Computational methods towards image-based biomarkers and beyond

Evangelia I. Zacharaki

▶ **To cite this version:**

Evangelia I. Zacharaki. Computational methods towards image-based biomarkers and beyond. Medical Imaging. Université Paris-Est, 2017. tel-01648583

HAL Id: tel-01648583
https://inria.hal.science/tel-01648583

Submitted on 26 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HABILITATION À DIRIGER DES RECHERCHES
DE
L'UNIVERSITÉ PARIS-EST

HABILITATION À DIRIGER DES RECHERCHES

Spécialité: **Informatique**

présentée par

Evangelia I. Zacharaki

Computational methods towards image-based biomarkers and beyond

Soutenue le 9 mars 2017 devant le jury composé par

Rapporteurs:

Dellingette Hervé	Directeur de recherche, INRIA
Schnabel Julia	Professeure, King's College, Londres, Royaume Uni
Deutsch Eric	Professeur, Institut Gustave Roussy

Examinateurs:

Thiran Jean-Philippe	Professeur, EPFL, Suisse
Thirion Bertrand	Directeur de recherche, Université Paris Saclay (CEA)
Paragios Nikos	Professeur, CentraleSupélec, INRIA
Najman Laurent	Professeur, ESIEE Paris

Candidature à l' habilitation à diriger des recherches (HDR)

Evangelia I. Zacharaki (PhD'04, M.Eng'99), CentraleSupélec, Paris-Saclay

Avant exercé depuis un an au Centre de Vision Numérique (CVN), CentraleSupélec, comme enseignant-chercheuse, j'ai travaillé comme associé de recherche à l'Université de Pennsylvanie, États-Unis (2005 - 2009), aux départements de physique médicale, Université de Patras (UPatras), Grèce (2009 - 2012) et de génie informatique, UPatras (2012-2015). J'ai reçu une bourse d'études doctorales (pour 3 années) de la *Fondation des Bourses de l'État Grec (IKY)*, et une bourse Marie Curie IRG pour poursuivre des recherches postdoctorales (aussi pour 3 années).

Ces onze dernières années j'ai enseigné de nombreux cours en combinant la theorie et le travail pratique. Actuellement j'enseigne le cours *Bases d' Apprentissage Statistique* (Master, CentraleSupélec, 3h/semaine) et pendant 2009-2015 j'ai enseigné à l' Institut de Technologie de la Grèce Occidentale (TEI) (12h/semaine total en moyenne): *Vision par Ordinateur et l'Infographie*, *Programmation d'Ordinateurs (Pascal, QBasic, Fortran)*, *Systèmes d'Exploitation et des Logiciels*, *Circuits Électriques*, *Mesures Électriques*. Au cours de cette période, j'ai aussi eu la chance de co-encadrer six doctorants, qui ont produit des résultats publiés dans des articles scientifiques et conférences internationales d'excellence. Ms. Angeliki Skoura a soutenu sa thèse en 2014, M. Vasileios Kanas a soumis sa thèse (date prévue de soutenance: Janvier 2017), M. Evgenios Kornaropoulos a commencé la rédaction de sa thèse, tandis que trois doctorants sont en cours (Evangelia Pippa, Alexia Tzalavra, Guillaume Chassagnon).

Les recherches que je mène sont dans le domaine de l' **analyse d'images biomédicales ou des biosignaux**, un domaine qui a des liens étroits avec les mathématiques appliquées, l'informatique et la technologie biomédicale. J'ai co-rédigé en collaboration plus de 70 publications scientifiques (h-index = 15, g-index = 29). Je suis membre du comité de rédaction de l' *International Journal of Radiology* et de *Dataset Papers in Science* journal, tandis que j'ai été rédacteur en chef adjoint de la revue scientifique *Medical Physics*.

Au cours de mes recherches, les méthodes d'extraction de données ont été explorées et validées par des études de cas réels y compris une variété de données biomédicales, parmi lesquels des biosignaux (tels que les EEG, ECG), des images spatiotemporelles (CT, IRM, dynamique améliorée IRM), ainsi que des structures moléculaires (gènes et protéines). Les contributions de recherche comprennent

- l'avancement des algorithmes d'analyse d'image en utilisant les données acquises

dans la pratique clinique qui est généralement de grande dimension, rares, et très variable, et

- l'application de techniques d'apprentissage statistique pour l'extraction de biomarqueurs en ce qui concerne les résultats cliniques ou la prédiction de la maladie.

Les travaux de recherche peuvent se résumer en deux parties, (i) la découverte de connaissances dans les milieux non supervisés ou semi-supervisés où l'apprentissage est possible en observant des populations et (ii) l'exploration de données en milieu supervisé où les modèles sont apprises afin de relier les variables cibles (tels que le type de maladie) avec les données cliniques disponibles et de trouver ainsi des relations qui font progresser notre compréhension. Celle-ci conduit généralement à la construction de modèles personnalisés formés pour des applications cliniques particulières. Un défi majeur dans la création de modèles statistiques à partir des populations de sujets est la standardisation (normalisation) des données dans un espace de référence commun. Surtout en ce qui concerne les données d'imagerie de différents sujets avec une progression de la maladie, la normalisation spatiale est généralement un problème mal défini, donc des efforts considérables ont été déployés à cette direction de recherche.

Les cadres informatiques ont été principalement développés pour l'analyse des données présentant des troubles neurologiques, tels que les tumeurs cérébrales, la sclérose en plaques, les maladies cérébro-vasculaires, et de l'épilepsie.

En proposant ma candidature pour la qualification au Poste de Professeur des Universités, j'espère développer davantage mes activités d'enseignement et de recherche.

Abstract

Over the last decades the use of artificial intelligence techniques for mining biomedical data has been incredibly popular and has opened numerous opportunities to improve medical diagnosis. In the current thesis foundational data mining methodologies are explored and validated through real case studies capturing a variety of biomedical data, from one-dimensional biosignals (such as EEG, ECG), to spatiotemporal images (CT, MRI, dynamic-enhanced MRI), as well as molecular structures (genes and proteins). Research contributions include (i) the advancement of image analysis algorithms using data acquired in clinical practice that usually are high-dimensional, sparse, and highly variable, and (ii) the application of machine learning techniques for extraction of biomarkers in respect to clinical outcome or disease prediction. The work can be summarized into two parts, (i) knowledge discovery in unsupervised or semi-supervised settings where learning is possible by observing group populations and (ii) data mining in supervised settings where models are learnt in order to link the target variables (such as disease annotations) with the available clinical data and thus find relationships that advance our understanding. The latter leads usually to the construction of personalized models trained for specific clinical applications. The two parts are interconnected since knowledge extracted from group populations can be subsequently used as prior distributions to guide and regularize the solutions in personalized data-driven scenarios.

A major challenge in the creation of statistical models from populations of subjects is the standardization (normalization) of the data in a common reference space. Especially in respect to imaging data from different subjects with disease progression, spatial normalization is usually an ill-defined problem, thus a significant amount of work is devoted to this research direction. In respect to supervised learning, the standard data mining pipeline is followed which includes the data representation phase where features are extracted or data are transformed into a new (more compact) space, and the learning phase where a model is learnt aiming to reproduce the available data or find correlations between the patient's profile and the underlying clinical condition.

The computational frameworks were mostly developed for the analysis of data with neurological disorders, such as brain tumors, multiple sclerosis, cerebrovascular disease, and epilepsy.

Contents

Abstract i

1 Background **1**
1.1 Introduction 1
1.2 Biomedical Data 2
1.3 Information fusion 4
1.4 Machine learning and knowledge discovery 5
1.5 Personal contributions 6

2 Spatial alignment **15**
2.1 Introduction to temporal and spatial alignment 15
2.2 Image registration using continuous models 17
2.3 Image registration based on discrete (graphical) models 27

3 Data representation and dimensionality reduction **45**
3.1 Feature extraction and fusion 45
3.2 Embedding methods 55
3.3 Integration of clinical, morphological and imaging characteristics 59

4 Unsupervised and semi-supervised learning **69**
4.1 Statistical modeling of imaging phenotypes 69
4.2 Fuzzy clustering or random walks 79
4.3 Spectral clustering of waveforms 84
4.4 Pattern similarity in biognals and molecular data 85

5 Supervised learning **93**
5.1 Pattern detection and classification in biosignals 93
5.2 Segmentation and classification in medical images 96
5.3 Single- and multi-label classification of molecular structures 104

6 Ongoing and future research **109**
6.1 Development of new quantitative imaging biomarkers for obstructive and interstitial lung diseases 109
6.2 Deep learning for image segmentation 111
6.3 Deep learning for protein structure classification 112

7 Curriculum Vitae **119**

Chapter 1: Background

1.1 Introduction

Among the most important objectives of biology and medicine today is the detection of associations between phenotype and genotype using multilevel analysis. Also, the discovery of patterns and associations between morphology and function (normal or pathological) of different organs and the effect of genes and proteins is an ongoing research problem. The biomedical variables that are analyzed by researchers nowadays, originate from various levels beginning from cellular and molecular level and going up to the level of organism. As a result, processes such as medical diagnosis, prognosis and treatment have started to be based on a combination of such variables. Taking into account the main types of biomedical data, such as biosignals, medical images, and molecular data, the aim of my research was to contribute to the development of innovative algorithmic methodologies targeting the discovery of new biomedical knowledge. The planning of experiments or clinical trials and related questions, such as sample size determination for randomized designs, was not part of this research and thus is not covered in this thesis.

Given a large number of parameters observed in various experiments, the developed tools may assist researchers in biology and medicine to focus on important regions of images and sequences or on the differences, similarities and associations among the experimental parameters. In biology, the research outcomes may result in a better comprehension of structure and function of biological sequences, expression and function of genes and possibly discovery of biomarkers, proteins responsible for diseases, etc. In medicine, a more complete analysis of patterns and associations among morphology and function or pathology can contribute to a better understanding of normal and disease states or response to treatment. Moreover, the application of the developed data mining tools on the nervous system contributes to a better understanding of various neurodegenerative diseases. Finally, apart from the expected benefits in biology and medicine, our research may be useful in other scientific fields where image mining takes place, such as computer vision.

In the next sections, first some background aspects are presented in respect to the computational analysis of biomedical data, which are categorized according to data type into one-dimensional biosignals, spatiotemporal images and molecular data. Then, a summary of the personal contributions of the applied research projects will follow. Although each individual research project was studied separately, the performed research led to the creation of a complete array of computational methods unifying the extracted characteristics of the different biomedical data with a common aim, i.e. the extraction of biomarkers for disease characterization and clinical decision support.

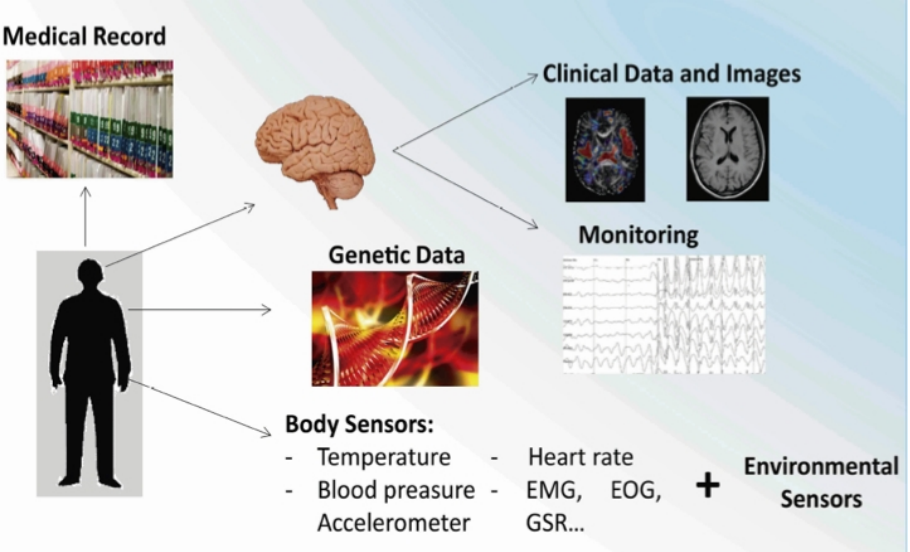


Figure 1.1: Multi-modal biomedical data analysis .

1.2 Biomedical Data

Biosignals

The biosignals are space-time recordings that measure body activity, such as heart rate, neural activity, blood pressure, respiration, blood glucose, and so forth. Biosignals are often examined for recognizable patterns that indicate change in pathology or underlying physiology and hence aid diagnosis [1]. The use of raw measurements for discriminating changes in underlying physiology is rarely efficient, thus signal analysis is usually performed aiming to represent the measured signals along a set of orthogonal basis functions (Fourier, wavelet, or principal components) for noise removal, pattern extraction, source separation (independent components), etc. Dimensionality reduction techniques are also applied aiming to find an appropriate representation of the data with reduced size for efficient storage, querying, streaming, and overall management [2].

The research work on biosignal analysis presented in this thesis focuses on the management and analysis of multimodal data from brain and body activities of epileptic patients and controls, such as multichannel Electroencephalography (EEG), Electrocardiogram (ECG), Electromyography (EMG). The aim was to design a personalized, medically efficient and economical monitoring system of patients with epilepsy. Due to its multifactorial causes and paroxysmal nature, epilepsy needs multi-parametric monitoring for purposes of accurate diagnosis, prediction, alerting and prevention, treatment follow-up and presurgical evaluation. Epileptic seizures differ with respect to motor, cognitive, affective and autonomic and EEG manifestations, thus their recognition and full understanding is the basis for the optimal management (including additional diagnostic tests and genetics) and treatment.

Medical Images

Medical imaging, like conventional magnetic resonance imaging (MRI), computerized tomography (CT), and functional MRI (fMRI) are commonly utilized in clinical practice and provide non-invasive representations of anatomy and function of living tissues. Prior to the search of imaging biomarkers that carry prognostic information on disease or therapy, certain computational steps might be necessary, such as image registration, segmentation, or modeling. The main part of the research presented in this thesis involves the development and application of such computational methods on medical imaging data, and is described analytically in the next chapters.

Briefly, image registration deals with the problem of morphological variability among subjects, motion and tissue deformation within the same subject (in different time points), or (multi-modal) information fusion. The aim is to calculate a mapping between homologous anatomical regions in two images (source and template); by applying this transformation on the source image, it becomes directly comparable to the template, making feasible any further (statistical) analysis. Image registration involves mainly three aspects: the choice of the transformation (global or local), the definition of a metric function (unimodal or multi-modal), and the optimization strategy to recover the parameters, i.e. by greedy or exhaustive search, or through estimation (discrete optimization, evolutionary computational methods).

Image segmentation deals with the problem of delineation of the boundary of certain structures, regions of interest, or the characterization of tissue. Often in the context of medical imaging, it combines prior knowledge (learnt geometric manifolds) with visual appearance. Data terms often exploit intensity homogeneity (acting either on the observation space, or a space derived from it), boundary discontinuities and more recently non-linear hypothesis separation.

Modeling anatomical and functional variability in medical images is achieved either using unsupervised or supervised methods. Unsupervised methods employ conventional dimensionality reduction techniques (PCA, ICA) or more recent embedding methods (LLE, ISOMAP, etc). In such a context, separation is achieved by looking into linear or non linear separation of the observations. Methods that have been used to classify and detect anomalies in medical images include among others wavelet transformation, fractal theory and statistical methods. In addition, methods have been proposed based on fuzzy set theory, Markov models and neural networks. Despite the significant advances in medical image analysis, the extraction of valuable biomarkers from medical images still remains an inexhaustible research field, as the methods have to be highly accurate and reproducible because they inherit high risk for people's health.

Molecular Data

The molecular data include the genetic expression profile, gene and protein sequences, proteomic structure and function. Our related research focused on the analysis of gene expression maps and classification protein structures. Previous work in the field of gene expression maps provides important insights on gene networks in unicellular systems using high-throughput multiplex gene expression methodologies, including microarrays, gene chips and serial analysis of gene expression. Although sophisticated techniques, such as voxelation, have been used in combination with microarrays for acquisition of genome-wide atlases of expression patterns (for the mouse brain), little has been done to take into account the location information of a gene's expression in brain.

Motivated by this idea, we analyzed gene expression maps, acquired by the technique of voxelation, using an atlas-based framework and employed the extracted spatial information to organize genes in significant clusters. Gene function enrichment analysis of clusters enabled exploration of the relationships among brain regions, gene expression and gene function.

Meanwhile, research in metagenomics, the field which combines the study of nucleotide sequences with their structure, regulation, and function, has been very productive the last years. While the number of newly discovered, but possibly redundant, protein sequences rapidly increases, experimentally verified functional annotation of whole genomes remains limited. Protein structure, i.e. the three-dimensional (3D) configuration of the chain of amino acids, is a very good predictor of protein function, and in fact a more reliable predictor than protein sequence. This is mainly because the chemistry required for the functionality of protein active sites arises from their 3D structure. Thus, as sequences diverge, only those residues required for the chemistry of the protein activity will be absolutely conserved whose 3D structure should also be conserved [3]. However, although the function of proteins is determined by their structure, external factors such as temperature increase or exposure to chemical denaturants might disrupt (unfold) the structure and cause loss of protein's activity [4]. These factors impose difficulties in the prediction of protein's function. By now, the number of proteins with functional annotation and experimentally predicted structure of their native state (e.g. by NMR spectroscopy or X-ray crystallography) is adequately large to allow learning training models that will be able to perform automatic functional annotation of unannotated proteins. Also, as the number of protein sequences rapidly grows, the overwhelming majority of proteins can only be annotated computationally, and this task was one of our latest research goals.

1.3 Information fusion

Although much progress has been achieved in acquisition of multimodal biological and medical data, most researchers still treat each modality separately, and integrate the results at the application stage. One reason for this is that the roles of multiple modalities and their interplay remain to be quantified and scientifically understood. Additionally, many open issues remain in processing each modality individually. Data fusion is a common technique that combines information from multiple sources in order to achieve more efficient and more accurate inferences than using a single source, for improving the quality of raw data (in pre-processing) or of the extracted knowledge and for handling data uncertainty and inconsistency [5]. There are two main approaches for fusing data from different modalities: feature level fusion and decision level fusion. In the decision level, features are classified for each modality by its local classifier and the results from these local classifiers are later fused in the decision layer. Characteristic examples of fusion based on integration of classifiers are the boosting and bagging techniques. In feature level fusion, the data is fused directly after feature extraction and classified by one global classifier [6].

Fusion of biosignals, or so-called multi-sensor data fusion, targets the combination of data collected by multiple sensors. From the neuroimaging point of view, the effective integration of data from different neuroimaging modalities, such as EEG, and fMRI, is currently considered to be an essential task in modeling human brain activity, aiming to achieve more accurate spatiotemporal patterns than is possible with any of the modalities alone. EEG has high temporal resolution

but limited spatial resolution unless the conductivity profile of each individual subject is known with high precision, a difficult and rather impractical task for routine applications. On the other hand, fMRI provides very good spatial but relatively poor temporal resolution. Several efforts recently attempted a full integration, often referred to as fusion, by inverting a generative model that explains all types of data. Such a model relates the same hypothetical neural causes to all types of data. In these approaches conflicting requirements must be resolved most notably between detail of the model and simplicity. On the one hand the model must accurately describe both the details that influence the diverse data recorded and the ones that the modeler is specifically interested in. On the other hand the model must be simple enough to avoid overfitting a large number of parameters and provide through the results a transparent view of the underlying system.

In the field of medical image fusion, the fusion of anatomical and functional information acquired by various modalities such as CT, MRI, digital subtraction radiography, positron emission tomography (PET), and single-photon emission computerized tomography (SPECT), has been proposed using anatomical atlases and warping models. A review summarizing the works in medical image fusion based on the applied fusion algorithm, imaging modalities, and regions (organs) of interest, can be found in [7]. The main challenge in image information fusion is the requirement of prior alignment of the data in the same space (more details on image registration in chapter 2). Recent works take advantage of the increase of data availability which can be used for training and propose decision level fusion frameworks that fuse local learners to completely bypass the need for accurate, yet computationally expensive, atlas-target registrations [8].

1.4 Machine learning and knowledge discovery

In the early nineties, biomedical knowledge discovery was mainly based on deductive, rule-based (expert) systems, as well as traditional statistical analysis. With increasing availability of data and advances in hardware and software, machine learning and artificial intelligence methods start to provide computer-assisted solutions for pattern discovery in complex data and unbiased inference. The most common techniques applied in biomedical research include artificial neural network (ANNs), Support Vector Machines (SVMs), Hidden Markov Models, and Bayesian approaches. Fig.1.2 (reproduced from [9]) shows the number of recent papers in PubMed which apply these supervised learning techniques in the biomedical domain [9].

In the recent years, there has been a drastic development of large-scale databases (genomics, proteomics, imaging) and of computational tools for analyzing large sets of data, and powerful methods for characterizing patients. This progress has opened new opportunities for biomarker identification and drug discovery. Moreover, the massive accumulation of data has given the opportunity to apply deep learning techniques, such as deep neural networks and deep belief networks. Particularly in computer vision, convolutional neural networks (CNN) have rapidly become the tool of choice for the analysis of images. Their still limited, use in biomedical research [10][11] has also shown very promising results.

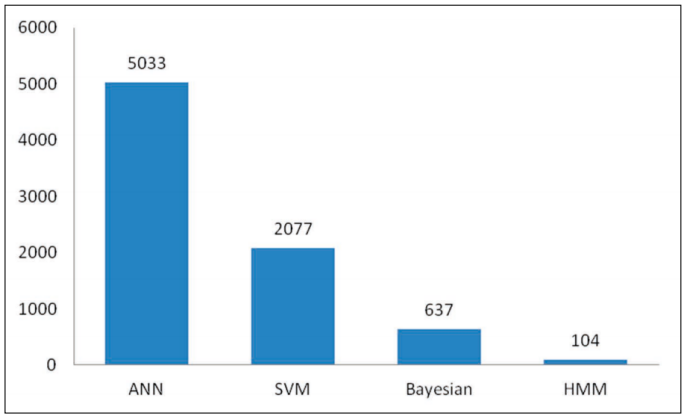


Figure 1.2: Popular machine learning techniques in biomedical research in the years 2011-2015 (from [9]).

1.5 Personal contributions

The main goal of the works presented in this thesis was to develop or apply computational methods for the analysis of biomedical data, and mainly medical images, in order to extract biomarkers of disease. This thesis is organized thematically rather than according to the temporal evolution of the research. Most of work was performed within small groups consisting of engineers or computer scientists and medical experts. In this thesis emphasis is given on research mostly leaded by myself in respect to the technical achievements. The clinical contribution including data annotation and clinical interpretation is attributed to our medical partners. Whenever my personal involvement was limited (usually reflected in publications where I am not within the first two authors or last author), I clarify my contribution at the end of the corresponding section.

The significance of work can be attributed to either technical contributions or novel (by the time of investigation) applications of computational methods aiming to solve specific biomedical problems. The key points are summarized next (including some representative references). More details are provided in the following chapters along with the rest of the studies not described here.

Technical contributions

Statistical modeling of high-dimensional data The aims of our work in [12][13][14][15] was to estimate the statistical variation of a normative set of volumetric images from healthy individuals, and then identify morphological or structural abnormalities as deviations from normality. A major issue during synthesizing valid images is the significantly larger dimensionality (number of voxels) compared to the usually available number of training data (images). Most anomaly detection methods in image processing extract only a few discriminatory features from regions of interest assuming data stationarity, thus do not suffer from the curse of dimensionality. Our purpose in this work was not develop an application-specific detector by modeling intensity-based or deformation-based class distributions, but to develop a general method of learning appearance

models in an unsupervised fashion. In order to deal with the large dimensionality, the method was based on partitioning the images into subspaces, i.e., locally coherent overlapping blocks. It was assumed that for each location the subspaces were generated from a Gaussian distribution and located in a tight cluster. As suggested also in other studies, high-dimensional data exhibit distributions that are highly sparse and can be represented by lower dimensional manifolds.

Under this scope, two fundamental contributions in discovering abnormality were made. In [12][15] the proposed method iteratively samples a large number of lower dimensional subspaces that capture image characteristics ranging from fine and localized to coarser and relatively more global. The marginal probability density functions pertaining to the selected features was estimated through a principal component analysis (PCA) model, in conjunction with an *estimability* criterion that limits the dimensionality of the estimated probability densities according to the available sample size and the underlying anatomy variation. A test sample is iteratively projected to the subspaces of these marginals as determined by the PCA models, and its trajectory until convergence delineates potential abnormalities (deviations from the normative database). Also, a *target-specific* feature selection strategy was introduced within each subspace to further reduce the dimensionality, by considering only imaging characteristics that are present in a test subject's image, rather than all possible characteristics found in the entire population [15].

A similar objective function for each local subspace was used in [13][14]. The main contribution in this work is that the local subspace estimates are fused into a globally optimal estimate that satisfies coupling constraints. Specifically, the maximum likelihood estimation problem was solved in a distributed setting using dual decomposition based on an algorithm developed for solving large-scale problems.

Registration of topologically non-equivalent images Most of the available registration methods in neuroimaging are designed to register a normal atlas with generally normal neuroanatomies. However there are many applications requiring the spatial normalization of images with pathology, such as brain images with tumors. Spatial normalization of such images could help to build statistical models of disease evolution, which could further guide neurosurgical treatment planning. Registration of images with tumors is especially challenging because some fundamental assumptions in image registration are violated: (i) the topological equivalence between the atlas and the patient's image, which is almost ubiquitous in deformable registration methods, is violated due to the anatomical changes caused by tissue death and tumor emergence, (ii) the confounding effects of edema and tumor infiltration cause changes in the image intensities and render the task of finding correspondences very difficult, and (iii) the large distortions caused by the mass effect of a growing tumor violate the usual assumption of deformation field smoothness. The developed methodology was based on the idea of first creating a topologically equivalent atlas image by replacing part of the healthy tissue with a tumor seed and then decoupling the total deformation (between atlas and patient's image) into two components, i.e., the deformation caused by the tumor mass effect and the deformation due to the inter-subject differences. We started with more simplified approaches (spherical tumors, homogeneous material, uniform deformation strategy) and gradually built into more elaborated frameworks (tumor seeds of arbitrary shape, optimized parameters) [16][17][18][19][20][21][22]. Overall the contributions of these methods involve (i) the estimation of the tumor-induced deformation by collecting statistics obtained from mass effect

simulations, (ii) the optimization of the tumor simulation parameters through coupling of mass effect models with registration, (iii) the formulation of an appropriate energy function for registration under the presence of uncertainties (i.e. around the tumor), and (iv) the acceleration of computations through asynchronous parallel pattern search.

High-order graphical models for image registration In the recent years, graphical models [23] have witnessed an enormous progress due to the development of efficient optimization/inference methods coupled with advanced machine learning techniques. In our works in [24][25][26][27] image registration was considered as a (undirected) graph optimization problem based on a Markov Random Fields (MRF) formulation in a discrete context. The aim in [25] was to solve jointly multi-atlas registration and segmentation. The agreement of the deformed segmentation masks with the underlying estimated segmentation was taken into account during the estimation of the deformation fields, leading to more accurate correspondences and consequently, improved segmentation. Furthermore, we aimed to improve the final segmentation by taking into account class specific appearance priors. The motivation behind incorporating prior segmentation probabilities lies in the fact that image registration is often trapped in local minima when trying to match areas of high anatomical variability (e.g., brain cortex). In such a setting, appearance information constitutes an alternative, more reliable, cue that can robustly guide segmentation. My personal involvement to these contributions (in [25]) was associated mainly to the overview of the method. Much more pronounced was my involvement in the works [26][27], which included the model formulation, energy cost definition and method evaluation. The aim in [26][27] was to co-register a group of images having intensity changes related to some tissue properties. The idea was to develop a group-wise registration method that incorporates a physiological model (reflecting the intensity change over time) as part of the image similarity definition [26]. Additionally, based on the assumption that the underlying parameter causing the spatial variation in intensity change should be spatially smooth itself, in [27] we introduced a high-order MRF model that simultaneously calculates the set of deformation fields and imposes spatial constraints on this parameter. Similarly to the joint registration-segmentation problem, our basic premise here was that by allowing the unknown variables (deformation and intensity) to interact, the produced representations become anatomically more meaningful.

Application-specific achievements

Extraction of imaging biomarkers for brain tumor assessment Infiltrative brain tumors, particularly high grade gliomas, have extremely poor prognosis, largely due to the fact that the neoplastic tissue has typically infiltrated beyond the treated tumor mass, without necessarily significantly changing imaging characteristics that are conventionally used to outline the tumor. Our work has explored multiparametric MRI, using an extensive protocol including conventional sequences in addition to perfusion and diffusion tensor imaging (DTI). We evaluated multiparametric voxel-wise imaging signatures that help to identify neoplastic and edematous tissue [28] or potentially predict future tumor recurrence [29], as well as more global-scale imaging characteristics of the tumors, captured by tumor shape and multi-scale texture properties, which differentiate between different tumor types [30][31] or predict the patient's survival [32]. The developed

computer-assisted tumor classification framework [31] could achieve higher accuracy than most reported studies using MRI; also, by exploring attribute selection jointly with classification, it investigated the diagnostic value of each image sequence. Moreover, the calculated prediction models based on data-mining algorithms in [32] could provide a more accurate predictor of prognosis in malignant gliomas than histopathologic classification (used in clinical practice as reference standard for tumor assessment).

Pattern recognition in biosignals As part of the ARMOR project [33] we have addressed the needs of epileptic patients and healthcare professionals, aiming at the design and development of a non-intrusive personal health system for the monitoring and analysis of epilepsy-relevant multi-parametric data, (i.e. EEG, ECG, EMG, skin conductance data) and the documentation of the epilepsy related symptoms. The similar seizure-like reactions of epileptic and non-epileptic events, such as psychogenic non-epileptic seizures and vasovagal syncopal attacks, make their diagnosis a difficult task. In clinical practice, the diagnosis is based on historical information assisted by specific tests. We performed online and offline analysis of data with the help of medical databases and the patient's medical file for the purpose of seizure detection/prediction [34][35][36] and differentiation between epileptic and non-epileptic events [37][38][39][40][41] for prevention and remote management. The processing and interpretation of the extracted patterns were used for accurate alerting and signaling of risks and for supporting healthcare professionals in their decision making [42][43].

Moreover, we studied the EEG brain activity during whole night sleep, since sleep is recognized as a major precipitator of epileptic activity and in many types of randomly occurring seizures, sleep EEG can be very revealing regarding the epileptic focus. In [44][45] a method for detecting spikes (epileptiform discharges) in interictal sleep EEG was developed, whereas in [46] an automatic scheme was proposed for the detection of a specific brain waveform, the K-complex, which is engaged in information processing, sleep protection, and memory consolidation.

Deep learning for protein structure classification In the past few years, data-driven deep learning models have become very popular because they tend to be domain agnostic and attempt to learn additional feature bases that cannot be represented through any handcrafted features. Building upon our recent work [47], we currently exploit experimentally acquired structural information of enzymes through deep learning techniques in order to produce models that predict enzymatic function based on structure [48]. The novelty of the proposed method lies first in the representation of the 3D structure as a bag of atoms (amino acids) which are characterized by geometric properties, and secondly in the exploitation of the extracted feature maps by deep convolutional neural networks. We hypothesize that by combining amino acid specific descriptors with the recent advances in deep learning we can boost model performance. Although assessed for enzymatic function prediction, the method is not based on enzyme-specific properties and therefore can be extended to other 3D molecular structures, thus providing a useful tool for automatic large-scale annotation.

Outline

Chapter 2 presents contributions in specific problems of medical image registration based on continuous or discrete optimization strategies. Chapter 3 addresses the problem of compact representation of high-dimensional biomedical data through feature extraction and embedding. Chapter 4 presents methods for abnormality segmentation in medical images, pattern similarity in biosignals and (co)-clustering in molecular data using unsupervised, or semi-supervised learning techniques, whereas Chapter 5 focuses on discriminative methods applied in a supervised fashion for detection and classification. Finally in chapter 6 ongoing research directions are discussed with the main interest in deep learning for tissue characterization and protein structure classification.

Bibliography

[1] M. J, *Biomedical signal analysis. standard handbook of biomedical engineering and design*, 2004.

[2] M. Krawczak and G. Szkatula, "Generalized nets model of dimensionality reduction in time series", in *Intelligent Systems'2014 - Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications*, 2014, pp. 847–858.

[3] J. S. Fetrow and J. Skolnick, "Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and t1 ribonucleases", *Journal of molecular biology*, vol. 281, no. 5, pp. 949–968, 1998.

[4] D. Dunaway-Mariano, "Enzyme function discovery", *Structure*, vol. 16, no. 11, pp. 1599–1600, 2008.

[5] B. Khaleghi, A. M. Khamis, F. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art", *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[6] P. Kumari and A. Vaish, "Feature-level fusion of mental task's brain signal for an efficient identification system", *Neural Computing and Applications*, vol. 27, no. 3, pp. 659–669, 2016.

[7] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art", *Information Fusion*, vol. 19, pp. 4–19, 2014.

[8] A. J. Asman, Y. Huo, A. J. Plassard, and B. A. Landman, "Multi-atlas learner fusion: An efficient segmentation approach for large-scale data", *Medical Image Analysis*, vol. 26, no. 1, pp. 82–91, 2015.

[9] L. Sacchi, J. Holmes, *et al.*, "Progress in biomedical knowledge discovery: A 25-year retrospective", *IMIA Yearbook*, 2016.

[10] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology", *Molecular Systems Biology*, vol. 12, no. 7, p. 878, 2016.

[11] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, "Medical image deep learning with hospital pacs dataset", *ArXiv preprint arXiv:1511.06348*, 2015.

[12] G. Erus, E. I. Zacharaki, N. Bryan, and C. Davatzikos, "Learning high-dimensional image statistics for abnormality detection on medical images", in *Computer Vision and Pattern Recognition Workshops (CVPR), 2010 IEEE Computer Society Conference on*, IEEE, 2010, pp. 139–145.

[13] E. I. Zacharaki and A. Bezerianos, "Segmentation of pathology by statistical modeling and distributed estimation", in *Biomedical Engineering, 2011 10th International Workshop on*, IEEE, 2011, pp. 1–3.

[14] ——, "Abnormality segmentation in brain images via distributed estimation", *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 3, pp. 330–338, 2012.

[15] G. Erus, E. I. Zacharaki, and C. Davatzikos, "Individualized statistical learning from medical image databases: Application to identification of brain lesions", *Medical image analysis*, vol. 18, no. 3, pp. 542–554, 2014.

[16] E. I. Zacharaki, D. Shen, A. Mohamed, and C. Davatzikos, "Registration of brain images with tumors: Towards the construction of statistical atlases for therapy planning", in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, IEEE, 2006, pp. 197–200.

[17] A. Mohamed, E. I. Zacharaki, D. Shen, and C. Davatzikos, "Deformable registration of brain tumor images via a statistical model of tumor-induced deformation", *Medical image analysis*, vol. 10, no. 5, pp. 752–763, 2006.

[18] E. I. Zacharaki, D. Shen, S.-K. Lee, and C. Davatzikos, "Orbit: A multiresolution framework for deformable registration of brain tumor images", *Medical Imaging, IEEE Transactions on*, vol. 27, no. 8, pp. 1003–1017, 2008.

[19] E. I. Zacharaki, C. S. Hogea, G. Biros, and C. Davatzikos, "A comparative study of biomechanical simulators in deformable registration of brain tumor images", *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 3, pp. 1233–1236, 2008.

[20] E. I. Zacharaki, C. S. Hogea, D. Shen, G. Biros, and C. Davatzikos, "Parallel optimization of tumor model parameters for fast registration of brain tumor images", in *Medical Imaging: Image Processing*, International Society for Optics and Photonics, vol. 6914, 2008, pp. 1–10.

[21] ——, "Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth", *Neuroimage*, vol. 46, no. 3, pp. 762–774, 2009.

[22] C. Davatzikos, E. I. Zacharaki, A. Gooya, and V. Clark, "Multi-parametric analysis and registration of brain tumors: Constructing statistical atlases and diagnostic tools of predictive value", in *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011, pp. 6979–6981.

[23] D. Koller and N. Friedman, *Probabilistic graphical models: Principles and techniques*. MIT press, 2009.

[24] N. Paragios, E. Ferrante, B. Glocker, N. Komodakis, S. Parisot, and E. I. Zacharaki, "(hyper)-graphical models in biomedical image analysis", *Medical Image Analysis*, 2016, in press.

[25] S. Alchatzidis, A. Sotiras, E. I. Zacharaki, and N. Paragios, "A discrete mrf framework for integrated multi-atlas registration and segmentation", *International Journal of Computer Vision*, pp. 1–13, 2016.

[26] E. Kornaropoulos, E. I. Zacharaki, P. Zerbib, C. Lin, A. Rahmouni, and N. Paragios, "Deformable group-wise registration using a physiological model: Application to diffusion-weighted mri", in *Image Processing (ICIP), 2016 IEEE International Conference on*, IEEE, 2016, pp. 1–5.

[27] ——, "Optimal estimation of diffusion in dw-mri by high-order mrf-based joint deformable registration and diffusion modeling", in *Biomedical Image Registration (WBIR) (in conjunction with CVPR), 2016 IEEE 7th International Workshop on*, IEEE, 2016, pp. 4321–4328.

[28] V. G. Kanas, E. I. Zacharaki, C. Davatzikos, K. N. Sgarbas, and V. Megalooikonomou, "A low cost approach for brain tumor segmentation based on intensity modeling and 3d random walker", *Biomedical Signal Processing and Control*, vol. 22, pp. 19–30, 2015.

[29] R. Verma, E. I. Zacharaki, Y. Ou, H. Cai, S. Chawla, S.-K. Lee, E. R. Melhem, R. Wolf, and C. Davatzikos, "Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of mr images", *Academic radiology*, vol. 15, no. 8, pp. 966–977, 2008.

[30] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme", *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[31] E. I. Zacharaki, V. G. Kanas, and C. Davatzikos, "Investigating machine learning techniques for mri-based classification of brain neoplasms", *International journal of computer assisted radiology and surgery*, vol. 6, no. 6, pp. 821–828, 2011.

[32] E. Zacharaki, N. Morita, P. Bhatt, D. O'rourke, E. Melhem, and C. Davatzikos, "Survival analysis of patients with high-grade gliomas based on data mining of imaging variables", *American Journal of Neuroradiology*, vol. 33, no. 6, pp. 1065–1071, 2012.

[33] *Advanced multi-parametric monitoring and analysis for diagnosis and optimal management of epilepsy and related brain disorders*, ARMOR (FP7-ICT-287720), 2011-11-01 to 2015-04-30. [Online]. Available: `http://cordis.europa.eu/project/rcn/101414_en.html`.

[34] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, M. Richardson, and V. Megalooikonomou, "Seizure detection using eeg and ecg signals for computer-based monitoring, analysis and management of epileptic patients", *Expert Systems with Applications*, vol. 42, no. 6, pp. 3227–3233, 2015.

[35] I. Mporas, V. Tsirka, E. Zacharaki, M. Koutroumanidis, and V. Megalooikonomou, "Evaluation of time and frequency domain features for seizure detection from combined eeg and ecg signals", in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, ACM, 2014, p. 28.

[36] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, and V. Megalooikonomou, "Online seizure detection from eeg and ecg signals for monitoring of epileptic patients", in *Artificial Intelligence: Methods and Applications*, Springer, 2014, pp. 442–447.

[37] E. Pippa, E. I. Zacharaki, M. Koutroumanidis, and V. Megalooikonomou, "Data fusion for paroxysmal events' classification from eeg", *Journal of Neuroscience Methods*, 2017, in press.

[38] E. Pippa, E. I. Zacharaki, V. Tsirka, M. Koutroumanidis, and V. Megalooikonomou, "Eeg-based classification of epileptic and non-epileptic events using multi-array decomposition", *International Journal of Monitoring and Surveillance Technologies Research*, 2016, in press.

[39] E. Pippa, E. I. Zacharaki, I. Mporas, V. Tsirka, M. P. Richardson, M. Koutroumanidis, and V. Megalooikonomou, "Improving classification of epileptic and non-epileptic eeg events by feature selection", *Neurocomputing*, vol. 171, pp. 576–585, 2016.

[40] V. G. Kanas, E. I. Zacharaki, E. Pippa, V. Tsirka, M. Koutroumanidis, and V. Megalooikonomou, "Classification of epileptic and non-epileptic events using tensor decomposition", in *Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Conference on*, IEEE, 2015, pp. 1–5.

[41] E. Pippa, E. I. Zacharaki, I. Mporas, V. Megalooikonomou, V. Tsirka, M. Richardson, and M. Koutroumanidis, "Classification of epileptic and non-epileptic eeg events", in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, IEEE, 2014, pp. 87–90.

[42] V. Megalooikonomou, D. Triantafyllopoulos, E. I. Zacharaki, and I. Mporas, "Cyberphysical systems for epilepsy and related brain disorders: Multi-parametric monitoring and analysis for diagnosis and optimal disease management", in, S. N. Voros and P. C. Antonopoulos, Eds. Springer International Publishing, 2015, ch. DSMS and Online Algorithms, pp. 271–279.

[43] ——, "Cyberphysical systems for epilepsy and related brain disorders: Multi-parametric monitoring and analysis for diagnosis and optimal disease management", in, S. N. Voros and P. C. Antonopoulos, Eds. Springer International Publishing, 2015, ch. Offline Analysis Server and Offline Algorithms, pp. 239–254.

[44] E. I. Zacharaki, I. Mporas, K. Garganis, and V. Megalooikonomou, "Spike pattern recognition by supervised classification in low dimensional embedding space", *Brain Informatics*, vol. 3, no. 2, pp. 73–83, 2016.

[45] E. I. Zacharaki, K. Garganis, I. Mporas, and V. Megalooikonomou, "Spike detection in eeg by lpp and svm", in *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, IEEE, 2014, pp. 668–671.

[46] E. I. Zacharaki, E. Pippa, A. Koupparis, V. Kokkinos, G. K. Kostopoulos, and V. Megalooikonomou, "One-class classification of temporal eeg patterns for k-complex extraction", in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, IEEE, 2013, pp. 5801–5804.

[47] A. Amidi, S. Amidi, D. Vlachakis, N. Paragios, and E. I. Zacharaki, "A machine learning methodology for enzyme functional classification combining structural and protein sequence descriptors", in *Bioinformatics and Biomedical Engineering*, Springer, 2016, pp. 728–738.

[48] E. I. Zacharaki, "Prediction of protein function using a deep convolutional neural network ensemble", *PeerJ Computer Science*, submitted, 2016.

Chapter 2: Spatial alignment

2.1 Introduction to temporal and spatial alignment

In biomedical research it is often required to combine data obtained (i) from different subjects in order to perform population analysis or statistical inference, (ii) from different time points or therapy stages for longitudinal analysis or treatment assessment, or (iii) from various modalities or sensors for information fusion. This process of data integration and comparison requires the data to be in the same domain, which is performed by (i) time synchronization in the case of biosignals [1][2], (ii) sequence alignment in the case of molecular structures, and spatial (or spatiotemporal) alignment in the case of volumetric (or dynamic) images.

Time synchronization of signals: The requirement for time synchronization is more common in sensory networks [3] where sensor nodes need to coordinate their operations and collaborate to achieve a complex sensing task, or after power-saving modes, when the nodes should sleep and wake up at coordinated times. Also, time synchronization might be required between different medical devices that interoperate with each other [2], or to improve body sensor networks energy efficiency, such as by exploiting physiological rhythm (e.g. heartbeat) information, instead of using periodic synchronization beacons [1]. Time synchronization is not required in offline exploration of biosignals since the data are temporally aligned during acquisition based on the time stamps of their frames. Thus time synchronization methods are not discussed in this thesis.

Sequence alignment: Sequence alignment is performed to identify similar regions between different DNA, RNA, or protein sequences that may be a consequence of functional, structural, or evolutionary relationships between the sequences [4]. Sequences can be aligned in a pairwise fashion to find homologues in databases, or they can be multiply aligned to visualize the effect of evolution across a whole protein family [5]. The pairwise alignment method can be based on similarity across the full extent of the sequences providing a ”global” alignment, or it can focus only on the regions where the similarity is present in some regions of the sequences providing a ”local alignment”. Global alignment is useful for evolutionary comparisons whereas local alignment helps in structural predictions, or comparison of sequences that share similarity only in a part of the sequence. Multiple alignments constitute an extremely powerful means of revealing the constraints imposed by structure and function on the evolution of a protein family. In our studies we performed local pairwise alignment of protein sequences for similarity assessment using standard algorithms. Since no contribution was made on the alignment algorithm itself, we don’t elaborate

more on this subject, but refer to [5] for a discussion on the potential strengths and weaknesses of the most widely used multiple alignment packages and to [1] for a more recent review of sequence alignment algorithms for next-generation sequencing.

