Advances in Analysis and Exploration in Medical Imaging

Nicolau Gonçalves

Advances in Analysis and Exploration in Medical Imaging

Nicolau Gonçalves

Doctoral dissertation for the degree of Doctor of Science in Techology to be presented with due permission of the School of Science for public examination and debate in Auditorium TU1 at the Aalto University School of Science (Espoo, Finland) on the 5th of December 2014 at 12 noon.

Aalto University
School of Science
Department of Information and Computer Science
Neuroinformatics Research Group

Supervising professor
Aalto Distinguished Prof. Erkki Oja

Thesis advisor
Doc. Ricardo Vigário

Preliminary examiners
Prof. Miika Nieminen, University of Oulu, Finland
Prof. Nikola Kasabov, Auckland University of Technology, New
Zealand

Opponent
Dr. Kristoffer H. Madsen, Danish Research Centre for Magnetic
Resonance, Denmark

Abstract

With an ever increasing life expectancy, we see a concomitant increase in diseases capable of disrupting normal cognitive processes. Their diagnoses are difficult, and occur usually after daily living activities have already been compromised. This dissertation proposes machine learning methods for the study of the neurological implications of brain lesions. It addresses the analysis and exploration of medical imaging data, with particular emphasis to (f)MRI. Two main research directions are proposed. In the first, a brain tissue segmentation approach is detailed. In the second, a document mining framework, applied to reports of neuroscientific studies, is described. Both directions are based on retrieving consistent information from multi-modal data.

A contribution in this dissertation is the application of a semi-supervised method, discriminative clustering, to identify different brain tissues and their partial volume information. The proposed method relies on variations of tissue distributions in multi-spectral MRI, and reduces the need for *a priori* information. This methodology was successfully applied to the study of multiple sclerosis and age related white matter diseases. It was also showed that early-stage changes of normal-appearing brain tissue can already predict decline in certain cognitive processes.

Another contribution in this dissertation is in neuroscience meta-research. One limitation in neuroimage processing relates to data availability. Through document mining of neuroscientific reports, using images as source of information, one can harvest research results dealing with brain lesions. The context of such results can be extracted from textual information, allowing for an intelligent categorisation of images. This dissertation proposes new principles, and a combination of several techniques to the study of published fMRI reports. These principles are based on a number of distance measures, to compare various brain activity sites. Application to studies of the default mode network validated the proposed approach.

The aforementioned methodologies rely on clustering approaches. When dealing with such strategies, most results depend on the choice of initialisation and parameter settings. By defining distance measures that search for clusters of consistent elements, one can estimate a degree of reliability for each data grouping. In this dissertation, it is shown that such principles can be applied to multiple runs of various clustering algorithms, allowing for a more robust estimation of data agglomeration.

Preface

Several years ago, when I decided to start my doctoral studies, I could never imagine the many challenges I would encounter. Most of those challenges stemmed from exciting new problems, that did not have obvious solutions. My kind of challenges. One thing is certain, all research I have done throughout the years has been remarkably interesting and engaging. Now, looking back at the years I spent working on this thesis, I feel that a new chapter of my life is about to begin.

Still on a work related note, I wish to thank my colleagues in the department for their friendship throughout the years. I want to acknowledge in particular: José Caldas, Francesco Corona, Jussi Gillberg, Mark van Heeswijk, Patrik Hoyer, Tiina Lundh-Knuutila, Oskar Kohonen, Yoan Miche, Mari-Sanna Paukkeri, Antti Sorjamaa and Tommi Suvitaival. I would be remiss not to thank also my co-authors: Timo Erkinjuntti, Hanna Jokinen-Salmela, Jari Lipsanen and Janne Nikkilä. I also want to thank the secretaries of the department, for always helping me solve the practical issues of a doctoral student life with a friendly smile.

On a more personal note, I want to extend my heartfelt thanks to all my friends in Finland, who have made me feel at home, in a country far from my own. I also want to express my sincere gratitude to my closest friends: Luís Gonçalves, Joana Godinho, Lidia Rychlinska and Miguel Silva. Friends like you are a rare gift.

Last, but not least, I wish to thank, from the bottom of my heart, all the support given by my family. They might not have understood the reasons why I chose Finland to do my doctoral degree, but that has not stopped them from helping in any way they could, especially my mother. My most grateful thank you, *o meu mais sincero obrigado*.

Espoo, October, 2014,

Nicolau Gonçalves

Contents

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Nicolau Gonçalves and Ricardo Vigário. Clustering through SOM Consistency. In *Lecture Notes on Computer Science - Image Analysis and Recognition*, A. Campilho and M. Kamel (Eds.): Proceedings of the 9th International Conference on Image Analysis and Recognition, ICIAR 2012, pages 61-68, DOI 10.1007/978-3-642-31295-3_8, June 2012.

II Nicolau Gonçalves, Janne Nikkilä and Ricardo Vigário. Partial Clustering for Tissue Segmentation in MRI. In *Lecture Notes in Advances in Neuro-Information Processing*, M. Köppen and N. Kasabov and G. Coghill (Eds.): Proceedings of the 15th International Conference on Neuro-Information Processing, ICONIP 2008, pages 559-566, DOI 10.1007/978-3-642-03040-6, November 2008.

III Nicolau Gonçalves, Janne Nikkilä and Ricardo Vigário. Self–supervised MRI Tissue Segmentation by Discriminative Clustering. *International Journal of Neural Systems*, Volume 24, Number 1, 16 pages, DOI 10.1142/S012906571450004X, January 2014.

IV Hanna Jokinen, Nicolau Gonçalves, Ricardo Vigário, Jari Lipsanen, Franz Fazekas, Reinhold Schmidt, Frederik Barkhof, Philip Scheltens, Sofia Madureira, José M. Ferro, Domenico Inzitari, Leonardo Pantoni, Timo Erkinjuntti and the LADIS Study Group. A novel multispectral MRI tissue segmentation approach reveals early stage white matter lesions and predicts cognitive decline. *Submitted to journal publication*, October 2014.

V Nicolau Gonçalves, Hanna Jokinen, Franz Fazekas, Reinhold Schmidt, Frederik Barkhof, Philip Scheltens, Sofia Madureira, Ana Verdelho, Domenico Inzitari, Leonardo Pantoni, Timo Erkinjuntti and Ricardo Vigário. Prediction of cerebral white matter lesion evolution through self-supervised tissue segmentation. *Submitted to journal publication*, October 2014.

VI Jayaprakash Rajasekharan, Ulrike Scharfenberger, Nicolau Gonçalves and Ricardo Vigário. Image Approach towards Document Mining in Neuroscientific Publications. In *Lecture Notes in Advances in Intelligent Data Analysis IX*, Paul R. Cohen and Niall M. Adams and Michael R. Berthold (Eds.): Proceedings of the 9th International Symposium, Intelligent Data Analysis 2010, pages 147-158, DOI 10.1007/978-3-642-13062-5_15, May 2010.

VII Nicolau Gonçalves, Gabriela Vranou and Ricardo Vigário. Towards automated image mining from reported medical images. In *Computational Vision and Medical Image Processing IV*, Proceedings of VipIMAGE 2013 - 4th ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, pages 255-261, DOI 10.1201/b15810-46, October 2013.

VIII Nicolau Gonçalves, Erkki Oja and Ricardo Vigário. Document mining combining image exploration and text characterization. In *Lecture Notes in Artificial Intelligence*, Sašo Džeroski, Panče Panov, Dragi Kocev and Ljupčo Todorovski (Eds.): Proceedings of the 17th International Conference on Discovery Science, Discovery Science 2014, pages 99-110, DOI 10.1007/978-3-319-11812-3_9, October 2014.

Author's Contribution

Publication I: "Clustering through SOM Consistency"

In Publication I the theory behind retrieving consistent information using multiple runs of Self Organizing Maps is explained. The author was responsible for all the experiments and most of the written text. The idea was developed and discussed between both authors.

Publication II: "Partial Clustering for Tissue Segmentation in MRI"

Publication II is a study on the application of discriminative clustering (DC) on simulated brain magnetic resonance imaging (MRI). The author was responsible for the method's implementation and conducting all the experiments. The article itself was written together with the last author, while J. Nikkilä provided support on the DC application to the task.

Publication III: "Self–supervised MRI Tissue Segmentation by Discriminative Clustering"

On Publication III, the theory behind the self-supervised tissue segmentation method is explained and the method is compared to other state-of-the-art algorithms. The idea behind the methodology was discussed among the first and last author. The fist author conducted the experiments and wrote most of the text, with the other co-authors helping in text revision.

Publication IV: "A novel multispectral MRI tissue segmentation approach reveals early stage white matter lesions and predicts cognitive decline"

Publication IV proves that the DC-based segmentation results correspond to real changes in cognitive scores of patients with neurodegenerative diseases. This study was proposed by N. Gonçalves and R. Vigário, and discussed with H. Jokinen and T. Erkinjuntti. The first two authors cooperated equally in the writing of this publication, with the first author being responsible for the statistical analysis of the results and the second author for the segmentation results. All other co-authors helped revising the publication text.

Publication V: "Prediction of cerebral white matter lesion evolution through self-supervised tissue segmentation"

Following the results of Publication IV, Publication V shows that it is possible to use the tissue segmentation method proposed by the author to predict neurodegenerative lesion evolution. The author was responsible for all the experiments and for writing most of the article. The other co-authors provided a substantial contribution to the publication text.

Publication VI: "Image Approach towards Document Mining in Neuroscientific Publications"

Publication VI proposed the idea of using images from neuroscientific articles as a means to extract information regarding the corresponding studies. The author was responsible for running the SOM algorithmic experiments and part of the text.

Publication VII: "Towards automated image mining from reported medical images"

It was shown in Publication VII that with a clear topic in mind, and using the original idea of Publication VI, it was possible to retrieve clear differences between studies. The author developed the algorithms and ran many of the experiments. The article was mostly written by the author.

Publication VIII: "Document mining combining image exploration and text characterization"

Publication VIII expands on Publications VI and VII, introducing text analysis as means to categorize the different clusters obtained using image analysis. The author was responsible for developing the algorithms and running the experiments. The article was written by the author, with cooperation with the co-authors.

List of Acronyms

TE	Time of Echo
TR	Time of Repetition
idf	Inverse Document Frequency
pdf	Probability Distribution Function
tf	Term Frequency
$tf\text{-}idf$	Term Frequency-Inverse Document Frequency
AIC	Akaike Information Criteria
ANN	Artifical Neural Network
BMU	Best Matching Unit
BOLD	Blood Oxigenation Level Dependent
BSS	Blind Source Separation
CBIR	Content-Based Information Retrieval
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
cSVD	Cerebral Small Vessel Disease
CT	Computed Tomography
DC	Discriminative Clustering
DMN	Default Mode Network
DoG	Difference of Gaussian
DTI	Diffusion Tensor Imaging
ECM	Evolving Clustering Method
EEG	Electroencephalography
EM	Expectation-Maximization
FLAIR	FLuid Attenuated Inversion Recovery
fMRI	Functional Magnetic Resonance Imaging
GM	Grey Matter
ICA	Independent Component Analysis
kNN	K-Nearest Neighbours

LADIS	Leukoaraiosis And DIsability Study
LoG	Laplacian of Gaussian
MDL	Minimum Description Length
MEG	Magnetoencephalography
MPRAGE	Magnetization Prepared Rapid-Acquisition Gradient-Echo
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MT	Magnetization Transfer
NLP	Natural Language Processing
PD	Proton Density
PDF	Portable Document Format
PET	Positron Emission Tomography
PMI	Pointwise Mutual Information
PNS	Peripheral Nervous System
PV	Partial Volume
PVE	Partial Volume Effect
RF	Radio Frequency
ROI	Region of Interest
SIFT	Scale-Invariant Feature Transform
SNR	Signal-to-Noise Ratio
SOM	Self-Organizing Map
SPECT	Single-Photon Emission Computed Tomography
SVM	Support Vector Machine
T_1c	T_1 sequence with gadolinium contrast
WM	White Matter
WML	White Matter Lesion

List of Symbols

$\cent$	Class or label of a data point
κ_c	Conformity coefficient
κ_d	Dice coefficient
N_d	Number of documents in a collection
N_w	Number of words in a collection
T_1	T_1 relaxation time, also know as spin-lattice relaxation time
T_2	T_2 relaxation time, also know as spin-spin relaxation time
$\cdot$	Dot product
$\oplus$	XOR logical operator
$\overline{x}$	Mean of a data vector x
$\|\cdot\|_p$	L_p-norm of a data vector or matrix
$\|\cdot\|$	Euclidean norm of a data vector or matrix
FN	Number of false negatives points in a classification task
FP	Number of false positive points in a classification task
TN	Number of true negative points in a classification task
TP	Number of true positive points in a classification task
x	Data vector
$\vee$	OR logical operator
$\wedge$	AND logical operator
$c_f(w_n)$	Collection frequency; total number of occurences of term w_n in a collection
$d_f(w_n)$	Document frequency; number of documents where term w_n appears
d_j	Document in a collection
$f_j(w_n)$	Term frequency (*tf*); frequency of term w_n in document d_j
h_{ck}	SOM neighbourhood function
t	Iteration step in an algorithm

w_n	Word, term or n-gram in a document collection
w_{nk}	Membership of a data point to a cluster
$\mathbf{E}[\cdot]$	Expected value of a random variable
$|\cdot|$	absolute value of a number
κ_j	Jaccard index
η_{sbl}	Sensibility measure
η_{svt}	Sensitivity measure
η_{spf}	Specificity measure
N	number of samples in a data set
n	Sample in a data set
rms	root mean square error
S	dimensionality of a data set
$\boldsymbol{X}$	a data matrix $\in \mathbb{R}^{S \times N}$

1. Introduction

"Begin at the beginning," the King said gravely,
"and go on till you come to the end: then stop."
— Lewis Carroll, *Alice in Wonderland*

The past decades have seen a tremendous amount of technological break-throughs, from the world wide web to the human genome mapping. In particular, recent advances in medical sciences and technology have lead to an ever increasing life expectancy. This leads to a concomitant increase in diseases capable of disrupting normal cognitive processes. The socio-economic impact of these issues is so crucial for the modern world that the European Union has openly supported research in this area, through their two last Framework Programmes.

Examples of diseases with a tremendous impact in the society include Alzheimer's, Parkinson's, Autism and multiple sclerosis (MS). Their diagnosis is often difficult, and occurs usually after cognitive impairments have already compromised daily living activities. This results in serious costs in palliative care, and a significant decrease in the quality of life, due to the progressive nature of those diseases. Therefore, the development of technical solutions capable of helping health-care specialists in their tasks is continuously required.

Such problems are within the scope of neuroinformatics and computational neuroscience. Among other tasks, these two research fields focus on the development of methods suitable for the organisation and analysis of neuroscience data.

1.1 Scope of the dissertation

The topics discussed in this dissertation are related to Neuroinformatics. Two main research directions are discussed, focusing on the use of machine learning approaches to study problems with great neuroscientific relevance. They address the analysis and exploration of medical imaging data, with a clear focus on MRI processing.

The first research direction deals with the development of a brain tissue segmentation method. Such a method should be able to detect early progressive brain lesions, enabling a pro-active intervention, reducing therapy costs, and allowing for a concomitant improvement in the well-being for the elderly. The field of brain image segmentation has evolved rapidly in recent years, but a reliable identification of brain tissues and lesions still poses significant challenges. Manual segmentation methods, still considered the best ground-truth source, are subjective and expensive (Commowick and Warfield, 2010). While several methods that avoid the use of manual segmentation have been developed (Klauschen et al., 2009; Mortazavi et al., 2012; Valverde et al., 2014), the problem is far from solved. In particular, the detection of new foci of lesions (Elliott et al., 2013) and the study of lesion evolutions (de Groot et al., 2013) are still under heavy research. In this dissertation, a new brain tissue segmentation procedure is proposed, capable of identifying specific brain tissues and of early lesion detection.

While the segmentation and detection of brain lesions are essential when studying the effects of a new medication and/or surgical intervention, it is only half of the coin. Research on the neurological implications of different lesions is crucial to understand the human brain. This is closely related to the second research direction discussed in this dissertation: neuroscience meta-research. With the vast amount of data being produced by several researchers, the task of how to integrate and analyse such data is highly demanding. When a seasoned neuroscientist searches for the explanation of an unexpected brain activation, or a researcher attempts to validate a newly proposed analysis method, it is rather common to spend a considerable amount of time scanning through a vast list of publications, in search for comparable experimental outcomes. A proper compilation of such reported information is therefore crucial. Several technologies have been developed to handle such task, such as brain atlases (Laird et al., 2005; Sunkin et al., 2013; Van Essen and Dierker, 2007) and standard formats

for neuroscientific data (Kruggel and von Cramon, 1999; Talairach and Tournoux, 1988). Nonetheless, they are all limited by the lack of publicly available information. By conducting an exploratory meta-research on the widely available reports of neuroscientific studies, it is possible to harvest brain imaging research findings, as well as the neuroscience relevant outcomes therein. Such an approach, based on visual information extraction of functional magnetic resonance imaging reports, using text to help categorize such information, is another contribution of this dissertation.

Despite the apparent difference in focus of the two main research directions explored in this dissertation, both rely on machine learning clustering approaches. Albeit being thoroughly studied and with many proven applications, the field of data clustering still faces many research challenges. While supervised methods typically require an extensive and accurate set of training data, completely unsupervised methods are often dependent on initialisation and parameter settings. On the other hand, semi-supervised approaches, where only a limited amount of training data is available, have been shown to improve clustering accuracy (Chapelle et al., 2006). The third contribution presented in this dissertation is a method to obtain training labels, by estimating data consistency across several runs of unsupervised clustering methods.

1.2 Scientific contributions of the thesis

The core contributions of this thesis, illustrated in the eight publications appended and summarised in this section, are three-fold. First, the retrieval of consistent information from data clustering is proposed. Relying on such an approach, the other contributions are a brain tissue segmentation method and a document mining framework.

1.2.1 Data clustering consistency estimation

One of the main problems in clustering approaches is parameter selection, *e.g.* the selection of the number of clusters to find, or the geometry of the clustering space. While there are several techniques to alleviate such limitation, this problem is still under heavy research. The first contribution of this dissertation, proposed in **Publication I**, is a different approach to solve the aforementioned problem. It relies on searching for consistency among different runs of clustering algorithms. In that pub-

lication, a procedure to exploit the intrinsic variability of stochastic clustering algorithms is discussed. Using such an approach it is possible to identify a subset of the data that reliably represents different clusters.

1.2.2 Self-supervised brain tissue segmentation

The second contribution of this dissertation is a semi-supervised method to identify different brain tissues, including brain lesions, and their partial volume information. Such method was first proposed in **Publication II**, and it relies on the differences between intensity value distributions of each tissue, with minimal to no added *a priori* information. This approach is improved and extended in **Publication III**, where it is compared to other tissue segmentation methods, in both simulated and real data. This methodology was also applied to the study of age related white matter diseases in **Publication IV**, where it was shown that barely visible changes of normal-appearing brain tissue can already predict the decline in cognitive processes. Furthermore, **Publication V** demonstrated, in a longitudinal study, that it is possible to predict and detect early foci of lesion using the proposed methodology.

1.2.3 Multi-modal mining of neuroscientific documents

In the field of neuroscience meta-research, most methods rely solely in textual information, discarding the information visually reported. The third contribution of this thesis is a framework that extracts and characterises visually reported brain activity from neuroscientific publications, specifically dedicated to the meta-analysis of functional magnetic resonance imaging (fMRI) reports. While **Publication VI** proposed a clustering method solely based on the creation of feature vectors from fMRI images, this was later extended to an analysis of three-dimensional representations of reported brain activity in **Publication VII**. The framework was further improved in **Publication VIII**, by using multi-modal information to characterise summarising maps of brain activity.

1.3 Structure of the thesis

After the introduction in the first chapter, Chapter 2 provides the reader with essential concepts in brain anatomy, neuroimaging, functional connectivity and brain lesions. Such concepts are required for a proper understanding of the topics discussed throughout the dissertation. Furthermore, several core methodologies of image processing are used in this dissertation, and Chapter 3 provides an overview of such techniques. It focuses mainly on biomedical signal analysis approaches, in particular methods applied to the pre-processing of magnetic resonance images.

The two main research directions proposed in this dissertation rely on clustering approaches, where data is grouped into different clusters, according to a distance measure. Chapter 4 discusses several such techniques and some of the methods used to measure similarity between data elements. The last section of that chapter presents a consistency based analysis of clustering results, which is the first contribution of this thesis.

Chapter 5 is dedicated to the field of brain tissue segmentation. This field deals with the reliable identification of brain tissues and lesions. Several segmentation methods are briefly reviewed, as well as some of the concepts behind them. After presenting the problem of tissue segmentation, the last sections of Chapter 5 describes the second contributions of this thesis, a self-supervised brain tissue segmentation method, and contains some exemplary results.

Chapter 6 deals with the topic of data mining and information retrieval. In particular, it addresses the problem of filtering and obtaining relevant content from various multi-modal sources, with both text and images. Several methods of information retrieval are detailed, with the core concepts of the field also explained. Particular relevance is given to approaches dealing with document mining of neuroimaging publications, which is the topic of the third contribution of this thesis. The framework of that contribution, as well as some of the results obtained, are presented in the last section of Chapter 6.

Finally, Chapter 7 summarises the proposed methods and discusses their advantages and limitations.

A summarising schematic of the different chapters of this dissertation can be seen in Fig. 1.1, where the relations between the several topics discussed are also illustrated.
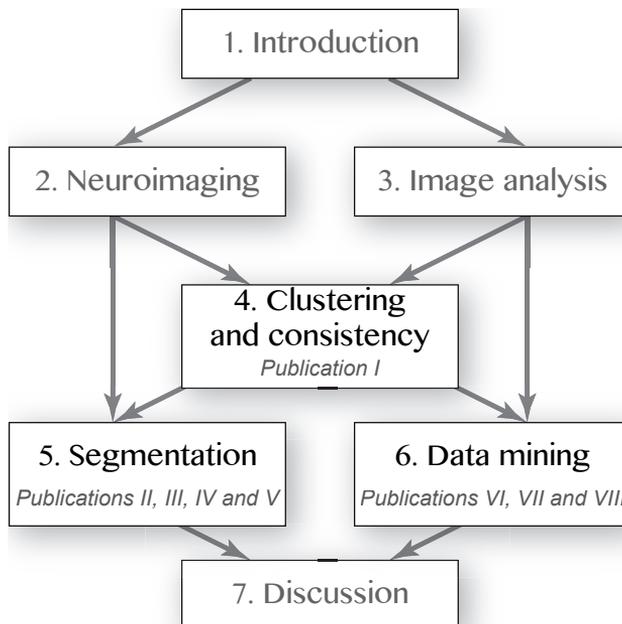
Figure 1.1. Flowchart depicting a summarising view of the different chapters of this dissertation, and the relation between the topics discussed therein.

The main aims of this thesis can thus be summarized as the following:

- methodology to analyse clustering consistency.

- self-supervised approach for tissue segmentation

  - theoretical and practical implementation
  - clinical applications

- neuroscientific data mining

  - extraction of brain activity from fMRI images
  - creation of a multi-modal brain ontology.

2. Brain and neuroimaging

Starting with the *Edwin Smith Papyrus*, ca. 1500 BC, and perhaps earlier, the brain has been a major topic of study. In biomedicine, psychology and information theory it plays a major role in research. With the recent advances in imaging and analysis techniques, our knowledge of the human brain has increased tremendously. In this chapter, the anatomy of the brain is briefly reviewed, as well as an overview of current neuroimaging techniques, with particular focus on magnetic resonance. This introduction is mostly based in Kalat (2012) and Kandel et al. (2000), which provide a comprehensive analysis of the human brain. The last two sections of this chapter detail several brain lesions and describe the default mode network.

2.1 Brain

In vertebrates, the nervous system can be divided in the central nervous system (CNS) and the peripheral nervous system (PNS). The PNS is responsible for connecting the CNS to the rest of the body. It controls involuntary and vital activities, such as respiration and the functioning of the *cardio-vascular* system. The human brain is the core of the CNS in humans, and the primary control centre for the PNS (Kandel et al., 2000). Thought and reasoning, the so called conscious activities, are some of responsibilities of the brain (Simon, 1999).

The brain is typically represented from three perspectives, as shown in Fig. 2.1: *axial*, along the horizontal plane that travels from the anterior to the posterior parts of the brain; *sagittal*, travelling along the lateral plane, from left to right; and *coronal*, along the frontal plane, that travels from dorsal to ventral.

On a typical adult human, the brain weighs around 1.5 kg and has an
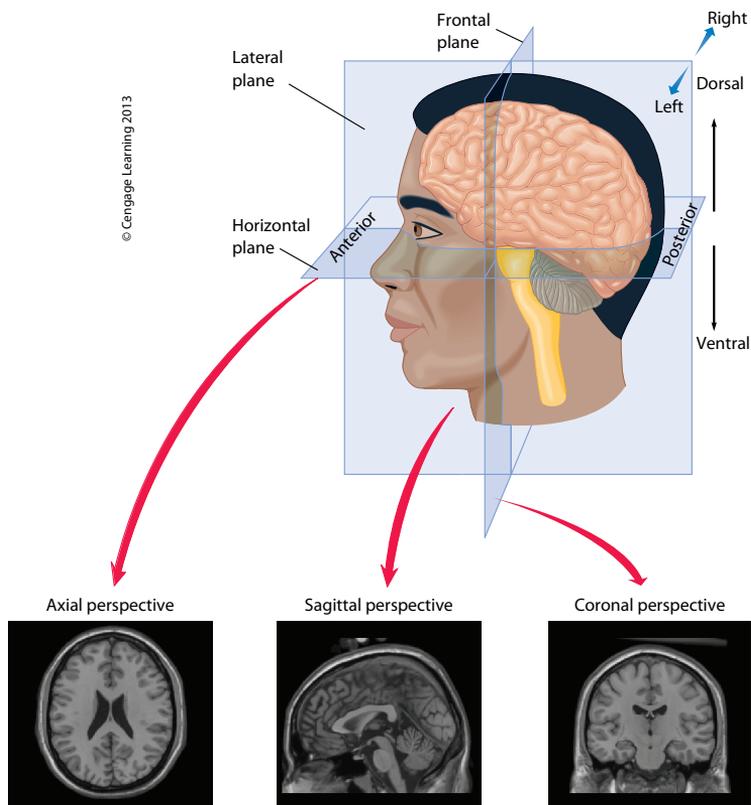
Figure 2.1. Anatomical directions used to depict the brain. Adapted from Kalat (2012).

average volume of 1,600 cm^3 (Crossman and Neary, 2006). It can be divided in three main parts: the hindbrain, the midbrain, and the forebrain. The *cerebellum*, together with the *pons* and the *medulla*, compose the hindbrain. The midbrain (*tectum*, *tegmentum* and *substantia nigra*), together with the *medulla* and *pons* are often referred collectively as the brain stem. These structures are almost completely surrounded by the forebrain, with only the *medulla* visible as it merges with the spinal cord. The forebrain can be divided in two hemispheres, left and right, with the cerebral cortex corresponding to its outer portions. Fig. 2.2 shows the basic anatomical structure of the brain. A typical division of the cerebral cortex corresponds to four "lobes": the frontal, parietal, occipital, and temporal lobes. These lobes are separated by sulci and fissures, and the cells in each lobe have a specific structure and function:

> *frontal lobe*: typically associated with reasoning, higher level cognition, and expressive language. It also includes the motor cortex, which is responsible for body movements and motor skills.
>
> *parietal lobe*: includes the somatosensory cortex, responsible for pro-

cessing sensory information such as pressure, touch, and pain.

temporal lobe: contains the primary auditory cortex, that deals with sound and language, and the hippocampus, which is heavily associated with the formation of memories.

occipital lobe: is associated with visual stimuli and information through the visual cortex.

Other brain structures worthy of note include the precuneous, usually associated with self-consciousness, and the cingulate gyrus, which is involved in emotion processing, learning, and memory.
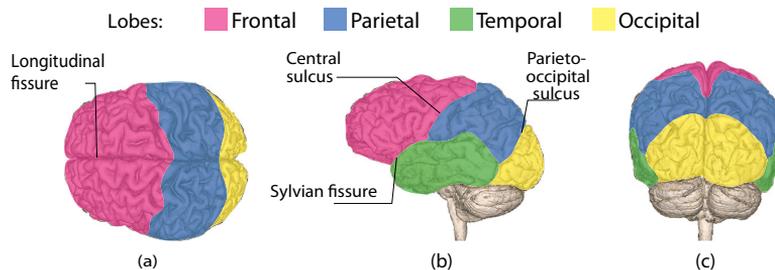
Figure 2.2. Some of the main anatomical and functional structures of the brain, including the four lobes and many relevant foci and sulci. (a) is shown from a top perspective, while (b) and (c) are taken from sagittal and coronal planes, respectively.

2.1.1 Brain tissues

From the various different types of tissues in the brain, this section will focus on those that are of interest for this dissertation, which are highlighted in Fig. 2.3. In that figure, one can also see an area of lesion.
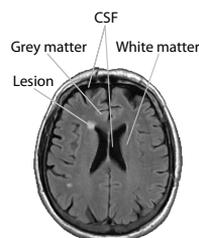
Figure 2.3. Brain tissues of interest to this dissertation, including lesioned areas.

In the early stages, the nervous system is a tube surrounding a fluid canal (Kalat, 2012). This canal develops into a central canal, a fluid-filled channel in the centre of the spinal cord, and into the ventricles, four fluid-filled cavities within the brain. The *choroid plexus* cells, inside the four ventricles, produce cerebrospinal fluid (CSF), filling the ventricular space (Nolte, 2008), and can be clearly seen as the darkest tissue in Fig. 2.3. The *subarachnoid* space, which separates the soft tissues of the brain and spinal cord from the hard sur-

rounding bones (skull and *vertebræ*), is also filled with CSF. This fluid absorbs shocks to the brain and also provides buoyancy. It is very similar to blood plasma and is the main supplier of nutrients to the brain.

The cerebrum is an area of the brain composed by the cerebral cortex, the basal ganglia and the limbic system (Kalat, 2012). As a whole, it is responsible for functions such as conscious decisions and movement (Kalat, 2012; Kandel et al., 2000). It is composed of an outer layer of grey matter (GM), corresponding to areas densely packed with cell bodies and dendrites. On the other hand, the cerebrum is internally supported by deep brain white matter (WM), consisting mostly of myelinated axons. The task of the myelin is to insulate nerve endings and enable brain signals to move smoothly (Kalat, 2012).

2.2 Neuroimaging

Neuroimaging is the field dedicated to techniques that allow to either directly or indirectly image the brain (Mazziotta et al., 2000; Zimmerman et al., 2000). This type of medical imaging may be divided in two categories: structural and functional neuroimaging. The former deals with imaging brain structures, and a precise diagnosis of intra-cranial structural pathologies. The latter attempts to obtain images of the functional organisation of the brain, mapping mental processes to regions in the brain (Huettel et al., 2008).

