th training sample, b is the bias so that $b/\|\vec{w}\|$ is the offset of the hyperplane from the origin in the direction of $\vec{w}$, and λ is the regularization term, used to keep the coefficients in $\vec{w}$ as small as possible. Many methods exist to perform this optimization, for example, gradient descent or primal and dual loss (using quadratic programming).

Despite the fact that SVMs are mostly used in its linear form, an extension to non-linearly separable problems can be made using the kernel trick. By using kernels, we can implicitly transform data to a higher-dimensional space, but without needing to work directly in that transformed space, simply replacing all dot products in the SVM computation with the kernel. A kernel is a function $K: \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ so that $K(\vec{x}_i, \vec{x}_j) = \langle \phi(\vec{x}_i), \phi(\vec{x}_j) \rangle_M$, where $\langle \cdot, \cdot \rangle_M$ is the dot product in $\mathbb{R}^M$ and $\phi(\vec{x}_i)$ is a function that transform $\vec{x}_i$ to that higher-dimensional space. Some common kernel functions are the polynomial kernel $K(\vec{x}_i, \vec{x}_j) = (\gamma\langle \vec{x}_i, \vec{x}_j \rangle + r)^d$ or the radial basis function (RBF) $K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma\left(|\vec{x}_i - \vec{x}_j|^2\right)$.

4.1.2. Ensembles of classifiers

Ensembles of classifiers are a recent approach to machine learning in which, instead of choosing the best classification algorithm, we train and test different classifiers with different properties, and then combine their outputs to obtain the prediction. In the simplest technique, called *bagging*, we combined the results of the classifiers by voting. This is the strategy used in most neuroimaging works, for example in [11] where the brain was divided into subvolumes and a classifier was assigned to each one. In *boosting* weights are assigned to training samples, and then varied so that new classifiers focus on the samples where previous classifiers failed. Finally, in *stacking,* the outputs of individual classifiers are used as features in a new classifier that combines them. This was the approach used in *spatial component analysis (SCA)* [12], where an ensemble of *M* SVM was trained on *M* anatomically defined regions, and their outputs were combined to define a new decision function based on Bayesian networks.

4.2. Image decomposition

Signal decomposition algorithms are the first feature extraction algorithms that we will deal with. They are aimed at modelling a set of samples as a linear combination of latent variables. These latent variables or components can be thought of as the basis of an *n*-dimensional space where each sample is projected and represented as an *n*-dimensional vector. In a general case applied to our neuroimaging data, the source images **X** can be modelled as

$$\mathbf{X} = s_0 \vec{\boldsymbol{w}}_0 + s_1 \vec{\boldsymbol{w}}_1 + \cdots + s_N \vec{\boldsymbol{w}}_N + \epsilon = \vec{\boldsymbol{s}}\boldsymbol{W} + \epsilon \tag{3}$$

where s_i is the coordinate (or component score) of the current image in the *i*th dimension of the new space defined by all the base vectors $\vec{\boldsymbol{w}}_i$ (component loadings), and ϵ is the error of the estimation. **Figure 3** shows an illustration of the process.

Signal decomposition techniques are widely used in many applications, ranging from one-dimensional signals such as audio or electroencephalography (EEG) to multidimensional arrays, and are frequently applied as feature reduction to overcome the *small sample-size problem*, that is, the loss of statistical power due to a larger number of features compared to the number of samples.

4.2.1. Principal component analysis (PCA)

Principal component analysis (PCA) is a decomposition method where each component (here known as principal components) is intended to model a portion of the variance. The order of the components is set so that the greatest variance in the component space comes to lie on the first coordinate, the second greatest variance on the second coordinate, and so on. The set of principal component loadings, **W**, is the set of eigenvectors obtained when applying the eigenvalue decomposition of the covariance matrix of the signal, that is, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. It is not uncommon for neuroimaging specialist to refer to these eigenvectors as *eigenbrains*.