Spatial alignment of images: The image alignment consists of establishing spatial correspondences between different images acquired by the same or different subjects, and is referred to as image registration. Image registration of different subjects in a common domain is also called spatial normalization. During the past two decades, much of the research in medical image analysis was devoted to image registration [6][7], producing a large number of free software solutions [8]. The goal of registration is to estimate the optimal transformation that maximizes an energy of the form

$$Sim(I_T, I_S \circ D) + R(D)$$

where Sim is the similarity (matching) criterion that quantifies the level of alignment between a target image I_T and a source image I_S using transformation D, and R is a regularization term that aims to favor any specific properties in the solution that the user requires, and seeks to tackle the difficulty associated with the ill-posedness of the problem. Regularization and deformation models are closely related. Thus, an image registration algorithm involves three main components: a deformation model, an energy function, and an optimization method [6]. The rest of this chapter will focus on deformable image registration in which a nonlinear dense transformation is sought, as opposed to linear or global, which is parametrized by a small number of variables and is inherently smooth. The need for deformable models is due to the fact that almost all anatomical parts, or organs of the human body are deformable structures. In deformable registration, where the number of variables is enormous, the solution space is usually reduced by evaluating the energy function over a set of control points, which might be defined on a regular grid or on salient locations. These points are moved in the direction that minimizes the energy cost, defining local deformations. Transformation between control points is propagated by interpolation. The energy cost is then minimized by an optimization method, which can be continuous or discrete, based on the nature of the variables used for inference. In continuous optimization problems the variables assume real values, whereas in discrete optimization problems the variables take values from a discrete set. Both classes of methods are constrained with respect to the nature of the objective function as well as the structure to be optimized. Heuristic and metaheuristic methods, on the contrary, can handle a wide range of problems and explore large solution spaces. However, they do not possess theoretical guarantees regarding the optimality of the solution.

This chapter continues first with our work on spatial normalization of medical images using deformation variables that take real values. Optimization was performed in a heuristic greedy strategy in which the locally optimal choice is made at each step. Being gradient free and intuitive, it was applied to tackle the problem of feature-driven image registration. The registration framework was developed with the ultimate goal of constructing statistical models of brain tumor evolution [9][10]. Statistical atlases have been used in a variety of studies of normal brain development and aging, as well as of brain diseases [11], but they have rarely been used in studies of brain cancer. Most of the available registration methods in neuroimaging [12] were designed to register a normal atlas with generally normal neuroanatomies. Direct application of these methods

to images of tumor patients can lead to poor registration around the tumor region, due to large deformations and lack of clear definition of anatomical detail in a patient's images.

Subsequently, the chapter continues with our work on image registration based on discrete optimization (section 2.3). The developed methods are based on a Markov Random Fields (MRF) formulation, where inference is expressed as a (undirected) graph optimization problem acting on a predefined graph structure (fixed number of nodes and connectivity) associated with a discrete number of variables. Two main directions are presented: (i) a generic method that integrates multi-atlas registration and segmentation [13], and (ii) a joint deformable registration and diffusion modeling approach aiming to improve estimation of the apparent diffusion coefficient (ADC) in diffusion-weighted MRI (DW-MRI) [14][15].

2.2 Image registration using continuous models

Deformable registration has been an active topic of research for over a decade. Its clinical applications are numerous. In particular, it is used for spatial normalization of functional images, group analysis, and statistical parametric mapping. It is also used in computational anatomy as a means for measuring structures, by adapting an anatomical template to individual anatomies, or as a means for image data mining in lesion-deficit studies, as well as in stereotaxic neurosurgery for mapping anatomical atlases to patient images. Many image analysis methodologies have been developed to tackle this issue [6], which fall in two general categories. The first family of methods involves feature-based matching, i.e., transformations that are calculated based on a number of anatomical correspondences established manually, semiautomatically, or fully automatically on a number of distinct anatomical features, such as distinct landmark points or curves and surfaces. The second family of methods is based on volumetric transformations, which seek to maximize the (voxel-wise) similarity between a source image and a template.

Our work builds upon the hierarchical attribute matching mechanism for elastic registration (HAMMER) [16] aiming to account for non homologous mappings, as in the case of images with pathology. HAMMER performs three-dimensional (3D) warping of brain images using feature-based similarity. It focused on minimizing the effect of local minima of the energy function being optimized, and on determining anatomically meaningful image correspondence. The presence of local minima is due primarily to three reasons: 1) the very high dimensionality of the space formed by the coordinates of all voxels in a volumetric image; 2) ambiguities inherent in the matching function being maximized and related to the practically uniform image intensities within each tissue type throughout the brain; and 3) the complex nature of brain anatomy. Let us assume that $I_T(x)$ is the intensity of the voxel $x \in V_T$ in the template image, and $I_S(y)$ is the intensity of the voxel $y \in V_S$ in the individual's image. The displacement field $d(x)$ defines the mapping from the coordinate system of the template I_T to the subject I_S, thus quantifies amount of movement, while the deformation field $D(x) = x + d(x)$ defines the mapping that transforms the template I_T into the shape of the subject I_S, i.e. stores the new point locations. HAMMER uses a sequence of lower dimensional energy functions to ultimately approximate the following multivariate energy

function to be minimized:

$$\mathcal{E} = \sum_{x \in V_T} w_T(x) \left(\frac{\sum_{z \in n(x)} e(z)(1 - Sim(a_T(z), a_S(D(z))))}{\sum_{z \in n(x)} e(z)} \right) + \sum_{y \in V_S} w_S(y) \left(\frac{\sum_{z \in n(y)} e(z)(1 - Sim(a_T(D^{-1}(z)), a_S(z)))}{\sum_{z \in n(y)} e(z)} \right) + \lambda \sum_x \left\| \bigtriangledown^2 d(x)) \right\| \tag{2.1}$$

The first energy term quantifies the dissimilarity of features (attributes) extracted from the template image and subject's image (a_T and a_S respectively) during the transformation $D(\cdot)$ from the template to the subject. For each template voxel x, the similarity of the attribute vectors is calculated in a spherical neighborhood $n(x)$, rather than on individual voxels, with each voxel z in the neighborhood contributing according to the weight $e(\cdot)$, which depends on the distinctiveness of the voxel, as will be described later. The second energy term is similar to the first, but it is defined in the subject domain and used to constrain the inverse transformation $D^{-1}(\cdot)$, from the subject to the template. The third term is a smoothness constraint on the displacement field $d(\cdot)$ used for regularization. The parameter λ controls the smoothness of the deformation field.

Registration of images with pathology

Building statistical models of the evolution of a disease, such as brain tumors, can help gain insight into the underlying physiological mechanisms. For example, population-based statistical atlases can potentially indicate whether a multi-parametric imaging, or proximity to certain fiber tracts, profile suggests higher likelihood of tumor progression in a particular direction. Moreover, augmenting these models with parameters related to the disease, such as tumor size and location relative to brain structures, could further improve predictive accuracy. The construction of such statistical models requires the integration of a variety of patient data, such as conventional MRI, perfusion, and DTI of a large number of patients, into the same space and also requires the linking of all these data to outcome measures. For this purpose, a registration method is needed that can map all the imaging data to a common space (normal atlas). Besides their potential value in predicting tumor progression, anatomical and functional statistical atlases or models are also useful in neurosurgical treatment planning, since they can integrate diverse information about anatomical and functional variability, thereby helping design treatment plans that minimize the risk for significant functional impairment of the patient or facilitate safe dose escalation. Here again, a method that can register a statistical atlas (image without disease) to the patient-specific brain tumor image, is required.

While the problem of co-registering brain images of healthy subjects has been addressed often in the literature, the normalization of tumor diseased images into a common template space, is still a very challenging problem that has motivated our work [17][18][9][19][20][10][21]. Most of the available registration methods in neuroimaging [12] are designed to register a normal atlas with generally normal neuroanatomies. In the images with tumor, the fundamental assumption of topological equivalence between the atlas and the patient's image, which is almost ubiquitous

in deformable registration methods, is violated due to the anatomical changes caused by tissue death and tumor emergence. Moreover, the confounding effects of edema and tumor infiltration, which cause changes in the image intensities, render the task of finding correspondences very difficult. Finally, the large distortions caused by the mass effect of a growing tumor violate the usual assumption of deformation field smoothness.

The framework we followed to facilitate the registration process was based on the creation of an atlas image that is as similar as possible to the patient's image, i.e. it contains a tumor and mass effect similar to those in the patient's image. The idea was to first create a topologically equivalent atlas image by replacing part of the healthy tissue with a tumor seed and then decoupling the total deformation (between atlas and patient's image) into two components, i.e., the deformation caused by the tumor mass effect and the deformation due to the inter-subject differences. The tumor-induced deformation can be calculated by a biomechanical model of soft tissue deformation [22], whereas the inter-subject deformation can be calculated by a deformable registration method. In other words, the tumor modeling component aims to resolve both the geometric discrepancies from the physiologic process of tumor growth and the image differences from the tumor emergence. The registration component is then based on the assumption that (i) there is equivalent image content between the atlas with simulated tumor and the patient's image, and (ii) the deformation between the tumor-bearing images is smooth, similar to normal image registration. Similar approaches of brain tumor image registration had already been proposed. However the approaches on that time were based on oversimplified assumptions, such as (i) tumor growth can be simply modeled as a radial expansion process, (ii) the tumor seed can be simply estimated by calculating the mass center of tumor or shrinking the tumor, or (iii) the morphological variability across individuals can be captured by an affine transformation. The work presented in [17][9][20][10][21], addresses all of these three issues, but special emphasis is given to the last two, since the first issue was addressed in previous work [22].

A first approach of our group for registering brain tumor images to a normal atlas, which avoids the previous simplifications, has been presented in [18]. We proposed a maximum likelihood framework for estimating the tumor model parameters by collecting statistics obtained from mass effect simulations. Briefly, statistical properties of the sought deformation map from the atlas to the image of a tumor patient were first obtained through tumor mass-effect simulations on normal brain images. This map was decomposed into the sum of two components in orthogonal subspaces, one representing inter-individual differences in brain shape, and the other representing tumor-induced deformation. For a new tumor case, a partial observation of the sought deformation map was obtained via deformable image registration and was decomposed into the aforementioned spaces in order to estimate the mass-effect model parameters. Using this estimate, a simulation of tumor mass-effect was performed on the atlas image in order to generate an image that is similar to tumor patient's image, thereby facilitating the atlas registration process.

Although this method reported promising results for small quasi spherical tumors, more experiments on different kind of tumors revealed some limitations, which motivated us to examine alternative methodologies. Firstly, in this method normal-to-normal registration was applied, as part of both the model parameters estimation and the final warping of the tumor-bearing images, whereas newer approaches [9][10] used a registration method developed for brain images under the presence of tumors (in corresponding or close locations to each other). Also, the method in [18]

was a statistical approach; it did not apply optimization and therefore did not provide any measure of confidence on the final solution (optimality criterion). Additionally, it retrieved the tumor model parameters based only on the tumor-induced deformation. Especially under the assumption of the brain being a homogenous material, that is, having the same material properties for white and gray matter, local information is lost. As an improvement, in [9][10], additionally to the information from the tumor-induced deformation fields, we incorporated local information from the image content, thus increasing the sensitivity of the optimization procedure. Finally, perhaps most importantly, the statistically-based method of [18] required a very large number of simulations of tumor growth, in order for shape statistics to be gathered. This renders this approach quite limiting in many practical situations.

Registration framework

Built upon the idea of HAMMER registration algorithm developed for normal brain registration [16], in [9][17] we present a framework for *Optimization (of tumor model parameters) and Registration of Brain Images with Tumors*, referred as ORBIT. The main advantages of the ORBIT algorithm are (i) the incorporation of a similarity criterion that uses two kinds of information, namely tissue properties and spatial location relative to tumor, and (ii) the development of a deformation strategy that is robust to unreliable matches caused by the presence of tumor. Moreover, one of the novelties of the proposed framework is the estimation of the optimal tumor-related parameters (including the origin of the tumor and the amount of tissue death), via optimization of a criterion reflecting the elastic stretching energy and the image similarity. Robustness is achieved by applying the optimization in a multi-resolution scheme. Also, efficiency is achieved by replacing the expensive non-linear biomechanical model with a PCA-based model of tumor growth, as well as by refining the warping throughout the optimization only in the regions of low confidence.

The basic components of ORBIT include: (i) a simulation model for tumor growth and mass effect, (ii) deformable registration for tumor-bearing images, and (iii) optimization/estimation of the parameters of the tumor growth and mass effect model. Fig.2.1 summarizes a closed-loop process for registering a normal brain atlas to a tumor-bearing image, using all three components. The illustrated mapping is useful for transferring information from the atlas to the patient's space. The normalization of the patient's image into a common atlas space can be performed by the reverse mapping.

The elastic deformation field is calculated according to the hierarchical approximation of an energy function, which consists of the similarity matching criterion defined in the template space, a constraint on the inverse matching, and smoothness constraints on the displacement field, following the general framework of the HAMMER algorithm. Critical parts of our formulation of the registration of tumor-bearing images are (i) the definition of the similarity criterion, (ii) the deformation mechanism, and (iii) the mechanism for improving the robustness of the method to slightly inaccurate estimates of the tumor simulation parameters.

Similarity function The deformation field that spatially warps the template to the patient's image was calculated by maximizing an attribute-based similarity criterion. The similarity of two voxels x and y was defined as the weighted summation of a similarity criterion matching the brain

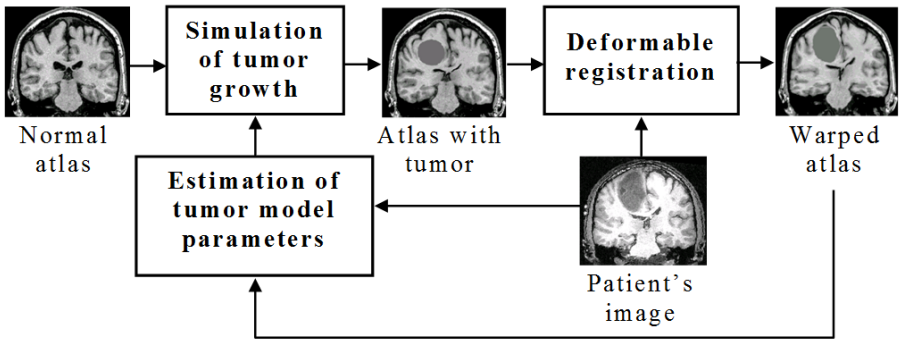


Figure 2.1: The ORBIT framework for registration of a normal brain atlas with an image with tumor. The total deformation is the composition of deformation caused by tumor growth and the deformation due to inter-subject variability.

structures, Sim_B, and a similarity criterion matching the tumor geometry, Sim_T, as given below:

$$Sim(x,y) = (1 - w(x,y)) \cdot Sim_B(x,y) + w(x,y) \cdot Sim_T(x,y)$$

where $w(x,y)$ is a weighting factor which decreases with the distance of x or y from the tumor. If at least one of the two images is normal (without tumor), w becomes zero and the similarity criterion matches only the brain structures, similar to HAMMER. The use of spatially adapted weights ensures that the identification of corresponding points is driven mainly by one of the two matching criteria, whereas the spatially smooth decrease of w makes the total similarity smooth. The two similarity criteria, Sim_B and Sim_T, reflected distance of attribute vectors defined for each voxel in the image. Sim_B was designed to match the brain structures by capturing the anatomical context around each voxel. It used edge type and tissue type information, as well as shape information based on geometric moment invariants (zero-order regular moments) from all tissue types. The tissue types include white matter (WM), gray matter (GM), ventricular cerebrospinal fluid (CSF), and cortical CSF. Sim_T quantified matching accuracy between the tumor in the patient's image and the simulated tumor in the atlas, and used as attributes the signed distance from the tumor boundary and the angular location with respect to the tumor center. Fig.2.2 demonstrates how such a similarity criterion can distinguish between different parts of a tumor-bearing brain image, which might otherwise be indistinguishable. Thus maximization of this similarity criterion would lead to the identification of correct anatomical correspondences.

Moreover, the optimal correspondence of a voxel is not determined by the similarity of only one voxel, but by integrating the similarity of all voxels within a small spherical neighborhood around this voxel. It is reasonable to expect that deforming larger regions of the image would result in more robust deformation schemes. This is especially important in medical image registration due to the highly nonconvex nature of the underlying energy functions resulting from the complex nature of anatomy and inter-individual differences. In fact, similarity matching is even more difficult when there are fundamental morphological differences between source and template image. In such cases where no good matches are found, the algorithm is designed to relax the matching forces, and the deformation is driven primarily by the deformations of the control

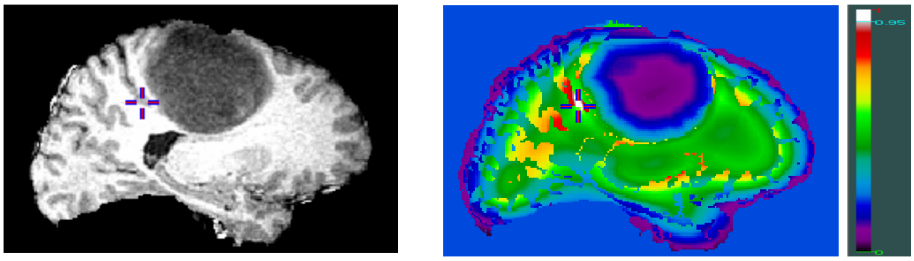


Figure 2.2: Distinctiveness of attribute vectors capturing anatomical and shape information in brain tumor images. The attribute-based similarity between the voxel indicated by a cross in left figure and every other voxel in the 3D data, is shown in right figure in color scale with white reflecting high similarity. For simplicity of illustration, edge type information is not included in the similarity calculation. The crosses correspond to the same coordinates in the two images.

points in the neighboring structures. Thus in order for a match to be enforced, high similarity of the attribute vectors must be present.

Image deformation mechanism The geometric transformation in the whole image domain was derived from interpolation theory, rather than by a physical model, such as elastic body, viscous fluid flow, diffusion, or curvature [23]. In interpolation theory, displacements considered known in a restricted set of locations in the image, are interpolated in the rest of the image domain. These models are rich enough to describe local displacements, while having low degrees of freedom, thus facilitating the inference of parameters. In ORBIT, similarly to HAMMER, Gaussian radial basis functions ϕ were used to find the displacement at an interpolation point x from the known displacements at points p_i:

$$d(x) = \sum_i \phi(\|x - p_i\|) d(p_i)$$

The points p_i that drive the deformation (control points) are not predefined (e.g. on a regular grid) but are selected hierarchically according to the distinctiveness of their attributes in order to reduce ambiguity in finding correspondence. Distinctive points lie for example on the roots of sulci, crowns of gyri, and strong isolated edges. The registration process starts by registering voxels with the most salient features and as the registration process proceeds, additional control points are selected to increase local accuracy. Especially for the tumor area, the selection of control points is not only based on the saliency of features, but also depends on the necessity of warping of the tumor volumes. When registration is used as part of the estimation process of the tumor model parameters, control points are selected on the tumor boundaries in order to facilitate the warping of the tumor volumes. Upon parameters estimation and tumor growth simulation, the registration is performed by relaxing all forces that prioritize the matching of the tumor boundaries. The reason is that the final registration should not be affected by (i) not accurately determined tumor boundaries and (ii) the residual variability in the tumor vicinity which is primarily due to fundamental differences in the growth process between a real and a simulated tumor.

The displacement of control points is interpolated in the neighborhood via a Gaussian kernel function. Upon interpolation of the displacement everywhere, a Laplacian-based smoothing is ap-

plied to ensure locally smooth displacement fields. The smoothing reduces with time, as the level of confidence in estimating the tumor model parameters increases. Also, registration is performed in a coarse to fine resolution scheme in order to speed up the algorithm, reduce susceptibility to local minima in both registration and estimation of tumor model parameters, and achieve robustness.

Learning a PCA-based model for tumor mass effect simulation The simulation of tumor mass effect is based on the application of a statistical model that describes the variation in deformability of the atlas brain due to different tumor model parameters. One way of modeling tumor growth and mass effect is via biomechanical simulators. Two examples of such simulators are based on (i) finite element models of nonlinear elasticity [22] and (ii) incremental linear elasticity models on regular grids [24]. However, incorporating such biomechanical models (especially the first one) into an iterative registration procedure could be computationally prohibitive. In [9], we have circumvented this limitation by using the mass effect model of [22] to train a PCA-based mass effect model and then using it in the registration framework. It is worth noting that the PCA-based tumor growth and mass effect simulation is extremely fast, since it can be achieved via linear combination of a relatively small number of principal components (deformations), thereby leaving the burden of simulation to off-line training using the costly biomechanical model. Our experiments have showed that this approach leads to very efficient approximation of the types of deformations caused by tumor growth, especially in view of the very approximating nature of any such modeling method.

In detail, consider the displacement maps $u_i(x)$, $i = 1, ..., N$, at the 3D Cartesian coordinates x due to the tumor mass effect, simulated by a biomechanical model with parameters θ_i (such as tumor seed location and size) in the atlas image. The displacement maps are first defined in a coordinate system x', centered at each tumor center, in order to make the domain of all the maps the same and allow point-to-point comparison and collection of statistics. The domain is restricted inside a region around each tumor center, where non-zero displacement due to mass effect is expected. Under the assumption of a Gaussian distribution, each $u_i(x')$ can be represented as

$$u_i = \mu + V z_i,$$

where μ the mean of the displacement at each voxel location, V the matrix containing the eigenvectors of the covariance matrix that correspond to the M largest eigenvalues, and z_i the vector with the M coefficients. The statistical parameters μ and V are determined from the training set. Then, for any new θ the displacement map $\hat{u}(x')$ can be calculated as $\hat{u} = \mu + V\hat{z}$, if $\hat{z}$ is known. If assume that the coefficients change smoothly for small variations of θ, we can approximate each coefficient in $\hat{z}$ by interpolating between the corresponding coefficients of the training samples in θ-space. This makes it possible to produce an estimator with continuous values of the model parameters. For this purpose, two scattered interpolation methods (based on either distance weighting or local fitting) were implemented and compared. The inverse distance weighting interpolation achieved the smallest reconstruction error and thus was chosen. Finally, the tumor-induced deformation map $\hat{u}(x)$ was calculated by re-centering the displacements at the original coordinate system.

Estimation of the tumor model parameters θ The optimal set of tumor growth parameters θ is not known for a particular patient; it must be estimated from the patient's image. The pattern of deformation around the tumor can be indicative of the accuracy in estimating θ. If the estimation of θ is wrong and thus the tumor is incorrectly simulated in the atlas, unrealistic and severe deformations are expected around the tumor region, when trying to match the atlas with the patient's image. Conversely, if tumor location and mass effect in the atlas are estimated in agreement with those in the patient, a relatively smooth deformation can be obtained. Additionally, due to the smoothness constraints applied during registration, the similarity of the two co-registered images is expected to be low around the tumor if the estimation of θ is inaccurate. Accordingly, we use the characteristics of the deformation field and the anatomical characteristics of the co-registered atlas and patient's images around the tumor to define an optimality criterion, E, for estimating θ. Specifically E is defined as the combination of three normalized measures: (i) the residual volume of overlap of the co-registered atlas and patient's images (E_1), (ii) the distance of attribute vectors (E_2), and (iii) the Laplacian of the deformation field that reflects smoothness properties of the deformation field (E_3), as mathematically given below.

$$\theta = \underset{\theta}{\operatorname{argmin}} \sum_{k=1}^{3} \sum_{x \in \Omega_E} c_k h_k(x) E_k(x; \theta)$$

The constants c_k are used to assign different weights on different measures, and $h_k(x)$ is used to assign different weights according to the voxel's location x. The constants c_k and the weighting functions $h_k(x)$ are learned by a heuristic strategy based on sensitivity criteria and confidence measures [9].

Application example An example of spatial normalization performed by ORBIT is shown in Fig.2.3. The tumor consists of an enhancing and a non-enhancing region, as illustrated by the T1-weighted image with contrast-enhancement. The non-enhancing region indicates the presence of edema or tumor infiltration. The patient's image is warped to the normal atlas space by reversing the deformation field produced by ORBIT. This warping causes relaxation of the mass effect and correction of the inter-subject differences facilitating the detection of the two tumorous regions: (i) the initial seed (as estimated by ORBIT) showing the tissue that is replaced by tumor, and (ii) the surrounding region that is infiltrated by tumor or edema.

Improving tumor simulation performance and estimation speed

In order to achieve equivalent image content between atlas and patient's image, tumor growth was simulated in the atlas image using subject-specific PCA models in the ORBIT registration framework [9]. This reduced the high computational cost of the finite element based biomechanical models for tumor growth simulation leaving the burden of simulations to off-line training. For same reasons a statistical approach for tumor growth simulation was also chosen in [18]. Statistical models, however, are not very accurate and also are limited by the parameters used during training. For example, training a model for irregularly shaped seeds would require an inhibitive large number of training cases. During that period, Hogea et al. [25] proposed a biomechanical model

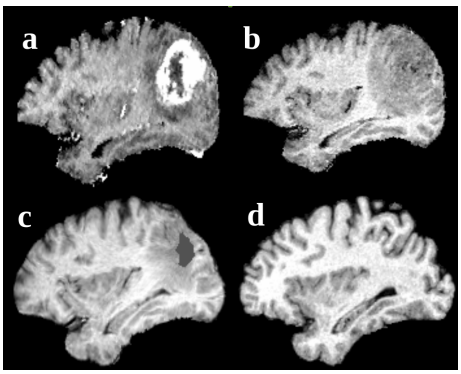


Figure 2.3: Spatial normalization example using ORBIT. The top row shows the T1-weighted patient's image (a) with and (b) without contrast-enhancement rigidly registered to the atlas in (d). The image in (b) after deformable registration to the normal atlas is shown in (c). The initial tumor seed representing tissue death is shown with gray color in (c) and indicates the location of initial tumor appearance, as defined in the atlas. The surrounding peri-tumor edema or infiltration, as mapped in the normal atlas, is also visible.

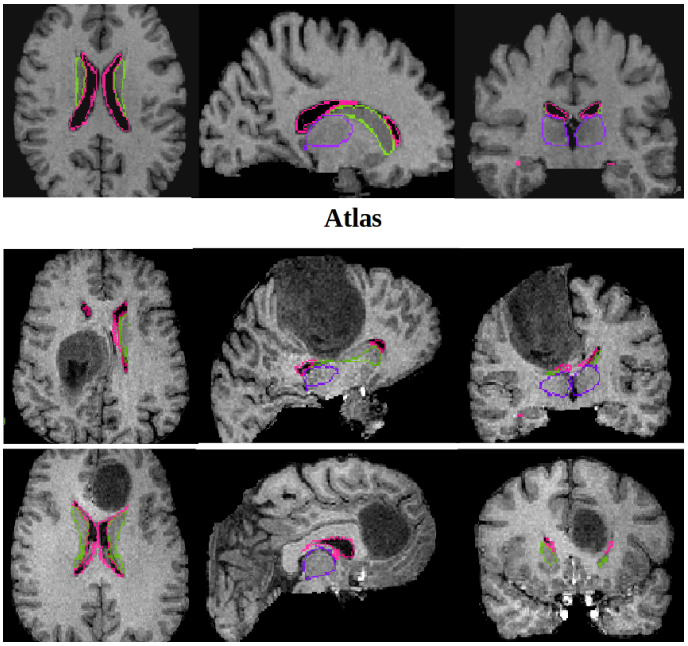


Figure 2.4: Atlas-based segmentation of brain images with tumor using the method in [10]. Contours of the thalamus, caudate nuclei, and ventricles, shown with lila, green, and pink, respectively, are registered and superimposed from the normal atlas (top row) to two patient's images (middle and bottom row). Each row shows the axial, sagittal and coronal view, as well as a zoomed snapshot of the coronal view.

developed in an Eulerian formulation and solved using regular grids, which is significantly faster than common finite element models. Thus, in [20][10] we employed this new tumor growth model within a model-based registration framework that simulates tumor emergence and tumor growth, and also simplistically differentiates between tumor mass effect and tumor infiltration. Also, we focused on increasing the speed of the estimation of the tumor model parameters by optimizing the objective function with the parallel optimization method, APPSPACK (Asynchronous Parallel Pattern Search) [26]. In detail, the new framework [10] added to our previous work [18][17][9], in the following aspects:

1. *Tumor simulation*: The new framework [10] uses a piecewise linear elasticity model and regular grids (PLE simulator), versus a finite element model of nonlinear elasticity and unstructured meshes in [18] and a local-PCA based model in [9]. PCA-based tumor growth simulations are extremely fast, since they are achieved via linear combination of principal components (deformations), thereby leaving the burden of simulations to off-line training using costly biomechanical models. However PCA-based simulations are not very accurate, since the expansion coefficients are not known in advance and can only be approximated. Moreover, PCA-based models are not flexible, since they can only reproduce deformations in the range of parameters used during training. On the other hand, nonlinear biomechanical simulators are flexible and more accurate, but also computationally very expensive. As shown in [19], the use of the robust and computationally efficient PLE simulator did not significantly affect the final registration accuracy in comparison to nonlinear biomechanical simulators. Therefore the PLE simulator seems to provide the best trade off between accuracy, flexibility and computational cost and was therefore chosen here. It should also be noted that the PLE simulator, is a stand-alone program which does not require the use of commercial packages, such as ABAQUS.

2. *Optimization of the tumor model parameters*: It applies the APPSPACK optimization package for parallel optimization (using mpi) of 5 tumor-related parameters, versus the Downhill Simplex method in [9] for serial optimization of 4 parameters, and a statistical approach in [18], not based on optimization, applied for a different sets of parameters (including edema). This allowed us to search over a large range of parameters in a computationally efficient way.

3. *Definition of the tumor model parameters*: It applies irregular shaped seeds, versus spherical seeds in [18]. The use of irregular seeds allows the creation of an atlas with tumor that is more similar to the subject's image.

4. *Registration method*: It maximizes a similarity criterion that matches both the brain structures and the tumor geometry using locally adapted weights, as in [9], versus the regular HAMMER registration algorithm developed for normal brains in [18]. Since the tumor in atlas and patient's image might be in non corresponding anatomical locations (e.g. if the estimate of the tumor location is inaccurate), it is not guaranteed that a diffeomorphic deformation field, that maps one image to the other, will exist. Therefore, the registration method in [10] and in [9], applies different deformation strategy close and far from the tumor, in order to maintain robust registration of the healthy part of the brain.

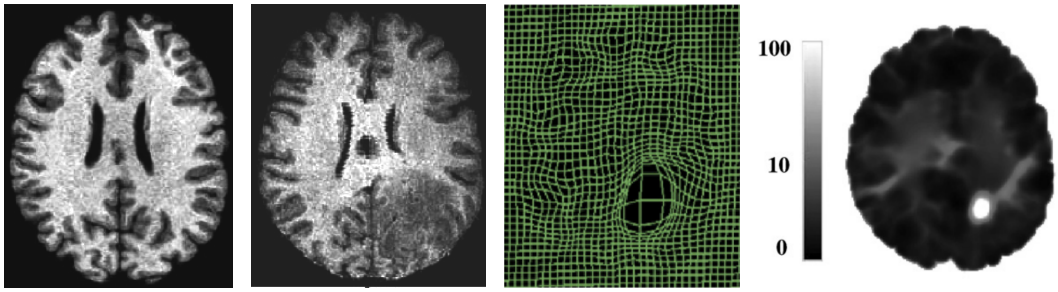


Figure 2.5: Mass effect estimation by the Jacobian determinant. A slice of the 3D volumetric data is illustrated for the (a) template image, (b) patient's image, (c) atlas-to-subject displacement field, (d) normalized Jacobian determinant (standard score).

Mass effect characterization A clinical application of this registration framework involves the quantitative characterization of the tumor mass effect [10]. The tumor mass effect has been used as a descriptor for classifying gliomas according to their clinical grade [27] or as an independent predictor of survival [28] and is therefore an important factor in the characterization of brain tumors. In this study, we investigate how well the estimated parameters (tumor seed and deformation field) help predicting the mass effect. In order to obtain an indicator of mass effect based on the proposed registration method, we evaluated how much the calculated deformation fields deviated from the range observed in a normal population. Specifically we calculated the Jacobian determinant of the deformation fields that spatially warp the atlas (used in this study) to the brain tumor images, as well as to a population of healthy subjects (60 individuals with age less than 68 years). All Jacobian images were smoothed with a Gaussian filter to reduce the noise and small mis-registration effects, and the voxel-wise standard score was calculated, in order to normalize the amount of deformation at each brain location, since the tumors appear in different locations in the brain for different subjects. The sum of the standard scores over the tumor seed defined in the common atlas space is a quantity that represents the distance between the total deformation and the population mean, and was used as an indicator of mass effect. The analysis was performed for 21 subjects and results were compared against visual ratings provided by two expert neuroradiologists. The correlation between our scores estimating tumor mass effect and the visual ratings, used as gold standard, was 0.763 and 0.618 for each of the two raters respectively, and 0.744 for the average ratings, whereas the correlation between the two visual ratings was 0.679. The relatively high correlation between the two measurements (through our modeling framework or by expert ratings) indicates that the measurements are consistent and therefore acts as a means of validation of our registration/estimation framework, which can be used to study the tumor growth process among populations of subjects based on reproducible and rater independent techniques.

2.3 Image registration based on discrete (graphical) models

In the recent years, graphical models [29] have witnessed an enormous progress due to the development of efficient optimization/inference methods coupled with machine learning algorithms and

the availability of large scale training data. While probabilistic graphical models have a variety of useful variants, here we will focus on a MRF formulation, where inference is often expressed as a graph optimization problem acting on a predefined graph structure (fixed number of nodes and connectivity) associated with a discrete number of variables [30].

A wide variety of tasks in medical image analysis can be formulated as discrete labeling problems. In very simple terms, a discrete optimization problem can be stated as follows: we are given a discrete set of variables $\mathcal{V}$, all of which are vertices in a graph $\mathcal{G}$. The edges of this graph (denoted by $\mathcal{E}$) encode the variables' relationships. We are also given as input a discrete set of labels $\mathcal{L}$. We must then assign one label from $\mathcal{L}$ to each variable in $\mathcal{V}$. However, each time we choose to assign a label, say, l_{p_1} to a variable p_1, we are forced to pay a price according to the so-called *singleton* potential function (unary) $g_p(l_p)$, while each time we choose to assign a pair of labels, say, l_{p_1} and l_{p_2} to two interrelated variables p_1 and p_2 (two nodes that are connected by an edge in the graph $\mathcal{G}$), we are also forced to pay another price, which is now determined by the so called *pairwise* potential function $f_{p_1p_2}(l_{p_1}, l_{p_2})$. Both the singleton and pairwise potential functions are problem specific and are thus assumed to be provided as input. Our goal is then to choose a labeling which will allow us to pay the smallest total price. In other words, based on what we have mentioned above, we want to choose a labeling that minimizes the sum of all the MRF potentials, or equivalently the MRF energy. This amounts to solving the following optimization problem:

$$\arg\min_{\{l_p\}} \mathcal{E}(g, f) = \sum_{p\in\mathcal{V}} g_p(l_p) + \sum_{(p_1,p_2)\in\mathcal{E}} f_{p_1p_2}(l_{p_1}, l_{p_2}). \tag{2.2}$$

The use of such a model can describe a number of challenging problems in medical image analysis. However these simplistic models can only account for simple interactions between variables, a rather constrained scenario for high-level medical imaging perception tasks. One can augment the expressive power of this model through higher order interactions between variables, or a number of cliques $\{C_i, i \in [1, n] = \{\{p_{i^1}, \cdots, p_{i|C_i|}\}\}$ of order $|C_i|$ that will augment the definition of $\mathcal{V}$ and will introduce hyper-vertices:

$$\arg\min_{\{l_p\}} \mathcal{E}(g, f) = \sum_{p\in\mathcal{V}} g_p(l_p) + \sum_{(p_1,p_2)\in\mathcal{E}} f_{p_1p_2}(l_{p_1}, l_{p_2}) + \sum_{C_i\in\mathcal{E}} f_{p_1\ldots p_n}(l_{p_{i^1}}, \ldots, l_{p_{i|C_i|}}). \tag{2.3}$$

where $f_{p_1\ldots p_{i|C_i|}}$ is the price to pay for associating the labels $(l_{p_{i^1}}, \ldots, l_{p_{i|C_i|}})$ to the nodes $(p_1, \ldots, p_{i|C_i|})$. Parameter inference, addressed by minimizing the problem above, is the most critical aspect in computational medicine and efficient optimization algorithms are to be evaluated both in terms of computational complexity as well as of inference performance. State of the art methods include deterministic and non-deterministic annealing, genetic algorithms, max-flow/min-cut techniques and relaxation. These methods offer certain strengths while exhibiting certain limitations, mostly related to the amount of interactions which can be tolerated among neighborhood nodes. In the area of medical imaging where domain knowledge is quite strong, one would expect that such interactions should be enforced at the largest scale possible.

The remainder of this chapter focuses on graph-based medical image registration/segmentation problems with personal involvement. A broader review on graphical models in biomedical image analysis can be found in [30].

Integrated multi-atlas registration and segmentation

Segmentation, or the process of assigning voxels to distinct anatomical regions or tissue types, is a fundamental task in medical image analysis. The accurate delineation of anatomical structures is the cornerstone of quantitative analysis that aims, among other, to understand normal and diseased anatomical variability. Important research efforts have been focused on developing automatic segmentation algorithms. Among the wealth of the developed techniques, segmentation via registration [31] stands as a unique example in medical image processing. Registration is used to map the grayscale image of an atlas to the query image, while the estimated deformation is subsequently used to propagate the available labels and provide an estimate of the segmentation. However, a single atlas is limited with respect to its ability to capture the variability of the population.

Extending registration-based segmentation by incorporating multiple atlases has thus emerged as a natural extension towards tackling the aforementioned limitation [32]. Multi-atlas segmentation, fueled by the maturity of the available registration techniques and advancements in computer hardware that partly alleviate its high computational cost, has gained significant popularity and found numerous applications in medical image analysis [33]. In multi-atlas segmentation, multiple atlas images are registered to the query image and their labels are propagate and fused, e.g. by majority voting, to produce the final labeling of the query image. A common principle behind most approaches is that registering the atlas images to produce candidate solutions and segmenting the query image are treated separately, in two independent steps. However, registration could benefit from taking into account the underlying segmentation towards establishing more accurate correspondences. Thus, approaches that treat registration and segmentation through fusion as an inter-weaved process have recently appeared.

In [13], we build upon the work presented in [34] by extending the formulation, providing a more detailed description of the method and reporting a more extensive validation setting. Specifically, in previous work [34] a discrete formulation based on Markov Random Field theory was introduced for integrated registration and label fusion segmentation. Latent variables included the displacements of the grid nodes of a B-Spline transformation model as well as voxel segmentation variables. Segmentation additionally took into account class likelihoods produced by a discriminatively trained classifier. Constraints were imposed by taking into account how congruent the proposed segmentations were with respect to the proposals of the rest of the atlases as well as the classifier produced likelihoods. Registration and segmentation variables were coupled by using an appearance-based weighting similar to the one used in local fusion strategies. As a consequence, as votes are weighted, atlases that do not match well locally would have a minor contribution to the inferred segmentation mask, resulting in a local ”soft” atlas selection scheme.

In [13], we estimate membership fields by additionally introducing a local atlas selection scheme. This scheme, explicitly models variables for selecting parts of each atlas labeling by comparing them directly to the estimated underlying segmentation. As a consequence, membership field images are directly produced by optimizing the model. The fact that the atlas selection is achieved by comparing segmentations, facilitates the use of images of different modalities to be part of the dataset.

Problem Formulation We consider a dataset of N annotated images $\mathbf{A} = \{A_0, \ldots, A_{N-1}\}$. Each image comes with a corresponding segmentation mask where the anatomical regions of interest have been annotated, forming the set $\mathbf{S} = \{S_0, \ldots, S_{N-1}\}$. Each voxel in the segmentation image is assigned to a segmentation label corresponding to one of M anatomical classes, $S_i(x) \in \{0, \ldots, M-1\}$. In this paper, we refer to an *atlas* as the aggregation of an intensity image A_i and its corresponding segmentation mask S_i.

Moreover, we consider that a (target or query) image I is given as input to be segmented into anatomical regions. The output of the proposed algorithms comprises: i) a set of membership field images $\mathbf{F} = \{F_0, \ldots, F_{N-1}\}, F_i(x) \in \{0, 1\}$ denoting if an atlas influences a point x in the query image; ii) the segmentation mask S_I corresponding to the target image; and iii) a set of deformation fields $\mathbf{D} = \{D_0, \ldots, D_{N-1}\}$, where D_i denotes the deformation field mapping of A_i to I.

The goal of the proposed method is to simultaneously solve for the parameters of the final segmentation S_I of the query image and the set of deformation fields $\mathbf{D}$. Hence, the agreement of the deformed segmentation masks with the underlying estimated segmentation may be taken into account during the estimation of the deformation fields, leading to more accurate correspondences and consequently, improved segmentation. Our basic premise is that *by allowing the two problems to interact, the quality of the respective solutions will rise due to the additional available information.*

Furthermore, we aim to improve the final segmentation S_I by taking into account class specific appearance priors. The motivation behind incorporating prior segmentation probabilities lies in the fact that image registration is often trapped in local minima when trying to match areas of high anatomical variability (e.g., brain cortex). In such a setting, appearance information constitutes an alternative, more reliable cue that can robustly guide segmentation [35].

Energy Function We formulate the problem as an energy minimization one. The proposed energy consists of three components: i) a registration component comprising a similarity matching term (Sim), that quantifies the level of alignment between each atlas and the query image, and a regularization term (R_d) that enforces the smoothness of the deformation field; ii) a segmentation component comprising an appearance prior term (S_P), that measures the log-likelihood of the segmentation with respect to the probabilities (π) learned during a training phase. iii) a coupling term (C) that takes into account the labeling that is proposed by the atlases, over the domain indicated by $\mathbf{F}$, and encourages their agreement with the estimated segmentation S_I. Finally, a regularization term (R_f) is imposed on the membership fields allowing the atlases to influence the derived segmentation in a smooth spatially varying fashion.

The continuous energy has the following form:

$$E(\mathbf{D}, \mathbf{A}, \mathbf{S}, S_I) = \underbrace{Sim(\mathbf{D}, \mathbf{A}, I) + R_d(\mathbf{D})}_{Registration} + \underbrace{S_P(S_I, \pi)}_{Segmentation} + \underbrace{C(\mathbf{D}, \mathbf{F}, I, \mathbf{S}, S_I) + R_f(\mathbf{F})}_{Coupling} \quad (2.4)$$

The first two energy terms correspond to the standard energy that is commonly minimized in multi-atlas segmentation frameworks, while the third term is common in segmentation frameworks. The fourth term introduces the main novelty of this work, i.e. the coupling between the segmentation and the multi-atlas registration. Let us now detail each term of the previous energy.

MRF Graph Structure We use Markov Random Field (MRF) theory to formulate the above minimization problem in a discrete context. The constructed graph encode the multi-atlas registration, the segmentation and the constraints that integrate the two problems. The problem is represented by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of nodes that encode the latent variables, and $\mathcal{E}$ the set of edges that encode the interactions between the variables.

- *Multi-Atlas Registration*: The deformation model is parametrized by N deformation grids. This is encoded in the MRF graph $\mathcal{G}$ by a set of N isomorphic grid graphs $\mathcal{G}_D = \{\mathcal{G}_{D_0}, \ldots, \mathcal{G}_{D_{N-1}}\}$. For every control point in the deformation grid that is superimposed onto image A_i, there is a node $p_i \in \mathcal{V}_{D_i}$ that represents its displacement. Since grids are isomorphic, p indexes a common control point position, while i indexes the grid. The edge system of each grid $\mathcal{E}_{D_i}$ is created by assuming a regular 6-connectivity scheme. The label set $\mathcal{L}_D$ for this set of variables is a quantized version of the displacement space. A label assignment $l^d_{p_i} \in \mathcal{L}_D$ (with $p_i \in \mathcal{V}_{D_i}$) is equivalent to displacing the control point p_i by displacement $\mathbf{d_{p_i}}$.

- *Segmentation*: An additional graph $\mathcal{G}_S = (\mathcal{V}_S)$ is employed to model segmentation. A node $p_s \in \mathcal{V}_S$ encodes a random variable and corresponds to a voxel in the target image whose position is indexed by the subscript s. We should also emphasize the fact that the nodes that form the segmentation graph are **not** connected to one another. The set of possible solutions $\mathcal{L}_S$ represents the set of anatomical regions augmented by the background label. We refer to a potential anatomical label in $\mathcal{L}_S$ by l^s.

- *Coupling*: Integrating segmentation and multi-atlas registration is achieved by coupling segmentation and deformation graphs. The set of edges $\mathcal{E}_C$ connects nodes of $\mathcal{V}_S$ with nodes of $\mathcal{V}_D$. In order to create the coupling edge system, we connect every node $p \in \mathcal{V}_D$ with nodes of $\mathcal{V}_S$ that correspond to voxels that are influenced by a displacement of p.