Examples of structural imaging techniques are computed tomography (CT) and anatomical MRI, while electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), single-photon emission computed tomography (SPECT) and fMRI deal with functional imaging. CT makes use of x-rays, taken from many directions, where the transversed rays are detected on the other side of the brain. Using a computerised algorithm, three dimensional images of the brain can be reconstructed (Webb et al., 2005), with very high spatial resolution. Both EEG (Schomer and da Silva, 2011) and MEG (Hämäläinen et al., 1993; Hansen et al., 2010) have millisecond resolution, and are used to see responses to sensory stimuli, and mostly to observe typical rhythmic activity in the brain. EEG was the first truly non-invasive neuroimaging technique discovered, and uses an electroencephalograph to measure electrical fields in the cerebral cortex, via electrodes attached to the scalp (Nunez and Srinivasan, 2005). MEG, on the other hand,

measures magnetic fields, indirectly associated with the electrical activity. Many healthy recordings are already well known (Head, 2002), and any abnormalities may suggest epilepsy, tumours or other medical pathologies. PET is a neuroimaging technique able to target very specific functional processes of the brain (Wahl, 2002). It works by detecting photons resulting from the annihilation of electrons and positrons (Valk et al., 2004). The positrons are released by the decay of radio labelled compounds (radio tracers) that are usually injected into the subject's bloodstream. It presents spatial and functional resolutions between the aforementioned methods, making it also an interesting approach when dealing with the localisation of particular neural functions. A similar approach is used in SPECT (Rosenthal et al., 1995) but, unlike in PET, the tracers emit gamma radiation that is measured directly. This method is often combined with CT (Mariani et al., 2010), increasing its specificity and spatial resolution.

The next subsections will focus on the MRI techniques mentioned previously, which stand at the core of this thesis.

2.2.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that uses strong magnetic fields and radio waves to produce high quality three-dimensional images of biological tissues (Meacham, 1995). While this technique can be used in a wide variety of settings, such as cardiovascular imaging (La Gerche et al., 2013; van der Geest and Reiber, 1999) and musculoskeletal analysis (Junno et al., 2013; Khoo et al., 2011), this dissertation will focus on its application in neuroimaging.

In order to create MRI images, an oscillating magnetic field is used. When this field is turned off, the atomic nuclei that had absorbed electromagnetic energy in a specific frequency release that energy (Webb, 1988). Sensors read these emissions and the target's images are created. In comparison with other techniques, MRI has the advantage of being almost completely harmless to the subject's health and allow for the distinction between soft tissues (Huettel et al., 2008). In addition, MRI produces images with high resolution, comparable to the ones obtained through CT scans. One disadvantage of this technique is its inability to be used in subjects with some kind of electronic implants, due to their magnetic sensitivity, like pacemakers (Kanal et al., 2002), and the slowness of the process. MRI is used to create images of both surface and subsurface

stationary structures, with a high degree of anatomical detail. Since it is beyond the scope of this chapter to provide a full description of the physics behind MRI, only a small overview of the main concepts will be detailed in this section.

MRI is based on the principles of nuclear magnetic resonance. It measures spatial variations in the phase and frequency of the radio frequency energy being absorbed and re-emitted by the imaged object (Westbrook and Kaut, 1998).

There is a huge amount of atoms in the human body. The ones with nuclei with an odd number of both protons and neutrons possess a property called spin (Griffiths, 2004), which is its intrinsic *magnetic angular momentum* and can attain any value multiple of $\pm 1/2$. Due to the charge in the atomic nuclei, the spinning motion causes a magnetic dipole moment in the direction of the spinning axis. The strength of the magnetic moment is a property of the type of nucleus. The hydrogen atom is an ideal atom for MRI, because it possesses a strong magnetic moment in its nucleus and there is a large quantity of them in biological tissues (Huettel et al., 2008). When atomic nuclei are placed in a uniform magnetic field, their magnetic moments have a tendency to align with that field. The predominant direction of non-zero spins of the atomic nuclei in the tissue is one of two, in respect to the magnetic field: parallel and anti-parallel. The majority of the spins will be in parallel state, which is a lower energy state than the anti-parallel one. The spin axes are not exactly aligned with the magnetic field, but precess around it with a frequency characteristic of the nucleus type. This frequency is called the Larmor frequency or resonant frequency, and corresponds to the precession frequency of the dipole moments.

Two important properties of any magnetic field are its *field homogeneity* and its *field strength*. To generate images that do not depend on the MRI scanner or body positioning, the fields employed need to be uniform, in space and time. Also, and to generate such strong and stable fields, modern MRI scanners use superconducting electromagnets, at temperatures near zero.

After exposing the nuclei to a magnetic field, a transient radio frequency (RF) pulse at a specific frequency is briefly applied, in a plane perpendicular to the main magnetic field. The RF pulse excites some of the spins in the lower-energy state at their resonant frequency, disturbing the aligned hydrogen nuclei, thus causing a disruption of the equilibrium. The phe-

nomenon caused by this *excitation* is called magnetic resonance (Westbrook and Kaut, 1998).

The return of the nuclei to their equilibrium state is known as *relaxation* (Bloembergen et al., 1948). The resulting release of energy can be detected by radio-frequency coils, in a process known as *reception*. This detected electromagnetic pulse defines the raw MR signal.

The relaxation time is mainly governed by two physical processes (McRobbie et al., 2002): T_1 and T_2 relaxation, characterised by two magnetic recovery times. T_1 relaxation is the realignment of spins with the external magnetic field, while T_2 corresponds to the transverse relaxation of the component of the nuclear magnetisation. Both relaxation times are characteristic of the type of tissue considered (see Table. 2.1).

Table 2.1. Examples of T_1, T_2 and proton density[1] values of different tissues in the brain at field strength 1.5T. From Huettel et al. (2008) and Bradley and Stark (1999).

Tissue	WM	GM	CSF
T_1 (ms)	600	900	4000
T_1 (ms)	80	100	2000
PD (H)	0.61	0.69	1

[1] Assumes CSF has the proton density of water (H=0.11 moles of hydrogen/cm^3).

The obtained MR-signal does not contain any spatial information. To encode all needed information, additional gradient coils are used to impose three linear orthogonal field gradients on the magnet generating the external magnetic field. One common problem in MR scanners is that both the external and the field gradients are rarely completely linear, which is clearly a source of image distortion. Therefore, modern scanners use shimming coils to compensate for inhomogeneities. Unlike other coils, these shimming coils are also adjusted for each subject, since the anatomy of each person affects the fields differently.

In MRI, a slice through the studied object is typically selected. Each slice is divided into volume elements (voxels), usually with 1 or 2 mm^3 (Huettel et al., 2008). Each one of them relating to a picture element (pixel) in the MR image. The pixel value (usually between 0 and 255) is determined according to the signal originating from the corresponding voxel. Let us

assume that we have a main gradient field along the z axis. Then, for a given frequency of the RF pulse, only the orthogonal x-y plane will be excited. Everywhere else, the sample receives the wrong resonance frequency of excitation. This technique allows for a given slice to be selected from the whole sample, with thickness determined by the magnetic field gradient strength.

Once the slice is selected, the MR signal data has to be combined with each individual voxel, in order to get the values for each pixel. To achieve this, phase and frequency encoding gradients are needed. With a slice selected and excited as described previously, the current is switched to one of the two remaining gradient coils, creating the phase encoding gradient. This gradient is applied along one side of the image, perpendicularly to the external magnetic field and to the slice selection gradient. The gradient is, however, switched off after a short period, letting the nuclei gain their original frequency. A significant difference arises: the nuclei are out of phase, each row possessing a phase of its own, leading to phase encoding. The remaining coil then produces a frequency encoding gradient, perpendicular to the other two gradients. This gradient will increase the precessing frequencies of the nuclei in the same way as the phase encoding gradient. This has the effect of spatially encoding the excited slice along one axis, so that columns of spins perpendicular to the axis precess at slightly different Larmor frequencies.

Thus, using three orthogonal magnetic field gradients, the signal emitted from a specific voxel will have a specific frequency and phase. Using inverse Fourier transform, the individual signals from each voxel, together with their locations can be extracted from the MR signal and thereby construct the MR image. By combining different slices, three-dimensional MRI images can then be obtained.

Multi-spectral MRI

One great advantage of MRI over other brain imaging techniques, is the ability to design different imaging sequences (spectra), which allow for discrimination of different tissues. By changing two user-selectable delay times for the RF pulse, different effects can be highlighted. The echo delay time or time of echo (TE) corresponds to the interval between excitation and acquisition of the signal. On the other hand, the sequence time of repetition (TR) corresponds to the amount of time between successive excitation pulse sequences. A third factor that contributes to the

MR signal is the proton density. It is independent of TE and TR, but proportional to the effective number of hydrogen nuclei per unit volume, *i.e.* it depends directly on the tissue type. Since each tissue type has a different set of relaxation times, as shown in Table. 2.1, variations in TR and TE result in changes in the relative contrast between tissues, depicted in Fig. 2.4 (Edelman et al., 2005).
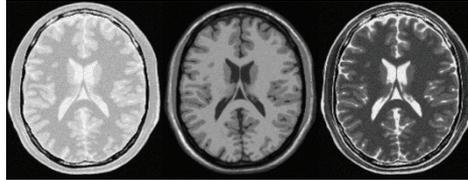
Figure 2.4. Example of three different MRI sequences. From left to right, PD, T_1 and T_2.

With a long TR, all tissues recover their longitudinal magnetisation, thus the differences in T_1, across tissues, will not influence the signal. Conversely, a short TE implies that the differences due to T_2 do not have enough time to become observable. Thus the images formed have different appearances, depending on the time constants. Short TR and TE lead to T_1 weighted images, and long TR and TE to T_2 weighted ones.

Another form of differentiation between tissue types, is to use proton density (PD) imaging. Using a very long TR and very short TE, the resulting images are mostly influenced by the differences in proton density (Edelman et al., 2005; Huettel et al., 2008).

When dealing with damaged tissues, other contrast sequences can be used. Damaged tissues tend to develop edemas, which makes a T_2 sequence sensitive to the distinction of pathological tissue from normal tissue (Tofts, 2003). With some additions to the RF pulse and specific manipulations of the magnetic gradients, a T_2 sequence can be converted into fluid attenuated inversion recovery (FLAIR), in which free water is now dark, but edematous tissues remain bright. This sequence is particularly suited to the evaluate demyelinating diseases in the brain (Hori et al., 2003; Khademi et al., 2012), such as multiple sclerosis.

Several other sequences have also been used in clinical settings, such as T_1 sequence with gadolinium contrast (T_1c) (Caravan et al., 1999; Geraldes and Laurent, 2009), magnetization transfer (MT) (Symms et al., 2004) or magnetization prepared rapid-acquisition gradient-echo (MPRAGE) (Brant-Zawadzki et al., 1992).

The acquisition of different sequences from the same brain, as shown in Fig. 2.4, is typically called multi-modal or multi-spectral data. The

different sequences acquired provide complementary information regarding the brain tissues, allowing for a better evaluation and distinction between tissue types. Nonetheless, this data has some drawbacks. Namely, it is rather expensive in time and use of scanning devices, and requires careful registration between the collected images. This registration usually results in taking the lowest spatial resolution dataset as the overall target (Sotiras et al., 2013).

2.2.2 Functional Magnetic Resonance Imaging

The knowledge of brain structures provides much information. Nonetheless, the active functioning of the brain cannot be studied with anatomical MRI, due to the rather long times used for image collection. Therefore, researchers turn to functional neuroimaging to study which parts of the brain are associated with particular mental processes, and how these interact with each other. A typical functional report displays images associating brain activity with a given mental task.

One of the most common modern methods of functional neuroimaging is functional magnetic resonance imaging (fMRI) (Huettel et al., 2008). Most fMRI scanners allow for the measurement of brain activity while subjects react to stimuli (Buxton, 2001). Consequently, fMRI can be used to reveal brain regions and processes associated with perception, thought and action.

Since neural information processing activity requires higher metabolism, the vascular system supplies this energy locally. Most of this energy comes from oxygen, which is bound to hæmoglobin molecules. Therefore, changes in the oxygen consumption are directly related to the concentration of de-oxygenated hæmoglobin. These changes can be easily seen in blood oxigenation level dependent (BOLD) (Ogawa et al., 1992) contrast MR images.The contrast in those images is based on paramagnetic property changes of water molecules, reflecting the concentration of deoxygenated hæmoglobin. Therefore, fMRI images do not give a direct measure of brain activity, but only a correlated measure of such activity, obtained with different stimuli conditions.

BOLD contrast differences between images are quite small, and changes directly related to neuronal activity are even smaller, when compared to other spatio-temporal variations. Therefore, a careful analysis of the resulting functional images is required. Modern scanners are capable of capturing full head volumes within a few seconds interval, but the resolu-
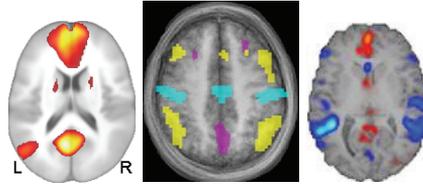
Figure 2.5. Example of fMRI images. On the leftmost image (taken from Johnson et al. (2007)), activity is present in the occipital, left temporal and frontal areas of the brain, and the activity is reported using the *hot* colour scale. The activity on the second image (adapted from Esposito et al. (2010)) is shown in three different uniform colours, while the third image (adapted from Yli-paavalniemi and Vigário (2008)) shows a combination of *hot* and *cold* colour scales, corresponding to increase and decrease of activity, when compared to a baseline reference.

tion is about two or three millimetres at present. The resolution of fMRI images limits the ability to distinguish between different functional brain regions in close proximity of each other, and is restricted by the spatial spread of the hæmodynamic response to neural activity (Jezzard et al., 2003). Furthermore, due to the slowness and variability of the vascular system response to neuronal activity, the estimation of brain activity timings is also somewhat limited. Other problems, such as head movements and physiological changes during measurement, will also distort the images. Therefore, fMRI acquisition is a compromise between fast scanning and high resolution images or, in other words, it has a good spatial and temporal imaging resolution. Nonetheless, and also in spite of its inability to identify specific brain receptors associated with particular neurotransmitters (Wahl, 2002), fMRI has surpassed PET in the study of brain functional activation patterns.

Typical fMRI images display changes in brain activity using colour overlays onto anatomical MRI, as shown in Fig. 2.5. These can reflect either the degree of change, as depicted one the left frame, or simply the region where functional change occurred, as seen in the middle frame. Often, to show also the sign of neuronal activity changes, many researchers use *hot* and *cold* colourmaps to show an increase or decrease of activity, respectively. This is shown in the rightmost frame of Fig. 2.5.

2.3 Brain disorders

With an increasing elderly population in the modern world, disorders of the human nervous system have become one the most debilitating and devastating human illnesses. These not only physically affect the suffer-

ing patient, but also comprise their ability to function in society. As such, brain disorders increase the burden, not only of patients, but also their families and the society at large, due to the economic and social adjustments they require. Hence, brain disorders have become a main focus in health-care research. Accordingly, the use and development of neuroinformatics techniques has become such an invaluable tool in diagnosing, monitoring and studying brain diseases (Mazziotta et al., 2000).

Although there are many types of brain diseases, a detailed explanation of them is beyond the scope of this chapter. Therefore, only four types of brain disorders will be briefly described in this section, since they play an important role in this dissertation. In addition, due to their progressive nature and relevance for this thesis, tumours or neoplasms are also described here. All disorders mentioned in this section are deemed progressive diseases, due to the progressive changes they cause in the brain structures and cells. Based on those changes, and the very high spatial resolution of the imaging technique, MRI is often used to study the aforementioned diseases. It should be noted, as well, that an early diagnosis of these diseases may lead to a significant improvement in the quality of life of a patient, as well as an increase in the impact of its treatment.

2.3.1 Neurodegenerative diseases

Neurodegenerative diseases are the result of progressive structural or functional neuron deterioration (Jolles and Stutzmann, 1994). Alzheimer's, Parkinson's and Huntington's are typical examples of such diseases. As we age, the brain starts to fail and the incidence of neurodegenerative diseases increases greatly (Jolles and Stutzmann, 1994). In particular, Alzheimer is already one of the most common causes of death in the developed world, and is likely to keep rising, due to the continuous increase in life expectancy.

2.3.2 Vascular diseases

Any disease that affects the circulatory system, is a vascular disease. They can affect any blood vessel in the body, and occur due to endothelial cell dysfunction. Examples of vascular diseases include aneurysms, Buerger's disease and Peripheral Artery disease. When dealing with the brain, cerebral small vessel disease (cSVD) is the most common cause of vascular cognitive impairment and dementia(Ferro and Madureira, 2002;

Jokinen et al., 2011). One of the main signs of cSVD are white matter lesions (WMLs), together with lacunar infarcts, micro-bleeds, and brain atrophy. These can all influence clinical and cognitive outcome (de Groot et al., 2001; Jokinen et al., 2012; Muller et al., 2011; Poels et al., 2012). Age-related WML can be observed through changes in white matter, which result in deviations of the grey levels in MRI (Mäntylä et al., 1997). Even the smallest age-related WML can cause executive dysfunction and progressive cognitive decline, as show in Publication IV.

2.3.3 Demyelinating diseases

As mentioned in subsection 2.1.1, WM is one of the tissues in the human brain, in which myelin is an important constituent. This molecule plays an essential role in the propagation of action potentials across axons. Demyelinating diseases are disorders that affect the myelin sheath covering the axons, and hence impair the normal functioning of the brain. MS belongs to this group of diseases(Compston and Coles, 2008; Keegan and Noseworthy, 2002). MS is an autoimmune chronic disease, with a wide range of symptoms (Grossman and McGowan, 1998), including fatigue, visual problems, balance problems and altered sensations. While not directly lethal, many of its side effects lead to a noticeable deterioration in quality of life. The main cause of this disease is not known, but genetic changes are suspected to be a partial cause (Hill, 2003). Furthermore, MS is hard to diagnose, because the symptoms vary heavily from one patient to another.

2.3.4 Chronic psychiatric disorders

The most common example of chronic psychiatric disorders is Schizophrenia. Schizophrenia is a chronic, severe, and disabling neural disorder (van Os and Kapur, 2009), that leads to abnormal social behaviour and a misinterpretation of reality. Several factors seem to be responsible for Schizophrenia, such as genetic predisposition and the environment. The onset of symptoms occurs usually during the first stages of adulthood.

2.3.5 Neoplams/Tumours

Neoplasia, from the dictionary definition, is the formation of tumours. These are collections of tissues, resulting from an abnormal growth or division of cells. Their causes are diverse, but usually derive from genetic
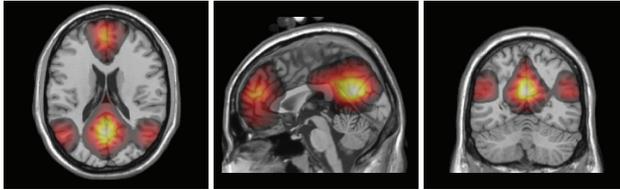
Figure 2.6. Average brain activity for the DMN from axial, sagittal and coronal perspectives, respectively. Most of the activity is located in the occipital, temporal and frontal areas of the brain.

mutations (Weinberg, 2007). Tumours can be benign or malignant (cancer). Since tumorous cells have a slightly different composition than other cells (Weinberg, 2007), and hence react differently to contrast agents, MRI and PET have been widely used for its diagnosis and monitoring.

2.4 Default Mode Network

The default mode network (DMN) was first described in 2001 by Raichle et al.. Using fMRI, certain areas of the brain exhibited a decrease in activity when "the control state was passive visual fixation or eyes closed resting". After that first publication, the topic of resting state and default mode network (DMN) became one of the most researched fields in the field of neuroscience. The DMN comprises areas such as the occipital, temporal and frontal areas of the brain, as seen in Fig. 2.6. They are active when the individual is not performing any goal-oriented task, and suppressed during activity (Buckner et al., 2008; Deco et al., 2011; Raichle et al., 2001). In spite of the great attention given to the study of those networks, scientific research of the brain's "default state" still poses various conceptual and methodological difficulties (Morcom and Fletcher, 2007; Snyder and Raichle, 2012; van Oort et al., 2014).

The DMN has been hypothesised to be heavily involved in some brain disorders, such as Alzheimer (Damoiseaux et al., 2012; Johnson et al., 2007; Koch et al., 2012; Wen et al., 2013), Schizophrenia (Dodell-Feder et al., 2014; van Os and Kapur, 2009; Whitfield-Gabrieli et al., 2009; Yu et al., 2012) or autism (Paakki et al., 2010; Starck et al., 2013). Many of those studies focus on comparing the areas composing the DMN in healthy and pathological brains, and how these differences influence cognitive and functional performances. Furthermore, decline in connectivity of the DMN in ageing brains has been a recurrent finding (Damoiseaux et al., 2008).

3. Image analysis

When dealing with imaging data, there are several factors of image quality to take into account. Before any further analysis, images need to be preprocessed to remove artifacts. To do this, the detection of regions/objects of interest is often of great importance. Furthermore, to allow for a proper statistical analysis and comparison between images, their overall characteristics need to be determined. The first section of this chapter briefly describes several preprocessing techniques, typically used in imaging data. It is based on Rangayyan (2004), which provides a comprehensive account of biomedical signal analysis techniques.

Several factors may jeopardise the signal-to-noise ratio (SNR) in MRI images. An increase in the magnetic field strength typically improves the SNR, but it also increases the probability of contamination from other sources, such as artifacts or external noise. Besides the intrinsic noise of both the subject and the scanners, also the scanner drift, produced by uncompensated changes in the magnetic field, increases the system noise. Nonetheless, the major causes for artifacts are related to motion artifacts. Head movement or the cardiac activity can cause misalignments and changes in the acquired images. Although this chapter deals with analysis techniques applicable to images in general, the last section is dedicated to typical MRI processing techniques, which aim at compensating for the aforementioned problems.

3.1   Image preprocessing

Data preprocessing is one of the most important steps in data analysis. As often mentioned and exemplified in the sentence "garbage in, garbage out", without proper handling of raw data, misleading results can and will be produced. Real-world data is usually incomplete, inconsistent, and/or

lacks certain behaviours or trends, and is likely to contain errors such as out of bounds data. Data pre-processing focus on different tasks, such as normalisation, filtering, feature extraction and selection. The aim of pre-processing is to remove uninteresting variability and prepare data for further processing. In imaging data this often equates to detecting the main regions of interest, remove artifacts that compromise a proper analysis, filtering and feature detection. This section will give a brief overview of typical techniques used in the aforementioned tasks, focusing on medical imaging preprocessing used in the studies reported in this dissertation.

3.1.1 Histogram analysis

In gray level images, the dynamic range of intensities provides information on the spread of intensities, but does not describe how the gray level is distributed in that range. The histogram, on the other hand, fully describes the distribution of the gray levels of an image.

In order to enhance the appearance of an image, several histogram-based methods for image enhancement have been proposed Rangayyan (2004). The most common and widely used being histogram equalisation, where the principle is to give equal probability to each gray level. This method increases the contrast of images, and is frequently used when all gray levels of image are closely grouped. By transforming the cumulative distribution of the image into a linear function, the intensity values are spread across all possible gray level values.

An advantage, as well as a major limitation in histogram equalisation, is that there is no control over the procedure or the result, since the transformation aims always at a uniformly distributed probability distribution function. Often, one requires a histogram close to that of another image, a procedure known as histogram specification. This method does have some limitations, such as possible non-invertibility of the transformation, and need for the specification of the desired histogram. Furthermore, the histogram of the resulting image is only an approximation of the desired one, albeit one as close as possible within the potential transformations.

These types of histogram-based methods are often restricted to particular biomedical applications, *e.g.* mammography (Langarizadeh et al., 2011), since their underlying assumptions are not practical in many other situations. For instance, they may increase the contrast of background noise, while decreasing the interesting signal. Nonetheless, they can sometimes provide reasonable improvements in some imaging conditions.

An example of a positive use of such a transformation appears when dealing with MRI sequences acquired at different times, such as the ones in Publication V. In that situation, there are several factors that need to be taken into account. One of the most important deals with intensity distribution differences, observed in the same MRI sequences. As an example, two FLAIR sequences, collected at different times, can have different overall mean intensities, as shown in Fig. 3.1A. Such difference has no physical grounds, but technical ones, hence the need for correction. To compensate for these disparities, a histogram adjustment procedure is used. While the first step corresponds to a simple mean adjustment (Fig. 3.1B), the second one performs a histogram specification, which minimises the differences between the histograms of both sequences (Fig. 3.1C). By avoiding local transformations, a histogram specification procedure allows for a good compromise between reducing image differences and changes to the original histograms.
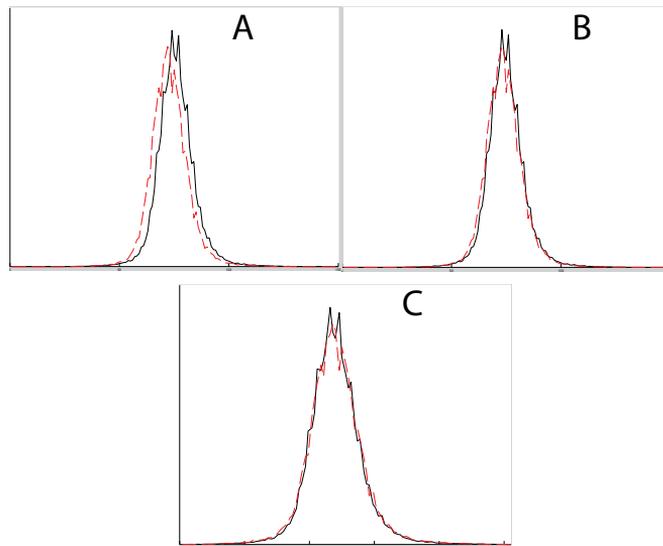
Figure 3.1. Histogram adjustment. Frame (A) shows the original histograms, while frame (B) depicts the mean-adjusted histograms. Frame (C) shows the normalised histograms. The histograms from the first and second acquisitions are shown in black and red, respectively. Adapted from Publication V.

3.1.2 Detection of regions of interest

In computer vision, feature detection, extraction and matching are important parts of data processing. Feature detection corresponds to finding interesting landmarks of an image. It allows for an abstract representa-

tion of visual information elements.

Since the classification of an interesting feature can be quite subjective, many methods have been proposed for feature detection. Visual image landmarks can often be grouped in three main types:

> *Edges*: Points defining a boundary between regions, which are defined by large changes in intensity levels;

> *Interest points*: Also called corners, these points usually have a local two dimensional structure;

> *Regions of interest (ROIs)*: Regions of an image in which some properties are similar throughout.

The type of feature detection and analysis conducted depends heavily on the particular problem at hand.

In a diagnostic situation, medical experts typically focus on small local regions, which are examined in search of abnormalities. These regions are called ROIs, and correspond to a portion of an image that is identified for a specific purpose. Once these regions are detected, the subsequent tasks deal with the characterisation and processing of the data. Typical examples of two-dimensional ROIs in medical imaging include:

- Tumours and masses,

- Calcifications in mammograms,

- Brain activity areas in fMRI, often known as 'blobs'.

In order to detect ROIs, two major concepts are used: discontinuity and similarity. Discontinuity based approaches detect abrupt changes in intensity levels, such as edges, while similarity approaches resort to finding homogeneous parts of the image. This homogeneity can be found through grey-level thresholding or region growing.

The nature of the images, and the ROIs themselves usually determine if edge detection is performed, or if the ROIs are approximated by growing regions. ROIs can also be described using particular points of interest, such as corners or other heavily structured areas.

While the techniques mentioned in this section are described for images, they are easily expanded to three-dimensional data.

Thresholding

Thresholding methods assume that all pixels lying within a specific range of values correspond to the same class. When the gray levels of a particular ROI are known *a priori*, or can be determined from image his-

tograms, then thresholding can be used. Threshold values can be based upon the "valleys" in the histograms. However, even with optimal thresholding techniques, their identification is not easy (Gonzalez and Woods, 2008; Sahoo et al., 1988). Other limitations of these algorithms stem from relying solely on pixel values, with disregard for any spatial information. This often causes problems when dealing with noisy or blurred images.

Edge detection

Another approach to identify ROIs is based on edge detection. Edges are defined as image discontinuities, where there is a large change in intensity levels. This discontinuity can be found in any direction, depending on the edge orientation. The most simple and intuitive methods to detect edges are based on gradients, since they measure rates of change.

In image processing, it is typical to express an operator in terms of odd-sized masks, centred on the pixel to process. One of the most commonly used method for this type of edge detectors is the Prewitt operator. Another commonly used edge detection is the Sobel operator. It is similar to the Prewitt operator, but gives a larger weight to the pixels in the same row or column as the pixel being processed. Using rotated versions of both Prewitt and Sobel operators allows for the detection of diagonal edges.

Yet another method for edge detection is the laplacian of Gaussian (LoG) operator (Gonzalez and Woods, 2008; Rangayyan, 2004). To increase computational efficiency, the LoG can be approximated by a difference of Gaussian (DoG) operator (Lowe, 1999).

In 1986, John Canny proposed a multi-stage approach for edge detection (Canny, 1986). This method, known as Canny filter or *optimal detector*, satisfies three main criteria:

- low error rate, corresponding to low probabilities of false edge detection and missing real edges;
- good localisation, meaning low distance of the detected edge from the true edges;
- minimal response, with only one detection per edge.

The optimisation function in Canny's detector corresponds to the sum of four terms, but can be approximated by the first derivative of a Gaussian. One advantage of Canny's method, when compared to the LoG, is that it avoids derivatives in uninteresting angles, making it less sensitive to noise effects.

Region growing

Region growing is a pixel-based procedure (Gonzalez and Woods, 2008; Rangayyan, 2004), where groups of pixels are grouped into regions according to their similarity. This approach starts with a *seed* pixel, and determines if the neighbouring pixels of the *seed* should be added to the region. This selection is based on a specified homogeneity criterion, which leads to different types of regions. Unlike thresholding, this procedure is heavily based on spatial intensity continuity.

A simplified region growing procedure is as follows:

1. Select a *seed* pixel;

2. Append neighbouring pixels of the region that have a specified property (gray level, color, etc.) that is similar to that of the *seed*;

3. Stop when there are no more neighbouring pixels that fulfil the criterion.

The selection of the *seed*, the measure of similarity and the method used to find the neighbouring pixels all influence the final result of the method.

When dealing with intensity-based segmentation, the simplest similarity measure is based on additive gray level tolerance. If the difference between the *seed* and a neighbouring pixel $\boldsymbol{I}(m,n)$ is within a tolerance level τ:

$$|\boldsymbol{I}(m,n) - \boldsymbol{seed}| \leq \tau, \tag{3.1}$$

then that pixel can be appended to the region.