http://dx.doi.org/10.5772/64980

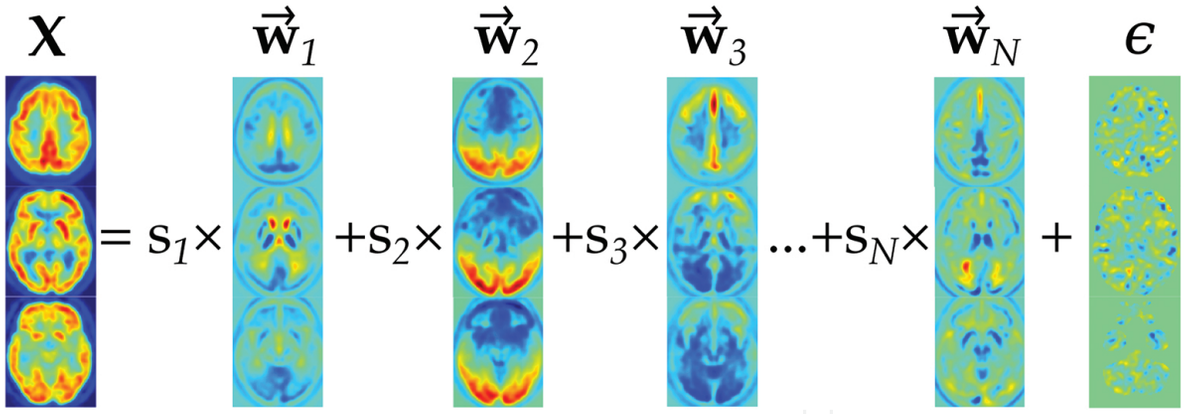


Figure 3. Illustration of a decomposition procedure of a PET-FDG brain image using PCA.

The most frequent approach to compute PCA on a given dataset is by using the singular value decomposition (SVD). This performs the decomposition of **X** in eigenvalues and eigenvectors in the form

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^{\mathrm{T}} \tag{4}$$

Here, **U** and **W** are square matrices, respectively, containing *n* and *p* orthogonal unit vectors of length *n* and *p*, known as left and right singular vectors of **X**; and **Σ**is an *n*-by-*p* rectangular diagonal matrix that contains the singular values of **X**. When compared to the eigenvector factorization of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, it is patent that the right singular vectors **X** of are equivalent to the eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, while **Σ**, the singular values of **X**, are equal to the square roots of the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, therefore, the set of principal component scores **S** of **X** in the principal component space can be performed as

$$\mathbf{S} = \mathbf{X}\mathbf{W} \tag{5}$$

As can be seen here, PCA does not account for the noise independently, but it integrates it in the model as another source of variance. The component-loading matrix can be truncated (i.e., only the first *m* components are used) to reduce the dimension of the principal component space, which is its most used application.

PCA has been used in many neuroimaging works, mainly used for feature reduction in a classification pipeline of nuclear imaging [13–15], but also in structural MRI [16, 17], functional MRI [18] or EEG signals [8].

4.2.2. Independent component analysis (ICA)

Independent component analysis (ICA) performs a decomposition of the source images, but, unlike PCA, it assumes that the components are non-Gaussian and statistically independent

from each other. Independence of the components is assessed via either the minimization of the mutual information or checking the non-Gaussianity of the components (motivated by the central limit theorem).

There exist a number of algorithms that implement ICA, from which the most extended are FastICA [57] and InfoMax [58], although other alternatives, such as CuBICA, JADE or TDSEP, are available. Most algorithms use an initial data preprocessing involving centring (create a zero mean signal by subtracting the mean), whitening and dimensionality reduction (with eigenvalue decomposition, SVD or PCA). Afterwards, the algorithm performs a decomposition of each of the m components by minimizing the cost function based on the independence criteria, usually the InfoMax, the Kullback-Leibler divergence or maximum entropy in the case of mutual information algorithms and kurtosis or negentropy for the non-Gaussianity-based algorithms. Most ICA algorithms require the number of sources m as an input.

The terms used in ICA vary from the ones used in PCA decomposition, although they roughly represent the same concepts. It is all based on a mixing model, where the original data **X**, known as mixed signal, is considered a mixture of several unmixed sources **S**, and therefore the matrix **W** that projects **X** to the ICA space **S** is known as the unmixing matrix.