- *Local Atlas Selection*: We parametrize membership fields by taking into account the spatial support of the deformation nodes. Voxels within the support of a control point share the same membership state ([0, 1]). To model this, we augment the label set of the deformation nodes by considering the Cartesian product between the deformation label set $\mathcal{L}_D$ and a binary selection label set $\mathcal{L}_E = \{0, 1\}$. Thus, for a node p a label l^d indexes a pair $(\mathbf{d_p}, e^d_p)$, where $e^d_p \in \mathcal{L}_E$. A node p is *selected* when $e^d_p = 1$, otherwise it is *deselected*. If p is deselected, it will not penalize inconsistent candidate segmentations and it will not be influenced by them.

MRF Energy The continuous energy in Eq. 2.4 is mapped to a discrete MRF energy of the form in Eq. 2.2. In short, we map i) the matching term Sim to the unary potentials of the deformation variables (Eq. 2.5), ii) the deformation smoothness penalty term R_d to pairwise potentials between deformation variables (Eq. 2.6), iii) the segmentation prior S_P to the unary potential of the segmentation variables (Eq.2.7), iv) the coupling penalty C to one pairwise potential between registration and segmentation variables and one unary potential over deformation variables (first and second part of the right hand side of Eq. 2.9, correspondingly), and finally v) the membership

field smoothness penalty term R_f to an additional pairwise potential between deformation nodes (Eq. 2.10).

- *Multi-Atlas Registration*: Multi-atlas registration is performed by registering in a pairwise fashion all atlases to the target image. Formulating pairwise registration in a discrete setting has been shown in [36]. For completeness reasons, we briefly discuss here how the matching term M and the regularization term R of Eq. 2.4 are mapped to unary and pairwise potentials.

  Concerning the matching term, we are interested in quantifying how well the assignment of a displacement label $l^d_{p_i} \in \mathcal{L}_D$ to a node $p_i \in \mathcal{V}_{D_i}$ aligns atlas A_i to the target image. This is measured by the following unary potential:

$$g^M_{p_i}(l^d_{p_i}) = \int_\Omega \hat{\omega}_{p_i}(x)\rho(A_i \circ D_{p_i}, I(x))dx. \tag{2.5}$$

  D_{p_i} is the transformation induced by the movement of the control point p in the i-th deformation grid by the displacement $l^d_{p_i}$. The weighting function $\hat{\omega}_{p_i}$ determines the contribution of the point x to the unary potential of the control point p.

  Regarding the regularization term, [36] shows that it can be efficiently modeled by pairwise potentials. A discrete approximation of the gradient of the spatial transformation can be computed by taking the vector difference between the displacements of neighboring nodes that belong to the same deformation grid:

$$f^R_{p_iq_i}(l^d_{p_i}, l^d_{q_i}) = \|\mathbf{d}_{p_i} - \mathbf{d}_{q_i}\|, \tag{2.6}$$

  where $\mathbf{d}_{p_i}$ is the displacement applied to node p in the i-th deformation grid, indexed by l^d_{pi}.

- *Segmentation*: In order to assign a class label to every voxel of the target image, we take into account learned appearance model for every class. The appearance model is encoded in the form of a probability distribution $\pi(l)$ and can be naturally incorporated in the MRF model by setting the unary potentials of the segmentation grid for every label to the negative log-probability of the respective class:

$$g^{SP}_{q_S}(l^s_{q_S}) = -log(\pi(l^s_{q_S})). \tag{2.7}$$

- *Integrated Segmentation and Multi-Atlas Registration*: We want to encourage the agreement between the estimated segmentation and the warped segmentation masks. Thus, we penalize control point displacements of grid $\mathcal{G}_{D_i}$ that result in warping the segmentation mask of the corresponding atlas i in a fashion that does not agree with our final segmentation:

$$f^C_{p_iq_S}(l^d_{p_i}, l^s_{q_S}) = \hat{\omega}_{p_i}(s) \cdot Ind(S_i \circ D_{p_i}(s), l^s_{q_S}), \tag{2.8}$$

  where p_i belongs to the grid $\mathcal{G}_{D_i}$, q_S belongs to $\mathcal{G}_S$ and $Ind(x, y)$ equals 0 if $x = y$, otherwise it equals 1.

Due to the sparse way we model selection variables, any candidate deformation $\mathbf{d}_p$ corresponds to a segmentation mask that agrees at a certain percentage to the segmentation variables within the support of p. In order to model local atlas selection, we need to determine a set of such percentages below which a candidate segmentation is unacceptably incongruent with the consensus segmentation, and thus should lead to a node being deselected. We call these percentages *agreement percentages* and index them with a_p. Note that the agreement percentage does not depend on the grid index i but only on the inter-grid control point index p. Agreement percentages are estimated by locally comparing each segmentation mask with an estimate of the consensus segmentation, followed by a clustering scheme to arrive at robust per control point estimates (more details in [13]). Given an agreement percentage a_p, atlas selection may be enforced by introducing an additional unary cost for deformation nodes. Thus the total coupling cost becomes

$$f^{A_S}_{p_i q_S}(l^d_{p_i}, l^s_{q_S}) = e^d_{p_i} \cdot f^C_{p_i q_S}(l^d_{p_i}, l^s_{q_S}) + (1 - e^d_{p_i}) \cdot (1 - a_p), \tag{2.9}$$

where $e^d_{p_i}$ is equal to 1 when p_i is enabled, and 0 otherwise. As a consequence, nodes for which the (pairwise) segmentation cost f^C is very high when warping an atlas i, are disabled for this atlas. When a node p_i is disabled the cost to be paid is $1 - a_p$ regardless of the level of disagreement of the deformation with its corresponding segmentation nodes. Thus, a disabled deformation node will not affect segmentation variables and conversely will not be affected by them.

Furthermore, membership fields are encoded over deformation nodes. We want to enforce smoothness over the fields to achieve concise regions of influence for every atlas:

$$f^S_{p_i,q_i}(l^d_{p_i}, l^d_{q_i}) = Ind(e^d_{p_i}, e^d_{q_i}). \tag{2.10}$$

To conclude the outline of the discrete energy, we summarize the terms along with the parameters controlling each term's weight:

$$E_{MRF}(\mathbf{l}) = g^M_{p_i}(l^d_{p_i}) + \lambda f^R_{p_i q_i}(l^d_{p_i}, l^d_{q_i}) + \sigma g^{SP}_{q_S}(l^s_{q_S}) + \alpha f^{A_S}_{p_i q_S}(l^d_{p_i}, l^s_{q_S}) + \beta f^S_{p_i,q_i}(l^d_{p_i}, l^d_{q_i}),$$

λ controls the deformation field smoothness, σ encodes the prior likelihood weight, α specifies how much registration influences segmentation and reversely how much segmentation affects registration, and β regulates the smoothness of membership fields.

MRF Optimization through Dual Decomposition DD-MRF [37] has been introduced as a framework for MRF optimization, offering global optimality guarantees. Its flexibility in terms of possible energy types, its ability to report the quality of the final solution as well as its optimality guarantees are the merits we considered in opting for its use. DD-MRF works by receiving as input a decomposition of the initial graph (primal problem) into subgraphs (dual problems). It initializes the costs of the dual problems using the costs of the primal problem. It then proceeds by iteratively finding a global optimum for each subproblem, compare the subproblem solutions and update their costs.

In short, at each iteration: i) dual subproblems are solved in an optimal fashion (usually by dynamic programming); ii) the dual energy, defined as the sum of the energies of the optimal dual solutions, is computed; iii) a solution to the primal is inferred by the multitude of, possibly conflicting, dual solutions; iv) the difference between the energy of the primal and the sum of the duals is computed (referred to as primal-dual gap); v) the primal dual gap is used to update subproblem costs. This way agreement is induced between the next subproblem solutions leading iteratively to a coherent, globally optimal solution. The way dual costs are updated guarantees that the euclidean distance of the current solution to the set of the globally optimum solutions will decrease monotonically.

Validation We have validated our method using leave-one-out cross-correlation over publicly data available on the Internet Brain Segmentation Repository (IBSR). We specifically used the skull stripped version of the dataset provided in [38] after linearly registering images and masks. Assessment was based on labels of 14 symmetric and 3 non-symmetric brain structures that were annotated in more than half of the images of the dataset. The method was compared against: i) pairwise registration fused using majority voting , ii) pairwise registration fused using appearance-based locally (over a patch) weighted fusion, and iii) the previous approach [34], which coupled registration and segmentation through the use of appearance-based local soft selection. The proposed approach outperformed all other three methods in terms of Dice coefficient and Symmetric Mean Surface Distance (SMSD). More details on the implementation and the quality of the obtained results can be found in [13].

Spatiotemporal MRF-based deformable registration

In the previous section a solution to the joint problem of multi-atlas registration and segmentation was proposed. In this section we study the special case in which the multiple images are from the same subject but acquired over different time points by varying some imaging parameters. Additionally to possible subject motion or organ deformation due to physiological processes (respiration or cardiac pulse), the change of imaging parameters causes an alteration in image intensities that is different in healthy and diseased tissue. The intensity change can often be described by a physiological or temporal model applied on a voxel-wise basis in the group of images. However the voxel correspondences are unknown, imposing the need for a group-wise deformable registration coupled with the computation of the model parameters. As an application paradigm, the registration of diffusion-weighted magnetic resonance (DW-MR) images can be chosen, which is performed by scanning the patient with different b-values (a parameter determining the strength and duration of the diffusion gradients). The apparent diffusion coefficient (ADC), which reflects the gradient of water diffusivity in the body, can be subsequently computed at each voxel and displayed as a parametric map. ADC has been examined not only for the characterization and staging of lesions but also for the response to treatment [39]. The calculation of ADC was performed without motion or artifact correction in the previous studies. It is reported though that ADC is a parameter susceptible to artifacts, the most frequent of all being patient's motion and breathing, resulting in misregistration of the images obtained with different b-values [40], [41]. Therefore,

in order to correctly visualize diffusion in biological tissues, image registration (non-linear due to organ movement) should precede the ADC calculation.

Two approaches were implemented for this purpose. First we proposed a group-wise registration method that incorporates a temporal model (reflecting diffusion) within the similarity term (data cost) [15]. Subsequently, in order to accurately quantify the diffusion process during image acquisition, we introduced a high-order MRF model that jointly registers the DW-MR images and models the spatiotemporal diffusion [14]. Spatial smoothness on the estimated diffusion variables, as well as spatiotemporal deformation smoothness, are imposed towards producing anatomically meaningful representations.

Let us consider a sequence of m DW-MR images, each one described by intensity values $s_t(x), t = 1, ..., m$, with $x \in \Omega_t$, $\Omega_t \in \boldsymbol{R}^3$. On top of that, let's consider an extra image $z(x)$ with $x \in \Omega_z$ which represents the ADC template corresponding to the previous image sequence. This ADC template is regarded as the reference frame of an optimal alignment among the DW-MR images. We therefore seek for a set of displacement fields d_t, $t = \{1, ..., m\}$ which map mutually corresponding points from the m-image spaces to the same point of the reference frame Ω_z. Having acquired the images with large b-values ($> 50s/mm^2$), we can use the exponential diffusion model proposed in [42] to express the relation between the ADC image (referred to with the letter $\boldsymbol{z}$) and the DW image $\boldsymbol{s_t}$ acquired with b-value b_t :

$$\boldsymbol{s_t} = \boldsymbol{s_0} e^{-b_t \cdot \boldsymbol{z}} \tag{2.11}$$

where $\boldsymbol{s_0}$ the DW image for $b = 0$. Denoting with $\boldsymbol{y_t}$ the natural logarithm of the image vector $\boldsymbol{s_t}$, and by assuming also the presence of noise $\boldsymbol{n_t}$, the relation between image intensities and ADC is expressed as:

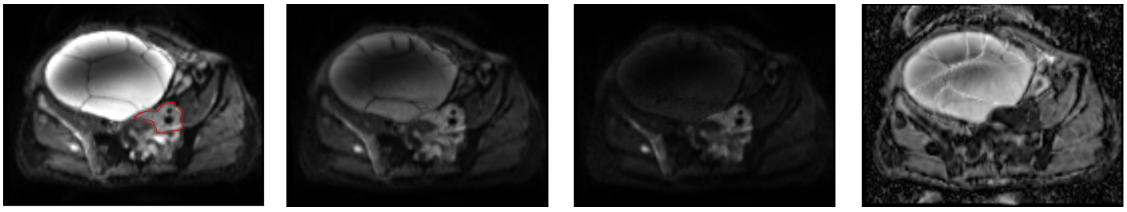
$$\boldsymbol{y_t} = -b_t \boldsymbol{z} + \boldsymbol{y_0} + \boldsymbol{n_t} \tag{2.12}$$

If all m DW-MR images are perfectly aligned, the standard way of estimating ADC is by linear regression against the b-values:

$$\hat{\boldsymbol{z}} = \frac{\sum_{t=1}^{m} (b_t - \bar{b})(\boldsymbol{y_t} - \bar{\boldsymbol{y}})}{\sum_{t=1}^{m} (b_t - \bar{b})^2}, \quad \hat{\boldsymbol{y}}_{\boldsymbol{0}} = \bar{b}\hat{\boldsymbol{z}} + \bar{\boldsymbol{y}} \tag{2.13}$$

where $\bar{b}$ denotes the mean b-value and $\bar{\boldsymbol{y}}$ the mean of the logarithmic images.

This model is incorporated in two different ways. In both methods [15][14] the registration approach we propose aims at (i) correcting deformations due to (local) organ deformations or motion, (ii) ensure temporal consistency across the deformations, and (iii) impose spatial consistency on the deformation fields. However in [14], additionally to the previous criteria, we aim at a joint mapping and ADC refinement in terms of spatial consistency and smoothness. Specifically we model the ADC values as additional variables and impose spatial consistency on the derived ADC map. The diffusion model used for the calculation of the ADC, as well as the deformation model, are jointly optimized to define the reference pose. The main differences in energy cost formulation between the two approaches are described next. An illustrative example is shown in Fig.2.6.



(a) (b) (c) (d)

Figure 2.6: An axial slice of the DWI volume acquired with b-values (a) 50, (b) 400, (c) 800 s/mm^2. The estimated ADC map is shown in (d). The red contour in image (a) denotes the tumor boundary.

Group-wise registration using a model of intensity change

In [15], we propose a computationally efficient group-wise registration scheme that does not require choosing a reference template. The latent variables of the model are m deformation fields (B-spline polynomials), $\{\boldsymbol{D_1}, \cdots, \boldsymbol{D_m}\}$, which are obtained using the discrete formulation introduced in [43]. The deformation variables are connected with the observations towards ensuring meaningful temporal correspondence among the DW images. They are also inter-connected in order to decrease the cost of pairwise comparisons between individual images. This is expressed by the following energy function that maximize a similarity function Sim and some regularization functions that enforce smoothness between neighboring nodes in space (R_t) and in time ($R_{t,t+1}$).

$$\begin{aligned}\hat{\boldsymbol{D}}_{\mathbf{1}}, \cdots, \hat{\boldsymbol{D}}_{\boldsymbol{m}} = argmax_{\boldsymbol{D_1}, \cdots, \boldsymbol{D_m}} \{ & Sim(\boldsymbol{s_1} \circ \boldsymbol{D_1}, \cdots, \boldsymbol{s_m} \circ \boldsymbol{D_m}) \\ & + \sum_{t=1}^{M} R_t(\boldsymbol{D_t}) + \sum_{t=1}^{M-1} R_{t,t+1}(\boldsymbol{D_t}, \boldsymbol{D_{t+1}}) \}\end{aligned} \tag{2.14}$$

Linear programming and duality [44] are used to determine the optimal solution of the problem. Our work was inspired by the approach of [45]. The main difference between the two approaches is found in the encoding of the global similarity of the population. A statistical compactness criterion has been used in [45], whereas we incorporate within the registration process the physiological model of Eq.2.11 representing the temporal intensity change. This intensity modeling step removes the ambiguity during the search of anatomic correspondences and thus increases the mapping accuracy. Specifically, the similarity term Sim uses some interpolation functions and a matching criterion ρ similarly to Eq. 2.5. Here it is expected that as the images are jointly aligned, the derived optimal pose would express more accurately the diffusion process, thus ρ is defined as the regression error over all images:

$$\rho(\boldsymbol{s_1}, \cdots, \boldsymbol{s_m}) = \sum_{t=1}^{m} (\boldsymbol{s_t} - \hat{\boldsymbol{s}}_{\boldsymbol{t}})^2 \tag{2.15}$$

where $\hat{\boldsymbol{s}}_{\boldsymbol{t}} = e^{\hat{\boldsymbol{y}}_{\mathbf{0}} - b_t \cdot \hat{\boldsymbol{z}}}$ is calculated by substituting Eq.2.13.

It is obvious that this matching criterion requires the joint registration of all images $\boldsymbol{s_t}$. We approximate the group-wise registration solution by formulating a deformation mechanism based

on two terms: a global unary term and a inter-image pair-wise term. The global unary term represents a global cost for the deformations by assuming that for each node p of a given deformation field/image t, the rest of the images do not get deformed within the current iteration. This assumption is considered for all nodes and for all deformation fields within a given iteration, an assumption that is common in minimizing graphical models through expansion moves. local pair-wise comparisons between members of the population are examined too. The inter-image pair-wise term is assessed between each pair of images by allowing them to be deformed, while keeping the rest $m-2$ images static within the current iteration. These inter-image pair-wise comparisons are performed only between the immediate neighbors (t, $t+1$) in the temporal domain. More details on the registration mechanism, the regularization cost and the optimization method, can be found in [15].

Joint deformable registration and diffusion modeling

In the work described above the aim was to calculate a set of deformation fields by minimizing the regression error, while $\boldsymbol{z}$ and $\boldsymbol{y_0}$ were derived by acquired solution (Eq.2.13). Regression is performed on a voxel-wise basis, thus interactions among neighboring ADC values are not taken into account. In [14] we use some priors on context dependencies for $\boldsymbol{z}$ and $\boldsymbol{y_0}$ and solve the joint problem:

$$\begin{aligned}\hat{\boldsymbol{z}}, \hat{\boldsymbol{y}}_{\mathbf{0}}, \hat{\boldsymbol{D}}_{\mathbf{1}}, \cdots, \hat{\boldsymbol{D}}_{\boldsymbol{m}} = argmax_{\boldsymbol{z}, \boldsymbol{y_0}, \boldsymbol{D_1}, \cdots, \boldsymbol{D_m}} \{ & Sim(\boldsymbol{z}, \boldsymbol{y_0}, \boldsymbol{s_1} \circ \boldsymbol{D_1}, \cdots, \boldsymbol{s_m} \circ \boldsymbol{D_m}) \\ & + \sum_{t=1}^{M} R_t(\boldsymbol{D_t}) + \sum_{t=1}^{M-1} R_{t,t+1}(\boldsymbol{D_t}, \boldsymbol{D_{t+1}}) \\ & + R_{\boldsymbol{z}}(\boldsymbol{z}) + R_{\boldsymbol{y_0}}(\boldsymbol{y_0}) \}\end{aligned} \tag{2.16}$$

where $R_{\boldsymbol{z}}$, $R_{\boldsymbol{y_0}}$ express the spatial regularization (smoothness) of $\boldsymbol{z}$ and $\boldsymbol{y_0}$, respectively. The energy terms can be derived by the Maximum a Posteriori (MAP) technique. If we assume that the noise is zero mean i.i.d., the multivariate pdf of $\boldsymbol{n_t}$ is given by:

$$P(\boldsymbol{n_t}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma_n^N} exp\{-\frac{1}{2\sigma_n^2} \boldsymbol{n_t^T n_t}\} \tag{2.17}$$

where N the size of the linearized noise vector and σ_n^2 denotes the variance of the noise process. Therefore from equations 2.17 and 2.12 we get

$$P(\boldsymbol{y_1}, \cdots, \boldsymbol{y_m} \mid \boldsymbol{z}, \boldsymbol{y_0}) \sim \sum_{t=1}^{m} (-\frac{\|\boldsymbol{y_t} + b_t \boldsymbol{z} - \boldsymbol{y_0}\|^2}{2\sigma_n^2}) \tag{2.18}$$

This equation is true if there is no deformation between the images $\boldsymbol{y_t}$. Assuming now that some motion has occurred during the acquisition process of the DW-MR images, Eq. 2.12 takes the form:

$$\boldsymbol{y_t} \circ \boldsymbol{D_t} = -b_t \boldsymbol{z} + \boldsymbol{y_0} + \boldsymbol{n_t} \tag{2.19}$$

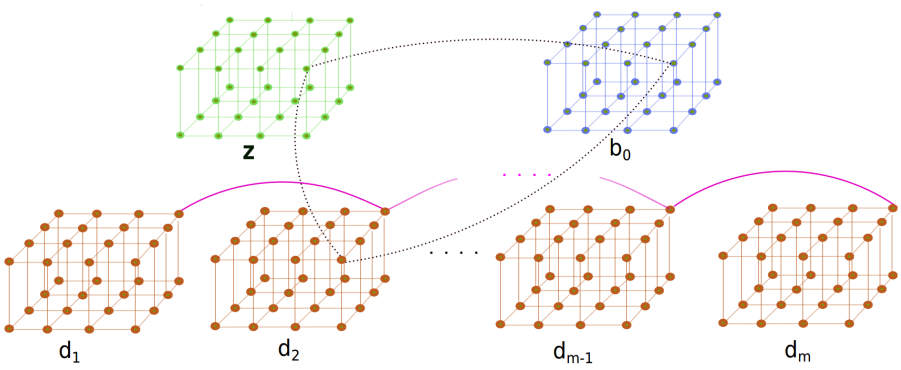


Figure 2.7: The node and the edge system of all the connected graphs. With green, blue and brown color the relationship within the grid nodes belonging to $\boldsymbol{z}$ (ADC), $\boldsymbol{y_0}$, and each of the deformation fields, respectively. These connections ensure smoothness within each grid. The black dotted edge represents a third-order dependency among deformation field and diffusion model (required in the data term), whereas pink edges denote the temporal relationship between successive (in time) deformation fields.

The joint posterior probability for the diffusion model parameters and the deformation fields can be calculated from Bayes' rule:

$$P(\boldsymbol{z}, \boldsymbol{y_0}, \boldsymbol{D_1}, \cdots, \boldsymbol{D_m} \mid \boldsymbol{y_1}, \cdots, \boldsymbol{y_m}) = \frac{P(\boldsymbol{y_1}, \cdots, \boldsymbol{y_m} \mid \boldsymbol{z}, \boldsymbol{y_0}, \boldsymbol{D_1}, \cdots, \boldsymbol{D_m}) P(\boldsymbol{z}, \boldsymbol{y_0}, \boldsymbol{D_1}, \cdots, \boldsymbol{D_m})}{P(\boldsymbol{y_1}, \cdots, \boldsymbol{y_m})} \tag{2.20}$$

By ignoring the denominator since it is not a function of the variables to be estimated, and assuming that the random fields $\boldsymbol{D_t}$, $\boldsymbol{z}$ and $\boldsymbol{y_0}$ are statistically independent, we can arrive at the MAP solution by updating Eq. 2.18 using Eq. 2.19 and introducing it into Eq. 2.20:

$$\begin{aligned} \hat{\boldsymbol{z}}, \hat{\boldsymbol{y}}_{\mathbf{0}}, \hat{\boldsymbol{D}}_{\mathbf{1}}, \cdots, \hat{\boldsymbol{D}}_{\boldsymbol{m}} = argmax_{\boldsymbol{z}, \boldsymbol{y_0}, \boldsymbol{D_1}, \cdots, \boldsymbol{D_m}} \{ \sum_{t=1}^{m} (-\frac{\| \boldsymbol{y_t} \circ \boldsymbol{D_t} + b_t \boldsymbol{z} - \boldsymbol{y_0} \|^2}{2\sigma_n^2}) \\ + \log P(\boldsymbol{z}) + \log P(\boldsymbol{y_0}) + \log P(\boldsymbol{D_1}, \cdots, \boldsymbol{D_m}) \} \end{aligned} \tag{2.21}$$

The priors on the diffusion model variables, $P(\boldsymbol{z})$ and $P(\boldsymbol{y_0})$, correspond to the regularization terms $R_{\boldsymbol{z}}$ and $R_{\boldsymbol{y_0}}$ in Eq. 2.16, whereas the prior on the deformation variables, $P(\boldsymbol{D_1}, \cdots, \boldsymbol{D_m})$, models deformation dependencies in space (R_t) and in time ($R_{t,t+1}$).

MRF Formulation of Joint Problem We formulate our joint problem of DW-MRI registration and ADC smoothness using Markov Random Fields (MRF). The joint model is parameterized by a set of $m+2$ isomorphic grid graphs $G = \{G_1, \ldots, G_{m+2}\}$, each of the first m superimposed onto the corresponding DW-MR image, whereas the last two grid graphs are superimposed onto the ADC map that we want to compute and the $\boldsymbol{y_0}$ image vector respectively. To this end, we define

three sets of labels, a quantized version of the displacement space, and quantized versions of the $\boldsymbol{z}$ and $\boldsymbol{y_0}$ values, respectively. Every node in each grid represents either a displacement if the grid is a deformation grid or the ADC or the $\boldsymbol{y_0}$ value in case of the last two grids, respectively. The nodes in the graph are also connected with a set of edges $\mathcal{E}$ that encode the interactions between the deformation and diffusion model variables, as illustrated in Fig.2.7.

Perspectives of graphical models

During the past decade the mainstream effort was dedicated to the development of efficient inference methods for low-rank (pair-wise) user-defined application-driven graphical models involving limited interactions between the nodes of the graphical models. Such models have been very powerful in addressing a number of low and mid-level computational perception problems but failed to cope with high-level tasks. One can claim that this was due to:

- constraints being imposed by the low-rank connections between graph nodes (pair-wise models offer limited interaction between variables) that was mostly the outcome of lacking efficient optimization methods to address higher order interactions.
- absence of leveraging large scale manually annotated data towards defining the optimal inference problem. Obviously recovering the optimal solution of an arbitrary objective function that does not correspond to the one that should be used towards getting the optimal visual perception answer is not that useful.
- absence of introducing high-level knowledge on the structure of the considered graphs towards facilitating the process of describing the space of solutions and making inference feasible at reasonable computational time with strong optimality guarantees.

An interesting future perspective will be to introduce a novel discrete, data-driven, higher order, structured computational framework to address visual perception and its applications in large scale modeling and biomedical image analysis. Medicine is a field where human understandable predictions could have a much stronger adoption rate from the clinical experts. Therefore coupling the advances of machine learning able to provide quite powerful individual predictions (like for example deep learning), with human understandable structured representations (graphs learned from examples, tree structures being the derivation of structured models like for example grammars) and efficient distributed optimization methods could be a major breakthrough towards reproducing or even surpassing human intelligence.

Personal contribution

My contribution in the work presented in section 2.2 was major and involved the formulation and implementation of the registration method designed for brain images with pathology, as well as the integration with the tumor growth model, whereas the actual modeling of the biomechanics due to tumor growth (i.e. the PLE simulator or the finite element model) was not part of my work. In the presented framework the two problems of registration and tissue characterization were coupled together and solved in an iterative brute-force type approach [9][10]. A framework for joint

segmentation-registration was later proposed in [46]. In respect to section 2.3 , the methods in the area of discrete optimization and inference algorithms are mainly attributed to co-authors, while my personal contribution relied mostly in the MRF model formulation and energy cost definition (first step in the implementation framework). Specifically, given an interpretation objective, the first step consists in describing the space of solutions through a parametric mathematical model. The parameters of this model are then associated with the measurements through the definition of an objective cost function, expressed as the MRF energy in our graphical models formulation. My overview was deeper concerning the work described in section 2.3 and involved the Maximum a Posteriori (MAP) solution and joint graph formulation.

Bibliography

[1] H. Li and J. Tan, "Heartbeat-driven medium-access control for body sensor networks", *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 44–51, 2010.

[2] E. Imhof, G. Konrad, J. R. Long, P. E. Pash, and R. E. Reinke, *Time synchronization improvements for interoperable medical devices*, US Patent 8,977,883, Mar. 2015.

[3] F. Sivrikaya and B. Yener, "Time synchronization in sensor networks: A survey", *IEEE network*, vol. 18, no. 4, pp. 45–50, 2004.

[4] W. M. David, "Bioinformatics: Sequence and genome analysis", *Bioinformatics*, vol. 28, 2001.

[5] C. Notredame, "Recent progress in multiple sequence alignment: A survey", *Pharmacogenomics*, vol. 3, no. 1, pp. 131–144, 2002.

[6] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey", *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[7] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: A review", *Computer methods in biomechanics and biomedical engineering*, vol. 17, no. 2, pp. 73–93, 2014.

[8] (2011). Neuroimaging informatics tools and resources clearing, [Online]. Available: `https://www.nitrc.org/` (visited on 06/02/2011).

[9] E. I. Zacharaki, D. Shen, S.-K. Lee, and C. Davatzikos, "Orbit: A multiresolution framework for deformable registration of brain tumor images", *Medical Imaging, IEEE Transactions on*, vol. 27, no. 8, pp. 1003–1017, 2008.

[10] E. I. Zacharaki, C. S. Hogea, D. Shen, G. Biros, and C. Davatzikos, "Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth", *Neuroimage*, vol. 46, no. 3, pp. 762–774, 2009.

[11] P. M. Thompson, J. N. Giedd, R. P. Woods, D. MacDonald, A. C. Evans, and A. W. Toga, "Growth patterns in the developing brain detected by using continuum mechanical tensor maps", *Nature*, vol. 404, no. 6774, pp. 190–193, 2000.

[12] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi, "Multi-modal image set registration and atlas formation", *Medical image analysis*, vol. 10, no. 3, pp. 440–451, 2006.

[13] S. Alchatzidis, A. Sotiras, E. I. Zacharaki, and N. Paragios, "A discrete mrf framework for integrated multi-atlas registration and segmentation", *International Journal of Computer Vision*, pp. 1–13, 2016.

[14] E. Kornaropoulos, E. I. Zacharaki, P. Zerbib, C. Lin, A. Rahmouni, and N. Paragios, "Optimal estimation of diffusion in dw-mri by high-order mrf-based joint deformable registration and diffusion modeling", in *Biomedical Image Registration (WBIR) (in conjunction with CVPR), 2016 IEEE 7th International Workshop on*, IEEE, 2016, pp. 4321–4328.

[15] ——, "Deformable group-wise registration using a physiological model: Application to diffusion-weighted mri", in *Image Processing (ICIP), 2016 IEEE International Conference on*, IEEE, 2016, pp. 1–5.

[16] D. Shen and C. Davatzikos, "Hammer: Hierarchical attribute matching mechanism for elastic registration", *IEEE transactions on medical imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.

[17] E. I. Zacharaki, D. Shen, A. Mohamed, and C. Davatzikos, "Registration of brain images with tumors: Towards the construction of statistical atlases for therapy planning", in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, IEEE, 2006, pp. 197–200.

[18] A. Mohamed, E. I. Zacharaki, D. Shen, and C. Davatzikos, "Deformable registration of brain tumor images via a statistical model of tumor-induced deformation", *Medical image analysis*, vol. 10, no. 5, pp. 752–763, 2006.

[19] E. I. Zacharaki, C. S. Hogea, G. Biros, and C. Davatzikos, "A comparative study of biomechanical simulators in deformable registration of brain tumor images", *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 3, pp. 1233–1236, 2008.

[20] E. I. Zacharaki, C. S. Hogea, D. Shen, G. Biros, and C. Davatzikos, "Parallel optimization of tumor model parameters for fast registration of brain tumor images", in *Medical Imaging: Image Processing*, International Society for Optics and Photonics, vol. 6914, 2008, pp. 1–10.

[21] C. Davatzikos, E. I. Zacharaki, A. Gooya, and V. Clark, "Multi-parametric analysis and registration of brain tumors: Constructing statistical atlases and diagnostic tools of predictive value", in *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011, pp. 6979–6981.

[22] A. Mohamed and C. Davatzikos, "Finite element modeling of brain tumor mass-effect from 3d medical images", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2005, pp. 400–408.

[23] J. Modersitzki, *Numerical methods for image registration*. Oxford University Press on Demand, 2004.

[24] C. Hogea, F. Abraham, G. Biros, and C. Davatzikos, "A framework for soft tissue simulations with applications to modeling brain tumor mass-effect in 3-d images", in *3rd Canadian Conf. Comput. Robot Vision*, 2006, pp. 24–33.

[25] C. Hogea, G. Biros, F. Abraham, and C. Davatzikos, "A robust framework for soft tissue simulations with application to modeling brain tumor mass effect in 3d mr images", *Physics in medicine and biology*, vol. 52, no. 23, p. 6893, 2007.

[26] G. A. Gray and T. G. Kolda, "Algorithm 856: Appspack 4.0: Asynchronous parallel pattern search for derivative-free optimization", *ACM Transactions on Mathematical Software (TOMS)*, vol. 32, no. 3, pp. 485–507, 2006.

[27] G.-Z. Li, J. Yang, C.-Z. Ye, and D.-Y. Geng, "Degree prediction of malignancy in brain glioma using support vector machines", *Computers in Biology and Medicine*, vol. 36, no. 3, pp. 313–325, 2006.

[28] M. Lacroix, D. Abi-Said, D. R. Fourney, Z. L. Gokaslan, W. Shi, F. DeMonte, F. F. Lang, I. E. McCutcheon, S. J. Hassenbusch, E. Holland, *et al.*, "A multivariate analysis of 416 patients with glioblastoma multiforme: Prognosis, extent of resection, and survival", *Journal of neurosurgery*, vol. 95, no. 2, pp. 190–198, 2001.

[29] D. Koller and N. Friedman, *Probabilistic graphical models: Principles and techniques*. MIT press, 2009.

[30] N. Paragios, E. Ferrante, B. Glocker, N. Komodakis, S. Parisot, and E. I. Zacharaki, "(hyper)-graphical models in biomedical image analysis", *Medical Image Analysis*, 2016, in press.

[31] T. Rohlfing, R. Brandt, R. Menzel, D. Russakoff, and C. Maurer, "Quo Vadis, Atlas-Based Segmentation?", in *Handbook of Biomedical Image Analysis*, ser. Topics in Biomedical Engineering International Book Series, J. Suri, D. Wilson, and S. Laxminarayan, Eds., Springer US, 2005, pp. 435–486.

[32] "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion", *NeuroImage*, vol. 33, no. 1, pp. 115–126, Oct. 2006.

[33] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey", *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[34] S. Alchatzidis, A. Sotiras, and N. Paragios, "Discrete Multi Atlas Segmentation using Agreement Constraints", in *British Machine Vision Conference*, Sep. 2014.

[35] M. Sdika, "Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote", *Medical Image Analysis*, vol. 14, no. 2, pp. 219–226, 2010.

[36] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, "Deformable Medical Image Registration: Setting the State of the Art with Discrete Methods*", *Annual Review of Biomedical Engineering*, vol. 13, no. 1, pp. 219–244, 2011.

[37] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 3, pp. 531–552, Mar. 2011.

[38] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable.", *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 153–163, Feb. 2012.

[39] C. Lin, A. Luciani, E. Itti, T. El-Gnaoui, A. Vignaud, P. Beaussart, S.-j. Lin, K. Belhadj, P. Brugières, E. Evangelista, *et al.*, "Whole-body diffusion-weighted magnetic resonance imaging with apparent diffusion coefficient mapping for staging patients with diffuse large b-cell lymphoma", *European radiology*, vol. 20, no. 8, pp. 2027–2038, 2010.

[40] D. Albano, C. Patti, L. La Grutta, F. Agnello, E. Grassedonio, A. Mulè, G. Cannizzaro, U. Ficola, R. Lagalla, M. Midiri, *et al.*, "Comparison between whole-body mri with diffusion-weighted imaging and pet/ct in staging newly diagnosed fdg-avid lymphomas", *European journal of radiology*, vol. 85, no. 2, pp. 313–318, 2016.

[41] T. C. Kwee, T. Takahara, M. A. Vermoolen, M. B. Bierings, W. P. Mali, and R. A. Nievelstein, "Whole-body diffusion-weighted imaging for staging malignant lymphoma in children", *Pediatric radiology*, vol. 40, no. 10, pp. 1592–1602, 2010.

[42] D. Le Bihan, E. Breton, D. Lallemand, M. Aubin, J. Vignaud, and M. Laval-Jeantet, "Separation of diffusion and perfusion in intravoxel incoherent motion mr imaging", *Radiology*, vol. 168, no. 2, pp. 497–505, 1988.

[43] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through mrfs and efficient linear programming", *Medical image analysis*, vol. 12, no. 6, pp. 731–741, 2008.

[44] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies", *Computer Vision and Image Understanding*, vol. 112, no. 1, pp. 14–29, 2008.

[45] A. Sotiras, N. Komodakis, B. Glocker, J.-F. Deux, and N. Paragios, "Graphical models and deformable diffeomorphic population registration using global and local metrics", in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, Springer, 2009, pp. 672–679.

[46] A. Gooya, K. M. Pohl, M. Bilello, L. Cirillo, G. Biros, E. R. Melhem, and C. Davatzikos, "GLISTR: glioma image segmentation and registration", *IEEE Trans. Med. Imaging*, vol. 31, no. 10, pp. 1941–1954, 2012.

Chapter 3: Data representation and dimensionality reduction

One the fundamental problems in machine learning and pattern recognition is to discover compact representations of high-dimensional data. Medical data is typically sparse, high-dimensional (especially the 3D or 4D imaging data) and noisy. Also, very often the number of observations (available subjects) is low whereas data variability (e.g. across subjects) is high. All these challenges make inference especially hard imposing the need for dimensionality reduction. One way of making the data more compact is to extract features related to repeated patterns or signal/image properties that are usually encoded within the data. Dimensionality reduction can overall be achieved either by selecting a subset of the original features (variable selection) or by mapping through linear and/or nonlinear transformation (embedding methods) the original data or extracted features onto a new, lower-dimension space with the aim of eliminating irrelevant and redundant features, which is a key component for both supervised and unsupervised classification or regression problems. Embedding methods explore local data structure which is a very important property, especially when the data lie on non-linear manifolds. However, a disadvantage of data transformation over simple variable selection is that the generation of new features makes the original variables uninterpretable, and information about how much each original variable contributes is often lost. Additionally, the above feature reduction methods are not always capable of handling all types of variables, such as categorical variables. Moreover, dimensionality reduction through data transformation can be applied only under certain conditions. The following sections summarize our work in respect to feature extraction and embedding methods for representation and analysis of biosignals, medical images and molecular data dedicated to the clinical applications we have studied.

3.1 Feature extraction and fusion

Analysis of biosignals

Common problem in signal processing, and more specifically in neuroinformatics, is how to monitor for anomalies, detect signals beyond the limits, discover typical patterns of activity followed by a pathological event, such as a seizure, predict atypical patterns before they occur, etc. The subjective evaluation of biosignals makes automatically extracted parameters (computer-based) highly useful for diagnostics. The research performed under this scope aimed particularly at the

development of a system for monitoring of epileptic patients based on single or multi-channel biosignals, such as EEG and ECG.

Seizure detection Over the last decade clues to the unknown process that produces seizure have began to emerge from quantitative analysis of the electroencephalogram (EEG). In most of the parametric approaches found in the literature the analysis is based on the estimation of the EEG channels' spectral magnitude [1][2]. Other EEG features that have been reported are the autoregressive filter coefficients, the continuous and discrete wavelet transform, as well as energy per brain wave (delta, theta, alpha, beta, gamma) bands [1][2]. Furthermore, time domain features have been proposed, such as zero-crossing rate and statistics of the EEG samples per channel. Additionally to the EEG signals, it has been shown that seizures are often associated with cardiovascular and respiratory alterations [1][3]. Specifically, the study of electrocardiographic (ECG) signals can offer valuable information related to the seizure discharges [1]. Due to the difficulty of visual assessment of multiparametric recordings (in this case time-synchronous EEG and ECG) and in combination with the progress of signal processing and pattern recognition technology, approaches for automatic detection of seizures have been proposed in the literature. The ECG features are mainly based on the heart rate estimation (based on R-R intervals) and statistics of it, i.e. heart rate variability, as well as morphological features of the ECG signal [2][3].

In our work [4][5][6] we performed a large scale evaluation of time-domain and frequency domain features of EEG and ECG signals for offline analysis and online monitoring. Assuming that the data recorded by the sensors have been synchronized and transmitted as streams of multidimensional signals, frame blocking of the incoming streams to epochs of constant length is applied, and features are extracted from each epoch separately for each of the $n = 21$ EEG channels and ECG channel. In particular, each of the channels was parameterized using the following features:

1. EEG time-domain features: minimum value, maximum value, mean, variance, standard deviation, percentiles (25%, 50%-median and 75%), interquartile range, mean absolute deviation, range, skewness, kyrtosis, energy, Shannon's entropy, logarithmic energy entropy, number of positive and negative peaks, zero-crossing rate, and

2. EEG frequency-domain features: 6-th order autoregressive-filter (AR) coefficients, power spectral density, frequency with maximum and minimum amplitude, spectral entropy, delta-theta-alpha-beta-gamma band energy, discrete wavelet transform coefficients with mother wavelet function Daubechies 16 and decomposition level equal to 8.

3. ECG-based heart rate statistics: absolute value, minimum value, maximum value, mean, variance, standard deviation, percentiles (25%, 50% and 75%), interquartile range, mean absolute deviation, range. The heart rate estimation was based on Shannon energy envelope estimation for R-peak detection algorithm, implemented as in [7].

Feature extraction resulted to a feature vector of dimensionality equal to 55 for each of the 21 EEG channels and 12 for the ECG channel. The concatenated feature vectors were used to train seizure detection models. In a further step, the discriminative ability of the extracted features was examined. For the estimation of the importance of each feature in terms of their binary classification ability, the ReliefF algorithm [8] was used. ReliefF is one of the commonly used feature

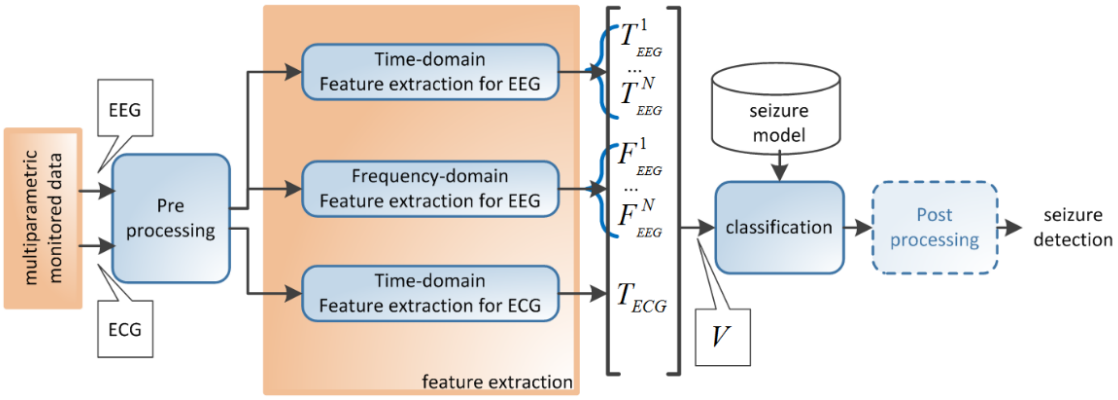


Figure 3.1: Block diagram of the seizure detection architecture based on time and frequency domain EEG and ECG features.

ranking algorithms due to its simplicity and effectiveness [9][10] (only linear time in the number of given features and training samples is required), noise tolerance and robustness in detecting relevant features effectively, even when these features are highly dependent on other features [9]. The cumulative evaluation results across the 21+1 electrodes revealed as most discriminative features the logarithmic energy entropy, the 2nd 3rd and 7th AR coefficients, the zero-crossing rate and the standard deviation. These features were ranked within the 10-best for all three evaluated subjects. For two of the subjects the median value of the ECG signal (i.e. the 50% percentile) was evaluated within the 20-best ranked features, indicating the existence of underlying information, related to seizure characteristics, within electrocardiographic signal, which is in agreement with previous studies [1][3]. For all examined subjects the use of subset of features reduced the precision of the seizure detector. However, the exclusion of the approximately 30% worst features still offered performance comparable to the best achieved and in combination with the reduction of the computational load of the detection architecture (both in the feature extraction stage and the classification stage) could be a valuable solution for online scenarios.