One problem with the simple scheme described above is that all pixels are compared with the *seed*, even when they are not spatially close to it. This causes problems, especially when the *seed* corresponds to a noisy pixel. To solve this issue, candidate pixels can be compared to the mean gray level of the region already formed. This value is called the running mean μ_R. Another issue with this method is that the tolerance level may not be adequate to a particular *seed* value. In these cases, a multiplicative tolerance level τ' can be applied:

$$\frac{|\boldsymbol{I}(m,n) - \mu_R|}{\mu_R} \leq \tau'. \tag{3.2}$$

The main factor for an accurate region growing segmentation is the selection of the homogeneity criterion, which depends mostly on the problem at hand. Several criterions have been proposed, besides the simple gray-level tolerance, such as regional feature analysis (Chang and Li, 1994),

Bayesian probability modelling of images (LaValle and Hutchinson, 1995), Markov random fields (Won and Derin, 1992), and seed-controlled homogeneity competition (Adams and Bischof, 1994).

Image features

Image corners and other structured areas of an image can provide important descriptors for an image. These can be used to identify particular types of objects in an image, and help performing image comparison, even under changes in image scale, noise and illumination.

One of the most widely used methods for object recognition is the scale-invariant feature transform (SIFT) (Lowe, 2004; Szeliski, 2010). SIFT can robustly identify objects, even among clutter and under partial occlusion. It is invariant to translations, rotations and scaling transformations, and robust to affine transformations.

There are four main stages to generate the feature descriptors:

Scale-space extrema detection: Use a DoG function, over all scales, to find potential points of interest. This stage is invariable to scale and orientation.

Key-point localisation: Detect "stable" points, based on a model fit to determine exact location and scale. Noise sensitive points are removed at this stage.

Orientation assignment: Assign an orientation to each key-point, based on local image gradient directions. All further operations are conducted in a transformed space, providing invariance to orientation, scale, and location.

Key-point descriptor: Compute a position-dependent histogram of local gradient directions around the interest point. The vectorised version of this histogram corresponds to the key-point descriptor.

SIFT has also been extended to work with colour images (Van De Sande et al., 2010) and video data (Laptev and Lindeberg, 2006). In Publication III, SIFT was used to detect fMRI image landmarks, in order to facilitate image comparison.

Recently, other methods have been proposed for object recognition. Speeded Up Robust Features (SURF) was proposed in 2008 by Bay et al.. It is inspired in the concepts behind the SIFT descriptors, but is based on sums of 2D Haar wavelets, and uses integral images (Crow, 1984). SURF tends to have similar performance to the SIFT operator, but is computationally

faster. Dalal and Triggs (2005) proposed Histogram of Oriented Objects (HOG) as feature descriptors. This method counts occurrences of gradient orientation in localised portions of an image and uses overlapping local contrast normalisation to increase accuracy. Unlike SIFT, this method produces regional image descriptors, but is not rotationally invariant.

3.1.3 Artifact removal

Any kind of signal acquisition, in particular medical imaging, is subject to various types of noise and artifacts. They have different origins, such as physiological, *e.g.* cardiac activity, or technical, *e.g.* power-line interference. Since they cause a variety of problems for any algorithm, their removal is essential. There are typically two approaches for artifact removal. The first consists in rejecting any portion of data where they are present. This is rather crude, and is often hard to conduct since one desires to retain as much data as possible. The alternative is to reduce or cancel artifacts from the data, but this poses a significant challenge if one wants no distortion or loss of the desired information.

The most common artifact is noise, either random or structured (Rangayyan, 2004). Random noise usually refers to interferences arising from random processes, such as thermal noise in the equipment. It is characterised by a probability distribution function (*pdf*) of a random variable, often assumed to be Gaussian (Barrett and Swindell, 1996). Structured noise, on the other hand, is defined as non-random signal contributions (Bellon et al., 1986), such as power-line interference, and field inhomogeneities. Due to their non-random structure, and since their effects are typically known in advance, it is usually possible to minimise or eliminate their contribution to the signal.

Processing published images

In published fMRI results, researchers typically annotate the reported figures with useful information for the readers, *e.g.* patient identification and positioning markers or highlighting relevant activity points/regions. These can interfere in the analysis of such images. Since this type of artifacts are human made, they are heavily structured. Ideally, this allows for a systematic removal of such annotations, but, occasionally, they may be connected or even overlapping the brain image. When the artifacts are not connected with the fMRI image, *i.e.* there are no connecting pixels between the image and the artifact, a simple method of region detection,

based on size, and particular image geometry can be applied. To facilitate this type of detection, a conversion from color-scale to gray levels can be conducted. This conversion is typically used when the data is of high color depth, since lower depth results in smaller computational requirements.

There are many methods to convert from color space to gray levels. The most widely used method is conversion through luminance. Luminance, in RGB color models, can be calculated by the weighted sum:

$$Y = 0.2989R + 0.5870G + 0.1140B. \tag{3.3}$$

This conversion method is applied in standard color TV and video systems such as PAL, SECAM and NTSC.

Other typical artifacts present in fMRI published reports are grids or frames used to position the subject. Furthermore, authors tend to use identifying lines to emphasise particular points or regions of brain activity change. Since the shape of these analytically formed objects does not vary enormously, their detection is facilitated.

The best method to perform line detection is the classical Hough transform (Hart, 2009; Hough, 1959). Straight lines in image space (m,n) can be represented by the slope equation:

$$n = am + b, \tag{3.4}$$

where a is the incline and b the position where line intercepts the n axis. This results in a space, the Hough space, where lines are characterised by the parameter space (a,b). In order to improve computation and avoid the unbounded parameter problem, Duda and Hart (Duda and Hart, 1972) proposed the use of angle-radius parameters as:

$$\rho = m \cos\theta + n \sin\theta, \tag{3.5}$$

where θ is limited to $[0, 2\pi]$ and, ρ relates to the image size. The reference origin can be centred anywhere in the image. In 1981, Ballard (1981) proposed a generalised Hough Transform, where shapes that cannot be represented analytically in a two dimensional space can also be detected.

The main limitation of the Hough transform is that it is heavily dependent on data quality. In noisy images, where edges are hard to identify, the detection of lines is troublesome. Furthermore, the selection of the number of accumulator cells is of great importance, since the detection of lines is jeopardised if votes are spread throughout many bins.

3.2 MRI processing

In this section, several methods dedicated to prepare anatomical MRI data for analysis are detailed.

When dealing with raw fMRI, there are many types of preprocessing requirements, other than the ones detailed in this section, including slice time correction or temporal filtering. Since those are out of the scope of this thesis, due to the use of already processed images, the author suggests the book Huettel et al. (2008) as a good overview of such methods.

3.2.1   Image registration

In brain imaging, data is typically collected from different subjects, at different times, from different orientations of the brain. In order to conduct meaningful overall research on brain images, one should be able to somehow relate all those images. The main requirement for this is that each voxel corresponds to a fixed and unique location, regardless of the variations in acquision conditions. This is the goal of image registration.

Image registration is the process of aligning images so that corresponding voxels can be easily related (Hajnal et al., 2001). Registration algorithms are designed to establish spatial correspondence between points or regions within the images. This correspondence involves spatial transformations that relate information in one image to another, using an image similarity metric, and an optimisation algorithm. The most widely used packages for brain image registration are the Statistical Parametric Mapping (SPM, FIL Methods Group) toolbox and FSL's FLIRT (Greve and Fischl, 2009; Smith et al., 2004).

Usually, the registration process uses one image as reference, and places the coordinates of the target images in relation to that particular reference. The realignment of images involves first the estimation of a spatial, usually a 'rigid-body' affine, transformation. There are many methods to calculate the parameters of this transformation, but the most commonly used is based on maximising the mutual information between the transformed and the targeted images. After the spatial transformation is estimated, the images need to be re-sampled, to match the grid of the reference image. This involves an image interpolation step, which matches the reference and transformed image to the same coordinate space. This step is typically executed using tri-linear, sinc or spline interpolation (Friston, 2003; Hajnal et al., 2001).

3.2.2 Bias correction and noise reduction

One important preprocessing step in MRI images is to deal with noise and non-uniformities. Even when using strong magnetic fields, noise is always present in MR images, usually following a Rician distribution (Gudbjartsson and Patz, 1995), and can jeopardise image analysis. One simple way to deal with such noise would be to acquire several samples of the same images and perform a simple average over all images. But this would lead to a huge increase in acquisition time and cost. A more effective approach for noise reduction are image denoising techniques, including the use of wavelet-domain filtering methods (Nowak, 1999), as well as non-local means filters (Buades et al., 2005; Liu et al., 2010).

Another source of noise is related to poor coil uniformity, and to changes in the field due to the anatomy of patients. These lead to intensity non-uniformities, which can cause as much as 20% of variation in the grey level values of one tissue (Sled et al., 1998). This leads to significant problems when applying automatic segmentation techniques that assume tissue intensity homogeneity. The most widely used technique to solve MRI non-uniformities is Non-parametric Non-uniform intensity Normalisation ($N3$, Sled et al. (1998)). This method has the advantage of being independent of the type of MRI sequence acquired, and is insensitive to pathological data. It is based on an iterative approach that estimates both the multiplicative bias field and the distribution of the true tissue intensities.

3.2.3 Skull stripping

A specific type of preprocessing that is used when dealing with MRI images is skull-stripping. Since the skull and non-brain voxels are usually of no interest to neuroscientific studies, it is therefore useful to remove them from any further computation. This leads to a significant improvement in performance, and also avoids unbalanced results due to the huge amount of background voxels that are present in the images. Fig. 3.2 shows an example of skull-stripping. The rightmost image is the result of skull-stripping, where voxels corresponding to the skull have been masked out.

Skull-stripping methods can be usually grouped in three categories: manual methods, morphology-based techniques (Höhne and Hanson, 1992), and brain surface modelling (Smith, 2002). A manual segmentation usually leads to better results than most automated methods, since humans can take into account the complex information required for this task.

Nonetheless, the time and expertise requirements make it a nonviable solution for most studies. The most common class of skull-stripping techniques are the ones based on morphology (FIL Methods Group; Hahn and Peitgen, 2000; Shattuck et al.). The first step usually consists in a foreground/background segmentation using a simple intensity threshold. However, the detected foreground is almost always connected to non-brain tissues such as the eyeballs. To separate these non-brain tissues, morphological filtering is usually applied, by eroding links between brain and non-brain regions of the image. These morphology-based methods are mostly semi-automated, since the definition of thresholds needs to be done manually. The third type of skull-stripping methods are the ones resorting to deformable surface models (Dale et al., 1999; Popescu et al., 2012; Smith et al., 2004). A typical example of a surface model is a tessellated mesh of triangles. The model is fitted to the brain surface in the image, by iteratively deforming the surface from a starting position, until an optimal solution is found. Despite the higher computational power required, the results of this type of methods are usually robust and reach manual-segmentation levels of accuracy. Other methods exist that combine both morphological techniques with deformed models (Galdames et al., 2012; Iglesias et al., 2011; Ségonne et al., 2004), or methods based in non-local means MRI denoising (Coupé et al., 2011; Eskildsen et al., 2012).

(a) Original image.   (b) Skull-stripped image.
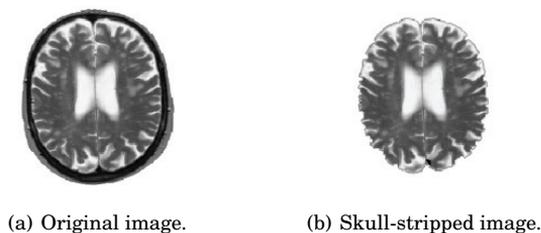
Figure 3.2. Example of the masking procedure. The frame on the left (a) shows the original MRI image, while (b) depicts the skull-striped brain.

4. Clustering and consistency

Data clustering in an essential methodology in statistical data analysis. Typical examples of research fields where it is commonly applied include data mining, big data analysis, image processing and bio-informatics.

A simple definition of clustering is to partition a set of data into different groups, *i.e.* clusters. These clusters share a similarity of some kind (Jain, 2010), and the definition of such similarity, usually assessed by a pre-defined distance measure, is at the core of any clustering method. The number of clusters can be selected *a priori*, or decided from the data itself through a number of model selection criteria, such as Akaike information criteria (Akaike, 1974) or minimum description length (Rissanen, 1978).

In this chapter, several clustering methods, and different distance measures are detailed. Clustering methods can be categorised in various ways. They can be divided according to their active principle, where most clustering methods are based either on competitive learning (Ahalt et al., 1989; Rumelhart and Zipser, 1985), or density estimation (Bishop, 2006; McLachlan and Basford, 1987). The former are more data-driven, while the latter build upon previously constructed models. Another type of data-driven approaches are the evolving clustering method (ECM) (Song and Kasabov, 2001).

Clustering results can also take the form of hard clustering, where each data element belongs to one and only one cluster; or soft clustering, where each element may belong to more than one cluster, with a varying degree of membership. Another form of categorising different clustering methods is according to their clustering results. These can be hierarchical, non-hierarchical, or graph-based.

In most clustering methods, the selection of parameters and the setting of initial conditions influences significantly the clustering outcome, and it is not always trivially solved. As shown in the last section of this chapter,

Table 4.1. Typical distance measures used to measure data similarity.

Measure	Distance	Measure	Distance
Euclidean	$\sqrt{(\boldsymbol{x} - \boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y})}$	Correlation	$1 - \dfrac{(\boldsymbol{x} - \overline{\boldsymbol{x}}) \cdot (\boldsymbol{y} - \overline{\boldsymbol{y}})}{\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|}$
Manhattan	$\sum\limits_{k=1}^{N} |x_k - y_k|$	Cosine	$1 - \dfrac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$
Chebychev	$\max_k |x_k - y_k|$	Jaccard[1]	$\dfrac{\sum_{k=1}^{N} x_k \wedge y_k}{\sum_{k=1}^{N} x_k \vee y_k}$
Hamming[1]	$\dfrac{1}{N} \sum\limits_{k=1}^{N} x_k \oplus y_k$	Dice[1]	$\dfrac{2 \sum_{k=1}^{N} x_k \wedge y_k}{\|\boldsymbol{x}\|_0^2 + \|\boldsymbol{y}\|_0^2}$

[1] typically used when $\boldsymbol{x}$ and $\boldsymbol{y}$ are binary vectors.

several techniques exist to alleviate this problem. One of those techniques is suggested in Publication I, where the extraction of consistent information from randomly initialised clustering methods is proposed.

4.1 Distance measures

In order to perform any type of clustering, the definition of the distance function $d(\boldsymbol{x}, \boldsymbol{y})$, between two data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is required. There are many examples of such distance functions, with some examples shown in Table. 4.1. The selection of the distance is mainly dependent on the task at hand, as well as the statistics of the data and the feature extraction methods used (Turney and Pantel, 2010). Most distances are based on the L_p norm (Deza and Deza, 2009; Dunford and Schwartz, 1958), such as Euclidean, Manhattan and Chebychev distances. These measures are commonly used throughout many fields. The cosine (Turney and Pantel, 2010) and the correlation (Galton, 1886; Stigler, 1989) distances measure the angle between two vectors. While correlation is based on the Pearson's correlation coefficient, the cosine distance can be derived from the Euclidean dot product. When dealing with binary vectors or sets, popular distances include the Hamming distance (Hamming, 1950), the Jaccard index (Jaccard, 1912) and the Dice coefficient (Dice, 1945), although the latter does not possess the triangle inequality property.

4.2 Competitive learning methods

In competitive learning methods, the premise is that the system can learn representations of data, through training of a set of parameters. After learning, the network is ready for generalisation.

Competitive learning is an unsupervised form of learning, akin in nature to Hebbian learning (Hebb, 1949). The basic procedure for a competitive learning method is (Rumelhart and Zipser, 1985):

1. Start with a set of randomly weighted network "units"

2. Limit the "strength" of each unit, typically done with a neighbouring function;

3. Allow the units to compete, usually through weight updating, for the right to respond to a given subset of inputs.

Therefore, as more data is fed to the units, each of them will converge to the "centre" of a particular subset of input data, responding with higher strength for input data "close" to that subset.

Typical competitive learning methods include K-means (Hartigan, 1975) and self-organizing maps (SOMs) (Kohonen, 2001). These methods are also examples of non-hierarchical clustering algorithms, where the result is a "flat" clustering of the input data.

4.2.1 *K*-means clustering

In K-means (Hartigan, 1975; Lloyd, 1982) the objective is to assign data points to K cluster centroids, which serve as cluster prototypes. The learning occurs by minimising the squared distances inside each cluster.

K-means is a particular case of the K-medoids algorithm, when the Euclidean distance is used, which results in spherical clusters. Other measures return different shapes for the clusters, such as cubical clusters for the Manhattan distance.

Formally, for a particular distance function $d(\boldsymbol{x}, \boldsymbol{y})$, the objective function H that K-medoids optimises for a set of data vectors $\boldsymbol{X} = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)\}$ is given by:

$$H = \sum_{n=1}^{N} \sum_{j=1}^{K} w_{nk} \, d(\boldsymbol{x}(n), \boldsymbol{m}_j) \, , \qquad (4.1)$$

where vector $\boldsymbol{m}_j$ represents the centroid of cluster j and w_{nk} is the membership of data point n to cluster k.

Since K-means can be derived from the expectation-maximization (EM) algorithm (Dempster et al., 1977), the optimisation of H consists of a two-step iterative optimisation:

Expectation or assignment step: Data points are assigned to the closest cluster centroid:

$$w_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j d(\boldsymbol{x}(n), \boldsymbol{m}_j) \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Maximisation or update step: The centroids are updated according to the data points that are assigned to them. This update can be done in batch mode:

$$\boldsymbol{m}_k = \frac{\sum_n w_{nk} \boldsymbol{x}(n)}{\sum_n w_{nk}} \,, \tag{4.3}$$

or, for faster convergence, in a sequential mode:

$$\boldsymbol{m}_k^{\text{new}} = \boldsymbol{m}_k^{\text{old}} + \lambda_n w_{nk}(\boldsymbol{x}(n) - \boldsymbol{m}_k^{\text{old}}) \,, \tag{4.4}$$

where λ_n is a learning parameter that decreases monotonically.

Despite its simplicity, K-means has several limitations. Since it is a heuristic algorithm, there is no guarantee of convergence to the global optimum, and the result is heavily dependent on the initialisation of the algorithm. Furthermore, K-means tends to find clusters of similar spatial extent, which is rarely the case for real data.

Several modifications and improvements to K-means have been proposed. In K-means, each data element is assigned to a single cluster, resulting in a hard clustering. A well-known variant of K-means is the fuzzy C-means method (Bezdek, 1981; Dunn, 1973), where each data element can be a member of multiple clusters, with a membership value assigned to each cluster (soft clustering). Other approaches include kernel K-means (Schölkopf et al., 1998) and X-means (Pelleg et al., 2000).

4.2.2 Self-Organizing Maps

Self-organizing maps (SOMs) can be described in an unsupervised artificial neural network framework (Kohonen, 2001). They are typically used for high-dimensional data visualisation. SOM may be formally described as a non-linear, ordered, smooth mapping of high-dimensional input data manifolds onto the elements of a regular, low-dimensional array. It performs a lattice projection that preserves similarity information in the input space, through competitive learning, and with an Hebbian learning

rule (MacKay, 2002). SOM has been shown to be comparable to the best methods of information visualisation, in terms of precision, but not in recall (Nybo et al., 2007).

As an input pattern is given as input to the network, the only map unit activated by it is the winning neuron, best representing the input pattern, called best matching unit (BMU). During the learning process, the neighbourhood of the winning neuron is also taken into account, by updating the location of the neighbour neurons. The learning process can be considered to take place in two distinct phases. During the ordering phase, the weight vectors organise themselves topologically. Afterwards, a fine-tuning of the feature map is carried out in order to provide an accurate statistical quantification of the input space (Haykin, 1999). After training, the result is a topographic map of the input patterns. In this map, the spatial locations of the neurons in the lattice are indicative of intrinsic statistical features contained in the input patterns. Thus, the continuous input space is mapped on a discrete set of prototype vectors.

SOM is a parametric method, where the dimensions and shape of the map need to be selected from the beginning. The map of prototype vectors is organised in a grid, composed of rectangular or hexagonal units. In case of a rectangular shape, map units have four neighbours, while units in a hexagonal map have six. Units at the edge of a map have a lower amount of neighbours, except if the opposite sides of the map are connected. This connection can be one sided only, resulting in a cylindrical shape, or through both sides, where the map shape is a toroid.

In SOM, the winning neuron c is found using a minimum-distance criterion, usually Euclidean distance, between the input vector $\boldsymbol{x}(n)$ and the weight vectors of the map:

$$c = c(\boldsymbol{x}(n), t) = \arg\min_k \|\boldsymbol{x}(n) - \boldsymbol{m}_k(t)\| \;, \tag{4.5}$$

where $\boldsymbol{m}_k(t)$ denotes the centroid vector of neuron k at iteration step t.

During the learning process, both the BMU and its neighbouring neurons are pulled closer to the input vector, by adjusting the centroid:

$$\boldsymbol{m}_k(t+1) = \boldsymbol{m}_k(t) + \alpha(t)h_{ck}(t)[\boldsymbol{x}(n) - \boldsymbol{m}_k(t)] \;, \tag{4.6}$$

where $h_{ck}(t)$ is the neighbourhood function of neuron k centred around the winning neuron c at iteration t. Typically, a Gaussian neighbourhood function is used:

$$h_{ck}(t) = \alpha(t)\exp^{(-\|\boldsymbol{m}_c(t)-\boldsymbol{m}_k(t)\|^2/2\sigma^2)} \;. \tag{4.7}$$

where $0 < \alpha(t) < 1$ is an exponential decaying factor, and $\sigma^2(t)$ the neighbourhood radius. Both of these terms decrease during the learning stage. Other neighbourhood functions include the bubble and the cut Gaussian.

The algorithm can also run in batch mode (Kohonen, 2001). In this case, the whole dataset is processed before the weight adaptation stage. Instead of (4.6), the update formula takes the following form:

$$\boldsymbol{m}_k(t+1) = \frac{\sum_n h_{ck}(t)\boldsymbol{x}(n)}{\sum_n h_{ck}(t)} \; . \tag{4.8}$$

This results in prototype vectors that are a weighted average of the data samples, with weights given by the neighbourhood function.

Aside from the specification of the SOM geometry, and before training, also the SOM prototypes need to initialised. This initialisation can be random,*e.g.* taken from random input vectors, or sampled from the two principal eigenvectors of the input data (linear initialisation). Although SOM is very robust regarding initialisation, using linear initialisation typically results in faster algorithmic convergence.

In order to use SOM as a clustering technique, interesting groups of map units, *i.e.* clusters, must be selected (Vesanto and Alhoniemi, 2000). Clustering the map units, instead of the original data, has the significant advantage that the set of prototypes can be significantly smaller than the original data set, resulting in a reduction of the computational cost. In clustering based on local minima, the centroids of the clusters are chosen to be the local minima of the SOM (Vellido et al., 1999; Vesanto and Alhoniemi, 2000). A map unit is a local minimum if its average distance to the neighbouring map units is smaller than any of the corresponding distances of its neighbours. The rest of the map units are then assigned to the cluster of the nearest centroid, in the Euclidean sense.

Because of its ease of use, as well as the ability to efficiently map high-dimensional data into a 2D lattice, SOM has been widely used in many applications, with over 10.000 published papers (Laaksonen and Honkela, 2011), most using the SOM toolbox (Alhoniemi et al.). Furthermore, many variants of SOM have also been proposed. Examples include TASOM (Shah-Hosseini and Safabakhsh, 2003), where adaptive learning rates and neighbourhood functions are employed, and Neural Gas (Martinetz et al., 1991), in which weights are adapted independently of any topological arrangement. Other researchers have also proposed a supervised version of SOM, where labelled data is used to train the map (Hagenbuchner and Tsoi, 2004; Kohonen, 2001).

4.3 Density estimation methods

Another set of clustering techniques is based in the estimation of under-lying probability density function (*pdfs*) (Silverman, 1986). These algorithms assume a generative model for the data, where the data is generated by a mixture of underlying probability distributions. Hence, each cluster can be described by one or more mixture components (McLachlan and Basford, 1987).

Most density estimation methods are based on the EM algorithm (Dempster et al., 1977). After estimating the parameters, *e.g.* through maximisation of the log-likelihood function, the model can then be used to predict subsequent clustering results. Recently, Bayesian-based approaches have also been developed, to improve the mixture models for data clustering, leading to, *e.g.* the latent Dirichlet allocation method (Blei et al., 2003).

When comparing statistical models to competitive learning, the choice of clustering criterion is less arbitrary and the approach includes rigorous statistical testing. Furthermore, density based methods have the advantage of being able to cope with arbitrarily shaped clusters. Yet, the definition of the underlying models is not always easy to set. Another typical limitation of these methods relates to the curse of dimensionality, which hinders their performance, when dealing with high-dimensional data.

4.3.1 Discriminative clustering

A particular method called discriminative clustering (DC) (Kaski et al., 2005; Sinkkonen and Kaski, 2002), is of particular relevance for this thesis. The main goal of DC is to cluster all data X, using the information within a subset of pairs $(x(n), \mathfrak{c}(x(n)))$, where $\mathfrak{c}(x(n))$ is the label of $x(n)$ (Sinkkonen and Kaski, 2002).

DC can be seen as a density estimation algorithm (Kaski et al., 2005), since it is based on a generative distributional clustering model. Nonetheless, a Hebb-like competitive learning stage is conducted, to estimate the clustering assignments. It performs semi-supervised clustering (Chapelle et al., 2006; Zhou, 2011; Zhu, 2006), where training is done in unlabelled data, with the help of a small amount of auxiliary data (labels).

The clusters obtained by DC are local in that a data element belongs to a cluster k, defined as a Voronoi region V_k, $x(n) \in V_k$, if the distance to the centroid defining that region, m_k, is less than or equal to its distance to any other centroid. Using the same notation as in Section 4.2, this can be

formally described as a binary membership of data points to a cluster:

$$
w_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\boldsymbol{x}(n) - \boldsymbol{m}_j(t)\| \\ 0 & \text{otherwise} \end{cases} \tag{4.9}
$$

The Voronoi regions are also homogeneous in terms of their labels, *i.e.* a cluster should not group elements with two different labels. DC assigns a distributional prototype

$$
\boldsymbol{\psi}_k = \mathbf{E}_{V_k}[p(\boldsymbol{\mathbb{c}}|\boldsymbol{x})] \, , \tag{4.10}
$$

to each V_k. The parameters of such prototype are fitted to cluster all the existing data, given the label.

The basic DC model (Sinkkonen and Kaski, 2002) is a piecewise-constant generative model for $\boldsymbol{\mathbb{c}}$, conditioned in $\boldsymbol{x}$, with the likelihood given by:

$$
L = \sum_k \sum_n w_{nk} p(\boldsymbol{\mathbb{c}}|\boldsymbol{x}(n))] \, , \tag{4.11}
$$

where the probability is given by a cluster-specific multinomial distribution.

Instead of predicting the classes by prototypes $\boldsymbol{\psi}_k$, DC partitions directly the primary data space, by searching for the set of all $\boldsymbol{m}_k$, $\boldsymbol{M}$, that maximises the marginalised posterior:

$$
\mathrm{MAP}_{DC} = p(\boldsymbol{M}|\boldsymbol{\mathbb{c}}, \boldsymbol{x}) = \int_{\boldsymbol{\psi}} p(\boldsymbol{M}, \boldsymbol{\psi}|\boldsymbol{\mathbb{c}}, \boldsymbol{x}) d\boldsymbol{\psi} \, . \tag{4.12}
$$

Kaski et al. (2005) suggest the use of:

$$
p(\boldsymbol{m}, \boldsymbol{\psi}) \propto p(\boldsymbol{\psi}) = \prod_k p(\boldsymbol{\psi}_k) \, , \tag{4.13}
$$

where the factors $p(\boldsymbol{\psi}_k) \propto \prod_i^C \psi_{ki}^{n_i^0 - 1}$ are Dirichlet priors with parameters n_i^0, and C is the total number of classes. For computational convenience, typically $n_i^0 = 1, \forall i$, as suggested in Kaski et al. (2005).

The logarithm of the objective function (4.12), after applying the Bayesian rule and marginalising for $\boldsymbol{\psi}$, can not be optimised by gradient-based approaches, since they would only be affected by samples at the border of the clusters, which would have zero probability. This is avoided by introducing a smoothing approach, as in Sinkkonen and Kaski (2002), consisting of a Gaussian transformation of the data sample memberships:

$$
w_{nk} = Z^{-1}(\boldsymbol{X}_k) \exp^{(-\|\boldsymbol{x}(n) - \boldsymbol{m}_k\|^2 / 2\sigma^2)} \, . \tag{4.14}
$$

The parameter σ controls the membership smoothness and Z normalises to unity the sum of all elements belonging to V_k.

In this DC formulation, categories may still over-fit to apparent dependencies in small data sets. One possible regularisation approach consists in favouring an equal number of data points per cluster. This reduces over-fitting and overcomes the "dead unit" problem, common in *K-means* with bad initialisation, where clusters can end up empty. The "equalised" objective function is then (Kaski et al., 2005):

$$\log(\mathbf{MAP}_{DC})^{EQ} \propto \sum_i \sum_k \log \Gamma \left(n_i^0 + n_{ik} \right)$$

$$-(1 + \lambda_{\mathbf{EQ}}) \sum_k \log \Gamma \left(N^0 + \sum_{\boldsymbol{x}} w_{nk} \right) . \qquad (4.15)$$

where $\lambda_{\mathbf{EQ}} >= 0$ and $N^0 = \sum_i n_i^0$. The smoothed number of samples, n_{ik}, is defined as $\sum_{\boldsymbol{c}(\boldsymbol{x})=i} w_{nk}$. Γ corresponds to the Gamma function.

The parameter $\lambda_{\mathbf{EQ}}$ governs the amount of regularisation applied. Higher values of $\lambda_{\mathbf{EQ}}$ favour solutions with an equal amount of samples per cluster. DC does not require a high percentage of labelled data, but prefers a good balance between the number of labels per class.

4.4 Graph-based methods

Since graph-based methods are not studied in this thesis, but are nonetheless worth mentioning, only a brief explanation about them is given in this section. For a good overview on this topic, the author suggests the book by Mirkin (2012).

Graph-based algorithms, sometimes referred to as spectral clustering methods, represent the data points as nodes in a graph (Donath and Hoffman, 1973; Fiedler, 1973; Jain, 2010). Graphs are structures formed by a set of nodes, or vertices, and their connecting edges (Gross et al., 2013; Schaeffer, 2007). An essential definition in graph clustering is *cut size*, *i.e.* the sum of the weights assigned to the edges connecting different clusters. The edge weights of a graph correspond to distances between data samples. Graph-based clustering is the task of grouping graph nodes into clusters, with the goal of minimising the *cut size* of the graphs.