A basic version of the FastICA algorithm is intended to find a unit vector $\vec{w}_p$ for the pth component so that the projection $\vec{w}_p^T\vec{x}$ maximizes non-Gaussianity, measured using an approximation of negentropy $J(\vec{w}_p^T\vec{x})$. To do so, first, we initialize the weight vector $\vec{w}_p$ to random. Then, using the derivative of the negentropy estimates, we perform the update:

$$\vec{w}_p^{+} = \mathrm{E}\{\vec{x}g(\vec{w}_p\vec{x})\} - \mathrm{E}\{g(\vec{w}_p\vec{x})\}\vec{w}_p \tag{6}$$

$$\vec{w}_p = \vec{w}_p^{+} / \| \vec{w}_p^{+} \| \tag{7}$$

This is performed for all components and iterated until convergence. The outputs $\vec{w}_1^T\vec{x}, \ldots, \vec{w}_n^T\vec{x}$ must be decorrelated after each iteration to prevent different vectors from converging to the same maxima. This is achieved, after pooling all $\vec{w}_1, \vec{w}_2, \ldots \vec{w}_n$ to the unmixing matrix **w** in a matricial form, using the alternative found in [57]

$$\mathbf{W} = \mathbf{W} / \sqrt{\| \mathbf{W}\mathbf{W}^T \|} \tag{8}$$

and repeating

$$\mathbf{W} = \frac{3}{2}\mathbf{W} - \frac{1}{2}\mathbf{W}\mathbf{W}^T\mathbf{W} \tag{9}$$

until convergence.

Many neuroimaging works have also applied ICA for feature extraction and reduction in classification pipelines [19, 20], but its main application today is in the processing of EEG signals [21, 22].

4.3. Other feature extraction techniques

In this section, we address feature extraction techniques other than the aforementioned decomposition algorithms. The philosophy behind them is still to provide higher-level features that allow a feature space reduction to overcome the small sample-size problem, but they are intended to extract and quantify information that otherwise would not be available.

4.3.1. Texture analysis

Texture analysis is any procedure intended to classify and quantify the spatial variation of voxel intensity throughout the image. In neuroimaging, they are more commonly used to classify images or to segment them (which can be also considered a form of classification). Depending on the number of variable studies, we can divide the methodology into first-, second- and higher-order statistical texture analysis. In *first-order statistics*, only voxel intensity values are considered, and values such as average, variance, kurtosis or frequency (histogram) of intensity values are computed.

Second-order statistical texture analysis is by far the most popular form. It is based on the probability of finding a pair of grey level at a certain distance and orientation on a certain image. Most algorithms are based in the grey level co-occurrence matrix (GLCM), in which is known as *Haralick texture analysis* [59]. The GLCM can be thought of as a representation of the frequency of the adjacent pixel variations in a certain spatial direction and distance. For a three-dimensional $n \times m \times k$ brain image $\mathbf{I}$ and two different grey levels i and j, at a given offset Δ the co-occurrence value is defined as

$$\mathbf{C}_{\Delta}(i,j) = \sum_{p=(1,1,1)}^{(n,m,k)} \begin{cases} 1, & \text{if } \mathbf{I}(p) = i, \mathbf{I}(p+\Delta) = j \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $p = (x, y, z)$ is the spatial position where $x = 1 \ldots n, y = 1, \ldots m, z = 1, \ldots k$, and $\Delta = (d_x, d_y, d_z)$ is the offset vector, which accounts for the direction and distance between voxels. The size of the GLCM matrix $\mathbf{C}_{\Delta}$ whose elements were defined before depends on the number of grey levels used in the image posterization, usually 8 or 16. Twelve Haralick texture features were proposed in [59] using a normalized co-occurrence matrix $\mathbf{P}_{\Delta} = \mathbf{C}_{\Delta} / \sum_{i,j} \mathbf{C}_{\Delta}(i,j)$ however, many more have been developed throughout the years. We provide formulas for energy $(\sum_{i,j} \{\mathbf{P}_{\Delta}(i,j)\}^2)$ entropy $(-\sum_{i,j} \mathbf{P}_{\Delta}(i,j) \log \mathbf{P}_{\Delta}(i,j))$ or contrast $(\sum_{i,j} |i-j|^2 \mathbf{P}_{\Delta}(i,j))$ among others. A 3D co-occurrence matrix was used for the analysis of MRI [23], segmentation [24] and analysis of SPECT images [25].

Higher-order analyses include the grey-level run length method (GLRLM) [26] that computes the number of grey-level runs of various run lengths. Each grey-level run is a set of consecutive and collinear voxels having the same grey-level value. Other texture-based feature extraction methods that have been applied in neuroimaging are the wavelet transform [27], the Fourier transform, used for segmentation [28] and characterization of MRI images [29], or local binary patterns (LBPs) [30, 31].