Classification of episodes of paroxysmal loss of consciousness The common causes of episodes of paroxysmal loss of consciousness are mainly that of epileptic seizures, commonly manifested by generalized spike wave discharges (GSW) [11], possible psychogenic non-epileptic seizures (PNES) [12] and vasovagal syncopal attacks (VVS) [13]. The similar seizure-like reactions of both epileptic and non-epileptic events make their diagnosis a difficult task. In clinical practice, the diagnosis is based on historical information assisted by specific tests. An example of the different EEG manifestations of epileptic and non-epileptic events is shown in Fig.3.2. Only a few studies have been proposed in the literature for automated classification between epileptic and non-epileptic pathological events from EEG. Poulos et al. [14] proposed an algorithm which estimates a number of auto-correlated coefficients extracted from an appropriately selected epileptic EEG segment and examines whether these coefficients are correlated with the coefficients of the unknown EEG segments in order to classify the latest into epileptic or non-epileptic. Papavlasopou-

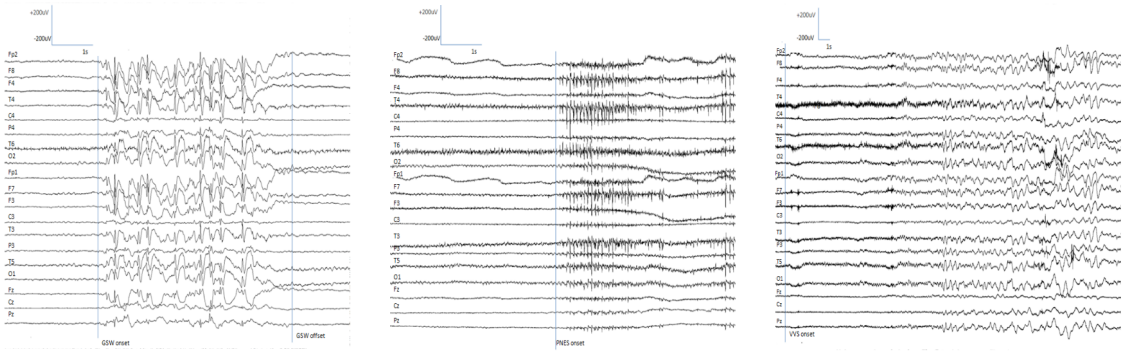


Figure 3.2: EEG recordings during epileptic and non-epileptic events. Left: epileptic seizure (GSW); middle: psychogenic non-epileptic seizure (PNES); right: vasovagal syncopal attack (VVS). The left and right marker in each figure indicate the onset and offset of the event, respectively.

los et al. [15] trained a LVQ1 neural network on an appropriately extracted set of auto-correlation coefficients (codebook) and used the resulting model to classify the corresponding feature vectors of the unknown EEG segments.

We used the same feature set extracted for seizure detection from temporal and spectral analysis over multiple EEG channels but this time trained classification models for epileptic and non-epileptic events [16][17]. Due to large number of features, feature ranking and selection was performed prior to classification using the ReliefF ranking algorithm within two different voting strategies. The classification models using feature subsets, achieved higher accuracy compared to the models using all features, as well as to methods of others [16].

Decision level fusion Spatiotemporal analysis of EEG is commonly used for detection or classification of EEG events in a multi-channel setting since it allows capturing dependencies across the EEG channels. However the significant increase of dimensionality, when features from different sensors are combined into a single feature vector, makes learning of classification models difficult, especially when using only a small number of samples usually available in clinical studies [4][16]. In [18], we investigated the classification of epileptic and non-epileptic events from multi-channel EEG data through the application of three different schemes for fusing information across EEG channels. We compare the common *feature level fusion* - which leads to the highest dimensionality of the feature vector and aims to capture the total spatiotemporal context prior to the classification step - with two *decision fusion* schemes performing per channel classification when: (i) the temporal context varies significantly across channels, thus local (sensor-dependent) training models are required, and (ii) the spatial variations are negligible in comparison to the inter-subject variation, thus only the temporal variation is modeled using a single global (sensor-independent) classifier. The three schemes are illustrated in Fig.3.3. The framework is applied on events that manifest across most EEG channels, as in the case of generalized epilepsy, PNES and VVS. Evaluation of the three classification architectures on EEG epochs from 11 subjects in an inter-subject cross-

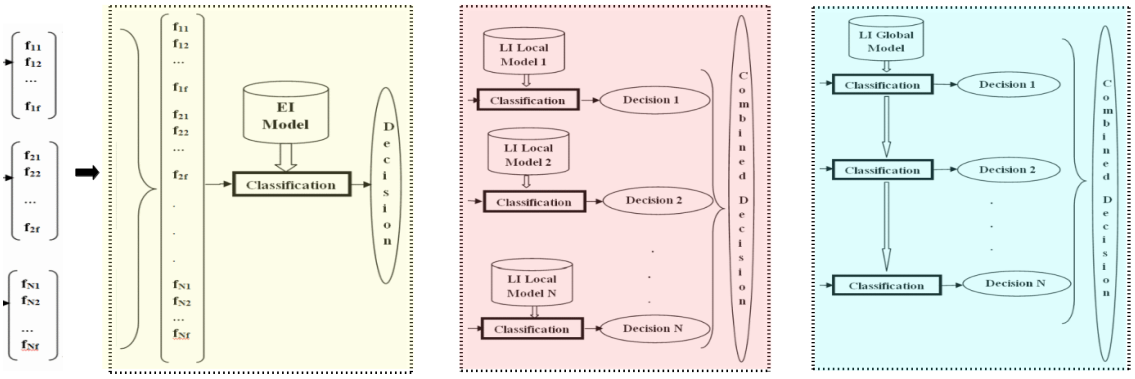


Figure 3.3: Different pattern analysis schemes. *Left*: Fusion of features, known as early integration (EI), *Middle/Right*: Fusion of classifier decisions, known as late integration (LI), using sensor-specific (LI Local) training models (*middle*) or a sensor-independent (LI global) model (*right*).

validation setting showed that the fusion in the decision level with a global (sensor-independent) classification model outperformed the other two schemes.

Fusion in the decision level was also performed in [19] for sleep spindle detection. Sleep spindles are characteristic transient oscillations that appear on the EEG during non-rapid eye movement (non-REM) sleep. The amount and the distribution of the sleep spindles is essential for describing the morphology of the sleep EEG, thus the assessment of the distribution of sleep spindles over a whole sleep cycle is needed. The low amplitude of some spindles, compared to the background EEG activity, renders the detection of sleep spindles difficult even for sleep experts. Visual detection and manual annotation of spindle occurrences for a whole sleep cycle recording is time consuming and affected by the subjectivity and degree of experience of the sleep technician. Therefore, automatic detection of sleep spindles is essential for reducing the workload and allowing the processing of large enough amount of data. In [19] we present a scheme for the automatic detection of sleep spindles, which is based on the combination of discriminative and statistical machine learning methods. The sequence of computed feature vectors was used as input for both a discriminative SVM model and a statistical hidden Markov model (HMM). Each of the two models estimates whether the i-th incoming feature vector corresponds to sleep spindle or not, i.e. providing binary classification results with the corresponding recognition score for each of the two classes. The second stage of the sleep spindle detection scheme exploits the SVM-based and the HMM-based predictions estimated at the first stage, in order to combine them and provide a final decision for each frame of the EEG signal. The fusion model was implemented with the SVM algorithm.

Analysis of 3D volumetric images

In [20][21] conventional MRI and perfusion MRI were integrated via a pattern classification technique into a multi-parametric imaging profile and used for differential diagnosis of brain neo-

plasms. Tumor characterization is difficult, because neoplastic tissue is often heterogeneous in spatial and imaging profiles, and for some imaging techniques often overlaps with normal tissue (especially the infiltrating part) [22]. Gliomas might show mixed characteristics, for example demonstrating both low and high grade features. The reference standard for characterizing brain neoplasms is currently based on histopathologic analysis following surgical biopsy or resection, but this is an invasive procedure which also has limitations including sampling error and variability in interpretation. In [20], we explore the heterogeneous regions of brain tumors by combining imaging features from several sequences and extract morphological and texture characteristics. Our analysis requires 3 regions of interest (ROIs), which define the neoplastic and necrotic region on contrast enhanced T1-weighted MRI (T1), and edematous region on Fluid-Attenuated Inversion Recovery (FLAIR) image.

The images were preprocessed following a number of steps including noise reduction, bias-field correction, and rigid intra-subject registration using the public software package FSL [23]. Co-registration of all sequences (T1, contrast enhanced T1-weighted images (T1ce), T2, FLAIR, relative cerebral blood volume (rCBV) maps calculated from the perfusion sequence), required in order to extract features from the ROIs, was performed with the rigid registration algorithm FLIRT [24] from FSL. The intensity levels were made comparable across subjects by histogram matching. For this purpose skull stripping was first performed using BET [25] to generate a brain tissue mask from the T1 image which was then used to extract the brain region from all other co-registered sequences. A linear transformation of the intensities (translation and scaling) was applied in order to minimize the L2-norm of the histogram difference between each subject and a template image. Histogram matching was not applied to the rCBV maps.

We chose a large number of features for investigation which included age, tumor shape characteristics, image intensity characteristics within some of the ROIs and Gabor texture features.

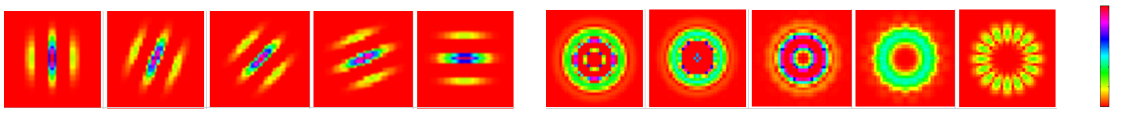
- **Shape and statistical characteristics of tumor:** Five shape features of the total tumor area were investigated, i.e. the tumor circularity, irregularity, rectangularity, the entropy of radial length distribution of the boundary voxels, and the surface-to-volume ratio. Also three statistical features were calculated, i.e. the ratio of enhancing, necrotic, and edematous tumor volume versus total (enhancing and non-enhancing) tumor volume.

- **Image intensity characteristics:** The mean and variance of image intensities of T1, T1ce, T2 are calculated in the central and marginal area of the different ROIs under certain conditions taking into account uncertainty due to possible neoplastic infiltration (we refer to [20] for more details).

- **Gabor texture:** The voxel-wise texture features of image $I(x, y, z)$ are extracted at each tomographic slice of the 3D ROI by convoluting with 2D Gabor filters [26][27] and averaging inside the ROI. The 2D Gabor filters are mathematically described at location (x, y, z) as

$$g_{\lambda,\theta,\sigma,\psi}(x, y) = exp\left(-\frac{x_\theta^2 + \gamma^2 y_\theta^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}x_\theta + \psi\right)$$

  where $x_\theta = xcos(\theta)+ysin(\theta)$, $y_\theta = -xsin(\theta)+ycos(\theta)$, $\lambda = 1/f$ is the wavelength, θ the orientation, γ the spatial aspect ratio which determines the eccentricity of the convolution



(a) (b) (c)

Figure 3.4: Examples of Gabor filters for same frequency in five different orientations (a) and five extracted rotation-invariant filters (b) displayed in color scale (c).

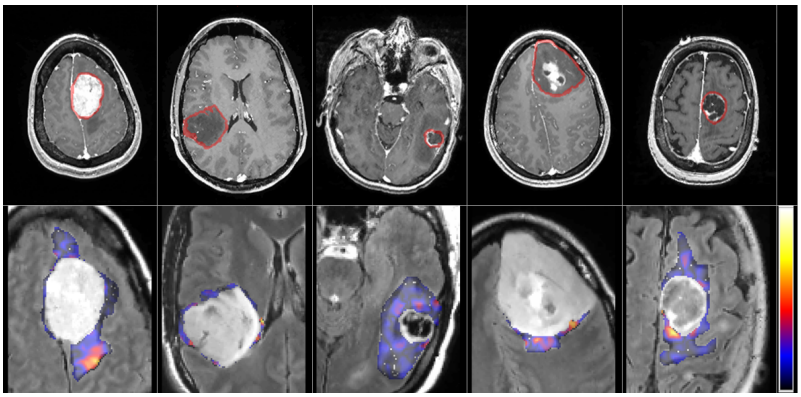
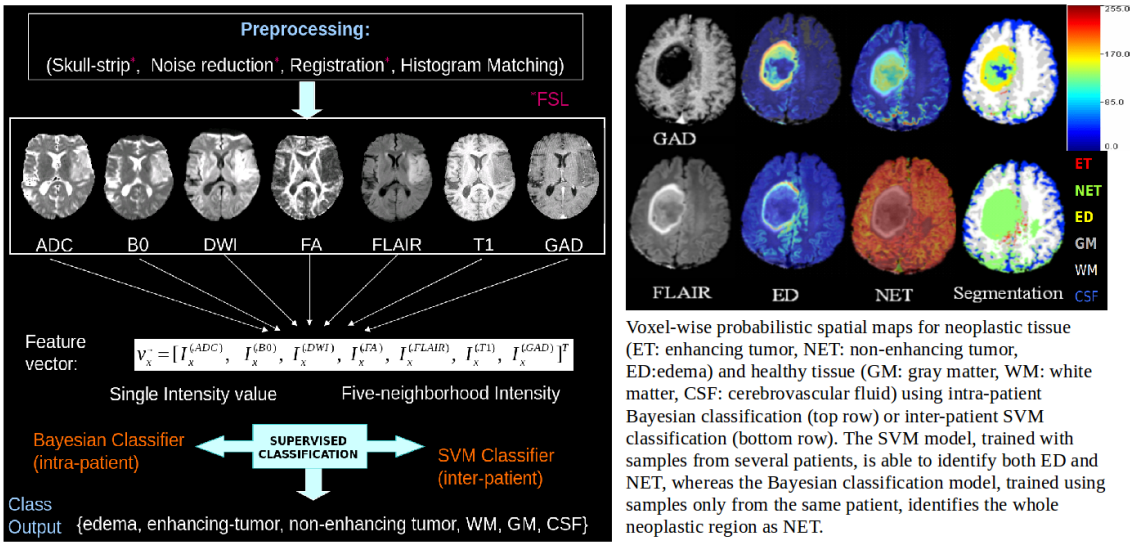


Figure 3.5: MR images of different of brain tumor types and an example of a texture pattern extracted from FLAIR in the edematous area. From left to right: meningioma, glioma grade II, grade III, grade IV and metastasis. 1st row: T1ce image with the tumoral ROI. 2nd row: FLAIR image (zoomed in the tumor region) overlaid with one of the textural patterns ($\lambda = 8$). The average texture values (calculated before FFT) proved to be significant in discrimination of meningiomas.

kernel, and ψ the phase offset which determines the symmetry of the Gabor function, and the ratio σ/λ the spatial frequency bandwidth. The texture was calculated by combining the output of a symmetric and anti-symmetric gabor kernel using the L2-norm. Then, in order to make the average Gabor features rotation invariant, for each radial frequency f Fast Fourier Transform (FFT) was performed across orientation θ [26] and the unique coefficients for each frequency were retained. Further details on the calculation of texture features are provided in [20].

Fig.3.4 illustrates in the first row the Gabor filter for a single frequency across the first 5 (out of 8) orientations and in the 2nd row the rotation-invariant filters after FFT for the same frequency. Fig.3.5 illustrates an example for each brain tumor type and a texture pattern extracted from the edematous area from FLAIR. The illustration shows the voxel-wise texture before averaging over the area of interest. The features were normalized to have zero mean and unit variance and then feature selection was applied to select a small set of effective features for classification in order to improve the generalization ability and the performance of the subsequent classifier. An intensity-based tissue profile was also used to produce healthy and neoplastic tissue probability maps and maps for tumor recurrence [27][28]. The purpose was to quantify the multiparamet-



Voxel-wise probabilistic spatial maps for neoplastic tissue (ET: enhancing tumor, NET: non-enhancing tumor, ED:edema) and healthy tissue (GM: gray matter, WM: white matter, CSF: cerebrovascular fluid) using intra-patient Bayesian classification (top row) or inter-patient SVM classification (bottom row). The SVM model, trained with samples from several patients, is able to identify both ED and NET, whereas the Bayesian classification model, trained using samples only from the same patient, identifies the whole neoplastic region as NET.

Figure 3.6: Tissue characterization based on multiparametric MRI. The analysis framework including preprocessing, fusion of multiple sequences and supervised (Bayesian or SVM) classification is illustrated on the left and some segmentation results on the right. GAD: gadolinium-enhanced T1-weighted image (same as T1ce).

ric imaging profile of neoplasms by integrating structural MRI (B0, FLAIR, T1, and T1ce) and Diffusion Tensor Imaging (DTI) metrics, such as Fractional Anisotropy (FA) and Apparent Diffusion Coefficient (ADC), via statistical image analysis methods to potentially capture complex and subtle tissue characteristics that are not obvious from any individual image or parameter. Results demonstrated that this multiparametric tissue characterization helps to better differentiate among neoplasm, edema, and healthy tissue (3.6), and to identify tissue that is likely to progress to neoplasm in the future.

Furthermore, in [29] we examined whether there are any correlations between FA, ADC and regional tumor blood volume (rTBV) values within the enhanced lesion in T1ce imaging. A negative correlation between normalized values of TBV and FA ($p < 0.05$, Spearman's test) was observed for the examined data.

Analysis of 4D spatiotemporal images

As example of spatiotemporal imaging data we will use the Dynamic contrast-enhanced MR imaging (DCE-MRI) applied for the characterization of breast cancer. DCE-MRI involves administration of a gadolinium-based contrast agent, followed by the acquisition of a temporal sequence of MR images of the breast under investigation. High permeability of tumor capillaries allows the contrast agent to diffuse faster in a tumor, leading to better enhancement of a tumor relative to the surrounding breast tissue. In the DCE-MRI image, malignant and benign tumors have been found to exhibit major types of spatiotemporal difference. Various features have been proposed in literature [30]. However, their use is mostly limited in practice because either they describe only

one type of difference (capturing dynamics or morphology) or they are too coarse to contain rich information on the tumor. In [30], we describe the spatiotemporal enhancement pattern (STEP) as a comprehensive set of these features. By viewing serial contrast-enhanced MR images as a single spatiotemporal image, we formulated the STEP as a combination of (1) dynamic enhancement and architectural features of a tumor, and (2) the spatial variations of pixelwise temporal enhancements. Although the latter has been widely used by radiologists for diagnostic purposes, it has rarely been employed for computer-aided diagnosis. The main contribution of this work in respect to data representation is that the STEP features are introduced to capture jointly the temporal enhancement and its spatial variations. This is essentially carried out through Fourier transformation and pharmacokinetic modeling of various temporal enhancement features, followed by the calculation of moment invariants and Gabor texture features.

We first modeled the temporal enhancement curves of breast tissues by Fourier transform. A pixelwise 1D discrete Fourier transform (DFT) was performed on the enhancement curve of each pixel p,

$$C(p,t) = \frac{I(p,t) - I(p,0)}{I(p,0)},\ (t = 1, ..., T-1),$$

where $I(p,t)$ denotes the intensity of p at a scanning time t and T is the total number of time sections. Thus we obtained $T-1$ DFT coefficients for each pixel p. Consequently, for a given tumor, each DFT coefficient yields a distinctive temporal enhancement map, which collectively represents the frequency content of the corresponding temporal enhancements. In practice, we select N_t enhancement maps corresponding to the lower order DFT coefficients. Once the tumor dynamics are modeled and the temporal enhancement maps are constructed, they are utilized to capture the spatial variations within a tumor. Since the orientation of a tumor sample is not related to its type, the features representing morphological and spatial structure should be rotation invariant. Accordingly, we employed rotation-invariant moment features [31] to capture the global structure, and Gabor rotation-invariant texture to capture the local spatial behavior. For the extraction of the gabor texture features a similar procedure as the one described in section 3.1 was followed. Image moments are computed as the particular weighted averages of pixel intensities to explain some global spatial distribution of image intensities. Hu′s seven moment invariants [31] are, for instance, defined as a polynomial equation of some scale-normalized centralized moments. With regard to our work, the two-dimensional centralized moments of an $M \times N$ temporal enhancement map, denoted as $f(x,y), (x,y) \in \Omega$, are defined as

$$m_{\alpha\beta} = \sum_{x=1}^{N} \sum_{y=1}^{M} (x - \bar{x})^{\alpha} (y - \bar{y})^{\beta} f(x,y),$$

where

$$\bar{x} = \frac{\sum_{x=1}^{N} \left(x \sum_{y=1}^{M} f(x,y) \right)}{\sum_{x=1}^{N} \sum_{y=1}^{M} f(x,y)}$$

and

$$\bar{y} = \frac{\sum_{y=1}^{M} \left(y \sum_{x=1}^{N} f(x,y) \right)}{\sum_{x=1}^{N} \sum_{y=1}^{M} f(x,y)}.$$

The two-dimensional scale-normalized centralized moment is defined as $\eta_{\alpha\beta} = m_{\alpha\beta}/m_{00}^{\gamma}$, where $\gamma = (\alpha+\beta)/2+1, \forall(\alpha+\beta) \geq 2$. This readily leads to $H_m = 7$ moment invariants for each of the N_t temporal enhancement maps. Consequently, each tumor sample was represented by $H_m \times N_t$ moment-invariant features. By including both moment invariants and Gabor texture features (H_g), we obtain a total of $(H_m + H_g) \times N_t$ features for characterizing the spatiotemporal profile of dynamic contrast-enhanced images.

More recently we investigated the application of decomposition for characterization of the spatial variation of temporal enhancement in DCE-MRI [32]. Multiresolution analysis has emerged a useful framework for many image analysis tasks in which the discrete wavelet transform (DWT) played a major role [33]. However, a drawback of the DWT is that it is not shift invariant. The stationary wavelet transform (SWT) [34] is a wavelet transform algorithm designed to overcome the lack of shift invariance of the DWT. More specifically, the DWT of a signal is defined as its inner product with a family of functions, which form an orthonormal set of vectors, a combination of which can completely define the signal. For the implementation of the DWT, only the coefficients of a low-pass and a high-pass half-band filter are required. SWT is similar to the DWT, but no downsampling is performed. Instead, upsampling of the low-pass and high-pass filters is carried out. For 2D images, the 2D SWT can be used which consists of a SWT on the rows of the image and a SWT on the columns of the resulting image. The decomposition of the image yields four subimages (one approximation and three detail images) for every level of decomposition. The detail subimages contain the textural information in horizontal, vertical, and diagonal orientations. We used as texture features the mean and entropy of the absolute value of the detail subimages. The approximation subimages were not used for texture analysis because they are rough estimates of the original image.

Analysis of molecular structures

Research in metagenomics, the field which combines the study of nucleotide sequences with their structure, regulation, and function, has been very productive the last years. While the number of newly discovered, but possibly redundant, protein sequences rapidly increases, experimentally verified functional annotation of whole genomes remains limited. Protein structure, i.e. the 3D configuration of the chain of amino acids, is a very good predictor of protein function, and in fact a more reliable predictor than protein sequence. This is mainly because the chemistry required for the functionality of protein active sites arises from their 3D structure. Thus, as sequences diverge, only those residues required for the chemistry of the protein activity will be absolutely conserved whose 3D structure should also be conserved [35]. By now, the number of proteins with functional annotation and experimentally predicted structure of their native state (e.g. by NMR spectroscopy or X-ray crystallography) is adequately large to allow learning training models that will be able to perform automatic functional annotation of unannotated proteins.

The building blocks of proteins are amino acids which are linked together by peptide bonds into a chain. The polypeptide folds into a specific conformation depending on the interactions between its amino acid side chains which have different chemistries. Many conformations of this chain are possible due to the rotation of the chain about each carbon (Cα) atom. The main type of protein structure representation we used was the shape of the protein backbone. The utilized

geometric descriptors are invariant to global translation and rotation of the protein, therefore previous protein alignment was not required. In [36][37] we also extracted sequence-based features, whereas in [38] the aim was to explore only structure, thus instead of sequence, pairwise amino acid distances were examined.

- **Torsion angles density:** The shape of the protein backbone was expressed by the two torsion angles of the polypeptide chain which describe the rotations of the polypeptide backbone around the bonds between N-Cα (angle ϕ) and Cα-C (angle ψ). The probability density of the torsion angles ϕ and $\psi(\in [-180, 180])$ was estimated based on the 2D sample histogram of the angles (also known as Ramachandran diagram). An example is shown in Fig.3.7 In [36][37] the torsion angles density was extracted for the whole protein, whereas in [38] feature maps were extracted separately for every amino acid type in the protein (therefore characterizing local interactions) and then stacked as a multi-dimensional array. Smoothness in the density function was achieved by moving average filtering, i.e. by convoluting the density map with a 2D gaussian kernel.

- **Density of amino acid distances:** All pairwise distances between each amino acid type (including standard and ambiguous) in the protein were calculated based on the coordinates of the Cα atoms for the residues and stored as an array. Since the size of the proteins varies significantly, the length of this array is different across proteins, thus not directly comparable. In order to standardize measurements, the sample histogram of these pairwise distances was extracted and smoothed by convolution with a 1D gaussian kernel.

- **Sequence-based features:** The similarity of each pair of sequences and can be quantified using the scoring matrix that is produced by a sequence alignment algorithm. The Smith-Waterman sequence alignment algorithm [39] has been preferred over the Needleman-Wunsch algorithm [40] due to the assessment of sequence similarity based on local alignment, which allows to take into consideration mutations that might have happened in amino acid sequences. The highest score in the previous matrix, which reflects the success of alignment of two sequences, is used as similarity criterion. The class probabilities for a given protein are expressed as the maximum similarity across all training samples within class, normalized over the sum of maximum similarities for all classes.

Both structural and amino acid sequence (AA) information are related to the enzymatic activity. In order to take into consideration these two properties, fusion of information was performed in both feature level and decision level. Assessment on single-labeled and multi-labeled enzymes showed that decision fusion outperformed feature fusion [37].

3.2 Embedding methods

Large feature sets, some of which may be irrelevant or even misleading, slow the machine learning algorithms down and make finding global optima difficult. In special cases, high-dimensional data set may contain many features that are all measurements of the same underlying cause, so are closely related. Revealing the low-dimensional representation of such high-dimensional data sets

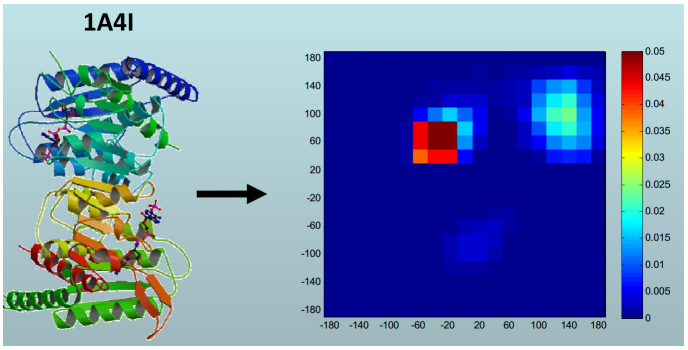


Figure 3.7: The 3D model of a protein structure (1A4I) and its corresponding torsion angles density map.

(manifold learning) not only leads to a more compact description of the data, but also enhances our understanding of the intrinsic data structure [41]. Many manifold learning techniques have been proposed, such as Isomap [42], Locally Linear Embedding (LLE) [43], and Laplacian Eigenmaps [44]. They have outperformed classical methods like Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS) [45] in harnessing non-linear data structures [46]. However, their high time and memory complexity impose severe limitations on their scalability [47].

Locality preserving projections for representation of waveforms

In [48][49] we compared several manifold learning techniques for compact representation of a specific EEG pattern, the interictal spike. The detection of epileptiform discharges (spikes) in interictal EEG is important for diagnosis of epilepsy. We proposed a methodology that increases specificity in a two stages process incorporating pattern classification. Similarly to most pattern detection methods in signal processing, the amount of data processed is reduced by first extracting candidate waveforms based on low level detection analysis (by feature extraction), while subsequently classification is performed to maximize specificity of the overall method [50]. Specifically, the proposed method first detects candidate spikes by breaking down the EEG signal around major peaks into half-waves and extracting distinctive attributes of the waveforms, such as height and duration, mimicking the criteria used by the neurophysiologists. If the raw signal (waveform around the primary vertex) is used as representation for the candidate spikes, classification is deemed to fail due to the high dimensionality of the input pattern. When the number of parameters increases, the volume of the space grows so fast that the concept of similarity, distance or nearest neighbor may not even be qualitatively meaningful, thus impeding clustering or classification. Thus classification of the detected waveforms is performed after embedding them in a low dimensional space.

We used the Locality Preserving Projections (LPP) [51] to embed the data in a low dimensional space. LPP is a linear approximation of the nonlinear Laplacian Eigenmap [52]. It finds a transformation matrix A that maps a set of points $x_i \in \mathcal{R}^d$, $(i = 1, ..., m)$ into a set of points $y_i \in \mathcal{R}^l$, $y_i = A^T x_i$ such that $l \ll d$. LPP is designed to preserve local structure, thus it is likely that a nearest neighbor search in the low dimensional space will yield similar results to that in

the high dimensional space. The intrinsic dimensionality (l) of the transients is unknown but we used the maximum likelihood estimation (MLE) method to obtain an estimated value. The MLE method gives a good estimate of the unknown parameters by maximizing the likelihood of the data we observe. It is a widely used estimation method showing essential properties with increasing number of samples, such as consistency, efficiency and asymptotic normality. In details, the LPP algorithm is as follows. Let X be the $d \times m$ matrix including the m waveforms. The samples constitute the nodes of a graph connected with edges having weights that depend on the samples' distance. If W is the $m \times m$ weighting matrix and D is a diagonal matrix whose entries are column sums of W, the eigenvectors α_k and eigenvalues λ_k, $(k = 0, ..., l-1)$ are computed for the generalized eigenvector problem :

$$XLX^T\alpha_k = \lambda_k XDX^T\alpha_k$$

where $L = D - W$ is the Laplacian matrix. The $n \times l$ transformation matrix A is formed by the l column vectors α_k ordered according to the corresponding eigenvalues.

The transformation matrix A was calculated by first concatenating training and test data in order to ensure that the extracted manifold will encompass also the test data. However since LPP supports exact out-of-sample extension, the matrix A could also be calculated by using the training data alone and then used to transform any new data set.

The method was assessed on 9 hours recordings including 101 marked spikes. The LPP method has been compared against other dimensionality reduction techniques [41]. The Linear Local Tangent Space Alignment (LLTSA) [53] performed slightly better in respect to $F - score$ but we chose LPP due to its highest sensitivity which is more important given the small value of false positive rate. The Neighborhood Preserving Embedding (NPE) [54], PCA, Maximally Collapsing Metric Learning (MCML) [55], Stochastic Proximity Embedding (SPE) [56] and Diffusion Maps [57] also had high sensitivity with increased however false positive rate.

Tensor decomposition of spatiotemporal signals

Raw EEG signals are naturally born with more than two modes (dimensions) of time and space and represented by a multi-way array (tensor). In addition, the process of feature extraction produces structured high-order multi-way arrays that are usually very high dimensional, with large amount of redundancy, while occupying only a subspace of the input space. However, most previous research works in epileptic and non-epileptic events classification treated EEG features as concatenated vectors (i.e. matrix representation with observations in the rows and features in the columns) in a very high-dimensional space neglecting the inherent structure and correlation in the original feature space. Although matrix representation is suitable for many datasets, it is not always a natural representation because it assumes the existence of a single target variable and lacks a means of modeling dependencies between other features [58][59].

Motivated by the above, in [60], we compare the commonly used matrix representation in which features are concatenated from all channels in order to capture the total spatiotemporal context with a tensor-based scheme which extracts signature features to feed the classification models. TUCKER decomposition [24] is applied to learn the essence of original, high-dimensional domain of feature space and extract a multi-linear discriminative subspace. The proposed scheme

reduced dramatically the computational complexity of the subsequent classification step, which now was performed efficiently in a lower dimensional feature space. The advantage in terms of computational cost relied on the notion that once the mapping (from the original feature space to a reduced space) was learned, its application to unknown EEG segments would only require a few matrix multiplications. The block diagram of the proposed tensor-based scheme is shown in Fig.3.8.

The parameterization of the brain signals was based on the temporal and spectral information in the EEG channels derived from linear and nonlinear signal analysis [4]. Then TUCKER-2 was applied to extract the discriminative multi-linear subspace. Given a third order tensor $F_V \in \mathcal{R}^{M \times Q \times K}$, where M denotes the number of channels, Q the number of extracted (temporal and spectral) EEG features for each of the M channels and K the total number of windowed frames (epochs), its TUCKER-2 model, expressed as a decomposition into two basis factors and a core tensor, is defined as:

$$F_V = G \times_1 A \times_2 B$$

with the symbol $\times_n$ denoting the n-mode product of a tensor with a matrix along the mode-n (i.e. tensor unfolding in the direction of the n-th dimension) [59]. $A \in \mathcal{R}^{M \times R_1}$ and $B \in \mathcal{R}^{Q \times R_2}$ are the basis factors (projection filters) and $G \in \mathcal{R}^{R_1 \times R_2 \times K}$ the core tensor. R_1 and R_2 are the number of selected components (set to two in our experiments). The core tensor G consists of signature features of F_V projected onto the factor subspace spanned by A and B. This low-dimensional tensor was matricized and used to train the classification model. We used Tucker decomposition instead of canonical polyadic decomposition (CPD) [61] due to its superior flexibility. Tucker model enables all the components of each mode (dimension) to interact with each other through the mean of the core tensor, whereas in CPD a component in a certain mode can be linked to only a single component in another mode.

Locally linear embedding of imaging profiles

In [62], we proposed a novel framework which extends manifold learning techniques to help discriminate brain lesions from various normal tissue image profiles established as patterns of normal appearances in healthy subjects. The method used locally linear embedding (LLE) [43] to create normality distributions in different locations of the brain where the distances of new mapped images are considered as deviations from healthy areas. To select the K-closest neighbor for each image patch, we adopted an approach similar to a fuzzy block matching approach which avoids the constraint of a strict one-to-one pairing [63]. A smooth mapping function estimating the relationship between ambient and manifold spaces was calculated as a joint distribution to map unseen test images. Towards this end, densities were replaced by kernel functions using Nadaraya-Watson kernel regression, under a conditional expectation setting [64]. From the intrinsic coordinates of the new point, we approximated it's deviation from the healthy distribution by calculating the geodesic distance based on the shortest path to the tangent plane of the manifold medoid point (center of the manifold). The originality of our method resides on the training using only healthy tissue images, which allows to detect the presence of brain abnormalities, such as white matter lesions (WMLs) in test images, and the definition of physiological difference from normal images

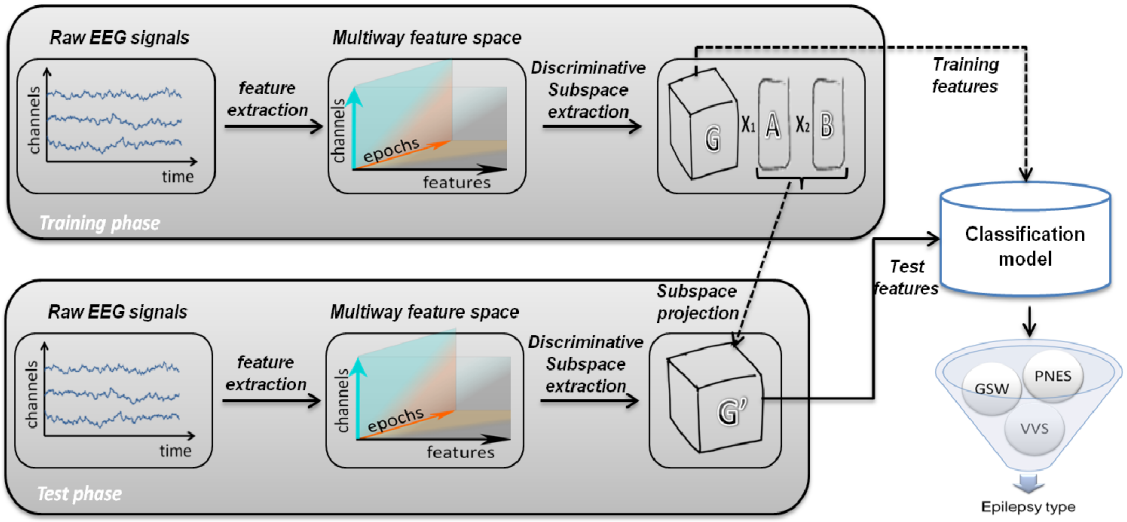


Figure 3.8: Tensor decomposition for EEG-based classification of episodes of transient loss of consciousness.

within manifold space. We presented WML detection results on FLAIR MRI scans of elderly patients with diabetes obtained in a clinical study.

3.3 Integration of clinical, morphological and imaging characteristics

One main challenge researchers in biomedical data mining face is the heterogeneity of the data suggested over different studies as relevant for a specific diagnostic task. The difficulty relies not only in the collection of all possible data, but also in their analysis since many of the standard data mining techniques are not capable of handling all types of variables, such as categorical variables. In the following, two studies [65][66] are presented in which different type of information is integrated for computer-assisted diagnosis of high grade gliomas.

The prediction of prognosis in high grade gliomas is poor in the majority of patients. In [65] our aim was to test whether multivariate prediction models constructed by machine-learning methods provide a more accurate predictor of prognosis in high grade gliomas than histopathologic classification. We calculated and analyzed the relationship of 55 categorical or continuous variables, which included clinical findings and tumor pathology descriptors obtained by visual inspection of conventional MR imaging and also imaging characteristics calculated from DTI and rCBV maps. Typical variables that have been used by others include localization, mass effect, T1 contrast enhancement, T2, diffusion, and perfusion signal intensity. We have combined such nominal variables ranked by an expert with attributes extracted automatically from ROIs, to achieve a more complete representation. The scoring pattern of the selected variables is described in [65]. A variable selection method was applied to identify the overall most important variables. These

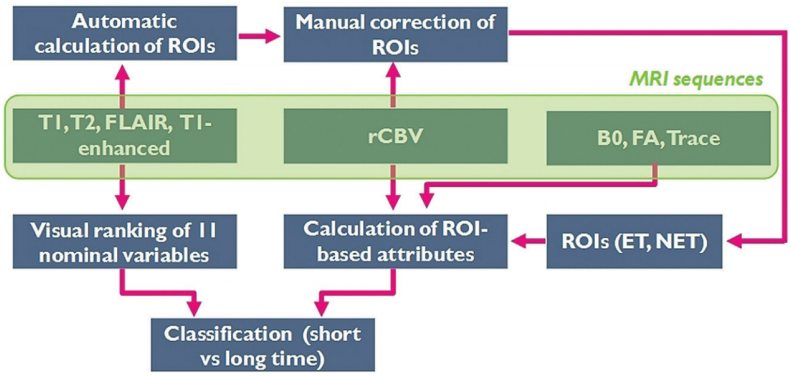


Figure 3.9: Diagram of the computer-based methodology for prediction of survival (long/short = more/less than 18 months). Conventional MR imaging was used to visually rank 11 variables characterizing the tumor, whereas DTI and rCBV were used to extract ROI-based imaging attributes.

variables were then used to construct a prediction model based on a decision tree algorithm. Decision trees provide a good solution for heterogeneous data integration due to their flexibility in handling both numerical and categorical data. Data analysis was performed in several stages, as shown in Fig.3.9.

We tested several variable selection algorithms and selected the one with overall best classification performance. The selected method searched over the variables following the scatter search algorithm [67] and defined the predictive value of each subset of variables by using a wrapper approach[68]. Scatter search operates on a group of subsets of variables, which constitute good solutions, in respect to special criteria such as diversity. The subsets are linearly combined, and a local search procedure is applied to update the initial group and incorporate good solutions. On each subset of variables, the wrapper builds a classifier by applying an induction learning algorithm. The variables subset with the highest classification accuracy, estimated by cross-validation, is selected. These steps are repeated until a stopping condition is met. Kaplan-Meier survival curves on 74 high grade gliomas showed that the constructed prediction model classified malignant gliomas in a manner that better correlates with clinical outcome than standard histopathology.

Moreover, in [66], a retrospective analysis of MR images from glioblastoma (GBM) patients was conducted aiming to assess the ability of quantitative and qualitative imaging features in predicting the O6-methylguanine-DNA-methyltransferase gene (MGMT) methylation status noninvasively. The MGMT methylation status has been shown to be associated with improved outcomes in patients with GBM and may be a predictive marker of sensitivity to chemotherapy. However, determination of the MGMT promoter methylation status requires tissue obtained via surgical resection or biopsy. A noninvasive and reliable surrogate method of determining MGMT status could serve as an alternative (or a complement) to biopsy. There has been only one report on predicting MGMT methylation status in GBM patients from MRI data by space-frequency analysis [69]. In that study, signal intensities and mean spectra from MR images were used to predict MGMT methylation status with a maximum accuracy of 71%. To the best of our knowledge, that

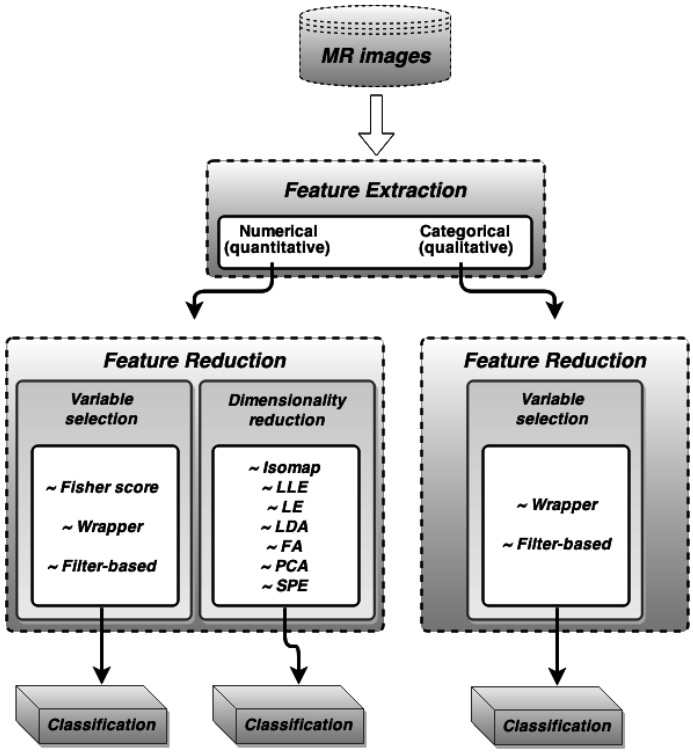


Figure 3.10: Dimensionality reduction techniques used in our recent work [66]. Isomap, isometric feature mapping; LLE, locally linear embedding, LE, Laplacian eigenmaps; LDA, linear discriminant analysis; FA, factor analysis; PCA, principal components analysis; SPE, stochastic proximity embedding.

was the first investigation of the potential of MR 3D volumetrics to predict MGMT methylation status. In clinical practice, volumetric analysis remains a difficult task, and clinically relevant MRI features are typically assessed qualitatively. In [66], we hypothesized that certain features derived from standard MRI sequences reflect differences in MGMT promoter methylation status in GBM patients.

For each patient, three board-certified neuroradiologists independently reviewed T1 and T1ce images as well as FLAIR images and categorized 24 qualitative variables for qualitative imaging features (based on the VASARI feature set for GBM) describing the size, location, and morphology of the tumoral region. Quantitative image analysis was performed with image segmentation and registration software (semi-automated process), instead of explicit manual tumor delineation, using the open-source 3D Slicer package (slicer.org). Hence, we were able to evaluate subtleties in a consistent manner and reduce the potential for intra- and inter-rater bias, increasing the likelihood of accurate noninvasive prediction of MGMT status. The T1ce image was registered and rigidly aligned with the FLAIR image, followed by resampling when the voxel sizes of the FLAIR

and T1ce images were different. Three distinct regions (edema/invasion, tumor enhancement, and necrosis) were automatically segmented based on relevant MRI sequences and subsequently reviewed by trained neuroradiologists until consensus was reached. Then, volumetric analysis was performed by measuring the volume of each of the three abnormal regions. Ten quantitative features (additional to the 24 qualitative features described in the previous section) were calculated from the volumes. The lengths of the major and minor axes of the tumor were also measured. We used different feature reduction methods to separately process numerical (quantitative) and categorical (qualitative) MRI variables as illustrated in Fig.3.10. The feature reduction methods were assessed by extensive computational tests of the accuracy of classification of GBM tumors by MGMT methylation status.

Personal contribution

In respect to biosignal analysis, my personal contribution was less significant in the works [19][70][71], in which I developed tools for data processing and pattern discovery mostly for offline analysis, whereas my involvement in the data stream management systems for online monitoring was reduced. The application of manifold learning techniques for compact representation of EEG pattern within a supervised classification framework [49] was fully my contribution from computational point of view, since data selection and annotation was performed by medical experts. In respect to medical image analysis, I mainly leaded the studies for brain tumor analysis [27][20][21][65]. More work on brain tumor classification using imaging features and machine learning algorithms [72] will be presented in Chapter 5. In [62] I contributed mostly in data preprocessing and in the semi-supervised framework formulation (where data only from healthy subjects are used for training) in which my main goal was to segment the pathological region rather than classify whole image patches into healthy or abnormal. More relevant personal work will be presented in chapter 4. For the analysis of spatiotemporal imaging data [30], I did not contribute to the graph-cut based tumor segmentation refinement algorithm and therefore I don't present it in this thesis. I only built upon the methods related to feature extraction, feature selection and classification, which I later also extended and applied for brain tumor classification [20]. In respect to molecular data analysis I was leading the studies [36][37] and presented my personal work in [38].