Shi and Malik (2000) first proposed an efficient approximate graph-cut based clustering algorithm with cluster size constraints, called Normalised Cut. This was extended to a multi-class version (Yu and Shi, 2003) and to handle an arbitrary number of clusters (Meila and Shi, 2001). Other approaches to spectral clustering include the representation of data using the normalised eigenvectors of a kernel matrix (Ng et al., 2002), and the graph Laplacian (Belkin and Niyogi, 2001; Luxburg, 2007). Spec-

tral clustering has also been applied to the co-clustering problem (Dhillon, 2001) or to transductive learning (Joachims, 2003).

The main limitation of graph-based methods is that they often require the computation of a distance matrix between all samples. When dealing with high-sample data, this can be computationally intractable. Even recent spectral clustering methods such as ClusterReg (Soares et al., 2012), DaSpec (Shi et al., 2009), or diffusion maps (Coifman et al., 2005) suffer from this limitation.

4.5 Evolving clustering method

Yet another clustering method is the evolving clustering method (ECM) (Song and Kasabov, 2001). This algorithm performs a simple evolving, adaptive, maximum distance-based clustering, with a fast one-pass approach. ECM is an unsupervised learning method, that clusters an input stream of data, without pre-defining the number of clusters of such data. The maximum distance between the centre of a cluster created by ECM and the samples belonging to that cluster cannot be larger than a threshold value. This value is a preset clustering parameter, and is responsible for defining the number of possible evolved clusters. Such parameter can also be adjusted during the clustering process, according to some optimisation.

An extension to this method, called ECMc (Kasabov and Song, 2002), uses the results from ECM as initial values, and further optimises the clusters in an off-line mode with a predefined objective function. By alternating between the adaptive clustering with ECM and off-line optimisation of ECMc, it is possible to obtain improved clustering results, although only in batch mode.

4.6 Consistency analysis

Typical clustering algorithms have, at least, a few parameters. Furthermore, the algorithms may approach the solution via different paths, depending on the algorithm's initial conditions. The selection of such parameters, as well as the optimal definition of the initial conditions is not always trivial, and the clustering results may heavily depend on them.

Ensemble learning (Bishop, 2006; Granger and Ramanathan, 1984; Polikar, 2012), *i.e.* using multiple runs of learning algorithms, is a typi-

cal machine learning paradigm used to solve this problem. It includes methods such as bagging (Breiman, 1996), random tree forests (Breiman, 2001), boosting (Freund and Schapire, 1995) and cross-validation (Stone, 1974). Ensemble learning is typically used when one has access to some ground-truth, although some approaches have been used in an unsupervised framework (Fern and Lin, 2008; Gao et al., 2013; Hong et al., 2008).

As mentioned before in Section 4.3.1, another related approach is semi-supervised clustering. In such approaches, the burden is again placed on the label selection. While typically a small set is manually selected, this selection becomes again subjective and might not be available for all classes. To avoid the infection of *a priori* or subjective information, labels could be created using a automatic data-driven approach.

As detailed in Publication I, a method based on an analysis of consistency in clustering methods can be utilised to automatically estimate data labels. Such analysis is particularly relevant for any algorithm where parameter selection and initialisation may lead to different optimisation paths. A particular example is SOM, with its intrinsic stochastic nature and dependence on the initialisations.

The approach proposed in this dissertation is based on exploiting the variability of different clustering runs. With different random initialisations, cluster representations of data may vary. Therefore, data points sharing the same cluster in a given run may be assigned to different clusters in another. Consistent clusters are therefore defined as those comprising elements that are grouped together in a large number of runs.

An illustration of this concept is given in Fig. 4.1, using toy-data. On the leftmost example, the numbered circles are all in the same cluster. In the other two cases, circle 3 appears in different clusters from the other two. Consistency in clustering membership, such as 1 and 2, is indicative of high similarity between the two elements. Circle 3 has too much variability in its clustering grouping to be considered consistent.

The overall similarity of the clusters can be assessed via a variety of distance measures (Vinh et al., 2010). In this dissertation, two such measures are proposed. One is based on the elements included in each cluster (4.16), and a complementary one is based on the distributional information of the elements belonging to different clusters (4.18).

After running Z clustering runs on data X, with dimensions $[S \times N]$, the first measure assessing the distance between two clustering assignments
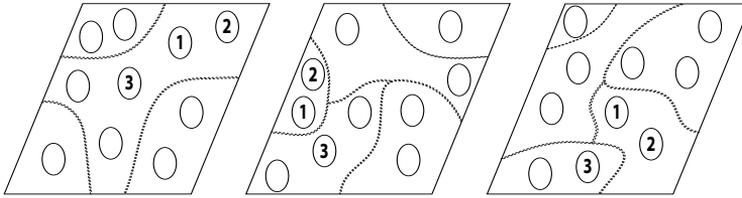
Figure 4.1. Clustering results with three different initialisations, using toy data. The circles represent different data points, while the dashed lines define cluster borders. Circles 1 and 2 always appear together for different runs, representing one same class. Circle 3 won't be picked to represent any such class due to the variability of its clustering. Figure taken from Publication I.

is defined as:

$$d^1(\boldsymbol{c}_i, \boldsymbol{c}_j) = \frac{1}{N} \frac{\sum_n^N \boldsymbol{c}_i(n) \wedge \boldsymbol{c}_j(n)}{\sum_n^N \boldsymbol{c}_i(n) \vee \boldsymbol{c}_j(n)} \ , \tag{4.16}$$

where

$$\boldsymbol{c}_z(n) = \begin{cases} 1 & \text{if element } n \text{ belongs to cluster } z \\ 0 & \text{otherwise} \end{cases} \tag{4.17}$$

$\wedge$ and $\vee$ are the AND and OR logic operators, respectively.

The main objective of d^1 is to group clusters according to the elements contained therein. In practice, this measure attempts to remove the redundancy obtained by running several times similar clustering methods. Clusters with $d^1(\boldsymbol{c}_i, \boldsymbol{c}_j) \geq D^1$ are merged together, where the value D^1 is mainly dependent on the parameter variation of the clustering methods used, and is typically over 0.8.

Once redundant clusters have been merged, the next stage needs to deal with clusters that have identical distributions. For example, when dealing with data with three well defined clusters, and selecting $K = 4$, the three true clusters will be split artificially. Since the methodology is unsupervised, there is no way to avoid these types of splits. Therefore, a measure based on distribution distance is required to reduce the number of split clusters.

In the case of Gaussian-like distributions, one distance measure can be defined as:

$$d^2(\boldsymbol{c}_i, \boldsymbol{c}_j) = \|\mathbf{E}_{\boldsymbol{c}_i}[\boldsymbol{x}(n)] - \mathbf{E}_{\boldsymbol{c}_j}[\boldsymbol{x}(n)]\| + \|\mathsf{Var}_{\boldsymbol{c}_i}[\boldsymbol{x}(n)] - \mathsf{Var}_{\boldsymbol{c}_j}[\boldsymbol{x}(n)]\| \ . \tag{4.18}$$

where $\mathbf{E}_{\boldsymbol{c}_p}[\boldsymbol{X}]$ and $\mathsf{Var}_{\boldsymbol{c}_p}[\boldsymbol{X}]$ represent the row-wise expected value and variance of all elements $\boldsymbol{x}(n)$ belonging to cluster p. Due to the data normalisation, $0 \leq \|\boldsymbol{x}(n)\| \leq 1$, the proposed distance, corresponding to the

sum of mean and variance differences, gives a higher weight to the former than the latter. The computation of mean and variances is computationally efficient, therefore the calculation of this distance avoids the complete derivation of the distribution.

Clusters with $d^2(\boldsymbol{c}_i, \boldsymbol{c}_j) \geq D^2$ are again merged. D^2 is usually ≥ 0.9, to avoid grouping together clusters that might represent different classes. This still allows for some variability in a cluster distribution, taking into account noise and other artifacts in the observed data.

In the case of non-Gaussian distributed data, d^2 might not be adequate. Nonetheless, the idea behind the use of a generic Gaussian-like measure rests on the central limit theorem. Since we often deal with a composition of identically distributed clusters, the resulting clusters should be approximately normally distributed.

Other distributional comparison methods can be adapted to replace (4.18), including the Kullback-Leibler (Kullback and Leibler, 1951) or the Jensen-Shannon (Manning and Schütze, 1999) divergences. The choice of d^2 depends mainly on the data set and computation requirements.

The pseudo-code for clustering consistency analysis proposed in this dissertation is given in Alg. 1. It describes how to use the proposed measures to retrieve data points consistently grouped together. Using distance measure d^1, the first stage clusters correspond to an intersection of only those cluster elements that are consistently grouped together. The second measure, d^2, allows for some variability in cluster distributions, and since it occurs only after the first, it does not decrease the clustering consistency. Note that, unlike typical clustering algorithms, not all data points will be assigned to a cluster, since not all will be consistent in their assignments. Furthermore, besides the clustering method used as basis for the consistency search, the dimensions of the data to analyse also heavily influence the number of runs needed to produce reliable results. Nonetheless, this procedure can be used as an estimation of representative points of the different classes present in the data.

Algorithm 1: Pseudo-code for consistency analysis of different clustering runs. D^1 and D^2 are two user selected thresholds, for distance d^1 and d^2, while Z is the number of different clustering runs.

Data: $\boldsymbol{X}$, $[S \times N]$, data matrix

Result: $\tilde{C}$, $[N \times L]$ clusters composed of data elements consistently
grouped together

Estimation of consistent elements:

    for $z = 1$ **to** Z **do**

        random initialisation of parameters

        Obtain clustering assignments $\boldsymbol{C}_z$

    end

    $\hat{C} = [\boldsymbol{C}_1 \ \ldots \ \boldsymbol{C}_Z]$

    while *any* $d^1(\hat{\boldsymbol{c}}_i, \hat{\boldsymbol{c}}_j) > D^1$ **do**

        $\check{\boldsymbol{c}}_p = \hat{\boldsymbol{c}}_i \wedge \hat{\boldsymbol{c}}_j$

    end

    $\check{C} = [\hat{\boldsymbol{c}}_1 \ \ldots \ \hat{\boldsymbol{c}}_P]$

    while *any* $d^2(\check{\boldsymbol{c}}_p, \check{\boldsymbol{c}}_q) > D^2$ **do**

        $\tilde{\boldsymbol{c}}_l = \check{\boldsymbol{c}}_p \vee \check{\boldsymbol{c}}_q$

    end

    $\tilde{C} = [\tilde{\boldsymbol{c}}_1 \ \ldots \ \tilde{\boldsymbol{c}}_L]$

---

5. Brain tissue segmentation

The field of medical imaging has been evolving rapidly in recent years, leading to a wide variety of tools to image the brain. In particular, the resolution increase in MRI-based techniques allows for better diagnosis and studies of the brain. Nonetheless, medical image segmentation remains a challenging problem, due to image complexity and absence of anatomical models that fully capture deformations in the brain structures (Ashburner and Klöppel, 2011; Kapur et al., 1996; Mortazavi et al., 2012; Sotiras et al., 2013). In addition, different MRI field strengths can affect segmentation results (West et al., 2013).

Manual segmentation methods are subjective and it is common to find disagreements and variability between independent experts (Warfield et al., 2004). Therefore, several semi-automated and automated methods in brain tissue segmentation have been developed to alleviate these problems. In this chapter, several tissue segmentation methods are briefly reviewed, with a clear focus on lesion detection. Also, because they are needed for a proper characterisation of each tissue, as well as the proposed segmentation, different tissue distributions and some typically used measures of segmentation accuracy are presented.

The methodology described in the Section 5.3 is one of the main contributions of this dissertation, and is detailed in Publication III. It consists of a self-supervised tissue segmentation approach, based on a discriminative strategy, and avoids the extensive use of *a priori* information, rendering it very versatile, and able to cope with different tissue types. Furthermore, it also returns tissue probabilities for each voxel, crucial for a good characterisation of the evolution of brain lesions, as well the characterisation of regions of transition between tissues. The last section in this chapter shows examples of the application of this methodology, using simulated data and real data, where ground-truth is available, similar to

the results detailed in Publication III. Publications IV and V were conducted with a more pragmatic objective in mind, and the segmentation was performed on a real data cohort. The contributed approach is used to show that changes in cognitive scores of patients with neurodegenerative disease are correlated with early-stage lesions. Furthermore, the development of WML was shown to be preceded by quantifiable changes in normal-appearing white matter. The last exemplary result comes from a tumour segmentation, where the self-supervised tissue segmentation method proposed in this dissertation was able to segment several regions of the tumour.

5.1 Background

Vannier et al. (1985) first proposed using multi-spectral data to the statistical analysis of MRI. Since then, major developments have happened in the brain segmentation field (Cuingnet et al., 2010; Heckemann et al., 2008; Iglesias et al., 2011; Klauschen et al., 2009; Valverde et al., 2014).

The simplest method to segment brain tissues is manual tracing, which is very subjective and time-consuming (Warfield et al., 2004). Computer-based methods allow for faster and more objective tissue segmentations. They are also more reliable, especially when dealing with pathological conditions (Mortazavi et al., 2012), but do not always use all the information available. Several of these methods still rely on manual tracing to create ground-truth data or labels for segmentation (Choi et al., 1991; Harvard Medical School; Shattuck et al., 2008; Wismüller et al., 2004). To avoid human intervention, other approaches have been developed to create subject-specific automatic labels based on clustering (Vovk et al., 2011), mixture models (Lee et al., 2009; Zhang et al., 2001) or atlas registration (Cocosco et al., 2003).

Segmentation methods tend to use several different basic approaches, rendering a proper categorisation hard. Nonetheless, they can be broadly grouped in three types: data-driven, statistical analysis and neural or fuzzy networks.

Data-driven methods were among the first methods developed to perform brain segmentation. Several rely on intensity thresholds, to detect the different tissues (Lim and Pfefferbaum, 1989; Schnack et al., 2001), especially when dealing with brain lesions (Anbeek et al., 2004; Khademi et al., 2012). These often require human intervention to set the thresh-

olds, leading to subjectivity and loss of generalisation. Other data-driven approaches include level-set methods (Fang et al., 2011), deformable models (Kapur et al., 1996), region growing (Hojjatoleslami et al., 1999) and hierarchical techniques (Pachai et al., 1998). While simple, the accuracy of these methods is limited, and they are very sensitive to noise.

Statistical methods are among the most widely used methodologies to perform brain segmentation techniques. Approaches based on the EM algorithm (Leemput et al., 1999, 2001; Manjón et al., 2010; Zhang et al., 2001) or non-parametric k-nearest neighbours (kNN) methods (Anbeek et al., 2004; de Boer et al., 2009; Vrooman et al., 2007; Wu et al., 2006) work by estimating the probability maps of the brain tissues. Other methodologies segment images through Parzen windows (Sajja et al., 2006), rule-based methods (García-Lorenzo et al., 2009) or support vector machines (SVMs) (Lee et al., 2005; Vovk et al., 2011). The main disadvantage in most statistical approaches is the assumption of normal distributions which, in the case of brain lesions, is seldom verified.

The third main category of brain segmentation methods are the fuzzy and the neural networks. These cover a wide range of techniques, from artifical neural networks (ANNs) (Alirezaie et al., 1998; Dyrby et al., 2008; Reddick et al., 1997; Zijdenbos et al., 2002) to fuzzy clustering (Admiraal-Behloul et al., 2005; Brandt et al., 1994; Lin et al., 2010; Nakamura and Fisher, 2009; Pham and Prince, 1998; Seghier et al., 2008; Shen et al., 2010). The main issue for these classifiers is the excessive training time, as well as the careful selection of training data. Also, as with intensity-based methods, noise presents many difficulties for segmentation.

Recent advances in MRI, namely whole brain coverage, high spatial resolution, and good contrast-to-noise ratios, have led to an increased usage of brain atlases, with standard prior tissue probabilities (Klauschen et al., 2009). The majority of modern brain segmentation methods register the images to segment to such atlases. In particular, most brain image segmentation software packages are atlas-based (BrainVoyager (Goebel et al., 2006); SPM2/SPM5 (Ashburner and Friston, 2005); FSL (Smith et al., 2004) and FreeSurfer (Fischl et al., 1999)). One major drawback of employing atlas priors happens when significant anatomical changes occur, *e.g.* due to brain lesions, regions with high degree of variability, in elderly people, with brain atrophy or in the case of infants. In such conditions, it is difficult to establish *a priori* the anatomy and number of tissues to be analysed (Cardoso et al., 2011; Pachai et al., 1998). To avoid strong

anatomical priors, several methods employ more general priors, based on intensity distributions or other statistical relations.

The study of lesions has received an increased focus in previous years, with particular emphasis to MS detection (Cruz-Barbosa and Vellido, 2011; Dyrby et al., 2008; Khademi et al., 2012; Sajja et al., 2006; Zijdenbos et al., 2002) and tumour segmentation (Cruz-Barbosa and Vellido, 2011; Dou et al., 2007; Gordillo et al., 2013; Prastawa et al., 2003). A common trend in modern methods of lesion detection is to consider lesion voxels as distribution outliers (García-Lorenzo et al., 2009; Leemput et al., 2001; Schmidt et al., 2012; Shiee et al., 2010). Nevertheless, since the quantity of MS voxels is much smaller, when compared to other classes, the estimation of its *pdf* may not be as accurate as for other classes. These outlier approaches rely on the correct identification of all healthy tissues, typically with the aid of probability atlases. Another recent imaging technique, diffusion tensor imaging (DTI) (Le Bihan et al., 2001; Merboldt et al., 1985), has also seen an increased usage, when dealing with brain lesions, due to its ability to reveal abnormalities in the white matter structure. Several studies have proposed to use DTI, or a combination of it and standard MRI images to perform brain tissue segmentation (Commowick et al., 2008; Sage et al., 2009; Trivedi et al., 2006; Zhan et al., 2009).

5.2 Tissue intensity distributions

When dealing with brain tissue segmentation, as well as in several machine learning approaches, a core concept is feature selection. Most tissue segmentation methods use pixel intensities of single or multi-spectral MRI as data features. There, images are represented as gray values, typically ranging from 1 to 255. Fig. 5.1 shows the intensity distributions of simulated MRI data, using 3 sequences. As shown in the figure, several tissue distributions overlap in the data space, even if their main peaks may be rather clearly identified.

While the data shown in Fig. 5.1 is simulated, it is based on a high-resolution, high-SNR MRI volume of a normal subject (Collins et al., 1998). Besides being anatomically realistic, it also models partial volume (PV) making it one of the best benchmark sets in MRI.

After basic preprocessing, *cf.* Chapter 3, all imaging sequences can be vectorised, each data element becoming a vector composed of the gray values of each voxel in all available sequences. This is the basic feature
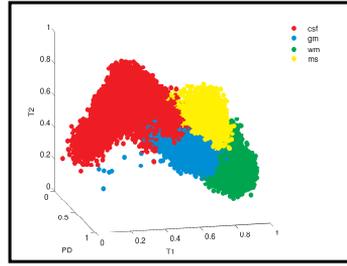
(a) 3D distribution of the gray levels of the different tissues.

(b) $\{T_1, T_2\}$ intensity distribution histogram.

(c) $\{T_1, PD\}$ intensity distribution histogram.

(d) $\{T_2, PD\}$ intensity distribution histogram.
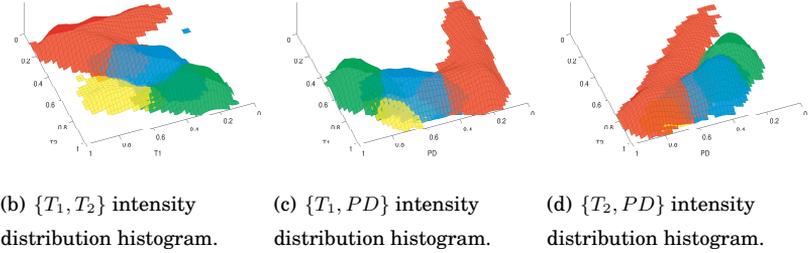
Figure 5.1. Intensity distribution images for simulated MRI data with lesion and 3% noise. (a) shows the three-dimensional plot of all the data points in the $\{PD, T_1, T_2\}$ space. (b)-(d) are projections of the 3D sequence space in a 2D histogram of intensities, where the z-axis is proportional to the number of elements projected to each (x,y) coordinate.

vector for brain imaging data. Other features can be added to increase spatial information, such as coordinates (Anbeek et al., 2004; Mayer and Greenspan, 2009), local region mean (Akselrod-Ballin et al., 2009) or textures describing neighbouring voxels (Theocharakis et al., 2008).

5.2.1 Partial Volume Effects

The intensity of a specific MRI voxel depends on the contents of the corresponding anatomical volume and the sequence used. Due to the finite spatial resolution of MRI, and since voxels may contain more than one tissue, this poses a major problem. When a voxel is composed by only a single tissue, the signal intensity will be characteristic of that tissue type. On the other hand, when one voxel represents more than one tissue type, the signal will be a combination of the contributions of the different tissues. This phenomenon is called partial volume effect (PVE) (Ángel González Ballester et al., 2002).

When the brain region to be analysed corresponds to tissue transitions, or degenerative lesions, the PVE is particularly evident (Mortazavi et al., 2012). Several segmentation methods ignore this effect, giving only a

hard-classification for all voxels, assigning one tissue for each voxel (Amato et al., 2003; Garcia-Sebastian et al., 2009).

Direct estimation of PVE is a difficult problem, without a unique solution. Several methods consider PVE a corrupting phenomenon, and try to correct it, while recent methodologies focus on its estimation, to obtain sub-voxel accuracy. This estimation can be done by solving a linear system, using the information provided by multi-spectral MRI (Brandt et al., 1994; Choi et al., 1991; Soltanian-Zadeh et al., 1993). Or, more commonly, the estimation is based in modelling the statistical distributions of tissues, and fit them to the imaging data (Khademi et al., 2012; Leemput et al., 2003; Manjón et al., 2010; Pham and Prince, 1998; Santago and Gage, 1993; Ángel González Ballester et al., 2002). Most such approaches also limit the number of possible tissues per voxel.

In Publication III, a statistical approach is also followed. In that case, the PVE is taken into account, by directly estimating the probabilities of tissue representation in each voxel. This approach requires different tissue distributions in a multi-spectral setting, but does not limit the number of possible tissue combinations. Despite some clear advantages, and because the method is based only on tissue intensities, the estimation accuracy decreases with noise. On the other hand, it does not impose any existing information on the data, rendering it a rather generic approach for tissue segmentation.

5.2.2 Contrast-enhancement through ICA

In order to improve the separability of different tissues, contrast-enhancing methods can be applied, the most typical of which being high-pass image filtering. In MRI, different tissues react differently to changes in the imaging parameters (Huettel et al., 2008). In addition, one can assume that the intensity of each voxel is a weighted combination of the contributions from the different tissues present in such voxel. If the number of sequences equals or surpasses the number of tissues represented in those images, then independent component analysis (ICA) can be applied to this data. ICA is a statistical and computational technique used to reveal hidden factors that underlie sets of random variables, measurements, or signals (Hyvärinen et al., 2001). This technique is probably the most widely used method to solve the blind source separation (BSS) (Cardoso, 1990; Jutten and Cardoso, 1995) problem, and is implemented in many algorithms, such as FastICA (Gävert et al.; Hyvärinen and Oja, 1997) and

Infomax (Amari et al., 1996; Bell and Sejnowski, 1995). Since a thorough overview of BSS and ICA methods is beyond the scope of this dissertation, only a brief explanation of ICA on the innovations will be detailed. The books by Comon and Jutten (2010) and Hyvärinen et al. (2001) provide more details on the aforementioned topics.

In the basic ICA model, the independent components are considered random variables with no time structure. The assumption of independence is the active principle behind the model identification. ICA estimates properly the sources by optimising a number of possible contrast functions, typically based on high-order statistics (Hyvärinen and Oja, 1997).

To relax the aforementioned assumption, considering time-dependent stochastic processes, instead of random variables, one can use the innovation of a process, *i.e.* the new information fed to the process at a given time point. This leads to ICA on the innovations (Hyvärinen, 1998), which can also be adapted to deal with spatial-dependent data. The innovation process $\tilde{s}(n)$ is defined as the error of the best prediction of a stochastic process $s(n)$. Thus, the innovation process is defined by:

$$\tilde{s}(n) = s(n) - \mathbf{E}\big[s(n)\big|n, s(n-1), s(n-2), \ldots\big] . \tag{5.1}$$

The innovation process, in the simplest case, can be reasonably approximated by the difference process:

$$\tilde{s}(n) \approx \Delta s(n) = s(n) - s(n-1) . \tag{5.2}$$

The independent component analysis on the innovations is based on the lemma (Hyvärinen, 1998) that states that if $x(n)$ and $s(n)$ follow $x(n) = As(n)$, then the innovation processes follows the same model:

$$\tilde{x}(n) = A\tilde{s}(n) . \tag{5.3}$$

Therefore, it is enough that $\tilde{s}(n)$ has independent components.

The benefit of applying ICA on the innovation process rather than on the original signals is that the innovations are usually more independent and more non-Gaussian than the original processes.

When applying ICA on innovations to brain imaging data, the objective is to increase the contrast between tissues, which allows for a better segmentation. The main restriction is, as mentioned before, that the number of independent components needs to equal, at most, the number of sequences used. Even when dealing with multi-spectral data, this is seldom the case. Therefore, the use of ICA is often limited,*cf.* Publication II.

Other methods to improve contrast, while reducing noise, include non-linear anisotropic (Gerig et al., 1992), and wavelet (Nowak, 1999) filtering. Unfortunately, these do not take into account multi-spectral data, rendering them impractical. When applied to each sequence separately, they would cause a difference of intensity values in the voxels of each tissue.

5.3   Self-supervised segmentation

One of the main contributions of this dissertation is the application of a self-supervised methodology to brain tissue segmentation. This approach is rooted on a discriminative clustering strategy, as shown in Section 4.3.1. The method is suited for the identification and segmentation of various brain tissues, including lesions such as MS and tumours. Fig. 5.2 illustrates an overall view of the methodology.
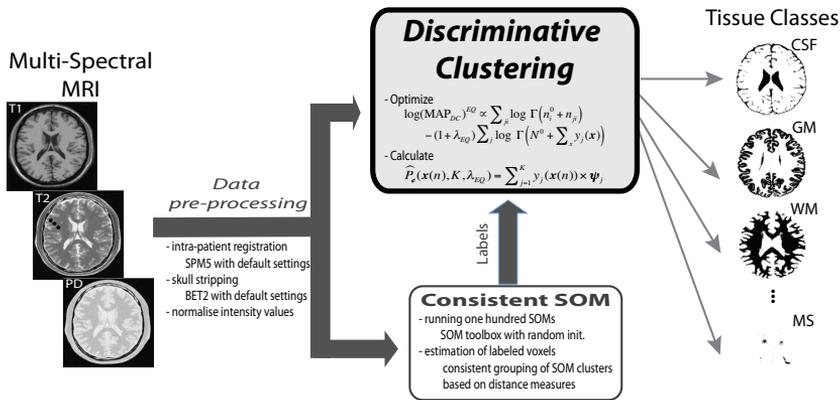
Figure 5.2. Flowchart depicting the proposed self-supervised brain segmentation method.

The discriminative clustering method, as described in Publication III, estimates the probability of membership of each voxel to all tissue classes, by maximising the difference between tissue intensities, while minimising the intensity variance within each tissue distribution. This probability of membership, or representation, is a direct measure of PVE. The main requirements of such approach is the existence of multi-spectral data, where the tissue distributions in the various sequences are as independent as possible, as well as the existence of a small set of labelled data elements.

The "self-supervision" stems from the fact that the labelled data required by DC is based on an analysis of clustering consistency of SOM (*cf.* Publication I and Section 4.6).

Clusters computed by DC are built based on labels for the different tis-

sues. Those clusters may represent a single tissue, or a weighted combination of various tissues. Such combination compensates for possible minor errors in the construction of labels, as well as allow for the existence of clusters corresponding to tissue transitions. In MRI, such transitions can arise from two different sources. The first caused by partial volume effects, due to which a voxel can contain more than one tissue. The other case occurs when a particular tissue starts losing some of its "standard" properties, typical in the case of progressive diseases.

As seen in Section 4.3.1, each voxel has a specific membership degree to the different clusters, obtained by DC. By combining the probability of a DC cluster to comprise a given tissue, with the membership of a voxel to all clusters, an estimation of tissue probability per voxel is obtained.

5.3.1 Segmentation robustness

DC, as many other clustering algorithms, is a parametrised method. Therefore it requires an appropriate selection of such parameters. Using an approach similar to the one used for consistent clustering analysis, see Section 4.6, DC can be run with several different initialisations and an average estimation calculated. Such approach produces a more accurate estimate of clustering outcomes, while providing some information on the stability of the solutions.

Since the methodology proposed is developed for tissue segmentation, where the number of tissues is never smaller than three (CSF, GM, WM), the minimum value for the number of clusters K is also set to three. Furthermore, when dealing with the segmentation of healthy tissues and lesions, a maximum of five clusters can be set. Higher values would lead to an over-segmentation. These values can be adapted to the study at hand, allowing for a general segmentation, whether or not lesion is present.

When dealing with brain lesions, the number of lesion voxels is typically much smaller than that of other tissues, in particular white matter. The parameter λ_{EQ} tries to compensate this disparity. Therefore, different values for this parameter, ranging from 0 to 1, can also be used. To avoid such imbalances, another approach is to use the same number of training voxels for each tissue, at the risk of having only a few samples per tissue.

To prune out possible misuses of existing label information, runs resulting in clusters merging two or more different labels are discarded. The results of the remaining runs can then be averaged out to obtain the final tissue segmentation result.

The pseudo-code version of algorithm is detailed in Algorithm 2.

Algorithm 2: Pseudo-code of discriminative clustering.

Discriminative clustering:

> **for** $K = 3$ **to** 5 **do**
>
> > create K random cluster prototypes
> >
> > **for** $\lambda_{\mathrm{EQ}} = 0.4$ **to** 1 **do**
> >
> > > optimise (4.11)
> > >
> > > calculate (4.12)
> >
> > **end**
> >
> > $\hat{P}_{\varrho}\left(\boldsymbol{x}(n), K\right) = \mathbf{E}_{\lambda_{\mathrm{EQ}}}[\hat{P}_{\varrho}\left(\boldsymbol{x}(n), K, \lambda_{\mathrm{EQ}}\right)]$
>
> **end**
>
> – Select only results for which the labels were not mixed in different clusters
>
> $\hat{P}_{\varrho}\left(\boldsymbol{x}(n)\right) = \mathbf{E}_{K}[\hat{P}_{\varrho}\left(\boldsymbol{x}(n), K\right)]$

5.4 Error measures

When dealing with segmentation techniques, a key concept is error calculation. Due to the complexity of the human brain anatomy, ground-truth or "gold standards" are not always available. Even when using manually delineated data, different independent experts may not always agree (Warfield et al., 2004), leading to inter-expert variability. In these cases, one can apply methods such as STAPLE (Commowick and Warfield, 2010; Warfield et al., 2004), where several expert segmentations are used to estimate a reference standard.