4.3.2. Spherical brain mapping (SBM)

The *spherical brain mapping (SBM)* is a framework that performs feature reduction of neuroimaging from three-dimensional volumes to two-dimensional maps representing a certain feature [32]. This is done by establishing a spherical coordinate system centred at the anterior commissure (AC) and defining a mapping vector $\vec{v}_{\theta,\varphi}$ at an elevation (θ) and azimuth (ϕ) angles. The sets of voxels $V_{\theta,\varphi}$ crossed by this vector are used to compute a value such as the average, variance or entropy, in each coordinate, and finally these values are used as pixels in a two-dimensional map representing that amount. This converts million-voxels volumes to several-thousand images, which achieves a significant computational load reduction as well. Maps for average, entropy or kurtosis of an MRI image are shown in **Figure 4**.

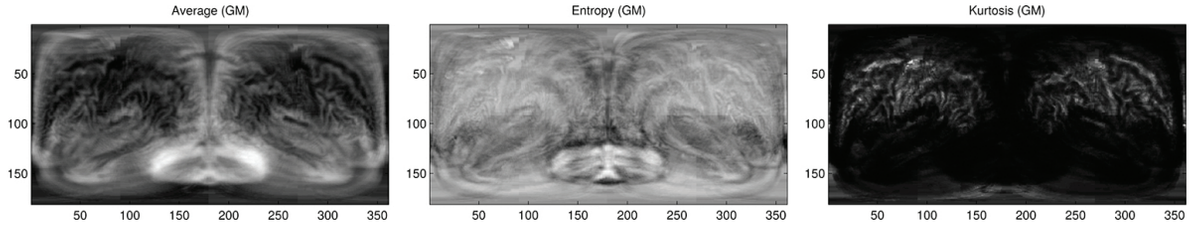


Figure 4. Example of the average, entropy and kurtosis maps generated by SBM.

An extension using volumetric LBP was recently proposed [31], and another utility that uses hidden Markov models to compute the path in the direction (θ,ϕ) using similar-intensity measures was later proposed [33].

4.3.3. Artificial neural networks (ANN)

Artificial neural networks (ANN) were inspired by the analysis of biological neural networks in the second half of the twentieth century, but then diverged from their original goal and now are a diverse area of research in machine learning. ANNs are composed by neurons, which model the behaviour of biological neurons by adding up the input signals $\boldsymbol{x}_i$, using different weightings w_i and producing an output by means of an activation function $f(x)$. Over the time, activation functions such as sigmoid or tan$h$ have been used, although nowadays the most common is rectified linear unit (ReLU), that computes the function $f(x) = \max(0, x)$ (see **Figure 5**).

The architecture of an ANN is a series of interconnected layers containing neurons. The perceptron, one of the most basic approaches, comprises an input layer (no neurons), a hidden

http://dx.doi.org/10.5772/64980

layer (composed by n neurons connected with all inputs and all neurons in the output layers) and one output layer. ANNs have been proven to be universal approximators. This means that given any continuous function $f(x)$, and some error $\epsilon > 0$, there exist an ANN $g(x)$ with at least one hidden layer such that $|f(x) - g(x)| < \epsilon$. The addition of more layers (a deeper network) allows the ANN to gain more abstract knowledge in the training process, especially in convolutional networks, where different types of layers coexist. In these networks, the deeper the ANN is, the more abstract the internal representation of the data will be.