Bibliography

[1] M. Valderrama, S. Nikolopoulos, C. Adam, V. Navarro, and M. Le Van Quyen, "Patient-specific seizure prediction using a multi-feature and multi-modal eeg-ecg classification", in *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, Springer, 2010, pp. 77–80.

[2] S. Nasehi and H. Pourghassem, "Seizure detection algorithms based on analysis of eeg and ecg signals: A survey", *Neurophysiology*, vol. 44, no. 2, pp. 174–186, 2012.

[3] B. R. Greene, G. B. Boylan, R. B. Reilly, P. de Chazal, and S. Connolly, "Combination of eeg and ecg for improved automatic neonatal seizure detection", *Clinical neurophysiology*, vol. 118, no. 6, pp. 1348–1359, 2007.

[4] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, M. Richardson, and V. Megalooikonomou, "Seizure detection using eeg and ecg signals for computer-based monitoring, analysis and management of epileptic patients", *Expert Systems with Applications*, vol. 42, no. 6, pp. 3227–3233, 2015.

[5] I. Mporas, V. Tsirka, E. Zacharaki, M. Koutroumanidis, and V. Megalooikonomou, "Evaluation of time and frequency domain features for seizure detection from combined eeg and ecg signals", in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, ACM, 2014, p. 28.

[6] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, and V. Megalooikonomou, "Online seizure detection from eeg and ecg signals for monitoring of epileptic patients", in *Artificial Intelligence: Methods and Applications*, Springer, 2014, pp. 442–447.

[7] M. S. Manikandan and K. Soman, "A novel method for detecting r-peaks in electrocardiogram (ecg) signal", *Biomedical Signal Processing and Control*, vol. 7, no. 2, pp. 118–128, 2012.

[8] I. Kononenko, "Estimating attributes: Analysis and extensions of relief", in *European conference on machine learning*, Springer, 1994, pp. 171–182.

[9] T. Ditterrich, "Machine learning research: Four current direction", *Artificial Intelligence Magzine*, vol. 4, pp. 97–136, 1997.

[10] Y. Sun and D. Wu, "A relief based feature extraction algorithm.", in *SDM*, SIAM, 2008, pp. 188–195.

[11] R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, "Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe)", *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.

[12] A. Krumholz, "Nonepileptic seizures: Diagnosis and management.", *Neurology*, vol. 53, no. 5 Suppl 2, S76–83, 1998.

[13] A. Moya, R. Sutton, F. Ammirati, J.-J. Blanc, M. Brignole, J. B. Dahm, J.-C. Deharo, J. Gajek, K. Gjesdal, A. Krahn, *et al.*, "Guidelines for the diagnosis and management of syncope (version 2009)", *European heart journal*, vol. 30, no. 21, pp. 2631–2671, 2009.

[14] M. Poulos, F. Georgiacodis, V. Chrissikopoulos, and A. Evagelou, "Diagnostic test for the discrimination between interictal epileptic and non-epileptic pathological eeg events using auto–cross-correlation methods", *American journal of electroneurodiagnostic technology*, vol. 43, no. 4, pp. 228–240, 2003.

[15] S. Papavlasopoulos, M. Poulos, and A. Evangelou, "Feature extraction from interictal epileptic and non-epileptic pathological eeg events for diagnostic purposes using lvq1 neural network", in *Proceedings of seventh International Conference on Mathematics Methods in Scattering Theory and Biomedical Technology*, World Scientific, 2005.

[16] E. Pippa, E. I. Zacharaki, I. Mporas, V. Tsirka, M. P. Richardson, M. Koutroumanidis, and V. Megalooikonomou, "Improving classification of epileptic and non-epileptic eeg events by feature selection", *Neurocomputing*, vol. 171, pp. 576–585, 2016.

[17] E. Pippa, E. I. Zacharaki, I. Mporas, V. Megalooikonomou, V. Tsirka, M. Richardson, and M. Koutroumanidis, "Classification of epileptic and non-epileptic eeg events", in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, IEEE, 2014, pp. 87–90.

[18] E. Pippa, E. I. Zacharaki, M. Koutroumanidis, and V. Megalooikonomou, "Data fusion for paroxysmal events' classification from eeg", *Journal of Neuroscience Methods*, 2017, in press.

[19] I. Mporas, P. Korvesis, E. I. Zacharaki, and V. Megalooikonomou, "Sleep spindle detection in eeg signals combining hmms and svms", in *Engineering Applications of Neural Networks*, Springer, 2013, pp. 138–145.

[20] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme", *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[21] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Mri-based classification of brain tumor type and grade using svm-rfe", in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, IEEE, 2009, pp. 1035–1038.

[22] J. M. Provenzale, S. Mukundan, and D. P. Barboriak, "Diffusion-weighted and perfusion mr imaging for brain tumor characterization and assessment of treatment response 1", *Radiology*, vol. 239, no. 3, pp. 632–649, 2006.

[23] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, *et al.*, "Advances in functional and structural mr image analysis and implementation as fsl", *Neuroimage*, vol. 23, S208–S219, 2004.

[24] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images", *Medical image analysis*, vol. 5, no. 2, pp. 143–156, 2001.

[25] S. M. Smith, "Bet: Brain extraction tool", *FMRIB TR00SMS2b, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain), Department of Clinical Neurology, Oxford University, John Radcliffe Hospital, Headington, UK*, 2000.

[26] T. Tan, "Rotation invariant texture features and their use in automatic script identification", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 7, pp. 751–756, 1998.

[27] E. I. Zacharaki, R. Verma, S. Chawla, E. R. Melhem, R. Wolf, and C. Davatzikos, "Towards predicting neoplastic recurrence with multi-parametric mr", in *ISMRM 16th Scientific Meeting and Exhibition 2008*, 2008.

[28] R. Verma, E. I. Zacharaki, Y. Ou, H. Cai, S. Chawla, S.-K. Lee, E. R. Melhem, R. Wolf, and C. Davatzikos, "Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of mr images", *Academic radiology*, vol. 15, no. 8, pp. 966–977, 2008.

[29] N. Morita, M. Harada, E. Zacharaki, P. Bhatt, S. Chawla, E. Melhem, and H. Nishitani, "Correlation between diffusion tensor and perfusion imaging in segmented enhancing lesion with high grade glioma", in *Joint Annual Meeting ISMRM-ESMRMB 2010*, 2010.

[30] Y. Zheng, S. Englander, S. Baloch, E. I. Zacharaki, Y. Fan, M. D. Schnall, and D. Shen, "Step: Spatiotemporal enhancement pattern for mr-based breast tumor diagnosis", *Medical physics*, vol. 36, no. 7, pp. 3192–3204, 2009.

[31] M.-K. Hu, "Visual pattern recognition by moment invariants", *IRE transactions on information theory*, vol. 8, no. 2, pp. 179–187, 1962.

[32] A. G. Tzalavra, E. I. Zacharaki, N. N. Tsiaparas, F. Constantinidis, and K. S. Nikita, "A multiresolution analysis framework for breast tumor classification based on dce-mri", in *Imaging Systems and Techniques (IST), 2014 IEEE International Conference on*, IEEE, 2014, pp. 246–250.

[33] B. Siva Kumar and S. Nagaraj, "Discrete and stationary wavelet decomposition for image resolution enhancement", *International Journal of Engineering Trends and Technology (IJETT)*, vol. 4, no. 7, pp. 2885–2889, 2013.

[34] J. E. Fowler, "The redundant discrete wavelet transform and additive noise", *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 629–632, 2005.

[35] J. S. Fetrow and J. Skolnick, "Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and t1 ribonucleases", *Journal of molecular biology*, vol. 281, no. 5, pp. 949–968, 1998.

[36] A. Amidi, S. Amidi, D. Vlachakis, N. Paragios, and E. I. Zacharaki, "A machine learning methodology for enzyme functional classification combining structural and protein sequence descriptors", in *Bioinformatics and Biomedical Engineering*, Springer, 2016, pp. 728–738.

[37] ——, "Automatic single- and multi-label enzymatic function prediction by machine learning", *IEEE/ACM Trans. Computational Biology and Bioinformatics*, 2016, under review.

[38] E. I. Zacharaki, "Prediction of protein function using a deep convolutional neural network ensemble", *PeerJ Computer Science*, submitted, 2016.

[39] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences", *Journal of molecular biology*, vol. 147, no. 1, pp. 195–197, 1981.

[40] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins", *Journal of molecular biology*, vol. 48, no. 3, pp. 443–453, 1970.

[41] L. van der Maaten, E. Postma, and H. van den Herik, "Dimensionality reduction: A comparative review", Tilburg, Netherlands: Tilburg Centre for Creative Computing, Tilburg University, Technical Report: 2009-005, Tech. Rep., 2009.

[42] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[43] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[44] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[45] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

[46] R. Timofte and L. Van Gool, "Iterative nearest neighbors for classification and dimensionality reduction", in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, IEEE, 2012, pp. 2456–2463.

[47] T. Kohonen, "Self-organizing maps, vol. 30 of springer series in information sciences", *Ed: Springer Berlin*, 2001.

[48] E. I. Zacharaki, K. Garganis, I. Mporas, and V. Megalooikonomou, "Spike detection in eeg by lpp and svm", in *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, IEEE, 2014, pp. 668–671.

[49] E. I. Zacharaki, I. Mporas, K. Garganis, and V. Megalooikonomou, "Spike pattern recognition by supervised classification in low dimensional embedding space", *Brain Informatics*, vol. 3, no. 2, pp. 73–83, 2016.

[50] C. J. James, "Detection of epileptiform activity in the electroencephalogram using artificial neural networks", 1997.

[51] X. Niyogi, "Locality preserving projections", in *Neural information processing systems*, MIT, vol. 16, 2004, p. 153.

[52] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering.", in *NIPS*, vol. 14, 2001, pp. 585–591.

[53] T. Zhang, J. Yang, D. Zhao, and X. Ge, "Linear local tangent space alignment and application to face recognition", *Neurocomputing*, vol. 70, no. 7, pp. 1547–1553, 2007.

[54] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding", in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, IEEE, vol. 2, 2005, pp. 1208–1213.

[55] A. Globerson and S. T. Roweis, "Metric learning by collapsing classes", in *Advances in neural information processing systems*, 2005, pp. 451–458.

[56] D. K. Agrafiotis, "Stochastic proximity embedding", *Journal of computational chemistry*, vol. 24, no. 10, pp. 1215–1221, 2003.

[57] R. R. Coifman and S. Lafon, "Diffusion maps", *Applied and computational harmonic analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[58] D. Tao, X. Li, W. Hu, S. Maybank, and X. Wu, "Supervised tensor learning", in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, IEEE, 2005, 8–pp.

[59] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[60] V. G. Kanas, E. I. Zacharaki, E. Pippa, V. Tsirka, M. Koutroumanidis, and V. Megalooikonomou, "Classification of epileptic and non-epileptic events using tensor decomposition", in *Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Conference on*, IEEE, 2015, pp. 1–5.

[61] R. Bro, "Parafac. tutorial and applications", *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.

[62] S. Kadoury, G. Erus, E. I. Zacharaki, N. Paragios, and C. Davatzikos, "Manifold-constrained embeddings for the detection of white matter lesions in brain mri", in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, IEEE, 2012, pp. 562–565.

[63] F. Rousseau, P. A. Habas, and C. Studholme, "Human brain labeling using image similarities", in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, IEEE, 2011, pp. 1081–1088.

[64] B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi, "Population shape regression from random design data", *International journal of computer vision*, vol. 90, no. 2, pp. 255–266, 2010.

[65] E. Zacharaki, N. Morita, P. Bhatt, D. O'rourke, E. Melhem, and C. Davatzikos, "Survival analysis of patients with high-grade gliomas based on data mining of imaging variables", *American Journal of Neuroradiology*, vol. 33, no. 6, pp. 1065–1071, 2012.

[66] V. G. Kanas, E. I. Zacharaki, G. A. Thomas, P. O. Zinn, V. Megalooikonomou, and R. R. Colen, "Learning mri-based classification models for mgmt methylation status prediction in glioblastoma", *Computer Methods and Programs in Biomedicine*, 2016, under review.

[67] M. Laguna and R. Marti, *Scatter search: Methodology and implementations in c*. Springer Science & Business Media, 2012, vol. 24.

[68] R. Kohavi and G. H. John, "Wrappers for feature subset selection", *Artificial intelligence*, vol. 97, no. 1, pp. 273–324, 1997.

[69] S. Drabycz, G. Roldán, P. De Robles, D. Adler, J. B. McIntyre, A. M. Magliocco, J. G. Cairncross, and J. R. Mitchell, "An analysis of image texture, tumor location, and mgmt promoter methylation in glioblastoma using magnetic resonance imaging", *Neuroimage*, vol. 49, no. 2, pp. 1398–1405, 2010.

[70] V. Megalooikonomou, D. Triantafyllopoulos, E. I. Zacharaki, and I. Mporas, "Cyberphysical systems for epilepsy and related brain disorders: Multi-parametric monitoring and analysis for diagnosis and optimal disease management", in, S. N. Voros and P. C. Antonopoulos, Eds. Springer International Publishing, 2015, ch. DSMS and Online Algorithms, pp. 271–279.

[71] ——, "Cyberphysical systems for epilepsy and related brain disorders: Multi-parametric monitoring and analysis for diagnosis and optimal disease management", in, S. N. Voros and P. C. Antonopoulos, Eds. Springer International Publishing, 2015, ch. Offline Analysis Server and Offline Algorithms, pp. 239–254.

[72] E. I. Zacharaki, V. G. Kanas, and C. Davatzikos, “Investigating machine learning techniques for mri-based classification of brain neoplasms”, *International journal of computer assisted radiology and surgery*, vol. 6, no. 6, pp. 821–828, 2011.

Chapter 4: Unsupervised and semi-supervised learning

Generally, the problems of machine learning is based on the extraction of important patterns and trends and aims at function estimation for classification, prediction or modeling. The process of learning from data can be unsupervised, supervised, or semi-supervised. In unsupervised learning the goal is to describe the associations and patterns among a set of input measures. Based on the problem (clustering or prediction) and the background knowledge of the space sampled, various methods can be used, such as density estimation (to estimate some underlying probability density function for prediction), k-means clustering (to group unlabeled real valued data), k-modes clustering (to group unlabeled categorical data), and others. In supervised learning, the goal is to predict the value of an outcome measure, that could be categorical (in classification) or quantitative (in regression), based on a set of input variables (features). The prediction model is trained using a set of annotated examples (feature/label pairs). A very popular supervised learning technique is the Support Vector Machine (SVM). Since labeled data require human effort and are often difficult or costly to generate, whereas unlabeled data are abundant, semi-supervised learning, a technique that lies in between unsupervised and semi-supervised, is often the only feasible solution. Semi-supervised learning uses a large amount of unlabeled data, together with a few labeled data, to build better classifiers. Although convenient because it needs less labeled data, it requires a good matching of problem structure with model assumption to account for the lack of training data.

This chapter focuses in applications using unsupervised, or semi-supervised learning, whereas in the next chapter supervised learning techniques are applied for pattern detection and classification, as well as segmentation of regions of interest. In the following, two methodologies are presented that aim to segment pathological regions in brain images without the use of disease annotations. The method learns the probability density function (pdf) of normal imaging phenotypes (images without disease) in an unsupervised fashion and identifies abnormalities as deviations from normality. Two approaches are described that handle the high-dimensionality of the data. Then, in the next sections, clustering and pattern similarity applications are presented using biosignals, medical images, gene expression data or biological networks.

4.1 Statistical modeling of imaging phenotypes

We have developed a method for capturing the statistical variation of normal imaging phenotypes, with emphasis on brain structure [1][2]. The method aims to estimate the statistical variation of

a normative set of images from healthy individuals, and identify abnormalities as deviations from normality. The present approach goes beyond the standard anomaly detection techniques in that it not only characterizes the data vector as normal or abnormal, but also locates which part of the vector includes the anomaly. It is used in a scenario where data vary in their biggest part according to an expected or predictable distribution that can be statistically modeled from a set of normal data and also vary in small abnormal areas that cannot be explained by the same statistical model. Such abnormalities might be due to structural or morphological differences beyond the expected morphological variability and might indicate damage, disease, or any kind of artifacts.

A direct estimation of the statistical variation of the entire volumetric image is challenged by the high-dimensionality of images relative to the typically much smaller sample sizes. To overcome this limitation, the proposed method iteratively samples a large number of lower dimensional subspaces that capture image characteristics ranging from fine and localized to coarser and relatively more global. The marginal probability density functions pertaining to the selected features is estimated through a PCA model, in conjunction with an "estimability" criterion that limits the dimensionality of the estimated probability densities according to the available sample size and the underlying anatomy variation. A test sample is iteratively projected to the subspaces of these marginals as determined by the PCA models, and its trajectory until convergence delineates potential abnormalities (deviations from the normative database). The method is applied to the segmentation of various types of brain lesions, and to simulated data on which the superiority of the iterative method over straight PCA is demonstrated. In [2] a "target-specific" feature selection strategy was introduced within each subspace to further reduce the dimensionality, by considering only imaging characteristics that are present in a test subject's image, rather than all possible characteristics found in the entire population. The method is described with more details next.

Consider n medical images of a normative population coregistered to a common domain Ω as realizations of a d-dimensional random vector I, consisting of d scalar random variables $[x_1, x_2, \ldots, x_d]$ corresponding to image voxels. The joint *pdf* of I,

$$\phi(I) = \phi(x_1, x_2, ..., x_d)$$

describes the relative likelihood for I to be observed. Images for which $\phi(I) \geq t$ lie in a hypervolume, which (for the purposes of this paper in which we consider anatomical brain images) we call the subspace of normal anatomy, in the d dimensional space. If we can estimate $\phi(\mathcal{I}^t)$, for any given test image $\mathcal{I}^t$ to be compared with the normative population, the likelihood of being abnormal can be calculated for this new image. Also, if $\mathcal{I}^t$ is abnormal, $\mathcal{I}^*$, the most similar image to $\mathcal{I}^t$ that is within normal variation, can be obtained by projecting $\mathcal{I}^t$ to this hypervolume. The difference between $\mathcal{I}^t$ and $\mathcal{I}^*$ provides an image of the abnormality patterns in $\mathcal{I}^t$. In most practical applications, however, $\phi(I)$ is unknown, while only a small set of $n << d$ training samples is available for estimating it.

Estimating $\phi(I)$ is a challenging problem, particularly when d is very high. Furthermore, the underlying distribution of the data is generally unknown. Previous studies have shown that images of complex objects, e.g. faces, brains, lie on a lower dimensional nonlinear submanifold embedded in the high dimensional space. Many manifold learning methods have been proposed to capture the low dimensional manifold structure from the data samples in the high dimensional space [3],

[4]. They have also been used for representing brain images based on diffeomorphic deformations [5][6]. However, nonlinear models are difficult to estimate, and require a large number of samples for learning the underlying manifold structure. For this reason, we take a different approach, motivated by our primary aim being to detect abnormalities in a test image. In particular, we sample from the image domain a large number of lower-dimensional subspaces, and estimate the distribution of the data by estimating *pdf*s of each subspace using a linear model, assuming that regional imaging statistics can be approximated to a large extent by Gaussian distributions. The likelihood of a new image is evaluated by testing the likelihoods of its projections to the smaller subspaces in order to detect patterns of abnormality on it. An image of potential abnormalities is formed by an iterative procedure that finds a path from $\mathcal{I}^t$ to the normal hypervolume.

While $\phi(I)$ cannot be estimated accurately from a limited number of training images, if a subspace is small enough, the *pdf* of the image projection on it can be estimated more reliably. If the projection is selection of certain voxels, effectively, this would estimate the marginal pdf involving those voxels. Assume that a large number of partially overlapping subspaces $S = \{\omega_1, ..., \omega_r\}$ are sampled from Ω, where

$$\Omega = \bigcup \omega_i, i \in \{1, .., r\},$$

with a high degree of redundancy, i.e., each voxel is included in many ω_i's.

The likelihood of a new test image's projection $\mathcal{I}^t_\omega$ to a subspace $\omega \in S$ can be calculated by estimating the *pdf* $\phi(I^t_\omega)$ using the set of n training samples. This estimation is defined according to the statistical modeling approach used herein. From $\mathcal{I}^t$, an image without abnormalities (i.e. an image that represents the closest point in the hypervolume of normals according to the estimated subspace models) can be reconstructed by minimizing the following energy function:

$$E(\mathcal{I}^*) = \|\mathcal{I}^* - \mathcal{I}^t\|^2_{l_2} + \alpha \sum_{i=1}^{r} L(\mathcal{I}^*_{\omega_i}) \tag{4.1}$$

The data term reflects the image dissimilarity, computed as the l_2-norm of the difference between the test image and the reconstructed image. The model term reflects the likelihood of being abnormal according to the subspace model and is defined as the Mahalanobis distance exceeding a predefined threshold (t) for normality [2].

For solving eq. 4.1, we use an iterative strategy employing a block-coordinate descent technique, where in order to minimize a multivariable function, optimization with respect to different (smaller) subsets of variables is iteratively achieved. We solve the optimization problem for one image subspace ω_i at a time by calculating the optimal reconstruction of $\mathcal{I}^t$ within ω_i (Fig.4.1), and iterate over a large number of subspaces. The motivation for selecting an iterative approach is two-fold: first we do not partition the image domain into a pre-defined number of disjoint blocks, as done in block-PCA [7], 2D PCA [8], or the spectral graph partitioning based approach used in [9]. Instead, in order to capture image characteristics in different scales and locations, a large number of possibly overlapping image patches are sampled from Ω in a random fashion, which makes an iterative approach more appropriate. Second, we assume that if a solution is within normal range from the perspective of a sufficiently high number of marginals, it is likely to be within normal range with respect to the overarching high dimensional *pdf*. Convergence of this

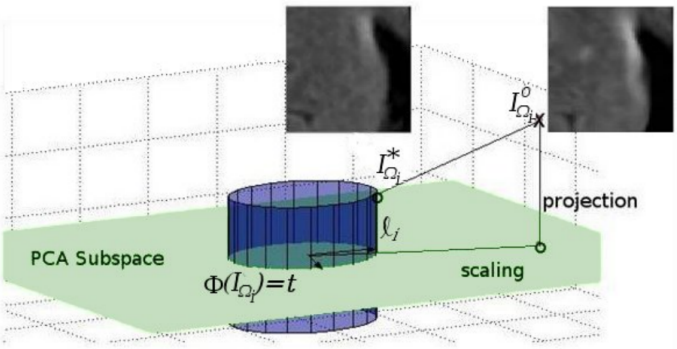


Figure 4.1: Illustration of the reconstruction of an image patch.

process can be monitored, and the process can be terminated when convergence has been achieved, without a priori knowledge of which and how many subspaces need to be used.

Subspace Representation S may be obtained in various ways. Generally speaking, an image subspace may consist of any subset of arbitrary image voxels or set of values derived from image voxels. Here we limit ourselves to rectangular image patches of varying sizes, which correspond to local neighborhoods around voxels in different spatial positions. Neighboring voxels are highly correlated in natural images, and the intrinsic dimensionality of image patches are generally much lower than their actual dimensionality, which makes them more reliably estimable.

An image patch consists of voxels in a rectangular block of size $s_i \in \Re^3$ around a selected seed voxel $p_i \in \Re^3$. At each iteration i, a patch is drawn from Ω by random selection of the seed location and patch size. An edge detector [10] is first applied on $\mathcal{I}^t$ to restrict the set of all possible seed points, $\mathcal{P}$, to voxels with relatively richer information content. To make sure that the space is well sampled, a weighted random selection strategy is used. A weight value is assigned to each $p \in \mathcal{P}$, and the selection is done such that p has a probability of being selected proportional to its weight. Initially, an equal weight value is assigned to each p. At the end of each iteration, the weights of all $p \in \mathcal{P}$ that are within the selected image patch are decreased by a constant factor, so that these points will have a lower probability of being selected in subsequent iterations, to encourage this process to sample the entire domain fairly evenly.

Block size s_i varies randomly in an interval bounded by predefined minimum and maximum values. In that way, the set of all subspaces capture image characteristics ranging from fine and localized to coarser and relatively more global at different spatial locations of the image domain.

Target Specific Feature Selection We use a target-specific feature selection strategy, i.e. a strategy that aims to estimate from the training set only those imaging aspects that are necessary to process the test image (i.e. the target). The feature selection method first expands the test image patch in a basis that better allows us to determine the important components of it, and then selects an estimable subset of these components for statistical group analysis. Feature selection is based on wavelet transform. A detailed exposition of wavelet based compression techniques is available in [11]. A wavelet transform produces as many coefficients as there are voxels in the image.

However, it provides a more compact representation, such that most of the information is concentrated in a small fraction of the coefficients. Also, wavelet coefficients with larger magnitude are correlated with salient features in the image data. The compression is performed by applying a thresholding operator to the coefficients in order to select a subset of coefficients with the largest values.

Let $\mathbf{d}^t$ be a vector of voxel intensities extracted from image patch ω on $\mathcal{I}^t$, and $\mathbf{d}_i^{Train}, i \in \{1, .., n\}$ be the voxels from the same patch on each training image. By applying a wavelet transform, $\mathbf{d}^t$ may be represented as a linear combination of m predefined wavelet basis functions ψ_j, where m is the dimensionality of $\mathbf{d}^t$ (the number of voxels in the image patch). If we choose the p basis functions with largest coefficients ($p < m$), $\mathbf{d}^t$ is expressed as

$$\mathbf{d}^t = \sum_{j=1}^{p} \mathrm{a}_j \psi_j + \mathcal{R}^t \tag{4.2}$$

where a_j are wavelet coefficients sorted in descending order of absolute value and $\mathcal{R}^t$ is the residual. The feature vector $\mathbf{a}^t \doteq \{\mathrm{a}_j\}_{j=1}^p$ can be used to reconstruct the image after wavelet thresholding. In our application, we apply an iterative algorithm for determining p: starting from $p = m$, the value of p is decreased gradually until the selected coefficients satisfy the estimability criteria (discussed in a subsequent section), which reflect whether the corresponding *pdf* of these coefficients can be estimated from the training normal data. In this way we select the largest set of coefficients whose *pdf* is estimable. p takes different values for different image patches, but it always satisfies $m > p > n$.

Note that the basis wavelets $\{\psi_j\}_{j=1}^p$ are selected based on the target data only. Each training image patch $\mathbf{d}_i^{Train}$ is projected to the selected basis to obtain the training feature matrix $\mathbf{A} \in \Re^{n \times p}$ which has, in each row, p wavelet coefficients from each training image patch.

Statistical Model Constrained Reconstruction Within a Subspace Various modeling approaches may be used for estimating the *pdf* of the selected coefficients within an image patch. We applied PCA, where a number of principal components that account for as much of the variability in the data as possible are calculated, based on the assumption that the data follow a Gaussian distribution. One of our assumptions here is that regional statistics can be approximated by Gaussian distributions, even if the distribution of the entire image is highly non-Gaussian.

Let $\mathbf{A} \in \Re^{n \times p}$ be the data matrix, consisting of feature vectors obtained from n training samples and $\bar{\mathbf{a}} \in \Re^{1 \times p}$ the mean of the rows of $\mathbf{A}$. Let $\mathbf{C}$ be the sample covariance matrix of the mean-centered data matrix. An eigenvalue decomposition is applied to calculate $\boldsymbol{\lambda}$, the vector of $n - 1$ non-zero eigenvalues of $\mathbf{C}$, sorted in descending order, and $\mathbf{Q} \in \Re^{p \times (n-1)}$ the $n - 1$ eigenvectors of $\mathbf{C}$. When projected to the space spanned by $\mathbf{Q}$, a data vector $\mathbf{a}^t \in \Re^{1 \times p}$, which consists of the coefficients selected from the target image, can be represented by its projection vector (or feature vector) $\mathbf{v}$:

$$\mathbf{v} = \mathbf{Q}^T (\mathbf{a}^t - \bar{\mathbf{a}})^T \ . \tag{4.3}$$

As the application of PCA diagonalizes the covariance matrix $\mathbf{C}$, the *pdf* of $\mathbf{a}^t$ can be calculated by

$$\phi(\mathbf{a}^t) = \phi(\mathbf{v}) = c \;\; exp\{-\frac{1}{2}\sum_{j=1}^{n-1}\frac{v_j^2}{\lambda_j}\} \;, \tag{4.4}$$

where c is the normalization coefficient, and v_j and λ_j are the j^{th} elements of $\mathbf{v}$ and $\boldsymbol{\lambda}$ respectively. Consequently, an image patch that has a low likelihood can be constrained to have a desired likelihood t on the PCA subspace by scaling down its projection coefficients by a scalar factor q

$$\mathbf{v}^* = q\mathbf{v} \;\; s.t. \;\; \phi(\mathbf{v}^*) = t \tag{4.5}$$

The reconstructed coefficients $\mathbf{a}^*$ can be obtained by projecting $\mathbf{v}^*$ back to the original space:

$$\mathbf{a}^* = (\mathbf{Q}\mathbf{v}^*)^T + \bar{\mathbf{a}} \;. \tag{4.6}$$

Estimability of a Subspace We aim to select features whose respective *pdf's* are reliably estimable from the limited set of training data. We consider that the *pdf* of the coefficient vector derived from a subspace is estimable if a significant fraction γ_v of the overall variance of the data can be explained by a small fraction γ_e of eigenvectors. The thresholds γ_v and γ_e are parameters to be chosen based on the particular application. Normalized eigenvalues

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda} / \sum_{j=1}^{n-1}\lambda_j \tag{4.7}$$

represent the fraction of variance contributed by each eigenvector. We calculate

$$\xi = \left(\underset{x\in\{1,..,n-1\}}{\operatorname{argmin}} \sum_{j=1}^{x}\hat{\lambda}_{ik} \geq \gamma_v \right) /(n-1) \tag{4.8}$$

A feature vector is considered estimable if $\xi < \gamma_e$.

Generation of Voxelwise Abnormality Maps Training and testing images are first registered to a common template using a non-linear registration algorithm that is robust to the presence of abnormalities, since it uses the concept of attribute vectors [12]. A histogram matching method is used to eliminate intensity variations due to scanner differences. The overall procedure for minimizing the energy function defined in eq. 4.1 is applied on each test image $\mathcal{I}^t$, and the closest image being part of normal variation without pathology, $\mathcal{I}^*$, is obtained.

An intensity abnormality map, $\mathcal{I}^{abn}$, is computed as the voxelwise difference between $\mathcal{I}^t$ and $\mathcal{I}^*$. Voxel values on $\mathcal{I}^{abn}$ reflect the amount of abnormality, measured as the difference between observed signal intensity and the estimated normal intensity according to the method. Since different brain regions display different levels of image variations, we adaptively normalized the abnormality map for the test sample at each voxel using abnormality values obtained from the training samples. Specifically, for each training image an abnormality map is computed using the proposed approach with leave-one-out cross validation. At each voxel, the mean and standard

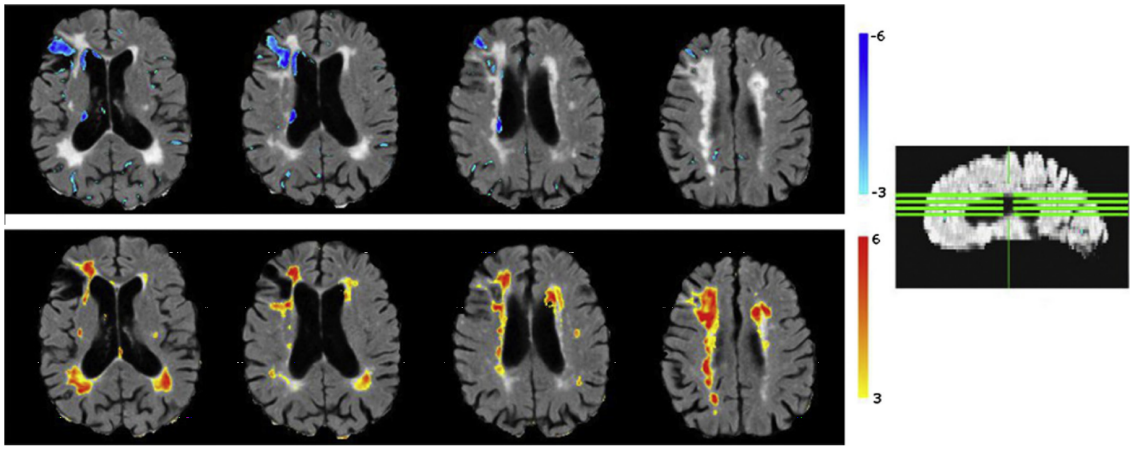


Figure 4.2: Segmentation of abnormalities on a sample image with cortical infarcts, periventricular lesions and atrophy. Positive (red) and negative (blue) normalized abnormality maps overlaid on the image. Positive abnormality map is thresholded at $t \geq 3.1$ (calculated through cross-validation), and negative map at $t \geq 3.5$.

deviation of the abnormality values from all training samples is calculated. $\mathcal{I}^{abn}$ is normalized by calculating the standard (z) score of the raw abnormality score at each voxel, with respect to the distribution estimated from the training data.
The method is applied on the whole MR scans for segmenting a set of common abnormalities of the brain, which may be the result of different brain diseases, such as multiple sclerosis and cerebrovascular disease, or may appear due to normal aging. We were specifically interested in segmenting white matter lesions (WMLs), large cortical infarcts, and periventricular atrophy. On FLAIR images WMLs show up as hyperintensities with respect to surrounding healthy white matter tissue. Cortical infarcts have a necrotic part (with intensity similar to the cerebrospinal fluid) in the cortex surrounded by a hyperintense rim. The periventricular atrophy is the enlargement of the ventricles as a result of the atrophy in surrounding brain tissue. Fig.4.2 shows segmented abnormalities on a test sample with cortical infarcts and periventricular ischemic lesions. Thresholded abnormality maps are overlaid on the original FLAIR image for the visualization of segmented abnormalities on the image. We observe that the method successfully segments brain pathologies of various types.

Learning using distributed estimation

Similarly to the previous section, the aim of the work in [13][14] is to introduce a novel unsupervised scheme for abnormality detection and segmentation in medical images. In order to tackle the high-dimensionality problem, we treat every image as a network of locally coherent image partitions (overlapping blocks). While in [1][2] optimization is performed locally, in [13][14] we formulate and maximize a strictly concave likelihood function estimating abnormality for each partition and fuse the local estimates into a globally optimal estimate that satisfies the consistency constraints, based on a distributed estimation algorithm. The likelihood function, $l(x)$, consists

given initial price vector v that satisfies $E^T v = 0$ (*e.g.* $v = 0$)
repeat

Optimize blocks (separately) to obtain (s_i^*, y_i^*) *for each block i*
$x_i^* = (s_i^*, y_i^*) = argmax_{s_i, y_i}\left(l_i(s_i, y_i) - v_i^T y_i\right)$

Compute average of public variables over each net.
$\hat{z} := (E^T E)^{-1} E^T y^*$

Update the dual variables
$v := v - \alpha_k (E\hat{z} - y^*)$

Figure 4.3: Algorithm solving the maximum likelihood estimation problem in a distributed setting using a subgradient method.

of three term and is formulated as a quadratic programming problem. The first two terms statistically model normality and are used to make the image look like if abnormality were removed. The first term maximizes the probability density function (e.g. assuming a multivariate Gaussian distribution) of the lower dimensional representation in a basis W. The second term reduces the solution space by constraining the solution to remain close to the subspace spanned by W, i.e. it minimizes the residual error. The third term is used to constrain the reconstructed image to be as similar as possible to the original image based on the assumption that the majority of the voxels in the test image are normal. Since this method is unsupervised for the abnormal class and aims to generalize for any kind of abnormality, we do not incorporate any prior for the abnormal areas. However, we focus on the normal class and introduce a confidence measure on the estimation ability of the calculated statistical model. Regions with large variability are much more difficult to model than uniform areas. A confidence map or vector shows the degree of certainty we have on the reconstruction of each parameter (voxel intensity). Parameters with high uncertainty in estimation should not deviate significantly from their original values. This is achieved by penalizing any change on those parameters more than on other parameters. The uncertainty vector is calculated as the average reconstruction error at each location over all training images obtained by leave-one-out cross validation.

Optimization The maximum likelihood estimation problem in a distributed setting is solved using dual decomposition based on the algorithm presented in [15] and also described here briefly for completeness. Let's assume that k blocks (partitions) are extracted from an image and that the k blocks are coupled through n_c consistency constraints that require the image intensities in overlapping voxels to be equal. The variables that are constraint to be equal across different blocks are denoted as *public* variables. The variables that are local to each block and are not common in other blocks are denoted as *private* variables. Let's assume that s_i and y_i are the unknown private and public variables (image intensities) of block i, respectively. If we concatenate s_i and y_i, we get the vector x_i, indicating all variables (private and public) in block i. For each block a local (strictly) concave log-likelihood function, $l_i(x_i)$ or $l_i(s_i, y_i)$, is maximized. The public variables for all blocks are collected together into one vector $y = (y_1, ..., y_k) \in \Re^p$, where $p = p_1 + ...p_k$, is the total number of public variables. A vector $z \in \Re^{n_c}$ is introduced to give the common values

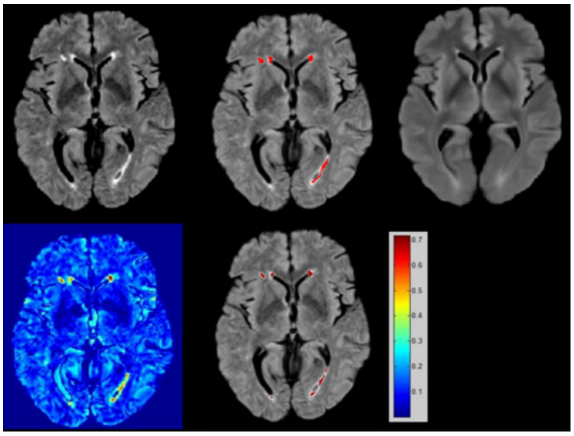


Figure 4.4: Segmentation of white matter lesions on a diabetes patient. *Top row from left to right*: FLAIR image, expert-defined lesion mask (in red) overlaid on the FLAIR image, and reconstructed image without abnormalities by the proposed method. *Bottom row*: calculated abnormality score map in color scale, segmentation mask (in dark red) after thresholding the abnormality score map.

of the public variables in each consistency constraint. The constraints are expressed as $y = Ez$, where $E \in \Re p \times n_c$ specifies the set of coupling constraints for the given block interaction,

$$E_{ij} = \begin{cases} 1 \text{ if } (y)_i \text{ is in constraint} j \\ 0 \text{ otherwise} \end{cases} \tag{4.9}$$

Lagrange multipliers, $v \in Re^p$, are introduced for the coupling constraints and a projected subgradient method is used to solve the dual master problem. Using these dual variables, optimization is independently performed in each block, and later on, the net variables are updated using the optimal values of the public variables of the blocks adjacent to that net. The dual variables are then updated, in a way that brings the local copies of public variables into consistency. The algorithm is summarized in Fig.4.3. A measure of the inconsistency of the current values of the public variables (consistency constraint residual) is given by the norm of the vector computed in the last step, $|E\hat{z} - y^*|$.

The method is applied for automatically segmenting brain pathologies, such as simulated brain infarction and dysplasia, as well as real lesions in diabetes patients. The data preprocessing steps include image smoothing, skull-stripping to extract the brain region [18], inhomogeneity correction [19], intensity normalization based on histogram matching, and deformable registration to a common space (template image). An example of lesion segmentation in an elderly individual with diabetes is shown in Fig.5.5.

Learning using Manifold Regularization

While tissue characterization has mainly been employed for separation of classes (healthy versus diseased), it is expected that it may also help in quantifying the deviation of tissue from *healthiness*

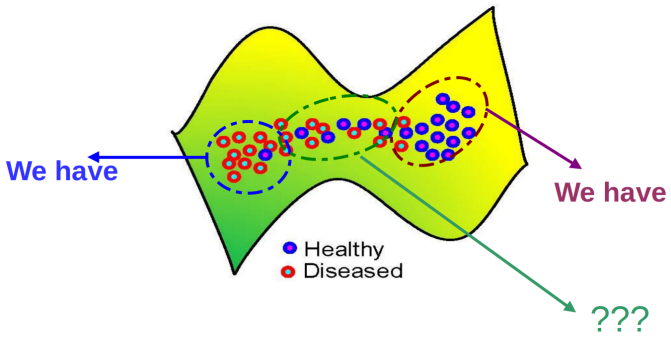


Figure 4.5: Distribution of samples that vary on a smooth manifold encompassing the normal cluster at one end to the abnormal cluster at the other.

and provide the degree of abnormality. In [16], we consider tissue abnormality characterization as a regression problem. We assume that there is a smooth manifold encompassing normal and lesion tissue with normal appearing abnormal tissue lying on this manifold. Furthermore, abnormality score should be continuous and smooth on manifold and spatially on the image. However, there are some issues to be addressed: first, conventional regression methods like Support Vector Regression (SVR) does not provide leverage to control the smoothness; second, abnormality in the absence of consistent ground truth may be easier to be characterized using combination of MR protocols rather than a single one. We have some samples from healthy brain and some samples from lesion part of diseased brain as labeled training samples. Voxels of the brain which are to be tested are considered as unlabeled samples. Taking advantage of Laplacian Regularized Least Square (LapRLS)[17] formulation as a semi-supervised regression method, we associate a continuous abnormality score pertaining to each voxel of the brain implemented as an embedding graph consisting of labeled and unlabeled voxels. Training samples and unlabeled voxels set up vertices of an embedding graph. Associations between neighborhood voxels are taken care of by using a proper edge weighting scheme between vertices of embedding graph. Since a smoothness constraint is imposed on the cost function of regression, the result of such a functional optimization could be treated as a qualifier of tissue which provides the abnormality characterization. Employing LapRLS, we propose a method which can handle both the criteria (spatial and manifold smoothness) in one framework, so that a continuous abnormality score is obtained. The framework is applied to multi-parametric data acquired on MS patients with the idea of characterizing not only the lesions diseased or healthy tissue, but also the WM that is progressing to abnormality based on the stage of the disease.

Laplacian Regularized Least Square Given a set of labeled example (x_i, y_i), $i = 1, ..., l$, (in our case, x_i's are voxels with multi-parametric intensities that have been labeled (y_i) as diseased or healthy), our aim is to find an (abnormality) function f, which satisfies the following condition:

$$f^* = \text{argmin}_{f \in H_k} \frac{1}{l} \sum_{i=1}^{l} C(x_i, y_i, f) + \lambda_R \|f\|_K^2 + \lambda_M \|f\|_M^2 \tag{4.10}$$

where H_k is a Reproducing Kernel Hilbert Space of functions. The first term, $C(x_i, y_i, f)$, penalizes the error for labeled samples, y_i, which, in the case of regression, could be the square loss function $C(x_i, y_i, f) = (y_i - f(x_i))^2$. The second term, $\lambda_R \|f\|_K^2$, and third term, $\lambda_M \|f\|_M^2$, together impose different smoothness conditions on the abnormality function. First one imposes smoothness such that normal and lesion samples would not be mixed together. The latter one takes care of smoothness of abnormality score between labeled and unlabeled samples. In accordance, the weighting factors λ_R and λ_M control the complexity of the function in the ambient space and intrinsic geometry, respectively. f^* is the abnormality score derived from minimization of Eq.4.10 in which $f^* > 0$ indicates abnormal and $f^* < 0$ indicates normal. With some reasonable mathematical assumptions described in [17] the last term, $\|f\|_M^2$, can be approximated by the graph Laplacian which is constructed based on labeled and unlabeled samples. The optimized function becomes:

$$f^* = \text{argmin}_{f \in H_k} \frac{1}{l} \sum_{i=1}^{l} C(x_i, y_i, f) + \lambda_R \|f\|_K^2 + \frac{\lambda_M}{(u+l)^2} \hat{f}^T L \hat{f} \tag{4.11}$$

where $\hat{f}$ is a vector containing outcome (class label) for labeled and unlabeled samples. Since we do not have outcome for unlabeled samples, corresponding elements in $\hat{f}$ will be zeros. The matrix $L = D - W$ is the graph Laplacian matrix, W is a matrix containing edge weights of the embedding graph and D is a diagonal matrix with elements equal to the sum of columns of W. As it is shown in [17], the decision function can be expressed in the form of $f(x) = \sum_{i=1}^{u+l} \alpha_i K(x_i, x)$. In this study, we have used the RBF kernel.

As the square loss function, $(y_i - f(x_i))^2$, has been used, the resulting optimization problem corresponds to the Laplacian Regularized Least squared (LapRLS) which is a type of regression problem. In that case, the optimal solution for α_i can be derived from the following linear system:

$$\alpha^* = (JK + \lambda_R l I + \frac{\lambda_M l}{(u+l)^2} LK)^{-1} Y \tag{4.12}$$

where Y is a vector containing the labels for training samples (+1,-1,0 for lesion, healthy and unlabeled samples respectively). $J = diag(y_1, y_2, ..., y_l, 0, 0, ..., 0)$ is a diagonal matrix holding labels for labeled samples and zero for unlabeled samples on diagonal elements and I is an identity matrix.