Assuming ground-truth or "gold standards" are available, several measures can be used. Two of the most typically used distances in tissue segmentation are the Dice coefficient (κ_d) (Dice, 1945):

$$\kappa_d = \frac{2TP}{2TP + FP + FN} \, , \tag{5.4}$$

and the Jaccard index (κ_j) Jaccard (1912)), *cf.* Section 4.1:

$$\kappa_j = \frac{TP}{TP + FP + FN} \, , \tag{5.5}$$

where the parameters are the number of true-positive (*TP*), true-negative (*TN*), false-positive (*FP*) and false-negative (*FN*) pixels, or voxels. Both

Dice and the Jaccard measures are limited between 0, for complete mis-classifications, and 1 for perfect performance.

These two measures are often associated with the sensitivity (η_{svt}):

$$\eta_{svt} = \frac{TP}{TP + FN} \ . \tag{5.6}$$

and the specificity (η_{spf}) coefficients:

$$\eta_{spf} = \frac{TN}{TN + FP} \ . \tag{5.7}$$

Sensitivity and specificity characterise how many data points are correctly segmented and how many of the ones outside the ROI are correctly excluded, respectively.

One problem with specificity is that it involves the number of true negatives, hence becoming directly dependent on the data set size. To avoid this problem, sensibility (η_{sbl}) can be used instead:

$$\eta_{sbl} = 1 - \frac{FP}{TP + FN} \ . \tag{5.8}$$

While both Dice and Jaccard distances provide a good summarising index of an overall classification, their sensitivity to high values of classification is rather low. Conformity (κ_c, Chang et al. (2009)) is a more sensitive and rigorous measure, allowing for a better discrimination of small variability in segmented images. It is defined as:

$$\kappa_c = 1 - \frac{FP + FN}{TP} \ , \text{ if } TP > 0. \tag{5.9}$$

Despite their advantages, both conformity and sensibility can take negatives values, which are hard to interpret. In particular, when the segmentation and the ground-truth do not have any overlap, κ_c takes the value of negative infinity, while η_{sbl} is negative for classification regions larger than the reference.

All the aforementioned measures can be adapted for fuzzy classification (Cardoso et al., 2011; Crum et al., 2006; Shattuck et al., 2001), but their definition is not particularly suited for those cases. In such circumstances, the rms error can be employed:

$$rms = \sqrt{\frac{1}{N} \sum_{\cent} (\hat{P}_{\cent} \boldsymbol{x}(n) - P_{\cent} \boldsymbol{x}(n))^2} \ , \tag{5.10}$$

where $\hat{P}_{\cent} \, \boldsymbol{x}(n)$ and $P_{\cent} \, (n)$ are the estimated and true probabilities of $\boldsymbol{x}(n)$ to belong to class $\cent$, respectively. This measure can be computed considering that each data point is assigned to only one class, by thresholding (hard classification, rms^h), or that each voxel can have contributions from several classes (soft classification, rms^s).

5.5 Experimental results

In this section, are shown some examples of results obtained when applying the self-supervised segmentation methodology. For more details on these and other results, the reader is directed to Publications II to V.

5.5.1 Simulated data

The performance of the proposed self-supervised segmentation method was evaluated on simulated data from the Brainweb site[1], where ground-truth is available.

This data set included brain both healthy and with MS lesion, all with multi-spectral images, comprising PD, T_1- and T_2-weighted sequences.
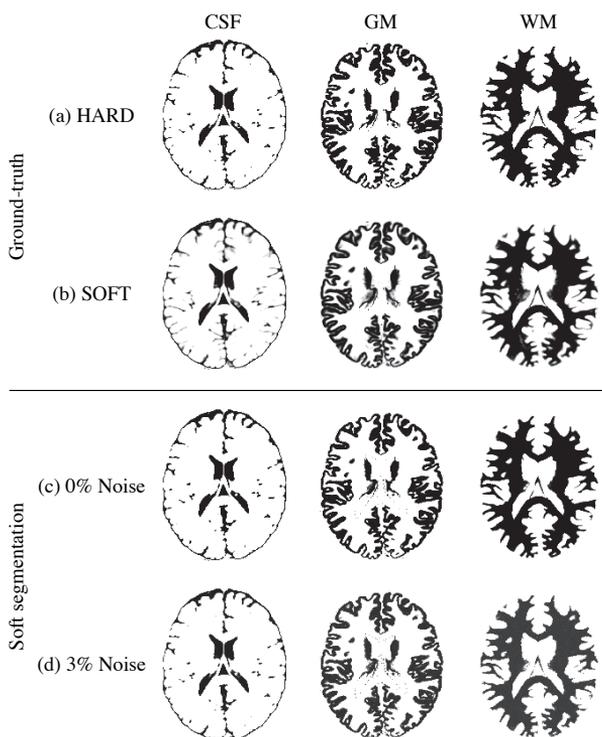
Figure 5.3. Results for the segmentation using simulated healthy brain data, with 0% and 3% noise levels and no inhomogeneity. The first two rows show the ground truth, both with hard (a) and soft (b) class assignments. The soft segmentation results for noiseless and noisy data are displayed in rows (c) and (d), respectively. From left to right, the classes shown are CSF, GM and WM, respectively. Adapted from Publication III.

In Fig. 5.3 are shown the results obtained when segmenting simulated healthy brains. Similar results for simulated brains with MS lesion are

---

[1]http://brainweb.bic.mni.mcgill.ca/brainweb/, visited 06/2014

shown in Fig. 5.4. In both figures, sub-figures (a) and (b) show the ground truth, in hard and soft perspectives, respectively. The proposed method's segmentation is depicted in sub-figures (c) and (d) for the noiseless and 3% noise data. Visual inspection shows a clear agreement between the soft ground-truth segmentation and the estimated soft segmentation, for both noiseless and noisy data. One can also see, especially in the borders of the WM, that noise has a slight detrimental effect on the results, but that effect is quite mild.
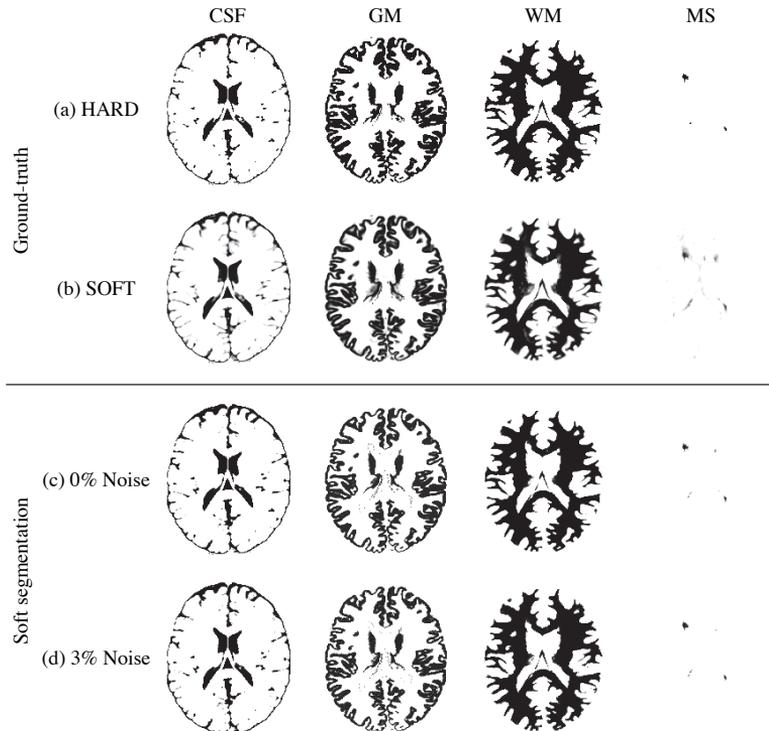
Figure 5.4. Similar results as the one presented in Fig. 5.3, but including a class of MS lesion, depicted in the right-most column.

To quantitatively evaluate the results of the proposed method, since ground-truth is available, one can compute a series of error measures. Such measures are shown in Table. 5.1, for healthy brains, and brains with MS lesion, both with 3% noise level. In the case of the simulated healthy brains, κ_d coefficients over 0.96 were reached for all tissues. The high values of sensitivity and specificity show that the method is very accurate, even in the presence of noise. For the data with simulated brain lesions, the results are again quite good, with even the lesion tissue obtaining a κ_d score of 0.89. Note that, in this simulated data, the wrongly classified voxels obtained correspond mostly to partial volume voxels, with

Table 5.1. Quantitative evaluation of the segmentation results for simulated data for both healthy and brains with lesion, with 3% noise and no inhomogeneity.

		κ_j	κ_d	κ_c	η_{svt}	η_{spf}	η_{sbl}	$\left(\frac{rms^h - rms^s}{rms^s}\right)$ (%)
Healthy	CSF	0.94	0.97	0.94	0.97	0.99	0.97	48.4
	GM	0.93	0.96	0.92	0.96	0.97	0.96	68.6
	WM	0.94	0.97	0.94	0.96	0.99	0.98	41.0
Lesion	CSF	0.86	0.93	0.84	1.00	0.98	0.85	55.9
	GM	0.89	0.94	0.86	0.90	0.99	0.99	55.0
	WM	0.95	0.97	0.94	0.98	0.97	0.96	55.5
	MS	0.79	0.89	0.74	0.87	1.00	0.91	29.7

a particular high level of noise. Of remarkable importance is the ability of the proposed segmentation method to handle the PVE, shown in the values of rms improvement, comparing a hard classification to a soft one. This improvement is highlighted in the last column of Table. 5.1.

5.5.2 Grand challenge

Fig. 5.5 shows the results obtained for one patient of the GrandChallenge dataset[2] (Styner et al., 2008), where the objective was to identify voxels corresponding to MS, in a population of 54 subjects.
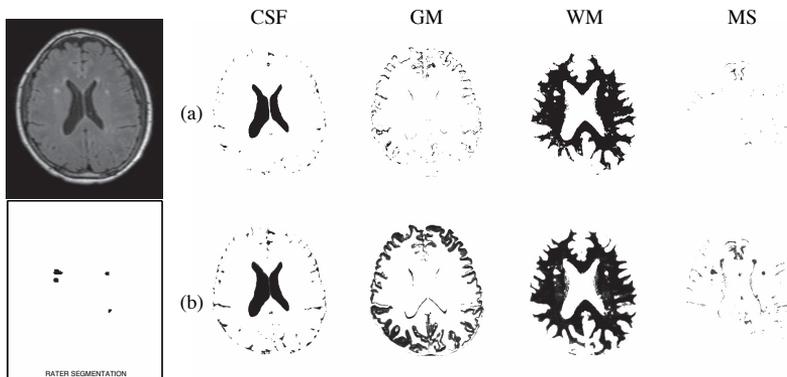
Figure 5.5. Segmentation results for a patient from the Grand Challenge data set. The labelled voxels are shown in (a) and the soft segmentation obtained by DC is displayed in (b). The classes exhibited are CSF, GM, WM and MS, from left to right respectively. The lesion segmentation, provided by an expert rater, is shown on the left bottom frame. Adapted from Publication III.

In this study, ground-truth was given by two expert raters. Although the segmentation is in-line with the one from the rater depicted in the figure, there are a few differences. As an example, the proposed method suggests

---

[2] http://www.ia.unc.edu/MSseg/results_table.php, visited 04/2014
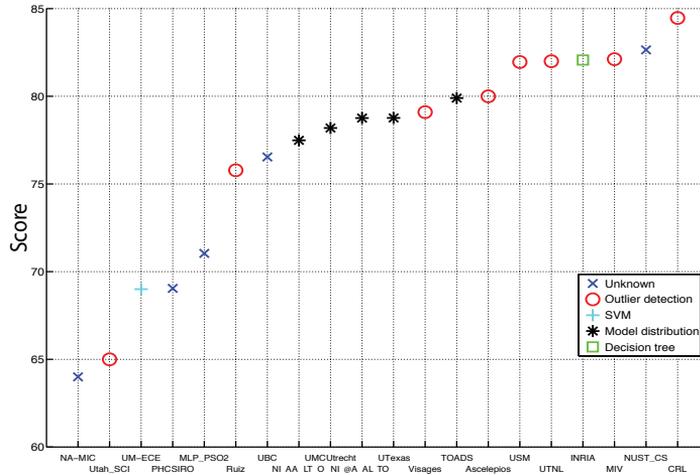
Figure 5.6. Results obtained by several methods, when applied to the test data set of the MS Lesion GrandChallenge.

a set of lesion voxels that are not considered by the rater. Nevertheless, it is important to note that those have the same multi-spectral gray level behaviour as the lesion. In any case, by visual inspection, and taking into account only tissue intensities, the results are quite reasonable, despite some artefactual lesions found in the frontal area of the brain.

Fig. 5.6 shows a plot with the segmentation score values of different methods on the GrandChallenge dataset.

The proposed method scores 77.5%, without resorting to any anatomical prior. Using a mild, and rather general prior, the score improves to 78.6%. While these results keep the method in the middle of the table, it is worth mentioning that these scores are almost the best from all model distributions methods. Other approaches tend to rely on very specific priors, rendering them quite specific to the study of MS. Furthermore, the segmentation provided by the two expert raters is quite different. Calculating the Dice scores between the manual segmentations for all training patient, results in an agreement of only $27.13\%\pm2.1\%$. This value is bigger than the difference between the scores of the best method and the worst, making a method comparison hard to evaluate.

5.5.3 LADIS data

The proposed segmentation was also applied to data from the Leukoaraiosis and Disability (LADIS) study (LADIS Study Group, 2011), with multi-spectral MRI images from patients that suffer from mild to severe WML. The extent of hyperintensities on white matter regions was evaluated on

FLAIR images only, both by visual rating scales and a semi-automated volumetric analysis, as detailed in van Straaten et al. (2006). The lesions classified through the aforementioned method are referred in this dissertation as conventionally estimated WML.
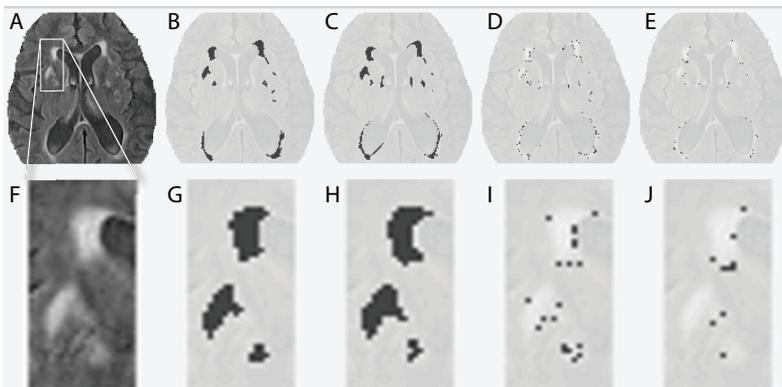
Figure 5.7. Lesion probabilities for a patient of leukoaraiosis and disability study (LADIS) study. (A) FLAIR image for a given subject suffering from WML, at a middle level height. (B) conventionally estimated white matter lesion. (C-E) estimated white matter lesion, using the proposed segmentation algorithm, for full, intermediate and small proportion of lesion. (F-J) present similar images for the zoomed portion depicted by the white box in A. Figure taken from Publication IV.

Fig. 5.7 shows one original FLAIR sequence of a subject in the LADIS study suffering from WML, both the full image slice (A) and a zoomed portion of it (F), together with the rater's segmentation (B and G) and the lesion segmentation results (C-E and H-J). The proposed method classified each lesion voxel as having a small (E and J), intermediate (D and I), or high (C and H) proportion of lesion tissue. The evolution around the foci of lesion, from fully blown, in the centre, to the intermediate stage and small proportion of lesion, at the edges, can be seen in frames H-J. One can see a small difference between the rater's segmentation and the full+intermediate segmentation results, *i.e.* , the voxels that would be considered lesions through a hard classification approach. Yet, a careful analysis of the image intensities in frame F suggests an overestimation of the lesion in B. A major advantage of this method is shown in frames E and J, where the voxels shown are indicative of possible locations of future lesions, and only present in a soft classification approach.

The data from the LADIS study included MRI sequences taken at two different time points, t0 and t3, separated by 3 years. Using the proposed method to segment the images at both time points, it is possible to show how the detection of early changes in the brain suggest possible future
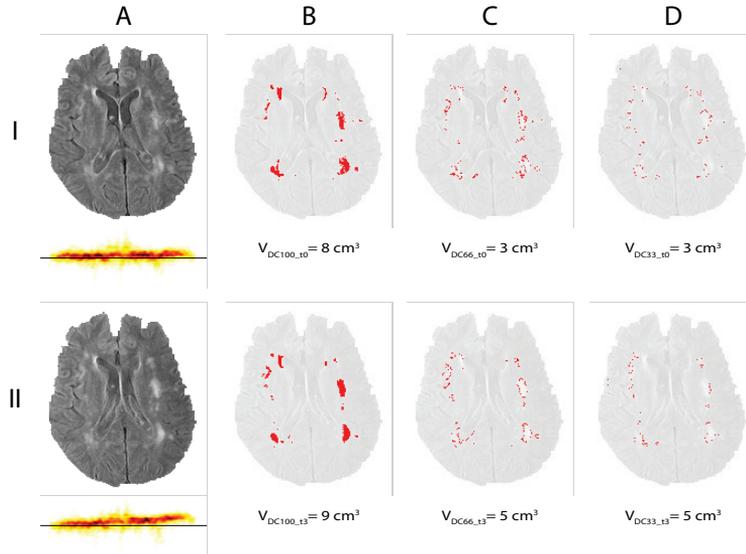
Figure 5.8. Follow-up results obtained for one patient of the LADIS study. Rows (I) and (II) show the images related to the first and second MRI acquisition times, respectively. Column (A) corresponds to FLAIR images, with the heat-map intensity distribution shown below it. Columns (B-D) show the estimated white matter lesions, for full, intermediate and small probability, respectively. Below the pictures is the total volume detected for each lesion probability. Figure taken from Publication V.

locations of lesion, and how those lesions evolve with time. An example of such study is shown in Fig. 5.8. The top row shows, from left to right, the FLAIR sequence (A), the voxels with complete (B), intermediate (C) and small (D) probability of being lesion, all at time t0, whereas the bottom row shows the same results for t3. Below each FLAIR sequence, a heat-map of gray intensity distribution is shown. The heat-maps display variations between the gray level distributions in the two time instances. In particular, at t3, it clearly shows a non-homogeneous gradient. The volumetric estimates for each of the lesion levels are shown below the corresponding images. The most important finding is that most voxels with a small probability of being lesion at t0 have developed into full-blown lesion at t3. Furthermore, the lesion has spread according to the areas suggested at t0, by all three estimated volume areas.

5.5.4 Tumour segmentation

The methodology in this dissertation can also be applied to tumour segmentation. After an initial segmentation, the tumour region was focused in a similar approach as detailed in Publication III. The results of the

segmentation of this ROI can be seen in Fig. 5.9. Frames A and B show, respectively, the sequences T_1c and FLAIR, at a height where the tumour is particularly evident. In frame C is the hard segmentation proposed by DC, and frames D to G show the soft segmentation for the non-enhancing, necrotic, enhancing and edema regions of the tumour, respectively.
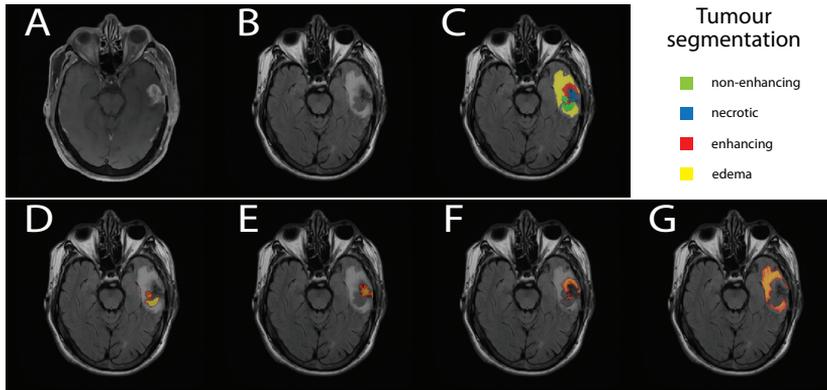
Figure 5.9. Tumour segmentation using the proposed methodology. Frames A and B show sequences T_1c and FLAIR, respectively, while C depicts the hard segmentation obtained. Frames D to G show soft segmentations for different regions of the tumour: non-enhancing, necrotic, enhancing and edema, in that order.

Without any specific prior or changes to the methodology, it is clear that the approach can be used to segment different types of brain lesion.

6. Data mining

The recent development of technology has paved the way for an exponential increase in the availability of data from many different types. We can take, as one of the most recent examples, the experiments conducted in the Large Hadron Collider, which produce around 15 Peta-bytes of data per year. This poses tremendous challenges both in storing such data, as well as in its analysis.

The field of data mining deals with the problem of filtering and obtaining relevant content from various sources, such as news feeds, books, websites, photo collections and other databases. The automatic extraction and analysis of such information is often a hard task, and constitutes a major subject in information retrieval and data mining.

Several data mining methods are text-based, where information is extracted using pattern recognition techniques. Despite the advances in natural language processing (NLP), using text is often limiting, since it does not make use of the intrinsic multi-modal nature of many data sets, which provides ample scope for mining information at various levels.

Besides text, information can also be encoded, for example, through visual content or multi-modal databases. Concomitantly, several approaches have been proposed to handle visual information retrieval (Gupta and Jain, 1997; Lew et al., 2006; Müller et al., 2010) and database analysis (Derrfuss and Mar, 2009; Günay et al., 2009; Laird et al., 2005, 2009). Even when textual annotations are present, these content-based methods can improve overall retrieval accuracy.

One particular application of data mining methods relates to the extraction of information from neuroscientific publications and other reports. Since most researchers do not have access to the original data reported in published studies, articles typically constitute their main source of information about current research findings. One particular topic of inter-

est are publications of fMRI studies, where crucial information is present both in textual and visual formats.

This chapter gives a very brief overview of current approaches in data mining, starting from the most typical textual perspective, and continuing to visual information retrieval. For more details on these topics, the author advises the books by Kao and Poteet (2007) and Baeza-Yates et al. (1999). Following the research done in Publications VI and VII, Publication VIII presents a document mining method of multi-modal information retrieval and analysis, where each document represents one fMRI report. That approach, another main contribution present in this dissertation, is detailed in the third section of this chapter. Finally, some of the results obtained using the document mining methodology proposed are presented in the last section.

6.1 Natural language processing and text mining

Text mining is the process of extracting information from textual data (Kao and Poteet, 2007; Tan et al., 1999; Weiss et al., 2010). It encompasses everything from information retrieval to text classification, including semantic analysis. A closely related research field is NLP (Bates, 1995; Indurkhya and Damerau, 2012), where the main goal is the extraction of meaning from text. By applying NLP techniques towards text mining, it is possible to perform information retrieval and text classification in meaningful ways.

In order to extract interesting content from written text, the first NLP methods made use of various knowledge representations, such as lexicons of words, grammar rules and thesaurus of synonyms or abbreviations (Goddard, 2011; Miller, 1995; Navigli and Ponzetto, 2012). Modern NLP methods (Indurkhya and Damerau, 2012; Manning and Schütze, 1999), on the other hand, use statistical machine learning to learn such rules. Machine learning approaches offer many advantages over rule-based algorithms, such as scalability and robustness to unfamiliar inputs. Nonetheless, thesaurus and grammar rules are often useful in guiding the learning stages of many algorithms.

A document[1] is a text document, in a natural language, with a clear internal structure. Each document tends to focus on a particular topic or

---

[1]In this thesis, a document may contain multi-modal information, such as images. Hence, a scientific publication is also considered a document.

idea, but it can also be a discussion on several topics. To process text as a whole, typical text mining approaches treat a document or a piece of text as an unstructured bag of words (Salton et al., 1975), *i.e.* only the count of words that are present in the text is relevant, discarding information such as positions with respect to each other, or other structural relationships. This model is widely used, and can be represented in a document-term matrix format. A document-term matrix is a representation of the occurrences of all words present in a document collection.

Instead of using only simple words, one can also use *n*-grams, which take into account word context. In *n*-grams notation, words correspond to unigrams (1-grams), two consecutive words represent bigrams (2-grams) and a sequence of *n* consecutive words is an *n*-gram. This notation is widely used in statistical NLP, allowing for the creation of extremely effective models of language data. A limitation of this model is that term frequencies need to be high enough for a proper statistical analysis. Nonetheless, with the increased availability of data, this problem is alleviated.

Table. 6.1 shows a toy-example of a dictionary composed of *n*-grams. The term frequency is displayed for each document. The first document seems to deal mostly with anatomical regions, while the second deals with Schizophrenia. Doc. III probably mentions changes related to activity in the cingulate gyrus in Schizophrenic subjects. Both "a" and "the" are very common words and do not contribute to understanding the text.

Table 6.1. Example of a document-term matrix with three documents and 5 terms.

	a	cingulate gyrus	the	Grey matter	Schizophrenia	$\cdots$
Doc. I	147	61	158	49	3	
Doc. II	104	3	125	12	53	$\cdots$
Doc. III	162	30	133	1	24	

6.1.1 Term weighting

As shown in Table. 6.1, some words tend to have high frequency values throughout all documents. This causes problems, for instance, when applying clustering techniques to this type of data. On the other end of the spectrum, also very rare words can hinder any statistical analysis of the word collection. In order to remove such words, two approaches are commonly used. The first is to use a list of *stop-words* as a reference for all

words to be ignored in the analysis. Instead of such lists, or in addition to them, one can give proper relevance to important terms through weighting schemes (Manning and Schütze, 1999; Salton and Buckley, 1988). The weights can be local, where terms are analysed per document, or global, where the term weights depend on the whole collection. Various such weighting methods are listed in Table. 6.2.

The most widely used local weighting approach is term frequency (*tf*) or, to avoid highly skewed distributions, its logarithmic version. This measure counts the number of occurrences of a word w_n in a document d_j. Global measures are evaluated on the whole document collection. While document frequency counts the number of documents that contain the term w_n, inverse document frequency (*idf*) gives less weight to frequent terms in the document corpus. To combine global and local weightings, a few different measures can be used. A widely used approach is *tf-idf* (Jones, 1972; Salton and Buckley, 1988), which combines the logarithmic versions of both *tf* and *idf*. Other combination measures, Okapi BM25 (Robertson et al., 1996) and pointwise mutual information (PMI) (Church and Hanks, 1990), are more focused on document/topic approaches, while term frequency-inverse document frequency (*tf-idf*) is better suited to give more weight to terms relevant in each document.

6.1.2 Stemming

Another important aspect in information retrieval is the process of stemming. Stemming reduces terms to their "stem" or "root" form, *e.g.* "running>run". While this corresponds typically to a shorter form of a written word, sometimes the stem is not identical to the morphological word root, such as "went>go". By treating words with the same stem as synonyms, also known as conflation, text mining algorithms reduce the dimensionality of dictionaries, allowing for better analysis of the textual data.

Several stemming algorithms have been proposed, from basic look-up tables, through rule-based methods to stochastic algorithms. The most widely used stemmer was developed by Porter (1980). The major limitation of most stemming algorithms is the reliance on language based rules, rendering them very language-specific. On the other hand, stochastic algorithms use training sets to identify stems, and are heavily dependent on the chosen training data. Despite their limitations, stemming algorithms remain an essential part of most text information retrieval methods.

Table 6.2. Commonly used weighting schemes in text mining. In a term-document matrix format, each document d_j is represented by the values of the column features $f_j(w_n)$, which correspond to the number of occurrences of term w_n in document d_j. The total number of documents and terms present in a collection are defined as N_d and N_w, respectively.

Local weighting

Term frequency, $tf(w_n)$	$f_j(w_n)$
Logarithmic $tf(w_n)$	$\log(1 + f_j(w_n))$

Global weighting

Document frequency[1], $d_f(w_n)$	$\sum_{j=1}^{N_d} I(f_j(w_n))$
Collection frequency, $c_f(w_n)$	$\sum_{j=1}^{N_d} f_j(w_n)$
Inverse document frequency, $idf(w_n)$	$\frac{N_d}{d_f(w_n)}$

Combined

Term frequency-inverse document frequency, $tf\text{-}idf$	$\log(1 + f_j(w_n)) \log\left(\frac{N_d}{d_f(w_n)}\right)$
Pointwise mutual information, PMI[2]	$\log\left(\frac{f_j(w_n)}{\frac{c_f(w_n)}{N_w} \cdot \frac{1}{N_d}}\right)$
BM25[3]	$idf(w_n) \cdot \frac{f_j(w_n) \cdot (k_1 + 1)}{f_j(w_n) + k_1 \cdot (1 - b + b \cdot \frac{\#d_j}{\mathsf{E}_j[\#d_j]})}$

[1] $I(A) = 1$, if $A > 0$, and 0 otherwise.

[2] In statistical notation, $\text{PMI} = \log(\frac{p(w_n, d_j)}{p(w_n)p(d_j)})$.

[3] k_1 and b are free parameters, usually selected as $k_1 \in [1.2, 2.0]$ and $b = 0.75$. $\#d_j$ is the length of document d_j in words.

6.2 Content-based image information retrieval

With the rapid increase in the creation of electronic visual content, driven by the pervasive availability of cameras, from mobile phones to web-cams, the volume of visual information is increasing exponentially. While many technical problems associated with the generation of visual data have been solved, the task of content detection and subsequent classification is still unsolved. This is the field of content-based information retrieval (CBIR) methods (Müller et al., 2010), which focus on automatic analysis of image contents. Such methods are crucial when textual information is nonexistent or incomplete, but can also improve retrieval accuracy in the presence of text annotations (Lew et al., 2006).

Early approaches used text queries, and employed representation schemes like relational models, frame models, and object-oriented models to search for content (Colombo et al., 1999). They relied mostly on manual tags

to construct databases that could then be queried. Such approaches had several problems, from the human intensive tagging requirements, to subjective annotation strategies. To overcome those limitations, new methods were proposed, based on computer vision techniques. By focusing on feature-based similarities over images (Laaksonen et al., 2000), video (Flickner et al., 1995), it is possible to find content without relying on text annotations. Although feature-based methods provide a reliable way of retrieving information, they are usually not intuitive.

Recent CBIR approaches make use of multi-modal information (Schroff et al., 2007; Sjöberg et al., 2008), typically by using both text annotations and visual content. These allow them to diminish the so called *semantic gap*, *i.e.* the difference between low-level features and human semantic concepts. By constructing a feature vector composed of both visual and textual features, the search for similarities can be conducted at a semantic level. This stems from the natural languages of the textual data, which are closer to human semantic perception.