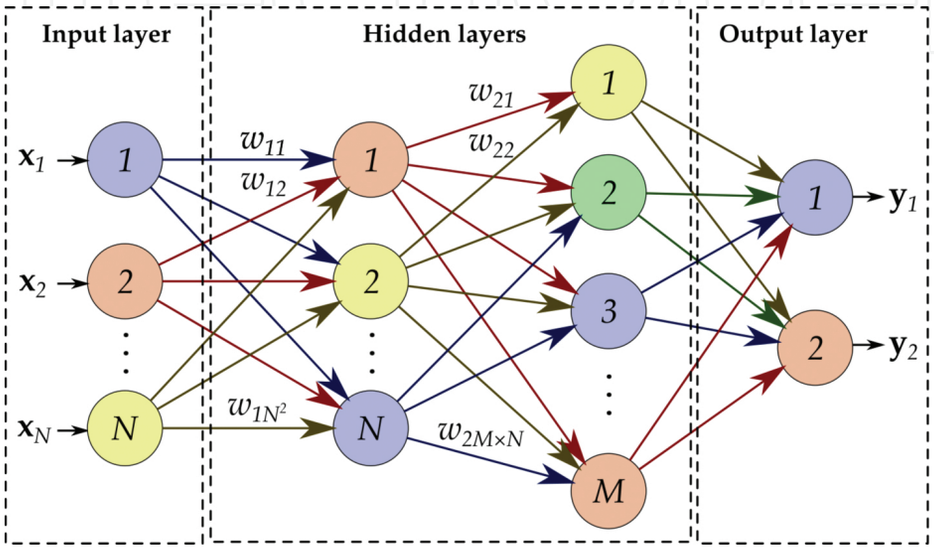


Figure 5. Illustration of the architecture of a perceptron with two hidden layers.

The field is extremely vast and there exist countless architectures, but the most common when applying them to neuroimaging are the *multilayer perceptron* [34] and *self-organizing maps* (SOMs) for segmentation [35] and pattern recognition in functional imaging [36], and more recently *convolutional networks*, which are the base of deep learning, used in segmentation [37], feature extraction [39] and classification [38, 40].

5. Conclusions

Computer-aided diagnosis systems are currently a thriving area of research. The bases are already established and contained in widely used software such as SPM or FreeSurfer. The neuroimaging community already uses these in their daily work, both at research and at clinical practice, with great benefit for the patients. These pieces of software usually include preprocessing (registration, intensity normalization, segmentation, etc.) and posterior automated procedures, such as ROI analysis, VBM or SPM, just like we saw in Sections 2 and 3.

In addition to this, state-of-the-art CAD systems involve the use of advanced techniques to characterize neuroimaging data. The field is still being developed and relevant breakthroughs are still to be made. Advances are being made in a daily basis, with the development of new

image modalities involving highly specific radiotracers, advanced registration, correction of inhomogeneities or application of existing machine learning and large data algorithms.

In this review, we have revealed a tendency towards fully automated tools capable of processing neuroimaging data, extract information and even predict the likelihood of having a specific condition. It is very likely that neuroimaging techniques will continue to increase its resolution and usage, and in this scenario the amount of data available will grow exponentially. CAD systems involving most of the topics that we covered in this chapter will be therefore crucial in clinical practice to provide understanding of all available information, otherwise intractable. Only this way can we address a major challenge: to discover meaningful patterns related to behaviour or diseases that ultimately help us to understand how the brain works.

Acknowledgements

This work was partially supported by the MINECO/FEDER under the TEC2015-64718-R and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the P11-TIC-7103 Excellence Project.

Author details

Francisco J. Martínez-Murcia, Juan Manuel Górriz* and Javier Ramírez

*Address all correspondence to: gorriz@ugr.es

Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain

References

[1] Penny W, Friston K, Ashburner J, Kiebel S, Nichols T, editors. Statistical Parametric Mapping: The Analysis of Functional Brain Images. 1st ed. London: Academic Press; 2006. 656 p.

[2] Reuter M, Rosas HD, Fischl B: Highly accurate inverse consistent registration: a robust approach. Neuroimage. 2010;53(4):1181–1196. DOI: 10.1016/j.neuroimage.2010.07.020.

[3] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy R, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004;23(S1):208–219. DOI: 10.1016/j.neuroimage.2004.07.051.

[4] Ashburner J, Friston KJ: Voxel-based morphometry—the methods. NeuroImage. 2000;11(6):805–821. DOI: 10.1006/nimg.2000.0582.

[5] Stoeckel J, Ayache N, Malandain G, Koulibaly PM, Ebmeier KP, Darcourt J. Automatic classification of SPECT images of Alzheimer's disease patients and control subjects. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004; September 26–29, 2004; Saint-Malo, France. Springer Berlin Heidelberg; 2004. p. 654–662. DOI: 10.1007/978-3-540-30136-3_80.

[6] Salas-Gonzalez D, Górriz JM, Ramírez J, López M, Illan IA, Segovia F, Puntonet CG, Gómez-Río M: Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines. Neuroscience Letters. 2009;461(1):60–64. DOI: 10.1016/j.neulet.2009.05.056.