Moreover, in [16] prior knowledge was incorporated in terms of weights between samples. Specifically, edge weights were assigned using an additive weighting scheme imposing spatial and feature space smoothness.

4.2 Fuzzy clustering or random walks

Modified fuzzy c-means with spatial priors

In [18], we present a semi-supervised segmentation methodology that detects and classifies cerebrovascular disease in multi-channel magnetic resonance (MR) images. The method combines intensity based fuzzy c-means (FCM) segmentation with spatial probability maps calculated from

a normative set of images from healthy individuals. Unlike common FCM-based methods which segment only healthy tissue, we have extended the fuzzy segmentation to include patient-specific spatial priors for both pathological conditions (lesions and infarcts). These priors are calculated by estimating the statistical voxel-wise variation of the healthy anatomy, and identifying abnormalities as deviations from normality. False detections are reduced by knowledge-based rules.

The FCM algorithm [19] is based on minimizing an objective function with respect to a fuzzy membership U and set of cluster centroids V. Let $x_j \in \Re^n$ be the feature representation (e.g. multi-channel intensities with $n = 2$) of voxel $j \in \Omega$. FCM clusters the data by computing the fuzzy membership $u_{ij} \subseteq U$ at each voxel j to the i-th class, such that $u_{ij} \in [0, 1]$ and $\sum_{i=1}^{C} u_{ij} = 1, \forall j \in \Omega$. The number of classes C is assumed to be known. In the conventional approach the objective function expresses the degree of intensity based dissimilarity between the data values x_j and the cluster centroids, $v_i \subseteq V$. In this modified approach the objective function J_m includes an additional term reflecting the distance between the fuzzy membership u_{ij} and prior tissue probability, p_{ij} at each voxel j, which is calculated as described in the previous section. Thus J_m is expressed as follows:

$$J_m(U, V) = \sum_{j \in \Omega} \sum_{i=1}^{C} \left[u_{ij}^m d_v^2(x_j, v_i) + w_i (u_{ij} - p_{ij})^2 \right] \tag{4.13}$$

where d_v is a distance function, $m \in (1, \infty)$ is a weighting exponent on each fuzzy membership controlling the degree of fuzziness and w_i, $i = 1, ..., C$ is a normalization weight controlling the contribution of each distance term per cluster. The use of different weights w_i allows controlling the penalty on dissimilarity for each class. For unbiased results w_i can take the default value of 1, as performed in this study. For the two terms in the objective function to be comparable (intensity dissimilarity and spatial inconsistency) the intensity values x are scaled in the range $[0, 1]$, similarly to the probability range. The constrained optimization of J_m is expressed using Lagrange multipliers and u_{ij} and v_i are determined by setting the derivative of J_m to zero. For $m = 2$ (as usually chosen in image segmentation), the fuzzy membership and the cluster center are calculated as follows:

$$v_i = \frac{\sum_j u_{ij}^2 x_j}{\sum_j u_{ij}^2} \tag{4.14}$$

$$u_{ij} = w_i p_{ij} q_{ij} - \frac{\sum_i w_i p_{ij} q_{ij} - 1}{\sum_i q_{ij}} q_{ij} \tag{4.15}$$

where $q_{ij} = \frac{1}{d_v^2(x_j, v_i) + w_i}$. The algorithm iteratively optimizes the objective function in Eq. 4.13 with the continuous update of u_{ij} and v_i, until convergence is reached. The prior tissue probability maps are utilized as the membership at first iteration. In regular FCM, d_v is the Euclidean distance. In order not to restrict the detection to data classes with hyper-spherical shape, we used the Mahalanobis distance.

Construction of patient-specific spatial priors The methodology for calculating prior tissue probability at each voxel is illustrated in Fig.4.6 and can be separated into two branches, the construction of spatial priors for normal classes and abnormal classes. The spatial priors for abnormal

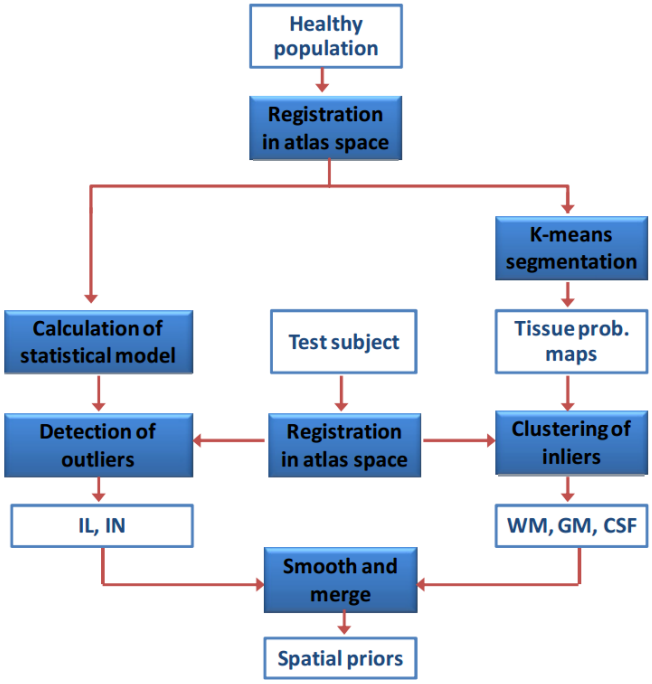


Figure 4.6: Methodology of the calculation of spatial priors for normal tissue (WM, GM and CSF), and abnormal tissue (IL and IN).

classes (hyperintense infarct lesions (IL) and infarct necrosis (IN)) are constructed following an outlier detection approach. Assuming that voxel intensities are statistically independent and follow a multivariate normal distribution we build a statistical atlas from a training set of healthy subjects. The healthy tissue priors are calculated by segmenting the patient's image into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) using the k-means algorithm. Since abnormalities are present, the algorithm is not randomly initialized because abnormal tissue might influence the cluster center estimation of healthy tissue. The estimation is thus guided by the healthy training data. Specifically, the healthy images are segmented by k-means, one by one, updating each time the estimate of the clusters centers location. The average (over all subjects) center location for WM, GM and CSF clusters is finally used as initial estimate in the segmentation process of the patient's image. Therefore, the effect of the presence of abnormalities is reduced. More details on the construction of the patient-specific spatial priors for normal classes and abnormal classes can be found in [18].

Graph-based image segmentation

Based on similar ideas, in [20][21] we exploit conventional MR modalities in order to segment brain images with neoplasms using a *nearly unsupervised* learning strategy. Healthy tissue clustering, outlier detection and geometric and spatial constraints are applied to perform a first segmentation using global features and then the segmentation boundaries are locally refined by a

modified 3D Random Walker segmentation algorithm [22]. In our preliminary work [20], we integrated graph-based segmentation with intensity modeling in order to characterize brain tissue in 2D MR images, whereas in [21] we expand the analysis into 3D and improve the methodology to achieve more accurate segmentation, applicable to enhancing masses. In [21] the method segments normal and abnormal tissues, including WM, GM, CSF, non-necrotic enhancing neoplastic tissue (ET) and non-enhancing tissue (non-enhancing neoplastic, edema and necrosis) (EDM/NET), in a multiclass tissue characterization problem.

We classify the method as *nearly unsupervised* because it only requires to depict from the test patient some image slices including only healthy tissue. There is no requirement for precise tissue annotation on the voxel level. These image slices are used to learn the (patient-specific) intensity distribution of healthy tissue. The FCM algorithm was applied to differentiate between healthy tissue types based on the T1 and T2 sequences. The Mahalanobis distance (MD) to the distribution of T1ce and FLAIR intensities was used to express the probability of each voxel to be an outlier (neoplastic tissue) in respect to each healthy tissue class $l_i \in \{WM, GM, CSF\}$. Abnormalities were detected by sampling the overlap of high confidence regions of the three MD maps, i.e. regions for which $MD_{l_i}^2 \geq thr$, where thr is a density threshold above which a region is considered abnormal. In [21] the threshold value thr was determined by selecting the 35% highest probability. Based on experimental results the segmentation process was not too sensitive to the choice of this value.

Furthermore, some intensity constraints were enforced followed by connected-component analysis to reduce false positives. Enhancing neoplastic tissue is identified as the cluster (based on FCM clustering) with the highest average T1ce value while the remaining voxels are labeled as EDM/NET. Finally, all image voxels corresponding to the abnormal and healthy tissue clusters are used as pre-labeled seeds for the modified 3D Random Walker. The whole framework is illustrated in Fig.4.7.

Random Walker algorithm The Random Walker algorithm is a graph-based image segmentation technique, used in combination with user-defined seeds. Assuming that the image is a given graph $G = (V, E)$, each voxel in the image represents a vertex $v \in V$, which is connected with neighboring voxels through an edge $e \in E$, representing the relationship of these voxels. In a weighted graph, a weight $r_{jk} \in \Re$ is assigned between the nodes j, $k \in V$, based on imaging features. The random walk method consists of computing the probability, that a "random walker", starting at an arbitrary node, first reaches a pre-assigned node with a particular label [22]. In image segmentation problems, given a subset of pre-defined marked vertices V_L, a graph-theoretic Random Walker optimization labeling problem is defined and solved by

$$x_{opt} = \underset{x}{\operatorname{argmin}} \left(\frac{1}{2} \sum_{e_{jk} \in E} r_{jk} (x_j^{l_i} - x_k^{l_i})^2 \right) \tag{4.16}$$

where $x_v^{l_i}$ indicates the probability of vertex v to belong to a specific class l_i, ($i \in [1, C]$), where C is the number of classes (here $C = 6$ including the background). Consequently, the solution x_{opt} is a $V \times C$ matrix containing the class probabilities of all vertices. Eq.4.16 is solved through

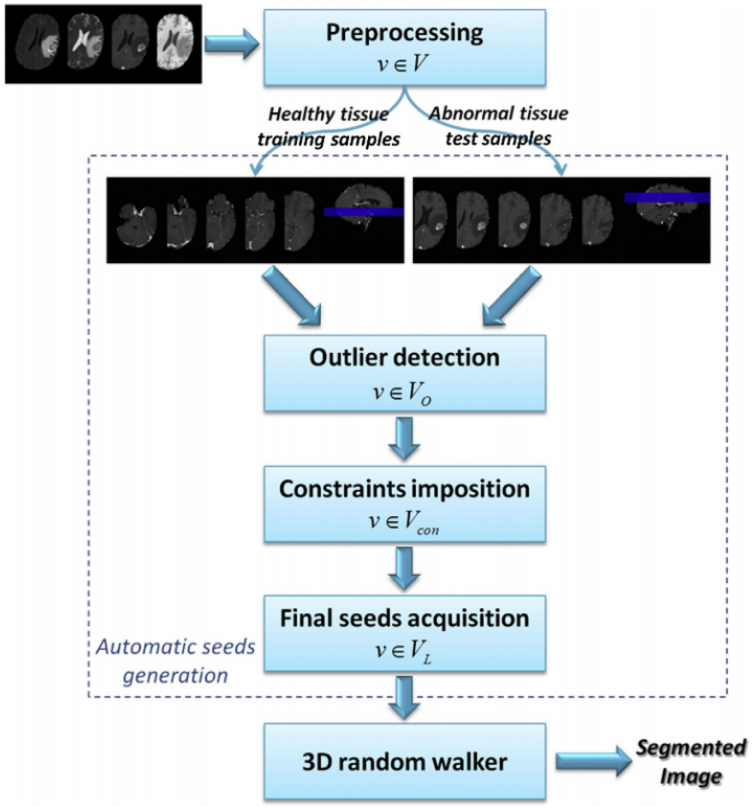


Figure 4.7: The framework for neoplastic and healthy brain tissue segmentation in [21].

a sparse linear system. In order to acquire the final undirected labeled (i.e. segmented) graph, the label with the highest probability is assigned to each vertex.

In [21], we adapt the original 2D Random Walker algorithm to 3D MR images in order to capture also relations across axial slices in brain structure. A 6-neighbor connectivity was used to handle the volumetric datasets. A Gaussian weighting function was utilized to connect the feature vectors h_j and h_k of voxels voxel j and voxel k, defined as $r_{jk} = e^{-\lambda|h_j - h_k|^2}$ where λ is a constant penalizing dissimilarity (the only free parameter in the algorithm). Low edge weights indicate high probabilities of region margin evidence between two neighboring voxels thus prevent a Random Walker crossing these boundaries. In our analysis, the intensities of T1ce, T1 and FLAIR sequences were used as 3-dimensional feature vectors. The low computational cost, the automated nature of the proposed method and the requirement of only routine MRI (advanced imaging techniques have inherently low signal-to-noise ratio and also are not performed in clinical routine) can be considered as the main strengths of the method. The main limitations of the method are that it is not appropriate for non-enhancing tumors and also necrosis is included in the same class with the non-enhancing tissue.

4.3 Spectral clustering of waveforms

Analysis of biosignals, such as EEG, has been widely performed as a diagnostic method in sleep studies or studies of disorders since it provides the means for the identification of conspicuous and quite repeated individual brain waves and rhythms. Due to the extreme size (duration) of this data, visual recognition of the sought waveforms is almost prohibitive in a routine setting, thus the necessity of an automatic detection method becomes apparent. Apart from labor-intensive, visual marking is also highly scorer dependent mainly due to high intra- and inter-subject variability and variation in human perception. In [23][24] we propose a two-step methodology for detection of transient waveforms, such as K-complexes, in sleep EEG. The method first identifies all possible candidates according to pre-defined expert-based rules using multi-channel information and then reduces false detections by applying semi-supervised classification. The novelty of the method mainly relies on the classification step which (i) exploits both the time-varying signal and the spectral content of it, (ii) applies a novel outlier detection methodology based on graph partitioning by spectral clustering, and (iii) uses time-frequency (TF) representations that describe the spatiotemporal characteristics of the waveform via their time-varying spectral content.

The detection of candidate waves using empirical rules and fundamental features is described in details in [23]. Since these knowledge-based rules define only basic patterns, one-class classification is performed to distinguish between subtle morphological characteristics. First each extracted EEG segment is represented by two temporal patterns: (i) the amplitude change over time (signal representation) and (ii) the frequency content over time, which is calculated as the power spectral density (TF representation) integrated in the frequency range of the target waveform (e.g. $0.5 - 5$Hz for K-complexes). Then we learn the distribution of these features vectors from a set of annotated samples and calculate the probability of the test sample to be an inlier in the distribution.

Training phase: Since different patterns of morphology exist for each transient type, the distribution of the training vectors in space is not expected to be around a unique center and neither have a known shape (that for example could be modeled by a Gaussian distribution). We only assume that enough examples exist from each pattern of the same class and that these examples lie closer to each other in space than examples from different classes. In order to detect these patterns, we apply a graph partitioning technique, known as spectral clustering [25], to partition the training set into a set of clusters. Spectral clustering divides graph nodes into groups so that connectivity is maximized between nodes in the same cluster and the connectivity is minimized between nodes in different clusters. Connectivity is measured by some affinity (similarity) measure. We define as affinity measure between sample i and j, the quantity

$$A_{ij} = \exp(-\frac{D_{ij}^2}{2\sigma^2}) \tag{4.17}$$

where D_{ij} is the distance of sample i and j (we use the Euclidean distance) and σ some normalization constant. Examples of the cluster averages are illustrated in Fig.4.8. For each training cluster c, $c = 1, ..., N_c$, where N_c is the number of clusters, the cumulative histogram of all pairwise affinities is calculated, $p_c(A)$, as a probability measure of a sample with affinity value A to be an inlier of the distribution. Linear interpolation is used to calculate the probability at intermediate

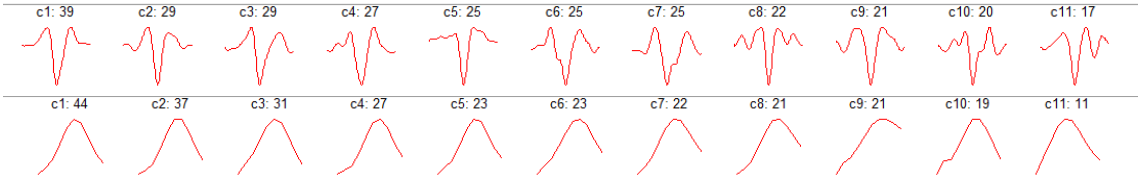


Figure 4.8: Training patterns for K-complexes calculated by spectral clustering. The amplitude and power of signal over time are averaged over all samples per cluster and illustrated in 1st and 2nd row, respectively. The titles show the number of samples per cluster ($N_c = 11$).

values (not present in the training set). The higher p_c, the smaller the significance (p-value) and thus the possibility of the sample to be an outlier (false positive detection).

Testing phase: For each test vector i, we calculate its distance to each training cluster c as the distance to the closest sample in cluster c and calculate the affinity A_{ic} according to Eq.4.18:

$$A_{ic} = \exp(-\frac{(\text{argmin}_{j \in \ cluster\ c}\, D_{ij})^2}{2\sigma^2}) \tag{4.18}$$

The probability $p_{ic}(A_{ic})$ of the sample to belong to cluster c is then calculated based on the *pdf* estimated in the training phase, whereas the total probability to be an inlier (e.g. a true K-complex) is calculated as the maximum probability across all clusters:

$$p_i = \underset{c}{\text{argmax}}\, p_{ic}(A_{ic}) \tag{4.19}$$

The estimation of the *pdf* is performed independently for each set of feature vectors, x^t and x^f, representing signal and frequency content, respectively. The two probabilities, let's denote them as p_i^t and p_i^f, for sample i are fused and thresholded to reach the final decision. Fusion is performed by Fisher's method [26] which combines p-values from several independent tests into one test statistic that has a chi-squared distribution:

$$p_i = F_{chi2}(-2(ln(1-p_i^t) + ln(1-p_i^f)) \tag{4.20}$$

where p_i is the total probability combing signal and frequency content and F_{chi2} is the cumulative distribution function evaluated at the test statistic.

4.4 Pattern similarity in biognals and molecular data

Similarity search in biosignals

In [27] we study the general problem of similarity search in databases of time series, such as biosignals, and we propose a novel multiresolution indexing (i.e., representation) and retrieval method for time series similarity search. Our approach is motivated by the idea that if we examine a time series at different resolution levels, we could possibly acquire further insights about the data. The proposed algorithm adopts a combined, two-step pruning (filtering) strategy to further reduce

data dimensionality by discarding irrelevant time series (i.e., false alarms). At a first level, the time series are represented by line segments and filtered by the triangular inequality property. Then, a Vector Quantization like scheme is applied to encode data and thus to reduce dimensionality.

MultiResolution Piecewise Vector Quantization (MR-PVQ) Our MR-PVQ method [27], extends the Piecewise Vector Quantized Approximation (PVQA) dimensionality reduction technique presented previously [28], in multiple resolutions. The lower the resolution level is, less number of segments are used to encode time series data. To achieve this, we propose to use a two-level pruning (filtering) strategy in order to decrease the number of objects that will be encoded during the next step. In each level of this strategy, we use an approximation function for the time series. The first filter applies a property that all the indexing schemes require to hold, triangular inequality discarding the non-qualifying objects. Here, the time series data are approximated with first degree polynomials. The second filtering level is based on the lower bounding lemma of the Generic Multimedia Indexing (GEMINI) algorithm [29] and a vector quantization technique [30] is used to encode the testing set. Our work is motivated by the observation that although global information of a time series is kept after the encoding by PVQA in one resolution, important local information of the time series is lost. The idea of using multiple resolution levels, gives us the opportunity to retain both global and local information. Combining this fact with the representation of the time series using polynomials and the application of the previously described filters that enhance the pruning power of the algorithm, can improve substantially the performance of similarity search in time series databases. The outline of the MR-PVQ method is described as follows.

For each resolution level i:

1. Split the time series into l segments.
2. Represent each segment with a first degree polynomial.
3. Discard the time series that are not close to the query using the above representation (first level pruning).
4. Encode the remaining time series using PVQA.
5. Discard the time series that correspond to false alarms (second level pruning).
6. Move on to the next resolution level and repeat steps 1-6.

We tested and demonstrated the performance of the proposed method, analyzing EEG time series data for retrieval of one of the constituent brain waveforms in EEG recordings, the K-complex, but the method can as well be applied for retrieval of other patterns of interest in time series analysis. The automatic detection and categorization of the EEG patterns can facilitate the correlation analysis of large amounts of data, and help towards the differential diagnosis of epilepsy or related disorders, as well as treatment evaluation.

My personal contribution in this work [27] related mostly to the data representation, clinical application and overall overview.

Biological networks

It has been observed that similar gene expression patterns among patients with regard to known gene markers cannot guarantee similar phenotype (i.e. disease outcome). Late studies implicated that alterations in gene expression might perturb the higher-level organization of the interactome, affecting so the disease outcome [31]. To investigate this hypothesis, we explored how the temporal dynamics of transcriptional behavior in a specific treatment scheme reforms the protein interactome. Our target was to reveal how interactome 'areas', in the form of modules/sub-networks, are perturbed in response to drug over time. To achieve this goal, we integrated gene expression and protein-protein interaction (PPI) data, a strategy recently established as fruitful in providing information of specific genes/proteins on disease-specific pathophysiology. Towards this orientation, many studies have combined multiple data types [32][33] or scored pathways based on the similarity of the expression values of the participating pathway genes [34]. For example, interesting studies like [35] detected sub-networks of highly co-expressed genes on the protein graph by starting from a random gene with the use of a greedy algorithm, which cannot guarantee completeness. Other studies like [36] integrated gene expression, PPI and phenotype data to identify dense modules with the provision of incorporating additional constraints from a variety of datasets. However, this approach is primarily designed for finding protein complexes from protein interaction data, is sensitive to gene expression noise and promotes the detection of dense modules.

In this work [37], we have illustrated the efficacy of our integrative methodology [38] in capturing the dynamic modular transitions in response to tamoxifen. To achieve this goal, we reinforced the protein graph structure, via weighting scheme, with time series microarray data descending from an in vivo study [39]. Next, the Detect Module from Seed Protein (DMSP) algorithm defined modules on the composite protein network starting from specific 'seed' proteins. An important feature of this algorithm is that the overlaid gene expression information, in the form of weight, reassures the entrance of certain interactions into the modules, even if they are not favored by the topology. Also, DMSP saves many interactions among proteins that interact closely (e.g. complexes) even if they show dissimilar or inverse expression trends, through the rest weighted neighbors of such an interaction. Our time-evolving modules report that the response to tamoxifen is a highly dynamic process and raise several biological questions regarding the recruitment of several known pathways. Finally, our findings corroborate towards the integration of heterogeneous data and the detection of discriminative temporal sub-networks that serve as hallmarks of disease-specific states.

My personal contribution in the analysis of biological networks [37] related to the processing of the datasets. Specifically, the datasets (estradiol and estradiol plus tamoxifen treatment) were normalized after background correction with loess normalization approach with the use of limma package in Bioconductor [40]. The expression value of each gene was computed by taking the average of the corresponding probe sets and all values were normalized with respect to the first day. A set of proteins ($n = 883$) related to breast cancer were selected from G2SBC and dbDEPC databases and mapped to our final gene list. This subset defined the 'seed' list that was used as input to the DMSP algorithm. The protein interaction network was weighted with the gene expression values from all time points. The initial step of the weighting scheme includes clustering of the temporal expression profiles of both datasets. Clustering was performed by the K-means

algorithm which was able to process fast and transparently the datasets.

Co-clustering of gene expression and function

About 40% of the proteins encoded in eukaryotic genomes are proteins of unknown function. Thus an important bioinformatics problem is to associate gene expression with gene function information, which will allow to reveal the function of unannotated genes [41]. The methodologies for the analysis of Gene Expression Maps (GEMs) involve the application of feature extraction techniques combined with data mining methods such as clustering, classification and similarity search. Furthermore, gene information from other sources, such as Gene Ontology, is usually employed to validate biological hypothesis or to strengthen the fidelity of research outcomes. In order to explore gene function and gene expressions differences with regard to brain regions, in [42][43] we propose an anatomy-oriented framework for the analysis of GEMs obtained by voxelation. The voxelation technique allows acquisition of expression images in parallel, simplifying cross-analysis of multiple genes and also is less expensive and faster than traditional approaches. Voxelation data however have much lower resolution (e.g. $1mm^3$) than single cell resolution data, and thus suffer from partial volume effect in which the acquired expression values represent an average over the gene expression of all cells in each voxel. Firstly, we examine if the down-weighting of inconsistent measurements, such as in voxels with high partial volume effects helps generate more informative clusters relevant to function categories. Afterwards, we identify clusters containing genes whose expressions display similar anatomical distribution in respect to specific brain regions such as white matter, gray matter and the hippocampal region. We then investigate the hypothesis whether gene clusters with similar expression patterns have also similar gene function. A summarized illustration of the analysis steps is shown in Fig.4.9.

Definition of gene expression and gene function similarity The gene expression maps similarity between two genes is defined as the squared weighted Euclidean distance function. The weight vector is used to emphasize dissimilarity on selective spatial locations. The first case scenario allows investigating whether by down-weighting the measurements on locations with high uncertainty a more informative similarity measure is formed that is not affected by partial volume (PV) and artifacts due to ventricles. This is tested by using the confidence map as weight vector and the method is then denoted as global approach with PV correction. The second case scenario allows investigating whether gene function correlates with gene expression in specific anatomic locations. Here we investigate whether genes with similar expression in some anatomic locations have similar gene functions. Similarity of expression in a group of genes is defined as the average similarity of all genes in the group to the group center.

The gene function similarity between two functions is calculated using Lin's method [44] to evaluate function distance in Gene Ontology structure. This method applies an information theoretic definition of similarity as long as there is a probabilistic model. The similarity values are publicly available and obtained within each of the three categories of Gene Ontology that refer to *Cellular Component*, *Molecular Function* and *Biological Process*. The values are based on frequencies from the Mouse Genome Informatics annotation dataset. The function similarity in a group of genes is calculated as average pairwise similarity.

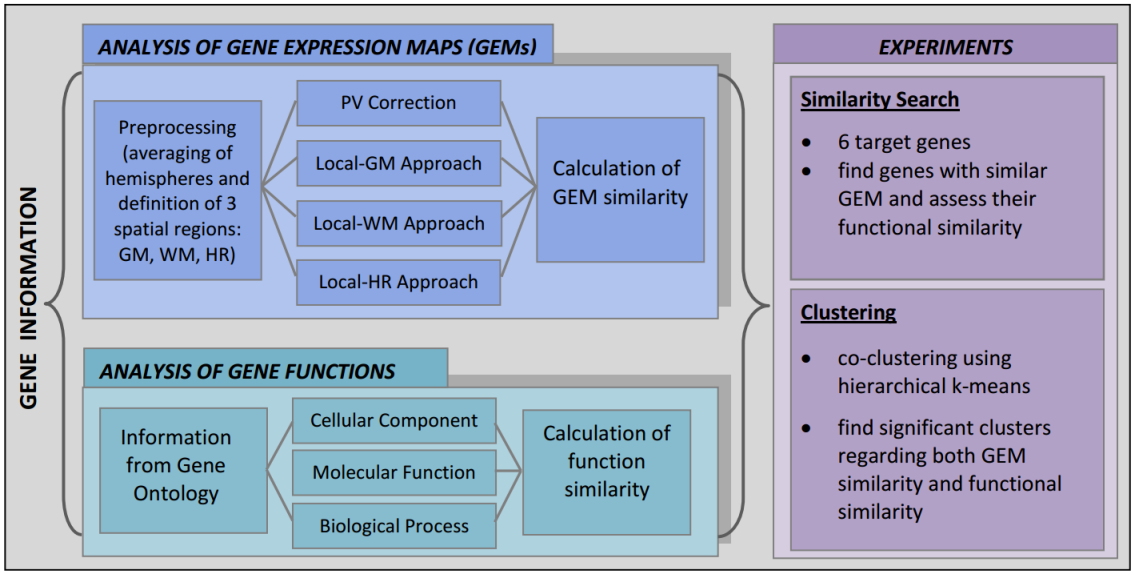


Figure 4.9: illustration of the analysis steps in [42]. When we use spatial maps to investigate local regions in the brain, the method is denoted as local-GM, local-WM and local-HR, respectively.

Clustering analysis is performed in order to classify the genes into clusters that have both similar GEMs and similar gene functions using tests of significance (*p-value*) and a hierarchical k-means algorithm [45]. Our investigation concluded that clusters of genes with similar localized expression patterns display function similarity. The results indicated that our work has the potential to create comprehensive atlases of gene expression in the mammalian brain and to provide insight into the identification of unannotated genes based on the analysis of their GEMs.

Bibliography

[1] G. Erus, E. I. Zacharaki, N. Bryan, and C. Davatzikos, "Learning high-dimensional image statistics for abnormality detection on medical images", in *Computer Vision and Pattern Recognition Workshops (CVPR), 2010 IEEE Computer Society Conference on*, IEEE, 2010, pp. 139–145.

[2] G. Erus, E. I. Zacharaki, and C. Davatzikos, "Individualized statistical learning from medical image databases: Application to identification of brain lesions", *Medical image analysis*, vol. 18, no. 3, pp. 542–554, 2014.

[3] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[5] S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, R. Whitaker, A. D. N. Initiative, *et al.*, "Manifold modeling for brain population analysis", *Medical image analysis*, vol. 14, no. 5, pp. 643–653, 2010.

[6] J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos, "Gram: A framework for geodesic registration on anatomical manifolds", *Medical image analysis*, vol. 14, no. 5, pp. 633–642, 2010.

[7] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular pca approach", *Pattern Recognition Letters*, vol. 25, no. 4, pp. 429–436, 2004.

[8] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: A new approach to appearance-based face representation and recognition", *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[9] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum, "Multiscale 3-d shape representation and segmentation using spherical wavelets", *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 598–618, 2007.

[10] J. Canny, "A computational approach to edge detection", *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[11] S. Mallat, *A wavelet tour of signal processing: The sparse way*. Academic press, 2008.

[12] D. Shen and C. Davatzikos, "Hammer: Hierarchical attribute matching mechanism for elastic registration", *IEEE transactions on medical imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.

[13] E. I. Zacharaki and A. Bezerianos, "Segmentation of pathology by statistical modeling and distributed estimation", in *Biomedical Engineering, 2011 10th International Workshop on*, IEEE, 2011, pp. 1–3.

[14] ——, "Abnormality segmentation in brain images via distributed estimation", *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 3, pp. 330–338, 2012.

[15] S. Samar, S. Boyd, and D. Gorinevsky, "Distributed estimation via dual decomposition", in *Control Conference (ECC), 2007 European*, IEEE, 2007, pp. 1511–1516.

[16] K. Batmanghelich, X. Wu, E. Zacharaki, C. E. Markowitz, C. Davatzikos, and R. Verma, "Multiparametric tissue abnormality characterization using manifold regularization", 16, International Society for Optics and Photonics, vol. 6915, 2008, pp. 1–6. DOI: 10.1117/12.770837. [Online]. Available: http://dx.doi.org/10.1117/12.770837.

[17] M. Belkin, P. Niyogi, and V. Sindhwani, "On manifold regularization.", in *AISTATS*, Citeseer, 2005.

[18] E. I. Zacharaki, G. Erus, A. Bezerianos, and C. Davatzikos, "Fuzzy multi-channel clustering with individualized spatial priors for segmenting brain lesions and infarcts", in *Artificial Intelligence Applications and Innovations*, Springer, 2012, pp. 76–85.

[19] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[20] V. G. Kanas, E. I. Zacharaki, E. Dermatas, A. Bezerianos, K. Sgarbas, and C. Davatzikos, "Combining outlier detection with random walker for automatic brain tumor segmentation", in *Artificial Intelligence Applications and Innovations*, Springer, 2012, pp. 26–35.

[21] V. G. Kanas, E. I. Zacharaki, C. Davatzikos, K. N. Sgarbas, and V. Megalooikonomou, "A low cost approach for brain tumor segmentation based on intensity modeling and 3d random walker", *Biomedical Signal Processing and Control*, vol. 22, pp. 19–30, 2015.

[22] L. Grady, "Random walks for image segmentation", *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[23] E. I. Zacharaki, E. Pippa, A. Koupparis, V. Kokkinos, G. K. Kostopoulos, and V. Megalooikonomou, "One-class classification of temporal eeg patterns for k-complex extraction", in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, IEEE, 2013, pp. 5801–5804.

[24] E. Zacharaki, E. Pippa, A. Koupparis, G. Kostopoulos, and V. Megalooikonomou, "Classification of eeg waveforms by spectral clustering", in *5th Panhellenic Conference on Biomedical Technology*, 2013, pp. 93–94.

[25] A. Y. Ng, M. I. Jordan, Y. Weiss, *et al.*, "On spectral clustering: Analysis and an algorithm", *Advances in neural information processing systems*, vol. 2, pp. 849–856, 2002.

[26] R. A. Fisher, *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 1925.

[27] A. Charisi, F. D. Malliaros, E. I. Zacharaki, and V. Megalooikonomou, "Multiresolution similarity search in time series data: An application to eeg signals", in *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, ACM, 2013, p. 33.

[28] Q. Wang and V. Megalooikonomou, "A dimensionality reduction technique for efficient time series similarity analysis", *Information systems*, vol. 33, no. 1, pp. 115–132, 2008.

[29] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, *Fast subsequence matching in time-series databases*, 2. ACM, 1994, vol. 23.

[30] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 2012, vol. 159.

[31] I. W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris, and J. L. Wrana, "Dynamic modularity in protein interaction networks predicts breast cancer outcome", *Nature biotechnology*, vol. 27, no. 2, pp. 199–204, 2009.

[32] I. A. Maraziotis, K. Dimitrakopoulou, and A. Bezerianos, "An in silico method for detecting overlapping functional modules from composite biological networks", *BMC Systems Biology*, vol. 2, no. 1, p. 93, 2008.

[33] I. Ulitsky and R. Shamir, "Identifying functional modules using expression profiles and confidence-scored protein interactions", *Bioinformatics*, vol. 25, no. 9, pp. 1158–1164, 2009.

[34] L. Tian, S. A. Greenberg, S. W. Kong, J. Altschuler, I. S. Kohane, and P. J. Park, "Discovering statistically significant pathways in expression profiling studies", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 38, pp. 13 544–13 549, 2005.

[35] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis", *Molecular systems biology*, vol. 3, no. 1, p. 140, 2007.

[36] E. Georgii, S. Dietmann, T. Uno, P. Pagel, and K. Tsuda, "Enumeration of condition-dependent dense modules in protein interaction networks", *Bioinformatics*, vol. 25, no. 7, pp. 933–940, 2009.

[37] K. Dimitrakopoulou, G. Dimitrakopoulos, E. I. Zacharaki, I. A. Maraziotis, K. Sgarbas, and A. Bezerianos, "Revealing the dynamic modularity of composite biological networks in breast cancer treatment", in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, IEEE, 2012, pp. 5432–5436.

[38] I. A. Maraziotis, K. Dimitrakopoulou, and A. Bezerianos, "Growing functional modules from a seed protein via integration of protein interaction and gene expression data", *Bmc Bioinformatics*, vol. 8, no. 1, p. 1, 2007.

[39] K. J. Taylor, A. H. Sims, L. Liang, D. Faratian, M. Muir, G. Walker, B. Kuske, J. M. Dixon, D. A. Cameron, D. J. Harrison, *et al.*, "Dynamic changes in gene expression in vivo predict prognosis of tamoxifen-treated patients with breast cancer", *Breast Cancer Research*, vol. 12, no. 3, p. 1, 2010.

[40] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, *et al.*, "Bioconductor: Open software development for computational biology and bioinformatics", *Genome biology*, vol. 5, no. 10, p. 1, 2004.

[41] K. Horan, C. Jang, J. Bailey-Serres, R. Mittler, C. Shelton, J. F. Harper, J.-K. Zhu, J. C. Cushman, M. Gollery, and T. Girke, "Annotating genes of known and unknown function by large-scale coexpression analysis", *Plant physiology*, vol. 147, no. 1, pp. 41–57, 2008.

[42] E. I. Zacharaki, A. Skoura, D. J. Smith, S. H. Faro, L. An, and V. Megalooikonomou, "Combining gene expression and function in a spatially localized approach", in *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*, IEEE, 2012, pp. 1–8.

[43] E. I. Zacharaki, A. Skoura, L. An, D. J. Smith, and V. Megalooikonomou, "Using an atlas-based approach in the analysis of gene expression maps obtained by voxelation", in *Artificial Intelligence Applications and Innovations*, Springer, 2012, pp. 566–575.

[44] D. Lin, "An information-theoretic definition of similarity.", in *ICML*, Citeseer, vol. 98, 1998, pp. 296–304.

[45] L. An, H. Xie, M. H. Chin, Z. Obradovic, D. J. Smith, and V. Megalooikonomou, "Analysis of multiplex gene expression maps obtained by voxelation", *BMC bioinformatics*, vol. 10, no. 4, p. 1, 2009.

Chapter 5: Supervised learning

In supervised learning a set of input and output variables are given and used to learn a function (often called model) that predicts the output based on the input features. When the response variables are quantitative the prediction task is called *regression*, whereas when they are qualitative (descriptive labels) the prediction task is called *classification*. Our work focuses mainly on pattern detection and classification problems and is divided according to the medical application into methods developed for the analysis of biosignals, medical images and molecular structures.

5.1 Pattern detection and classification in biosignals

Visual analysis of biosignals is rater-dependent and time consuming, especially for long-term recordings, while computerized methods can provide efficiency in reviewing long recordings, such as EEG. In this section we will present several supervised classification techniques that have been applied for detection of patterns in EEG, such as epileptic spikes and spindles, as well as pattern classification, e.g. epileptic or non-epileptic.

Support vector machines for spike detection In [1][2] a method is presented for spike detection incorporating pattern classification. Similarly to most pattern detection methods in signal processing, the amount of data processed is reduced by first extracting candidate waveforms based on low level detection analysis (by feature extraction), while subsequently classification is performed to maximize specificity of the overall method. Specifically, the proposed method first models coarsely the shape of the spike by breaking down the signal around major peaks into half-waves. Thresholding of shape characteristics extracted from the half-waves, such as amplitude and duration, is applied to generate a number of candidate spike locations. Subsequently, the method classifies the candidate spikes by embedding the data in a low dimensional space using the locality preserving projections (LPP) algorithm [3] and applying supervised classification in the embedding space.

The whole framework is illustrated in 5.1. Details on the application of LPP are provided in section 3.2. After embedding, the mapped data are introduced to a support vector machines (SVM) classifier [4]. SVM is an extremely popular algorithm that captures complex relationships between the data points and finds an optimal boundary between the class outputs. Although SVMs can be said to have started when statistical learning theory began with Vapnik and Chervonenkis (1974), the algorithm for optimal margin classifiers was introduced in [5]. The key point of the algorithm is that it finds a decision boundary that maximizes the (geometric) margin, based on

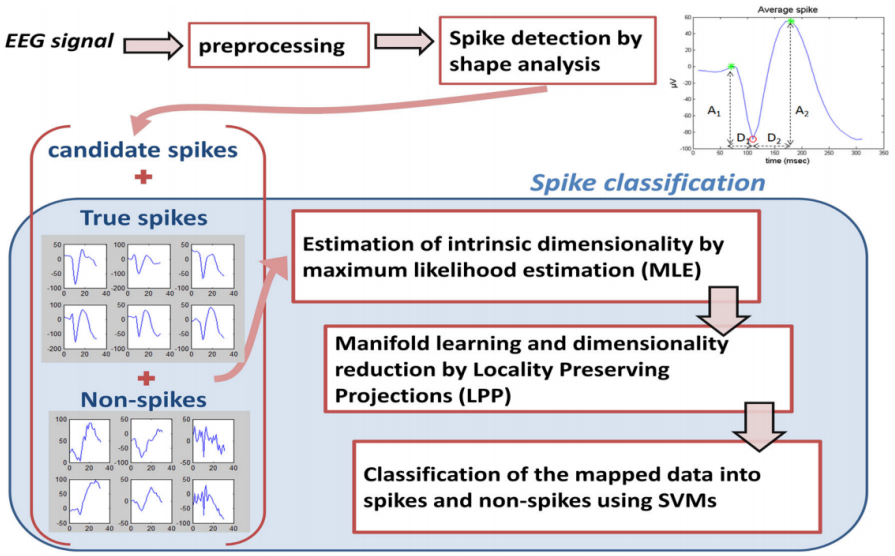


Figure 5.1: Spike detection framework. The 1st step of the method detects spike-like waveforms by extracting the two half-waves. The half-waves are defined between the negative peak (marked with a red circle) and the two positive peaks (marked with green stars) and are characterized by the amplitude difference (A_1, A_2) and duration (D_1, D_2). The 2nd step of the method classifies the detected waveforms into spikes and non-spikes by dimensionality reduction and supervised classification.

the idea that points near the decision surface represent very uncertain classification decisions. To make the algorithm less sensitive to outliers and appropriate for non-linearly separable datasets, the optimization problem is formulated such that it permits a few mistakes (samples inside or on the wrong side of the margin) by paying a misclassification cost. The second contribution of SVMs is that they extend to patterns that are non-linearly separable in the original feature space by mapping them to some higher-dimensional space (using a kernel function) where the training set is separable.

In our implementation we used a gaussian radial basis function as kernel to perform non-linear classification. The C and γ parameters, controlling the misclassification penalty and kernel size, respectively, were adjusted to account for unbalanced data. Since the sample size is rather small to produce balanced classes by subsampling the largest class, we used a weighted SVM and set the ratio of penalties for the two classes, C_1 and C_2, equal to the inverse ratio of the training class sizes. Thus we avoided bias toward the class with the largest training size. We defined γ to be adaptive to the dimensionality l, using the equation

$$\gamma = \frac{1}{(k \cdot l \cdot log(l))^2} \tag{5.1}$$

where k is a constant determined such that the fraction of the training samples contained in the kernel is approximately 20%.

The method [2] achieved high sensitivity with low false positive rate and outperformed the majority of the other approaches used for comparison. The contribution of the method is that (i) it is fully automated, i.e. no user interaction or manual intervention is required, (ii) it is template-free, thus it generalizes to any morphological patterns and shapes and can easily be applied for detection of other waveforms as long as some training patterns have been defined, (iii) it applies to all stages of sleep, therefore is appropriate for sleep monitoring, and (iv) it achieves high sensitivity with low false positive rate.

Hidden Markov Models and SVM for spindle detection Over the last decades sleep medicine is studying sleep for the purpose of sleep disorders treatment. The analysis of brain signals and the detection of specific patterns offers information related to sleep disorders. One such pattern is the sleep spindle. For the detection of sleep spindles we relied on the combination of discriminative and statistical models [6]. Specifically, the support vector machines and the hidden Markov models (HMMs) [7] were selected due to their advantageous performance in similar signal processing tasks. HMMs provide a conceptual toolkit for representing probability distributions over sequences of observations. They assume that (i) the observation at a given time was generated by some process whose state is *hidden* from the observer, (ii) the state of this hidden process satisfies the Markov property and (iii) the hidden state variable is discrete. The HMMs are useful for modeling time series data, such as biosignals, where an event can cause another event in the future, but not vice-versa.

The method for sleep spindle detection includes two stages. In the first stage the signal is pre-processed, parameterized and processed independently from the discriminative (SVMs) and the statistical (HMMs) models. In the second stage the output recognition results from each model are combined by a fusion method in order to provide the final sleep spindle detection results. For the HMM spindle models, we used a 3-state fully connected HMM model architecture, the states of which were modeled by a mixture of eight continuous gaussian distributions. HMM parameters were estimated using the Baum-Welch algorithm [8]. For the implementation we relied on the HTK software toolkit [9].

BayesNet and decision trees for seizure classification Correctly diagnosing generalized epileptic from non-epileptic episodes, such as psychogenic non epileptic seizures (PNES) and vasovagal or vasodepressor syncope (VVS), is rarely tackled in the literature despite its importance for the administration of appropriate treatment, life improvement of the patient, and cost reduction for patient and healthcare system. Usually clinicians differentiate between generalized epileptic seizures and PNES based on clinical features and video-EEG. In [10][11][12], we investigate the use of machine learning techniques for automatic detection and classification of generalized epileptic and non-epileptic events based only on multi-channel EEG data. Several classification algorithms are explored and evaluated on EEG epochs in an inter-subject cross-validation setting. The examined classification algorithms include BayesNet [13], Random Committee, Random Forest [14], IBk [15] and SMO [16] with RBF kernel, and were implemented by the WEKA machine learning toolkit [17]. The classifiers were selected in an attempt to evaluate representative algorithms for each one of the main categories of machine learning classification methods including probabilistic networks (BayesNet), decision trees (Random Forest), support vector machines (SMO), ensemble

classifiers (Random Committee and Random Forest), but also simple methods such as k-nearest neighbors (IBk). Due to large number of features feature ranking and selection is performed prior to classification using the ReliefF ranking algorithm [18] within two different voting strategies. The features introduced to the classifiers include signal characteristics in time and frequency domain. Features were combined across channels in order to characterize the spatio-temporal manifestation of seizures. More background on this clinical problem and details on representation of biosignals are provided in section 3.1.