While several CBIR methods do take into account multi-modal information, the visual content is usually seen as a whole, and they do not segment particular image regions. Since images can be composed by several different regions, a natural extension is to focus on information retrieval from segmented images with associated text (Barnard et al., 2003; Blei and Jordan, 2003). Despite recent advances, these approaches typically require images annotated with reliable ground truth information, to model the joint (text and visual) feature distributions of image regions.

6.3 Document mining in neuroscience

Neuroscience is a particular field in which CBIR methods have been of particular relevance. With a pool of thousands of different neuroimaging journal publications, it is crucial to develop tools to synthesise and aggregate such data (Derrfuss and Mar, 2009). This kind of meta-research tools collect consistent findings across different studies, and can be used to evaluate activity changes in brain regions or networks, according to particular tasks or brain lesions reported by many different researchers.

The majority of available neuroimaging meta-analysis approaches use a coordinate-based analysis to generate probabilistic mappings between cognitive processes and neural states. Other methods do resort to a visual mapping of brain activity, to create the meta-analysis outcomes (Lewis,

2006), but data collection is still done by hand. Furthermore, in such methods, the regions of brain activity are typically approximated by spheres, instead of accurate spatial representations. While most meta-research studies still rely on human curators (Laird et al., 2009; Levy and Glimcher, 2012), rendering them cumbersome and with limited applicability, automatic approaches have also been proposed. Yarkoni et al. (2011) combine text-mining, meta-analysis and machine-learning techniques to create a fully automated framework to retrieve functional information from neuroscientific publications, depicted in Fig. 6.1. In addition to requiring the existence of activation coordinates, which are not always reported, none of the vast visual information, such as figures and charts, is used to generate the aforementioned mappings. Furthermore, and despite the use of rigorous multiple comparisons corrections, neuroimaging studies typically suffer from restricted sample sizes, which leads to many "false positive" reports of brain activity in meta studies (Wager et al., 2009).
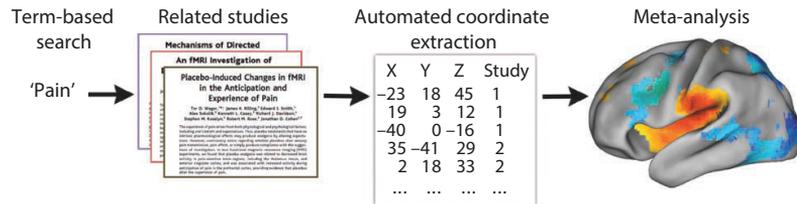
Figure 6.1. Flowchart describing a typical neuroscience meta-analysis method. First, all the coordinates reported in articles related to a particular term are extracted. Then, those peak coordinates are used to build a whole-brain map of activity probability. Adapted from Yarkoni et al. (2011).

As shown in the previous section, several content-based information retrieval methods do take into account multi-modal information, but they are not particularly suited to the analysis of neuroimaging data. Neuroimaging data can be considered as multi-layered information. The lowest or deepest layer is the anatomical brain itself, the brain activity is the next layer, while the combination of these layers leads to fMRI data. The final layer is a surrogate representations of such data, and corresponds to the fMRI images constructed by statistical analysis algorithms. Often, the last layer of information is the 2D representation of the full statistical maps, which are shown in neuroimaging publications. In order to extract as much brain activity[2] information as possible, one needs to devise a

---

[2]In this dissertation, the "brain activity" concept is interpreted in the absolute sense, but includes also differences in brain activity, when compared to a reference state.

method that goes beyond the imaging layer, and as close as possible to the activity itself. One of the main contributions of this dissertation is such a framework, allowing for the creation of an ontology of brain activity areas published in neuroscientific reports. It relies on the fact that typical neuroimaging studies report their data in a consistent way, where activation areas are overlayed on an anatomical image. In Publication VI, a CBIR method similar to the one proposed by Laaksonen et al. (2000) was used to extract, characterise and cluster fMRI images, based on their functional properties. This first approach was limited to a low-level feature characterisation of fMRI images, without trying to explore the "deepest layers" of such images. Publication VII improves that framework, creating and classifying three-dimensional brain activity maps estimated from studies dealing with effects of brain diseases in the DMN. Finally, Publication VIII extends the framework to a multi-modal approach, using textual content to characterise the estimated brain activities.

The following subsections give a general overview of the methodology proposed in Publications VI, VII and VIII, with a summarising flowchart shown in Fig. 6.2.

Figure 6.2. Flowchart of the document mining procedure. Starting with the 'blob' mining and clustering, the article information is used to group activity maps in a dendrogram, where each node can be characterised using textual information. Figure taken from Publication VIII.

6.3.1 'Blob' mining in neuroimaging articles

To retrieve the brain activity information from each neuroimaging publication, the first objective is to obtain fMRI activity reports depicted in images. As mentioned before, in neuroimaging reports, fMRI images display changes in brain activity using colour overlays onto anatomical MRI. These functional activity coloured regions are often known as 'blobs'.

Different authors of neuroscientific studies tend to report their results in slightly different styles. Furthermore, each document has inhomogeneous content information, such as multiple image frames per figure, other plots, annotations, or captions. These need to be handled before proceeding with the fMRI image analysis. After extracting the figures from the portable document format (PDF) files of publications, and in order to distinguish between different images in a publication, a compound figure segmentation method needs to be applied. The aim of this step is to identify, in each figure, the composing elements or images, as well as to characterise each of these. Several approaches can be used to detect such figure elements, including salient object segmentation (Zhou et al., 2013) and systematic detection and analysis of uniform space gaps (Chhatkuli et al., 2013). The methodology proposed in this dissertation uses a similar approach, but is mainly based in morphologically considerations, *cf.* Publication VII.
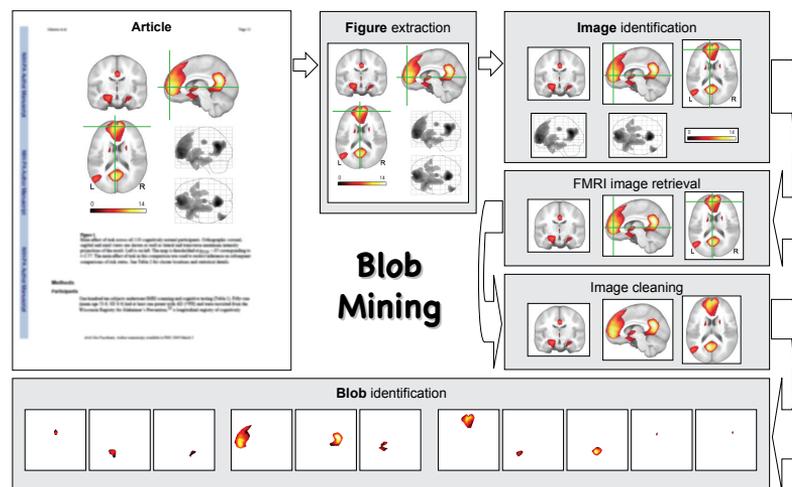

Figure 6.3. Flowchart describing the 'blob' mining procedure. First, figures are retrieved from articles (images adapted from (Johnson et al., 2007)). This is then followed by the detection and retrieval of fMRI images. After a cleaning procedure, the brain activity reported in those images is identified. Adapted from Publication VII.

After removing the background, and to identify each constituting image, figures are transformed to black and white, for the background and non-background areas respectively. Assuming that there is always background around images, and that those images have a minimum size, it is always possible to segment the figures into their compound images. This procedure is depicted in the "*Image identification*" frame of Fig. 6.3.

Since the goal is only to analyse fMRI images, the next step is to select them from the previously segmented images, as seen in frame *FMRI image retrieval* of Fig. 6.3. This selection is mostly done based on an ensemble of heuristic properties, including a minimum percentage of colour, and the aspect ratio typical of a brain image. The full detail of the properties used can be found in Publication VII.

Once in presence of images comprising a single brain with reported activity, the next step is to fully characterise such activity. Then, a few anatomical and functional parameters need to be estimated. These include several features of relevance, such as the type of section and the slice height, as well as the 'blobs'. After retrieving the fMRI images, and to allow for a proper comparison with anatomical templates, artifacts need to be removed. Common artifacts include text annotations and lines. Simple morphological techniques can be used to remove textual annotations, while Hough transform (Duda and Hart, 1972; Szeliski, 2010) deals very efficiently with lines, see Section 3.1.3.

After all pre-processing steps are performed, the last step consists in the estimation of the spatial location and intensity of the reported brain activity, as shown in the *Blob identification* frame of Fig. 6.3. To allow for a correct mapping of the fMRI data, one needs to identify which of the three standard planes are used in each image. As mentioned in Section 2.1, these brain views correspond to the axial, sagittal and coronal perspectives. The identification of such planes can be done using anatomical symmetry properties: axial sections are mostly symmetric about both axes, coronal sections are symmetric with respect to the vertical axis, and sagittal sections are clearly asymmetric.

Reports of fMRI activity areas are typically overlayed on anatomical MRI templates, usually in SPM (Friston, 2003) or Colin (Brett et al., 2002) formats. Colin volumes tend to have smaller voxel size and spacing than SPM ones, and therefore a measure of image complexity can be used to distinguish between them. In the proposed method, such measure is based in the *Canny* filter (Canny, 1986) edge detection.

After the volume and section are determined, all that remains is the identification of the slice where the brain activity occurred. With no other information characterising the location of brain activity besides the original slice image, this step finds the best matching slice from among the template images. As detailed in Section 3.1.2, there are many methods for image matching, from standard correlation to key-point matching (Bay et al., 2008; Lowe, 2004; Szeliski, 2010). When dealing with images from fMRI reports, one needs to take into account the varying sizes and rotations of the images, when comparing to the templates. This suggests that key-point matching methods would be suitable for this task. But, when dealing with electronic documents, image resolution is rather low, which limits the applicability of such techniques. Furthermore, due to the graphical characteristics of activity 'blobs', the number and type of features vary significantly between template and extracted images. Therefore, a combination of both correlation and SIFT matching is proposed. This can be done in a step-by-step approach, comparing all template slices from the volume and section previously identified, with a gray-scale version of the extracted images. First, a correlation procedure is performed. If the correlation between a scaled and/or rotated image and one of the template slices to which the image is compared to is above a high threshold, then correlation should be enough to find the correct match. If the correlation values are too low, then the slice with smallest SIFT feature distance should be selected as the best match.

Once the correct slice is found, the complete spatial coordinates are identified for the extracted image. While most coordinate-based meta-analysis methods rely on simple coordinates, often corresponding to the point of maximal activity, the proposed framework allows for the estimation of the full spatial information of the fMRI activity 'blobs'. Since these 'blobs' are typically reported as coloured regions, a segmentation based on *hue* information is enough for their identification. Unfortunately, the reporting style and colours used by different researchers can vary, as shown in Fig. 2.5. Furthermore, these 'blobs' can correspond to an increase or decrease in activity, usually through the *hot* and *cool* color scales respectively. Therefore, the task of detecting the correct colormap is far from trivial. One way to simplify the process is to consider that 'blobs' are only reports of activity, and ignore the type of change. This is a limitation of the proposed method, but since the current main goal is to find areas and their relations, it is not yet a crucial one.

With the spatial and intensity information from the 'blobs' determined, one can characterise those regions in more detail. For each 'blob', a set of features can be defined, as shown in Publication VI. The location of maximum activation, which may also be reported in textual format, can be determined by calculating the centroid of mass of the activation region. The size of such region is given by its area and perimeter. Other shape measures, such as compactness, extent or eccentricity can also be determined. Often, though, one may use, instead of those features, the activity patterns directly, *e.g.* when combining the information contained in various publications.

Although each 'blob' is considered an independent entity, the information about the originating articles, figures and images for each 'blob' is also stored. This preserves the original relations between 'blobs', allowing for comparisons across different structural levels, from articles to the 'blobs' themselves.

6.3.2 Summarising brain activity

The information retrieved in the previous section can easily be used to create a brain database. In order to create visual representations of the brain activity, most existing databases use fixed models, centred in the coordinates extracted from the publications (Lewis, 2006; Nielsen, 2003). With the proposed approach, a much more thorough information is obtained, since the complete activated area is recorded, which allows for a more precise estimate of brain activity.

Using the Colin (Brett et al., 2002) brain template as reference, all 'blob' intensity information can be mapped to their respective Talairach (Talairach and Tournoux, 1988) coordinates. The result is a four-dimensional intensity map $\boldsymbol{I}$, where each element $\boldsymbol{I}_n(x, y, z)$ corresponds to the intensity of 'blob' n, in its respective original image, at coordinates (x, y, z).

By superimposing these intensity maps over a common volume template, it is possible to obtain a visual summary of the brain activity reported in the analysed publications, as depicted in Fig. 6.4.

In the proposed approach, those summarising intensity maps use the *hot* colour scale, directly representing the estimated spatial intensity of each 'blob'.

In Publication VII, the main goal was to calculate an average intensity map for three different classes: Healthy, Alzheimer and Schizophrenia. After manually identifying which images corresponded to each of class,
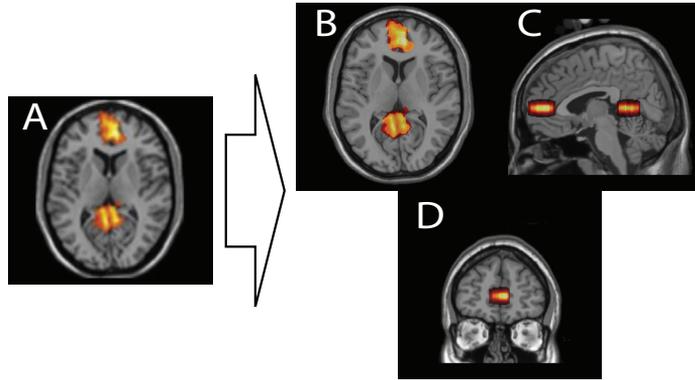
Figure 6.4. Example of the mapping procedure from a two dimensional image to a whole-brain activity map. On the left side is the fMRI image extracted from a publication (A), with the corresponding three dimensional activity map projected in axial (B), sagittal (C) and coronal (D) views on the right.

one summarising image per class could be produced. This resulted in three average intensity maps, where the intensity of each voxel corresponded to the average activity reported for each class of publications.

6.3.3 Intensity based clustering

After the construction of the intensity maps, and instead of directly using a classifying approach, one can start exploring relations between publications and different activity regions, as shown in Publication VIII.

In order to identify similar activation regions, reported in different articles, one can group all 'blobs' according to their spatial similarity. The main goal here being to group common activation patterns across various images, figures or even articles. In Publication VI, this grouping was performed based only on a pre-defined feature vector of each 'blob'. While that study demonstrated the feasibility of such approach, the feature definition and the wide range of some of the features pose crucial limitations. To avoid these limitations, in Publication VIII, a two-step clustering approach took only into account spatial information, such as location, 'blob' sizes, overlaps and sections.

To compare pairs of brain intensity maps, one needs to take into consideration the overlap and similarity between their intensity patterns. The measure proposed in Publication VIII takes into account those two factors, and is based on the *cosine* distance (Turney and Pantel, 2010):

$$d_1(n_1, n_2) = \frac{i_{n_1}^* \cdot i_{n_2}^*}{\|i_{n_1}^*\| \|i_{n_2}^*\|}, \tag{6.1}$$

where $i_{n_1}^*$ and $i_{n_2}^*$ are the vector forms of I_{n_1} and I_{n_2}, extracted from im-

ages n_1 and n_2, and for which both $\boldsymbol{I}_n$ have values greater than 0. By iteratively joining similar intensity maps, until a predefined threshold, a set of functional activation groups is obtained.

Since the intensity maps originated from 2D images, the comparison of 'blobs' from close, but different heights, is not easy to handle, using only the aforementioned distance. Also, the intensity similarity between maps originating from two different section types is hard to measure. To solve such limitations, and still join maps with similar spatial locations, a second clustering step can be performed. By calculating the average centroid of the 3D spatial coordinates of the 'blobs' belonging to each group, it becomes possible to cluster such groups, based on spatial coordinate proximity. Hence, one merger is based on overlapping brain activity, wheres the other is based on centroid distance proximity.

At this stage, the only information obtained is an average of how the brain activity in different brain regions looks like. In order to search for relations between different regions and different tasks, other types of information need to be explored.

6.3.4 Finding related activity regions

One may also assume that different articles reporting similar brain activity regions share a common pattern, which suggests the search for particular relations between the articles that built such overall regions. This step uses information regarding which articles built the previously mentioned clusters.

By hierarchically joining clusters based on the similarity of originating articles and their respective reported 'blobs', a dendrogram representative of article relation can be built. In such dendrogram, each branch is a consolidation of the two nodes that have the most overlap, in terms of articles pooled together in previous clusters, which correspond as well to similar brain activity regions.

This dendrogram can be seen as a summarising structure of co-occurring brain activity regions, where it is possible to search for article relations between regions. In Fig. 6.5, an example of such a dendrogram is presented, taken from Publication VIII. Four different article ensembles, corresponding to nodes with the least overlap of articles between them, but with common reported brain activity, are also shown in the figure and can be further studied.
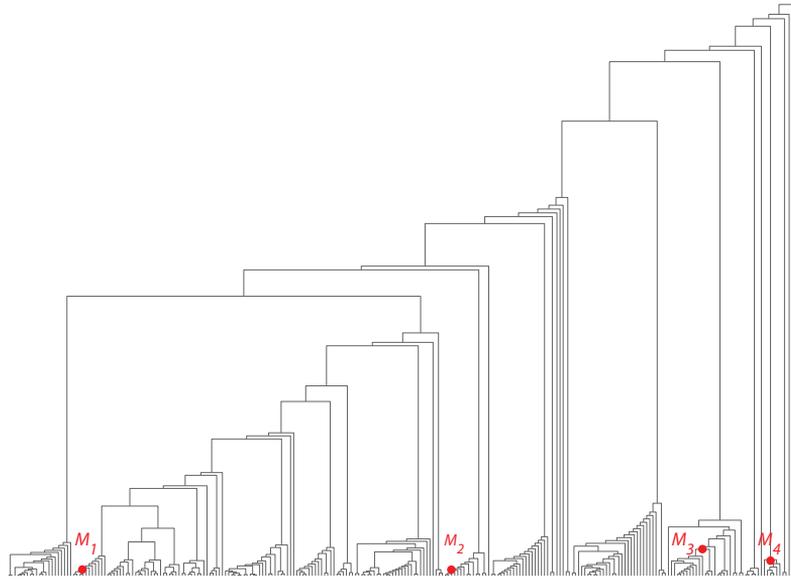
Figure 6.5. Dendrogram grouping articles based on their reported brain activity. Figure taken from Publication VIII.

6.3.5 Merging text and image data

While in Publication VII, all maps were averaged based on manually defined labels for three classes (Healthy, Alzheimer and Schizophrenia), the clustering procedure proposed above allows for a more complex search of relations between the reports of different articles, *cf.* Publication VIII.

To build the aforementioned dendrogram, only visual information was used. Yet, it is important to bridge the semantic gap, also in this context, and allow a proper brain activity ontology (Tegginmath et al., 2014). To achieve this goal, one can also use the textual information contained in the studied publications.

The search for interesting patterns allowing for a complete characterisation of all dendrogram nodes starts with the collection of text information from all publications, including the title, the abstract and the conclusion/discussion sections. Only this subset of text structures is used to avoid confounding topics, since it is typical for article introductions to mention a wide variety of literature survey information and methods not fully related to the topic studied in those articles.

To reduce the dimensionality of the data, the stemmed versions of words should be used to calculate term frequencies. Furthermore, common *stop words* should also be filtered out. To allow for a natural language analysis, the stemmed terms should also be replaced by the most common

originating word after the initial text processing.

Once the text is processed, a bag-of-words matrix can be calculated. To preserve textual structure and permit the occurrence of common neuroscientific terms composed of more than one word, the terms of such bag-of-words matrix should include words, bi-grams and tri-grams. By weighing the node terms using *tf-idf*, and selecting the most common ones, it is possible to build a textual characterisation of each node.

Once all this information is gathered, one can search for the most important terms in articles with brain activity in specific regions, or vice-versa. To simplify the visualisation of the most relevant terms in a group of articles, word-clouds can also be built (Rivadeneira et al., 2007). They easily depict the differences and commonalities between co-occurring brain activity regions and articles.

6.4 Experimental Results

In this section, some results, obtained using the proposed document mining approach, are shown. The publications used to show these results focused on the study of the resting state network, and the effects of Alzheimer and Schizophrenia in the human brain activity. For more detailed results, the reader is directed to Publications VII to VIII.

6.4.1 DMN and brain disorders

The average brain activity of 195 neuroscientific articles dealing with the DMN, in healthy patients is depicted in Fig. 6.6. The brain activity is shown at several axial heights and is superimposed over the Colin template. The extracted brain activity coincides with the typical subsystems reported to compose the DMN: the posterior cingulate/precuneous, the medial pre-frontal cortex and the inferior parietal lobes. Note that such an image is not present in any of the analysed articles. Yet, it can be produced by superimposing the reported images.

6.4.2 Characterising fMRI images

Figure 6.7 depicts four examples of brain activity regions found through the proposed document mining procedure, at various nodes of the dendrogram of Fig. 6.5, M_1 to M_4. Since the nodes do not contain all articles analysed, as shown in parentheses by the total number of articles
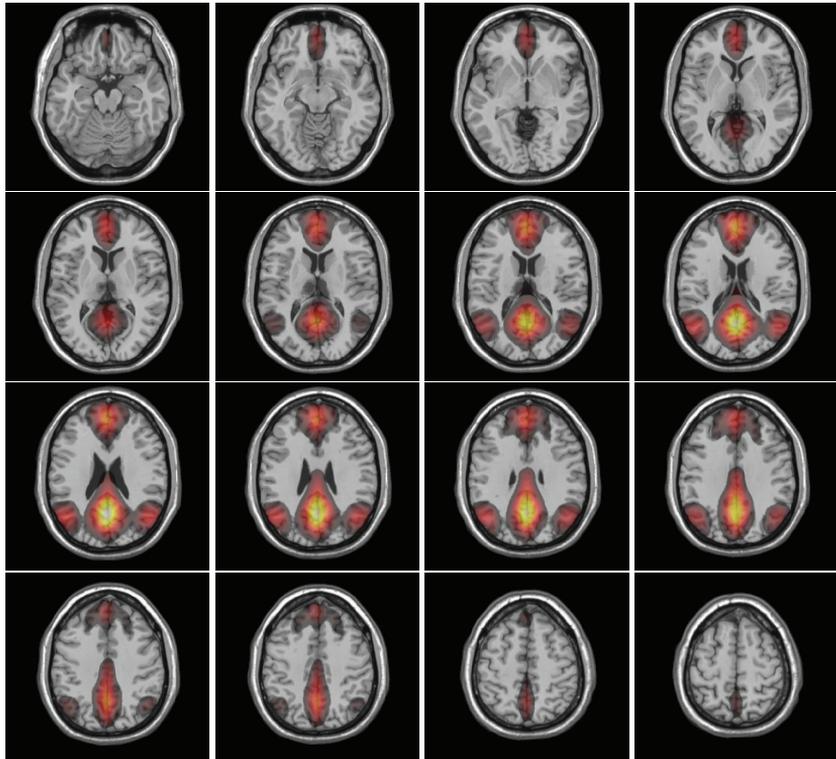
Figure 6.6. Average brain activity reported in publications dealing with healthy brains, superimposed on a Colin brain template, shown at various axial heights. Most of the activity is reported on the occipital, temporal and frontal areas of the brain, which correspond to the typical default mode network areas.

in participating in that node, the reported activity is different from the one shown in Fig. 6.6. From left to right, are displayed the axial, coronal and sagittal views of the activity map volumes, centred at their maximum value of intensity. Each of these regions is a graphical representation of the brain activity reported in four ensembles of articles, with as small as possible article overlap between them. As expected, they correspond to different co-occurring regions of brain activity. All volumes have activation in the posterior cingulate cortex, although each with a subtle change in location, while only M_2 and M_4 show considerable frontal activity. Notice that each node may correspond to a different activity function, since their corresponding articles may address a rather specific set of research questions. In particular, comparing those regions to the DMN activity map displayed in Fig. 6.6, none of the nodes, hence the articles that contributed to them, exhibited activity in the lateral parietal cortices.

Finally, using the textual information contained in the articles that built the previously shown brain activities, it is possible to build word-clouds
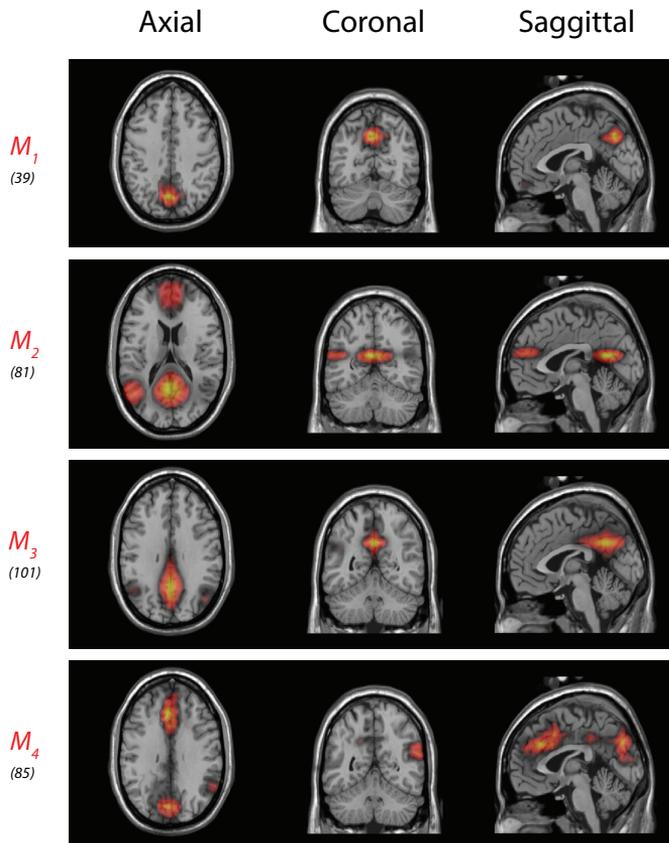
Figure 6.7. Average brain activity change reported in four different ensembles of articles (one per row), superimposed on a Colin-based brain template. From left to right are the axial, coronal and sagittal views of the nodes. In parentheses, the number of articles that built each set. Adapted from Publication VIII.

of the differences and commonalities between those node activity regions. Such word-clouds are shown in Fig. 6.8. The word-clouds displayed in the corners contain the n-grams that are most common in each set of articles, but that are not present in any of the other ensembles. The size of the n-grams is proportional to their weight in the bag of words corresponding to each set. While the M_1 word-cloud is characterised by n-grams mainly related with frequency analysis, M_2 contains mostly methodology considerations. M_4 seems to deal with Alzheimer and Parkinson's disease effects, while M_3 is somewhat harder to characterise. As expected, the terms that are prevalent in the cloud containing n-grams common to all article sets, shown in the middle of Fig. 6.8, are related to the main topics mentioned in the article collection, *e.g.* Alzheimer disease (*AD*), Schizophrenia, resting state network (*rsn*) or mild cognitive impairment (*mci*).

Figure 6.8. Word-clouds of four different article sets. On the corners are the word-clouds of n-grams representing the corresponding node, but not other nodes. The cloud in the middle contains the n-grams common to all four clusters. Figure taken from Publication VIII.

7. Discussion

This is not the end. It is not even the beginning of
the end. But it is, perhaps, the end of the beginning.
— Winston Churchill

The contributions proposed in this dissertation are related to different topics of research in Neuroinformatics. From a novel approach to perform brain segmentation, to a document mining procedure, developed to handle fMRI reports, both methods are valuable applications of machine learning techniques to the field of neuroscience. At the core of these methodologies rests one of the most widely used frameworks in data mining, cluster analysis. This dissertation contributes to all these topics, providing technical solutions to several problems that neuroscience researchers and health specialists face every day. Hopefully, such contributions will pave the way to new insights on the human brain, and the implications of brain lesions in its normal functioning.

7.1 Data clustering consistency estimation

The first contribution of this dissertation is an approach to identify which data elements are consistently clustered together, which allows for an estimation of the reliability of each data grouping. In this dissertation, such principles are applied to multiple runs of clustering algorithms, such as self-organizing maps and discriminative clustering, allowing for a robust estimation of data consistency across runs. In real data sets, typically due to noise, the number of data points consistently grouped together is often limited. Therefore, the groups of consistent data elements may not cover the whole data set. Nonetheless, these can be used as labels or training data for subsequent classification or semi-supervised clustering methods.

The greatest advantage of the proposed clustering consistency estimation is its ability to avoid parameter selection. Furthermore, by finding labels in a data-driven approach, one avoids an expensive and time consuming manual label estimation.

Despite its advantages, the application of such methodology has also some limitations. When dealing with unbalanced data, it is possible that some of the less populated clusters are not found. Clustering relies on clear differences between groups of data. If these differences are not significant, and the number of available samples of one of those groups is clearly insufficient to represent it, the results will not be optimal.

7.2 Self-supervised brain tissue segmentation

Another major contribution presented in this dissertation is a self-supervised method to perform brain tissue segmentation. Using conventional multi-spectral MRI data, and relying only on intensity differences, it performs tissue identification for many tissues present in the brain. Despite its non-specificity regarding the type of segmentation, the performance of the proposed self-supervised method is in-line with more dedicated methods. Unlike many other segmentation approaches, the proposed method handles partial volume effects, by estimating tissue probabilities for each voxel. Furthermore, it avoids the use of anatomical priors, allowing it to be flexible regarding variations in MRI machine settings and time/site of image collection. Another advantage is its ability to deal with different tissue types. Although the proposed method prefers a balanced number of voxels per class, it still works remarkably well when dealing with lesions, which often have a significantly smaller number of voxel count.

Since the segmentation is based on a semi-supervised clustering method, it requires labels to guide the learning process. This data can be acquired using the data clustering consistency estimation procedure proposed also in this thesis. Even when such an approach does not achieve perfect results, errors in the labels are easily compensated, due to the method's use of distributional information of the data. Furthermore, when a dedicated segmentation is sought, it is possible also to manually select labels.

This dissertation and publications contained herein show several applications of the proposed self-supervised segmentation method. The soft classification provided by the proposed method is perfectly suited to evaluate different stages of lesions, as well as assess lesion progression. It

also allows for the quantification of pathological brain changes, at very early stages, enabling the study of lesion evolution in longitudinal studies. This quantitative analysis is of remarkable use in clinical settings, and may help in the early diagnosis of brain lesions.

Although the proposed tissue segmentation method achieves accuracy values in-line with other state of the art approaches, it has some limitations as well. Of crucial importance, in any multi-spectral MRI setting, is the selection of the sequences used. Depending on the task at hand, the intensity distributions of the different tissues need to be as independent as possible, in order for the algorithm to distinguish them. Furthermore, image noise, resolution and movement artifacts are all factors that may jeopardise the segmentation outcome, especially when dealing with partial volume effects. All these factors tend to blur the differences between tissue intensity distributions, posing difficulties to the segmentation.