[7] Martínez-Murcia FJ, Górriz JM, Ramírez J, Puntonet CG, Salas-González D: Computer aided diagnosis tool for Alzheimer's disease based on Mann–Whitney–Wilcoxon U-test. Expert Systems with Applications. 2012;39(10):9676–9685. DOI: 10.1016/j.eswa.2012.02.153.

[8] Baumgartner R, Ryner L, Richter W, Summers R, Jarmasz M, Somorjai R: Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. Magnetic Resonance Imaging. 2000;18(1):89–94. DOI: 10.1016/S0730-725X(99)00102-2.

[9] Ryali S, Supekar K, Abrams DA, Menon V: Sparse logistic regression for whole-brain classification of fMRI data. Neuroimage. 2010;51(2):752–764. DOI: 10.1016/j.neuroimage.2010.02.040.

[10] Dickerson BC, Goncharova I, Sullivan MP, Forchetti C, Wilson RS, Bennett DA, Beckett LA, de Toledo-Morrell L: MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiology of Aging. 2001;22(5):747–754. DOI: 10.1016/S0197-4580(01)00271-8.

[11] Gorriz JM, Ramirez J, Lassl A, Salas-Gonzalez D, Lang EW, Puntonet CG, Alvarez I, Lopez M, Gomez-Rio M: Automatic computer aided diagnosis tool using component-based SVM. In: 2008 IEEE Nuclear Science Symposium Conference Record; October 19–25, 2008; Dresden, Germany. IEEE; 2008. p. 4392–4395. DOI: 10.1109/NSSMIC.2008.4774255cc.

[12] Illán IA, Górriz JM, Ramírez J, Meyer-Baese A: Spatial component analysis of MRI data for Alzheimer's disease diagnosis: a Bayesian network approach. Frontiers in Computational Neuroscience. 2014;8:00156. DOI: 10.3389/fncom.2014.00156.

[13] López M, Ramírez J, Górriz JM, Álvarez I, Salas-Gonzalez D, Segovia F, Chaves R, Padilla P, Gómez-Río M: Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. Neurocomputing. 2011;74(8):1260–1271. DOI: 10.1016/j.neucom.2010.06.025.

[14] Hansen LK, Larsen J, Nielsen FÅ, Strother SC, Rostrup E, Savoy R, Lang N, Sidtis J, Svarer C, Paulson OB: Generalizable patterns in neuroimaging: how many principal components? Neuroimage. 1999;9(5):534–544. DOI: 10.1006/nimg.1998.0425.

[15] Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ: Functional connectivity: the principal-component analysis of large (PET) data sets. Journal of Cerebral Blood Flow and Metabolism. 1993;13(1):5–14. DOI: 10.1038/jcbfm.1993.4.

[16] Ung H, Brown JE, Johnson KA, Younger J, Hush J, Mackey S: Multivariate classification of structural MRI data detects chronic low back pain. Cerebral Cortex. 2012. DOI: 10.1093/cercor/bhs378.

[17] Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F: Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. Neurocomputing. 2015;151(1):139–150. DOI: 10.1016/j.neucom.2014.09.072.

[18] Subasi A, Gursoy MI: EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications. 2010;37(12):8659–8666. DOI: 10.1016/j.eswa.2010.06.065.

[19] Illán IA, Górriz JM, Ramírez J, Salas-Gonzalez D, López MM, Segovia F, Chaves R, Gómez-Rio M, Puntonet CG: 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. Information Sciences. 2011;181(4):903–916. DOI: 10.1016/j.ins.2010.10.027.

[20] Calhoun VD, Adali T, Pearlson GD, Pekar JJ: A method for making group inferences from functional MRI data using independent component analysis. Human Brain Mapping. 2001;14(3):140–151. DOI: 10.1002/hbm.1048.

[21] Delorme A, Makeig S: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004;131(1):9–21. DOI: 10.1016/j.jneumeth.2003.10.009.

[22] Mammone N, La Foresta F, Morabito FC: Automatic artifact rejection from multichannel scalp EEG by wavelet ICA. IEEE Sensors Journal. 2011;12(3):533–542. DOI: 10.1109/JSEN.2011.2115236.

[23] Kovalev VA, Kruggel F, Gertz HJ, von Cramon DY: Three-dimensional texture analysis of MRI brain datasets. IEEE Transactions on Medical Imaging. 2001;20(5):424–433. DOI: 10.1109/42.925295.