5.2 Segmentation and classification in medical images

Image segmentation using Naive Bayes and SVM

The quantification and spatial localization of neoplastic tissue are of greatest importance for diagnosis, treatment planning and therapeutic monitoring. Brain lesion or tumor detection and tissue characterization is usually based on MRI which provides a great means for assessing the disease evolution and efficacy of therapy. Tissue classification generally requires information of several MR protocols and contrasts, as the axial 3D T1-weighted (T1) and T2-weighted (T2), Fluid Attenuated Inversion Recovery (FLAIR), axial 3D contrast enhanced T1 contrast-enhanced (T1ce) images. As an example, Fig.5.2 illustrates the multivariate image intensity of T1ce, T2 and FLAIR MR sequences of normal and abnormal tissue samples extracted from expert-defined regions of several patients. As shown, there is a distinction between healthy tissue and abnormalities, however not without overlap, making thresholding techniques not sufficient for accurate segmentation. There have been considerable efforts to develop automated or minimally interactive computer-based approaches for segmenting different tissue types in human brain using MRI data. Clustering algorithms, atlas-based methods, deformable models and (semi)supervised classification techniques are used in literature for brain tumor segmentation. A review of pattern recognition methods for (semi)automatic brain tumor segmentation based on human brain magnetic resonance images can be found in [19] and in [20]. Also in order to compare the different automated brain tumor segmentation methods the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* challenge was organized in conjunction with the international conference on Medical Image Computing and Computer Assisted Interventions (MICCAI) [21].

Our work on brain tumor segmentation is much earlier [23][24]. It incorporated high-dimensional intensity features created from multiple MRI acquisition protocols (structural MRI as well as DTI) into a pattern classification framework, to obtain a voxel-wise probabilistic spatial map. Moreover, guided by the follow-up scans, the likelihood of a region presenting tumor recurrence after treatment was determined. This study was one of the first to investigate integration of multiple MRI parameters via sophisticated nonlinear pattern classification methods to obtain a better characterization of the tumor and the surrounding tissue, as well as to investigate imaging profiles of tissue that are relatively more likely to present tumor recurrence in follow-up scans. We constructed two kinds of classifiers using two different nonlinear classification strategies optimized for the respective application: 1) intrapatient classifier: Bayesian classifiers [25] trained using a few expert defined training samples from within a single patient; and 2) interpatient classifier: SVM classifiers trained by combining tissue samples from several patients.

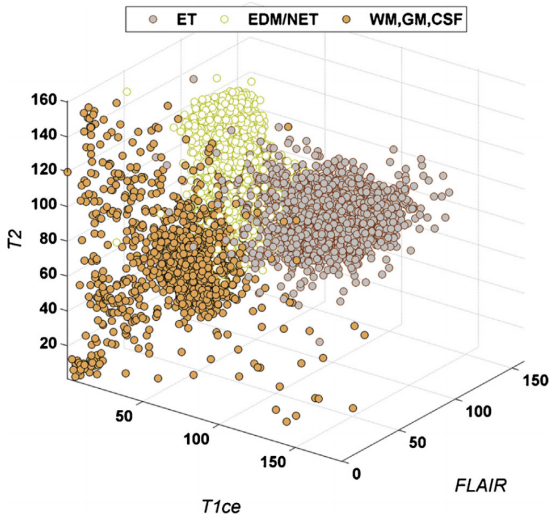


Figure 5.2: Multivariate intensity distributions using T1CE, FLAIR and T2 sequences for healthy tissue (WM, GM, CSF), non-enhancing or edematous tissue (EDM/NET) and enhancing tumor (ET) (from our work in [22]).

For intrapatient classification, we assumed multivariate Gaussian distribution for the features and used a Bayesian approach to design discriminant functions [25] for each of the six tissue classes. Different discriminant functions designed for each of the six tissue classes (i.e., ET, NET, ED, WM, GM, and CSF), evaluated at each voxel, provide the estimate of the probability of that voxel belonging to the respective class, and produce a three-dimensional voxel-wise probability map. This method of tissue classification is optimal when training samples are available for the patient whose tissue needs to be characterized. It effectively replicates the experts samples to identify regions that are similar. However, only tissue classes (ET, ED, NET) identified by the expert can be characterized for that patient, and because of the conservative nature of sample selection, expert identification may not be provided for all alternate tissue types. This requires pooling samples from several patients and, because of the high variability across individuals, Bayesian classification with its multinomial Gaussian assumption does not provide adequate classification.

For interpatient classification we combined training samples from across patients, to obtain more generalized tissue classification using SVM. Since SVMs are inherently two-class classifiers, a common strategy to do multiclass classification with SVMs is to build N *one-versus-rest* classifiers, where N is the number of classes (6 in our case), and to choose the class which classifies the data with greatest margin. Another way is to build a set of $N(N-1)/2$ *one-versus-one* classifiers, and to choose the class that is selected by the most classifiers. We chose the *one-versus-rest* strategy and converted the decision score for each class c returned by each binary classifier to a pseudo-probability score (p^c_{platt}) using Platt's method [16]. Then the pseudo-probability values were normalized so that they sum up to 1 for all classes: $p^c_{normalized} = p^c_{platt} / \sum_{c=1}^{N} p^c_{platt}$. These voxel-wise pseudo-probability scores form the tissue abnormality map pertaining to that classifier. Image segmentation was performed by assigning labels according to the maximum tissue

probability (after normalization).

More recently, in [26][22], we proposed a computationally efficient and (almost fully) automated method for segmenting neoplastic and healthy brain tissue using conventional multiparametric MRI. We did not use advanced MRI, such as DTI, because such advanced imaging techniques have inherently low signal-to-noise ratio compared to conventional MR modalities and, in addition, such acquisitions are not performed in clinical routine, especially during regular follow-up. Tissue segmentation was performed by combining a non-parametric intensity-based outlier detection scheme with the 3D Random Walker algorithm, and did not require expert-defined brain tissue samples in MR images. Due to its nearly unsupervised learning strategy, the method was presented in chapter 4.

Image-based disease classification

The malignancy of brain neoplasms is measured by the tumor grade which is determined by visually examining tissue sections (biopsies), based on guidelines determined by the World Health Organization (WHO). The classification of brain neoplasms is of critical clinical importance in making decisions regarding initial and evolving treatment strategies, for example high-grade gliomas are usually treated with adjuvant radioor chemotherapy after resection, whereas low-grade gliomas are not. The objective of our studies [27][28][29] was to provide an automated tool that integrates advanced MR with conventional MR imaging findings in order to assist in the radiological diagnosis of brain neoplasms by determining the glioma grade and differentiating between types, such as primary neoplasms (gliomas) from secondary neoplasms (metastases). Automated tools, if proven accurate, can ultimately be applied to (i) provide more reliable differentiation, especially when the neoplasm is heterogeneous and therefore cannot be adequately sampled by localized needle biopsy, (ii) circumvent invasive procedures such as biopsy, especially in cases where the risks outweigh the benefits, (iii) expedite or anticipate the diagnosis (histological examination is usually time consuming), and (iv) avoid the inter and intra observer variability observed when pathologists give different relative importance to each of the grading criteria. We explore the heterogeneous regions of brain tumors by combining imaging attributes from several sequences, extract morphological characteristics, and assess the significance of each attribute in classification. This approach incorporates imaging data which are acquired in a routine clinical protocol, such as multi-parametric conventional MRI and perfusion. The methods were was applied for pairwise classification, but also the multi-class classification problem was investigated for differentiating between the most common brain tumors: metastasis, meningioma (usually grade I), and gliomas (grade II, III, and IV) histopathologically diagnosed and graded according to the WHO system.

The investigated imaging attributes were similar in the two works [27][29] and included shape and multi-parametric intensity and texture characteristics over several regions of interest. However, experiments showed in [29] that accuracy did not significantly improve when textural characteristics were used, as in [27]. Thus, classification in [29] was finally based only on shape and intensity characteristics. In [27] we first reduced the number of features by eliminating the less relevant features using a forward selection method based on a ranking criterion, and then applied backward feature elimination using a feature subset selection method, such as the support vector machine recursive feature elimination (SVM-RFE) algorithm. Classification was performed by

starting with the more discriminative features and gradually adding less discriminative features, until classification performance no longer improved. Different pattern classification methods were investigated for comparison in both studies [27][29].

Attribute selection The attribute selection is a widely known process, during which a subset of the most informative attributes is chosen, so that the highest accuracy is achieved using the least number of variables. Attribute selection involves searching through all possible combinations of attributes in the data to find which subset of them works best for prediction. In [27] we first ranked and reduced the features based on the *t-statistics*, and then applied the support vector machine recursive feature elimination (SVM-RFE) algorithm for backward feature elimination. The goal of SVM-RFE is to find a subset of features that optimize the performance of the classifier. This algorithm determines the ranking of the features based on a backward sequential selection method that removes one feature at a time. At each time, the removed feature makes the variation of SVM-based leave-one-out error bound smallest, compared to removing other features. We applied the zero-order method for identifying the variable that produces the smallest value of the ranking criterion when removed, and used the weight magnitude $\|w^{(i)}\|$ as ranking criterion, defined as

$$\|w^{(i)}\|^2 = \sum_{j=1}^{N_s} \sum_{k=1}^{N_s} \alpha_k^{(i)} \alpha_j^{(i)} y_k y_j K^{(i)}(x_k, x_j) \tag{5.2}$$

where $K^{(i)}(x_k, x_j)$ is the Gram matrix for two feature vectors x_k and x_j when the variable i is removed and $\alpha^{(i)}$ is the corresponding solution of the SVM classifier.

For the purpose of comparison, we also evaluated the performance of constrained Linear Discriminant Analysis (CLDA) algorithm [30]. CLDA maximizes the discriminant capability between classes without transforming the original features, as done by traditional LDA or PCA.

The attribute selection algorithms are characterized by two components (i) the algorithm used to define the predictive value of each subset of attributes, denoted as *feature evaluator*, and (ii) the method determining the search over the attributes, denoted as *search method*. In [29], three evaluators were implemented in the WEKA platform [17]:

- a correlation-based feature selection (CFS) method [31], which evaluates the worth or merit of a subset of attributes by considering the individual predictive ability of each attribute along with the degree of redundancy between them.
- a method evaluating consistency in the class values [32] which evaluates the predictive value of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes. The consistency of any subset can never be lower than that of the full set of attributes.
- an approach based on wrappers [33] in which an induction learning algorithm is applied repeatedly on a distinct portion of the dataset using various feature subsets. A classifier is built on each feature subset using a set aside distinct portion of the dataset, and the feature subset with the highest performance (measured by some criterion) is used as the final set.

Also, three search methods were examined:

- The Best First [34] searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. It starts with the empty set of attributes and searches forward.

- Greedy Stepwise [35] method performs a greedy forward or backward search through the space of attribute subsets. It starts with a population of many significant and diverse subsets and stops when the accuracy is higher than a given threshold or there is not more improvement.

- Scatter Search [36] is an evolutionary method that combines solution vectors by linear combinations to produce new ones through successive generations.

Classification algorithms In [27] three pattern classification methods were implemented for comparison: LDA with Fisher's discriminant rule [37], k-nearest neighbor (kNN) [38], and nonlinear SVMs. In LDA, a transformation function is sought that maximizes the ratio of between-class variance to within-class variance. Since usually there is no transformation that provides complete separation, the goal is to find the transformation that minimizes the overlap of the transformed distributions. The kNN algorithm [38] finds the nearest (most similar) training samples to the test sample. The class label in majority among the k-nearest neighbors is assigned to the new sample.

In [29], a different set of classifiers was examined and compared against the SVM-based criteria used in [27] aiming to improve classification accuracy. The investigated classifiers were the kNN [38] (as in [27] but combined with different feature selection techniques), J48 tree [39], VFI [40] and Nave Bayes [41]. The classification algorithms are briefly described next.

- J48 [39] is an implementation of C4.5 algorithm that produces decision trees from a set of labeled training data using the concept of information entropy. It examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data into smaller subsets. To make the decision, the attribute with the highest normalized information gain is used.

- Learning in the VFI algorithm [40] is achieved by constructing feature intervals around each class for each attribute (basically discretization) on each feature dimension. Class counts are recorded for each interval on each attribute and classification is performed by a voting scheme.

- The Nave Bayesian Classifier [41] assumes that features are independent. Given the observed feature values for an instance and the prior probabilities of classes, the a posteriori probability that an instance belongs to a class, is estimated. The class prediction is the class with the highest estimated probability. The SVMs have been described in section 5.1.

Overall, the highest accuracy was achieved by the wrapper evaluator in combination with the Best First search method for both pairwise problems (low versus high-grade gliomas and gliomas versus metastases) and multiclass problems. The classifier achieving the highest accuracy was the kNN(k=3) or the VFI depending on the classification problem, but the kNN was preferred due to its simplicity and overall more stable performance. When the attribute selection methods presented in [29] were used in combination with an SVM classifier, the results were similar or worse than

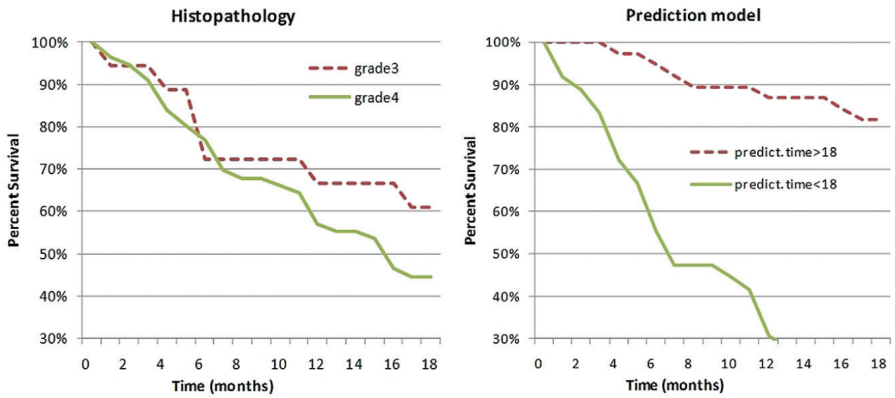


Figure 5.3: Kaplan-Meier survival curves of patients with 56 gliomas of grade 4 and 18 anaplastic gliomas of grade 3 [43]. Survival curves were plotted according to classifications based on either traditional histopathology (dashed line, grade 3; solid line, grade 4) or the class prediction model (dashed line, predicted survival time > 18 months; solid line, predicted survival time < 18 months).

those in previous work [27] using weighted SVMs [42]; however the accuracy increased when a VFI or kNN(k=3) classifier was applied instead of the SVMs. The good performance of the kNN classifier in [29] might be attributed to the significantly small number of retained attributes ($N = 2.4$ and $N = 2.7$ on the average for all pairwise classification problems when the Best First and the Greedy Stepwise search algorithms were used, respectively).

Prediction of survival by decision trees The prediction of prognosis in high-grade gliomas is poor in the majority of patients. In [43] our aim was to test whether multivariate prediction models constructed by machine-learning methods provide a more accurate predictor of prognosis in high-grade gliomas than histopathologic classification. The prediction of survival was based on the integration of clinical, morphological and imaging characteristics from DTI and rCBV measurements as an adjunct to conventional imaging. Overall survival was evaluated from the baseline to death or, for cases that were not followed until death (e.g. living patients) from the baseline to the time of last available follow-up. A time threshold of 18 months was defined to differentiate the patients into 2 groups, those with short- or long-term survival. We tested several variable selection algorithms and selected the one with overall best classification performance. The selected method searched over the variables following the scatter search algorithm [36] and defined the predictive value of each subset of variables by using a wrapper approach [33]. Classification of the datasets into short- and long-term survivors was performed with a J48 classification tree [39]. The variables selected as the most discriminative were the extent of resection, mass effect, volume (in cubic millimeters) of enhancing tissue, maximum B0 (baseline T2-weighted image) intensity in the region of non-enhancing tissue, and the mean trace intensity in non-enhancing tissue. Kaplan-Meier curves showed that when tumors were classified according to histopathology (grade 3 versus grade 4), the survival of patients was not significantly different ($P = 0.17$), whereas class distinctions

according to the calculated prediction model were significantly associated with survival outcome ($P < 0.00001$) (Fig.5.3).

Estimation of disease progression

Brain lesions, especially White Matter Lesions (WMLs), are associated with cardiac and vascular disease, but also with normal aging. Quantitative analysis of WML in large clinical trials is becoming more and more important. Because of the decreased contrast between WM and GM in MRI in elderly, techniques that require the segmentation of WM and GM for the extraction of the WMLs perform moderately well when applied to geriatric patients, especially when they were originally designed and trained to extract lesions in MS patients. Only a few methods have combined space and time into the lesion characterization process [8,9]. These approaches focused primarily on quantifying the temporal variations of multiple sclerosis lesions, important in differentiating active from chronic lesions. In contrast to the complicated MRI dynamics of lesions in multiple sclerosis, the monitoring of WMLs does not require spatiotemporal modeling, since the effects in WMLs are irreversible.

In [44], we present a computer-assisted WML segmentation approach that has been designed to process longitudinal MR scans of elderly diabetes patients [45]. Image intensities from multiple MR acquisition protocols, after co-registration, are used to form a voxel-wise feature vector that helps to discriminate lesion from various normal tissue image profiles. First, we jointly preprocess baseline and follow-up data. The preprocessing step includes co-registration of different MR modalities of the same patient, intra-modal registration of follow-up to baseline, skull-stripping, intensity normalization, as well as inhomogeneity correction. Then a supervised classification model is built via SVM and the AdaBoost algorithm [46] using training samples delineated by an expert reader on the baseline images. This model is then used in the testing stage to perform voxel-wise segmentation of the longitudinal images of a new subject. The key points of this approach [44] include the (i) histogram normalization based on temporal variance reduction, (ii) training via Adaboost to reduce the effect of unbalanced classes and (iii) false positive elimination via thresholding of feature vector distance in Hilbert space.

- *Histogram normalization*: A fundamentally important step in supervised classification is the standardization of features, such as image intensities. To this end, it is common to linearly align histograms to a template histogram. In order to achieve high temporal stability, we aligned follow-up histograms to their respective - standardized to the template - baselines, a problem that is relatively easy to solve, since baseline and follow-up images belong to the same individual. Thus, histograms between images of the same subject were aligned consistently, and the temporal variance was reduced. We referred to this as *temporal variance reduction (TVR)* approach, as opposed to the standard approach aligning baseline and follow-up images independently to the template histogram. The latter aims at reducing the *inter-subject variance* and was referred to as *IVR*. Since the inter-subject variability (between subject and template) is much larger than the intra-subject temporal variability, a global histogram matching based on *IVR* tends to produce more inconsistencies in temporal WML segmentation.

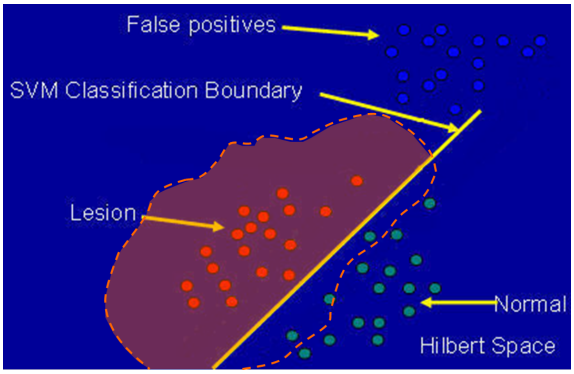


Figure 5.4: False positive reduction by combination of two solution spaces (shown as red area), from intra-class distances and linear SVM.

- *Training via AdaBoost*: Because the number of normal tissue voxels is far larger than the number of lesion voxels, it is essential to select only a representative set of normal tissue voxels comparable to the number of lesion voxels. This selection is not random, but it is rather guided by the classification results themselves, using the AdaBoost algorithm [46]. This approach is based on a sequence of classifiers that rely increasingly on misclassified voxels, since those are presumably the voxels on which the classifier must focus. During this adaptive boosting procedure, each sample receives a weight that determines its probability of being selected in a training set for the next iteration. If a training sample is accurately classified, then its likelihood of being used again in subsequent iterations is reduced; conversely, if a training sample is inaccurately classified, then its likelihood of being used again is increased.

- *Constraining SVM solution space*: Normal tissue is a heterogeneous class and therefore the multi-parametric features exhibit a large variability. The sampling process during training of the SVM model causes a significant under-representation of normal tissue with rare intensity profile. This type of tissue is far away from both normal and lesion tissue classes and often is classified as lesions by the linear SVM model. In order to eliminate these false positives, we learn the intra-class distances of the training samples and accordingly constrain the solution space of the SVM model. Fig.5.4 illustrates this concept.

It should be noted that all steps in the WML segmentation procedure are automated and the same parameters are used for all subjects. Only one parameter has been shown to be important and vary across subjects, which is the threshold for binarizing the abnormality map generated by the SVM classifier. This threshold is optimized for each subject in the training set by maximizing the *Jaccard* score. The average threshold maximizing the *Jaccard* score is then used as the default value for segmenting new data. Fig.5.5 shows the WML segmentation in baseline (top row) and follow-up (bottom row) images of an elderly subject. It can be noticed that the utilized segmentation algorithm [44] has high sensitivity and specificity. Moreover, lesion volume measurements showed that the proposed *TVR* approach for jointly normalizing the histograms of baseline and

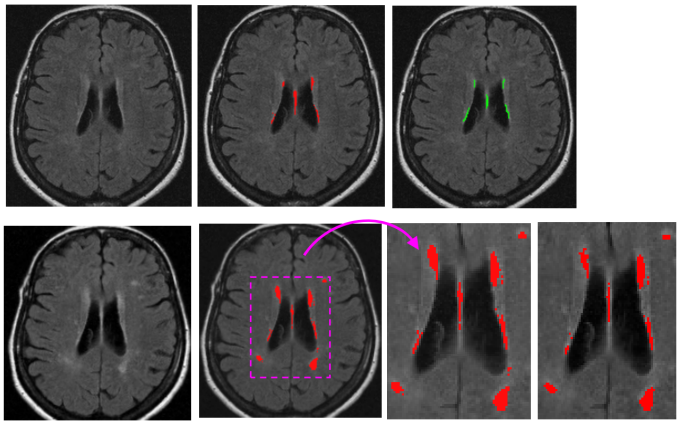


Figure 5.5: Identification of WM lesion progression in an elderly subject. Top row: The automatic WML segmentation (red) in baseline is very similar to the expert-defined WML (green). Bottom row: Segmentation in the follow-up image using two histogram equalization approaches, i.e. the proposed *TVR* (left) and the standard *IVR* (right).

follow-up images leaded to increasing lesion volume for all subjects. Since only disease progression is expected, the results confirmed that the *TVR* approach is more appropriate for measuring temporal WML change.

Note: In respect to personal implementation of computational algorithms, methods published more or less before 2010 were re-implemented (in C programming language or Matlab), such as the Adaboost algorithm [44] or the SVM-RFE [27], whereas the availability of WEKA platform after 2010 facilitated the analysis significantly and allowed large scale comparisons of feature selection and classification algorithms [29][43].

5.3 Single- and multi-label classification of molecular structures

The number of protein structures in the PDB database increased more than fifteen-fold since 1999. The creation of computational models predicting enzymatic function is of major importance since such models provide the means to better understand the behavior of newly-discovered enzymes when catalyzing chemical reactions. Until now, single-label classification has been widely performed for predicting enzymatic function limiting the application to enzymes performing unique reactions and introducing errors when multi-functional enzymes were examined. Indeed, some enzymes may be performing different reactions and can hence be directly associated with multiple enzymatic functions. Building upon our first work on single-label enzymatic function prediction [47], we developed a multi-label classification scheme that combines structural and amino acid sequence information [48]. The methodology was evaluated for general enzymatic function prediction indicated by the first digit of the Enzyme Commission (EC) code (6 main classes) on 40,034 enzymes from the PDB database.

The proteins were represented by structural descriptors capturing the shape of the protein

backbone and sequence-based features reflecting the success of local alignment of two sequences. Details on the utilized features are provided in section 3.1. Two classification techniques were investigated, the nearest neighbor (NN) and SVM. The classifiers were trained using a number of annotated examples and then tested on novel enzymes. Two types of classification models were produced: single-label models for enzymes performing unique reactions and multi-label models for the multi-functional enzymes. Fusion of structural and amino acid sequence information was performed in two different ways, i.e. in the feature level and in the decision level. The decision-level fusion approach associates class probabilities for structural information (f_{SI}) obtained by SVM or NN, with class probabilities for amino acid sequence (f_{AA}) through a heuristic fusion rule. The applied fusion rule performs weighted averaging of class probabilities, resulting to fused class probability $(1-\alpha)(f_{SI})+\alpha(f_{AA})$. The value of α was estimated for each classification method by minimizing the empirical error [47]. In the single-label classification, class assignment was based on the maximum a posteriori probability. This decision rule was not appropriate for the multi-label scenario. In the latter case, the class probabilities (based on fused features) were introduced into a multi-label SVM or multi-label NN classifier, which computed a 6-dimensional vector with binary values indicating if the test sample belongs to each class or not [48]. When both single- and multi-label enzymes were mixed during training, a slight improvement in prediction accuracy was observed.

Furthermore, we investigated techniques for learning from imbalanced classes and specifically, we tested an adaptive synthetic sampling approach (ADASYN) [49], in which a weighted distribution for the minority class examples is used. Results based on ADASYN didn't show any significant change in classification accuracy on the enzyme dataset.

Bibliography

[1] E. I. Zacharaki, K. Garganis, I. Mporas, and V. Megalooikonomou, "Spike detection in eeg by lpp and svm", in *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, IEEE, 2014, pp. 668–671.

[2] E. I. Zacharaki, I. Mporas, K. Garganis, and V. Megalooikonomou, "Spike pattern recognition by supervised classification in low dimensional embedding space", *Brain Informatics*, vol. 3, no. 2, pp. 73–83, 2016.

[3] X. Niyogi, "Locality preserving projections", in *Neural information processing systems*, MIT, vol. 16, 2004, p. 153.

[4] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[5] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers", in *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 1992, pp. 144–152.

[6] I. Mporas, P. Korvesis, E. I. Zacharaki, and V. Megalooikonomou, "Sleep spindle detection in eeg signals combining hmms and svms", in *Engineering Applications of Neural Networks*, Springer, 2013, pp. 138–145.

[7] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition", *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[8] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains", *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[9] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, *et al.*, "The htk book", *Cambridge university engineering department*, vol. 3, p. 175, 2002.

[10] E. Pippa, E. I. Zacharaki, I. Mporas, V. Tsirka, M. P. Richardson, M. Koutroumanidis, and V. Megalooikonomou, "Improving classification of epileptic and non-epileptic eeg events by feature selection", *Neurocomputing*, vol. 171, pp. 576–585, 2016.

[11] V. Megalooikonomou, D. Triantafyllopoulos, E. I. Zacharaki, and I. Mporas, "Cyberphysical systems for epilepsy and related brain disorders: Multi-parametric monitoring and analysis for diagnosis and optimal disease management", in, S. N. Voros and P. C. Antonopoulos, Eds. Springer International Publishing, 2015, ch. Offline Analysis Server and Offline Algorithms, pp. 239–254.

[12] E. Pippa, E. I. Zacharaki, I. Mporas, V. Megalooikonomou, V. Tsirka, M. Richardson, and M. Koutroumanidis, "Classification of epileptic and non-epileptic eeg events", in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, IEEE, 2014, pp. 87–90.

[13] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data", *Machine learning*, vol. 9, no. 4, pp. 309–347, 1992.

[14] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[15] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms", *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.

[16] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization", *Advances in kernel methods*, pp. 185–208, 1999.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update", *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[18] I. Kononenko, "Estimating attributes: Analysis and extensions of relief", in *European conference on machine learning*, Springer, 1994, pp. 171–182.

[19] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on mri brain tumor segmentation", *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.

[20] E.-S. A. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer-aided diagnosis of human brain tumor through mri: A survey and a new algorithm", *Expert systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.

[21] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)", *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[22] V. G. Kanas, E. I. Zacharaki, C. Davatzikos, K. N. Sgarbas, and V. Megalooikonomou, "A low cost approach for brain tumor segmentation based on intensity modeling and 3d random walker", *Biomedical Signal Processing and Control*, vol. 22, pp. 19–30, 2015.

[23] R. Verma, E. I. Zacharaki, Y. Ou, H. Cai, S. Chawla, S.-K. Lee, E. R. Melhem, R. Wolf, and C. Davatzikos, "Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of mr images", *Academic radiology*, vol. 15, no. 8, pp. 966–977, 2008.

[24] E. I. Zacharaki, R. Verma, S. Chawla, E. R. Melhem, R. Wolf, and C. Davatzikos, "Towards predicting neoplastic recurrence with multi-parametric mr", in *ISMRM 16th Scientific Meeting and Exhibition 2008*, 2008.

[25] G. McLachlan, *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, 2004, vol. 544.

[26] V. G. Kanas, E. I. Zacharaki, E. Dermatas, A. Bezerianos, K. Sgarbas, and C. Davatzikos, "Combining outlier detection with random walker for automatic brain tumor segmentation", in *Artificial Intelligence Applications and Innovations*, Springer, 2012, pp. 26–35.

[27] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme", *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[28] C. Davatzikos, E. I. Zacharaki, A. Gooya, and V. Clark, "Multi-parametric analysis and registration of brain tumors: Constructing statistical atlases and diagnostic tools of predictive value", in *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011, pp. 6979–6981.

[29] E. I. Zacharaki, V. G. Kanas, and C. Davatzikos, "Investigating machine learning techniques for mri-based classification of brain neoplasms", *International journal of computer assisted radiology and surgery*, vol. 6, no. 6, pp. 821–828, 2011.

[30] C. Zifeng, X. Baowen, Z. Weifeng, J. Dawei, and X. Junling, "Clda: Feature selection for text categorization based on constrained lda", in *International Conference on Semantic Computing (ICSC 2007)*, IEEE, 2007, pp. 702–712.

[31] M. A. Hall, "Correlation-based feature selection of discrete and numeric class machine learning", 2000.

[32] M. Dash and H. Liu, "Consistency-based search in feature selection", *Artificial intelligence*, vol. 151, no. 1, pp. 155–176, 2003.

[33] R. Kohavi and G. H. John, "Wrappers for feature subset selection", *Artificial intelligence*, vol. 97, no. 1, pp. 273–324, 1997.

[34] L. Xu, P. Yan, and T. Chang, "Best first strategy for feature selection", in *Pattern Recognition, 1988., 9th International Conference on*, IEEE, 1988, pp. 706–708.

[35] R. Caruana and D. Freitag, "Greedy attribute selection.", in *ICML*, Citeseer, 1994, pp. 28–36.

[36] M. Laguna and R. Marti, *Scatter search: Methodology and implementations in c*. Springer Science & Business Media, 2012, vol. 24.

[37] P. A. Lachenbruch and M. Goldstein, "Discriminant analysis", *Biometrics*, pp. 69–85, 1979.

[38] T. Cover and P. Hart, "Nearest neighbor pattern classification", *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[39] G. M. Gandhi and S. Srivatsa, "Adaptive machine learning algorithm (amla) using j48 classifier for an nids environment", *Adv Comput Sci Technol*, vol. 3, no. 3, pp. 291–304, 2010.

[40] G. Demiröz and H. A. Güvenir, "Classification by voting feature intervals", in *European Conference on Machine Learning*, Springer, 1997, pp. 85–92.

[41] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers", in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

[42] Y.-M. Huang and S.-X. Du, "Weighted support vector machine for classification with uneven training class sizes", in *2005 International Conference on Machine Learning and Cybernetics*, IEEE, vol. 7, 2005, pp. 4365–4369.

[43] E. Zacharaki, N. Morita, P. Bhatt, D. O'rourke, E. Melhem, and C. Davatzikos, "Survival analysis of patients with high-grade gliomas based on data mining of imaging variables", *American Journal of Neuroradiology*, vol. 33, no. 6, pp. 1065–1071, 2012.

[44] E. I. Zacharaki, S. Kanterakis, R. N. Bryan, and C. Davatzikos, "Measuring brain lesion progression with a supervised tissue classification system", in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, Springer, 2008, pp. 620–627.

[45] J. D. Williamson, M. E. Miller, R. N. Bryan, R. M. Lazar, L. H. Coker, J. Johnson, T. Cukierman, K. R. Horowitz, A. Murray, L. J. Launer, *et al.*, "The action to control cardiovascular risk in diabetes memory in diabetes study (accord-mind): Rationale, design, and methods", *The American journal of cardiology*, vol. 99, no. 12, S112–S122, 2007.

[46] R. E. Schapire and Y. Freund, *Boosting: Foundations and algorithms*. MIT press, 2012.

[47] A. Amidi, S. Amidi, D. Vlachakis, N. Paragios, and E. I. Zacharaki, "A machine learning methodology for enzyme functional classification combining structural and protein sequence descriptors", in *Bioinformatics and Biomedical Engineering*, Springer, 2016, pp. 728–738.

[48] ——, "Automatic single- and multi-label enzymatic function prediction by machine learning", *IEEE/ACM Trans. Computational Biology and Bioinformatics*, 2016, under review.

[49] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning", in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, IEEE, 2008, pp. 1322–1328.

Chapter 6: Ongoing and future research

6.1 Development of new quantitative imaging biomarkers for obstructive and interstitial lung diseases

Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. The goal of this thesis is to develop quantitative imaging biomarkers allowing (i) severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis (CF) as models for restrictive and obstructive lung disease, respectively. Two different approaches will be investigated: disease assessment by histogram or texture analysis and assessment of the regional lung elasticity through deformable registration.

Histogram and texture analysis Although morphological changes on CT have shown to be correlated with clinical endpoints such as survival, quality of life and exacerbation rate, CT remains unapproved by both European Medicines Agency (EMA) and Food and Drug Administration (FDA) as a secondary endpoint. Main drawbacks to its use are the radiation dose and the limitations of CT visual scoring methods. The development of automated scoring could solve the issues encountered with visual methods such as questionable repeatability, difficulty of use and time-consuming scoring and training, preventing their use in clinical practice. To date, no automated scoring method has been proposed. We have recently introduced an automated density-based scoring method to quantify CF-related lung disease.

Indeed, most morphological changes related to CF show attenuation values higher than that of the normal lung, and we hypothesized that an automated quantification of high attenuating structures in CF could reflect disease severity (Fig.6.1). The developed CT-Density score correlated well to the percentage of predicted forced expiratory volume in 1 second (FEV1) and to its evolution, and was only slightly inferior to visual scoring. However, our automated score is much faster (less than 2 minutes per CT) than to visual scoring, highly repeatable and does not require complex training. We validated the good correlation to FEV1 in two independent cohorts of adults CF patients. These first results have been submitted for publication [1]. My personal involvement on this analysis was reduced but I am very active in the extension of this work involving mainly (i) the

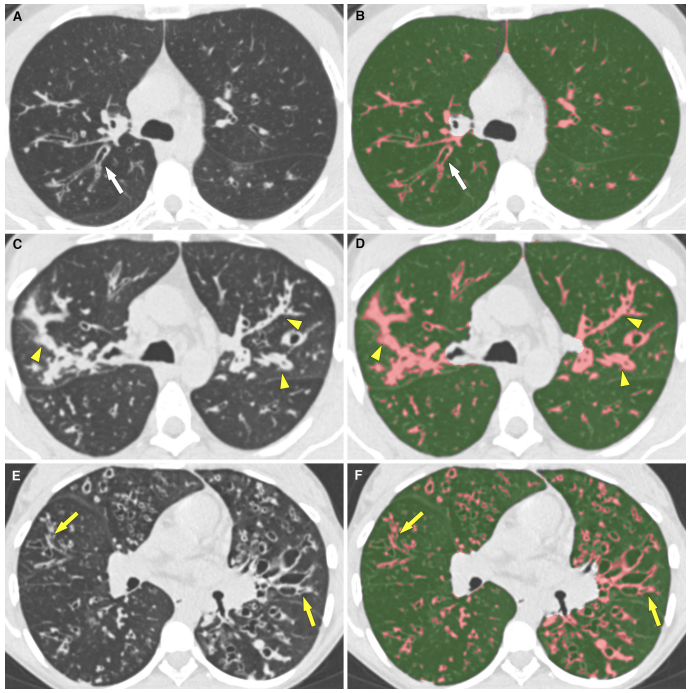


Figure 6.1: Automated CT scoring in patients of the validation cohort with various disease severity. (A) Axial CT image in a patient with mild lung disease (FEV1=77%). Bronchiectasis and bronchial wall thickening are seen in the posterior segment of the right upper lobe (white arrow). (B) Scoring with Mode+300 HU results in a CT-Density score of 4.4. (C) Axial CT image in a patient with moderate disease (FEV1=56%) shows bilateral mucus plugging (yellow arrowheads). (D) CT-Density score is 9.8. (E) Axial CT image in a patient with severe disease (FEV1=31%) shows diffuse bronchial bronchiectasis and bronchial wall thickening (yellow arrows). (F) CT-Density score is 14.5.

automated extraction of features, (ii) statistical analysis by advanced methods, and (iii) correlation to other (than FEV1) clinical scores.

Assessment of lung elasticity through deformable registration Fibrosis results in increased lung stiffness that induces a restrictive lung disease with decreased lung volumes on pulmonary function testing (PFT). We hypothesize that the assessment of lung elasticity will allow differentiation between fibrotic and healthy lung parenchyma.

Lung function depends on lung expansion and contraction during the respiratory cycle. In order to locally estimate regional lung volume change, respiratory-gated CT imaging and 3D image registration can be performed. The work in [2] aimed to assess lung function through registration of volumetric images obtained at inspiration and expiration, and specifically to classify the lung tissue which is abnormal as having air trapping only without emphysema, or as having emphysema. Also in [3][4] follow-up and baseline lung CT images were coregistered in order to obtain

an estimation of regional destruction of lung tissue for subjects with a significant difference in inspiration level between CT scans. To account for differences in lung intensity owing to differences in the inspiration level in the two scans rather than disease progression, they adjusted the density of lung tissue with respect to local expansion or compression such that the total weight of the lungs is preserved during deformation. In [5], the authors examine the reproducibility of Jacobian-based measures of lung tissue expansion in two repeat 4DCT acquisitions of mechanically ventilated sheep and free-breathing humans. The (voxel-wise) determinant of the Jacobian matrix of the deformation field (computed by deformable registration) provides a measure of local lung tissue expansion and contraction. The (per voxel) Jacobian values were calculated between end inspiration and end exhalation images. The reproducibility of the Jacobian values was found to be strongly dependent on the reproducibility of the subject's respiratory effort and breathing pattern. Global linear normalization was applied to globally compensate for breathing effort differences. Such a homogeneous scaling does not account for differences in regional lung expansion rates. Ding et al. [6] derived and compared three different registration based measures of regional lung mechanics: the specific air volume change calculated from the Jacobian, the specific air volume change calculated by the corrected Jacobian (SACJ), and the specific air volume change by intensity change. All three ventilation measures were evaluated by comparing to Xe-CT estimates of regional ventilation. Significant differences between the three measures were found with the SACJ providing better correlations with Xe-CT based sV than the other two measures, thus providing an improved.

The framework we propose is very similar to the one in [5]. Instead of using the nonrigid registration algorithm proposed in [7], we will apply the DROP algorithm [8], which uses graphical models and discrete optimization, and can be adapted to geometric (landmark based) or iconic (intensity based) matching criteria. Graphical models are powerful formalisms that allow us to model most vision problems in a scalable and modular way. As representation of lung tissue expansion and contraction we will examine the several proposed surrogates for regional ventilation based on the determinant of the Jacobian of the deformation field [6]. For each patient, registration will be applied between serial CT examinations (baseline and follow-up) avoiding the need to acquire expiratory CT images (the clinical protocol does not allow the acquisition of both inspiratory and expiratoty CT due to increased irradiation). The developed tools will be evaluated on data from Cochin hospital.

6.2 Deep learning for image segmentation

Disease assessment through tissue characterization is very common in medical image analysis. Several intensity or texture-based descriptors have been proposed, as well as subsequent classifiers to perform a voxel-wise characterization of the tissue. The latter usually apply machine learning techniques, in which a number of previously annotated samples (voxels) is used in order to train models that can subsequently classify new samples, e.g. as healthy or diseased. Most frameworks are application-specific and might not perform well under different conditions (patients' characteristics, image acquisition protocol or scanner used, noise level, preprocessing, etc).

In the past few years, deep learning techniques, and particularly Convolutional Neural Networks (CNN), have rapidly become the tool of choice for tackling many challenging computer

vision tasks, such as image classification [9], object detection, texture recognition and object semantic segmentation [10]. The main advantage of deep learning techniques is the automatic exploitation of features and tuning of performance in a seamless fashion, that simplifies the conventional image analysis pipelines. Deep belief networks and a convolutional neural networks were recently used for lung nodule classification in CT images [11][12][13]. In [14] a method is presented to classify imaging patterns on CT images with interstitial lung diseases. Six lung tissue types were considered: normal , emphysema, ground glass, fibrosis, micronodules, consolidation. Different from [15][16][17], their CNN based method is formulated as a holistic image recognition task that is also considered as a weakly supervised learning problem. In [18] emphysema classification is performed by employing an automatic feature extractor based on CNNs. High-resolution CT was used to classify lung tissue into normal, centrilobular emphysema and paraseptal emphysema.

We plan to apply a deep learning framework for segmentation of fibrotic tissue in scleroderma patients. Most of scleroderma-related interstitial lung disease (76%) typically manifests on CT as reticulations, ground-glass opacities and traction bronchiectasis with basal and peripheral predominance [19]. Unlike previous CNN-based methods, the model we propose performs convolution in 3D and does not require any alignment or registration steps at testing time. A previous method [20] was implemented in our group, which combined the output of a 2D CNN with a Markov Random Field in order to impose spatial (volumetric) homogeneity. It was applied for sub-cortical brain structure segmentation and showed very promising results on two different brain MRI datasets.

6.3 Deep learning for protein structure classification

There have been plenty machine learning approaches in the literature for automatic enzyme annotation. A systematic review on the utility and inference of various computational methods for functional characterization is presented in [21], while a comparison of machine learning approaches can be found in [22]. There has been little work in the literature on automatic enzyme annotation based only on structural information. In the past few years, data-driven CNN models have become very popular because they tend to be domain agnostic and attempt to learn additional feature bases that cannot be represented through any handcrafted features. CNNs have recently been used for protein secondary structure prediction [23][24]. In [23] prediction was based on the position-specific scoring matrix profile (generated by PSI-BLAST), whereas in [24] 1D convolution was applied on features related to the amino acid sequence. Also a deep CNN architecture was proposed in [25] to predict protein properties. This architecture used a multilayer shift-and-stitch technique to generate fully dense per-position predictions on protein sequences. To the best of our knowledge, deep CNNs have not been used for prediction of protein function so far.

In our current work (submitted for publication [26]) we exploit experimentally acquired structural information of enzymes through deep learning techniques in order to produce models that predict enzymatic function based on structure. The novelty of the proposed method lies first in the representation of the 3D structure as a bag of atoms (amino acids) which are characterized by geometric properties, and secondly in the exploitation of the extracted feature maps by deep CNNs. We hypothesize that by combining amino acid specific descriptors with the recent advances in deep learning we can boost model performance. Although assessed for enzymatic function prediction,

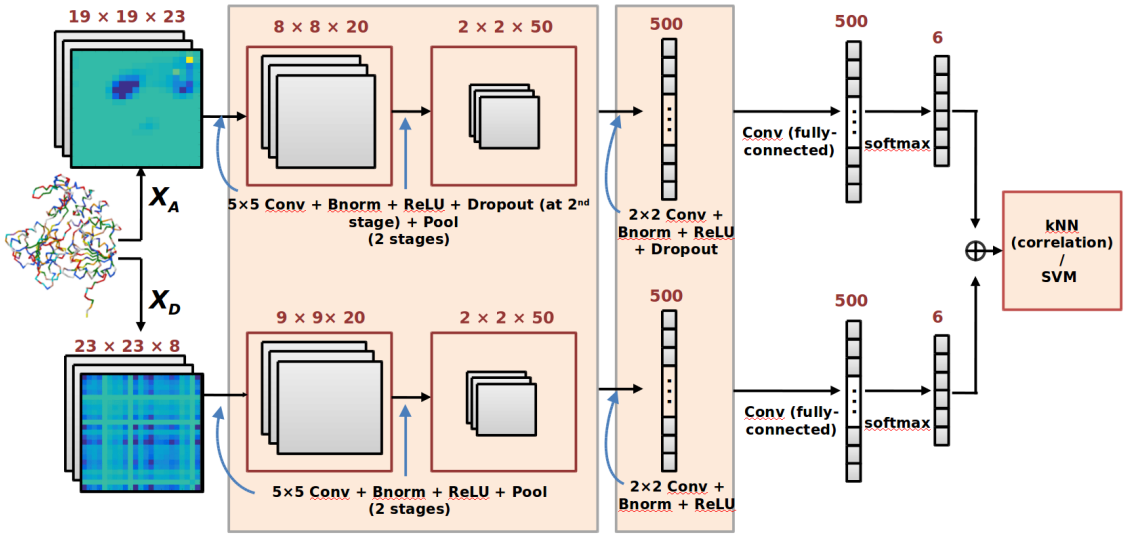


Figure 6.2: The deep CNN ensemble for protein classification. In this architecture each multi-channel feature set is introduced to a CNN and results are combined by kNN or SVM classification.

the method is not based on enzyme-specific properties and therefore can be extended to other 3D molecular structures, thus providing a useful tool for automatic large-scale annotation. The main advantage of our method is that it exploits complementarity in both data representation phase and learning phase. Regarding the former, the method uses an enriched geometric descriptor that combines local shape features with features characterizing the interaction of amino acids on this 3D spatial model. Shape representation is encoded by the local (per amino acid type) distribution of torsion angles [27]. Amino acid interactions are encoded by the distribution of pairwise amino acid distances. While the torsion angles and distance maps are usually calculated and plotted for the whole protein [27], in our approach they are extracted for each amino acid type separately, therefore characterizing local interactions. More details were provided in section 3.1. Thus, the protein structure is represented as a set of multi-channel images which can be introduced into any machine learning scheme designed for fusing multiple 2D feature maps.