Using the segmentation results of the proposed method, experts may obtain better diagnoses and dissipate dubious evaluations. Although the method was developed having neuroimaging in mind, it should be applicable to any multi-spectral setting where the different classes to be segmented are distinguishable through gray values distributions.

7.3 Multi-modal mining of neuroscientific documents

A content-based information retrieval system of neuroscientific journals is the third contribution of this dissertation. This framework retrieves fMRI images from published articles, and maps the activity reported therein to a template, allowing for an automatic summarisation and comparison between studies. With the huge amount of neuroimaging research data being produced every day, the task of gathering and analysing such data becomes highly complex. This task becomes even more daunting due to the very limited publicly available repositories of such data. To circumvent this limitation, and gather, indirectly, such information from all possible sources, current data mining approaches rely on meta-research of neuroscientific literature. While typical automatic meta-research methods involve the extraction of text-based brain coordinates from such publications, the proposed approaches try to access the layer of neuroimaging data information through their reports in scientific publications.

The proposed framework can be divided in two parts. The first consists in building databases of functional brain activity, by automatically de-

tecting fMRI images from published articles, and summarising the brain activity reported therein. The second part structures the database, by using several information elements present in the set of articles. Those structures collect relational information between articles/studies. Furthermore, textual information contained in those articles is used to characterise the resulting relations and activity maps.

Using a database built with the proposed approach, a neuroscientist can search for particular keywords, either from a functional, methodological or anatomical origin, to find interesting activity maps and their corresponding studies. Such a database can be used to test situations where there is either a clear agreement between different research reports or a challenge between theories. The former is a key aspect to the construction of functional neuro-atlases, whereas the latter may lead to true findings in neuroscience.

There are some limitations when using reported fMRI images to extract brain activity information. First, all images are assumed to represent a similar scale of activity, irrespectively of the number of subjects studied. Also, different thresholds and methods are used by different researchers. These problems impair the aforementioned database creation and any statistical analysis performed therein. To mitigate such problems, an extension of the proposed method could be implemented in online publishing systems. This extension would calculate the anatomical and functional characteristics reported in fMRI images, and enable authors to provide clear descriptions or keywords of such images. With minimal effort, the quality of the gathered information would improve tremendously.

Another restriction of the proposed framework is related to how textual information is used. At the moment, text is only a supporting "tool" that guides and characterizes visual information extracted from published articles. A future venue of research would be the unification of both information sources, possibly through the use of a textual brain ontology. Such ontology could then be extended and complemented with information automatically extracted from neuroscientific documents, both visual and textual, further improving the results obtained and the knowledge contained therein.

The extraction of information from medical images proposed in this dissertations is based on a few premises. The main one is that researchers don't have access to each others' data, only to their results in a printed format. Currently this premise still holds, with lack of publicly available

data and results. There are some efforts to fix this, like the recent support from the Elsevier publisher for authors to upload their neuroimaging data[1], or the openfMRI project[2], dedicated to the open sharing of fMRI datasets. Therefore, the ideas presented here are still timely and hopefully will be the precursor of several others in years to come, when publicly available online data is the norm.

---

[1]`http://www.elsevier.com/about/content-innovation/`
`3d-neuroimaging-data`, visited 05/2014.
[2]`https://openfmri.org/`, visited 07/2014.

Bibliography

R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

F. Admiraal-Behloul, D. M. J. van den Heuvel, H. Olofsen, M. J. P. van Osch, J. van der Grond, M. A. van Buchem, and J. H. C. Reiber. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. *Neuroimage*, 28(3):607–617, Nov 2005. doi: 10.1016/j.neuroimage.2005.06.061.

S. Ahalt, A. Krishnamurthy, P. Chen, and D. Melton. Competitive learning algorithms for vector quantization. *Neural Networks*, 3(3):277–290, September 1989. doi: 10.1016/0893-6080(90)90071-R.

H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, Dec 1974. ISSN 0018-9286. doi: 10.1109/TAC.1974.1100705.

A. Akselrod-Ballin, M. Galun, J. M. Gomori, M. Filippi, P. Valsasina, R. Basri, and A. Brandt. Automatic segmentation and classification of multiple sclerosis in multichannel MRI. *IEEE Transactions on Biomedical Engineering*, 56(10): 2461–2469, Oct. 2009. ISSN 0018-9294. doi: 10.1109/TBME.2008.926671.

E. Alhoniemi, J. Himberg, J. Parhankangas, and J. Vesantoen. SOM Toolbox. http://www.cis.hut.fi/projects/somtoolbox/. visited 07/2014.

J. Alirezaie, M. Jernigan, and C. Nahmias. Automatic segmentation of cerebral mr images using artificial neural networks. *IEEE Transactions on Nuclear Science*, 45(4):2174–2182, Aug 1998. ISSN 0018-9499. doi: 10.1109/23.708336.

S. Amari, A. Cichocki, and H. H. Yang. A New Learning Algorithm for Blind Signal Separation. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 757–763. The MIT Press, 1996.

U. Amato, M. Larobina, A. Antoniadis, and B. Alfano. Segmentation of magnetic resonance brain images through discriminant analysis. *Journal of Neuroscience Methods*, 131(1-2):65–74, Dec 2003. doi: 10.1016/S0165-0270(03)00237-1.

P. Anbeek, K. L. Vincken, M. J. P. van Osch, R. H. C. Bisschops, and J. van der Grond. Probabilistic segmentation of white matter lesions in MR imaging. *Neuroimage*, 21(3):1037 – 1044, 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2003.10.012.

J. Ashburner and K. J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, Jul 2005. doi: 10.1016/j.neuroimage.2005.02.018.

J. Ashburner and S. Klöppel. Multivariate models of inter-subject anatomical variability. *Neuroimage*, 56(2):422 – 439, 2011. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.03.059.

R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.

D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

K. Barnard, P. Duygulu, D. Forsyth, N. De Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, Feb 2003.

H. Barrett and W. Swindell. *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. Elsevier Science, 1996. ISBN 9780080572307.

M. Bates. Models of natural language understanding. *Proceedings of the National Academy of Sciences*, 92(22):9977–9982, 1995.

H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, June 2008. ISSN 1077-3142. doi: 10.1016/j.cviu.2007.09.014.

M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Neural Information Processing Systems (NIPS)*, volume 14, pages 585–591, 2001.

A. J. Bell and T. J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

E. Bellon, E. Haacke, P. Coleman, D. Sacco, D. Steiger, and R. Gangarosa. MR artifacts: a review. *American Journal of Roentgenology*, 147(6):1271–81, 1986.

J. C. Bezdek. Models for pattern recognition. In *Pattern Recognition with Fuzzy Objective Function Algorithms*, pages 1–13. Springer, 1981.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag New York Inc., 2006. ISBN 9780387310732.

D. M. Blei and M. I. Jordan. Modeling annotated data. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 127–134, New York, NY, USA, 2003. ACM. ISBN 1-58113-646-3. doi: 10.1145/860435.860460.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

N. Bloembergen, E. M. Purcell, and R. V. Pound. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Physical Review Letters*, 73(7):679–712, 1948. doi: 10.1103/PhysRev.73.679.

W. G. Bradley and D. D. Stark. *Magnetic resonance imaging*. Number v. 1 in Magnetic Resonance Imaging. C.V. Mosby Co. St. Louis, 1999. ISBN 9780815185185.

M. E. Brandt, T. P. Bohan, L. A. Kramer, and J. M. Fletcher. Estimation of CSF, white and gray matter volumes in hydrocephalic children using fuzzy clustering of MR images. *Computerized Medical Imaging and Graphics*, 18(1):25–34, 1994.

M. Brant-Zawadzki, G. Gillan, and W. Nitz. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. *Radiology*, 182(3):769–75, 1992.

L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

M. Brett, I. S. Johnsrude, and A. M. Owen. The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, 3(3):243–249, Mar 2002. doi: 10.1038/nrn756.

A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation*, 4(2):490–530, 2005.

R. Buckner, J. Andrews-Hanna, and D. Schacter. The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124:1–38, 2008.

R. B. Buxton. *Introduction to Functional Magnetic Resonance Imaging : Principles and Techniques*. Cambridge University Press, 2001.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986. ISSN 0162-8828. doi: 10.1109/TPAMI.1986.4767851.

P. Caravan, J. J. Ellison, T. J. McMurry, and R. B. Lauffer. Gadolinium (III) chelates as MRI contrast agents: structure, dynamics, and applications. *Chemical reviews*, 99(9):2293–2352, 1999.

J.-F. Cardoso. Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. In *Proc. ICASSP*, pages 2655–2658, 1990.

M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, S. Ourselin, and T. A. D. N. Initiative. LoAd: A locally adaptive cortical segmentation algorithm. *Neuroimage*, 56(3):1386–1397, Feb 2011. doi: 10.1016/j.neuroimage.2011.02.013.

H.-H. Chang, A. H. Zhuang, D. J. Valentino, and W.-C. Chu. Performance measure characterization for evaluating neuroimage segmentation algorithms. *Neuroimage*, 47(1):122–135, Aug 2009. doi: 10.1016/j.neuroimage.2009.03.068.

Y.-L. Chang and X. Li. Adaptive image region-growing. *IEEE Transactions on Image Processing*, 3(6):868–872, 1994.

O. Chapelle, B. Schölkopf, A. Zien, et al. *Semi-supervised learning*, volume 2. MIT press Cambridge, 2006.

A. Chhatkuli, A. Foncubierta-Rodríguez, D. Markonis, F. Meriaudeau, and H. Müller. Separating compound figures in journal articles to allow for subfigure classification. In *Proceedings of SPIE 2013*, volume 8674, pages 86740J–86740J–12, 2013. doi: 10.1117/12.2007897.

H. Choi, D. R. Haynor, and Y. Kim. Partial volume tissue classification of multichannel magnetic resonance images-a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407, 1991. ISSN 1558-254X. doi: 10.1109/42. 97590.

K. W. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29, 1990.

C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis*, 7(4):513 – 527, Dec 2003. ISSN 1361-8415. doi: 10.1016/S1361-8415(03)00037-9. Medical Image Computing and Computer Assisted Intervention.

R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7426–7431, 2005. doi: 10. 1073/pnas.0500334102.

D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, Jun 1998. doi: 10. 1109/42.712135.

C. Colombo, A. Del Bimbo, and P. Pala. Semantics in visual information retrieval. *IEEE Multimedia*, 6(3):38–53, 1999.

O. Commowick and S. K. Warfield. Estimation of inferential uncertainty in assessing expert segmentation performance from staple. *IEEE Transactions on Medical Imaging*, 29(3):771 –780, march 2010. ISSN 0278-0062. doi: 10.1109/TMI.2009.2036011.

O. Commowick, P. Fillard, O. Clatz, and S. K. Warfield. Detection of DTI white matter abnormalities in multiple sclerosis patients. *Medical Image Computing and Computer Assisted Intervention*, 11(Pt 1):975–982, 2008. doi: 10.1007/ 978-3-540-85988-8_116.

P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

A. Compston and A. Coles. Multiple sclerosis. *The Lancet*, 372(9648):1502 – 1517, 2008. ISSN 0140-6736. doi: 10.1016/S0140-6736(08)61620-7.

P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage*, 54(2):940 – 954, 2011. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.09.018.

A. R. Crossman and D. Neary. *Neuroanatomy*. Churchill Livingstone, 3rd edition, 2006.

F. C. Crow. Summed-area tables for texture mapping. *SIGGRAPH Computer Graphics*, 18(3):207–212, January 1984. ISSN 0097-8930. doi: 10.1145/ 964965.808600.

W. R. Crum, O. Camara, and D. L. G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, Nov 2006. doi: 10.1109/TMI.2006.880587.

R. Cruz-Barbosa and A. Vellido. Semi-Supervised Analysis of Human Brain Tumour from Partially Labeled MRS Information using Manifold Learning Models. *International Journal of Neural Systems*, 21(1):17–29, 2011. doi: 10.1142/S0129065711002626.

R. Cuingnet, E. Gérardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the adni database. *Neuroimage*, 56(2):766–781, 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.06.013.

N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893. IEEE, 2005.

A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis. I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, Feb 1999. doi: 10.1006/nimg.1998.0395.

J. S. Damoiseaux, K. E. Prater, B. L. Miller, and M. D. Greicius. Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiology of Aging*, 33(4):828.e19–828.e30, Apr 2012. ISSN 0197-4580. doi: 10.1016/j.neurobiolaging.2011.06.024.

J. Damoiseaux, C. Beckmann, E. Arigita, F. Barkhof, P. Scheltens, C. Stam, S. Smith, and S. Rombouts. Reduced resting-state brain activity in the "default network" in normal aging. *Cerebral Cortex*, 18(8):1856–64, 2008.

R. de Boer, H. A. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen. White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage*, 45(4):1151–1161, May 2009. doi: 10.1016/j.neuroimage.2009.01.011.

J. de Groot, F.-E. de Leeuw, M. Oudkerk, A. Hofman, J. Jolles, and M. Breteler. Cerebral white matter lesions and subjective cognitive dysfunction: The Rotterdam Scan Study. *Neurology*, 56:1539–1545, 2001.

M. de Groot, B. F. Verhaaren, R. de Boer, S. Klein, A. Hofman, A. van der Lugt, M. A. Ikram, W. J. Niessen, and M. W. Vernooij. Changes in normal-appearing white matter precede development of white matter lesions. *Stroke*, 44(4):1037–1042, 2013. doi: 10.1161/STROKEAHA.112.680223.

G. Deco, V. K. Jirsa, and A. R. McIntosh. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1):43–56, Jan 2011.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

J. Derrfuss and R. Mar. Lost in localization: The need for a universal coordinate database. *Neuroimage*, 48(1):1 – 7, 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.01.053.

M. Deza and E. Deza. *Encyclopedia of Distances*. Encyclopedia of Distances. Springer, 2009. ISBN 9783642002342.

I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 269–274. ACM, 2001.

L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945. doi: 10.2307/1932409.

D. Dodell-Feder, L. E. DeLisi, and C. I. Hooker. The relationship between default mode network connectivity and social functioning in individuals at familial high-risk for schizophrenia. *Schizophrenia Research*, 156(1):87 – 95, 2014. ISSN 0920-9964. doi: 10.1016/j.schres.2014.03.031.

W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, 1973.

W. Dou, S. Ruan, Y. Chen, D. Bloyet, and J.-M. Constans. A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images. *Image and Vision Computing*, 25(2):164 – 171, 2007. ISSN 0262-8856. doi: 10.1016/j.imavis.2006.01.025. Soft Computing in Image Analysis.

R. O. Duda and P. E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, January 1972. ISSN 0001-0782. doi: 10.1145/361237.361242.

N. Dunford and J. Schwartz. *Linear Operators: General theory*. Pure and applied mathematics. Interscience Publishers, 1958.

J. Dunn. Fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973. cited By (since 1996)1620.

T. B. Dyrby, E. Rostrup, W. F. C. Baaré, E. C. W. van Straaten, F. Barkhof, H. Vrenken, S. Ropele, R. Schmidt, T. Erkinjuntti, L.-O. Wahlund, L. Pantoni, D. Inzitari, O. B. Paulson, L. K. Hansen, G. Waldemar, and L.A.D.I.S. study group. Segmentation of age-related white matter changes in a clinical multi-center study. *Neuroimage*, 41(2):335–345, Jun 2008. doi: 10.1016/j.neuroimage.2008.02.024.

R. R. Edelman, J. Hesselink, and M. Zlatkin. *Clinical Magnetic Resonance Imaging*. W.B. Saunders Company, 2005.

C. Elliott, D. Arnold, D. Collins, and T. Arbel. Temporally Consistent Probabilistic Detection of New Multiple Sclerosis Lesions in Brain MRI. *IEEE Transactions on Medical Imaging*, 32(8):1490–1503, Aug 2013. ISSN 0278-0062. doi: 10.1109/TMI.2013.2258403.

S. F. Eskildsen, P. Coupé, V. Fonov, J. V. Manjón, K. K. Leung, N. Guizard, S. N. Wassef, L. R. Østergaard, D. L. Collins, and A. D. N. Initiative. BEaST: brain extraction based on nonlocal segmentation technique. *Neuroimage*, 59 (3):2362–2373, Feb 2012. doi: 10.1016/j.neuroimage.2011.09.012.

F. Esposito, G. Pignataro, G. Di Renzo, A. Spinali, A. Paccone, G. Tedeschi, and L. Annunziato. Alcohol increases spontaneous BOLD signal fluctuations in the visual network. *Neuroimage*, 53(2):534–43, 2010.

J. Fang, J. Yang, E. Tu, Z. Jia, N. Kasabov, and C. Liu. Statistical approaches to automatic level set image segmentation with multiple regions. *Optical Engineering*, 50(10):107001–107001–10, 2011. doi: 10.1117/1.3631042.

X. Z. Fern and W. Lin. Cluster ensemble selection. *Statistical Analysis and Data Mining*, 1(3):128–141, 2008.

J. M. Ferro and S. Madureira. Age-related white matter changes and cognitive impairment. *Journal of Neurological Science*, 203-204:221–225, 2002.

M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

FIL Methods Group. Statistical Parametric Mapping. http://www.fil.ion.ucl.ac.uk/spm/. visited 05/2014.

B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2): 195–207, Feb 1999. doi: 10.1006/nimg.1998.0396.

M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, et al. Query by image and video content: The QBIC system. *Computer*, 28(9):23–32, 1995.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

K. J. Friston. Statistical parametric mapping. In *Neuroscience Databases*, pages 237–250. Springer, 2003.

F. Galdames, F. Jaillet, and C. Perez. An accurate skull stripping method based on simplex meshes and histogram analysis for magnetic resonance images. *Journal of Neuroscience Methods*, 206(2):103–19, 2012.

F. Galton. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, pages 246–263, 1886.

J. Gao, F. Liang, W. Fan, Y. Sun, and J. Han. A graph-based consensus maximization approach for combining multiple supervised and unsupervised models. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):15–28, Jan 2013. ISSN 1041-4347. doi: 10.1109/TKDE.2011.206.

D. García-Lorenzo, J. Lecoeur, D. L. Arnold, D. L. Collins, and C. Barillot. Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts. *Medical Image Computing and Computer Assisted Intervention*, 12(2):584–591, 2009.

M. Garcia-Sebastian, C. Hernandez, and A. d'Anjou. Robustness of an adaptive MRI segmentation algorithm parametric intensity inhomogeneity modeling. *Neurocomputing*, 72(10-12):2146 – 2152, 2009. ISSN 0925-2312. doi: 10.1016/j.neucom.2008.07.014.

C. F. G. C. Geraldes and S. Laurent. Classification and basic properties of contrast agents for magnetic resonance imaging. *Contrast Media & Molecular Imaging*, 4(1):1–23, 2009. ISSN 1555-4317. doi: 10.1002/cmmi.265.

G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, 1992. doi: 10.1109/42.141646.

C. Goddard. *Semantic Analysis: A Practical Introduction*. Oxford Textbooks in Linguistics. OUP Oxford, 2011. ISBN 9780199560288.

R. Goebel, F. Esposito, and E. Formisano. Analysis of functional image analysis contest data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, 27(5):392–401, May 2006. doi: 10.1002/hbm.20249.

R. Gonzalez and R. Woods. *Digital Image Processing*. Pearson/Prentice Hall, 2008. ISBN 9780131687288.

N. Gordillo, E. Montseny, and P. Sobrevilla. State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging*, 31(8):1426 – 1438, 2013. ISSN 0730-725X. doi: 10.1016/j.mri.2013.05.002.

C. W. Granger and R. Ramanathan. Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204, 1984.

D. Greve and B. Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.

D. J. Griffiths. *Introduction to Quantum Mechanics*. Prentice Hall, 2004.

J. Gross, J. Yellen, and P. Zhang. *Handbook of Graph Theory, Second Edition*. Discrete Mathematics and Its Applications. Taylor & Francis, 2013. ISBN 9781439880180.

R. I. Grossman and J. C. McGowan. Perspectives on multiple sclerosis. *American Journal of Neuroradiology*, 19:1251–1265, 1998.

H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–4, 1995.

C. Günay, J. R. Edgerton, S. Li, T. Sangrey, A. A. Prinz, and D. Jaeger. Database analysis of simulated and recorded electrophysiological datasets with PANDORA's toolbox. *Neuroinformatics*, 7(2):93–111, 2009. ISSN 1539-2791. doi: 10.1007/s12021-009-9048-z.

A. Gupta and R. Jain. Visual information retrieval. *Commun. ACM*, 40(5):70–79, May 1997. ISSN 0001-0782. doi: 10.1145/253769.253798.

H. Gävert, J. Hurri, J. Särelä, and A. Hyvärinen. Fastica MATLAB™ package. http://www.cis.hut.fi/projects/ica/fastica. visited 04/2014.

M. Hagenbuchner and A. C. Tsoi. A supervised self-organizing map for structures. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 3, pages 1923–1928 vol.3, July 2004. doi: 10.1109/IJCNN.2004.1380906.

H. K. Hahn and H.-O. Peitgen. The Skull Stripping Problem in MRI Solved by a Single 3D Watershed Transform. In S. L. Delp, A. M. DiGoia, and B. Jaramaz, editors, *Medical Image Computing and Computer-Assisted Intervention*

– *MICCAI 2000*, volume 1935 of *Lecture Notes in Computer Science*, pages 134–143. Springer Berlin Heidelberg, 2000. ISBN 978-3-540-41189-5. doi: 10.1007/978-3-540-40899-4_14.

J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes. *Medical Image Registration*. CRC, 2001.

M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography– theory, instrumentation, and applications to non-invasive studies of the working human brain. *Reviews of Modern Physics*, 65: 413–497, Apr 1993. doi: 10.1103/RevModPhys.65.413.

R. W. Hamming. Error detecting and error correcting codes. *Bell System technical journal*, 29(2):147–160, 1950.

P. Hansen, M. Kringelbach, and R. Salmelin. *MEG: An Introduction to Methods*. Oxford University Press, USA, 2010. ISBN 9780199719136. doi: 10.1093/acprof:oso/9780195307238.001.0001.

P. Hart. How the Hough transform was invented [DSP History]. *IEEE Signal Processing Magazine*, 26(6):18–22, November 2009. ISSN 1053-5888. doi: 10.1109/MSP.2009.934181.

J. A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition, 1975. ISBN 047135645X.

Harvard Medical School. Internet Brain Segmentation Repository (IBSR). http://www.nitrc.org/projects/ibsr. visited 05/2013.

S. Haykin. *Neural Networks - A Comprehensive Foundation*. Prentice Hall International, 1999.

T. C. Head. *Essentials of Clinical Neurophysiology*. Butterworth-Heinemann, 2002.

D. O. Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 1st edition, 1949.

R. A. Heckemann, A. Hammers, D. Rueckert, R. I. Aviv, C. J. Harvey, and J. V. Hajnal. Automatic volumetry on MR brain images can support diagnostic decision making. *BMC Medical Imaging*, 8:9, 2008. doi: 10.1186/1471-2342-8-9.

B. A. Hill. *Multiple Sclerosis Q & A*. Avery, 2003.

K. H. Höhne and W. A. Hanson. Interactive 3D segmentation of MRI and CT volumes using morphological operations. *Journal of Computer Assisted Tomography*, 16(2):285–294, 1992.

S. Hojjatoleslami, F. Kruggel, and D. Y. von Cramon. Segmentation of white matter lesions from volumetric MR images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI'99*, pages 52–61. Springer, 1999.

Y. Hong, S. Kwong, Y. Chang, and Q. Ren. Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm. *Pattern Recognition*, 41(9):2742–2756, 2008.

M. Hori, T. Okubo, K. Uozumi, K. Ishigame, H. Kumagai, and T. Araki. T1-Weighted Fluid-Attenuated Inversion Recovery at Low Field Strength: A Viable Alternative for T1-Weighted Intracranial Imaging. *American Journal of Neuroradiology*, 24(4):648–651, 2003.

P. Hough. Machine Analysis Of Bubble Chamber Pictures. *Proceedings of the International Conference on High Energy Accelerators and Instrumentation*, C590914:554–558, 1959.

S. A. Huettel, A. W. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging, Second Edition*. Sinauer Associates, Inc, 2008.

A. Hyvärinen. Independent Component Analysis for Time-dependent Stochastic Processes. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of the International Conference on Artificial Neural Networks (ICANN´98)*, Perspectives in Neural Computing, pages 135–140, Svökde, Sweden, 1998. Springer London. ISBN 978-3-540-76263-8. doi: 10.1007/978-1-4471-1599-1_16.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.7.1483.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*, volume 46. John Wiley & Sons, INC., 2001. Editor Simon Haykin.

J. Iglesias, C. Liu, P. Thompson, and Z. Tu. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging*, 30(9):1617–34, 2011.

N. Indurkhya and F. J. Damerau. *Handbook of natural language processing*, volume 2. CRC Press, 2012.

P. Jaccard. The Distribution of the Flora in the Alpine Zone . *New Phytologist*, 11(2):37–50, Feb 1912.

A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651 – 666, 2010. ISSN 0167-8655. doi: 10.1016/j.patrec.2009.09.011. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 19th International Conference in Pattern Recognition (ICPR).

P. Jezzard, P. M. Matthews, and S. M. Smith. *Functional Magnetic Resonance Imaging: An Introduction to Methods*. Oxford University Press, 2003.

T. Joachims. Transductive learning via spectral graph partitioning. In *International Conference on Machine Learning (ICML)*, pages 290–297, 2003.

S. C. Johnson, M. L. Ries, T. M. Hess, C. M. Carlsson, C. E. Gleason, A. L. Alexander, H. A. Rowley, S. Asthana, and M. A. Sager. Effect of alzheimer disease risk on brain function during self-appraisal in healthy middle-aged adults. *Archives of General Psychiatry*, 64(10):1163–1171, Oct 2007. doi: 10.1001/archpsyc.64.10.1163.

H. Jokinen, A. A. Gouw, S. Madureira, R. Ylikoski, E. C. van Straaten, W. M. van der Flier, F. Barkhof, P. Scheltens, F. Fazekas, R. Schmidt, A. Verdelho, J. M. Ferro, L. Pantoni, D. Inzitari, T. Erkinjuntti, and L. S. Group. Incident

lacunes influence cognitive decline: the LADIS study. *Neurology*, 76(22):1872–1878, May 31 2011.

H. Jokinen, J. Lipsanen, R. Schmidt, F. Fazekas, A. A. Gouw, W. M. van der Flier, F. Barkhof, S. Madureira, A. Verdelho, J. M. Ferro, A. Wallin, L. Pantoni, D. Inzitari, T. Erkinjuntti, and L. S. Group. Brain atrophy accelerates cognitive decline in cerebral small vessel disease: the LADIS study. *Neurology*, 78(22): 1785–1792, May 29 2012.

G. Jolles and J. M. Stutzmann. *Neurodegenerative Diseases*. Academic Press, 1994.

K. S. Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.

J. Junno, M. Paananen, J. Karppinen, T. Tammelin, J. Niinimäki, E. Lammentausta, M. Niskanen, M. Nieminen, M. Järvelin, J. Takatalo, O. Tervonen, and J. Tuukkanen. Influence of physical activity on vertebral strength during late adolescence. *Spine Journal*, 13(2):184–9, 2013.

C. Jutten and J.-F. Cardoso. Source separation: really blind? In *Proceedings NOLTA*, pages 79–84, 1995.

J. W. Kalat. *Biological Psychology*. Cengage Learning, 11th edition, 2012.

E. Kanal, J. P. Borgstede, A. J. Barkovich, C. Bell, W. G. Bradley, J. P. Felmlee, J. W. Froelich, E. M. Kaminski, E. K. Keeler, J. W. Lester, et al. American College of Radiology white paper on MR safety. *American Journal of Roentgenology*, 178(6):1335–1347, 2002.

E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science*. McGraw-Hill Medical, 4th edition, January 2000.

A. Kao and S. R. Poteet. *Natural language processing and text mining*. Springer, 2007.

T. Kapur, W. L. Grimson, W. M. W. III, and R. Kikinis. Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis*, 1(2):109 – 127, 1996. ISSN 1361-8415. doi: 10.1016/S1361-8415(96)80008-9.

N. Kasabov and Q. Song. DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Transactions on Fuzzy Systems*, 10(2):144–154, Apr 2002. ISSN 1063-6706. doi: 10.1109/91.995117.

S. Kaski, J. Sinkkonen, and A. Klami. Discriminative clustering. *Neurocomputing*, 69(1-3):18–41, 2005. doi: 10.1016/j.neucom.2005.02.012.

B. M. Keegan and J. H. Noseworthy. Multiple sclerosis. *Annual Review of Medicine*, 53:285–302, 2002.

A. Khademi, A. Venetsanopoulos, and A. R. Moody. Robust white matter lesion segmentation in FLAIR MRI. *IEEE Transactions on Biomedical Engineering*, 59(3):860–871, Mar 2012. doi: 10.1109/TBME.2011.2181167.

M. M. Khoo, P. A. Tyler, A. Saifuddin, and A. R. Padhani. Diffusion-weighted imaging (DWI) in musculoskeletal MRI: a critical review. *Skeletal Radiology*, 40(6):665–681, 2011. ISSN 0364-2348. doi: 10.1007/s00256-011-1106-6.

F. Klauschen, A. Goldman, V. Barra, A. Meyer-Lindenberg, and A. Lundervold. Evaluation of automated brain MR image segmentation and volumetry methods. *Human Brain Mapping*, 30(4):1310–1327, Apr 2009. doi: 10.1002/hbm.20599.

W. Koch, S. Teipel, S. Mueller, J. Benninghoff, M. Wagner, A. L. Bokde, H. Hampel, U. Coates, M. Reiser, and T. Meindl. Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease. *Neurobiology of Aging*, 33(3):466–478, Mar 2012. ISSN 0197-4580. doi: 10.1016/j.neurobiolaging.2010.04.013.

T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2001. Editors: Thomas S. Huang, Teuvo Kohonen, Manfred R. Schroeder.

F. Kruggel and D. Y. von Cramon. Alignment of magnetic-resonance brain datasets with the stereotactical coordinate system. *Medical Image Analysis*, 3 (2):175 – 185, 1999. ISSN 1361-8415. doi: 10.1016/S1361-8415(99)80005-X.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 03 1951. doi: 10.1214/aoms/1177729694.

A. La Gerche, G. Claessen, A. Van de Bruaene, N. Pattyn, J. Van Cleemput, M. Gewillig, J. Bogaert, S. Dymarkowski, P. Claus, and H. Heidbuchel. Cardiac MRI A New Gold Standard for Ventricular Volume Quantification During High-Intensity Exercise. *Circulation: Cardiovascular Imaging*, 6(2):329–338, 2013.

J. Laaksonen and T. Honkela, editors. *Advances in Self-Organizing Maps, WSOM 2011*. Springer, Berlin, 2011. doi: 10.1007/978-3-642-21566-7.