[24] Kruggel F, Paul JS, Gertz HJ: Texture-based segmentation of diffuse lesions of the brain's white matter. NeuroImage. 2008;39(3):987–996. DOI: 10.1016/j.neuroimage.2007.09.058.

[25] Martinez-Murcia FJ, Górriz JM, Ramírez J, Moreno-Caballero M, Gómez-Río M: Parametrization of textural patterns in I-123-ioflupane imaging for the automatic detection of Parkinsonism. Medical Physics. 2014;41(1):012502. DOI: 10.1118/1.4845115.

[26] Galloway MM: Texture analysis using gray level run lengths. Computer Graphics and Image Processing. 1975;4(2):172–179. DOI: doi:10.1016/S0146-664X(75)80008-6.

[27] El-Dahshan EA, Hosny T, Salem ABM: Hybrid intelligent techniques for MRI brain images classification. Digital Signal Processing. 2010;20(2):433–441. DOI: 10.1016/j.dsp.2009.07.002.

[28] Székely G, Kelemen A, Brechbühler C, Gerig G: Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. Medical Image Analysis. 1996;1(1):19–34. DOI: 10.1016/S1361-8415(01)80003-7.

[29] Wedeen VJ, Reese TG, Tuch DS, Weigel MR, Dou JG, Weiskoff RM, Chessler D. Mapping fiber orientation spectra in cerebral white matter with Fourier-transform diffusion MRI. In: Proceedings of the 8th Annual Meeting of ISMRM; Denver: 2000. p. 82.

[30] Unay D, Ekin A, Cetin M, Jasinschi R, Ercil A. Robustness of local binary patterns in brain MR image analysis. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 22–26, 2007; Lyon. IEEE; 2007. p. 2098–2101. DOI: 10.1109/IEMBS.2007.4352735.

[31] Martinez-Murcia FJ, Ortiz A, Górriz JM, Ramírez J, Illán IA. A volumetric radial LBP projection of MRI brain images for the diagnosis of Alzheimer's disease. In: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2015; June 1–5, 2015; Elche, Spain. Berlin: Springer International Publishing; 2015. p. 19–28. DOI: 10.1007/978-3-319-18914-7_3.

[32] Martínez-Murcia FJ, Górriz JM, Ramírez J, Ortiz A: A spherical brain mapping of MR images for the detection of Alzheimer's disease. Current Alzheimer Research. 2016;13(5):575–588. DOI: 10.2174/1567205013666160314145158.

[33] Martínez-Murcia FJ, Górriz JM, Ramírez J, Ortiz A: A structural parametrization of the brain using hidden Markov models-based paths in Alzheimer's disease. International Journal of Neural Systems. 2016;26(6):1650024. DOI: 10.1142/S0129065716500246.

[34] Liu J, Shang S, Zheng K, Wen JR: Multi-view ensemble learning for dementia diagnosis from neuroimaging: an artificial neural network approach. Neurocomputing. 2016;195:112–116. DOI: 10.1016/j.neucom.2015.09.119.

[35] Reddick WE, Glass JO, Cook EN, Elkin TD: Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. IEEE Transactions on Medical Imaging. 2002;16(6):911–918. DOI: 10.1109/42.650887.

[36] Wiggins JL, Peltier SJ, Ashinoff S, Weng SJ, Carrasco M, Welsh RC, Lord C, Monk CS: Using a self-organizing map algorithm to detect age-related changes in functional connectivity during rest in autism spectrum disorders. Brain Research. 2011;1380:187–197. DOI: 10.1016/j.brainres.2010.10.102.

[37] Moeskops P, Viergever MA, Mendrik AM, de Vries LS: Automatic segmentation of MR brain images with a convolutional neural network. IEEE Transactions on Medical Imaging. 2016;35(5):1252–1261. DOI: 10.1109/TMI.2016.2548501.

[38] Plis SM, Hjelm DR, Salakhutdinov R, Allen EA, Bockholt HJ, Long JD, Johnson HJ, Paulsen JS, Turner JA, Calhoun VD: Deep learning for neuroimaging: a validation study. Frontiers in Neuroscience. 2014;8(00229). DOI: 10.3389/fnins.2014.00229.

[39] Suka HI, Lee SW, Shen D: Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage. 2014;101:569–582. DOI: 10.1016/j.neuroimage.2014.06.077.