Our method constructs an ensemble of deep CNN models that are complementary to each other. The deep network outputs are combined and introduced into a correlation-based k-nearest neighbor (kNN) classifier for function prediction. For comparison purposes, SVM were also implemented for final classification. Two system architectures were investigated in which the multiple image channels are considered jointly or independently, as will be described next. Both architectures use the same CNN structure which is illustrated in Fig.6.2.

Classification by deep CNNs

The CNN architecture employs three computational blocks of consecutive convolutional, batch normalization, rectified linear unit (ReLU) activation, dropout (optionally) and pooling layers,

and a final fully-connected layer. The convolutional layer computes the output of neurons that are connected to local regions in the input in order to extract local features. It applies a 2D convolution between each of the input channels and a set of filters. The 2D activation maps are calculated by summing the results over all channels and then stacking the output of each filter to produce the output 3D volume. The convolutional layer used neurons with receptive field size 5 for the first two layers and 2 for the third layer. The stride (specifying the sliding of the filter) was always 1. The number of filters was 20, 50 and 500 for the three layers, respectively. Batch normalization normalizes each channel of the feature map by averaging over spatial locations and batch instances. The ReLU layer applies an element-wise activation function, such as the $max(0, x)$ thresholding at zero. The dropout layer (applied at only one of the two feature sets) is used to randomly drop units from the CNN during training and reduce overfitting. The pooling layer performs a downsampling operation along the spatial dimensions (by $[2 \times 2]$ in our implementation); it is used to capture the most relevant global features with fixed length. The last layer is fully-connected and represents the class scores.

The output of each CNN is a vector of probabilities, one for each of the possible enzymatic functions. We can measure the CNN performance by a loss function which assigns a penalty to classification errors. The CNN parameters are learned to minimize this loss averaged over the annotated (training) dataset. The 'softmaxloss' function (i.e. the 'softmax' operator followed by the logistic loss) is applied to predict the probability distribution over categories. Optimization was based on an implementation of stochastic gradient descent. At the testing stage the features after 'softmax' normalization are used as class probabilities.

Fusion of CNN outputs

Two system architectures were implemented. In the first architecture the two feature sets based on torsion angles and amino acid distances (see section 3.1) are each introduced into a CNN, which performs convolution at all channels, and then the class probabilities produced for each set are combined into a feature vector. In the second architecture, each one of the channels of each feature set is introduced separately into a CNN and the obtained class probabilities for each channel are concatenated into a vector. Then the produced vectors for each feature representation are further combined into a single vector. For both architectures, kNN classification was applied for final class prediction using as distance metric between two feature vectors x_1 and x_2 the metric $1 - cor(x_1, x_2)$, where cor is the sample Spearman's rank correlation. The obtained results were also compared with linear SVM classification [28]. The code was developed in MATLAB environment and the implementation of CNNs was based on MatConvNet.

The method has been applied for the prediction of the primary EC number and achieved 90.1% accuracy, which is a considerable improvement over the accuracy achieved in our previous work [29] when only structural information was incorporated. Overall, our approach can provide quick protein function predictions on extensive datasets opening the path for relevant applications, such as pharmacological target identification. Moreover, the investigation of protein function based only on structure can reveal relationships hidden at the sequence level and provide the foundation to build a better understanding of the molecular basis of biological complexity.

Bibliography

[1] G. Chassagnon, C. Martin, P.-R. Burgel, D. Hubert, I. Fajac, N. Paragios, E. Zacharaki, P. Legmann, J. Coste, and M.-P. Revel, "Structural abnormalities in the cystic fibrosis lung: An automated computed tomography score", *American Journal of Respiratory and Critical Care Medicine*, vol. submitted, 2017.

[2] Y. Yin, P. Raffy, and S. A. Wood, *Visualization and quantification of lung disease utilizing image registration*, US Patent 20,150,332,454, Nov. 2015.

[3] V. Gorbunova, P. Lo, H. Ashraf, A. Dirksen, M. Nielsen, and M. de Bruijne, "Weight preserving image registration for monitoring disease progression in lung ct", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2008, pp. 863–870.

[4] V. Gorbunova, "Image registration of lung ct scans for monitoring disease progression", PhD thesis, KøbenhavnsUniversitetKøbenhavns Universitet, Det Natur-og Biovidenskabelige FakultetFaculty of Science, Datalogisk InstitutDepartment of Computer Science, 2010.

[5] K. Du, J. E. Bayouth, K. Cao, G. E. Christensen, K. Ding, and J. M. Reinhardt, "Reproducibility of registration-based measures of lung tissue expansion", *Medical physics*, vol. 39, no. 3, pp. 1595–1608, 2012.

[6] K. Ding, K. Cao, M. K. Fuld, K. Du, G. E. Christensen, E. A. Hoffman, and J. M. Reinhardt, "Comparison of image registration based measures of regional lung ventilation from dynamic spiral ct with xe-ct", *Medical physics*, vol. 39, no. 8, pp. 5084–5098, 2012.

[7] K. Cao, K. Ding, G. E. Christensen, and J. M. Reinhardt, "Tissue volume and vesselness measure preserving nonrigid registration of lung ct images", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2010, pp. 762 309–762 309.

[8] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through mrfs and efficient linear programming", *Medical image analysis*, vol. 12, no. 6, pp. 731–741, 2008.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs", *ArXiv preprint arXiv:1606.00915*, 2016.

[11] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique", *OncoTargets and therapy*, vol. 8, pp. 2015–2022, 2015.

[12] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in ct images", in *Computer and Robot Vision (CRV), 2015 12th Conference on*, IEEE, 2015, pp. 133–138.

[13] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification", in *International Conference on Information Processing in Medical Imaging*, Springer, 2015, pp. 588–599.

[14] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, *et al.*, "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks", *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–6, 2016.

[15] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network", in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, IEEE, 2014, pp. 844–848.

[16] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification", *IEEE transactions on medical imaging*, vol. 32, no. 4, pp. 797–808, 2013.

[17] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, Y. Wang, M. J. Fulham, and M. Chen, "Large margin local estimate with applications to medical image classification", *IEEE transactions on medical imaging*, vol. 34, no. 6, pp. 1362–1377, 2015.

[18] X. Pei, "Emphysema classification using convolutional neural networks", in *International Conference on Intelligent Robotics and Applications*, Springer, 2015, pp. 455–461.

[19] G. Bussone and L. Mouthon, "Interstitial lung disease in systemic sclerosis", *Autoimmunity reviews*, vol. 10, no. 5, pp. 248–255, 2011.

[20] M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, N. Paragios, and I. Kokkinos, "Sub-cortical brain structure segmentation using f-cnn's", in *13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2016.

[21] M. Sharma and P. Garg, "Computational approaches for enzyme functional class prediction: A review", *Current Proteomics*, vol. 11, no. 1, pp. 17–22, 2014.

[22] S. K. Yadav and A. K. Tiwari, "Classification of enzymes using machine learning based approaches: A review", *Machine Learning and Applications: An International Journal (MLAIJ)*, vol. 2, no. 3/4, 2015.

[23] M. Spencer, J. Eickholt, and J. Cheng, "A deep learning network approach to ab initio protein secondary structure prediction", *IEEE/ACM Trans. on Computational Biology and Bioinformatics (TCBB)*, vol. 12, no. 1, pp. 103–112, 2015.

[24] Y. Li and T. Shibuya, "Malphite: A convolutional neural network and ensemble learning based protein secondary structure predictor", in *IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, 2015, pp. 1260–1266.

[25] Z. Lin, J. Lanchantin, and Y. Qi, "Must-cnn: A multilayer shift-and-stitch deep convolutional architecture for sequence-based protein structure prediction", in *30th AAAI Conference on Artificial Intelligence*, 2016.

[26] E. I. Zacharaki, "Prediction of protein function using a deep convolutional neural network ensemble", *PeerJ Computer Science*, submitted, 2016.

[27] G. A. Bermejo, G. M. Clore, and C. D. Schwieters, “Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of nmr protein structures”, *Protein Science*, vol. 21, no. 12, pp. 1824–1836, 2012.

[28] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines”, *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[29] A. Amidi, S. Amidi, D. Vlachakis, N. Paragios, and E. I. Zacharaki, “A machine learning methodology for enzyme functional classification combining structural and protein sequence descriptors”, in *Bioinformatics and Biomedical Engineering*, Springer, 2016, pp. 728–738.

Chapter 7: Curriculum Vitae

I. EDUCATION

Nov'99 – May'04 **Ph. D., Electrical and Computer Engineering**

School of Electrical and Computer Engineering (ECE), National Technical University of Athens (NTUA), Greece. *Dissertation:* Development of algorithms for medical image registration and computer simulation of tumor evolution aiming at supporting clinical decisions in radiooncology.

Sep'94 – Jun'99 **Diploma in Electrical and Computer Engineering**

ECE, NTUA, Greece. *Diploma Thesis:* Simulation of tumor progress and response to radiotherapeutic schemes using Monte Carlo and control theory methods (7,68/10).

II. ACADEMIC ACTIVITY

Research Associate:

09/2015 – : Center for Visual Computing, Department of Applied Mathematics, CentraleSupélec, France, and Equipe GALEN, INRIA Saclay, Orsay, France

04/2009 – 08/2015 : University of Patras (UPatras), Greece

Department of Computer Engineering and Informatics (04/2012 – 08/2015)

Department of Medical Physics, School of Medicine (04/2009 – 03/2012)

02/2005 – 03/2009: Section of Biomedical Image Analysis (SBIA), Department of Radiology, School of Medicine, University of Pennsylvania (UPenn), USA.

Teaching Experience

After PhD:

- Lecturer of the course "Foundations of Machine Learning (theory and laboratory)" of the Master's Program "Data Sciences Business Analytics", CentraleSupélec, Paris Saclay for the (winter semester 2016/2017).
- Lecturer in T.E.I. of Western Greece (former T.E.I. of Patras), for the acad. years 2009-2015 (on the average 12h/week). *Laboratory courses*: Electrical Circuits I, Electrical Circuits II, Electrical Measurements, Computer Programming (Pascal, QBasic, Fortran), Introduction to operating systems and software.

- Lecturer of the course «Biosignal Processing and Medical Imaging» of the Master's Program "Informatics in Life Sciences (direction: Medical Informatics)" UPatras (summer semester 2010/2011, winter semester 2013/2014).
- Lecturer of the course «Digital signal and image processing for the combined study of the anatomy and function of the human brain» of the Master's Program "Informatics in Life Sciences (direction: Neuroinformatics)" UPatras (summer semesters 2009/2010, 2010/2011 and 2011/2012).
- Ass. Professor under 407/1980 contract, Dep.of Computer Engineering and Informatics, UPatras. *Course*: Computer Vision and Graphics (winter semesters 2009/2010 and 2010/2011).
- Lecturer of the course «Basics of Biomedical Instrumentation» of the Master's Program "Basic Medical Sciences", UPatras (May 2009).

During PhD:

- Lecturer, College of Pedagogic and Technological Education (ASPAITE), Athens, Oct'01 – Jun'04. Courses: Electrical Circuits (theory and laboratory), Automatic Control Systems, Electrical Measurements, Digital Systems Laboratory.
- Teaching Assistant, Biomedical Engineering Lab, School of Electrical and Computer Engineering, NTUA, spring semester (01/09/2001 - 31/08/2003).

Student Co-supervision

Phd students:

- Evgenios Kornaropoulos (2015-, 75%), CentraleSupélec, Advisor: Nikos Paragios.
  Topic: Deformable image registration and diffusion estimation in MRI.
- Guillaume Chassagnon (2016-, 25%), CentraleSupélec/Cochin, Advisor: Nikos Paragios.
  Topic: Development of new quantitative imaging biomarkers for obstructive and interstitial lung diseases.
- Angeliki Skoura (2012-2014, 40%), University of Patras, Advisor: V. Megalooikonomou.
  Topic: Data mining system for tree and network structures in medical images.
- Vasileios Kanas (2012-2016, 70%), University of Patras, Advisor: K. Sgarbas.
  Topic: Development and implementation of machine learning techniques for brain data analysis.
- Evangelia Pippa (2014-, 60%), University of Patras, Advisor: V. Megalooikonomou.
  Topic: Data mining techniques for multi-dimensional time series
- Alexia Tzalavra (2013-, 50%), National Technical University of Athens, Advisor: K. Nikita.
  Topic: Disease classification in spatio-temporal data.

Master students:

- Afshine Amidi (2015-, 100%), CentraleSupélec, Advisor: Nikos Paragios.
  Topic: Fast similarity searches over large databases.
- Shervine Amidi (2015-, 100%), CentraleSupélec, Advisor: Nikos Paragios.
  Topic: Fast similarity searches over large databases.

III. WORK AND RESEARCH EXPERIENCE

Senior Research Associate in the *Center for Visual Computing, Department of Applied Mathematics*, CentraleSupélec, France, and Equipe GALEN, INRIA Saclay, Orsay, France (Supervising Professor: Nikos Paragios).

- 1/9/2011-31/8/2016 1/4/2012-30/9/2015, ERC-STG-259112 "Discrete bioimaging perception for longitudinal organ modelling and computer-aided diagnosiS (Diocles)"

Consultanting services in research programs of *Section of Biomedical Image Analysis*, UPenn, USA (01/04/2009-31/12/2011 and 01/08/2012-01/03/2013).

Research Associate in the *Multidimensional Data Analysis and Knowledge Management Laboratory,* Department of Computer Engineering and Informatics, UPatras, Greece (Supervising Professor: Vasileios Megalooikonomou).

- 1/4/2012-30/9/2015, Thalis–UPatras "Mining Biomedical Data And Images: Development Of Algorithms And Applications (BIOMEDMINE)",
- 1/7/2012-31/12/2013, FP7-ICT-2011-7 "Advanced multi-paRametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and Related brain disorders (ARMOR)"

Research Associate and main (single) investigator in *Biosignal Processing Group*, School of Medicine, UPatras (Supervising Professor: Anastasios Bezerianos)

- 4/1/2009-4/1/2012, FP7-PEOPLE-IRG-2008, Proposal N°239247 "Detection of Brain Abnormality (DeBrA)"

Research Associate in the *Section of Biomedical Image Analysis* (Feb'05 – Mar'09), Department of Radiology, University of Pennsylvania School of Medicine, Philadelphia, USA (Supervising Professor: Christos Davatzikos).

- 1/2/2005, NIH (R01 NS042645) "Modeling/estimating deformation in tumor patients",
- 01/10/2004-30/9/2012, NIH/NIA (R01) "Early Markers of Alzheimer's Disease in Selected Baltimore Longitudinal Study of Aging Participants",
- 01/07/2003-30/04/2010 NIH (N01) "Accord-Mind MRI Sub Study",
- 30/9/2005- 30/06/2008 (R01 AG010785) "Age, Lead Exposure and Neurobehavioral Decline"

Research Engineer in ECE, NTUA (Nov'99 – Sep'04) in the following research projects funded by the Institute of Communication and Computer Systems (ICCS)/NTUA or the Greek General Secretariat of Research and Technology (GSRT).

- 1/5/2000-30/4/2002, ICCS/ARCHIMEDES "Development of a 4D simulation model of in vivo tumor growth and response to radiation therapy", Principal Investigator: Myrsini Makropoulou
- 1/1/2000-30/6/2001, GSRT/PENED 99ED124 "Experienced system supporting diagnosis of psychopathological entities and assessment of therapy interventions", Principal Investigator: Haralambos Papageorgiou
- 1/5/1999-31/4/2001, GSRT/EPET2 "Development of Virtual Simulation and Treatment Planning in Radiation Oncology (GALENOS)", Principal Investigator: Konstadina Nikita

IV. PROFESSIONAL ACTIVITIES

Editorial Board Member

- International Journal of Radiology (since 2014)

- Dataset Papers in Science (Radiology) (since 2013)

Associate Editor

- Medical Physics (guest editor, 2016)

Occasional Reviewer

- Medical Image Analysis (since 2016)
- IEEE Transactions on Medical Imaging (since 2006)
- Neuroimage (since 2009)
- International Journal of Image and Graphics (since 2009)
- IEEE Transactions on Information Technology in BioMedicine (from 2010 to 2013)
- Artificial Intelligence in Medicine (since 2010)
- International Journal of Computer Assisted Radiology and Surgery (since 2010)
- IEEE Transactions on Biomedical Engineering (since 2011)
- Magnetic Resonance in Medicine (since 2011)
- International Journal for Numerical Methods in Biomedical Engineering (since 2011)
- Medical Engineering & Physics (από 2012)
- Biomedical Signal Processing and Control (since 2012)
- Computer Methods and Programs in Biomedicine (since 2013)
- International Journal on Artificial Intelligence Tools (since 2013)
- Kuwait Journal of Science and Engineering (since 2013)
- Connection Science (since 2013)
- Journal of Biomedical and Health Informatics (since 2014)
- American Journal of Neuroradiology (since 2014)
- Expert Systems with Applications (since 2015)
- International Journal of Biomedical Imaging (since 2015)

Session chair in conferences

- 6th IEEE International Symposium on Biomedical Imaging (ISBI 2009)
- 10th International Workshop on Biomedical Engineering, (Bioeng 2011)

Organizing committee member

- 5th Int. Summer School on Emerging Technologies in Biomedicine, "High Throughput Communication between Brain and Machines", Patras, Greece, Sep 26 - Oct 1, 2010.
- 1st Int. Advanced Research Workshop on In Silico Oncology: Advances and Challenges, Sparta, Greece, Sep 9-11, 2004.

Presenter in scientific conferences/meetings

- 10 oral talks (2 times invited speaker) and 3 poster presentations in international scientific conferences
- Invited speaker in universities (University of Pennsylvania 11/2004, Ecole Centrale de Paris 11/2009) and in companies (iCAD Inc. Beavercreek OH 45431 USA, 11/2007).

Participant in scientific seminars/summer schools

- Seminar (36 hours duration) by Hewlett Packard – IFS Hellas AE on «Parallel programming of Supercomputers with the use of MPI».
- Summer school (12 days duration) by the Advanced Study Institute (ASI) on «Multisensor and Sensor Data Fusion», organized by NATO, June 25 -Jul 7, 2000, Pitlochry, Scotland.

- Symposium on «Consciousness and its Measures», organized by COST (European Cooperation in the field of Scientific and Technical Research), 29/11/2009-1/12/2009, Limassol, Cyprus.
- Seminar and Practice on the Analysis of EEG and MEG entitled «Foundation Themes for Advanced EEG/MEG Source Analysis: Theory and Demonstrations via Hands-on Examples», 2/12/2009-4/12/2009, Nicosia, Cyprus.
- 5th Int. Summer School on Emerging Technologies in Biomedicine, "High Throughput Communication between Brain and Machines", Sep 26 - Oct 1, 2010, Patras, Greece.

V. AWARDS- SCHOLARSHIPS

- Three year scholarship on Biomedical Engineering from Greek State Scholarship's Foundation (IKY) for Phd position in ECE, NTUA.
- Awards from Thomaidio Foundation for 4 scientific publications in the years 2000, 2002, 2003.
- Three year Marie Curie IRG fellowship (FP7-PEOPLE-IRG-2008) for postdoctoral research in UPatras.

VI. PUBLICATIONS

Dissertations

A1. ***Doctoral Thesis:*** "Development of algorithms for medical image registration and computer simulation of tumor behavior aiming at supporting clinical decisions in radiooncology (in greek)", ECE, NTUA, 2004.

A2. ***Diploma Thesis:*** " Simulation of tumor progress and response to radiotherapeutic schemes using Monte Carlo and control theory methods (in greek)", ECE, NTUA, 1999.

Journal Papers

J1. E. Pippa, E.I. Zacharaki, M. Koutroumanidis, V. Megalooikonomou, "Data fusion for paroxysmal events' classification from EEG," *Journal of Neuroscience Methods*, in press (Impact Factor = 2.053).

J2. E. Pippa, V.G. Kanas, E.I. Zacharaki, V. Tsirka, M. Koutroumanidis, V. Megalooikonomou, "EEG-based classification of epileptic and non-epileptic events using multi-array decomposition," *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, in press, 2016.

J3. N. Paragios, E. Ferrante, B. Glocker, N. Komodakis, S. Parisot, E.I. Zacharaki, "(Hyper)-Graphical Models in Biomedical Image Analysis," *Medical Image Analysis*, vol. 33, pp. 102–106, 2016. (Impact Factor = 4.565

J4. S. Alchatzidis, A. Sotiras, E.I. Zacharaki, N. Paragios, "A discrete MRF framework for integrated multi-atlas registration and segmentation," *International Journal of Computer Vision*, pp 1-13, 2016. (Impact Factor = 4.270)

J5. E.I. Zacharaki, I. Mporas, K. Garganis, V. Megalooikonomou, "Spike pattern recognition by supervised classification in low dimensional embedding space," *Brain Informatics*, vol. 3, no. 2, pp 73-83, 2016. (Impact Factor =1.959)

J6. E. Pippa, E.I. Zacharaki, I. Mporas, V. Tsirka, M. Richardson, M. Koutroumanidis and V. Megalooikonomou, " Improving classification of epileptic and non-epileptic EEG events by feature selection," *Neurocomputing*, vol.171, pp. 576-585, 2016. (Impact Factor = 2.005)

J7. V.G. Kanas, E.I. Zacharaki, C. Davatzikos, K.N. Sgarbas, V. Megalooikonomou, "A low cost approach for neoplastic and healthy brain tissue segmentation based on intensity modeling and 3D random walker," *Biomedical Signal Processing and Control*, vol.22, pp.19-30, 2015. (Impact Factor = 1.532)

J8. I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, M. Richardson, V. Megalooikonomou, "Seizure Detection using EEG and ECG Signals for Computer-based Monitoring, Analysis and Management of Epileptic Patients," *Expert Systems With Applications*, vol.40, pp.3227-3222, 2015. (Impact Factor = 1.965)

J9. G. Erus, E.I. Zacharaki, C. Davatzikos, "Individualized Statistical Learning from Medical Image Databases: Application to Identification of Brain Lesions, " *Medical Image Analysis*, vol. 18, pp.542-554, 2014. (Impact Factor = 4.565)

J10. E.I. Zacharaki, N. Morita, P. Bhatt, D.M. O'Rourke, E.R. Melhem, C. Davatzikos, "Survival analysis of patients with high grade gliomas based on data mining of imaging variables," *American Journal of Neuroradiology*, vol. 33, no. 6, pp.1065-71, 2012. (Impact Factor = 3.675)

J11. E.I. Zacharaki, A. Bezerianos, "Abnormality segmentation in brain images via distributed estimation," *IEEE Transaction on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 330-338, 2012. (Impact Factor = 2.072)

J12. E.I. Zacharaki, V.G. Kanas, C. Davatzikos, "Investigating machine learning techniques for MRI-based classification of brain neoplasms," *International Journal of Computer Assisted Radiology and Surgery*, vol.6, pp.821–828, 2011. (Impact Factor = 1.659)

J13. E.I. Zacharaki, S. Wang, S. Chawla, D.S. Yoo, R. Wolf, E.R. Melhem, C. Davatzikos, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine*, vol. 62, pp. 1609-18, 2009. (Impact Factor = 3.398)

J14. Y. Zheng, S. Englander, S. Baloch, E.I. Zacharaki, Y. Fan, M.D. Schnall, D. Shen, "STEP: Spatio-Temporal Enhancement Pattern for MR-based Breast Tumor Diagnosis," *Medical Physics*, vol. 36, no. 7, pp. 3192-3204, 2009. (Impact Factor = 3.012)

J15. E.I. Zacharaki, C.S. Hogea, D. Shen, G. Biros, C. Davatzikos, "Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth," *Neuroimage*, vol.46, no. 3, pp. 762-774, 2009. (Impact Factor = 6.132)

J16. R. Verma, E.I. Zacharaki, Y. Ou, H. Cai, S. Chawla, R. Wolf, S.-K. Lee, E.R. Melhem, C. Davatzikos, "Multi-parametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images," *Academic Radiology*, vol.15, pp. 966–977, 2008. (Impact Factor = 2.077)

J17. E.I. Zacharaki, D. Shen, S.-K. Lee and C. Davatzikos, "ORBIT: A Multiresolution Framework for Deformable Registration of Brain Tumor Images," *IEEE Trans. Medical Imaging*, vol. 27, no. 8, pp. 1003-1017, 2008. (Impact Factor = 3.799)

J18. E.I. Zacharaki, C.S. Hogea, G. Biros and C. Davatzikos, "A comparative study of biomechanical simulators in deformable registration of brain tumor images," *IEEE Trans. Biomedical Engineering*, vol.55, no.3, pp1233-1236, 2008. (Impact Factor = 2.233)

J19. A. Mohamed, E.I. Zacharaki, D. Shen, C. Davatzikos, "Deformable Registration of Brain Tumor Images via a Statistical Model of Tumor-Induced Deformation," *Medical Image Analysis (Special Issue)*, vol.10, pp.752-763, 2006 (invited paper). (Impact Factor = 4.565)

J20. E. Ntasis, M. Gletsos, N.A. Mouravliansky, E.I. Zacharaki, C. Vasios, S. Golemati, T.A. Maniatis and K.S. Nikita, "Telematics Enabled Virtual Simulation System for Radiation Treatment Planning," *Computers in Biology and Medicine*, vol. 35, no. 9, pp.765-781, 2005. (Impact Factor = 1.240)

J21. E.I. Zacharaki, G.S. Stamatakos, K.S. Nikita and N.K. Uzunoglu, "Simulating growth dynamics and radiation response of avascular tumour spheroids – Model validation in the case of an EMT6/Ro multicellular spheroid," *Computer Methods and Programs in Biomedicine*, vol.76, no.3, pp.193-206, 2004. (Impact Factor = 1.897)

J22. E.I. Zacharaki, G.K. Matsopoulos, P.A. Asvestas, K.S. Nikita, K. Gröndahl, H.-G. Gröndahl, "A digital subtraction radiography scheme based on automatic multiresolution registration," *Dentomaxillofacial Radiology*, vol.33, no.6, pp.379-390, 2004. (Impact Factor = 1.271)

J23. G.S. Stamatakos, D.D. Dionysiou, E.I. Zacharaki, N.A. Mouravliansky, K.S. Nikita, and N. Uzunoglu, "In Silico Radiation Oncology: Combining Novel Simulation Algorithms with Current Visualization Techniques," *Proceedings of the IEEE*, vol.90, no.11, pp.1764-1777, 2002 (invited paper). (Impact Factor = 5.466)

J24. G.S. Stamatakos, E.I. Zacharaki, M.I. Makropoulou, N.A. Mouravliansky, A. Marsh, K.S. Nikita, N. Uzunoglu, "Modeling Tumor Growth and Irradiation Response in Vitro: a Combination of High Performance Computing and Web Based Technologies Including VRML Visualization," *IEEE Trans. Inform. Technol. Biomed, vol.5, no.4, pp.279-289*, 2001. (Impact Factor = 2.072)

J25. G.S. Stamatakos, E.I. Zacharaki, N.K. Uzunoglu, K.S. Nikita, "Tumor Growth and Response to Irradiation in Vitro: a Technologically Advanced Simulation Model," *Int. J. Oncol. Biol. Phys*. vol.51, Suppl.1, pp.240-241, 2001. (Impact Factor = 4.258)

J26. G.S. Stamatakos, E.I. Zacharaki, M. Makropoulou, N. Mouravliansky, K.S. Nikita, N. Uzunoglu, "Tumour growth in vitro and tumour response to irradiation schemes: a simulation model and virtual reality visualization," *Radiotherapy and Oncology*, vol.56, Suppl.1, pp.179-180, Elsevier, 2000. (Impact Factor = 4.857)

Book chapters

B1. V. Megalooikonomou, D. Triantafyllopoulos, E.I. Zacharaki, I. Mporas, "*Offline Analysis Server and Offline algorithms*," N.S. Voros, C.P. Antonopoulos (eds.), Cyberphysical Systems for Epilepsy and Related Brain Disorders, 2015, pp.239-254, Springer International Publishing Switzerland (ISBN 978-3-319-20048-4).

B2. V. Megalooikonomou, D. Triantafyllopoulos, E.I. Zacharaki, I. Mporas, "*DSMS and Online Algorithms*," N.S. Voros, C.P. Antonopoulos (eds.), Cyberphysical Systems for Epilepsy and Related Brain Disorders, 2015, pp.271-279, Springer International Publishing Switzerland (ISBN 978-3-319-20048-4).

Conference Proceedings

High rank conferences and conferences publishing chapters in books

C1. A. Amidi, S. Amidi, D. Vlachakis, N. Paragios, E.I. Zacharaki, "A machine learning methodology for enzyme functional classification combining structural and protein

sequence descriptors," *Bioinformatics and Biomedical Engineering*, 2016, pp.728-738, Springer International Publishing.

C2. A. Tzalavra, K. Dalakleidi, E.I. Zacharaki, N. Tsiaparas, F. Constantinidis, K.S. Nikita, Comparison of Multi-resolution Analysis Patterns for Texture Classification of Breast Tumors Based On DCE-MRI, *7th Int. Workshop on Machine Learning in Medical Imaging (MICCAI workshop)*, October 17-18, 2016, Athens, Greece.

C3. E.N. Kornaropoulos, E.I. Zacharaki, P. Zerbib, C. Lin, A. Rahmouni, N. Paragios, "Deformable group-wise registration using a physiological model: application to diffusion-weighted MRI," *IEEE Int. Conf. on Image Processing (ICIP)*, September 25-28, 2016, Phoenix, USA.

C4. E.N. Kornaropoulos, E.I. Zacharaki, P. Zerbib, C. Lin, A. Rahmouni, N. Paragios, "Optimal estimation of diffusion in DW-MRI by high-order MRF-based joint deformable registration and diffusion modeling," *7th Int. Workshop on Biomedical Image Registration (WBIR)*, in conjunction with *CVPR*, July 1, 2016, Las Vegas, USA.

C5. V.G. Kanas, E.I. Zacharaki, E. Pippa, V. Tsirka, M. Koutroumanidis, V. Megalooikonomou, "Classification of Epileptic and Non-epileptic Events using Tensor Decomposition," *IEEE 15th Int. Conf. on Bioinformatics and Bioengineering (BIBE 2015)*, Nov. 2-4, 2015, Belgrade, Serbia.

C6. I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, V. Megalooikonomou, "Online Seizure Detection from EEG and ECG signals for Monitoring of Epileptic Patients," *Artificial Intelligence: Methods and Applications, Lecture Notes in Computer Science*, vol. 8445, 2014, pp. 442-447, Springer.

C7. A.G. Tzalavra, E.I. Zacharaki, N.N.Tsiaparas, F. Constantinidis, K.S.Nikita, "A Multiresolution Analysis Framework For Breast Tumor Classification Based On DCE-MRI," *IEEE Int. Conf. on Imaging Systems & Techniques (IST 2014)*, October 14-17, 2014, Santorini, Greece.

C8. E.I. Zacharaki, K. Garganis, I. Mporas, V. Megalooikonomou, "Spike detection in EEG by LPP and SVM ", *IEEE EMBS Int. Conf. on Biomedical and Health Informatics (BHI'2014)*, pp.668-671, 1-4 June, 2014, Valencia, Spain.

C9. Mporas, P. Korvesis, E.I. Zacharaki, V. Megalooikonomou "Sleep Spindle Detection in EEG Signals combining HMMs and SVMs,*" Engineering Applications of Neural Networks (EANN)*, vol. 384, 2013, pp 138-145, Springer.

C10. E.I. Zacharaki, E. Pippa, A. Koupparis, V. Kokkinos, G. Kostopoulos, V. Megalooikonomou, "One-class classification of temporal EEG patterns for K-complex extraction," *35th Int. Conf.of the IEEE Engineering in Medicine and Biology Society (EMBC '13)*, July 3-7, 2013, Osaka, Japan.

C11. E.I. Zacharaki, A. Skoura, L. An, D. Smith, V. Megalooikonomou, "Using an atlas-based approach in the analysis of gene expression maps obtained by voxelation," *Artificial Intelligence Applications and Innovations*, *IFIP Advances in Information and Communication Technology,* vol. 382, 2012, pp. 566-575, Springer.

C12. V.G. Kanas, E.I. Zacharaki, E. Dermatas, A. Bezerianos, K.Sgarbas, C. Davatzikos, "Combining outlier detection with random walker for automatic brain tumor segmentation," *Artificial Intelligence Applications and Innovations*, *IFIP Advances in Information and Communication Technology*, vol. 382 , 2012, pp. 26-35, Springer.

C13. E.I. Zacharaki, G. Erus, A. Bezerianos, C. Davatzikos, "Fuzzy multi-channel clustering with individualized spatial priors for segmenting brain lesions and infarcts," *Artificial*

Intelligence Applications and Innovations, *IFIP Advances in Information and Communication Technology*, vol. 382, 2012, pp 76-85, Springer.

C14. E.I. Zacharaki, A. Skoura, L. An, D.J. Smith, S. H. Faro, V. Megalooikonomou, "Combining gene expression and function in a spatially localized approach," *2012 IEEE Int. Conf.on Bioinformatics and Biomedicine (BIBM)*, October 4-7, 2012, Philadelphia, PA, USA.

C15. K. Dimitrakopoulou, G. Dimitrakopoulos, E.I. Zacharaki, I.A. Maraziotis, K. Sgarbas, A. Bezerianos, "Revealing the dynamic modularity of composite biological networks in breast cancer treatment," *34rd Int. Conf.of the IEEE Engineering in Medicine and Biology Society (EMBC '12)*, Aug 28- Sep 1, 2012, San Diego, USA.

C16. S. Kadoury, G. Erus, E.I. Zacharaki, N. Paragios, C. Davatzikos, "Manifold-constrained embeddings for the detection of white matter lesions in brain MRI," *9th IEEE International Symposium on Biomedical Imaging (ISBI 2012)*, Barcelona, 2-5 May 2012.

C17. C. Davatzikos, E.I. Zacharaki, A. Gooya, V. Clark, "Multi-parametric Analysis and Registration of Brain Tumors: Constructing Statistical Atlases And Diagnostic Tools of Predictive Value," *33rd Int. Conf.of the IEEE Engineering in Medicine and Biology Society (EMBC '11)*, Boston, MA, Aug 30- Sept 3, 2011, pp. 6979-81 (invited paper).

C18. G. Erus, E.I. Zacharaki, R.N. Bryan, C. Davatzikos, "Learning high-dimensional image statistics for abnormality detection on medical images," *IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA10)*, San Francisco, CA, June 14, 2010.

C19. E.I. Zacharaki, S. Wang, S. Chawla, D.S. Yoo, R. Wolf, E.R. Melhem, C. Davatzikos, "MRI-based classification of brain tumor type and grade using SVM-RFE," *6th IEEE International Symposium on Biomedical Imaging (ISBI 2009)*, Boston, Massachusetts, USA, June 28- July 1, 2009.

C20. E.I. Zacharaki, C.S. Hogea, D. Shen, G. Biros, C. Davatzikos, "Parallel optimization of tumor model parameters for fast registration of brain tumor images," *SPIE Medical Imaging 2008: Image Processing*, ed. J.M. Reinhardt, J.P.W. Pluim, vol. 6914, Issue 1, article 0K, pp. 1-10, 2008.

C21. E.I. Zacharaki, S. Kanterakis, R.N. Bryan, C. Davatzikos, "Measuring brain lesion progression with a supervised tissue classification system," *Medical Image Computing and Computer Assisted Intervention, Lecture Notes in Computer Science*, vol. 5241, 2008, pp. 620-627, Springer.

C22. N. Batmanghelich, X. Wu, E.I. Zacharaki, C.E. Markowitz, C. Davatzikos, R. Verma, "Multiparametric Tissue Abnormality Characterization using Manifold Regularization," *SPIE Medical Imaging 2008: Computer-Aided Diagnosis*, ed. M.L. Giger, N. Karssemeijer, vol. 6915, Issue 1, article 16, pp. 1-6, 2008.

C23. E.I. Zacharaki, R. Verma, S. Chawla, E.R. Melhem, R. Wolf, C. Davatzikos, "Towards predicting neoplastic recurrence with multi-parametric MR," *ISMRM 16th Annual Meeting and Exhibition*, May 3-6, 2008, Toronto, Canada.

C24. E.I. Zacharaki, D. Shen, A. Mohamed, C. Davatzikos, "Registration of brain images with tumors: Towards the construction of statistical atlases for therapy planning," *3rd IEEE International Symposium on Biomedical Imaging (ISBI 2006)*, Arlington, Virginia, USA, April 6-9, 2006.

C25. E.I. Zacharaki, G.S. Stamatakos, N.K. Uzunoglu, K.S. Nikita, "Stochastic modeling and validation of growth saturation and radiotherapeutic response of multicellular tumor

spheroids," *26th Annual Int. Conf.of the IEEE Engineering in Medicine and Biology Society*, pp.3039-3042, San Francisco, California, September 1-5, 2004

C26. .E.I. Zacharaki, G.K. Matsopoulos, K.S. Nikita, G.S. Stamatakos, "An application of multimodal image registration and fusion in a 3D tumor simulation model," *25th Annual Int. Conf.of the IEEE Engineering in Medicine and Biology Society*, pp.686-689, Cancun, Mexiko, September 17-21, 2003.

C27. E.I. Zacharaki, G.K. Matsopoulos, K.S. Nikita, G.S. Stamatakos, "3D image registration and fusion tools in simulating tumor evolution," *3rd IASTED Int. Conf.on Visualization, Imaging, and Image Processing*, pp.307-311, Benalmádena, Spain, September 8-10, 2003.

C28. N. Mouravliansky, E.I. Zacharaki, P. Asvestas, G. Matsopoulos, K. Delibasis, K.S. Nikita, "Image registration based on lifting process and genetic optimization: an application to dental imaging," *3rd IASTED Int. Conf. on Visualization, Imaging, and Image Processing*, pp.312-316, Benalmádena, Spain, September 8-10, 2003.

C29. N. Mouravliansky, G. Matsopoulos, K. Delibasis, E.I. Zacharaki, P. Asvestas, K.S. Nikita, "Image Registration Based on Lifting Process: an Application to Dental Imaging," *2nd European Medical & Biological Engineering Conference*, pp.852-853, Vienna, December4-8, 2002.

C30. E.I. Zacharaki, P. Asvestas, G.K. Matsopoulos, K.K. Delibasis, and K.S. Nikita, "An automatic registration scheme based on similarity measures: an application to dental imaging, " *23rd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol.3, pp.2429-2432, Istanbul, Turkey, October 25-28, 2001.

C31. G. Stamatakos, E. Zacharaki, N. Mouravliansky, K. Delibassis, K. Nikita, N. Uzunoglu and A.Marsh, "Using Web technologies and meta-computing to visualize a simplified simulation model of tumor growth in vitro," *IEEE EMBS Int. Conf. on Information Technology Applications in Biomedicine, ITIS-ITAB '99*, pp.31-32, 1999.

Ordinary conferences/workshops

C32. E. Pippa, E. I. Zacharaki, I. Mporas,V. Tsirka, M. Richardson, M. Koutroumanidis, V. Megalooikonomou, "Classification of Epileptic and Non-Epileptic EEG Events", *4th Int. Conf. on Wireless Mobile Communication and Healthcare (MOBIHEALTH 2014),* November 3–5, 2014, Athens, Greece.

C33. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, V. Megalooikonomou, "Evaluation of Time and Frequency Domain Features for Seizure Detection from Combined EEG and ECG signals," *7th Int. Conf. on PErvasive Technologies Related to Assistive Environments (PETRA 2014).*

C34. Charisi, F.D. Malliaros, E.I. Zacharaki, V. Megalooikonomou, "Multiresolution Similarity Search in Time Series Data: An Application to EEG Signals," *6th Int. Conf. on PErvasive Technologies Related to Assistive Environments (PETRA 2013)*, May 29-31, 2013, Rhodes Island, Greece.

C35. E.I. Zacharaki, A. Bezerianos, "Segmentation of pathology by statistical modeling and distributed estimation," *10th International Workshop on Biomedical Engineering*, October 5-7, 2011, Kos, Greece.

C36. N. Morita, M. Harada, E. Zacharaki, P.Bhatt, S. Chawla, E.R. Melhem, H.Nishitani, "Correlation between Diffusion Tensor and Perfusion Imaging in segmented enhancing lesion with high grade glioma," *Joint Annual Meeting ISMRM-ESMRMB*, Stockholm, Sweden, May 1-7, 2010

C37. S. Magnitsky, E.I. Zacharaki, R. Verma, R.M. Walton, J.H. Wolfe and H. Poptani, "Longitudinal Detection of Neuronal Stem Cells Labeled with Types of Iron Oxide Particles," *Joint Annual Meeting ISMRM-ESMRMB*, May 19-25, 2007.

C38. E.I. Zacharaki, G.S. Stamatakos, N.K. Uzunoglu, "Computer simulation of tumour spheroid behaviour as a platform for understanding cancer in silico," *1st International Advanced Research Workshop on In Silico Oncology: Advances and Challenges*, pp.54-55, Sparta, Greece, September 9-11, 2004 (invited speaker).

National journal and conference publications (in greek)

G1. K.S. Nikita, T.A. Maniatis, E. Ntasis, M. Gletsos, N. Mouravliansky, C. Vasios, E. Zacharaki, "The Virtual Simulation System «Galenos» (Το Πρόγραμμα Εικονικής Εξομοίωσης «Γαληνός»)," Oncological Review (*Ογκολογική Ενημέρωση*), vol.3, no.3, pp.180-186, 2001.

G2. E.I. Zacharaki, E. Pippa, A. Koupparis, G.K. Kostopoulos, V. Megalooikonomou, "Classification of EEG waveforms by spectral clustering," *5th Panhellenic Conference on Biomedical Technology*, pp. 93-94, Athens, April 4-6, 2013 (paper in english).

G3. E. Zacharaki, "Computer simulation of tumor growth and response to irradiation," *6thPanhellenic Conference on Radiotherapeutic Oncology*, pp.135-137, Paros, September 27-30, 2001 (invited speaker).

G4. K.S. Nikita, T.A. Maniatis, E. Ntasis, M. Gletsos, N. Mouravliansky, C. Vasios, E. Zacharaki, "The Virtual Simulation System «Galenos»," *10th Panhellenic Conference of Clinical Oncology*, Athens, March 29-31, 2001.

G5. G.S. Stamatakos, E. Zacharaki, N. Mouravliansky, M. Makropoulou, K. Nikita, N. Uzunoglu, "Simulation of in vitro tumor response to radiotherapeutic schemes," *2nd Panhellenic Conference on Biomedical Technology*, pp.136-141, Athens, November 5-6, 1999.

Submitted Papers (under review)

Z1. G. Chassagnon, C. Martin, P-R Burgel, D. Hubert, I. Fajac, N. Paragios, E.I. Zacharaki, P. Legmann, J. Coste, M-P Revel, "Structural abnormalities in the Cystic Fibrosis lung: an automated Computed Tomography score", *European Respiratory Journal*, submitted (Impact Factor = 8.332).

Z2. E.I. Zacharaki, "Prediction of protein function using a deep convolutional neural network ensemble" *PeerJ Computer Science,* under revision (Impact Factor = 2.1).

Z3. A. Amidi, S. Amidi, D. Vlachakis, N. Paragios, E.I. Zacharaki, "Automatic single- and multi-label enzymatic function prediction by machine learning," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, under revision. (Impact Factor = 1.609)

Z4. V.G. Kanas, E.I. Zacharaki, G.A. Thomas, P.O. Zinn, V. Megalooikonomou, R.R. Colen, "Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma," *Computer Methods and Programs in Biomedicine*, under revision. (Impact Factor = 1.897)

VII. LANGUAGES

- Greek (native language)
- English (*First Certificate in English* and 4.5 years residency in USA)

- German (*Grosses Deutsches Sprachdiplom* and 5 years residency in Germany)
- French (*Certificat de la Langue Française* and 1.5 years residency in France)