J. Laaksonen, M. Koskela, S. Laakso, and E. Oja. PicSOM – content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13–14): 1199 – 1207, 2000. ISSN 0167-8655. doi: 10.1016/S0167-8655(00)00082-9. Selected Papers from The 11th Scandinavian Conference on Image.

LADIS Study Group. 2001-2011: a decade of the LADIS (Leukoaraiosis And DISability) Study: what have we learned about white matter changes and small-vessel disease? *Cerebrovascular diseases (Basel, Switzerland)*, 32(6): 577–588, 2011.

A. R. Laird, J. L. Lancaster, and P. T. Fox. BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics*, 3(1):65–78, 2005.

A. R. Laird, J. L. Lancaster, and P. T. Fox. Lost in localization? the focus is meta-analysis. *Neuroimage*, 48(1):18–20, Oct 2009. doi: 10.1016/j.neuroimage.2009. 06.047.

M. Langarizadeh, R. Mahmud, A. Ramli, S. Napis, M. Beikzadeh, and W. Rahman. Improvement of digital mammogram images using histogram equalization, histogram stretching and median filter. *Journal of Medical Engineering and Technology*, 35(2):103–8, 2011. doi: 10.3109/03091902.2010.542271.

I. Laptev and T. Lindeberg. Local descriptors for spatio-temporal recognition. In *Spatial Coherence for Visual Motion Analysis*, pages 91–103. Springer, 2006.

S. M. LaValle and S. A. Hutchinson. A Bayesian segmentation methodology for parametric image models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(2):211–217, 1995.

D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546, 2001. ISSN 1522-2586. doi: 10.1002/jmri.1076.

C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner. Segmenting brain tumors with conditional random fields and support vector machines. In *Computer vision for biomedical image applications*, pages 469–478. Springer, 2005.

J.-D. Lee, H.-R. Su, P. E. Cheng, M. Liou, J. A. D. Aston, A. C. Tsai, and C.-Y. Chen. MR image segmentation using a power transformation approach. *IEEE Transactions on Medical Imaging*, 28(6):894–905, June 2009. ISSN 0278-0062. doi: 10.1109/TMI.2009.2012896.

K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, Oct 1999. doi: 10.1109/42.811270.

K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8):677–688, Aug 2001. doi: 10.1109/42.938237.

K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 22(1):105–119, Jan 2003. doi: 10.1109/TMI.2002.806587.

D. J. Levy and P. W. Glimcher. The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, 22(6):1027–1038, Jul 2012. doi: 10.1016/j.conb.2012.06.001.

M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2(1):1–19, February 2006. ISSN 1551-6857. doi: 10.1145/1126004.1126005.

J. W. Lewis. Cortical networks related to human use of tools. *The Neuroscientist*, 12(3):211–231, 2006. doi: 10.1177/1073858406288327.

K. O. Lim and A. Pfefferbaum. Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. *Journal of Computer Assisted Tomography*, 13(4):588–593, 1989.

G.-C. Lin, C.-M. Wang, W.-J. Wang, and S.-Y. Sun. Automated classification of multispectral MR images using unsupervised constrained energy minimization based on fuzzy logic. *Magnetic Resonance Imaging*, 28(5):721–738, Jun 2010. doi: 10.1016/j.mri.2010.03.009.

H. Liu, C. Yang, N. Pan, E. Song, and R. Green. Denoising 3D MR images by the enhanced non-local means filter for Rician noise. *Magnetic Resonance Imaging*, 28(10):1485–1496, 2010. ISSN 0730-725X. doi: 10.1016/j.mri.2010.06.023.

S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, Mar 1982. ISSN 0018-9448. doi: 10.1109/TIT.1982. 1056489.

D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

D. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150 –1157 vol.2, 1999. doi: 10.1109/ICCV.1999.790410.

U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4): 395–416, 2007. ISSN 0960-3174. doi: 10.1007/s11222-007-9033-z.

D. J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, 1st edition, 2002.

J. V. Manjón, J. Tohka, and M. Robles. Improved estimates of partial volume coefficients from noisy brain MRI using spatial context. *Neuroimage*, 53(2): 480 – 490, 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.06.046.

C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

R. Mäntylä, T. Erkinjuntti, O. Salonen, H. J. Aronen, T. Peltonen, T. Pohjasvaara, and C.-G. Standertskjöld-Nordenstam. Variable agreement between visual rating scales for white matter hyperintensities on MRI : Comparison of 13 rating scales in a poststroke cohort. *Stroke*, 28(8):1614–1623, 1997.

G. Mariani, L. Bruselli, T. Kuwert, E. Kim, A. Flotats, O. Israel, M. Dondi, and N. Watanabe. A review on the clinical uses of SPECT/CT. *European Journal of Nuclear Medicine and Molecular Imaging*, 37(10):1959–1985, 2010. ISSN 1619-7070. doi: 10.1007/s00259-010-1390-8.

T. Martinetz, K. Schulten, et al. *A "neural-gas" network learns topologies*. University of Illinois at Urbana-Champaign, 1991.

A. Mayer and H. Greenspan. An adaptive mean-shift framework for MRI brain segmentation. *IEEE Transactions on Medical Imaging*, 28(8):1238–1250, Aug. 2009. ISSN 0278-0062. doi: 10.1109/TMI.2009.2013850.

J. C. Mazziotta, A. W. Toga, and R. S. Frackowiak. *Brain Mapping: The Disorders: The Disorders*. Academic Press, 2000.

G. J. McLachlan and K. E. Basford. Chapters 1 and 2. In McLachlan and Basford, editors, *Mixture models: inference and applications to clustering*, pages 1–69. Marcel Dekker, Inc., 1987.

D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince. *MRI from Picture to Proton*. Cambridge University Press, 2002.

K. S. Meacham. *The MRI Study Guide for Technologists*. Springer, 1st edition, 1995.

M. Meila and J. Shi. A random walks view of spectral segmentation. *AI and Statistics (AISTATS) 2001*, 2001.

K.-D. Merboldt, W. Hanicke, and J. Frahm. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance*, 64(3):479–486, 1985.

G. A. Miller. WordNet: a lexical database for English. *Commun. ACM*, 38(11): 39–41, November 1995. ISSN 0001-0782. doi: 10.1145/219717.219748.

B. Mirkin. *Clustering: a data recovery approach*. CRC Press, 2nd edition, 2012.

A. M. Morcom and P. C. Fletcher. Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage*, 37(4):1073–1082, Oct 2007. doi: 10. 1016/j.neuroimage.2006.09.013.

D. Mortazavi, A. Z. Kouzani, and H. Soltanian-Zadeh. Segmentation of multiple sclerosis lesions in MR images: a review. *Neuroradiology*, 54(4):299–320, 2012. ISSN 0028-3940. doi: 10.1007/s00234-011-0886-7.

H. Müller, P. Clough, T. Deselaers, and B. Caputo. Experimental evaluation in visual information retrieval. *The Information Retrieval Series*, 32, 2010. doi: 10.1007/978-3-642-15181-1.

M. Muller, A. P. Appelman, Y. van der Graaf, K. L. Vincken, W. P. Mali, and M. I. Geerlings. Brain atrophy and cognition: interaction with cerebrovascular pathology? *Neurobiology of aging*, 32(5):885–893, May 2011.

K. Nakamura and E. Fisher. Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. *Neuroimage*, 44(3):769–776, Feb 2009. doi: 10.1016/j.neuroimage.2008.09.059.

R. Navigli and S. P. Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193(0):217 – 250, 2012. ISSN 0004-3702. doi: 10.1016/j.artint. 2012.07.001.

A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, volume 2, pages 849–856. MIT Press, 2002.

F. Å. Nielsen. The Brede database: a small database for functional neuroimaging. *Ninth Annual Meeting of the Organization for Human Brain Mapping*, 2003.

J. Nolte. *The Human Brain: An Introduction to Its Functional Anatomy*. C.V. Mosby, 6th edition, 2008. ISBN: 978-0323041317.

R. Nowak. Wavelet-based Rician noise removal for magnetic resonance imaging . *IEEE Transactions on Image Processing*, 8(10):1408 –1419, oct 1999. ISSN 1057-7149. doi: 10.1109/83.791966.

P. L. Nunez and R. Srinivasan. *Electric Fields of the Brain : The Neurophysics of EEG*. Oxford University Press, USA, 2005.

K. Nybo, J. Venna, and S. Kaski. The self-organizing map as a visual neighbor retrieval method. In *Proc. of the Sixth Int. Workshop on Self-Organizing Maps*, 2007.

S. Ogawa, D. Tank, R. Menon, J. Ellermann, S. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13):5951–5, 1992.

J. Paakki, J. Rahko, X. Long, I. Moilanen, O. Tervonen, J. Nikkinen, T. Starck, J. Remes, T. Hurtig, H. Haapsamo, K. Jussila, S. Kuusikko-Gauffin, M. Mattila, Y. Zang, and V. Kiviniemi. Alterations in regional homogeneity of resting-state brain activity in autism spectrum disorders. *Brain Research*, 1321(0): 169–79, 2010. ISSN 0006-8993. doi: 10.1016/j.brainres.2009.12.081.

C. Pachai, Y. M. Zhu, J. Grimaud, M. Hermier, A. Dromigny-Badin, A. Boudraa, G. Gimenez, C. Confavreux, and J. C. Froment. A pyramidal approach for automatic segmentation of multiple sclerosis lesions in brain MRI. *Computerized Medical Imaging and Graphics*, 22(5):399–408, 1998. doi: 10.1016/ S0895-6111(98)00049-4.

D. Pelleg, A. W. Moore, et al. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *ICML*, pages 727–734, 2000.

D. L. Pham and J. L. Prince. Partial Volume Estimation and the Fuzzy C-means Algorithm. In *Proceedings of the International Conference on Image Processing (ICIP98)*, pages 819–822, 1998. doi: 10.1109/ICIP.1998.999071.

M. M. Poels, M. A. Ikram, A. van der Lugt, A. Hofman, W. J. Niessen, G. P. Krestin, M. M. Breteler, and M. W. Vernooij. Cerebral microbleeds are associated with worse cognitive function: the Rotterdam Scan Study. *Neurology*, 78 (5):326–333, Jan 31 2012.

R. Polikar. Ensemble learning. In *Ensemble Machine Learning*, pages 1–34. Springer, 2012.

V. Popescu, M. Battaglini, W. Hoogstrate, S. Verfaillie, I. Sluimer, R. van Schijndel, B. van Dijk, K. Cover, D. Knol, M. Jenkinson, F. Barkhof, N. de Stefano, and H. Vrenken. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D {T1} images in multiple sclerosis. *Neuroimage*, 61(4):1484 – 1494, 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2012.03.074.

M. F. Porter. An algorithm for suffix stripping. *Program: electronic library and information systems*, 14(3):130–137, 1980.

M. Prastawa, E. Bullitt, N. Moon, K. Van Leemput, and G. Gerig. Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic radiology*, 10(12):1341–1348, 2003.

M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman. A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):676–682, Jan 2001. doi: 10.1073/pnas.98.2.676.

R. Rangayyan. *Biomedical Image Analysis*. Biomedical Engineering. Taylor & Francis, 2004. ISBN 9780203492543.

W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on Medical Imaging*, 16(6):911–918, Dec 1997.

J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465 – 471, 1978. ISSN 0005-1098. doi: http://dx.doi.org/10.1016/0005-1098(78)90005-5.

A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and D. R. Millen. Getting our head in the clouds: Toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 995–998, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240775.

S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, M. Gatford, et al. Okapi at TREC-3. *NIST Special Publication*, pages 109–126, 1996.

M. Rosenthal, J. Cullom, W. Hawkins, S. Moore, B. Tsui, and M. Yester. Quantitative SPECT imaging: a review and recommendations by the Focus Committee of the Society of Nuclear Medicine Computer and Instrumentation Council. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 36(8):1489—1513, August 1995. ISSN 0161-5505.

D. E. Rumelhart and D. Zipser. Feature discovery by competitive learning. *Cognitive Science*, 9(1):75–112, 1985.

C. A. Sage, W. V. Hecke, R. Peeters, J. Sijbers, W. Robberecht, P. Parizel, G. Marchal, A. Leemans, and S. Sunaert. Quantitative diffusion tensor imaging in amyotrophic lateral sclerosis: Revisited. *Human Brain Mapping*, 30:3657–3675, Apr 2009. doi: 10.1002/hbm.20794.

P. Sahoo, S. Soltani, and A. Wong. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233 – 260, 1988. ISSN 0734-189X. doi: 10.1016/0734-189X(88)90022-9.

B. R. Sajja, S. Datta, R. He, M. Mehta, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Annals of Biomedical Engineering*, 34(1):142–151, Jan 2006. doi: 10.1007/s10439-005-9009-0.

G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513 – 523, 1988. ISSN 0306-4573. doi: 10.1016/0306-4573(88)90021-0.

G. Salton, A. Wong, and C.-S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

P. Santago and H. Gage. Quantification of MR brain images by mixture density and partial volume modeling. *IEEE Transactions on Medical Imaging*, 12(3):566–574, Sep 1993. ISSN 0278-0062. doi: 10.1109/42.241885.

S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.

P. Schmidt, C. Gaser, M. Arsic, D. Buck, A. Förschler, A. Berthele, M. Hoshi, R. Ilg, V. J. Schmid, C. Zimmer, B. Hemmer, and M. Mühlau. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage*, 59(4):3774 – 3783, 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2011.11.032.

H. Schnack, H. H. Pol, W. Baaré, W. Staal, M. Viergever, and R. Kahn. Automated separation of gray and white matter from MR images of the human brain. *Neuroimage*, 13(1):230–237, 2001. ISSN 1053-8119. doi: 10.1006/nimg.2000.0669.

B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

D. Schomer and F. da Silva. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincott Williams & Wilkins, 6th edition, 2011. ISBN 9780781789424.

F. Schroff, A. Criminisi, and A. Zisserman. Harvesting image databases from the web. In *IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007. doi: 10.1109/ICCV.2007.4409099.

M. L. Seghier, A. Ramlackhansingh, J. Crinion, A. P. Leff, and C. J. Price. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*, 41(4):1253–1266, Jul 2008. doi: 10.1016/j.neuroimage.2008.03.028.

F. Ségonne, A. Dale, E. Busa, M. Glessner, D. Salat, H. Hahn, and B. Fischl. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3): 1060–75, 2004.

H. Shah-Hosseini and R. Safabakhsh. TASOM: a new time adaptive self-organizing map. *IEEE Transactions on Systems, Man, and Cybernetics*, 33 (2):271–82, 2003.

D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage*, 13(5):856–876, May 2001. doi: 10.1006/nimg.2000.0730.

D. W. Shattuck, A. A. Joshi, C. Bhushan, A. C. Krause, J. P. Haldar, and R. M. Leahy. Brainsuite. http://www.fil.ion.ucl.ac.uk/spm/. visited 05/2014.

D. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. Narr, R. Poldrack, R. Bilder, and A. Toga. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–80, 2008.

S. Shen, A. J. Szameitat, and A. Sterr. An improved lesion detection approach based on similarity measurement between fuzzy intensity segmentation and spatial probability maps. *Magnetic Resonance Imaging*, 28(2):245–254, 2010. ISSN 0730-725X. doi: 10.1016/j.mri.2009.06.007.

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

T. Shi, M. Belkin, and B. Yu. Data spectroscopy: Eigenspaces of convolution operators and clustering. *The Annals of Statistics*, pages 3960–3984, 2009.

N. Shiee, P.-L. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage*, 49(2):1524–1535, Jan 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.09.005.

B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

S. Simon. *The Brain*. HarperTrophy, 1999.

J. Sinkkonen and S. Kaski. Clustering based on conditional distributions in an auxiliary space. *Neural Computation*, 14(1):217–239, 2002. doi: 10.1162/089976602753284509.

M. Sjöberg, J. Laaksonen, T. Honkela, and M. Pöllä. Inferring semantics from textual information in multimedia retrieval. *Neurocomputing*, 71(13–15):2576 – 2586, 2008. ISSN 0925-2312. doi: 10.1016/j.neucom.2008.01.029. Artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006).

J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, Feb 1998. doi: 10.1109/42.668698.

S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, Nov 2002. doi: 10.1002/hbm.10062.

S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23 Suppl 1:S208–S219, 2004. doi: 10.1016/j.neuroimage.2004.07.051.

A. Z. Snyder and M. E. Raichle. A brief history of the resting state: The Washington University perspective. *Neuroimage*, 62(2):902 – 910, 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2012.01.044.

R. Soares, H. Chen, and X. Yao. Semisupervised classification with cluster regularization. *IEEE Transactions on Neural Networks and Learning Systems*, 23(11):1779 –1792, nov. 2012. ISSN 2162-237X. doi: 10.1109/TNNLS.2012.2214488.

H. Soltanian-Zadeh, J. P. Windham, and A. E. Yagle. Optimal transformation for correcting partial volume averaging effects in magnetic resonance imaging. *Nuclear Science, IEEE Transactions on*, 40(4):1204–1212, 1993.

Q. Song and N. Kasabov. ECM-A novel on-line, evolving clustering method and its applications. In *In M. I. Posner (Ed.), Foundations of cognitive science*, pages 631–682. Citeseer, The MIT Press, 2001. doi: 10.1.1.18.3108.

A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013. ISSN 0278-0062. doi: 10.1109/TMI.2013.2265603.

T. Starck, J. Nikkinen, J. Rahko, J. Remes, T. Hurtig, H. Haapsamo, K. Jussila, S. Kuusikko-Gauffin, M. Mattila, E. Jansson-Verkasalo, D. Pauls, H. Ebeling, I. Moilanen, O. Tervonen, and V. Kiviniemi. Resting state fMRI reveals a default mode dissociation between retrosplenial and medial prefrontal subnetworks in ASD despite motion scrubbing. *Frontiers in Human Neuroscience*, 7 (802), 2013. ISSN 1662-5161. doi: 10.3389/fnhum.2013.00802.

S. M. Stigler. Francis Galton's account of the invention of correlation. *Statistical Science*, 4(2):73–79, 05 1989. doi: 10.1214/ss/1177012580.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 111–147, 1974.

M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells, and S. Warfield. 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. In *MICCAI 2008 Workshop, MIDAS Journal*, September 2008.

S. M. Sunkin, L. Ng, C. Lau, T. Dolbeare, T. L. Gilbert, C. L. Thompson, M. Hawrylycz, and C. Dang. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic acids research*, 41 (D1):D996–D1008, 2013.

M. Symms, H. R. Jäger, K. Schmierer, and T. A. Yousry. A review of structural magnetic resonance neuroimaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(9):1235–1244, 2004. doi: 10.1136/jnnp.2003.032714.

R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010. ISBN 1848829345, 9781848829343.

J. Talairach and P. Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System : an Approach to Cerebral Imaging*. Thieme Publishers Series. Thieme, 1988. ISBN 9783137117018.

A.-H. Tan et al. Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, volume 8, pages 65–70, 1999.

S. Tegginmath, R. Pears, and N. Kasabov. Ontologies and machine learning systems. In N. Kasabov, editor, *Springer Handbook of Bio-/Neuroinformatics*, pages 865–872. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-30573-3. doi: 10.1007/978-3-642-30574-0_49.

P. Theocharakis, D. Glotsos, I. Kalatzis, S. Kostopoulos, P. Georgiadis, K. Sifaki, K. Tsakouridou, M. Malamas, G. Delibasis, D. Cavouras, and G. Nikiforidis. Pattern recognition system for the discrimination of multiple sclerosis from cerebral microangiopathy lesions based on texture analysis of magnetic resonance images. *Magnetic Resonance Imaging*, 27(3):417–422, 2008. doi: 10.1016/j.mri.2008.07.014.

P. Tofts, editor. *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*. John Wiley & Sons, 2003.

R. Trivedi, R. K. Gupta, A. Agarawal, K. M. Hasan, A. Gupta, K. N. Prasad, G. Bayu, D. Rathore, R. K. S. Rathore, and P. A. Narayana. Assessment of white matter damage in subacute sclerosing panencephalitis using quantitative diffusion tensor MR imaging. *American Journal of Neuroradiology*, 27(8): 1712–1716, Sep 2006.

P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1):141–188, January 2010. ISSN 1076-9757.

P. E. Valk, D. L. Bailey, D. W. Townsend, and M. N. Maisey. *Positron Emission Tomography : Basic Science and Clinical Practice*. Springer, 2004.

S. Valverde, A. Oliver, M. Cabezas, E. Roura, and X. Lladó. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *Journal of Magnetic Resonance Imaging*, 2014. ISSN 1522-2586. doi: 10.1002/jmri. 24517.

K. E. Van De Sande, T. Gevers, and C. G. Snoek. Evaluating color descriptors for object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1582–1596, 2010.

R. J. van der Geest and J. H. Reiber. Quantification in cardiac MRI. *Journal of Magnetic resonance imaging*, 10(5):602–608, 1999.

D. C. Van Essen and D. L. Dierker. Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron*, 56(2):209–225, Oct 2007. doi: 10.1016/j. neuron.2007.10.015.

E. van Oort, A. van Cappellen van Walsum, and D. Norris. An investigation into the functional and structural connectivity of the default mode network. *Neuroimage*, 90:381–389, Apr 2014. ISSN 1053-8119. doi: 10.1016/j.neuroimage. 2013.12.051.

J. van Os and S. Kapur. Schizophrenia. *The Lancet*, 374(9690):635 – 645, 2009. ISSN 0140-6736. doi: 10.1016/S0140-6736(09)60995-8.

E. C. van Straaten, F. Fazekas, E. Rostrup, P. Scheltens, R. Schmidt, L. Pantoni, D. Inzitari, G. Waldemar, T. Erkinjuntti, R. Mantyla, L. O. Wahlund, F. Barkhof, and L. Group. Impact of white matter hyperintensities scoring method on correlations with clinical data: the LADIS study. *Stroke*, 37(3): 836–840, Mar 2006.

M. Vannier, R. Butterfield, D. Jordan, W. Murphy, R. Levitt, and M. Gado. Multispectral analysis of magnetic resonance images. *Radiology*, 154(1):221–4, 1985.

A. Vellido, P. Lisboa, and K. Meehan. Segmentation of the on-line shopping market using neural networks. *Expert Systems with Applications*, 17(4):303–314, 1999. ISSN 0957-4174. doi: 10.1016/S0957-4174(99)00042-1.

J. Vesanto and E. Alhoniemi. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11(3):586, May 2000.

N. X. Vinh, J. Epps, and J. Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11:2837–2854, 2010. ISSN 1532-4435.

A. Vovk, R. W. Cox, J. Stare, D. Suput, and Z. S. Saad. Segmentation priors from local image properties: without using bias field correction, location-based templates, or registration. *Neuroimage*, 55(1):142–152, Mar 2011. doi: 10. 1016/j.neuroimage.2010.11.082.

H. A. Vrooman, C. A. Cocosco, F. van der Lijn, R. Stokking, M. A. Ikram, M. W. Vernooij, M. M. B. Breteler, and W. J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage*, 37(1):71–81, Aug 2007. doi: 10.1016/j.neuroimage.2007.05.018.

T. Wager, M. Lindquist, T. Nichols, H. Kober, and J. Van Snellenberg. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage*, 45(1):S210–21, 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage. 2008.10.061. Mathematics in Brain Imaging.

R. L. Wahl, editor. *Principles and Practice of Positron Emission Tomography*. Lippincott Williams & Wilkins, 2002.

S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, Jul 2004. doi: 10.1109/TMI.2004.828354.

S. Webb. *The Physics of Medical Imaging*. Taylor & Francis, 1st edition, 1988.

W. R. Webb, W. Brant, and N. Major. *Fundamentals of Body Ct*. W.B. Saunders Company, 3rd edition edition, 2005.

R. Weinberg. *The Biology of Cancer*. Number v. 1 in The Biology of Cancer. Garland Science, 2007. ISBN 9780815340768.

S. M. Weiss, N. Indurkhya, T. Zhang, and F. Damerau. *Text mining: predictive methods for analyzing unstructured information*. Springer, 2010.

X. Wen, X. Wu, R. Li, A. S. Fleisher, E. M. Reiman, X. Wen, K. Chen, and L. Yao. Alzheimer's disease-related changes in regional spontaneous brain activity levels and inter-region interactions in the default mode network. *Brain Research*, 1509(0):58 – 65, 2013. ISSN 0006-8993. doi: 10.1016/j.brainres.2013. 03.007.

J. West, I. Blystad, M. Engström, J. B. M. Warntjes, and P. Lundberg. Application of Quantitative MRI for Brain Tissue Segmentation at 1.5T and 3.0T Field Strengths. *PLoS One*, 8(9):e74795, 09 2013. doi: 10.1371/journal.pone. 0074795.

C. Westbrook and C. Kaut. *MRI In Practice*. Blackwell Publishers, 1998.

S. Whitfield-Gabrieli, H. W. Thermenos, S. Milanovic, M. T. Tsuang, S. V. Faraone, R. W. McCarley, M. E. Shenton, A. I. Green, A. Nieto-Castanon, P. LaViolette, and et al. Hyperactivity and hyperconnectivity of the default network in Schizophrenia and in first-degree relatives of persons with Schizophrenia. *Proceedings of the National Academy of Sciences*, 106(4):1279–1284, Jan 2009. ISSN 1091-6490. doi: 10.1073/pnas.0809141106.

A. Wismüller, F. Vietze, J. Behrends, A. Meyer-Baese, M. Reiser, and H. Ritter. Fully automated biomedical image segmentation by self-organized model adaptation. *Neural Networks*, 17(8-9):1327 – 1344, 2004. ISSN 0893-6080. doi: 10.1016/j.neunet.2004.06.015.

C. S. Won and H. Derin. Unsupervised segmentation of noisy and textured images using markov random fields. *CVGIP: Graphical Models and Image Processing*, 54(4):308–328, 1992.

Y. Wu, S. K. Warfield, I. L. Tan, W. M. Wells, D. S. Meier, R. A. van Schijndel, F. Barkhof, and C. R. G. Guttmann. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *Neuroimage*, 32(3):1205–1215, Sep 2006. doi: 10.1016/j.neuroimage.2006.04.211.

T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665–670, 2011. doi: 10.1038/nmeth.1635.

J. Ylipaavalniemi and R. Vigário. Analyzing consistency of independent components: An fMRI illustration. *Neuroimage*, 39(1):169 – 180, 2008. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.08.027.

Q. Yu, S. M. Plis, E. B. Erhardt, E. A. Allen, J. Sui, K. A. Kiehl, G. Pearlson, and V. D. Calhoun. Modular organization of functional network connectivity in healthy controls and patients with schizophrenia during the resting state. *Frontiers in Systems Neuroscience*, 5, 2012. ISSN 1662-5137. doi: 10.3389/fnsys.2011.00103.

S. X. Yu and J. Shi. Multiclass spectral clustering. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 313–319. IEEE, 2003.

W. Zhan, Y. Zhang, S. G. Mueller, P. Lorenzen, S. Hadjidemetriou, N. Schuff, and M. W. Weiner. Characterization of white matter degeneration in elderly subjects by magnetic resonance diffusion and FLAIR imaging correlation. *Neuroimage*, 47(Supplement 2):T58 – T65, 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.02.004.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, Jan 2001. doi: 10.1109/42.906424.

L. Zhou, C. Gong, Y. Li, Y. Qiao, J. Yang, and N. Kasabov. Salient object segmentation based on automatic labeling. In M. Lee, A. Hirose, Z.-G. Hou, and R. Kil, editors, *Neural Information Processing*, volume 8228 of *Lecture Notes in Computer Science*, pages 584–591. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-42050-4. doi: 10.1007/978-3-642-42051-1_72.

Z.-H. Zhou. When semi-supervised learning meets ensemble learning. *Frontiers of Electrical and Electronic Engineering in China*, 6(1):6–16, 2011. ISSN 1673-3460. doi: 10.1007/s11460-011-0126-2.

X. Zhu. Semi-supervised learning literature survey. http://www.cs.wisc.edu/ jerryzhu/pub/ssl_survey.pdf, 2006. visited 12/2013.

A. P. Zijdenbos, R. Forghani, and A. C. Evans. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Transactions on Medical Imaging*, 21(10):1280–1291, Oct 2002. ISSN 1558-254X. doi: 10.1109/TMI.2002.806283.

R. A. Zimmerman, W. A. Gibby, and R. F. Carmody, editors. *Neuroimaging: Clinical and Physical Principles*. Springer, 2000.

M. Ángel González Ballester, A. P. Zisserman, and M. Brady. Estimation of the partial volume effect in mri. *Medical Image Analysis*, 6(4):389 – 405, 2002. ISSN 1361-8415. doi: 10.1016/S1361-8415(02)00061-0.

DISSERTATIONS IN INFORMATION AND COMPUTER SCIENCE

Aalto-DD110/2014 Virtanen, Seppo
Bayesian latent variable models for learning dependencies between multiple data sources. 2014.

Aalto-DD120/2014 Bergström-Lehtovirta, Joanna
The Effects of Mobility on Mobile Input. 2014.

Aalto-DD127/2014 Zhang, He
Advances in Nonnegative Matrix Decomposition with Application to Cluster Analysis. 2014.

Aalto-DD138/2014 Sovilj, Dušan
Learning Methods for Variable Selection and Time Series Prediction. 2014.

Aalto-DD144/2014 Eirola, Emil
Machine learning methods for incomplete data and variable selection. 2014.

Aalto-DD149/2014 Äijö, Tarmo
Computational Methods for Analysis of Dynamic Transcriptome and Its Regulation Through Chromatin Remodeling and Intracellular Signaling. 2014.

Aalto-DD156/2014 Sjöberg, Mats
From pixels to semantics: visual concept detection and its applications. 2014.

Aalto-DD157/2014 Adhikari, Prem Raj
Probabilistic Modelling of Multiresolution Biological Data. 2014.

Aalto-DD171/2014 Suvitaival, Tommi
Bayesian Multi-Way Models for Data Translation in Computational Biology. 2014.

Aalto-DD177/2014 Laitinen, Tero
Extending SAT Solver with Parity Reasoning. 2014.

With an ever increasing life expectancy, we see a concomitant increase in diseases capable of disrupting normal cognitive processes. Their diagnoses are difficult, and occur usually after daily living activities have already been compromised. This leads to serious costs in palliative care, and a significant decrease in quality of life.

This dissertation proposes a set of machine learning methods, useful for the study of the neurological implications of brain lesions. It focuses on the analysis and exploration of magnetic resonance images. Two main research directions are proposed. The first, a brain tissue segmentation approach, identifies brain tissues, and provides early lesion detection. The second, a document mining framework, is applied to neuroscientific publications. This framework allows for an intelligent harvesting and summarization of research results dealing with neural activity.

The aforementioned research directions are based on the retrieval of consistent information from neuroimaging data.

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

**DOCTORAL
DISSERTATIONS**