[40] Ortiz A, Martínez-Murcia FJ, García-Tarifa MJ, Lozano F, Górriz JM, Ramírez J. Automated diagnosis of Parkinsonian syndromes by deep sparse filtering-based features. In: Innovation in Medicine and Healthcare 2016; Springer International Publishing; 2016. p. 249–258. DOI: 10.1007/978-3-319-39687-3_24.

[41] Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B: A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2001;356(1412):1293–1322. DOI: 10.1098/rstb.2001.0915.

[42] Jenkinson M, Smith S: A global optimisation method for robust affine registration of brain images. Medical Image Analysis. 2001;5(2):143–156. DOI: 10.1016/S1361-8415(01)00036-6.

[43] Scherfler C, Seppi K, Donnemiller E, Goebel G, Brenneis C, Virgolini I, Wenning GK, Poewe W: Voxel-wise analysis of [123I]β-CIT SPECT differentiates the Parkinson variant of multiple system atrophy from idiopathic Parkinson's disease. Brain. 2005;128(7):1605–1612. DOI: 10.1093/brain/awh485.

[44] Arndt S, Cizadlo T, O'Leary D, Gold S. Andreasen NC: Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain. Neuroimage. 1996;3(3):175–184. DOI: 10.1006/nimg.1996.0019.

[45] Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD: Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping. 1994;2(4):189–210. DOI: 10.1002/hbm.460020402.

[46] Salas-Gonzalez D, Górriz JM, Ramírez J, Illán IA, Lang EW: Linear intensity normalization of FP-CIT SPECT brain images using the α-stable distribution. NeuroImage. 2013;65(15):449–455. DOI: 10.1016/j.neuroimage.2012.10.005.

[47] Chang H, Fitzpatrick JM: A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. IEEE Transactions on Medical Imaging. 2002;11(3): 319–329. DOI: 10.1109/42.158935S.

[48] Weiskopf N, Suckling J, Williams G, Correia MM, Inkster B, Tait R, Ooi C, Bullmore ET, Lutti A: Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. Frontiers in Neuroscience. 2013;7. DOI: 10.3389/fnins. 2013.00095.

[49] Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B: A hybrid approach to the skull stripping problem in MRI. Neuroimage. 2004;22(3):1060–1075. DOI: 10.1016/j.neuroimage.2004.03.032.

[50] Dale AM, Fischl B, Sereno MI: Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage. 1999;9(2):179–194. DOI: 10.1006/nimg. 1998.0395.

[51] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2022;33(3):341–355. DOI: 10.1016/ S0896-6273(02)00569-X.

[52] Xu L, Groth KM, Pearlson G, Schretlen DJ, Calhoun VD: Source-based morphometry: the use of independent component analysis to identify gray matter differences with application to schizophrenia. Human Brain Mapping. 2009;30(3):711–724. DOI: 10.1002/hbm.20540.

[53] Bossa M, Zacur E, Olmos S, ADNI: Tensor-based morphometry with stationary velocity field diffeomorphic registration: application to ADNI. Neuroimage. 2010;51(3):956–969. DOI: 10.1016/j.neuroimage.2010.02.061.

[54] Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solorzano C: Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Transactions on Medical Imaging. 2009;28(8):1266–1277. DOI: 10.1109/ TMI.2009.2014372.

[55] Salas-Gonzalez D, Górriz JM, Ramírez J, López M, álvarez I, Segovia F, Chaves R, Puntonet CG: Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees. Physics in Medicine and Biology. 2010;55(10):2807. DOI: 10.1088/0031-9155/55/10/002.

[56] Ramírez J, Górriz JM, Segovia F, Chaves R, Salas-Gonzalez D, López M, álvarez I, Padilla P: Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. Neuroscience Letters. 2010;472(2):99–103. DOI: 10.1016/j.neulet.2010.01.056.

[57] Hyvärinen A, Oja E: Independent component analysis: algorithms and applications. Neural Networks. 2000;13(4–5):411–430. DOI: 10.1016/S0893-6080(00)00026-5.

[58] Lee TW, Girolami M, Sejnowski TJ: Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation. 1999;11(2):417–441. DOI: 10.1162/089976699300016719.

[59] Haralick RM: Statistical and structural approaches to texture. Proceedings of the IEEE. 1979;67(5):786–804. DOI: 10.1109/PROC.1979.11328.

IntechOpen

IntechOpen